

Analysis of the basis set superposition error in molecular dynamics of hydrogen-bonded liquids: application to methanol

Marc Van Houteghem, Toon Verstraelen, An Ghysels, Louis Vanduyfhuys, Michel Waroquier, and Veronique Van Speybroeck^{a)}

*Center for Molecular Modeling, QCMM Alliance Ghent-Brussels,
Ghent University, Technologiepark 903, B-9052 Zwijnaarde,
Belgium*

An efficient protocol is presented to compensate for the basis set superposition error (BSSE) in DFT molecular dynamics (MD) simulations using localized Gaussian basis sets. We propose a classical correction term that can be added a posteriori to account for BSSE. It is tested to what extension this term will improve radial distribution functions (RDFs). The proposed term is pairwise between certain atoms in different molecules and was calibrated by fitting reference BSSE data points computed with the counterpoise method. It is verified that the proposed exponential decaying functional form of the model is valid. This work focuses on hydrogen-bonded liquids, i.e. methanol, and more specific on the intermolecular hydrogen bond, but in principle the method is generally applicable on any type of interaction where BSSE is significant. We evaluated the relative importance of the Grimme-dispersion versus BSSE and found that they are of the same order of magnitude, but with an opposite sign. Upon introduction of the correction, the relevant RDFs, obtained from MD, have amplitudes equal to experiment.

^{a)}Electronic mail: veronique.vanspeybroeck@ugent.be

I. INTRODUCTION

A radial distribution function (RDF) is the elementary tool used to extract the structural information from computer simulations of molecular liquids.¹ The RDF $g(r)$ of a fluid describes how the liquid density varies as a function of the distance from a reference molecule. It reflects the correlations in the distribution of the molecules of the fluid arising from the forces they exert on each other. The RDF is a measure for the probability to find an atom pair separated by a distance r , relative to the probability one would expect for a uniform distribution at the same density.

In this work the RDFs of liquids are studied with the help of molecular dynamics (MD) simulations, based on density functional theory (DFT).^{2,3} We focus on a specific category of molecules which interact with each other by means of weak covalent interactions. A well-known example of this type of bonding is the hydrogen bond, which occurs in e.g. water, methanol or ethanol. In this study, we concentrate on methanol in its liquid form. Methanol is widely used in chemical industry, as a solvent or fuel. It is the smallest alcohol consisting of a hydrogen-bonding hydroxyl group and a hydrophobic methyl group. The characteristic alcohol group allows methanol to form hydrogen bonds that dominate the structural and dynamical behaviour of the liquid phase. Because of the hydrogen bonding, the O-H bonds are elongated^{4,5} in the liquid phase with respect to the gas phase.^{4,6,7}

The liquid phase of methanol is extensively studied. The RDFs of methanol have been determined experimentally⁸⁻¹¹ and have been computed by means of full ab initio MD,^{5,12-15} mixed empirical and ab initio interactions^{16,17} and force-fields.¹⁸⁻²⁶ Often, the first intermolecular peak in the RDF of methanol as well of water is overestimated compared to experiment.^{12,13,15-17,23-25,27-30} The search for plausible explanations of this quasi systematic overestimation forms the onset of a long debate, but until now it is not yet unraveled and what suits for one particular case does not suit for another case. When computing RDFs from molecular simulations based on first principles (e.g. DFT), several parameters may affect the final result: system-size effects,^{13,29,31} basis set superposition error (BSSE), dispersion interaction, differences in exchange-correlation functional,³¹⁻³³ differences in basis sets,³¹ nuclear quantum effects, etc. Unfortunately, it is not yet clear what the net impact really is from each ingredient on the RDF, and in most cases the corrections work in opposite sense. It means that there is some compensation effect: if we correct the RDF for one miss-

ing ingredient, it can be canceled by another. This implies that one should be very careful when correcting for one specific item. It can happen that the final result could be worse and thus it is very important to get detailed insight into the various factors contributing to the overall result of the RDF.

- (i) Nuclear quantum effects in liquid water are commonly investigated by methods based on Feynman’s path-integral molecular dynamics simulations to calculate finite-temperature equilibrium properties. It is well known that quantum effects become more significant in the intramolecular region. Quantum statistical mechanical simulations are able to reproduce a quantitative agreement with experiment for the heights and broadening of the intramolecular O-H and H-H peaks in liquid water.³⁴ A general feature is that nuclear quantum effects make the structure of the liquid water softer.^{35–40} One notices a broadening and lowering of the RDF compared to the distribution generated from classical molecular dynamics simulations.^{34,41–43} An alternative is the Car-Parrinello molecular dynamics method,⁴⁴ which is based on the description of the intermolecular interactions at the density functional level of theory. The success of COMPD simulations in water depends how accurate the density based functional describes the Born-Oppenheimer potential energy surface for the description of the hydrogen bond.²⁸
- (ii) The choice of the exchange-correlation functional does affect the density of the liquid and the radial distribution function. There is a tendency that BLYP produces weaker hydrogen bonds (underbinding) than PBE (overbinding).²⁷ The global performance of a specific functional cannot be assessed without taking into account the dispersion interaction. Therefore, a quantification of each effect separately is recommended to better understand the global atomic and molecular interactions.
- (iii) Dispersion interactions are indispensable to predict correct densities and RDFs, but many models exist to describe them and it is not always clear what impact these differences have on the RDFs. Most models work with empirical pairwise interatomic potentials of the C^6R^{-6} form with damping terms, which can differ from model to model.^{45–48} In DFT-D3 of Grimme⁴⁷ the empiricism of the model has been largely removed by fitting as much as possible on first principles derived properties. In addition eight-order terms have been added, as well as a three-body correction. In general,

more and more molecular dispersion interactions have been proposed in the literature determined from ab initio electronic structure calculations and in this way largely reducing the empirical input.^{49,50} Their overall effect is that they push the density in the correct direction, but – depending on the individual cases – the RDFs remain overstructured.

This work focuses on the role of BSSE on the RDFs. Most of the published work on this item focused on the impact of dispersion on the radial distribution functions, and as already reported they represent a significant step toward experimental agreement,^{27,28} although some discrepancies remain. As most of the simulations have been applied to liquid water, one should be very careful to extrapolate their conclusions to methanol. A combined study of the RDFs in methanol taking into account the two main effects influencing the interatomic potential : BSSE + dispersion , is indispensable in the attempt to reproduce correctly the RDF of methanol.

The BSSE results from the approximation of using a limited number of localized basis functions to describe atomic and molecular orbitals. When two molecules are far apart, they can only use their own basis functions. But when they approach each other, they may lower their energy by using each other’s basis functions. This causes an artificial strengthening of the intermolecular interactions and artificial shortening of intermolecular distances due to the overlap between basis functions. The effect is more pronounced for smaller basis sets. Note that BSSE does not occur with plane wave basis sets as they are not associated with any particular atom.

The BSSE is especially prominent in systems with intermolecular hydrogen-bonds,^{51–57} and is thus expected to influence the RDF of methanol. It has been shown that the calculation of the BSSE corrections can change their potential energy curves and surfaces,^{58–62} and can alter the bond distances.^{58,60,61,63,64} It is well known that BSSE can be a significant fraction of the binding energy of the water dimer, which could affect liquid simulations.³¹ Simon et al. analyze the effect of BSSE on the geometries and vibrational frequencies of 15 H-bonded systems at the B3LYP and MP2 levels of theory, using the 6-31++G(d,p) basis set.⁵³ It is shown that the intermolecular H-bond distance increases when correcting for BSSE, whereas the intermolecular H-bond stretching frequency decreases, and that the observed changes depend on the level of theory. The effect of BSSE on structural and dynamical properties of the liquid has not been investigated previously. In fact, there is little

information available concerning the performance of a Gaussian basis set in the context of MD simulations of liquids. So the influence of BSSE on RDFs remains an open question.

New in the model – presented in this work – is the proposal of a systematic correction of the BSSE with a parametrized classical force-field energy term. The parameters are fitted by calculating the BSSE error for a series of fluid configurations. These configurations have been created by *ab initio* MD simulations. Finally, MD simulations are performed including the BSSE correction and dispersion. The RDFs computed from MD runs are compared with the experiment before and after the corrections, and it is assessed to what extent the correspondence with experiment due to these corrections has improved. This geometrical approach to account for BSSE is supported by a very recent paper by Kruse⁶⁵ where a semi-empirical correction is introduced, constructed from overlap integrals over Slater functions and which employs computed measures for the incompleteness of the basis set.

In the following section, the technical details of the quantum MD simulations and the computational methodology to calculate the RDF and BSSE are given. Section III introduces the new BSSE correction model, i.e. its mathematical form and the calibration of the parameters. Section IV presents the results, which are further discussed in section V. In the last section the most relevant conclusions – in view of future work – are drawn.

II. COMPUTATIONAL DETAILS

A. *Ab initio* MD

In this work, all molecular dynamics calculations were performed with the CP2K code.⁶⁶ For the *ab initio* MD, the Quickstep module of CP2K⁶⁷ was used, employing the hybrid Gaussian and plane-wave (GPW) density functional method with a BLYP gradient-corrected functional.^{68,69} The dual GPW basis set consisted of a triple zeta TZVP Gaussian type orbital basis for the real space representation and an auxiliary plane wave basis expanded to a 400 Ry cut-off to compute the long-range periodic electrostatic interactions in the reciprocal space.⁷⁰ The core electron states are represented by the norm conserving Goedecker-Teter-Hutter (GTH) pseudopotential.^{31,71,72} The BLYP functional was chosen because of its good description of the structure and dynamics of water³² where hydrogen bonds are, as in liquid methanol, the dominant interactions. Furthermore, it has been shown that the BLYP

functional gives a proper description of solvation of methanol in water.⁷³

For the ab initio MD simulations that include the correction for the BSSE with the force-field energy terms, we used FIST, the molecular mechanics (MM) module of CP2K. Periodic boundary conditions were employed using the minimum image convention. At each timestep the total energy is computed as the sum of the ab initio energy and the energy calculated by the classical correction term.

Standard DFT approaches do not account for dispersion interactions.^{74,75} Dispersion is a long-range van der Waals interaction between electron clouds arising from instantaneous transition dipoles. These dispersion interactions are a quantum-mechanical electron correlation effect that is completely absent when considering classical charge distributions. To test the effects of dispersion on the fluid structure of methanol, we also performed MD simulations with dispersion-corrected DFT using the DFT-D3 scheme of Grimme for the BLYP functional.^{45-47,76} The combination of the two corrections for the BSSE and dispersion then provides us four different potential energy surfaces: no corrections (DFT), corrected for BSSE with the counterpoise (CP) correction (DFT + CP) as explained in section II C, corrected for dispersion (DFT-D3), corrected for BSSE and dispersion (DFT-D3 + CP).

Liquid methanol at room temperature was modeled in a periodic cubic simulation cell containing 129 methanol molecules. The simulation cell size, 20.58 Å, was chosen to correspond with the experimental density of 0.787 kg/m³. After equilibration, the MD runs lasted 8.4 ps at a temperature of 300 K with a Nosé thermostat in the canonical (NVT) ensemble.⁷⁷ The Verlet integrator was used with an integration time step of 1 fs.

B. The radial distribution function

The radial distribution function is defined as

$$g(r) = \frac{dn(r)}{dn_0(r)} \tag{1}$$

and represents the number of particle pairs $dn(r)$ in the spherical shell with radius r and $r + dr$, with r the interparticle distance (relative coordinate), relative to the number of particle pairs $dn_0(r)$, obtained in an uncorrelated ideal gas. With ρ_0 as the density of particles in the ideal gas case, $dn_0(r)$ may be expressed as $\rho_0 4\pi r^2 dr$.

In practice, the space around a given atom is discretized in concentric spherical shells

with width δr . The number of atoms n_i in each shell with finite volume $\delta V_{\text{shell},i}$ is counted:

$$\delta V_{\text{shell},i} = \frac{4}{3}\pi (r_i + \delta r)^3 - \frac{4}{3}\pi r_i^3 \simeq 4\pi r_i^2 \delta r \quad (2)$$

A histogram of the n_i values is built during the MD run and $g(r_i)$ is approximated as $n_i / \rho_0 4\pi r_i^2 \delta r$.

C. Counterpoise correction

In this section we describe a method to correct for BSSE in computations with localized basis sets that are centered at the nuclei. Atom-centered basis sets have a number of advantages. The most important one is that with relatively few basis functions one is able to describe the molecular orbitals to high accuracy. Such basis sets cannot be systematically enlarged in a simple way and the severity of the problem is proportional to the incompleteness of the basis sets.

From the suggested methods to correct for BSSE,^{78–83} the most widely used approach is the counterpoise method of Boys and Bernardi.⁸³ The uncorrected interaction energy of a complex AB can be defined as:

$$\Delta E_{\text{int}}(AB) = E^{AB}(AB) - E^A(A) - E^B(B) \quad (3)$$

where the superscripts denote the basis used and the symbol in parentheses denotes the chemical system considered. Boys and Bernardi suggested to correct Eq.(3) by estimating the amount of artificial stabilization for monomer A as the energy shift caused by a change of basis $A \rightarrow AB$:

$$E_{\text{BSSE}}(A) = E^{AB}(A) - E^A(A), \quad (4)$$

and similarly for monomer B. To evaluate the energy of monomer A in the AB basis, one places all the basis functions of monomer B on the atomic centers of monomer B while omitting the electrons and the nuclear charges of monomer B. The B-basis in the E^{AB} computation and the A-basis in the E^A are called *ghost atoms*. The energy of monomer A in the AB basis is lower than the energy of monomer A in the A basis ($E_{\text{BSSE}}(A) < 0$). The CP-correction is then defined as:

$$\Delta E_{\text{CP}} = E^A(A) + E^B(B) - E^{AB}(A) - E^{AB}(B). \quad (5)$$

When the estimated errors $E_{BSSE}(A)$ and $E_{BSSE}(B)$ are subtracted from the uncorrected interaction energy, we obtain the *counterpoise-corrected* interaction energy,

$$\Delta E_{int}^{CP}(AB) = E^{AB}(AB) - E^{AB}(A) - E^{AB}(B). \quad (6)$$

In this derivation we have assumed that the geometries of monomers A and B do not change as they approach each other and form the bimolecular complex.

The CP-computations are carried out on pairs of methanol molecules. These pairs are selected from snapshots created during the MD run. For the CP-computations the same specifications (CP2K software package, basis set, electronic structure method, ...) were used as those of the MD simulations of the methanol fluid. To accelerate convergence in the CP-computations of the pairs, the box size is decreased from 20.58 Å (in the MD run) to 10.0 Å.

III. DERIVATION OF THE CP FORCE-FIELD CORRECTION MODEL

A. Model

For the CP force field correction term, we consider a pairwise model between certain atoms in two different molecules. Because the BSSE converges by definition to zero at large distances, an exponential decaying model is proposed. The electron density of a molecule in vacuum exhibits an exponential decay. BSSE follows the same trend as the overlap of the electron density of two molecules and also decays exponentially. The goal of the correction term is to account for the artificial BSSE-stabilization, which may be responsible for the overestimation of the first intermolecular peak of the RDF. Therefore, the CP-model has to be repulsive.

We consider an intermolecular pairwise CP-correction model. To reduce the number of parameters, atom types are introduced. With each pair of atom types t , a unique set of parameters (A_t, B_t) is associated:

$$\Delta E_{CP}^{\text{mod}} = \sum_{\mu=1}^{N-1} \sum_{\nu=\mu+1}^N \sum_{\substack{i \in M_\nu \\ j \in M_\mu}} A_{t(ij)} \exp(-B_{t(ij)} r_{ij}), \quad (7)$$

where r_{ij} is the distance between the atoms i and j , M_ν is a set of atomic indices of molecule ν , μ and ν are molecule indexes and N is the total number of molecules. An example of

an intermolecular pair of atom types t is the pair O-H_m, consisting of the methyl hydrogen of monomer 1 (methanol molecule 1) and the oxygen of monomer 2 (methanol molecule 2). For liquid methanol, $M_\nu = 6 \forall \nu$, and $N = 129$ in our MD simulations.

The model is not to be confused with the functional form as it was introduced by Feller⁸⁴ to extrapolate energies, as well as properties, to the complete basis set (CBS) limit.^{85–88} In this work, the parameters of the proposed model (i.e. $A_{t(ij)}$ and $B_{t(ij)}$) are fitted to reference data, which will be further explained below, and validated by comparing calculated RDFs with experiment.

B. Parameter calibration

To calibrate the parameters of the model (Eq.(7)), reference data are needed. These training data are obtained from molecular dynamics simulations. A snapshot of the fluid is selected every 400 steps (= 400 fs) of the MD simulation (with dispersion, without CP-correction), corresponding to a time τ greater than the velocity auto correlation time of the centers of mass (COMs) of the molecules to ensure statistically independent samples. The correlation time was computed with the MD-TRACKS program^{89,90} and amounted to 86 fs. In this way, 21 frames of the MD simulation are selected. At each snapshot, the distances between the COMs of every possible pair of two methanol molecules are computed. The pairs of molecules for which this distance is smaller than 4 Å, are selected for the computation of reference data. Following this procedure, a total number of 3850 pairs are obtained. We then calculate the CP-correction for each of these pairs with Eq.(5).

The parameters ($A_{t(ij)}, B_{t(ij)}$) are estimated from the training data set of CP-computations. Several fitting procedures, linear and non-linear, have been considered. After careful consideration, however, only the non-linear optimized parameters are retained. For this non-linear fit, the in-house developed software package FFit2⁹¹ is used. FFit2 is a Python library for the calibration of force-field parameters based on electronic structure computations. Specifically, we make use of the submodule MFit2, which can be used to combine several types of calibration schemes for valence force fields.

In this work, the parameters are optimized with a non-linear least-squares fitting procedure based on the Levenberg-Marquardt algorithm.^{92,93} The Levenberg-Marquardt algorithm is an iterative technique for the minimization of a multivariate function that is expressed as

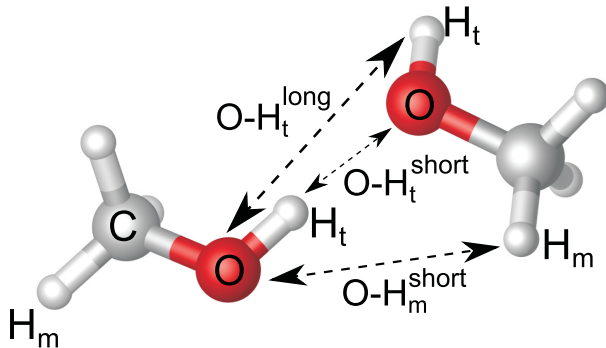


FIG. 1. Labels of the methanol atoms (C,O,H_m,H_t) and the most relevant intermolecular O-H pairs.

the sum of squares of non-linear real-valued functions.

Once calibrated parameters ($A_{t(ij)}, B_{t(ij)}$) are obtained, a new MD simulation is performed with the Quickstep and FIST modules of CP2K. The energy in the MD run is computed as the DFT-energy (Quickstep) plus the counterpoise force-field term (FIST).

IV. RESULTS

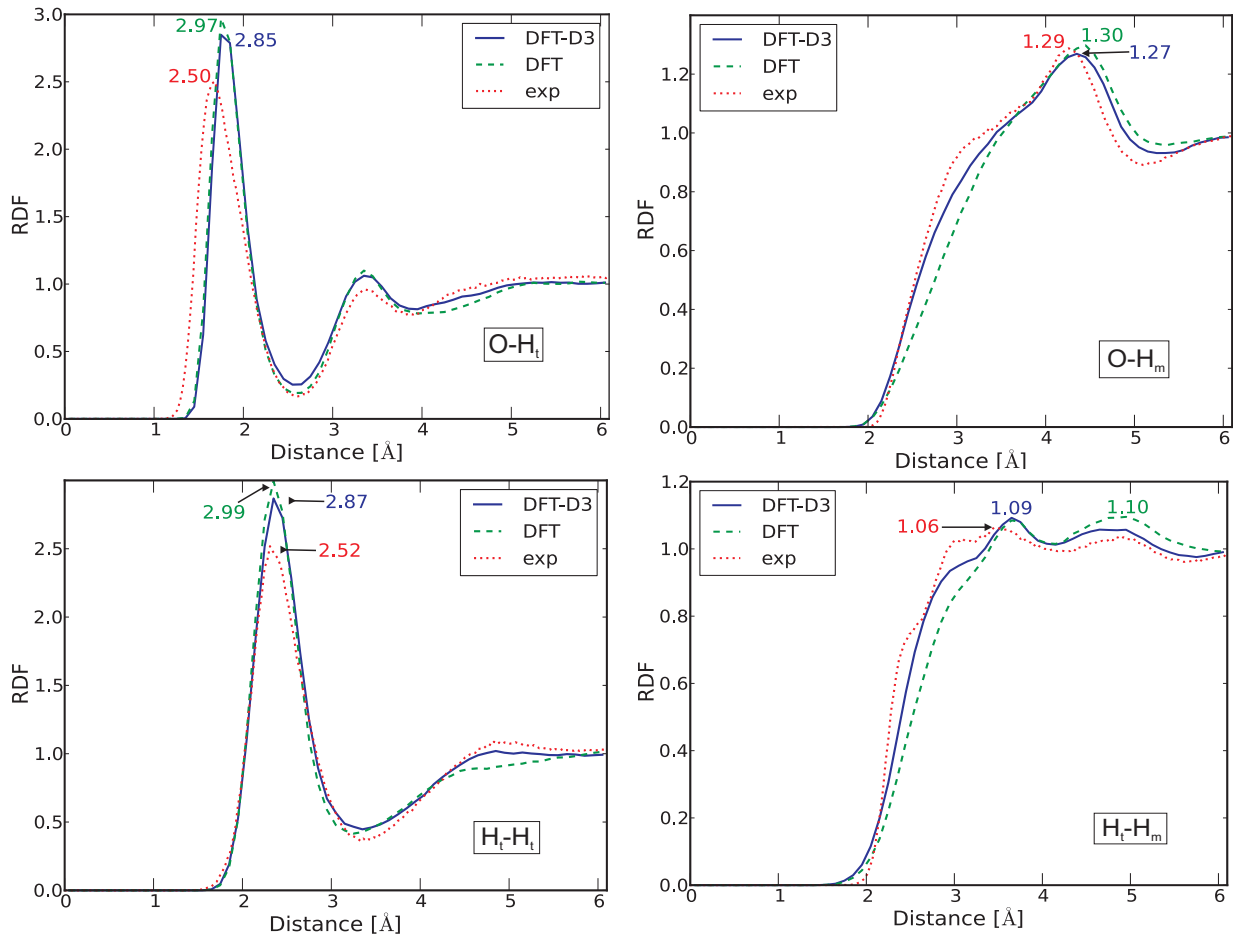
In this section, the intermolecular RDFs computed from MD runs with and without CP-correction are presented. The relevant RDFs between oxygen and the different types of hydrogen as well the intermolecular hydrogen radial distributions, are discussed in the main manuscript; all other RDFs are given in the Supporting Information (S.I.).

Fig. 1 clarifies the atom types of the methanol molecules as used in this paper. We distinguish between the hydroxyl hydrogen (H_t) and the methyl hydrogen (H_m), as they are situated in different chemical environments. We introduce two distances which depend on the intermolecular geometry: O-H_t^{short} (the shortest intermolecular O-H_t distance between two methanol dimers) and O-H_t^{long} (the longest intermolecular O-H_t distance).

A. RDFs without correction

The RDFs are first computed without correcting for BSSE. Fig. 2 shows the results for MD runs with and without dispersion correction. Inspection of Fig. 2 reveals that (i) the peaks of the O-H_t and the H_t-H_t RDF are too high compared to experiment, (ii) the first

FIG. 2. Radial distribution functions for oxygen-hydrogen and hydrogen-hydrogen with BLYP and BLYP-D3. Experimental curves from.^{8,9}



peaks are positioned at too large distances, and that (iii) the radial distribution is slightly overstructured. The inclusion of dispersion does not improve the correspondence with experiment. The absolute and relative characteristic differences in RDFs are numerically reported in Table I.

As can be seen from Fig. 2, the intermolecular peaks of the RDFs involving the hydroxyl hydrogen H_t are overestimated, while those with the methyl hydrogen H_m are not. This immediately suggests that a distinction between H_t and H_m should be made when the parameters (A_t, B_t) of the CP-correction model are fitted.

The deficiency encountered in the position of the first peak, in particular in the O- H_t RDF, will be discussed later.

TABLE I. Amplitudes of the first peak of the RDFs for O-H_t and H_t-H_t and the relative error compared to experiment.^{8,9}

	exp.	DFT	Δ_{DFT}	DFT-D3	$\Delta_{\text{DFT-D3}}$
O-H _t	2.50	2.97	15.7%	2.85	12.1%
H _t -H _t	2.52	2.99	15.8%	2.87	12.2%
O-H _m	1.29	1.30	0.7%	1.27	1.4%
H _t -H _m	1.06	1.09	2.9%	1.09	2.7%

B. Energy scan of two methanol monomers

To gain more insight in the pairwise interatomic interaction and to disentangle the contributions of dispersion and BSSE on the total interaction energy, an energy scan is performed of a complex, consisting of two methanol monomers, as a function of the intermolecular separation O-H_t^{short} distance to construct the potential energy surface (PES). First, the geometry of the two methanol molecules is optimized in the gasphase with CP2K, to form a stable dimer at a hydrogen bond distance of about 1.9 Å. Next, the O-H_t^{short} distance is gradually varied with a constrained geometry optimization with a step size of 0.1 Å ranging from 0.6 Å to 4.5 Å to obtain the energy scan in Fig. 3. The residual fluctuations that are visible in the curves are due to the well known egg-box effect.⁹⁴⁻⁹⁶ The same computational settings are used for the scan as for the MD runs (section II A) for consistency. This energy scan has been performed in four cases: with and without dispersion and with and without CP correction. The goal of this scan is threefold: (i) to study the interaction between two methanol molecules in more detail, (ii) to validate the proposed exponential model for the BSSE, (iii) to assess the relative importance of BSSE versus dispersion interaction. In Fig. 3 the total energy of the dimer E_{AB} is plotted against the O-H_t^{short} distance. All energies are relative with respect to the DFT binding energy of two isolated methanol monomers.

Inclusion of the D3 dispersion correction causes a slight shift to a smaller equilibrium distance. In contrast, BSSE, by means of the CP-correction term, creates the opposite effect on the equilibrium distance, but makes the potential some 3 kJ/mol less attractive. To get more insight into the relative importance of the CP correction terms versus the total dispersion energy of the dimer, both energy contributions are plotted in Fig. 4. Since the optimized energy for both monomers is not subtracted, the dispersion curve converges to

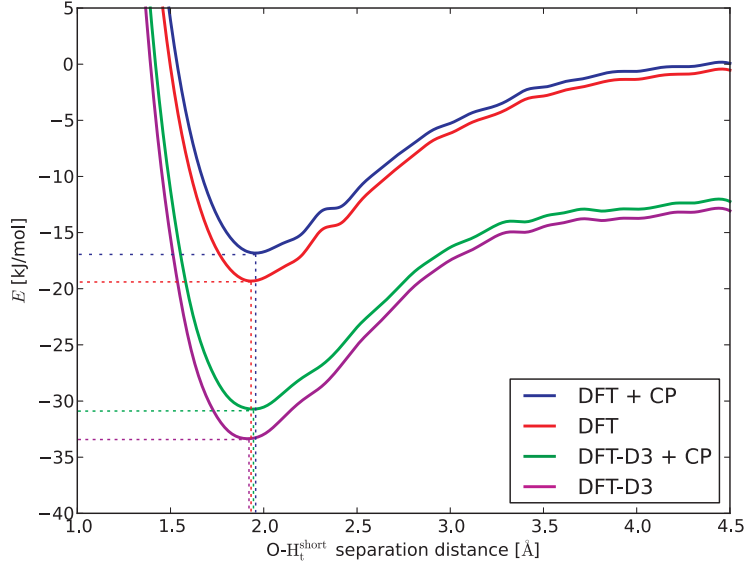


FIG. 3. Scan of a complex formed by two methanol monomers as a function of the separation distance $O-H_t^{\text{short}}$. All calculations are performed in gas phase with CP2K and with BLYP as functional.

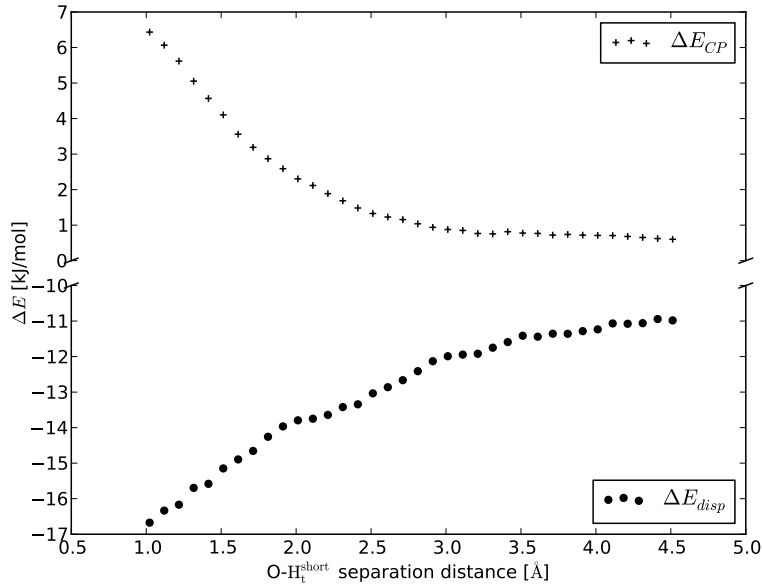


FIG. 4. CP-correction and dispersion energies derived from the $E_{\text{DFT}+\text{CP}}$ and $E_{\text{DFT}+\text{D3}}$ scans respectively, in function of the interatomic $O-H_t^{\text{short}}$ distance.

a value of $2 \times E_{disp}^{mon} = -7.53$ kJ/mol, where E_{disp}^{mon} is the dispersion energy of one optimized monomer. The CP-correction curve converges to zero at a O-H_t^{short} distance of 7.2 Å.⁹⁷

Some interesting conclusions can be drawn: (i) the shape of the CP-curves support the exponential model for the CP-correction, (ii) the dispersion correction is of the same order of magnitude as the BSSE, but with an opposite sign, (iii) in principle, the exponential trend as observed in Fig. 4 can be used for fitting the $(A_{O-H_t^{short}}, B_{O-H_t^{short}})$ parameters in Eq.(7). However, many plots as in Fig. 4 can be constructed. One could easily introduce many other intermolecular distances (such like O-H_t^{long}, O-H_m^{short}, etc.) and construct a potential energy scan following the same procedure as done in Fig. 4. Hence, for each pair of atom type t , one could construct a different set of (A_t, B_t) parameters. There is no reason why a particular parameter set is preferable to another. Moreover, there exist multiple dimer configurations with different spatial orientations of the methanol molecules and several paths to separate the two molecules from each other in the constrained geometry optimization. In the scans, performed in this work, only one single path is taken into consideration. It is obvious that this does not lead to a systematic way to derive the parameters. To remedy these shortcomings of the individual potential energy scans, it should be better to set up a fitting procedure that utilizes data from MD, allowing many possible orientations. This latter procedure provides a more statistically averaged result.

C. CP correction as a function of various O-H distances

In order to set up a fitting protocol which is based on MD, we make use of the data set of 3850 pairs of methanol molecules, obtained as described in section III B. Fig. 5 displays the logarithmic plot of the CP-energies as a function of the three O-H distances: O-H_t^{short}, O-H_t^{long} and O-H_m^{short} as defined in Fig. 1. For each specific pair of atom types t , we can construct an approximate model for the CP-correction:

$$\ln(\Delta E_{CP}(\mu, \nu)) \approx -B_{t(ij)} r_{ij} + \ln A_{t(ij)} \quad \forall \mu \neq \nu; i \in M_\nu, j \in M_\mu, \quad (8)$$

Eq.(8) is a special case of the general model expressed in Eq.(7): only one molecular pair is present and only one kind of interaction is considered. A straight line can be fitted to the logarithmic plots in Fig. 5. With each linear fit, the corresponding Pearson R^2 -value can be evaluated as a measure of correlation between the CP-energies and the oxygen-hydrogen

FIG. 5. CP-corrected energies as a function of various O-H distances based on geometries generated from MD: $\text{O-H}_t^{\text{short}}$, $\text{O-H}_t^{\text{long}}$, $\text{O-H}_m^{\text{short}}$.

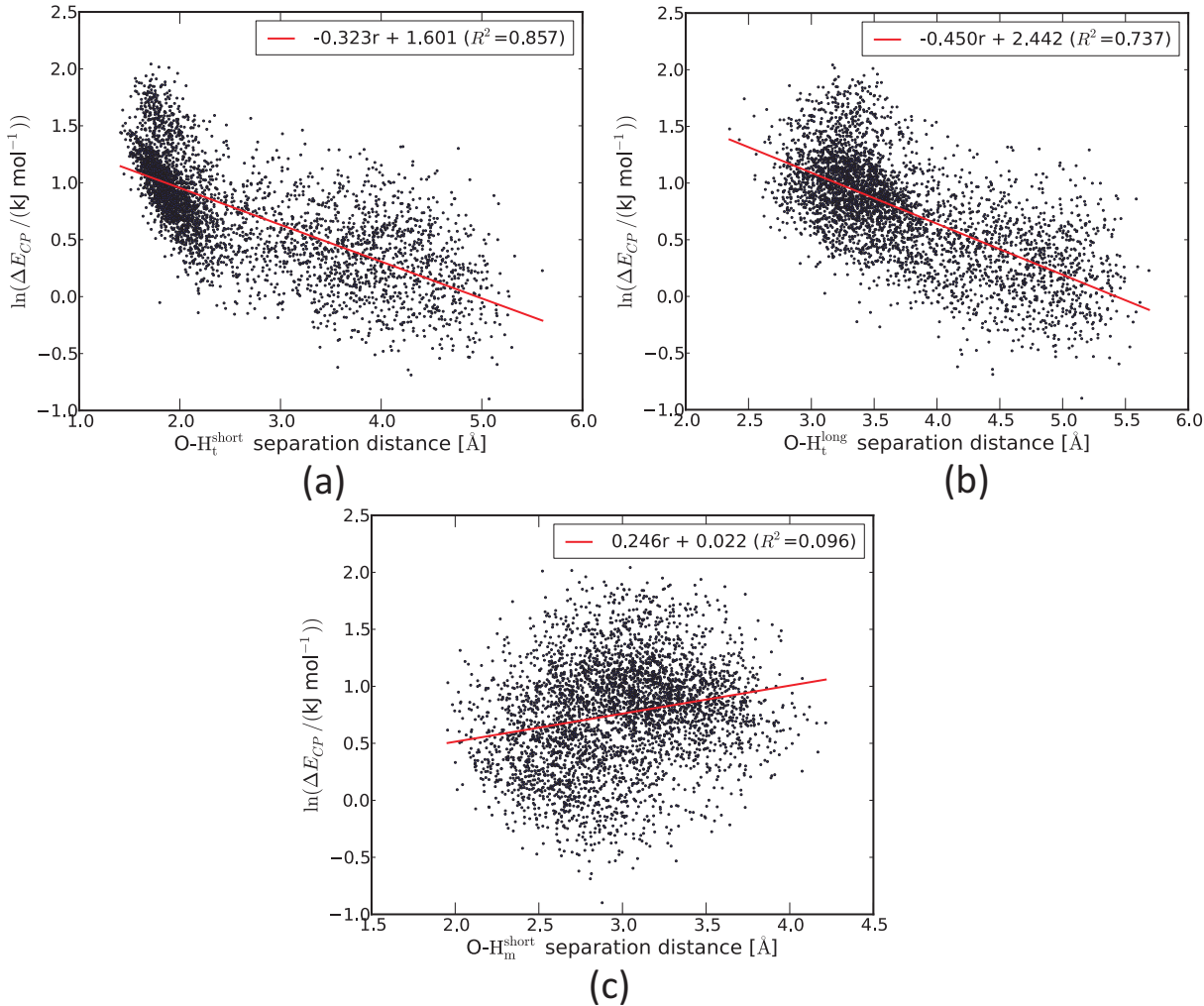


TABLE II. A_t, B_t parameters methanol by fitting CP-energies

method	interaction	A_t [kJ/mol]	B_t [$1/\text{\AA}$]	R^2
FIT/ $\text{O-H}_t^{\text{short}}$	$\text{O-H}_t^{\text{short}}$	4.96	0.323	0.857
FIT/ $\text{O-H}_t^{\text{long}}$	$\text{O-H}_t^{\text{long}}$	11.50	0.450	0.737

distances. It is also clear that it is worth to distinguish between the two kinds of O-H_t distances. The two fits are denoted by FIT/ $\text{O-H}_t^{\text{short}}$ and FIT/ $\text{O-H}_t^{\text{long}}$ and the parameters $A_{\text{O-H}_t}$ and $B_{\text{O-H}_t}$ are listed in Table II. On the contrary no linear correlation between $\ln \Delta E_{CP}$ and O-H_m distances could be found (Fig. 5(c)). Therefore, this specific interaction is left out of the final fitting procedure. So finally we are going to take two terms into

consideration to determine the parameters for the CP-correction model. It is emphasized that in Eq.(8) only one atom pair per dimer is considered to gain insight and can not be used to correct MD-data as it would be impossible to run a proper simulation with a term that includes only one kind of interaction (i.e. $\text{O-H}_t^{\text{short}}$ or $\text{O-H}_t^{\text{long}}$). As soon as the two O-H distances for a given pair interchange (short becomes long and vice versa) due to fluctuations on the distances, a discontinuity would appear in the forces. Such events are unphysical and violate energy conservation. For this reason, the linear fits in Fig. 5 are only indicative for the remainder of the analysis.

D. Final protocol for the determination of the CP-correction for the O.H_t interaction.

In the previous section, different parameter sets were determined for the pairs $\text{O-H}_t^{\text{short}}$ and $\text{O-H}_t^{\text{long}}$, which provided useful insights in the correlation of these pair distances with the CP correction. However, the distinction between $\text{O-H}_t^{\text{short}}$ and $\text{O-H}_t^{\text{long}}$ is not compatible with Eq.(7) because $\text{H}_t^{\text{short}}$ and H_t^{long} are not valid atom types. The distinction between long and short depends on the geometry of the pair, which may interchange during an MD simulation. In order to avoid discontinuities in the inter-molecular forces, the parameters for both $\text{O-H}_t^{\text{short}}$ and $\text{O-H}_t^{\text{long}}$ pairs must be identical. In line with Eq.(7), the CP model energy for a single pair of methanol molecules μ and ν (only considering the O-H_t terms) becomes:

$$\Delta E_{CP}^{\text{mod}}(\mu, \nu) = A_{\text{O-H}_t} \exp\left(-B_{\text{O-H}_t} r_{\text{O-H}_t^{\text{short}}}\right) + A_{\text{O-H}_t} \exp\left(-B_{\text{O-H}_t} r_{\text{O-H}_t^{\text{long}}}\right) \quad (9)$$

The parameters $A_{\text{O-H}_t}$ and $B_{\text{O-H}_t}$ are no longer determined by a linear fitting procedure due to the appearance of a non-linear term in the expression of $\ln\{E_{CP}^{\text{mod}}(\mu, \nu)\}$. We apply an in-house developed fitting procedure FFit2⁹¹ (section III B). The results are shown in Table III and are deduced from the DFT-D3 MD data set. As initial values for the parameters we have taken $A=1$ kJ/mol and $B=1 \text{ \AA}^{-1}$ which are reasonable initial values.

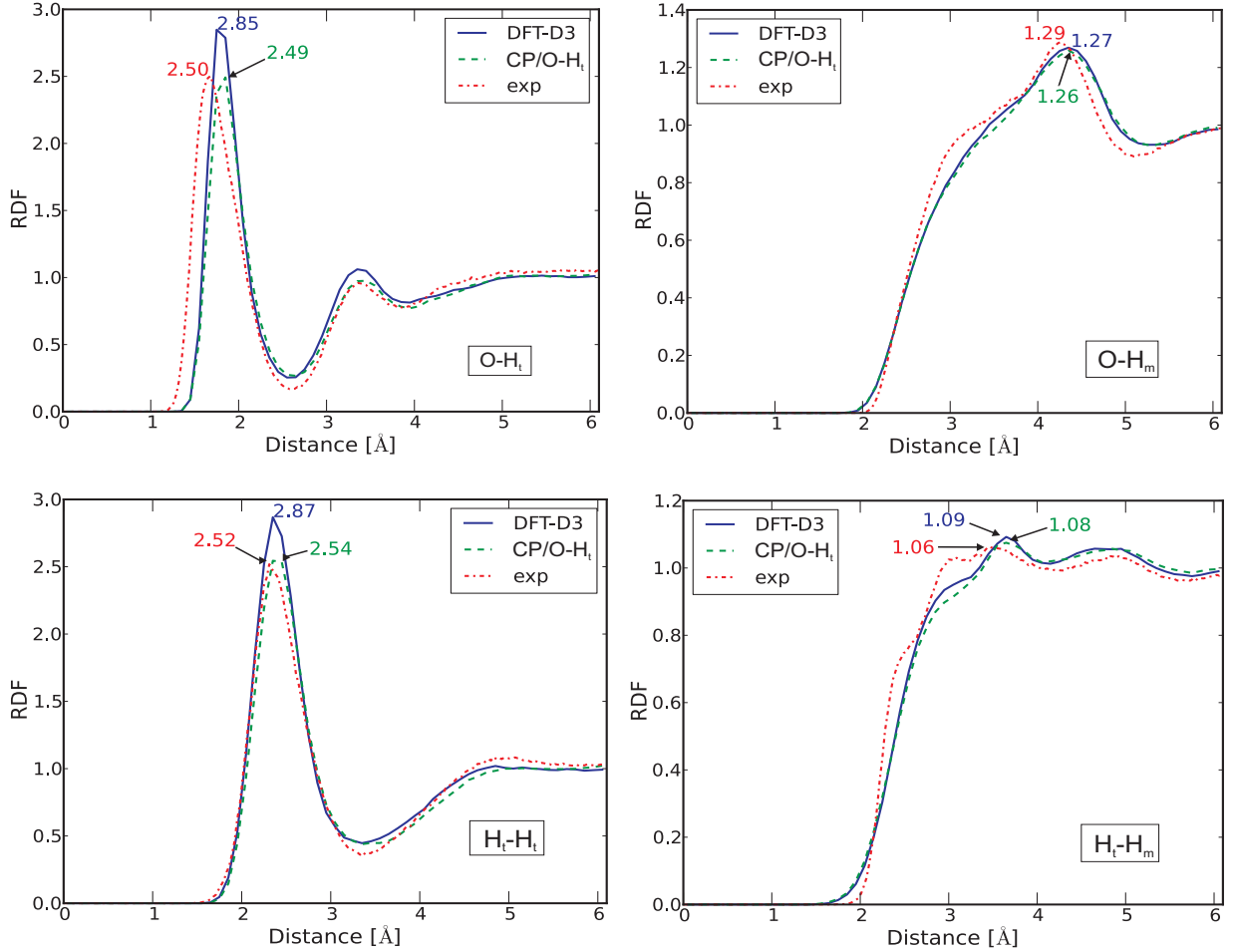
E. RDFs with correction

Fig. 6 presents the CP-corrected RDFs computed from a MD run following the new CP-protocol and taking into account DFT-D3 dispersion corrections. The DFT-D3 RDF is

TABLE III. $A_{\text{O-H}_t}$ and $B_{\text{O-H}_t}$ parameters methanol determined with FFit2 based on DFT-D3 MD.

method	interaction	optimized value	
		A [kJ/mol]	B [\AA^{-1}]
CP/O-H _t	O-H _t	4.47	0.453

FIG. 6. Comparison RDFs methanol: DFT-D3, CP/O-H_t, experiment.



shown as reference. The inclusion of BSSE according to our CP-model significantly improves the reproduction of the experimental peak for O-H_t and H_t-H_t. The amplitudes are tabulated in Table IV. As expected, the H_t-H_m and the O-H_m RDFs are not affected by adding the extra BSSE model term since the interactions with the H_m atoms were left out of the calibration procedure.

TABLE IV. Amplitudes of the first peak of the RDFs for O-H_t and H_t-H_t and the relative error compared to experiment.

	exp.	DFT-D3	$\Delta_{\text{DFT-D3}}$	CP/O-H _t	$\Delta_{\text{CP/O-H}_t}$
O-H _t	2.50	2.85	12.1%	2.49	0.4%
H _t -H _t	2.52	2.87	12.2%	2.54	1%

V. DISCUSSION

A first conclusion from the energy scan of the two methanol monomers (Figs. 3 and 4) is that the BSSE and dispersion interaction are almost equal in strength but opposite in sign. Dispersion is attractive and converges very slowly at large distances. At an O-H_t distance of 4.5 Å the dispersion energy still accounts for -11 kJ/mol and the slow increase of the dispersion curve (Fig. 4) demonstrates manifestly its long-range behavior. The CP-energy correction is repulsive and compensates partially the dispersion at short distances. The strongest effects of BSSE are noticed at distances shorter than 2.5 Å, and rapidly decays exponentially to about 1 kJ/mol at intermediate distances (3 – 3.5 Å), the structure of the RDFs at intermediate and long-range will be mainly determined by the dispersion. This is confirmed by the various RDF curves displayed in Figs. 2 and 6.

The most pronounced effect is the ability of the CP model to reduce the amplitude of the first peak to values which are comparable to experimental predictions (Table IV). In this perspective, it is instructive to compare our results with similar molecular dynamics simulations in the literature, although the influence of BSSE on RDFs has not been investigated in all these works. Special attention goes to the work of Schmidt et al. where DFT MD simulations have been carried out for water in near-ambient conditions with almost the same specifications as in our work, but within an isobaric-isothermal ensemble (NPT).²⁷ The authors used two functionals BLYP and PBE and reported some functional dependence in the position of the first peak of the oxygen-oxygen RDF in water. BLYP overestimates the position by about 0.1 Å toward larger values due to the tendency of BLYP to produce weaker hydrogen bonds (underbinding) than that predicted by the PBE functional (overbinding). In turn, the height for the first oxygen-oxygen peak is significantly larger for PBE. Inclusion of dispersion following the correction proposed by Grimme,⁴⁵ gives a substantial improvement in the RDF but this is not exactly what we observe in methanol. Different Grimme

corrections are used in both works, which troubles the comparison. Schmidt et al. makes use of DFT-D (D1 and D2) corrections described by damped interatomic potentials of the form C^6R^{-6} , while in this work the more recent DFT-D3 corrections were used. We expect that differences in the D-correction will affect the height and position of the first RDF peak. For the convenience of the reader the most relevant interatomic potentials are included in the S.I. for the three Grimme parameterizations D1, D2 and D3. We concentrate on the O-H interatomic potential⁹⁸ as we could expect that it affects the O-H_t RDF. Minima of this dispersion term are observed around 2.5 Å with an interaction energy of the order of -0.7 kJ/mol for D1 and D3. These features are too moderate to have a significant impact on the RDF. One notices a slight increase of the O-H radial distribution curves – DFT compared with DFT-D3 – in the range around 2.5 Å, which is counterintuitive given the attractive nature of the dispersion potential. In the region of the first peak, the dispersion correction to the interatomic O-H potential is small and, as one could expect, its influence on the height of the first peak would be negligible. Nevertheless, one must be very careful in oversimplifying the overall mechanism responsible in the reproduction of the first RDF peak. First, the total dispersion energy in the region of the first peak is of the order of -14 kJ/mol (based on the energy scan in Fig. 4), which contains contributions from multiple pairs. Second, our comments concern the methanol liquid while Schmidt et al. deal with water.

The importance of dispersion interactions in predicting structural and dynamical properties of liquid water is also demonstrated by Lin et al.²⁸ In this study both dispersion-corrected atom-centered potentials,⁴⁸ as well as empirical DFT-D1 Grimme corrections,⁴⁵ are used as two pragmatic ways to describe dispersion interactions in BLYP ab initio molecular dynamics simulations. They found an improved agreement of the RDFs with experiment, but significant deviations still persist.

Oxygen-oxygen and oxygen-hydrogen radial distribution functions for liquid water apparently improve when taking into account dispersion interactions. From our MD-BLYP calculations, we conclude that the effect of dispersion in a methanol liquid is less significant. The BSSE corrections turn out to be much more important for a correct description of the height of the first peak in the O-H_t and H_t-H_t RDFs. The deviation with experiment is largely reduced to respectively 0.4% and 1% from the experimental value (Table IV). These very small differences can be resolved by noting that the experimental RDF is only extracted

TABLE V. Positions of the first intermolecular RDF peak in Å from numerical simulations and experiment.

RDF	exp.	DFT	DFT-D3	CP/O-H _t
O-H _t	1.68	1.75	1.75	1.85
H _t -H _t		2.35	2.35	2.35

indirectly from experiment. It is computed from Monte Carlo simulations according to the empirical potential model refinement (EPSR) procedure. In this iterative technique, a potential energy function is refined by making use of empirical data as input:^{99–102} the Fourier transform of the difference between experimental and simulated structure factors is used solely for ensuring the best possible agreement between the simulated and measured partial structure factors.¹⁰³ The use of an interaction potential, from which the RDFs are extracted, may generate small discrepancies between this experimentally derived RDF and the ‘real’ RDF. Yamaguchi et al. illustrate how a different set of parameters of a single term of the potential can change the resulting site-site RDFs.⁹

The position of the first peak in the O-H_t and H_t-H_t RDFs is still overestimated with or without BSSE and/or dispersion correction (Table V), compared to experiment. Plausible explanations for this discrepancy are twofold. First, the shift of the peak position towards longer distances is in line with intuitive expectations since the CP model adds a repulsive interaction between the molecules. Second, the differences with experiment in peak positions could be due to the choice of the DFT-functional, BLYP, that was used. Schmidt et al. found that, for water, the position of the peaks shifted by about 0.1 Å toward larger values for BLYP as compared to PBE.²⁷ This shift confirms the tendency of the BLYP functional to produce weaker hydrogen bonds than the PBE functional.^{104,105}

One might suspect that the CP correction model may induce artificial distortions of the monomer geometries. We have compared the time averages and spreads of the intramolecular O-H_t distance and C-O-H_t angle (data not shown) between simulations without and with the CP correction model. No significant distortions have been noticed, indicating that the extra intermolecular forces by the model are small compared to the intramolecular forces.

The results so far clearly indicate that the CP model improves the O-H_t RDF, however, it also introduces minor deviations in the O-O RDF.¹⁰⁶ The height of the experimental peak of

the O-O RDF has an amplitude of 2.83, while the CP-correction lowers the BLYP-D3 peak from 2.89 to 2.48. This outcome is not surprising since the oxygen atom is covalently bonded to the H_t atom, and hence some correlation exists between the position of the oxygen and the hydroxyl hydrogen. A simultaneous parametrisation of the O-O and the O- H_t interaction could improve the agreement. However, this is a non-trivial task due to the non-linear character of Eq.(7) in the fitting-procedure. First, the system of equations to be solved in the Levenberg-Marquardt algorithm becomes more ill-conditioned when the parameter space is extended. Second, the parameters A_{O-H_t} , B_{O-H_t} and A_{O-O} , B_{O-O} are heavily correlated: the O-O distance and the O- H_t are not independent variables, causing their parameters to be also dependent. These problems render it hard to obtain meaningful parameters for the O-O and O- H_t interaction at the same time. Furthermore, parameter correlation impairs the physical meaning of the model parameters in Eq.(7) and reduces the predictive power of the model. When even more interaction terms are considered, the non-linear optimization problem becomes even more difficult.

VI. CONCLUSION AND PERSPECTIVES

The radial distribution function gives the probability of finding a pair of molecules at a distance r apart, relative to the expected probability for a completely random distribution at the same density. In DFT MD, using localized basis sets, the first intermolecular peak of the RDF of a liquid is often overestimated. In this paper it was investigated for liquid methanol to what extent this error is caused by the BSSE, which follows from the use of finite basis sets in quantum chemical calculations.

The relative importance of BSSE versus the dispersion interaction (DFT-D3) was assessed. The dispersion energy is of the same order of magnitude as the BSSE, but has an opposite sign.

Combinations of the two types of corrections – a correction term for BSSE and dispersion – leads to three types of DFT MD simulations that were carried out. For the simulations with and without dispersion, without correction for BSSE, errors on the first peak height of the RDFs between 12% and 16% were found. A classical correction model for BSSE is proposed and its parameters were derived from counterpoise (CP) computations on methanol dimers. This CP model can then be added to the DFT energy in subsequent MD simulations.

We confirmed that an exponential decaying functional form for the CP model works well. Because the counterpoise correction correlates stronger with the O-H_t distance, as compared to the O-H_m distance, the final form of the CP model is only based on O-H_t distances. Upon introduction of the correction, the errors on the heights of the first peak of the relevant O-H_t and H_t-H_t RDFs are reduced to 0.4% and 1%, respectively. A small shift of the peak position compared to experiment is still observed, which could be due to the added repulsive interaction of the CP model and the limitations of the BLYP functional.

We focused on hydrogen bridges in methanol by applying the CP model to the intermolecular interaction between the oxygen and the hydroxyl hydrogen, but the results should be extendable to other types of interaction where BSSE is important. However, including more interaction terms in the CP model complicates the optimization of the parameters because of the inherent parameter correlation and ill-conditioning of the set of equations that needs to be solved. It needs furthermore to be tested in how far the CP model is transferable to other protic solvents.

In this paper we concentrated on the effect of BSSE, and correcting for it, on a structural quantity. It would be highly valuable to obtain similar information for other quantities as well, e.g. dynamical properties such as the diffusion coefficient, or vibrational spectra. This way, it would be possible to obtain a complete picture of the influence of BSSE on molecular dynamics simulations that makes use of localized basis sets.

ACKNOWLEDGMENTS

This work is supported by the Fund for Scientific Research - Flanders (FWO), the Research Board of Ghent University (BOF) and BELSPO in the frame of IAP/6/27, the Belgian Prodex office ESA and the European Research Council under the European Community's Seventh Framework Programme (FP7(2007-2013) ERC grant agreement number 240483). The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure). AG and TV are post-doctoral researchers funded by the Foundation of Scientific Research - Flanders.

REFERENCES

- ¹M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids, Chapter 2*, 1st ed. (Oxford University Press, 1991).
- ²P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
- ³P. Hohenberg and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
- ⁴O. M3, M. Y3, and J. Elguero, J. Chem. Phys. **107**, 3592 (1997).
- ⁵E. Tsuchida, Y. Kanada, and M. Tsukada, Chem. Phys. Lett. **311**, 236 (1999).
- ⁶R. Ludwig, ChemPhysChem **6**, 1369 (2005).
- ⁷J. A. B. da Silva, F. G. B. Moreira, V. M. L. dos Santos, and R. L. Longo, Phys. Chem. Chem. Phys. **13**, 593 (2011).
- ⁸T. Yamaguchi, K. Hidaka, and A. Soper, Mol. Phys. **96**, 1159 (1999).
- ⁹T. Yamaguchi, K. Hidaka, and A. Soper, Mol. Phys. **97**, 603 (1999).
- ¹⁰T. Yamaguchi, C. J. Benmore, and A. K. Soper, J. Chem. Phys. **112**, 8976 (2000).
- ¹¹A. K. Adya, L. Bianchi, and C. J. Wormald, J. Chem. Phys. **112**, 4231 (2000).
- ¹²J.-W. Handgraaf, T. S. van Erp, and E. J. Meijer, Chem. Phys. Lett. **367**, 617 (2003).
- ¹³J.-W. Handgraaf, E. J. Meijer, and M.-P. Gaigeot, J. Chem. Phys. **121**, 10111 (2004).
- ¹⁴M. Pagliai, G. Cardini, R. Righini, and V. Schettino, J. Chem. Phys. **119**, 6655 (2003).
- ¹⁵M. J. McGrath, I.-F. Kuo, and J. I. Siepmann, Phys. Chem. Chem. Phys. **13**, 19943 (2011).
- ¹⁶M. E. Mart3n, M. L. S3nchez, F. J. del Valle, and M. A. Aguilar, J. Chem. Phys. **116**, 1613 (2002).
- ¹⁷M. L. S3nchez, M. E. Mart3n, M. A. Aguilar, and F. J. del Valle, J. Comput. Chem. **21**, 705 (2000).
- ¹⁸M. Haughney, M. Ferrario, and I. R. McDonald, J. Phys. Chem. **91**, 4934 (1987).
- ¹⁹L. Bianchi, O. N. Kalugin, A. K. Adya, and C. J. Wormald, Mol. Simulat. **25**, 321 (2000).
- ²⁰J. L. Thomas, D. J. Tobias, and A. D. MacKerell, J. Phys. Chem. B **111**, 12941 (2007).
- ²¹L. Zorani3, F. Sokoli3, and A. Perera, J. Chem. Phys. **127**, 024502 (2007).
- ²²T. Kosztol3ny, I. Bak3, and G. P3link3s, J. Chem. Phys. **118**, 4546 (2003).
- ²³T. Ishiyama, V. V. Sokolov, and A. Morita, J. Chem. Phys. **134**, 024509 (2011).
- ²⁴H. Yu, D. P. Geerke, H. Liu, and W. F. V. Gunsteren, J. Comput. Chem. **27**, 1494 (2006).
- ²⁵S. Patel and C. L. Brooks, J. Chem. Phys. **122**, 024508 (2005).

- ²⁶H. Nakano, T. Yamamoto, and S. Kato, *J. Chem. Phys.* **132**, 044106 (2010).
- ²⁷J. Schmidt, J. VandeVondele, I.-F. W. Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy, *J. Phys. Chem. B* **113**, 11959 (2009).
- ²⁸I.-C. Lin, A. P. Seitsonen, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger, *J. Phys. Chem. B* **113**, 1127 (2009).
- ²⁹J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli, *J. Chem. Phys.* **120**, 300 (2004).
- ³⁰T. Ishiyama and A. Morita, *J. Chem. Phys.* **131**, 244714 (2009).
- ³¹J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello, *J. Chem. Phys.* **122**, 014515 (2005).
- ³²M. Sprik, J. Hutter, and M. Parrinello, *J. Chem. Phys.* **105**, 1142 (1996).
- ³³A. D. Boese, N. L. Doltsinis, N. C. Handy, and M. Sprik, *J. Chem. Phys.* **112**, 1670 (2000).
- ³⁴G. S. Fanourgakis, G. K. Schenter, and S. S. Xantheas, *J. Chem. Phys.* **125**, 141102 (2006).
- ³⁵B. Guillot and Y. Guissani, *J. Chem. Phys.* **108**, 10162 (1998).
- ³⁶M. E. Tuckerman, D. M. Marx, M. L. Klein, and M. Parrinello, *Science* **275**, 817 (1997).
- ³⁷E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli, *J. Chem. Phys.* **121**, 5400 (2004).
- ³⁸C. Swalina, Q. Wang, A. Chakraborty, and S. Hammes-Schiffer, *J. Phys. Chem. A* **111**, 2206 (2007).
- ³⁹E. G. Noya, C. Vega, L. M. Sesé, and R. Ramirez, *J. Chem. Phys.* **131**, 124518 (2009).
- ⁴⁰E. Balog, A. L. Hughes, and G. J. Martyna, *J. Chem. Phys.* **112**, 870 (2000).
- ⁴¹J. A. Morrone and R. Car, *Phys. Rev. Lett.* **101**, 017801 (2008).
- ⁴²G. S. Fanourgakis and S. S. Xantheas, *J. Chem. Phys.* **128**, 074506 (2008).
- ⁴³H. A. Stern and B. J. Berne, *J. Chem. Phys.* **115**, 7622 (2001).
- ⁴⁴R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 24712474 (1985).
- ⁴⁵S. Grimme, *J. Comput. Chem.* **25**, 1463 (2004).
- ⁴⁶S. Grimme, *J. Comput. Chem.* **27**, 1787 (2006).
- ⁴⁷S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010).
- ⁴⁸I.-C. Lin, M. D. Coutinho-Neto, C. Felsenheimer, O. A. von Lilienfeld, I. Tavernelli, and U. Rothlisberger, *Phys. Rev. B* **75**, 205131 (2007).
- ⁴⁹A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.* **102**, 073005 (2009).

- ⁵⁰M. Mantina, A. C. Chamberlin, R. Valero, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. A* **113**, 5806 (2009).
- ⁵¹S. Simon, M. Duran, and J. J. Dannenberg, *J. Chem. Phys.* **105**, 11024 (1996).
- ⁵²K. N. Kirschner, J. B. Sorensen, and J. P. Bowen, *J. Chem. Educ.* **84**, 1225 (2007).
- ⁵³S. Simon, J. Bertran, and M. Sodupe, *J. Phys. Chem. A* **105**, 4359 (2001).
- ⁵⁴S. Scheiner, *Molecular Interactions. From van der Waals to Strongly Bound Complexes*, 1st ed. (John Wiley & Sons, 1997).
- ⁵⁵D. Hadži, *Theoretical Treatments of Hydrogen Bonding*, 1st ed. (John Wiley & Sons, 1997).
- ⁵⁶F. B. van Duijneveldt, J. G. C. M. van Duijneveldt - van de Rijdt, and J. H. van Lenthe, *Chem. Rev.* **94**, 1873 (1994).
- ⁵⁷A. Bende and S. Suhai, *Int. J. Quantum Chem.* **103**, 841 (2005).
- ⁵⁸T. van Mourik, *J. Phys. Chem. A* **112**, 11017 (2008).
- ⁵⁹D. Tzeli, I. D. Petsalakis, and G. Theodorakopoulos, *J. Phys. Chem. A* **111**, 8892 (2007).
- ⁶⁰A. E. Shields and T. van Mourik, *J. Phys. Chem. A* **111**, 13272 (2007).
- ⁶¹L. F. Holroyd and T. van Mourik, *Chem. Phys. Lett.* **442**, 42 (2007).
- ⁶²P. Hobza and Z. Havlas, *Theor. Chem. Acc.* **99**, 372 (1998).
- ⁶³R. Crespo-Otero, L. A. Montero, W.-D. Stohrer, and J. M. D. la Vega, *J. Chem. Phys.* **123**, 134107 (2005).
- ⁶⁴D. Tzeli and A. A. Tsekouras, *Chem. Phys. Lett.* **496**, 42 (2010).
- ⁶⁵H. Kruse and S. Grimme, *J. Chem. Phys.* **136**, 154101 (2012).
- ⁶⁶CP2K(2011), <http://cp2k.berlios.de>.
- ⁶⁷J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, *Comput. Phys. Commun.* **167**, 103 (2005).
- ⁶⁸A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- ⁶⁹C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ⁷⁰G. Lippert, J. Hutter, P. Ballone, and M. Parrinello, *J. Phys. Chem.* **100**, 6231 (1996).
- ⁷¹S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* **54**, 1703 (1996).
- ⁷²C. Harwigsen, S. Goedecker, and J. Hutter, *Phys. Rev. B* **58**, 3641 (1998).
- ⁷³E. J. Meijer and T. S. van Erp, *Chem. Phys. Lett.* **333**, 290 (2001).
- ⁷⁴P. Hobza, J. Šponer, and T. Reschel, *J. Comput. Chem.* **16**, 1315 (1995).
- ⁷⁵S. Kristyán and P. Pulay, *Chem. Phys. Lett.* **229**, 175 (1994).

- ⁷⁶S. Grimme, J. Antony, T. Schwabe, and C. Mück-Lichtenfeld, *Org. Biomol. Chem.* **5**, 741 (2007).
- ⁷⁷S. Nosé, *J. Chem. Phys.* **81**, 511 (1984).
- ⁷⁸I. Mayer and P. R. Surján, *Int. J. Quantum Chem.* **36**, 225 (1989).
- ⁷⁹I. Mayer and Á. Vibók, *Int. J. Quantum Chem.* **40**, 139 (1991).
- ⁸⁰J. Noga and Á. Vibók, *Chem. Phys. Lett.* **180**, 114 (1991).
- ⁸¹P. Valiron, Á. Vibók, and I. Mayer, *J. Comput. Chem.* **14**, 401 (1993).
- ⁸²Á. Vibók and I. Mayer, *Int. J. Quantum Chem.* **43**, 801 (1992).
- ⁸³S. F. Boys and F. Bernardi, *Mol. Phys.* **19**, 553 (1970).
- ⁸⁴D. Feller, *J. Chem. Phys.* **98**, 7059 (1993).
- ⁸⁵B. Paizs, P. Salvador, A. G. Császár, M. Duran, and S. Suhai, *J. Comput. Chem.* **22**, 196 (2001).
- ⁸⁶I. Alkorta, C. Trujillo, J. Elguero, and M. Solimannejad, *Comput. Theor. Chem.* **967**, 147 (2011).
- ⁸⁷T. Helgaker, W. Klopper, H. Koch, and J. Noga, *J. Chem. Phys.* **106**, 9639 (1997).
- ⁸⁸S. K. Min, E. C. Lee, H. M. Lee, D. Y. Kim, D. Kim, and K. S. Kim, *J. Comput. Chem.* **29**, 1208 (2008).
- ⁸⁹T. Verstraelen, M. Van Houteghem, V. Van Speybroeck, and M. Waroquier, *J. Chem. Inf. Mod.* **48**, 2414 (2008).
- ⁹⁰<http://molmod.ugent.be/software>.
- ⁹¹L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier, and V. Van Speybroeck(2012), in press, *J. Chem. Theory Comput.*, DOI: 10.1021/ct300172m.
- ⁹²K. Levenberg, *Quart. Appl. Math.* **2**, 164 (1944).
- ⁹³D. Marquardt, *SIAM J. Appl. Math.* **11**, 431 (1963).
- ⁹⁴E. Artacho, E. Anglada, O. Diéguez, J. D. Gale, A. García, J. Junquera, R. M. Martin, P. Ordejón, J. M. Pruneda, D. Sánchez-Portal, and J. M. Soler, *J. Phys.: Condens. Matter* **20**, 064208 (2008).
- ⁹⁵J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, *J. Phys.: Condens. Matter* **14**, 2745 (2002).
- ⁹⁶E. Anglada and J. M. Soler, *Phys. Rev. B* **73**, 115122 (2006).
- ⁹⁷ See Supplementary Material Document No. for an extended potential energy scan ranging from 0.6 Å to 7.5 Å in Fig. S.4.

- ⁹⁸ See Supplementary Material Document No. for dispersion contributions to the interaction potentials for oxygen-hydrogen pair in Fig. S.3.
- ⁹⁹ A. K. Soper, *Chem. Phys.* **202**, 295 (1996).
- ¹⁰⁰ A. K. Soper, *J. Phys. Condens. Matter* **9**, 2717 (1997).
- ¹⁰¹ A. K. Soper, *Mol. Phys.* **99**, 1503 (2001).
- ¹⁰² D. T. Bowron, J. L. Finney, and A. K. Soper, *J. Phys. Chem. B* **102**, 3551 (1998).
- ¹⁰³ A. K. Soper, *Phys. Rev. B* **72**, 104204 (2005).
- ¹⁰⁴ J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- ¹⁰⁵ J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **78**, 1396 (1997).
- ¹⁰⁶ See Supplementary Material Document No. for O-O RDF in Fig. S.2.