



Cite this: *Phys. Chem. Chem. Phys.*,
2014, 16, 20915

Quantum Monte Carlo for noncovalent interactions: an efficient protocol attaining benchmark accuracy

Matúš Dubecký,^{*a} René Derian,^b Petr Jurečka,^{*a} Lubos Mitas,^c Pavel Hobza^{ad} and Michal Otyepka^a

Reliable theoretical prediction of noncovalent interaction energies, which are important e.g. in drug-design and hydrogen-storage applications, is one of the longstanding challenges of contemporary quantum chemistry. In this respect, the fixed-node diffusion Monte Carlo (FN-DMC) method is a promising alternative to the commonly used “gold standard” coupled-cluster CCSD(T)/CBS method due to its benchmark accuracy and favourable scaling, in contrast to other correlated wave function approaches. This work is focused on the analysis of protocols and possible trade-offs for FN-DMC estimations of noncovalent interaction energies, and proposes an efficient yet accurate computational protocol using simplified explicit correlation terms with a favorable $O(N^3)$ scaling. It achieves results in excellent agreement (mean unsigned error ~ 0.2 kcal mol⁻¹) with respect to the CCSD(T)/CBS data on a number of complexes, including benzene/hydrogen, the T-shape benzene dimer, stacked adenine–thymine complex and a set of small noncovalent complexes (A24). The high accuracy and reduced computational costs predestinate the reported protocol for practical interaction energy calculations of large noncovalent complexes, where the CCSD(T)/CBS is prohibitively expensive.

Received 14th May 2014,
Accepted 31st July 2014

DOI: 10.1039/c4cp02093f

www.rsc.org/pccp

1 Introduction

Noncovalent interactions between molecular complexes and/or their parts are of key importance in many areas of chemistry, biology and materials science.^{1–3} Their striking manifestations include the properties of liquids, molecular recognition⁴ or the structure and function of bio-macromolecules, to name just a few.^{2,5} In general, experiments make it possible to obtain information on the strength of noncovalent interactions e.g. from dissociation and adsorption enthalpies. Nevertheless, direct information on their nature is usually unavailable. In order to characterise the noncovalent interaction/s of interest more precisely, one typically resorts to a combination of multiple techniques.^{1,6} Theory and high-accuracy calculations here usefully complement experiments by providing the detailed information necessary for their fundamental understanding.

As the binding energies of noncovalent complexes are weak (typically 0.5 to 30 kcal mol⁻¹ in small complexes) when compared to a typical covalent bond energy (100 kcal mol⁻¹), their computations require methods of exceptional quality. For benchmark purposes, only approaches with a degree of accuracy beyond the accepted chemical accuracy (1 kcal mol⁻¹) are considered reliable enough. Ideally, the level of subchemical accuracy⁷ (0.1 kcal mol⁻¹) should be reached. This task poses a long-standing challenge to modern computational chemistry.

The established “gold standard” of the quantum mechanical calculation of noncovalent interaction energies⁸ has been the CCSD(T)⁹ (coupled-cluster singles and doubles with perturbative triples) method, which guarantees the desired accuracy provided that large enough basis sets and/or complete basis set (CBS) extrapolations are employed. The benchmark capability of CCSD(T)/CBS has recently been verified on a set of small noncovalent complexes by a higher-order CCSDT(Q) method¹⁰ and in small bases also by the CCSDTQ, CCSDTQP and full CI (configuration interaction).¹¹ Nevertheless, because of the rapid growth of the CCSD(T) computational cost in the basis set size M , $\propto O(M^7)$, and even because of the more steeply growing demands of higher-order coupled-cluster approaches, their practical use remains limited to relatively small systems.^{12–16} For a summary of developments in this field, cf. ref. 17.

A promising alternative to solve the non-relativistic Schrödinger equation for electrons in the Born–Oppenheimer approximation is

^a Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacký University Olomouc, tř. 17 listopadu 12, 771 46 Olomouc, Czech Republic. E-mail: matus.dubecky@upol.cz, petr.jurecka@upol.cz

^b Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 845 11 Bratislava, Slovakia

^c Department of Physics and CHiPS, North Carolina State University, Raleigh, NC 27695, USA

^d Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Prague 6, Czech Republic

the fixed-node (FN) diffusion Monte Carlo (DMC/FN-DMC) method, a member of the quantum Monte Carlo (QMC) class of methods based on random sampling. FN-DMC solves an imaginary-time Schrödinger equation projecting out the exact ground-state within the constraints given by the nodal surface ($\Psi_T = 0$) of the best available trial wave function (Ψ_T), required to preserve the antisymmetry of the simulated electronic state.^{18,19} The FN-DMC approach is favourable for its direct treatment of many-body correlations (competitive to high-order approaches including full CI),^{18,19} low-order polynomial scaling ($\propto O(N^{3-4})$ with the number of electrons N)¹⁸ and its intrinsic massive parallelism.^{18,19} The representation of a wave function as an ensemble of walkers (electron real space position vectors) in QMC results in a CBS-equivalent mode and energies that are insensitive to basis set superposition errors (possibly present in Ψ_T).¹⁹ QMC makes it possible to sample sophisticated many-body wave function ansätze with explicit inter-electronic dependencies efficiently, as the wave functions are sampled stochastically.^{18–21} On the other hand, it suffers from disadvantages like the slowly convergent error bars that are intrinsic to stochastic sampling, the need to circumvent the difficulties in sampling antisymmetric wave functions and/or complicated estimations of quantities beyond energies. Despite the mentioned limitations, QMC approaches are behind some of the most paradigmatic results obtained for quantum systems such as the correlation energy of electron gas,²² widely used in computational electronic structure methods. For more details and background on QMC, we refer the reader to the published reviews.^{18,19,21}

In the domain of noncovalent interactions, there are two salient features of FN-DMC. First, the method accurately describes the dynamic correlation effects crucial for noncovalent interactions, since it recovers *all possible many-body correlations* within the constraints given by the nodal surface.¹⁸ Second, the finite FN error of Ψ_T used in DMC is expected to cancel out in energy differences nearly exactly,^{23–26} because the nodes in the region of the molecule A are essentially unchanged by the presence of the weakly interacting molecule B and *vice versa*. This trend has been qualitatively confirmed by a direct inspection of nodal surfaces.²⁶ For these reasons, FN-DMC based on single-determinant trial functions (containing *e.g.* Kohn–Sham orbitals) has been found to be very promising in small/medium^{25–27} and medium/large molecular systems,^{3,28–34} where conventional correlated methods are inapplicable due to the prohibitive computational cost. Note that the requirement of small error bars, a must in noncovalent interactions, leads to a large cost (large prefactor) of the calculations. For example, calculations of small complexes like those considered as a teaching set in the current work are much more costly in QMC than in CCSD(T)/CBS. Nevertheless, the situation turns to the contrary in large systems. Although systematic understanding of error cancellation in QMC energy differences is still far from complete, the data accumulated to date have begun to delineate the method performance for various types of systems.^{26,35,36} For more detailed presentations, see ref. 19, 23, 25 and 26.

In our previous study in this field,²⁶ we demonstrated the ability of the FN-DMC method to reach the CCSD(T)/CBS results

within 0.1 kcal mol^{−1} on a teaching set of six small noncovalent complexes: the dimers of ammonia, water, hydrogen fluoride, methane and ethene and the ethene–ethyne complex. The successive procedure was subsequently tested on complexes of benzene–methane, benzene–water and the T-shape benzene dimer, where the FN-DMC results deviated by no more than 0.25 kcal mol^{−1} with respect to the best available CCSD(T)/CBS interaction energy estimates. In a way, QMC performed much better than expected *a priori*.²⁵ Note that the genuine accuracy of CCSD(T) for these sizable systems has not yet been confirmed to subchemical accuracy at a higher level (CCSDT(Q)), and both approaches therefore provide independent interaction energy estimates that agree remarkably well. In addition, extended basis sets (*e.g.* aug-cc-pVTZ) are unavoidable in these CC calculations, because small bases lead to qualitatively incorrect results.¹⁰ For more of the complexes considered in the current work, such CC calculations are prohibitively expensive and if the FN-DMC calculations prove to be accurate enough, this route immediately provides a new means for addressing such difficult problems.

In the present work, we analyze the possible FN-DMC approaches and trade-offs and report on a considerably cheaper and yet accurate scheme based on a two-body explicit correlation (Jastrow) factor, which is a simplification with respect to the traditionally used three-body Jastrow term. The discovery of its ability to attain ~ 0.2 kcal mol^{−1} accuracy in the case of noncovalent interactions as reported in the current work represents an important advance. The favourable performance of the simplified scheme also directly confirms the previous observation that the fixed-node error cancellation is the main reason for the success of the one-determinant FN-DMC in weakly bound noncovalent complexes.²⁶

2 Molecular complexes

The calculations reported below were performed on a diverse set of hydrogen-bonded and/or dispersion-bound complexes for which reliable estimates of interaction energies have been published.^{10,12,14} The teaching set consists of the dimers of ammonia, water, hydrogen fluoride, methane and ethene, and the ethene–ethyne complex as in our previous work to allow a convenient comparison. The complexes used for testing purposes include benzene–H₂, benzene–methane, benzene–water, the T-shape benzene dimer, stacked adenine–thymine complex and the whole set of 24 small noncovalent complexes A24.¹⁰

The selection of the benzene/H₂ complex is motivated by our interest in the assessment of interaction energies between carbon-based materials and H₂, which is useful for hydrogen-storage applications. The adenine–thymine complex, on the other hand, serves as a first stringent test of our approach in DNA base-pair interactions, and for future reference in the assessment of interactions in larger DNA fragments.

3 Analysis of QMC protocols

The development of a FN-DMC-based methodology leading to accurate noncovalent interaction energies involves extensive

testing and the elimination of biases that affect the final results. Indeed, this has to be done in a stepwise manner since several sets of parameters enter the multistage approach.^{37,38}

(i) The ansatz for the trial wave function Ψ_T must be selected first. The “standard model” and frequently used choice is the Slater–Jastrow^{18,19} functional form, a product of determinant(s) and an explicit correlation term (Jastrow factor), as employed here. Then it is necessary to choose an effective core potential (ECP, if any) and a basis set (*e.g.*, aug-TZV). Subsequently, the Slater determinant(s) are constructed with DFT, Hartree–Fock (HF) or post-HF orbitals. Finally, the terms included in the Jastrow factor²⁰ must be specified. They can include, for instance, electron–electron (ee), electron–nucleus (eN), and electron–electron–nucleus (eeN) terms, which contain explicit functional dependencies on inter-particle distances and thus efficiently describe dynamic correlation effects.¹⁸

(ii) The variational (VMC) optimisation step consists of the selection of the VMC cost function and a parametric optimisation of Ψ_T , which may or may not include the adjustment of the nodal surfaces (which we avoid in the current work). It is possible to improve the nodes by the reoptimisation of the orbitals and/or determinant expansion coefficients.

(iii) The final FN-DMC ground-state projection calculations depend, in addition to the Ψ_T optimised in step (ii), on the parameters of the DMC simulation itself, including an imaginary time step, the treatment of ECPs, the target walker population/s and target error bar/s, to name but a few of the most important.

In general, the parameters and/or choices in all of points (i) to (iii) affect the final interaction energies obtained after the production DMC simulations in (iii), as the differences of the statistically independent total-energy expectation values (with associated error bars). The parameters in (i) and (ii), in addition to the energies accumulated during DMC runs, also modify the energy variance, thus determining the length of the DMC simulations to reach the fixed target statistical accuracy.

Table 1 shows the representative setups and the related results that helped us to trace the importance of the parameter changes considered and identify useful protocols; here, the preceding one²⁶ is labelled 3tj and the new one 2tj. These differ in the number of terms (ee, eN and eeN, *vs.* ee and eN) considered in the Jastrow factor (*cf.* the Methods section). The key observations from Table 1 may be summarised as follows.

Table 1 The FN-DMC interaction energies E (kcal mol^{−1}) obtained from various tested protocols compared to the CCSD(T)/CBS reference E_R (kcal mol^{−1}). The protocol-type attribute, if applicable, is indicated in column P. For clarity, analyzed features of the protocols with respect to the standard²⁶ 3tj protocol are indicated by the bold typeface. Abbreviations: Ψ_T – trial wave function, LC – a linear combination of energy (95%) and variance (5%), Dis. – a distinct Jastrow factor (see text)

Complex	E_R	P	Ψ_T			VMC opt. Cost function	FN-DMC		
			Basis set	Method	Jastrow		Time step/a.u.	ECP treatment	E
Ammonia dimer	−3.15 ^a	3tj	TZV	B3LYP	3tj	LC	0.005	T-moves	−3.33 ± 0.07 ^b
			QZV	B3LYP	3tj	LC	0.005	T-moves	−3.47 ± 0.07 ^b
			aug-TZV	B3LYP	3tj	LC	0.005	T-moves	−3.10 ± 0.06 ^b
			aug-QZV	B3LYP	3tj	LC	0.005	T-moves	−3.13 ± 0.07 ^b
			aug-TZV	HF	3tj	LC	0.005	T-moves	−3.12 ± 0.07 ^b
			aug-TZV	B3LYP	2tj	LC	0.005	T-moves	−3.15 ± 0.05
		2tj	aug-TZV	B3LYP	2tj	LC	0.01	T-moves	−3.14 ± 0.05
			aug-TZV	B3LYP	3tj	Variance	0.005	T-moves	−3.28 ± 0.04
			aug-TZV	B3LYP	3tj	LC	0.01	T-moves	−3.22 ± 0.07
			aug-TZV	B3LYP	3tj	LC	0.005	Locality	−3.27 ± 0.07
			aug-TZV	B3LYP	3tj	LC	0.01	Locality	−3.39 ± 0.06
Water dimer	−5.07 ^a	3tj	aug-TZV	B3LYP	3tj	LC	0.005	T-moves	−5.26 ± 0.08 ^b
			aug-TZV	B3LYP	3tj	LC	0.01	T-moves	−5.13 ± 0.08
			aug-TZV	B3LYP	3tj Dis.	LC	0.005	T-moves	−5.15 ± 0.08 ^b
			aug-TZV	B3LYP	3tj Dis.	LC	0.01	T-moves	−5.26 ± 0.08
		2tj	aug-TZV	B3LYP	2tj	LC	0.005	T-moves	−5.24 ± 0.09
			aug-TZV	B3LYP	2tj Dis.	LC	0.005	T-moves	−5.50 ± 0.08
			aug-TZV	B3LYP	2tj	LC	0.01	T-moves	−5.33 ± 0.09
			aug-TZV	B3LYP	2tj	LC	0.01	T-moves	−5.33 ± 0.09
Methane dimer	−0.53 ^a	3tj	aug-TZV	B3LYP	3tj	LC	0.005	T-moves	−0.44 ± 0.05 ^b
			aug-QZV	B3LYP	3tj	LC	0.005	T-moves	−0.55 ± 0.04
			aug-TZV	HF	3tj	LC	0.005	T-moves	−0.52 ± 0.10
		2tj	aug-TZV	B3LYP	2tj	LC	0.005	T-moves	−0.60 ± 0.07
			aug-TZV	HF	2tj	LC	0.005	T-moves	−0.64 ± 0.08
			aug-TZV	HF	2tj	LC	0.005	T-moves	−0.64 ± 0.08
Ethene dimer	−1.48 ^a	3tj	aug-TZV	B3LYP	3tj	LC	0.005	T-moves	−1.47 ± 0.09 ^b
			aug-QZV	B3LYP	3tj	LC	0.005	T-moves	−1.54 ± 0.09
			aug-TZV	B3LYP	3tj	LC	0.01	T-moves	−1.42 ± 0.09
		2tj	aug-TZV	B3LYP	2tj	LC	0.005	T-moves	−1.54 ± 0.09
			aug-TZV	B3LYP	2tj	LC	0.01	T-moves	−1.55 ± 0.09
			aug-TZV	B3LYP	2tj	LC	0.01	T-moves	−1.55 ± 0.09

^a Takatani *et al.*¹⁴ ^b Dubecký *et al.*²⁶

3.1 Basis sets

In the ammonia dimer complex, where the reference interaction energy amounts to $-3.15 \text{ kcal mol}^{-1}$,¹⁴ the TZV and QZV bases result in interaction energies of -3.33 ± 0.07 and $-3.47 \pm 0.07 \text{ kcal mol}^{-1}$, whereas the aug-TZV and aug-QZV bases lead to FN-DMC interaction energies of -3.10 ± 0.06 and $-3.13 \pm 0.07 \text{ kcal mol}^{-1}$,²⁶ respectively. The presence of augmentation functions in Ψ_T is therefore crucial,²⁴ whereas an increase of the basis set cardinality beyond the TZV level seems to play a smaller role than in the standard methods of quantum chemistry. In the methane dimer and ethene dimer complexes, the aug-TZV data are found to be statistically indistinguishable from the aug-QZV data as well. Since the aug-TZV basis set reaches the reference data within $0.1 \text{ kcal mol}^{-1}$ in the whole training set considered,²⁶ we have used it throughout the study.

Note that using the aug-QZV basis set with approximately two times more basis functions than the aug-TZV lowers the energy variance. For example, in the methane dimer complex, the variance is improved by 0.01 a.u. (from 0.19 to 0.18 a.u.), decreasing the sampling needed to reach the same error bar by about 10%. Nevertheless, the cost of the aug-QZV calculations for the fixed number of DMC steps is three times higher, making the aug-TZV approach still much more favourable when considering the overall cost/accuracy ratio.

3.2 Orbitals in the Slater part of Ψ_T

Since the FN-DMC energies depend primarily on the nodal surface of the trial wave function Ψ_T , we tested HF and B3LYP sets of orbitals in the Slater determinants. In the ammonia dimer complex,²⁶ we found that the HF and B3LYP orbitals provide FN-DMC interaction energies that are indistinguishable within the error bars, namely -3.12 ± 0.07 vs. $-3.10 \pm 0.06 \text{ kcal mol}^{-1}$, which are both in good agreement with the CCSD(T)/CBS reference ($-3.15 \text{ kcal mol}^{-1}$ (ref. 14)). Similar conclusions apply in the case of methane dimer in both of the types of schemes considered, 2tj and 3tj alike. The total energies from the B3LYP orbitals were always found to produce variationally lower total energy expectation values^{39,40} than those from HF orbitals (e.g. in the ammonia dimer by $\sim 0.001 \text{ a.u.}$). This indicates the better quality of the B3LYP nodal surfaces and therefore we used B3LYP orbitals for the rest of the calculations as well. We note that the results are expected to depend on the choice of the orbitals only very weakly due to the favourable FN error cancellation that takes place in weakly interacting complexes.^{23–26}

Since the concept of error cancellation is not limited to one determinant, similar behaviour is expected in the case of *non-covalent interactions* between open-shell systems where complete-active-space wave functions capturing multi-reference effects^{41,42} may be used instead of Hartree–Fock/Kohn–Sham determinants.

3.3 Jastrow factor

The considered variations of the Jastrow term include the reduction of terms (2tj, cf. the Methods section) with respect to the previously reported version (3tj), and the so-called distinct

Jastrow factor including distinct parameters (3tj Dis.) on the non-equivalent atoms of the same type. We have found that, e.g. in ammonia, water and ethene, the 2tj and 3tj protocols generate approximately the same results. On the other hand, this is not the case in the methane dimer. The water dimer is an example where the 3tj Jastrow factor is not sufficient and the corresponding interaction energy ($-5.26 \pm 0.08 \text{ kcal mol}^{-1}$) deviates considerably from the reference ($-5.07 \text{ kcal mol}^{-1}$ (ref. 14)). In order to reach a subchemical accuracy margin, a distinct 3tj Jastrow correlation factor must be considered (3tj Dis., $-5.15 \pm 0.08 \text{ kcal mol}^{-1}$). On the other hand, the use of the 2tj scheme in combination with the distinct feature (2tj Dis., Table 1) is (currently) not recommended, since in the studied case it produced a value that is too different from the reference value ($-5.50 \pm 0.08 \text{ kcal mol}^{-1}$). The reason is probably an insufficient number of sampling points used in the VMC optimisation (cf. ref. 43). A reasonable compromise between the accuracy and cost is thus provided by the 2tj scheme ($-5.24 \pm 0.09 \text{ kcal mol}^{-1}$), which is acceptable within the less stringent but acceptable target error criterion, e.g. $0.2\text{--}0.3 \text{ kcal mol}^{-1}$.

The comparison of timings (on 64 cores) for calculations with various Jastrow factors, including VMC optimization, DMC (to reach $0.1 \text{ kcal mol}^{-1}$ error margin) and the total timings to get a single point total energy are reported in Table 2. In addition, the DMC timings are illustrated in Fig. 1. The data in Table 2 show that the typical VMC optimization cost is negligible with respect to the DMC part of a typical calculation for noncovalent interaction energy purposes, i.e. the total timings are dominated by the DMC. The DMC timings indicate that the 2tj scheme cost scales as N^3 while the 3tj approximately scales as N^4 (cf. Fig. 1), where N is the number of electrons. Since the scaling of the 2tj approach is asymptotically much more favorable and generates results with comparable quality, we believe it will be very useful in calculations of large complexes where more complex Jastrow terms are very costly.

3.4 VMC cost function

The variance used as a cost function in the VMC optimization of the Jastrow term parameters leads to higher FN-DMC total energies when compared to energy minimisation (with 95% of energy and 5% of variance, cf. the Methods section).⁴⁴ For example, test calculations employing variance minimisation in ammonia lower the energy variance, as expected, but the total

Table 2 The timings (in hours of running on 64 cores) for VMC optimization t_{opt} , DMC (to reach $0.1 \text{ kcal mol}^{-1}$ error bar) t_{DMC} and total time t to get a total single point energy. N denotes the total number of electrons

Complex	N	Protocol	t_{opt}	t_{DMC}	t
Water monomer	8	2tj	0.1	0.5	1.9
		3tj	1.9	0.6	2.3
Water dimer	16	2tj	0.2	10.5	10.7
		3tj	4.7	24.4	29.1
		3tj Dis.	9.3	93.0	102.3
Benzene dimer	60	2tj	4	616	620
		3tj	208	4931	5139

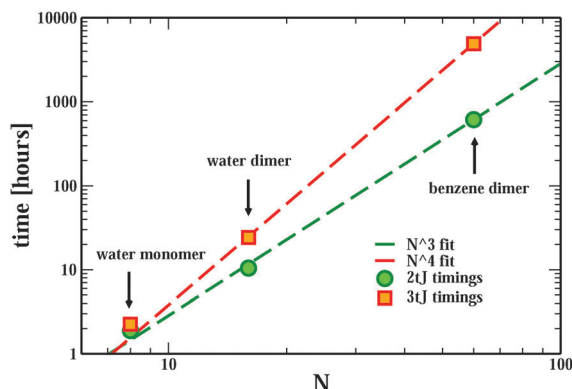


Fig. 1 The DMC timings (in hours) to reach 0.1 kcal mol⁻¹ error bars, for running on 64 cores, versus the total number of electrons N .

energy remains higher (by 0.0012 a.u. in the dimer), and the interaction energy produced in this way (-3.28 ± 0.1 kcal mol⁻¹) also deviates more from the reference. We thus recommend using a large fraction of energy in the VMC-optimisation cost function.

3.5 DMC time step

For completeness, we explored the DMC time step of 0.01 a.u. in addition to our standard conservative time-step setting of 0.005 a.u., used to avoid the extrapolation of energy to a zero time step.²⁵ A conclusive discussion of this point is, however, not possible because the error bars do not allow statements of statistical significance. Thus we also do not attempt zero time step extrapolations; we discuss the observations only qualitatively. In the ammonia dimer, we have observed a deterioration

of the final interaction energies in the case of 3tJ with an increase of the time step, whereas in the case of 2tJ, no apparent dependence arises. In the case of the water dimer, the calculation of energies using an increased time step shows that only the 3tJ Dis. scheme is able to approach the reference in the zero time step limit. In the ethene dimer, there is no significant dependence on the time step in either of the cases considered (2tJ and 3tJ). The time step of 0.005 a.u. is accurate enough for our purposes and it is used throughout the study.

3.6 ECP treatment in DMC

In order to reduce the numerical cost of the calculations, we have removed core electrons using ECPs. In the ammonia dimer, we have found that the T-moves⁴⁵ scheme used to treat ECPs in DMC produces more accurate results than the locality approximation⁴⁶ (where the error reaches ~ 0.2 kcal mol⁻¹), as expected, and we have thus combined T-moves with a short time step (0.005 a.u.).

3.7 Summary

To summarise the discussion related to the analysis of protocols and trade-offs relevant in the calculations of noncovalent interactions (between closed-shell complexes), we conclude that for general setups, the main group elements and the target ~ 0.3 kcal mol⁻¹ accuracy (in small systems, or formally per one bond), the use of the following scheme (2tJ protocol) is recommended: (i) single-determinant trial wave functions of Slater-Jastrow type using B3LYP orbitals and an aug-TZV basis set, (ii) an exhaustively optimised Jastrow factor (keeping the Slater determinant intact) with ee and eN terms, and, (iii) a FN-DMC

Table 3 The FN-DMC interaction energies E (kcal mol⁻¹) obtained by the new protocol based on the simplified Jastrow factor (2tJ, *cf.* Section 3), compared to the CCSD(T)/CBS reference interaction energies E_R (kcal mol⁻¹) and conservative 3tJ protocol data,²⁶ along with the corresponding differences Δ and statistics: mean error (ME, kcal mol⁻¹), mean unsigned error (MUE, kcal mol⁻¹) and relative unsigned error (RUE, %), for each subset and the whole considered set (total), respectively

Complex	E_R	$E/3tJ^d$	Δ_R	$E/2tJ^c$	Δ_R
(Teaching set)					
Ammonia dimer	-3.15 ^a	-3.10 \pm 0.06	-0.05	-3.21 \pm 0.09	0.06
Water dimer	-5.07 ^a	-5.26 \pm 0.08	0.19	-5.28 \pm 0.10	0.21
Hydrogen fluoride dimer	-4.62 ^b	-4.68 \pm 0.10	0.10	-4.77 \pm 0.12	0.19
Methane dimer	-0.53 ^a	-0.44 \pm 0.05	-0.09	-0.60 \pm 0.07	0.07
Ethene dimer	-1.48 ^a	-1.47 \pm 0.09	-0.01	-1.53 \pm 0.13	0.05
Ethene/ethyne	-1.50 ^a	-1.56 \pm 0.08	0.06	-1.54 \pm 0.14	0.04
		ME:	0.033	ME:	0.104
		MUE:	0.083	MUE:	0.104
		RUE:	4.78	RUE:	4.94
(Test set)					
Benzene-H ₂	-1.03 ^c	—	—	-1.08 \pm 0.10	0.04
Benzene-water	-3.29 ^a	-3.53 \pm 0.13	0.24	-3.21 \pm 0.15	-0.08
Benzene-methane	-1.45 ^a	-1.30 \pm 0.13	-0.15	-1.50 \pm 0.15	0.05
Benzene dimer T-shape	-2.71 ^a	-2.88 \pm 0.16	0.17	-2.53 \pm 0.23	-0.18
Adenine-thymine stacked	-11.66 ^a	—	—	-11.00 \pm 0.30	-0.66
				ME:	-0.167
				MUE:	0.203
				RUE:	4.44

^a Takatani *et al.*¹⁴ ^b CCSDT(Q), Řezáč *et al.*¹⁰ ^c This work. ^d Dubecký *et al.*²⁶

ground-state projection using the T-moves scheme⁴⁵ and a time step of 0.005 a.u. The error bars should be converged to 0.1–0.2 kcal mol^{−1} to obtain statistically meaningful results. To this end, DMC projection times of several thousands of a.u. with large walker ensembles are unavoidable.



























4 Benchmarks

The production FN-DMC results obtained by the scheme involving a simplified Jastrow term (2tj) are reported and compared to the CCSD(T)/CBS and more traditional protocol (3tj) data in Table 3. The mean error (ME) and the mean unsigned error (MUE) of the 2tj scheme, with respect to the CCSD(T)/CBS reference data, are both found to be about 0.1 kcal mol^{−1} respectively and the mean relative unsigned error (RUE) reaches 4.9%. The maximum interaction energy deviations with respect to the reference reach ~ 0.2 kcal mol^{−1}, observed in the hydrogen-bonded water and hydrogen fluoride dimers. For comparison, the MUE of the FN-DMC results obtained with the 3tj scheme amounts to 0.08 kcal mol^{−1} (RUE: 4.8%) and the maximum deviation of ≈ 0.2 kcal mol^{−1} is obtained only in the

case of the water dimer. This may be further improved by using the distinct Jastrow factor (*cf.* Table 1)²⁶ if required, as already mentioned. Nevertheless here we are interested in the computationally less expensive 2tj scheme. The remaining 2tj values differ by no more than 0.1 kcal mol^{−1} from the CCSD(T)/CBS reference.

The favourable performance of the 2tj scheme is further demonstrated when it comes to larger complexes, where the MUE amounts only to 0.2 kcal mol^{−1} (RUE: 4.4%). Overall, the data in Table 3 clearly show that the faster 2tj scheme is able to attain benchmark results, close to the subchemical accuracy in most systems. These results also presumably indicate that efficient FN-error cancellation²⁶ takes place in the considered set of complexes. On the other hand, if the 2tj scheme was used in larger complexes, a possible error-accumulation is possible in principle. Nevertheless the concept of chemical accuracy in these cases loses its importance since the absolute value of the interaction energy grows and what should be taken as measure of reliability of any theoretical approach is the relative energy deviation. The relative error of our approach does not noticeably increase with the size of the system which gives us confidence

Table 4 The FN-DMC interaction energies E (kcal mol^{−1}) of the A24 set, obtained by the 2tj protocol, compared to the CCSD(T)/CBS reference interaction energies E_R (kcal mol^{−1}) along with the corresponding differences Δ , illustrations⁴⁹ and statistics: mean error (ME, kcal mol^{−1}), mean unsigned error (MUE, kcal mol^{−1}) and relative unsigned error (RUE, %)

Label	Complex		E_R	$E/2tj$	Δ	Label	Complex		E_R	$E/2tj$	Δ
1	Water-ammonia		−6.493	−6.71 ± 0.07	−0.22	13	Ethene-ammonia		−1.374	−1.45 ± 0.07	−0.07
2	Water dimer		−5.006	−5.30 ± 0.05	−0.29	14	Ethene dimer		−1.09	−1.08 ± 0.07	0.01
3	HCN dimer		−4.745	−5.09 ± 0.08	−0.35	15	Methane-ethene		−0.502	−0.54 ± 0.06	−0.04
4	HF dimer		−4.581	−4.88 ± 0.05	−0.30	16	Borane-methane		−1.485	−1.48 ± 0.04	0.01
5	Ammonia dimer		−3.137	−3.30 ± 0.04	−0.17	17	Methane-ethane		−0.827	−0.85 ± 0.06	−0.03
6	Methane-HF		−1.654	−1.29 ± 0.07	0.37	18	Methane-ethane		−0.607	−0.68 ± 0.06	−0.07
7	Ammonia-methane		−0.765	−0.83 ± 0.06	−0.06	19	Methane dimer		−0.533	−0.63 ± 0.03	−0.10
8	Methane-water		−0.663	−0.57 ± 0.06	0.09	20	Ar-methane		−0.405	−0.18 ± 0.07	0.23
9	Formaldehyde dimer		−4.554	−4.90 ± 0.10	−0.35	21	Ar-ethene		−0.364	−0.19 ± 0.07	0.18
10	Water-ethene		−2.557	−2.55 ± 0.07	0.01	22	Ethene-ethyne		0.821	0.94 ± 0.07	0.12
11	Formaldehyde-ethene		−1.621	−1.70 ± 0.08	−0.07	23	Ethene dimer		0.934	1.06 ± 0.03	0.13
12	Ethyne dimer		−1.524	−1.74 ± 0.07	−0.21	24	Ethyne dimer		1.115	1.23 ± 0.03	0.11
										ME:	−0.05
										MUE:	0.15
										RUE:	12.1

for its use in large complexes. An exhaustive assessment of the general applicability and limits of the presented methods is one of the next goals of our work.

We note in passing that in the case of the benzene-H₂ complex, we report our own CCSD(T)/CBS reference value of $-1.033 \text{ kcal mol}^{-1}$ (Table 3), obtained by a standard CBS extrapolation/correction technique (*cf.* the Methods section) which improves upon the existing value of $-1.037 \text{ kcal mol}^{-1}$.⁴⁷ The obtained QMC value ($-1.08 \pm 0.1 \text{ kcal mol}^{-1}$) agrees (within the error bar) with the reference, as well as with the reported value of $0.96 \pm 0.08 \text{ kcal mol}^{-1}$ (ref. 48) obtained at a slightly different geometry.

Finally, we have extensively tested the new 2tj protocol on a set of 24 noncovalent complexes (A24).¹⁰ The results are summarized in Table 4. The overall MUE amounts to $0.15 \text{ kcal mol}^{-1}$, confirming our conclusions based on Table 3. One may observe that in general, the hydrogen-bound complexes deviate more from the reference than the mixed and dispersion-bound ones, showing the most favorable domain of the protocol applicability. The high RUE may be attributed to the high ratio between interaction energy amplitudes and noise. The overall performance of the 2tj protocol for the A24 group is nevertheless very good. Some of the cases with considerable deviations were further checked using the 3tj protocol (Table 5), which brings the interaction energies to within $\sim 0.1 \text{ kcal mol}^{-1}$ from the reference, as expected. The HCN dimer and the formaldehyde dimer represent the cases where the interaction energy remains off, even with the 3tj protocol. An analysis and fundamental understanding of these effects is beyond the scope of the current work, and remains as our next goal. We tentatively assign these deviations to the presence of multiple bonds in combination with hydrogen bonding, possibly causing nodal surface non-linearity effects³⁶ that may lead to the partial breakdown of the fixed-node error cancellation assumption.²⁶ On the other hand, the deviations still remain well below the chemical accuracy. The expected error margin of the 2tj protocol thus remains at $0.2\text{--}0.3 \text{ kcal mol}^{-1}$ per bond (formally).

Table 5 The FN-DMC interaction energies E (kcal mol^{-1}) of selected complexes from A24, obtained by the 2tj and 3tj protocols, the corresponding differences Δ , and the CCSD(T)/CBS reference interaction energies E_R (kcal mol^{-1})

Label	Complex	E_R	Protocol	E	Δ
3	HCN dimer	−4.745	2tj	-5.09 ± 0.08	−0.35
			3tj	-5.13 ± 0.06	−0.39
6	Methane-HF	−1.654	2tj	-1.29 ± 0.07	0.37
			3tj	-1.54 ± 0.04	0.11
9	Formaldehyde dimer	−4.554	2tj	-4.90 ± 0.10	−0.35
			3tj	-4.88 ± 0.06	−0.32
12	Ethyne dimer	−1.524	2tj	-1.74 ± 0.07	−0.21
			3tj	-1.62 ± 0.05	−0.10
21	Ethene-Ar	−0.364	2tj	-0.19 ± 0.07	−0.18
			3tj	-0.22 ± 0.04	−0.14

5 Conclusions

The analysis of the QMC FN-DMC-based protocols and related trade-offs provided in this work revealed a favourable computational scheme based on a simplified explicit correlation Jastrow term that results in a reduced computational cost scaling. The tests on a number of complexes including up to a stacked DNA base pair show good performances and nearly sub-chemical accuracy with respect to CCSD(T)/CBS, consequently enabling an easier access to reliable estimates of the interaction energies in large complexes.

Since the QMC is not limited to single-reference and “gas-phase” complexes and the reported protocol relies on the fixed-node error cancellation, it may also find application in cases that are intrinsically difficult for mainstream-correlated wavefunction approaches. These include estimates of noncovalent interactions between open-shell organometallic and/or metal-organic systems (using simple multi-determinant trial functions instead of a single Slater determinant), studies of periodic models like physisorption on metal surfaces and 2D materials, or the prediction of noncovalent crystal stability.³¹ In the domain of noncovalent interactions, the quantum Monte Carlo FN-DMC method therefore appears to be very promising for its benchmark accuracy, low-order polynomial scaling with the system complexity, low memory requirements and nearly ideal scaling across thousands of processors⁵⁰ in parallel supercomputing environments.

6. Methods

The geometries of the studied complexes were taken from the S22¹² and A24¹⁰ sets except for the benzene-H₂⁴⁷ complex. ECPs with the corresponding basis sets developed by Burkatzki *et al.*⁵¹ were used throughout the work, with the exception of the H where a more recent version was used.⁵² The augmentation functions were taken from the corresponding Dunning bases.⁵³ Single-determinant Slater-Jastrow²⁰ trial wave functions were constructed using B3LYP or HF orbitals from GAMESS.⁵⁴ The used Schmidt-Moskowitz⁵⁵ homogeneous and isotropic Jastrow factors,²⁰ including either the electron-electron and electron-nucleus terms (2tj), or 2tj with electron-electron-nucleus terms in addition (3tj), were expanded in a fixed basis set of polynomial Padé functions.²¹ The parameters of the positive definite Jastrow factor were optimised by the Hessian driven VMC optimisation of at least 10×10 iterations (*i.e.* a full VMC energy calculation after each 10 optimisation steps on a fixed walker population), using the variance optimisation⁵⁶ or a linear combination⁴⁴ of energy (95%) and variance (5%) as a cost function. The optimised trial wave functions were subsequently used in the production FN-DMC runs performed with a time step of $0.01/0.005 \text{ a.u.}$ within the locality approximation^{18,46} (LA, for testing purposes) or using the T-moves scheme⁴⁵ for the treatment of ECPs beyond LA. The target walker populations ranged from 5k (for small systems) up to about 20k (for the largest system). All QMC calculations were performed using the code QWalk.⁵⁷

The reference interaction energy for the benzene-H₂ complex was estimated by the basis set superposition error corrected standard CBS extrapolation technique¹² from HF/aug-cc-pV5Z, MP2/aug-cc-pVQZ/aug-cc-pV5Z and CCSD(T)/aug-cc-pVQZ results.

Acknowledgements

M.D. is grateful to Claudia Filippi for sharing an improved effective core potential for H and valuable discussions. The support from the Operational Programme (OP) Research and Development (R&D) for Innovations – European Regional Development Fund (ERDF, project CZ.1.05/2.1.00/03.0058) and the OP Education for Competitiveness – European Social Fund (projects CZ.1.07/2.3.00/30.0004 and CZ.1.07/2.3.00/20.0058) is gratefully acknowledged. This work was supported by the grants P208/10/1742 (P.J.) and P208/12/G016 (M.O.), from the Czech Science Foundation. R.D. acknowledges support from grants APVV-0207-11 and VEGA (2/0007/12). L.M. is supported by NSF OCI-0904794, DMR-1410639 and ARO W911NF-04-D-0003-0012 grants and by XSEDE computer time allocation at TACC. The calculations were in part performed at the Slovak infrastructure for high-performance computing (projects ITMS 26230120002 and 26210120002) supported by the OP R&D funded by the ERDF. Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme “Projects of Large Infrastructure for Research, Development, and Innovations” (LM2010005), is greatly appreciated.

References

- 1 P. Hobza and K. Müller-Dethlefs, *Non-Covalent Interactions*, The Royal Society of Chemistry, 2010.
- 2 K. E. Riley, M. Pitoňák, P. Jurečka and P. Hobza, *Chem. Rev.*, 2010, **110**, 5023–5063.
- 3 S. Grimme, C. Diedrich and M. Korth, *Angew. Chem., Int. Ed.*, 2006, **45**, 625–629.
- 4 L. M. Salonen, M. Ellermann and F. Diederich, *Angew. Chem., Int. Ed.*, 2011, **50**, 4808–4842.
- 5 V. Georgakilas, M. Otyepka, A. B. Bourlinos, V. Chandra, N. Kim, K. C. Kemp, P. Hobza, R. Zbořil and K. S. Kim, *Chem. Rev.*, 2012, **112**, 6156–6214.
- 6 P. Lazar, F. Karlický, P. Jurečka, M. Kocman, E. Otyepková, K. Šafářová and M. Otyepka, *J. Am. Chem. Soc.*, 2013, **135**, 6372–6377.
- 7 M. S. Schuurman, S. R. Muir, W. D. Allen and H. F. Schaefer III, *J. Chem. Phys.*, 2004, **120**, 11586–11599.
- 8 P. Hobza, *Acc. Chem. Res.*, 2012, **45**, 663–672.
- 9 K. Raghavachari and G. W. Trucks, *Chem. Phys. Lett.*, 1989, **157**, 479–483.
- 10 J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2013, **9**, 2151–2155.
- 11 L. Šimová, J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2013, **9**, 3420–3428.
- 12 P. Jurečka, J. Šponer, J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.
- 13 R. Podeszwa, K. Patkowski and K. Szalewicz, *Phys. Chem. Chem. Phys.*, 2010, **12**, 5974–5979.
- 14 T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall and C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 144104.
- 15 J. Řezáč, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2011, **7**, 2427–2438.
- 16 J. Brndiar and I. Štich, *J. Chem. Theory Comput.*, 2012, **8**, 2301–2309.
- 17 M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, *Chem. Rev.*, 2012, **112**, 632–672.
- 18 W. M. C. Foulkes, L. Mitas, R. J. Needs and G. Rajagopal, *Rev. Mod. Phys.*, 2001, **73**, 33–83.
- 19 B. M. Austin, D. Y. Zubarev and W. A. Lester, *Chem. Rev.*, 2012, **112**, 263–288.
- 20 R. Jastrow, *Phys. Rev.*, 1955, **98**, 1479–1484.
- 21 M. Bajdich and L. Mitas, *Acta Phys. Slovaca*, 2009, **59**, 81–168.
- 22 D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.*, 1980, **45**, 566–569.
- 23 M. Mella and J. B. Anderson, *J. Chem. Phys.*, 2003, **119**, 8225–8228.
- 24 C. Diedrich, A. Lüchow and S. Grimme, *J. Chem. Phys.*, 2005, **123**, 184106.
- 25 M. Körth, A. Lüchow and S. Grimme, *J. Phys. Chem. A*, 2008, **112**, 2104–2109.
- 26 M. Dubecký, P. Jurečka, R. Derian, P. Hobza, M. Otyepka and L. Mitas, *J. Chem. Theory Comput.*, 2013, **9**, 4287–4292.
- 27 N. A. Benedek, I. K. Snook, M. D. Towler and R. J. Needs, *J. Chem. Phys.*, 2006, **125**, 104302.
- 28 B. Santra, A. Michaelides, M. Fuchs, A. Tkatchenko, C. Filippi and M. Scheffler, *J. Chem. Phys.*, 2008, **129**, 194111.
- 29 J. Ma, D. Alfé, A. Michaelides and E. Wang, *J. Chem. Phys.*, 2009, **130**, 154303.
- 30 A. Tkatchenko, D. Alfé and K. S. Kim, *J. Chem. Theory Comput.*, 2012, **8**, 4317–4322.
- 31 K. Hongo, M. A. Watson, R. S. Sánchez-Carrera, T. Iitaka and A. Aspuru-Guzik, *J. Phys. Chem. Lett.*, 2010, **1**, 1789–1794.
- 32 K. Hongo, T. C. Nguyen and R. Maezono, *J. Chem. Theory Comput.*, 2013, **9**, 1081–1086.
- 33 D. Alfé, A. P. Bartók, G. Csányi and M. J. Gillan, *J. Chem. Phys.*, 2013, **138**, 221102.
- 34 A. Ambrosetti, D. Alfé, R. A. DiStasio and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2014, **5**, 849–855.
- 35 F. R. Petruziello, J. Toulouse and C. J. Umrigar, *J. Chem. Phys.*, 2012, **136**, 124116.
- 36 K. Rasch, S. Hu and L. Mitas, *J. Chem. Phys.*, 2012, **140**, 041102.
- 37 M. Dubecký, R. Derian, L. Horváthová, M. Allan and I. Štich, *Phys. Chem. Chem. Phys.*, 2011, **13**, 20939–20945.
- 38 L. Horváthová, M. Dubecký, L. Mitas and I. Š. Štich, *Phys. Rev. Lett.*, 2012, **109**, 053001.
- 39 J. Kolorenč, S. Hu and L. Mitas, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **82**, 115108.

- 40 M. C. Per, K. A. Walker and S. P. Russo, *J. Chem. Theory Comput.*, 2012, **8**, 2255–2259.
- 41 F. Schautz and C. Filippi, *J. Chem. Phys.*, 2004, **120**, 10931–10941.
- 42 M. Dubecký, R. Derian, L. Mitás and I. Štich, *J. Chem. Phys.*, 2010, **133**, 244301.
- 43 A. Zen, Y. Luo, S. Sorella and L. Guidoni, *J. Chem. Theory Comput.*, 2013, **9**, 4332–4350.
- 44 C. J. Umrigar and C. Filippi, *Phys. Rev. Lett.*, 2005, **94**, 150201.
- 45 M. Casula, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2006, **74**, 161102.
- 46 L. Mitás, E. L. Shirley and D. M. Ceperley, *J. Chem. Phys.*, 1991, **95**, 3467–3475.
- 47 M. Rubeš and O. Bludský, *ChemPhysChem*, 2009, **10**, 1868–1873.
- 48 T. D. Beaudet, M. Casula, J. Kim, S. Sorella and R. M. Martin, *J. Chem. Phys.*, 2008, **129**, 164711.
- 49 “Schrödinger, LLC”, The PyMOL Molecular Graphics System, Version 1.3r1, 2010.
- 50 R. J. Needs, M. D. Towler, N. D. Drummond and P. L. Riós, *J. Phys.: Condens. Matter*, 2010, **22**, 023201.
- 51 M. Burkatzki, C. Filippi and M. Dolg, *J. Chem. Phys.*, 2007, **126**, 234105.
- 52 M. Dolg and C. Filippi, *Private communication*.
- 53 T. H. Dunning Jr., *J. Chem. Phys.*, 1989, **90**, 1007–1023.
- 54 M. W. Schmidt, K. K. Baldridge and J. A. Boatz, *et al.*, *J. Comput. Chem.*, 1993, **14**, 1347–1363.
- 55 J. W. Moskowitz and K. E. Schmidt, *J. Chem. Phys.*, 1992, **97**, 3382–3385.
- 56 N. D. Drummond and R. J. Needs, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2005, **72**, 085124.
- 57 L. K. Wagner, M. Bajdich and L. Mitás, *J. Comput. Phys.*, 2009, **228**, 3390–3404.