Noncovalent Interactions by Quantum Monte Carlo: A Speedup by a Smart Basis Set Reduction

Matúš Duběcký
Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacký University Olomouc, Štefánikova 17, 772 12 Olomouc, Czech Republic
(Dated: January 28, 2015)

A fixed-node diffusion Monte Carlo (FN-DMC) method provides a promising alternative to the commonly used coupled-cluster (CC) methods, in the domain of benchmark noncovalent interaction energy calculations. This is mainly true for a low-order polynomial CPU cost scaling of FN-DMC and favorable FN error cancellation leading to benchmark interaction energies accurate to 0.1 kcal/mol. While it is empirically accepted that the FN-DMC results depend weakly on the one-particle basis sets used to expand the guiding functions, limits of this assumption remain elusive. Our recent work indicates that augmented triple zeta basis sets are sufficient to achieve a benchmark level of 0.1 kcal/mol. Here we report on a possibility of significant truncation of the one-particle basis sets without any visible bias on the overall accuracy of the final FN-DMC energy differences. The approach is tested on a set of seven small noncovalent closed-shell complexes including a water dimer. The reported findings enable cheaper high-quality FN-DMC benchmarks.

In the domain of benchmark ab-initio noncovalent interaction energy calculations, a fixed-node diffusion Monte Carlo (FN-DMC) provides a promising alternative to the commonly used coupled-cluster (CC) methods like CCSD(T) [1]. A truly accurate FN-DMC interaction energies (to 0.1 kcal/mol) have been recently reported on a number of small/medium noncovalent closed-shell complexes [2–3]. In addition to accuracy, the FN-DMC method is attractive for its favorable low-order polynomial CPU cost scaling [4–6] and favorable FN error cancellation that enabled its use in medium/large complexes [9–14] where the CC methods were intractable (in original formulation and reasonable basis set) until recently. An advent of local CC methods [15] may possibly change a situation in single-reference cases, nevertheless, a development of quantum Monte Carlo (QMC) approaches for quantum chemistry is still promising (and fruitful as shown below) for additional benefits like simple treatment of extended systems within the explicit periodic-boundary conditions or a possibility to treat multireference systems.

While it is empirically accepted that the FN-DMC results depend only weakly on the one-particle basis sets used to expand the guiding functions [4] (beyond a certain reasonable level, for the use of a real-space electron position representation), limits of this assumption remain unclear. Our recent work e.g. indicates that in DFT-based single-determinant guiding functions, the cardinality of basis sets is not as important as the presence of augmentation functions. An example of ammonia dimer complex well illustrates this behavior: a sequence of TZV, QZV and aug-TZV basis sets generates the interaction energies of -3.33±0.07, -3.47±0.07 and -3.10±0.06 kcal/mol [3], while the complete-basis-set CCSD(T) benchmark value for this complex at exactly the same geometry amounts to -3.15 kcal/mol [16, 17]. Augmented triple zeta basis sets were confirmed to be sufficient to achieve a challenging level of 0.1 kcal/mol in the final FN-DMC interaction energies in more than 30 cases [2–3]. However, is this a smallest basis set leading to the benchmark noncovalent interaction energies?

Here we report on a further possible significant truncation of the one-particle basis sets used to expand orbitals in the FN-DMC guiding functions. Namely, the rule identified to work well in the case of closed-shell noncovalent complexes is i) to remove all angular momentum basis functions higher than the first unoccupied shell from the TZV basis, and ii) add only the first corresponding [1s] function to the augmentation part. For example, in the case of a carbon atom, the aug-TZV basis set ([3s3p2d1f]+[1s1p1d1f]) coming with the effective-core potential (ECP, cf. below) contains 45 basis functions, while the proposed truncation ([3s3p2d]+[1s]), denoted as 1s-rTZV, keeps only 23 spherical harmonic basis functions in the calculation. Even though such a reduction might seem drastic, the results produced with such basis sets reveal no bias in energy differences with respect to the full aug-TZV values, as reported below. A proof of principle is demonstrated on a set of seven noncovalent complexes covering various types of bonding including pure dispersion, native hydrogen bonds and mixed cases. It turns out, that a high-quality benchmark FN-DMC numbers are available at lower costs than assumed to date.

The considers test set contains seven closed-shell noncovalent complexes from the A24 database [18]: ammonia dimer (AM...AM), water-ammonia (WA...AM), water dimer (WA...WA), ammonia methane complex (AM...MT), methane dimer (MT...MT), hydrogen fluoride dimer (HF...HF), and HCN dimer (HCN...HCN). The atomic cores in these complexes were replaced by ECPs developed by Burkatzki et al. [19, 20], and the cor-
responding TZV basis sets were used together with the augmentation functions [21] leading to aug-TZV quality bases or their truncated counterparts (1s-rTZV). Single-determinant Slater-Jastrow [5] trial wave functions were constructed using orbitals from B3LYP (GAMESS code) and the Schmidt-Moskowitz [23] homogeneous isotropic Jastrow factors [5] including electron-electron and electron-nucleus terms [8] were expanded in a fixed basis set of polynomial Padé functions [5]. The Jastrow parameters were refined by the Hessian driven VMC optimization of at least 10x10 iterations using a linear combination [24] of energy (95%) and variance (5%) as a cost function. The orbital parameters were frozen to keep the nodes of guiding functions intact. After optimization, these were subsequently used in the production FN-DMC runs using the so-called T-moves scheme to treat ECPs [25] and a time step of 0.005 a.u. The target walker populations in FN-DMC amounted to 16-32k. All QMC calculations were performed using the QWalk [26] code.

The FN-DMC results obtained with guiding functions expanded in aug-TZV and reduced 1s-rTZV basis sets are reported in Tab. 1. The interaction energies obtained with 1s-rTZV are clearly compatible with the aug-TZV data in all considered cases (Figure 1), even the total number of basis functions is reduced by more than two times. As mentioned above, the TZV or even QZV bases do not lead to correct energy differences. On the other hand, rTZV has smaller number of primitives than TZV. Therefore, what causes a surprising qualitative difference is the one [1s] augmentation function per atom. At the current stage, we expect that this leads to the enhancement of sampling in the interstitial regions. The nodes of guiding functions may hardly be improved and the FN error is thus operative [22]. This claim is also supported by the total energies which are higher by ∼0.001 a.u. (in dimers) with respect to aug-TZV values, but interestingly, the local energy variances (σ) essentially do not differ (only by ∼0.01 a.u. or less) probably because identical occupied-shell contractions are used in both bases. The truncated bases are thus promising not only for single-determinant but also for the multideterminant FN-DMC calculations, since they seem to encode the required information but produce more compact Fock spaces.

The CPU cost of the Slater matrix evaluation step scales as $O(MN^3)$ where $M$ is the number of basis functions and $N$ is the number of electrons. The expected speedup with reduced $M$ is thus proportional to the ratio $M_i/M_j$, if the statistical sampling efficiency remains constant (as here, cf. above). Nevertheless, the cost saving starts to be interesting only for a significant reduction of $M_j$. We note in passing that since the FN-DMC calculations are relatively expensive in general, and frequently millions of CPU-hours are invested to valuable projects (e.g. Ref. [12]), a two-fold reduction of $M_j$, as reported here, may result in considerable CPU time savings.

The following examples illustrate the real speedups: in HCN...HCN complex (20 electrons) and a Slater determinant guiding function in FN-DMC, the 1s-rTZV calculations are 1.5 times faster than the aug-TZV ones ($M_{\text{aug-TZV}}/M_{\text{1s-rTZV}} = 2.02$). The Slater-Jastrow run achieves a speedup of 1.22. The variance of local energy in the truncated case increases by a negligible value of 0.006 a.u (a value typical for all cases in Table 1). The speedup is far from expected, indicating that routines other than the Slater matrix value and Laplacian updates are in play, e.g. evaluation of Jastrow factor, pseudopotentials or distance matrix updates. In the case of larger systems like coronene...H$_2$, a complex with 24 carbon atoms and 110 electrons ($M_{\text{aug-TZV}}/M_{\text{1s-rTZV}} = 1402/692 = 2.03$), the speedup of a pure Slater run is 1.92-fold while the FN-DMC calculation using Slater-Jastrow guiding function achieves a speedup factor of 1.74. Clearly, in large systems, the theoretical limit of the speedup, $M_{\text{aug-TZV}}/M_{\text{1s-rTZV}}$, is smoothly achieved since the Slater matrix updates proportional to $O(MN^3)$ dominate.

In conclusion, a new one-particle basis set truncation scheme has been shown to considerably reduce computational requirements while retaining a full accuracy in energy differences. This finding extends range of applicability of FN-DMC method in noncovalent complexes.
ACKNOWLEDGMENTS

The author is grateful to Claudia Filippi for sharing an improved effective core potential for H and to Lubos Mitas, Petr Jurečka, and René Derian for stimulating and fruitful discussions. Support from the Ministry of Education, Youth and Sports of the Czech Republic (project LO1305) is gratefully acknowledged.

[1] K. Raghavachari and G. W. Trucks. Chem. Phys. Lett., 157:479–483, 1989.
[2] M. Dubécký, P. Jurečka, R. Derian, P. Hobza, M. Otyepka, and L. Mitas. J. Chem. Theory Comput., 9:4287–4292, 2013.
[3] M. Dubécký, R. Derian, P. Jurečka, L. Mitas, P. Hobza, and M. Otyepka. Phys. Chem. Chem. Phys., 16:20915–20923, 2014.
[4] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Rev. Mod. Phys., 73:33–83, 2001.
[5] M. Bajdich and L. Mitas. Acta Phys. Slovaca, 59:81–168, 2009.
[6] B. M. Austin, D. Y. Zubarev, and W. A. Lester. Chem. Rev., 112:263–288, 2012.
[7] M. Mella and J. B. Anderson. J. Chem. Phys., 119:8225–8228, 2003.
[8] C. Diedrich, A. Lüchow, and S. Grimme. J. Chem. Phys., 123:184106, 2005.
[9] S. Sorella, M. Casula, and D. Rocca. J. Chem. Phys., 127:014105, 2007.
[10] M. Körth, A. Lüchow, and S. Grimme. J. Phys. Chem. A, 112:2104–2109, 2008.
[11] B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, C. Filippi, and M. Scheffler. J. Chem. Phys., 129:194111, 2008.
[12] A. Ambrosetti, D. Allé, R. A. DiStasio, and A. Tkatchenko. J. Phys. Chem. Lett., 5:849–855, 2014.
[13] A. Benali, L. Shulenburger, N. A. Romero, J. Kim, and O. A. von Lilienfeld. J. Chem. Theory. Comput., 10:3417–3422, 2014.
[14] Y. S. Al-Hamdani, D. Allé, O. A. von Lilienfeld, and A. Michaelides. J. Chem. Phys., 141:18C530, 2014.
[15] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K.-L. Chan. Science, 345:640–643, 2014.
[16] P. Jurečka, J. Šponer, J. Černý, and P. Hobza. Phys. Chem. Chem. Phys., 8:1985–1993, 2006.
[17] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill. J. Chem. Phys., 132:144104, 2010.
[18] J. Řezáč and P. Hobza. J. Chem. Theory Comput., 9:2151–2155, 2013.
[19] M. Burkatzi, C. Filippi, and M. Dolg. J. Chem. Phys., 126:234105, 2007.
[20] M. Dolg and C. Filippi. Private communication.
[21] R.A. Kendall, T. H. Jr. Dunning, and R. J. Harrison. J. Chem. Phys., 96:6796, 1992.
[22] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, and et. al. J. Comput. Chem., 14:1347–1363, 1993.
[23] J. W. Moskowitz and K. E. Schmidt. J. Chem. Phys., 97:3382–3385, 1992.
[24] C. J. Umrigar and C. Filippi. Phys. Rev. Lett., 94:150201, 2005.
[25] M. Casula. Phys. Rev. B, 74:161102, 2006.
[26] L. K. Wagner, M. Bajdich, and L. Mitas. J. Comp. Phys., 228:3390–3404, 2009.