

## Benchmark all-electron ab initio quantum Monte Carlo calculations for small molecules

Norbert Nemec, Michael D. Towler, and R. J. Needs

Citation: *The Journal of Chemical Physics* **132**, 034111 (2010); doi: 10.1063/1.3288054

View online: <http://dx.doi.org/10.1063/1.3288054>

View Table of Contents: <http://aip.scitation.org/toc/jcp/132/3>

Published by the [American Institute of Physics](#)

---

### Articles you may be interested in

[Approaching chemical accuracy with quantum Monte Carlo](#)

*The Journal of Chemical Physics* **136**, 124116 (2012); 10.1063/1.3697846

[Calculating potential energy curves with fixed-node diffusion Monte Carlo: CO and N<sub>2</sub>](#)

*The Journal of Chemical Physics* **145**, 224308 (2016); 10.1063/1.4971378

[Influence of single particle orbital sets and configuration selection on multideterminant wavefunctions in quantum Monte Carlo](#)

*The Journal of Chemical Physics* **142**, 234103 (2015); 10.1063/1.4921984

[On the effectiveness of CCSD\(T\) complete basis set extrapolations for atomization energies](#)

*The Journal of Chemical Physics* **135**, 044102 (2011); 10.1063/1.3613639

[Energies of the first row atoms from quantum Monte Carlo](#)

*The Journal of Chemical Physics* **126**, 224110 (2007); 10.1063/1.2743972

[Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen](#)

*The Journal of Chemical Physics* **90**, 1007 (1998); 10.1063/1.456153

---

COMPLETELY

REDESIGNED!



PHYSICS  
TODAY

*Physics Today* Buyer's Guide  
Search with a purpose.

# Benchmark all-electron *ab initio* quantum Monte Carlo calculations for small molecules

Norbert Nemec,<sup>a)</sup> Michael D. Towler, and R. J. Needs

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

(Received 13 August 2009; accepted 15 December 2009; published online 21 January 2010)

We study the efficiency, precision and accuracy of all-electron variational and diffusion quantum Monte Carlo calculations using Slater basis sets. Starting from wave functions generated by Hartree–Fock and density functional theory, we describe an algorithm to enforce the electron–nucleus cusp condition by linear projection. For the 55 molecules in the G2 set, the diffusion quantum Monte Carlo calculations recovers an average of 95% of the correlation energy and reproduces bond energies to a mean absolute deviation of 3.2 kcal/mol. Comparing the individual total energies with essentially exact values, we investigate the error cancellation in atomization and chemical reaction path energies, giving additional insight into the sizes of nodal surface errors. © 2010 American Institute of Physics. [doi:10.1063/1.3288054]

## I. INTRODUCTION

*Ab initio* variational Monte Carlo (VMC) and diffusion Monte Carlo (DMC) methods have been used successfully for systems containing hundreds and sometimes thousands of electrons.<sup>1</sup> Such calculations typically retrieve 95% or more of the correlation energy within the fixed-node approximation based on a single determinant. Testing the accuracy of VMC and DMC results against experiment and other theoretical methods plays an important role in the development of computer codes such as the CASINO package.<sup>1</sup>

As a method that directly competes for accuracy and efficiency with deterministic quantum chemical methods, benchmark results are of great interest. Total electronic energies of atoms lend themselves to direct comparison as highly accurate reference values are available. DMC calculations can recover 99% or more of the correlation energy for first row atoms using multideterminant and backflow wave functions.<sup>2</sup> Another well-known set of benchmark data are the atomization energies of the G2 set of molecules.<sup>3</sup> These energies have been reproduced in DMC to high accuracy using pseudopotentials.<sup>4</sup> Calculations of a selection of small molecules consisting of first-row atoms have proven all-electron DMC to be nearly as accurate as CCSD(T)/cc-pVTZ.<sup>5</sup>

In every case, the dominant deviations were attributed to the fixed node approximation made in the DMC method. DMC satisfies the variational principle and therefore the energies are too high when an approximate nodal surface is used. Comparing total energies for benchmarking purposes, DMC typically performs extremely well. For atomization and chemical reaction energies, however, error cancellation in the energy differences becomes essential. Unfortunately, the quality of the nodal surface turns out to depend intri-

cately on the chemical structure of each molecule or atom and the error cancellation in DMC is less efficient and predictable than in competing methods.

In this paper we present benchmark results for the aforementioned 55 molecules of the G2 set from all-electron DMC calculations using wave functions based on Slater-type orbitals (STOs). These results are directly comparable to pseudopotential-based calculations. Beyond this, however, the availability of highly accurate reference data for the total energies of the same set of molecules<sup>6</sup> allows a deeper analysis of the relative nodal surface errors in various chemical species and permits an intuitive visualization of error cancellation in chemical reaction energies.

## II. SLATER-TYPE ORBITALS

STOs were an important tool in quantum mechanics long before the availability of computational tools in physics and chemistry.<sup>9</sup> Inspired by the analytic solution of the hydrogen atom, STO basis sets were the first choice for a number of important approximate studies in the early years of quantum chemistry. With the arrival of computers, however, it turned out that Gaussian basis functions (GTOs) (Ref. 10) allow a far simpler efficient implementation due to the possibility of factorizing Gaussian functions in Cartesian coordinates and the simplicity of evaluating multicenter integrals. They became the standard in quantum chemistry to the point that chemists typically discuss “basis sets,” implicitly referring to a GTO basis.

While the relative merits of the GTO and STO representations of orbitals in quantum chemistry are still under debate,<sup>12</sup> attempts have been made to reduce the computational effort of working with STO basis sets,<sup>13–16</sup> leading to the development of several STO-based electronic structure codes.<sup>17–19</sup> Of these, the only code active development and available for general use is ADF,<sup>19</sup> offering a full state-of-the-art implementation of Hartree–Fock (HF) and density functional theory (DFT) electronic structure calculations for mol-

<sup>a)</sup>Electronic mail: nn245@cam.ac.uk.

ecules and, via its sister program BAND for periodic systems.

Quantum Monte Carlo (QMC) methods have very different computational requirements from conventional nonstochastic electronic structure methods. Without the need to perform analytic integrations, the usual advantages of GTO become irrelevant. Instead, the bulk of the computational cost lies in evaluating the trial wave function and its derivatives at arbitrary positions in space. A basis set that achieves the same precision with a more compact representation (fewer basis functions) will gain a clear advantage. Furthermore, STO wave functions allow an exact treatment of the Kato cusp condition at nuclei<sup>21</sup> and do not suffer from divergent local energies at large distance from a molecule. For these reasons, STO basis sets have often been used in QMC.

In the past, the main disadvantage of STO basis sets in QMC calculations has stemmed from the use of a trial wave function commonly generated from a preliminary HF or DFT calculation; almost all suitable mainstream local basis set electronic structure codes for finite systems—particularly in the quantum chemistry community—use GTO. To exploit the advantages of STO for QMC calculations, one could therefore use either a conversion step<sup>5</sup> or optimize the orbitals directly within VMC.<sup>22</sup> The advent of the ADF code has allowed the generation of HF or DFT orbitals directly in an STO basis, which can then be used in QMC calculations.

GTO basis sets have been available in the CASINO program for over a decade. We have now implemented the additional capability to evaluate orbitals expanded in a STO basis, allowing the use of trial wave functions generated by ADF in VMC and DMC calculations.<sup>23</sup> In the following, we present details of the cusp constraint used in the converter and demonstrate the precision achievable with the combination of ADF and CASINO.

Each molecular orbital is expanded in the STO basis

$$\Psi(\mathbf{r}) = \sum_{i=1}^{N_{\text{bas}}} c_i \psi_i(\mathbf{r} - \mathbf{R}_i), \quad (1)$$

with  $N_{\text{bas}}$  basis functions of the form

$$\psi_i(\mathbf{r}) = Y_l^{m_i}(\vartheta, \varphi) r^{l+n_i} e^{-\zeta_i r}, \quad (2)$$

with  $\zeta_i > 0$  and the  $Y_l^m$  are the Laplace spherical harmonics. Abandoning orthogonality in favor of simplicity, the Laguerre polynomials present in the analytic solutions for the hydrogen atom are replaced by  $r^n$ .

The centers  $R_i$  of the basis functions usually coincide with the positions  $R_I$  of the  $N_{\text{nuc}}$  pointlike nuclei  $I$  of charge  $Z_I$ . In principle, a single value of  $\zeta$  would allow the construction of a complete basis set by including sufficiently high orders  $n$ . In practice, however, using a small number of different  $\zeta$  values is a more efficient means of improving the precision of the basis set. A further improvement in precision can be made by including high-angular-momentum basis functions to improve the description of polarization. In this work, we used four general-purpose basis sets from the ADF package, in increasing size and precision: single- $\zeta$  (SZ), double- $\zeta$  (DZ), triple- $\zeta$ -polarized (TZP), and quadruple- $\zeta$ -fourfold-polarized (QZ4P).

### III. CUSP CONSTRAINT

The exact wave function of particles interacting via a Coulomb interaction fulfills the Kato cusp condition whenever two pointlike particles coalesce.<sup>21</sup> For pairs of electrons, this condition gives rise to dynamic correlations that can be very efficiently represented by a Jastrow factor.<sup>26–29</sup> In all-electron calculations, each single-electron orbital  $\Psi$  should fulfill the cusp condition

$$\left\langle \frac{d}{dr} \Psi(\mathbf{r}) \right\rangle_{\Omega} \Big|_{\mathbf{r}=\mathbf{R}_I} = -Z_I \Psi(\mathbf{R}_I), \quad (3)$$

in the vicinity of each pointlike nucleus  $I$  of charge  $Z_I$  at position  $\mathbf{R}_I$ , where  $\langle \cdot \rangle_{\Omega}$  denotes the spherical average around the nucleus.

In methods such as DFT or HF, having the exact cusp at the nucleus is less important than the overall quality of the wave function, so smooth Gaussian functions can be used to represent the wave function. QMC, on the other hand, is based on evaluating the local energy which diverges at coalescences if the cusp condition is not exactly fulfilled.<sup>7</sup> When using GTO-based trial wave functions in QMC, the cusp condition is typically enforced artificially, either by modifying the GTO basis functions<sup>30</sup> or by directly constructing a correction to the single-electron orbitals.<sup>7,31</sup>

In contrast to smooth GTO basis functions, STO-based orbitals [Eq. (2)] are able to fulfill the cusp condition [Eq. (3)] exactly, leading to one linear constraint per nucleus  $I$  on the coefficients  $c_i$  of any molecular orbital  $\Psi$ . These  $N_{\text{nuc}}$  constraints can be expressed as a single matrix equation

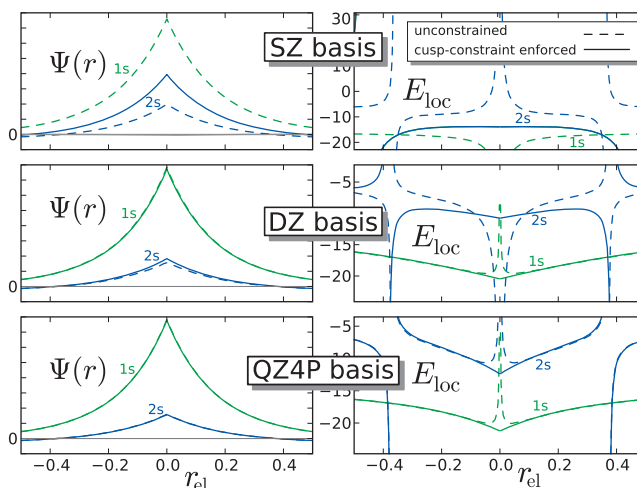


FIG. 1. Effect of the nuclear cusp constraint on orbitals  $\Psi(r)$  (left) and orbital local energies  $E_{\text{loc}} = \Psi^{-1} \mathcal{H} \Psi$  (right) for the carbon atom. Three different general-purpose basis sets from the ADF package were used (top to bottom with increasing size and precision). For atoms, only s orbitals require cusp correction and are here shown before and after the constraint for comparison. Within the single- $\zeta$  basis SZ the wave function is severely distorted when the coefficient of the single 1s basis function is adjusted; for the double- $\zeta$  basis DZ, the local energy is still strongly distorted; for the quadruple- $\zeta$  basis QZ4P, the divergence in the local energy at the nucleus is cleanly removed, otherwise preserving the orbitals and the local energy.

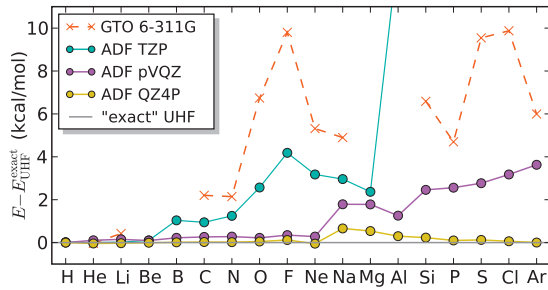


FIG. 2. Comparison of atomic unrestricted HF (UHF) energies. The reference energy (Ref. 6) is based on a cc-pV5Z-h GTO basis set and assumed to be exact within the published precision of 0.1 kcal/mol.

$$\sum_i \chi_i^j c_i = 0, \quad (4)$$

where the  $N_{\text{nuc}} \times N_{\text{bas}}$  elements of the constraint matrix are given by

$$\chi_i^j = \delta_{R_i, R_j} \delta_{i,0} [\delta_{n_i,0} (\zeta_i - Z_I) - \delta_{n_i,1} - Z_I (1 - \delta_{R_i, R_j}) \psi_i(\mathbf{R}_i - \mathbf{R}_I)].$$

In principle, these linear constraints can be enforced during a HF computation,<sup>32</sup> but this is rarely implemented in electronic structure codes. Instead a wave function originating from a code such as ADF can be cusp-corrected by a linear projection. To restrict the effect of the cusp correction to the vicinity of the nuclei, we fix all coefficients except those of the narrowest 1s basis function on each nucleus, and adjust the coefficients of the  $N_{\text{nuc}}$  remaining orbitals to fulfill Eq. (4). As the cusp conditions on different nuclei are nearly independent, this linear problem is always well-conditioned and has a unique solution.

Figure 1 demonstrates the effect of this cusp correction scheme on the orbitals and local energies. The single- $\zeta$  basis set clearly does not leave sufficient freedom for the correction and the single 1s basis function is severely distorted. For larger basis sets, however, the singularity in the local energy is cleanly removed with negligible distortion of the orbitals away from the cusp.

#### IV. ATOMIC TOTAL ENERGIES

We computed the HF energies of the first- and second-row atoms with ADF (see Fig. 2) to compare the quality of available basis sets. We found the ADF/QZ4P basis-set error to be below 0.1 kcal/mol for the first row atoms and below 1 kcal/mol for the second row atoms. The corresponding DMC energies are expected to be less sensitive to errors in the orbital basis than HF energies. For comparison, GTO-based HF energies were computed with the CRYSTAL program,<sup>33</sup> using a 6-311G basis set, also displayed in Fig. 2.

To evaluate the combined approach of using ADF and CASINO, the total energies of the first- and second-row atoms including correlation effects are compared to exact reference values in Fig. 3, as well as with earlier QMC data using wave functions defined on a radial grid.<sup>2</sup> A VMC calculation using the Slater-determinant wave function reproduced the

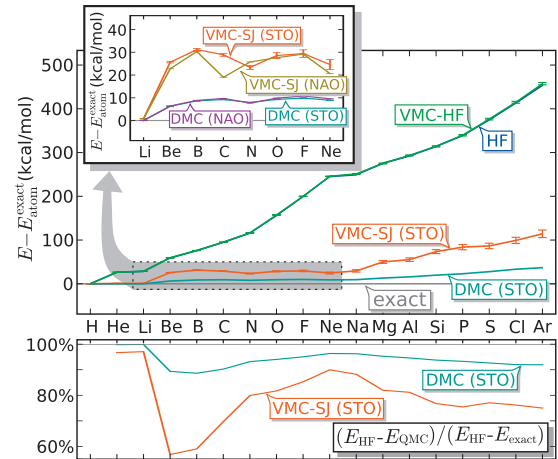


FIG. 3. Atomic total energies computed within HF (ADF/QZ4P) and QMC compared to the exact results (Ref. 11). The VMC run with the HF wave function reproduces the HF energy to within statistical error bars (demonstrating the negligible effect of the cusp correction on the energy for an accurate basis set). Optimized SJ wave functions recover roughly 60%–85% of the correlation energy within VMC and 90%–95% within DMC. The results agree very well with an earlier study based on numerical atomic orbitals (NAO) on a radial grid (Ref. 2).

HF energy, confirming that enforcing the cusp constraint preserves the overall quality of the wave function.

For each atom, we optimized a Jastrow factor<sup>26</sup> consisting of electron-electron, electron-nucleus and electron-electron-nucleus terms. The Jastrow factor was expressed as a power expansion in the interparticle distances with the parameters:  $C=3$ ,  $N_u=N_\chi=12$ ,  $N_f^{\text{en}}=2$ , and  $N_f^{\text{ec}}=3$  (see Eqs. 19, 20, and 21 of Ref. 29). For the VMC optimization, we minimized the mean absolute deviation (MAD) of the energy from the mean energy over a set of  $50\,000/N_{\text{elec}}$  configurations, performing five consecutive optimization iterations. This conservative choice of parameters allowed a reliable, automated optimization procedure. The resulting Slater-Jastrow (SJ) wave functions recover 60%–85% of the correlation energy.

DMC calculations using the optimized SJ wave functions recovered 90%–95% of the correlation energy for Be and heavier atoms. (Time step errors<sup>22</sup> were eliminated by linear extrapolation.) H and He do not have a nodal surface, so DMC produces the exact energy. The HF nodal surface for Li is extremely good, and DMC recovers more than 99.5% of the correlation energy in this case. The total energies of the first row atoms show excellent agreement with previous QMC results.<sup>2</sup> This indicates that the remaining error is due to the fixed-node approximation and presents the limit of the single determinant SJ method which can only be bettered by improving the nodal surface, e.g., by using backflow<sup>34</sup> or multideterminant wave functions.<sup>35</sup>

#### V. THE G2 SET OF MOLECULES

Further benchmarking was performed for the bond energies of the 55 molecules of the G2 set.<sup>3</sup> The bond energy  $E_{\text{bond}}$  is the difference between the molecular and atomized total energies, not including the zero point motion of the nuclei<sup>6</sup> or any further corrections.<sup>25</sup>



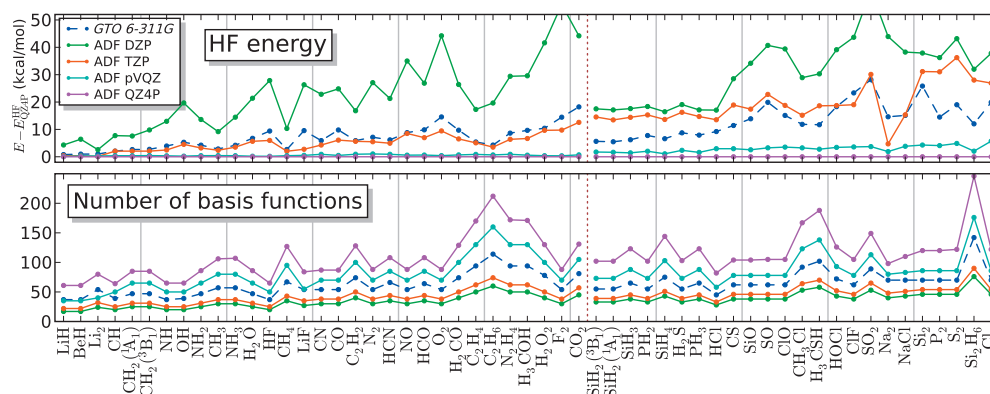


FIG. 4. Comparison of basis set errors in the HF energies for various STO basis sets. The QZ4P basis is used as a reference. Judging by Fig. 2, the remaining error is less than 1 kcal/mol for each second-row atom. The GTO 6–311G basis is roughly equivalent in precision to the STO TZP basis set which typically contains about 45% more basis functions. Geometries and GTO reference data were the same as those used in Ref. 7 [see supplementary material (Ref. 20)]. The dashed line separates the molecules containing only first-row atoms from those also containing second-row atoms.

First of all we used the G2 set to analyze the quality of the various basis sets by comparing the molecular HF energies (see Fig. 4). Based on this data and the atomic data, we chose the QZ4P basis set for all further work as it systematically gives the smallest basis set error.

To enable a direct comparison of results, the HF calculations (Fig. 4) were performed using the same molecular geometries as the previous GTO calculations reported in Ref. 7. For all other computations we used the more precise geometries from Ref. 25, where available, and for the remaining molecules we used those from Ref. 6, which were provided by the authors.<sup>20</sup> The geometry of  $\text{SiH}_2$  in the triplet state was obtained from Ref. 38.

For each molecule, we optimized a Jastrow factor with the same set of terms and parameters as used for the atoms. On average, we recovered 82% of the correlation energy within VMC.

The SJ wave functions thus obtained were then used as

the trial wave functions for DMC. For each system, two runs at time steps  $dt=0.01$  and  $dt=0.001$  were performed, allowing a linear extrapolation  $dt \rightarrow 0$ . We used a total of 40 CPU hours per electron (Intel XEON, 3 GHz) for each molecule, reaching a statistical precision of about  $25 \mu\text{hartree}/\text{electron}$ .

For comparison, the bond energies computed using different methods are shown in Fig. 5. The MAD of our STO-based bond energies from the experimental reference values is 3.2 kcal/mol. This is slightly larger than the deviation of 2.9 kcal/mol found in the pseudopotential-based DMC study of Grossman,<sup>4</sup> however, those values excluded the relativistic and spin-orbit corrections which shift individual reference values by up to 2 kcal/mol and would increase the MAD to 3.1 kcal/mol. Indeed, the MAD between Grossman's pseudopotential results and our all-electron results is just 2.0 kcal/mol, showing the strong correlation between the errors in the two sets of results. Our GTO-based all-electron DMC calcu-

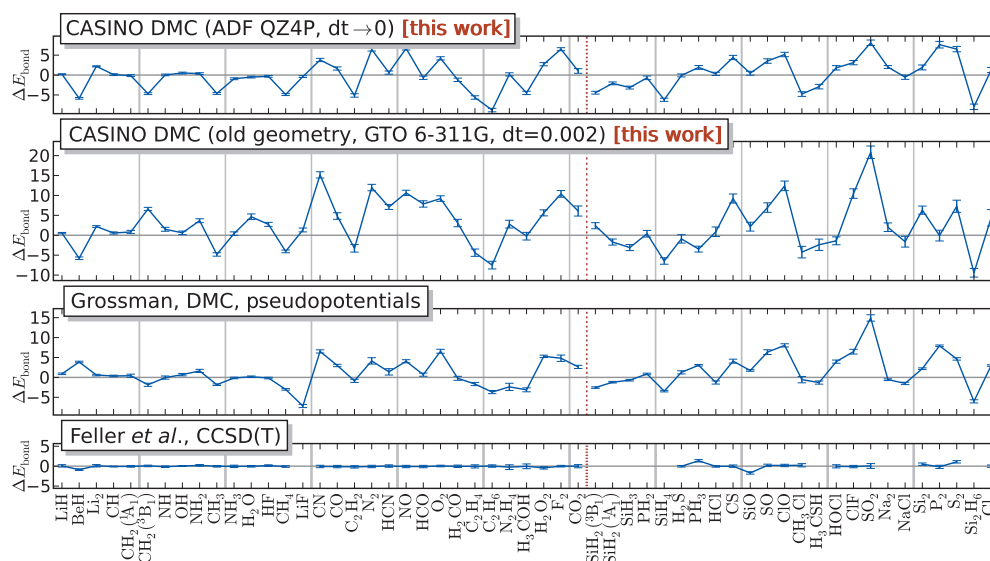


FIG. 5. Comparison of various calculated bond energies from *ab initio* computations. The deviation  $\Delta E_{\text{bond}} = E_{\text{bond}} - E_{\text{bond}}^{\text{ref}}$  from the experimental reference energy given in Ref. 24 (including zero-point motion, relativistic and spin-orbit corrections) is shown. Our GTO-based DMC calculations were performed with deliberately less effort than the STO-based ones. Grossman (Ref. 4) chose a DMC approach similar to ours but using pseudopotentials, whereas we have used an all-electron approach. The excellent values obtained by Feller *et al.* (Ref. 25) are based on fixed-core coupled cluster computations with a careful choice of basis set for each molecule including core-valence corrections.

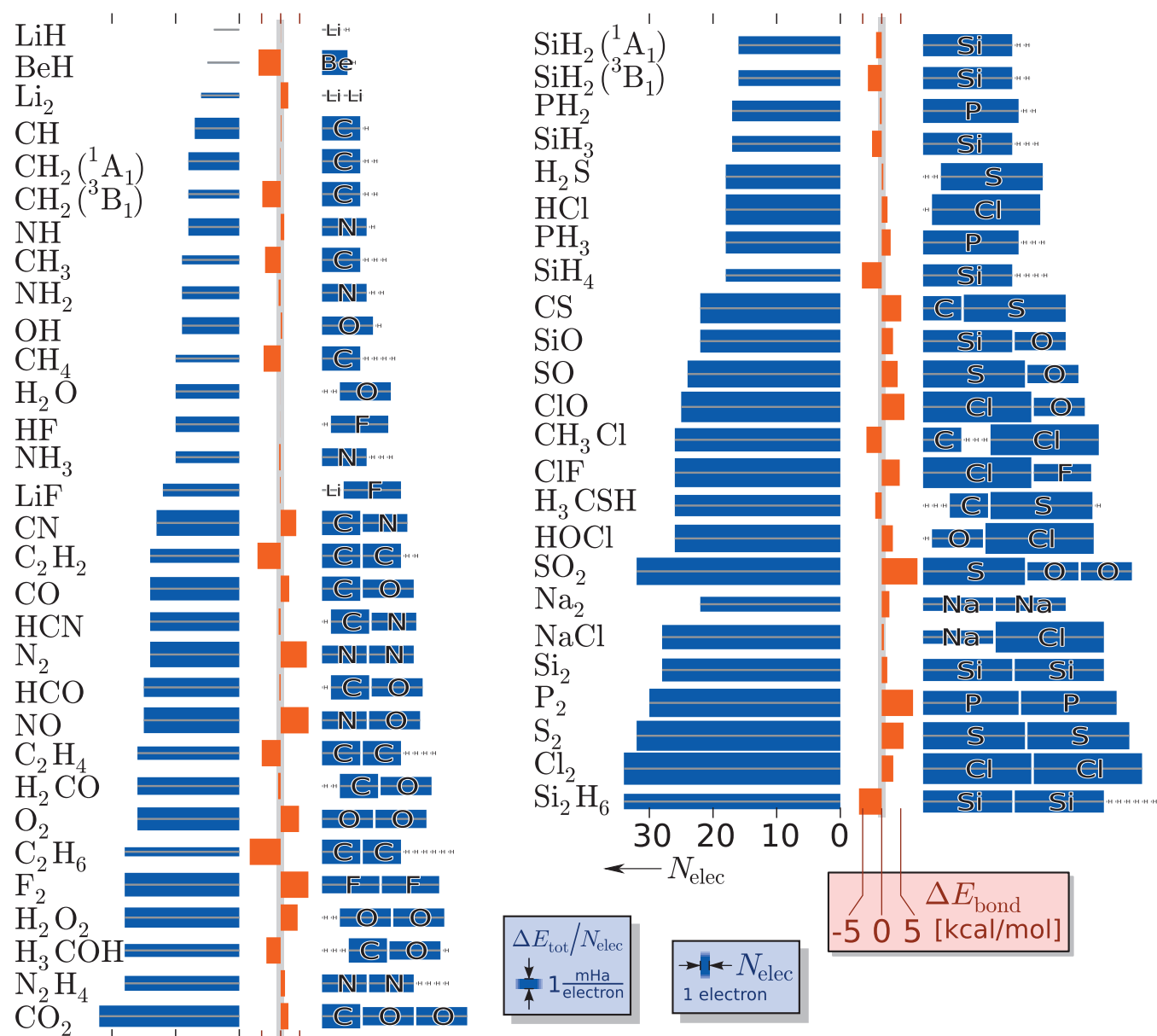


FIG. 6. Visualization of the error cancellation in bond energies. The horizontal bars on the left and right of each reaction correspond to the error in the total energy of each species  $\Delta E_{\text{tot}} = E_{\text{tot}}^{\text{dmc}} - E_{\text{tot}}^{\text{ref}}$  where the reference total energies are based on experimental atomization energies, atomic total energies and theoretical corrections (Ref. 6). Within the statistical precision, these errors are due almost entirely to the fixed node approximation. The vertical extent of each bar is the error per electron, which is in the range of 1–3 mHa/electron except for H, H<sub>2</sub>, He, Li, LiH, BeH, and Li<sub>2</sub> for which the nodal surface is exact or nearly so. The horizontal extent is the number of electrons. The difference in the areas on the right and left side of each reaction represents the error in the bond energy, which is shown in the center.

lations, which were deliberately performed without time step extrapolation or any kind of fine tuning of the basis sets or other computational parameters, gave a MAD from the reference values of 5.1 kcal/mol. This larger MAD shows that time-step extrapolations and a careful choice of basis sets are important in obtaining accurate single-determinant SJ DMC results.

Overall the obvious correlations between the errors of the three independent DMC-based attempts clearly suggest that a further systematic improvement can only be achieved by going beyond the fixed-node DMC approach with single-determinant SJ trial wave functions. This confirms the finding of Grossman that the fixed-node approximation dominates the remaining error.

## VI. ERROR CANCELLATION

The variational principle guarantees that the fixed node approximation leads to an overestimate of the total energy for each molecule and atom. The bond energy, being a difference of total energies, therefore shows significant error cancellation. Studying the bond energies in Fig. 5 reveals only very limited systematics about the sign and the magnitude of the errors in the bond energies. The picture becomes much clearer when we directly compare the errors in the total energies for the molecules and their constituent atoms. We found that the nodal surface error lies in the range of 1–3 mHa/electron for each atom and molecule in our test set, except for a very limited set of species for which the nodal

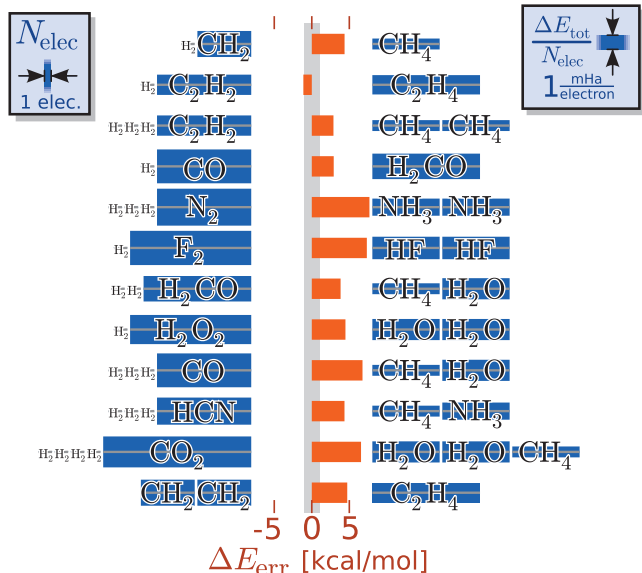


FIG. 7. The same visualization scheme as introduced in Fig. 6 applied to some of the chemical reactions studied in Ref. 36. In most cases, the nodal surface of the unsaturated molecule on the left is described significantly worse by single determinant wave function than the saturated molecules on the right, leading to a systematic overestimation of reaction energies as reported previously (Refs. 5 and 37).

surface is exact or nearly so. This allows us to visualize the error cancellation in a very compact and intuitive manner (Fig. 6).

We observe a limited number of cases where both molecule and constituent atoms are described well, leading to an accurate bond energy (LiH and Li<sub>2</sub>). In some cases, all species involved show similar nodal errors per electron, leading to strong error cancellation (e.g., CO, CO<sub>2</sub>, Na<sub>2</sub>, and Si<sub>2</sub>). In many other cases, however, the quality of the wave function is very different for the various species, which may lead to a large net error (e.g., NO and SO<sub>2</sub>), although the errors may also largely cancel (e.g., LiF and NaCl).

For most molecules containing hydrogen and carbon (e.g., CH<sub>3</sub> and CH<sub>4</sub>), one can observe that the wave function of the molecule is described significantly better than that of the carbon atoms, leading to a systematic overestimation of the bond energy. (Even though the absence of a nodal surface error for the hydrogen atom might intuitively suggest the opposite.) A similar effect can also be seen in the visualization of chemical reaction energies (Fig. 7). Here, one can observe that fully hydrogenated molecules are typically described better by the fixed node approximation than molecules containing double or triple bonds.

Finally, one can observe that second row atoms and their molecules typically show significantly larger nodal surface errors than first row atoms. The errors in bond energy, however, are not necessarily larger, indicating that these nodal surface errors arise mainly from the core electrons and that they cancel in energy differences.

## VII. CONCLUSIONS

To conclude, we have demonstrated the accuracy of STO trial wave functions generated by the ADF software packages in VMC and DMC calculations using the CASINO program.

TABLE I. Timing of VMC steps for the C<sub>2</sub>H<sub>6</sub> molecule. The STO/TZP basis set is of comparable precision to GTO and significantly more efficient to evaluate. For equal size of basis set, GTO without cusp correction performs as well as STO. The two types of cusp correction implemented in CASINO each add some computational overhead. The gpcc (Ref. 8) is a scheme that adds a correction function to each molecular orbital of arbitrary type, while the Gaussian cusp correction (Ref. 7) replaces the orbital close to nuclei by the exponential of a polynomial. For larger systems the computational cost is expected to grow linearly with the basis set size, promising up to a 45% performance gain for equivalent precision.

Type	Basis	$N_{\text{bas}}$	$T_{\text{cpu/step}}$ ( $\mu\text{s}$ )
STO	DZP	60	359
	TZP	74	391
	pVQZ	160	531
	QZ4P	212	643
Type	Cusp corr.	$N_{\text{bas}}$	
GTO	None	114	468
	gpcc	114	486
	Gaussian <sup>a</sup>	114	512

<sup>a</sup>Reference 7.

In direct comparison, STO basis sets promise a performance gain of 45% over GTO basis sets of comparable precision (see Table I). Using the QZ4P basis set from ADF, the basis set errors are below 0.1 kcal/mol for first-row atoms and below 1 kcal/mol for second-row atoms. DMC calculations for the G2 set of molecules recovered on average 95% of the correlation energy. Due to partial error cancellation, the atomization energies could be reproduced to a MAD from the experimental values of 3.2 kcal/mol.

The errors in the total energies of individual molecules and atoms, which originate—apart from statistical errors—almost entirely from the fixed node approximation, were then used to analyze and visualize the error cancellation in atomization and chemical reaction energies. While we find that the nodal error from the core electrons in the second row atoms largely cancels out, other error cancellations seem more coincidental than systematic.

As our bond energies are of similar quality to those obtained previously in pseudopotential calculations, we may assume that we have reached the limit in accuracy that is possible with a nodal surface obtained from single Slater determinant wave functions. Systematic studies of the nodal surface of multideterminant wave functions<sup>39</sup> indicate that significant improvements can be achieved with reasonable effort. Using backflow functions<sup>34</sup> or geminal wave functions<sup>40</sup> should also lead to higher accuracy QMC results, while retaining its excellent performance and scaling behavior.

The molecular geometries and the full set of results are provided in electronic form, available from supplementary material.<sup>20</sup>

## ACKNOWLEDGMENTS

We acknowledge the authors of the ADF software for permission to use the code, together with excellent support by S. van Gisbergen and E. van Lenthe. We thank D. O'Neill

and P. Gill for providing molecular coordinates and M. Korth for valuable discussions. This work was funded by the DAAD and the Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom (project number EP/F032773/1). M.D.T. acknowledges financial support from the Royal Society. The computations were performed using the facilities of the University of Cambridge High Performance Computing Service.

- <sup>1</sup>R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Ríos, *J. Phys.: Condens. Matter* **22**, 023201 (2010).
- <sup>2</sup>M. D. Brown, J. R. Trail, P. L. Ríos, and R. J. Needs, *J. Chem. Phys.* **126**, 224110 (2007).
- <sup>3</sup>L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Chem. Phys.* **94**, 7221 (1991).
- <sup>4</sup>J. C. Grossman, *J. Chem. Phys.* **117**, 1434 (2002).
- <sup>5</sup>S. Manten and A. Lüchow, *J. Chem. Phys.* **115**, 5362 (2001).
- <sup>6</sup>D. P. O'eill and P. M. W. Gill, *Mol. Phys.* **103**, 763 (2005).
- <sup>7</sup>A. Ma, M. D. Towler, N. D. Drummond, and R. J. Needs, *J. Chem. Phys.* **122**, 224322 (2005).
- <sup>8</sup>N. D. Drummond (unpublished), implemented in the CASINO code (gpcc) (Ref. 1).
- <sup>9</sup>J. C. Slater, *Phys. Rev.* **36**, 57 (1930).
- <sup>10</sup>S. F. Boys, *Proc. R. Soc. London, Ser. A* **200**, 542 (1950).
- <sup>11</sup>S. J. Chakravorty, S. R. Gwaltney, E. R. Davidson, F. A. Parpia, and C. F. Fischer, *Phys. Rev. A* **47**, 3649 (1993).
- <sup>12</sup>I. Shavitt, *Int. J. Quantum Chem.* **100**, 105 (2004).
- <sup>13</sup>H. J. Silverstone, *J. Chem. Phys.* **45**, 4337 (1966).
- <sup>14</sup>E. J. Baerends, D. E. Ellis, and P. Ros, *Chem. Phys.* **2**, 41 (1973).
- <sup>15</sup>A. Allouche, *Theor. Chem. Acc.* **42**, 325 (1976).
- <sup>16</sup>I. I. Guseinov, B. A. Mamedov, N. Sünel, and I. Guseinov, *J. Mol. Struct.: THEOCHEM* **593**, 71 (2002).
- <sup>17</sup>A. Bouferguene, M. Fares, and P. E. Hoggan, *Int. J. Quantum Chem.* **57**, 801 (1996).
- <sup>18</sup>J. F. Rico, R. López, A. Aguado, I. Ema, and G. Ramírez, *Int. J. Quantum Chem.* **81**, 148 (2001).
- <sup>19</sup>G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. F. Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler, *J. Comput. Chem.* **22**, 931 (2001).
- <sup>20</sup>See supplementary material at [10.1063/1.3288054](https://doi.org/10.1063/1.3288054) for the data presented in this work, including geometries and data cited for reference and comparison.
- <sup>21</sup>T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- <sup>22</sup>C. J. Umrigar, M. P. Nightingale, and K. J. Runge, *J. Chem. Phys.* **99**, 2865 (1993).
- <sup>23</sup>W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
- <sup>24</sup>D. Feller and K. A. Peterson, *J. Chem. Phys.* **110**, 8384 (1999).
- <sup>25</sup>D. Feller, K. A. Peterson, and D. A. Dixon, *J. Chem. Phys.* **129**, 204105 (2008).
- <sup>26</sup>R. Jastrow, *Phys. Rev.* **98**, 1479 (1955).
- <sup>27</sup>S. F. Boys and N. C. Handy, *Proc. R. Soc. London, Ser. A* **310**, 63 (1969).
- <sup>28</sup>K. E. Schmidt and J. W. Moskowitz, *J. Chem. Phys.* **93**, 4172 (1990).
- <sup>29</sup>N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. B* **70**, 235119 (2004).
- <sup>30</sup>J. Kussmann and C. Ochsenfeld, *Phys. Rev. B* **76**, 115115 (2007).
- <sup>31</sup>M. C. Per, S. P. Russo, and I. K. Snook, *J. Chem. Phys.* **128**, 114106 (2008).
- <sup>32</sup>P. T. A. Galek, N. C. Handy, A. J. Cohen, and G. K.-L. Chan, *Chem. Phys. Lett.* **404**, 156 (2005).
- <sup>33</sup>V. R. Saunders, R. Dovesi, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, N. M. Harrison, K. Doll, B. Civalieri, I. Bush, P. D'Arco, and M. Llunell, *CRYSTAL2003 User's Manual* (University of Torino, Torino, 2003).
- <sup>34</sup>P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. E* **74**, 066701 (2006).
- <sup>35</sup>E. Buendía, F. J. Gálvez, P. Maldonado, and A. Sarsa, *J. Chem. Phys.* **131**, 044115 (2009).
- <sup>36</sup>W. Klopper, K. L. Bak, P. Jorgensen, J. Olsen, and T. Helgaker, *J. Phys. B* **32**, R103 (1999).
- <sup>37</sup>M. Korth, oral presentation at QMC in the Apuan Alps, Vallico Sotto (2009).
- <sup>38</sup>A. Kalesos, T. H. Dunning, Jr., and A. Mavridis, *Mol. Phys.* **102**, 2597 (2004).
- <sup>39</sup>D. Bressanini, G. Morosi, and S. Tarasco, *J. Chem. Phys.* **123**, 204109 (2005).
- <sup>40</sup>M. Casula and S. Sorella, *J. Chem. Phys.* **119**, 6500 (2003).