Published in final edited form as:

Phys Chem Chem Phys. 2012 June 7; 14(21): . doi:10.1039/c2cp23714h.

Liquid Water Simulations with Density Fragment Interaction Approach

Xiangqian Hu[†], Yingdi Jin[‡], Xiancheng Zeng[†], Hao Hu[‡], and Weitao Yang^{†,*}

[†]Department of Chemistry, Duke University, Durham, NC 27708

[‡]Department of Chemistry, The University of Hong Kong, Hong Kong

Abstract

We reformulate the density fragment interaction (DFI) approach [J. Chem. Phys. 129, 054102 (2008)] to achieve linear-scaling quantum mechanical calculations for large molecular systems. Two key approximations are developed to improve the DFI efficiency and thus enable the calculations for large molecules: the electrostatic interactions between fragments are computed efficiently by means of polarizable electrostatic-potential-fitted atomic charges; and frozen fragment pseudopotentials, similar to the effective fragment potentials that can be fitted from interactions between small molecules, are employed to take into account the Pauli repulsion effect among fragments. Our reformulated and parallelized DFI method demonstrates excellent parallel performance based on the benchmarks for the system of 256 water molecules. Molecular dynamics simulations for the structural properties of liquid water also show a good agreement with experimental measurements including the heat capacity, binding energy per water molecule, and the radial distribution functions of for atomic pairs of O-O, O-H and H-H. With this approach, large-scale quantum mechanical simulations for water and other liquids become feasible.

Keywords

Electrostatic potential; Pauli repulsion; linear-scaling; effective potential

I. Introduction

Quantum mechanical (QM) calculations for molecules and materials, especially those based on density functional theory (DFT), are indispensable tools for exploring the structures, physical properties, kinetics and thermodynamics of complicated systems in science and engineering.¹⁻⁸ The computational cost for popular quantum chemistry methods usually scales as power of N (e.g., N^3), where N is the number of electrons in the molecule. The power scaling prohibits routine QM calculations for large molecules.⁹ To achieve linear-scaling for the computational cost versus system size, many methods have been developed. Among them, the divide-and-conquer (DAC) approach⁹⁻¹⁸ is one of the earliest methods pioneered by Yang and co-workers. DAC splits the entire molecular system into many subsystems according to the spatial positions of atoms: each subsystem contains a unique and non-overlapping core region, and a buffer region with atoms from other neighboring subsystems. Electronic structures are solved for each subsystem. With the density matrix partition functions, the electron densities from subsystems are pieced together to recover the

Corresponding authors: haohu@hku.hk and weitao.yang@duke.edu.

Supporting Information **Available**. The water dimer conformations used to fit the current FFP parameters can be downloaded at <http://www.chem.duke.edu/~xqhu/water2.tar.gz>. The DCD trajectory from 50 ps water simulations is available via http://www.chem.duke.edu/~xqhu/h2o_frag256.dcd.tar.gz with the psf file at http://www.chem.duke.edu/~xqhu/h2o_256.psf.

electron density of the entire molecule as well as to compute molecular properties. The success of DAC method is an elegant example for the principle of the nearsightedness of electrons. Following DAC, many other linear scaling algorithms have been proposed, for example, by reducing the computational cost of constructing the Fock matrix¹⁹ and even avoiding the diagonalization of large matrixes²⁰, and by fragmenting on the basis of functional groups derived from chemists' intuitions.²¹⁻²⁹ All these linear-scaling algorithms share the same principle of nearsightedness of electronic interactions. The nearsightedness makes the linear-scaling construction of the Fock matrix plausible; it also rationalizes the fragmentation of chemical functional groups, as well as the decomposition of the total electron density.

In this work, we focus on developing efficient algorithms for quantum chemistry calculations based on the concept of fragmentation. Specifically, we reformulated the original density fragment interaction (DFI) approach³⁰ and developed new schemes to efficiently and accurately compute the interactions between fragments. We elucidated that how this new DFI scheme can be derived and applied to simulate large molecular systems using liquid water as an example.

Water is ubiquitous and plays crucial roles in our life. A large number of studies^{8,21,26,31-49} has been carried out to determine the structures and properties of water with both experimental and theoretical methods. The water model used in theoretical researches can vary from classical, static point charge^{35,50} or polarizable models³⁶⁻⁴², to semi-empirical QM models^{21,26,44}, and to even highly accurate but very costly *ab initio* models⁴⁵⁻⁴⁹. Although the previous simulations using Car-Parrinello molecular dynamics (CPMD)⁴⁹ showed good agreements with experimental measurements for many physical properties of liquid water, the high computational cost severely limits the system size and time scale of the simulations. A small water box and a short trajectory may not be sufficient for simulating many important processes, such as the chemical reactions and solvation free energies in aqueous solutions. The reformulated DFI approach developed in the current work aims for applications to larger systems without covalently linked and well-defined fragments.

This paper is organized as following: In Sec. II, the original DFI approach is reformulated to reduce the computational cost. We show how one may compute the inter-fragment interactions in a very efficient way. Sec. III discusses the fitting procedure to determine the parameters introduced by the reformulations. Then several different fragmentation schemes on 256 water molecules were tested to demonstrate the parallel efficiency of our approach. The computed heat capacity, bind energy per water molecule, and radial distribution functions of OO, OH, and HH agree well with the experimental results. Sec. IV summarizes our work and discusses several issues for improvements.

II. Theory

II-A. Reformulations of density fragment interaction approach

In the Kohn-Sham DFT (KS-DFT) scheme, the total energy of the molecular system is⁹

$$E[\rho] = T_s[\rho] + \int V_{nuc}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + J[\rho] + E_{xc}[\rho] + E_{NN} \quad (1)$$

where ρ is the electron density of the system, $T_s[\rho]$ is the kinetic energy, V_{nuc} is the nuclear-electron attractive potential, $J[\rho]$ is the Coulombic energy, $E_{xc}[\rho]$ is the exchange-correlation energy, and E_{NN} is the nuclear-nuclear interaction energy. Let us consider one simple system that contains two non-overlapping fragments of A and B and without covalent bonds

between them. Suppose the true global electron density could be decomposed into the electron densities of subsystems,

$$\rho(\mathbf{r}) = \rho_A(\mathbf{r}) + \rho_B(\mathbf{r}) \quad (2)$$

Therefore, Eq. (1) could be rewritten by

$$\begin{aligned} E[\rho] &= T_s[\rho] + \int V_{nuc}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + J[\rho] + E_{xc}[\rho] + E_{NN} \\ &= T_s[\rho_A] + T_s[\rho_B] + \int V_{nuc,A}(\mathbf{r})\rho_A(\mathbf{r})d\mathbf{r} + \int V_{nuc,B}(\mathbf{r})\rho_B(\mathbf{r})d\mathbf{r} + \int V_{nuc,A}(\mathbf{r})\rho_B(\mathbf{r})d\mathbf{r} \\ &\quad + \int V_{nuc,B}(\mathbf{r})\rho_A(\mathbf{r})d\mathbf{r} + J[\rho_A] + J[\rho_B] + \iint \frac{\rho_A\rho_B}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_{xc}[\rho_A] + E_{xc}[\rho_B] \\ &\quad + E_{NN}^A + E_{NN}^B + E_{NN}^{AB} \\ &\quad + (E_{xc}[\rho] - E_{xc}[\rho_A] - E_{xc}[\rho_B]) + (T_s[\rho] - T_s[\rho_A] - T_s[\rho_B]) \end{aligned} \quad (3)$$

where $V_{nuc} = V_{nuc,A} + V_{nuc,B}$. The electrostatic potential (ESP) of fragment A is defined by

$$V_A^{ESP}(\mathbf{r}) = \sum_{i=1}^{N_A} \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}|} - \int \frac{\rho_A(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}' \quad (4)$$

where N_A is the number of atoms of fragment A, Z_i the nucleic charge, and $\mathbf{R}_i$ the nucleic positions. Eq. (3) can be simplified as,

$$E[\rho] = E_A^0[\rho_A] + E_B^0[\rho_B] + E_{NN}^{AB} - \iint \frac{\rho_A\rho_B}{|\mathbf{r} - \mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + \Delta E_{xc} + \Delta T_s \quad (5)$$

The extracted Coulombic crossing term is due to the double counting term in $E_A^0[\rho_A]$ and $E_B^0[\rho_B]$, and the last two terms arise from the non-additivity of the exchange-correlation functional and KS orbital kinetic energy. The energy expression of $E_A^0[\rho_A]$ is

$$E_A^0[\rho_A] = T_s[\rho_A] + \int V_A(\mathbf{r})\rho_A(\mathbf{r})d\mathbf{r} - \int V_B^{ESP}(\mathbf{r})\rho_A(\mathbf{r})d\mathbf{r} + J[\rho_A] + E_{xc}[\rho_A] + E_{NN}^A \quad (6)$$

Using Eq. (4), the two electrostatic terms in Eq. (5) can be rewritten as

$$\begin{aligned} E_{NN}^{AB} - \iint \frac{\rho_A\rho_B}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' &= \sum_{a=1}^{N_A} \sum_{b=1}^{N_B} \frac{Z_a Z_b}{|\mathbf{R}_a - \mathbf{R}_b|} + \int \rho_A d\mathbf{r} \left(- \int \frac{\rho_B}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}' + \sum_{b=1}^{N_B} \frac{Z_b}{|\mathbf{R}_b - \mathbf{r}|} - \sum_{b=1}^{N_B} \frac{Z_b}{|\mathbf{R}_b - \mathbf{r}|} \right) \\ &= \underbrace{\int \rho_A V_B^{ESP} d\mathbf{r} + \sum_{b=1}^{N_B} Z_b V_A^{ESP}(\mathbf{R}_b)}_{Term1} \left[\underbrace{\int \rho_B V_A^{ESP} d\mathbf{r} + \sum_{a=1}^{N_A} Z_a V_B^{ESP}(\mathbf{R}_a)}_{Term2} \right] \\ &= \frac{1}{2} \left(\int \rho_A V_B^{ESP} d\mathbf{r} + \sum_{b=1}^{N_B} Z_b V_A^{ESP}(\mathbf{R}_b) + \int \rho_B V_A^{ESP} d\mathbf{r} + \sum_{a=1}^{N_A} Z_a V_B^{ESP}(\mathbf{R}_a) \right) \end{aligned} \quad (7)$$

where a and b are the indices for atoms in fragments A and B. Similar equations have been derived before by Gao and co-workers within the framework of wave function theory.²¹ Eqs. (5) and (7) can be generalized to any number of non-overlapping fragments and the expression for total energy is

$$\begin{aligned}
E[\rho] &= \sum_{i=1}^{N_F} E_i^0[\rho_i] + \underbrace{\frac{1}{2} \sum_{i=1}^{N_F} \int \rho_i(\mathbf{r}) \sum_{j \neq i}^{N_F} V_j^{ESP}(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \sum_{i=1}^{N_F} \sum_{a=1}^{n_i} Z_a^i \sum_{j \neq i}^{N_F} V_j^{ESP}(\mathbf{R}_a^i)}_{E_{ESP}^{FF}[\{\rho_i\}] + \Delta E_{xc} + \Delta T_s} \\
&= \sum_{i=1}^{N_F} E_i^0[\rho_i] + E_{ESP}^{FF}[\{\rho_i\}] + \Delta E_{xc} + \Delta T_s
\end{aligned} \quad (8)$$

where i and j are the indices for fragments, $E_{ESP}^{FF}[\{\rho_i\}]$ contains two double-counting energy terms between fragments and can be computed with their ESPs, N_F is the number of fragments in the system and n_i is the number of atoms in fragment i . The general form of Eq. (6) is,

$$E_i^0[\rho_i] + T_s[\rho_i] + \int V_i(\mathbf{r}) \rho_i(\mathbf{r}) d\mathbf{r} - \int \sum_{j \neq i}^{N_F} V_j^{ESP}(\mathbf{r}) \rho_i(\mathbf{r}) d\mathbf{r} + J[\rho_i] + E_{xc}[\rho_i] + E_{NN}^i \quad (9)$$

The only question left for solving Eq. (8) is to compute the electrostatic interactions between fragments, and the non-additive terms of $\Delta E_{xc} + \Delta T_s$. These will be discussed immediately in the next sections.

II-B. Electrostatic interactions between fragments

Because each fragment interacts with other fragments through electrostatic potentials (Eq.

(9)), the calculations of the electrostatic integral for each fragment, $\int \sum_{j \neq i}^{N_F} V_j^{ESP}(\mathbf{r}) \rho_i(\mathbf{r}) d\mathbf{r}$, become the most time-consuming task.

In the original DFI work³⁰, the electrostatic integrals were computed analytically by the four-center two-electron repulsion and nuclear-electron attraction integrals as indicated by Eq. (4). Although the analytical integrations are exact, they are obviously too expensive if the number of fragments or the size of the fragments in the system is large, as the computational cost of this term approximately scales to N^4 .

Note that ESP values on a set of predefined grids, such as DFT grids, can be computed analytically for any fragment i . Therefore, an alternate and computationally efficient approach would be computing first the ESPs on a DFT grids, then carrying out numerical integrations with DFT grids. This scheme can reduce the computational costs of

$\int \sum_{j \neq i}^{N_F} V_j^{ESP}(\mathbf{r}) \rho_i(\mathbf{r}) d\mathbf{r}$ from N^4 to N^3 . This is, however, still time-consuming because of two reasons. The first reason is that the electrostatic contributions of a given fragment to other fragments need be calculated many times during the DFI self-consistent-field procedure (See Sec. II-D). The second reason is that even a single ESP calculation on one grid point in space could be expensive since its computational cost scales with the number of basis functions of each fragment.

In order to identify efficient schemes for the calculation of this inter-fragment electrostatic interactions term with low computational costs, we tested several approximated methods for computing $\{V_j^{ESP}(\mathbf{r})\}$. We will discuss the advantages and disadvantages, in terms of accuracy and computational cost, of each scheme we have tested. The first scheme is an indirect scheme (Scheme I): (i) a set of uniform grids are first generated to cover all

fragments; (ii) the $\{V_j^{ESP}(\mathbf{r})\}$ values on the uniform grids can be computed efficiently for each fragment; and (iii) the final $\{V_j^{ESP}(\mathbf{r})\}$ values on the DFT grids of each fragment can be computed by numerical interpolations with respect to the uniform grids. To make this scheme efficient, the number of uniform grids required in the interpolations must be significantly smaller than the total number of DFT grids of all fragments. However, based on the initial tests for the liquid water simulations, the number of uniform grids must be dense, for example, 1 point per 0.2 Å in order to describe the hydrogen bonding energies in water. The sharp cusps of ESP at the positions of nuclei require the dense grid points. Such dense grids make the calculation very slow and thus infeasible for molecular dynamics simulations. The second scheme is to compute the Coulomb potential term in Eq. (4) using the multipolar expansion⁵¹ based on the radial points of DFT grids for each fragment (Scheme II). To obtain accurate multipolar expansions, our tests showed that the number of radial points must be sufficiently large, e.g., 80 radial points are required to perform the multipolar expansions for the simulations of liquid water. This makes the calculation still expensive. The last scheme used in our simulations is to employ the fitted ESP charges ($\{Q_a^j\}$) for each fragment to compute $\{V_j^{ESP}(\mathbf{r})\}$ (Scheme III),

$$V_j^{ESP}(\mathbf{r}) \approx \sum_{a=1}^{N_j} \frac{Q_a^j}{|\mathbf{R}_a^j - \mathbf{r}|} \quad (10)$$

The HLY ESP charge fitting method⁵² is applied here since this method showed much improved numerical stability with less grids required for fitting. (Note that the X-POL method^{21,26} uses the simple Mulliken charges instead of ESP charges, which might show substantial difference in applications.)

Although Schemes I and II are not feasible for simulating liquid water, both could become practical in the simulations of other systems since the ESP values computed with both schemes are more accurate than the simple ESP-fitted point charge model in Scheme III. For liquid water simulations, when each fragment contains only one water molecule, the SCF calculation for each fragment, i.e., a water molecule embedded in external potential, is already fast. Therefore, the bottleneck for the parallel DFI calculation of liquid water resides in the computational and communication cost for computing the ESP values on DFT grids for each fragment. In this work, we found that the simple ESP-fitted point charges have the moderate accuracy to compute the electrostatic interactions between fragments. Numerical experiments indicated that the simplest scheme of ESP-fitted charges is the fastest among three schemes, at the similar level of accuracy for energy computations. Another advantage to employ the point charge model for liquid water simulations is that the periodical boundary condition can be easily implemented with local nearest-image convention. The accuracy of electrostatic interactions might be further improved by efficient multi-resolution descriptions of the ESP of fragments.

II-C. Frozen fragment pseudopotential approach

The two terms ΔE_{xc}^{nonad} and ΔT_s^{nonad} in Eq. (8) are difficult to compute, even though they have been explored by many previous studies.⁵³⁻⁵⁸ In the two terms, the magnitude of ΔE_{xc}^{nonad} is typically small when there are no covalent connections between fragments, but its value can be either negative or positive. The non-additive errors of the kinetic term in KS-DFT (ΔT_s^{nonad}) are often attributed to the non-orthogonality of KS molecular orbitals between fragments, which could also be interpreted with the Pauli Exclusion principle. In

the divide-and-conquer approaches⁹⁻¹⁸, the Pauli repulsion is taken care by using the buffer regions where the localized atomic orbitals are used in the calculation of electron density of the central system. The non-orthogonality problem of molecular orbitals also appears in the frozen-core approximations. After the core electrons are removed from the QM calculations, the *ab initio* effective core potential (ECP) approach⁵⁹⁻⁶³ was proposed and developed to model the interactions between core and valence electrons. In addition, the “pseudo-bond” approach^{64,65} applied for the boundary atoms in the quantum mechanical/molecular mechanical (QM/MM) interface utilizes the ECP forms to alleviate the boundary problems in QM/MM calculations.

In this work, we propose a frozen fragment pseudopotential (FFP), which has the same form as ECP, to approximate the contribution of the non-additive errors of kinetic term. FFP can be expressed as angular momentum l -dependent operators centered at atoms in fragments. The new FFP integrals are calculated by the half-numerical method for ECP.⁶⁶ Therefore, Eqs. (8) and (9) can be reformulated as,

$$E[\rho] = \sum_{i=1}^{N_F} E_i^F[\rho_i] + E_{ESP}^{FF} \quad (11)$$

$$\begin{aligned} E_i^F[\rho_i] = & T_s[\rho_i] + \int V_i(\mathbf{r})\rho_i(\mathbf{r})d\mathbf{r} - \int \sum_{j \neq i}^{N_F} V_j^{ESP}(\mathbf{r})\rho_i(\mathbf{r})d\mathbf{r} \\ & + \int \sum_{j \neq i}^{N_F} V_j^{FFP}(\mathbf{r})\rho_i(\mathbf{r})d\mathbf{r} + J[\rho_i] + E_{xc}[\rho_i] + E_{NN}^i \end{aligned} \quad (12)$$

where $V_j^{FFP}(\mathbf{r})$ can be expressed by atoms in fragment j ,

$$V_j^{FFP}(\mathbf{r}) = \sum_{a=1}^{N_j} U_a(\mathbf{r}, \mathbf{R}_a^j) \quad (13)$$

The formula for $U_a(\mathbf{r}, \mathbf{R}_a^j)$ for each atomic type, similar with ECP of atoms, is given by,

$$U(\mathbf{r}) = U_L(r) + \sum_{l=0}^{L-1} \sum_{m=-l}^{m=l} |S_{lm}\rangle U_l(r) \langle S_{lm}| \quad (14)$$

$$U_l(r) = \sum_{k=1}^{K_l} d_{kl} r^{n_{kl}-2} e^{-\zeta_{kl} r^2} \quad (15)$$

where $|S_{lm}(\theta, \phi)\rangle$ is the normalized real spherical harmonic functions. All parameters in Eqs. (14) and (15) such as L , K_L , d_{kl} , n_{kl} , and ζ_{kl} need to be fitted for each atomic type according to molecular properties. (More details about fitting are discussed in Sec. III-A.) In principle, by introducing the FFP approach, ΔE_{xc}^{nonad} and ΔT_s^{nonad} are expanded by,

$$\Delta E_{xc}^{nonad} + \Delta T_s^{nonad} = \sum_{i=1}^{N_F} \sum_{j \neq i}^{N_F} \int V_j^{FFP}(\mathbf{r}) \rho_i(\mathbf{r}) d\mathbf{r} \quad (16)$$

The accuracy of the expansion is determined by the completeness of the FFP expansions. Generally, the larger number of radial Gaussian functions in Eq. (14) is chosen, the better accuracy can be achieved to capture the non-additivity errors of ΔE_{xc}^{nonad} and ΔE_s^{nonad} .

II-D. Self-consistent field procedure of DFI

To minimize the DFI energy defined by Eq. (11), one can define the functional of the fragment orbitals,

$$\Omega[\{\psi_m^i\}] = E[\tilde{\rho}] - \sum_{i=1}^{N_F} \sum_{m=1}^{N_i^{MO}} \sum_{n=1}^{N_i^{MO}} \varepsilon_{mn}^i \int \psi_m^i(\mathbf{r}) * \psi_n^i(\mathbf{r}) d\mathbf{r} \quad (17)$$

where $\{\psi_m^i\}$ are the molecular orbitals of the i -th fragment and N_i^{MO} is the number of molecular orbitals. Note that the orthonormality of molecular orbitals is only required within each fragment because of the FFP approach. Based on the stationary condition, $\delta\Omega[\{\psi_m^i\}] = 0$ the KS orbital equations for DFI can be deduced after a unitary transformation of the molecular orbitals for fragment i ,

$$\left[-\frac{1}{2} \nabla^2 + V_{eff}^i \right] \psi_m^i = \varepsilon_m^i \psi_m^i \quad (18)$$

$$V_{eff}^i = V_i(\mathbf{r}) - \sum_{j \neq i}^{N_F} V_j^{ESP}(\mathbf{r}) + \sum_{j \neq i}^{N_F} V_j^{FFP}(\mathbf{r}) + \int \frac{\rho_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{xc}^i(\mathbf{r}) \quad (19)$$

Then the typical self-consistent-field procedure can be employed to minimize the total DFI energy. However, $\{V_j^{ESP}(\mathbf{r})\}$ in Eq. (19) needs to be updated in each SCF cycle, as discussed in Sec. II-B for the electrostatic interactions. Therefore, the DFI SCF procedure is designed as a sequential two-layer minimization procedure in the current work:

- a. Solving the optimal MOs $\{\psi_m^i\}$ with a fixed set of $\{V_j^{ESP}(\mathbf{r})\}$ by the normal SCF procedure for each fragment independently, namely SCF-MO;
- b. Using the new MOs to update $\{V_j^{ESP}(\mathbf{r})\}$ and go to step a) again until the total DFI energy defined in Eq. (11) is converged, namely SCF-DFI.

Step *a* is an intrinsic (inner-layer) SCF iteration for each fragment to obtain the optimal MOs for the fixed $\{V_j^{ESP}(\mathbf{r})\}$ while step *b* is an outer-layer iteration (i.e., SCF-DFI) to update the electrostatic potentials of each fragment and minimize the total DFI energy. This sequential minimization procedure also guarantees that $\delta\Omega[\{\psi_m^i\}] = 0$ when both the inner- and outer- iterations are converged. When the number of iterations in SCF-MO is set to be one, the foregoing two-layer sequential minimization procedure regresses to the one-step SCF-DFI procedure. Note there could be many possible schemes to carry out this two-layer SCF; some might provide better computational efficiency than the simple scheme we described here. Works are currently undergoing to improve DFI SCF efficiency.

II-E. Analytical energy gradients with respect to nuclear positions

In order to carry out molecular dynamics simulation with DFI, we need to calculate the energy gradient of each atom with respect to atomic positions. Let us consider the simple system with two fragments **A** and **B**. The total energy is

$$E[\rho] = T_s[\rho_A] + T_s[\rho_B] + \int V_A(\mathbf{r})\rho_A(\mathbf{r})d\mathbf{r} + \int V_B(\mathbf{r})\rho_B(\mathbf{r})d\mathbf{r} + \int V_A(\mathbf{r})\rho_B(\mathbf{r})d\mathbf{r} + \int V_B(\mathbf{r})\rho_A(\mathbf{r})d\mathbf{r} + J[\rho_A] + J[\rho_B] + \iint \frac{\rho_A(\mathbf{r})\rho_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_{xc}[\rho_A] + E_{xc}[\rho_B] + \int \rho_A(\mathbf{r})V_B^{FFP}(\mathbf{r})d\mathbf{r} + \int \rho_B(\mathbf{r})V_A^{FFP}(\mathbf{r})d\mathbf{r} + E_{NN}^A + E_{NN}^B + E_{NN}^{AB} \quad (20)$$

The energy gradient for atoms in fragment **A** is

$$\begin{aligned} \frac{\partial E[\rho]}{\partial \mathbf{R}_a^A} = & \frac{\partial T_s[\rho_A]}{\partial \mathbf{R}_a^A} + \int \frac{\partial V_A(\mathbf{r})}{\partial \mathbf{R}_a^A} \rho_A(\mathbf{r})d\mathbf{r} + \int V_A(\mathbf{r}) \frac{\partial \rho_A(\mathbf{r})}{\partial \mathbf{R}_a^A} d\mathbf{r} \\ & + \int \frac{\partial \rho_A(\mathbf{r})}{\partial \mathbf{R}_a^A} V_B(\mathbf{r})d\mathbf{r} + \int \frac{\partial \rho_A(\mathbf{r})}{\partial \mathbf{R}_a^A} \int \frac{\rho_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \frac{\partial J[\rho_A]}{\partial \mathbf{R}_a^A} + \frac{\partial E_{xc}[\rho_A]}{\partial \mathbf{R}_a^A} \\ & + \int \frac{\partial \rho_A(\mathbf{r})}{\partial \mathbf{R}_a^A} V_B^{FFP}(\mathbf{r})d\mathbf{r} + \frac{\partial E_{NN}^A}{\partial \mathbf{R}_a^A} + \int \frac{\partial V_A^{FFP}(\mathbf{r})}{\partial \mathbf{R}_a^A} \rho_B(\mathbf{r})d\mathbf{r} + \int \frac{\partial V_A(\mathbf{r})}{\partial \mathbf{R}_a^A} \rho_B(\mathbf{r})d\mathbf{r} + \frac{\partial E_{NN}^{AB}}{\partial \mathbf{R}_a^A} \\ & + \frac{\partial T_s[\rho_B]}{\partial \mathbf{R}_a^A} + \int V_B(\mathbf{r}) \frac{\partial \rho_B(\mathbf{r})}{\partial \mathbf{R}_a^A} d\mathbf{r} + \int V_A(\mathbf{r}) \frac{\partial \rho_B(\mathbf{r})}{\partial \mathbf{R}_a^A} d\mathbf{r} + \int \frac{\partial \rho_B(\mathbf{r})}{\partial \mathbf{R}_a^A} \int \frac{\rho_A(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \\ & + \frac{\partial J[\rho_B]}{\partial \mathbf{R}_a^A} + \frac{\partial E_{xc}[\rho_B]}{\partial \mathbf{R}_a^A} + \int \frac{\partial \rho_B(\mathbf{r})}{\partial \mathbf{R}_a^A} V_A^{FFP}(\mathbf{r})d\mathbf{r} \end{aligned} \quad (21)$$

In Eq. (21), all the terms without the $\rho_B(\mathbf{r})$ response with respect to $\mathbf{R}_a^A$ are

$$\begin{aligned} W_1 = & \frac{\partial E^F}{\partial \mathbf{R}_a^A} + \int \frac{\partial V_A^{FFP}(\mathbf{r})}{\partial \mathbf{R}_a^A} \rho_B(\mathbf{r})d\mathbf{r} + \int \frac{\partial V_A(\mathbf{r})}{\partial \mathbf{R}_a^A} \rho_B(\mathbf{r})d\mathbf{r} + \frac{\partial E_{NN}^{AB}}{\partial \mathbf{R}_a^A} \\ = & \frac{\partial E^F}{\partial \mathbf{R}_a^A} + \int \frac{\partial V_A^{FFP}(\mathbf{r})}{\partial \mathbf{R}_a^A} \rho_B(\mathbf{r})d\mathbf{r} + Z_a^A \frac{\partial V_B^{ESP}(\mathbf{R}_a^A)}{\partial \mathbf{R}_a^A} \end{aligned} \quad (22)$$

Note that the derivatives of $\{C_{\mu m}^i\}$ with respect to the atomic positions can be avoided by the orthonormality condition of molecular orbitals for each fragment, as in the conventional DFT or Hartree-Fock approach. Therefore, in Eq. (21), all the terms with the $\rho_B(\mathbf{r})$ response to $\mathbf{R}_a^A$ are

$$W_2 = - \sum_{\mu\nu}^{N_B^{AO}} Q_{\nu\mu}^B \frac{\partial S_{\mu\nu}^B}{\partial \mathbf{R}_a^A} = 0 \quad (23)$$

where N_B^{AO} is the number of atomic basis sets in fragment B, $Q_{\nu\mu}^B$ is the eigenvalue weighted density matrix of fragment B, and $S_{\mu\nu}^B$ is the overlap matrix of fragment B.

Overall, no response such as $\frac{\partial \rho_B(\mathbf{r})}{\partial \mathbf{R}_a^A}$ is needed to compute the analytical DFI energy gradients when the analytical V_j^{ESP} form is employed.

The analytical energy gradients for any number of fragments are

$$\begin{aligned}\frac{\partial E[\rho]}{\partial \mathbf{R}_a^k} &= \frac{\partial E_k^F[\rho_k]}{\partial \mathbf{R}_a^k} + \int \sum_{j \neq k}^{N_F} \frac{\partial V_k^{FFP}(\mathbf{r})}{\partial \mathbf{R}_a^k} \rho_j(\mathbf{r}) d\mathbf{r} + \sum_{j \neq k}^{N_F} \int \frac{\partial V_k}{\partial \mathbf{R}_a^k} \rho_j(\mathbf{r}) d\mathbf{r} + \sum_{j \neq k}^{N_F} \frac{\partial E_{NN}^{jk}}{\partial \mathbf{R}_a^k} \\ &= \frac{\partial E_k^F[\rho_k]}{\partial \mathbf{R}_a^k} + \int \sum_{j \neq k}^{N_F} \frac{\partial V_k^{FFP}(\mathbf{r})}{\partial \mathbf{R}_a^k} \rho_j(\mathbf{r}) d\mathbf{r} + Z_a^k \sum_{j \neq k}^{N_F} \frac{\partial V_j^{ESP}(\mathbf{R}_a^k)}{\partial \mathbf{R}_a^k}\end{aligned}\quad (24)$$

Eq. (24) gives the analytical gradients for DFI, which does not suffer the nonvariational problems in principle. However, Eq. (24) is only valid for exact $V_j^{ESP}(\mathbf{R}_a^k)$. In other words, the energy expression in Eq. (20) cannot be restored from Eq. (11) when $V_j^{ESP}(\mathbf{R}_a^k)$ is approximately computed, e.g. with the ESP-fitted point charge model in Eq. (10). The approximate ESP model always introduces errors in the energy and its analytical gradient (Note that the DFI energy in principle is variational as illustrated in Sec. II-D). The energy error can be simply tested by two terms in Eq. (7). Any approximate $V_j^{ESP}(\mathbf{R}_a^k)$ shows that the term 1 is not equal to the term 2 in Eq. (7). On the other hand, these errors can be significantly reduced by improving the accuracy of $V_j^{ESP}(\mathbf{R}_a^k)$, for instance, using the analytical $V_j^{ESP}(\mathbf{R}_a^k)$ instead of point charges for the atoms in proximity to fragment k .

III. Results and Discussion

The new DFI scheme was implemented in our in-house program QM^{4D}.⁶⁷ We first fitted the FFP parameters using a set of water dimer conformations to reproduce the total energies of full QM calculations. Then the parallel efficiency of our current DFI scheme was explored in details. Finally, the system with 256 water molecules was simulated by DFI and some liquid water properties were analyzed and were compared with the experimental results.

III-A. Fitting FFP parameters using a set of water dimer conformations

The FFP parameters are important to reproduce the correct total DFI energy when two fragments are proximal, for instance, when two water molecules form the hydrogen bond. Here, we chose several water dimer conformations to fit the FFP parameters. The object function is defined by the energy difference between DFI and full QM calculations for each water dimer conformation,

$$f = \sum_{i=1}^{N_{conf}} (E_i^{DFI} - E_i^{QM})^2 \quad (25)$$

Simulated annealing (SA)⁶⁸ was employed to minimize the above function.

Different forms can be selected for the radial functions in Eq. (15). The nonlinear fitting procedure is also complicated due to the complexity of Eqs. (15) and (25). Including in the training set sufficient water dimer conformations existing in liquid water simulations would lead to transferable fitted parameters. However, incomplete training sets may have the over-fitting issue. Furthermore, the FFP parameters, similar to ECP basis sets, depend on the specific basis sets in DFI calculations. Here, B3LYP/6-31G* combined with SG1⁶⁹ DFT grids were used in our DFI and full QM calculations. Fifteen water dimer conformations shown in the inset of Figure 1 were chosen as the training set. Although the FFP coefficients $\{d_{kl}\}$ in Eq. (15) can be either positive or negative, we constrained $\{d_{kl}\}$ to be positive with the fixed L , K_L , and $\{n_{kl}\}$ as shown in Table 1 for hydrogen and oxygen atoms, respectively, because the repulsive Pauli interactions dominate the non-additivity energy terms in Eq. (16). Finally, twelve parameters including $\{d_{kl}\}$ and $\{\zeta_{kl}\}$ (both are constrained to be

positive) for H and O were fitted using the forgoing procedure and the optimized parameters are listed in Table 1 with the Gaussian 94 format for ECP. As shown in Figure 1, the fitted FFP parameters can reproduce the entire curve for the hydrogen bonding between two water molecules as compared to full QM calculations. The inset of Figure 1 shows the details around the equilibrium distance and the points labeled by stars belong to the training sets. The largest error of energy between DFI and full QM calculations is 2.54 kcal/mol at 1.9 Å. The hydrogen bond for this set of training set is less stable in DFI compared to the full QM of B3LYP, which influences the accuracy of liquid water simulations as discussed below.

The fitting accuracy can be further improved with more parameters and different forms in FFP. A larger training set including water dimer, trimer, or even hexamer conformations might improve the transferability of FFP. In addition, as demonstrated in the optimization of effective atom centered pseudopotentials for dispersion forces by Rothlisberger and co-workers⁷⁰, one can even fit the DFI binding energies with respect to the binding energies from other methods with accurate dispersion interactions such as MP2. The works along these directions have been explored in our groups. In this work, we focus on demonstrating the new DFI scheme using the parameters in Table 1 to test the parallel efficiency and verify the liquid water properties.

III-B. Parallel efficiency

To simulate liquid water efficiently, parallel implementation must be considered here. The Message Passing Interface (MPI)⁷¹ protocol is implemented to parallelize our DFI scheme. Our implementation of DFI contains two-level parallelization: the first level is on the fragmentations of the entire system; and the second level is within each fragment. For instance, if the system with 256 water molecules is divided into 32 fragments with each fragment containing 8 water molecules, all the processors (e.g. 256) are first divided into 32 groups (i.e., the first-level parallelization). In the second layer, each fragment, as one independent QM calculation, is computed by a subgroup of 8 processors (i.e., the second-level parallelization). As such, two kinds of communications exist to collect the necessary information such as ESP charges and energy components in DFI: the inter-fragment communications and the intra-fragment communications. The parallel efficiency can be influenced by the two types of communications.

As shown in Table 2, for the system with 256 water molecules, the parallel efficiency was tested against different fragmentations with different number of processors. Each computing node has eight Intel Xeon X5560 2.80 GHz CPUs with the 24 GB physical memory and nodes are connected by InfiniBand⁷². Table 2 shows the average time cost per MD step in DFI including the energy and its gradient calculations without PBC. The last two columns of Table 2 demonstrate that the fragmentations from 32 fragments to 256 fragments significantly reduce the computational cost. For example, as shown in the last column of Table 2, for the same 512 processors, the time is reduced from 30.2 seconds to 3.6 seconds when changing 32 fragments (8 water molecules per fragment) to 256 fragments (1 water molecule per fragment). Particularly, only 3.6 seconds per MD step are required for 256 fragments with 512 CPUs including the communication time, which indicates that 20 ps MD simulations using DFI can be achieved for 256 water molecules in one day. For the given fragmentation scheme on 256 water molecules, the parallel efficiency is determined by the intra-fragment communications. For instance, for 32 fragments, the speed up is 10.2 up to 512 processors when the time of 32 processors is taken as 1.0. Generally, the parallel efficiency according to the speed up in Table 2 is larger than 80% up to 256 CPUs for 256 water molecules. For 512 CPUs, the parallel efficiency is still larger than 60% for all four different fragmentation schemes. Overall, the parallel DFI scheme shows excellent efficiency. We believe that the quantum mechanical simulations using DFI on large

molecular systems can be feasible with high efficiency and accuracy for non-covalently linked fragments.

III-C. Liquid water simulations

A cubic box of 256 water molecules was used as the model for liquid water. The system was divided into 256 fragments running on 256 CPU/cores to carry out the long time MD simulations in the NVT ensemble. The same QM calculations as the fitting procedure in Sec. III-A were employed. Long-range electrostatic interactions were not considered here (with a cut off radius of 9.86 Å). The box size is constrained to be $19.7286 \times 19.7286 \times 19.7286 \text{ Å}^3$. The temperature was maintained at 298 K with the Berendsen thermostat.⁷³ A total of 50 ps MD simulations were performed with the integration step 1 fs. The conformations were recorded every 10 fs, resulting 5,000 snapshots which were used to compute the thermodynamic properties.

The heat capacity computed by the energy fluctuation is 19.75 cal/(mol K), which is close to the value of 20.0 cal/(mol K)²¹ computed with TIP3P model. The experimental value is 17.99 cal/(mol K)^{74,75}. Compared to the charges of O (i.e., -0.834) and H (i.e., 0.417) in the TIP3P water model, the average ESP charges from DFI are -0.942 for O and 0.471 for H. The average total dipole moment for one water molecule is 2.72 ± 0.25 Debye, which is somewhat larger than the estimated values (the range is 2.3-2.5 Debye) from experiments.^{74,75} The average internal energy per water molecule is -8.78 kcal/mol, whose magnitude is slightly smaller than that of the experimental value of -9.91 kcal/mol.³⁵ The higher internal energy might be caused by the underestimated hydrogen bonding energy as shown in Figure 1. As discussed below, the weaker hydrogen bonding energy also influences the radial distinction functions of liquid water.

As one of the most important characterizations of liquid water structures, the radial distribution functions (RDFs) of OO, OH, and HH (i.e., $g_{OO}(r)$, $g_{OH}(r)$, and $g_{HH}(r)$) were plotted in Figures 2-4 along with the TIP3P (from the 80 ps simulation) and experimental results.⁸ The three peak positions of $g_{OO}(r)$ in DFI are 2.73 Å, 4.86 Å, and 6.96 Å with the heights 2.60, 1.07, and 1.02, compared to 2.73 Å, 4.50 Å, and 6.78 Å with the heights 2.75, 1.16, and 1.06 in experiments, and 2.76 Å with the height 2.68 from the TIP3P water model for the first peak. Note that the second and third peaks from TIP3P can barely be reproduced.³⁵ Although the first peak height (2.60) from DFI is slightly lower than the experimental one (2.75), the entire characteristics of $g_{OO}(r)$ computed by DFI agree much better with the experimental one than the peak from the TIP3P water model. Due to the weaker hydrogen bonding energy in DFI as shown in Fig 1, the dip of the first peak in $g_{OO}(r)$ is shallower than the experimental curve. In addition, the second and third peak positions in $g_{OO}(r)$ are shifted to the longer OO distance while the corresponding peak heights are also decreased in DFI. (Note that the experimental RDFs were used to fit the TIP3P model rather than the simple binding curves of water dimers to fit the parameters in DFI.) This analysis indicates that the FFP parameters as well as the ESP point charges in DFI need to be further tuned to reproduce the accurate first peak in $g_{OO}(r)$. For $g_{OH}(r)$ and $g_{HH}(r)$, except the first peak that need to include the nuclear quantum effects of hydrogen atoms⁴⁹, all the other characteristics of $g_{OH}(r)$ and $g_{HH}(r)$ from DFI are in good agreement with experiments. Compared to the TIP3P results, DFI significantly improve the descriptions of first peaks of $g_{OH}(r)$ and $g_{HH}(r)$. Overall, the structures and properties of liquid water computed by DFI with the current FFP parameters agree well with the measured ones.

In terms of the force field development for water molecule, the DFI scheme here can be viewed as a polarizable model. The charges in one water molecule are polarized during the simulations. The repulsive interactions between fragments can be characterized by the FFP

scheme. To fully test the DFI accuracy for liquid water simulations for other water properties (e.g., self-diffusion constant and dielectric constant) at different temperatures and pressures, we are currently developing new schemes to take the weak dispersion interactions into account and improve the charge models used for electrostatic interactions among water molecules. On the other hand, the DFI method in principle can be applied to study the properties of other liquids besides water as well as chemical reactions in solution. Here, our new DFI formulations only require fitting several FFP parameters with respect to binding energies of a set of water dimer conformations. When simulating liquid water as a proof-of-concept example, we demonstrated that DFI indeed can achieve excellent computing performance and improve the water descriptions compared with the TIP3P model.

Due to the similarities between DFI, FMO, X-POL, and DAC, it is worth comparing these linear-scaling approaches here. DAC is a rigorous approach to achieve linear scaling with controllable approximation, specially, the size of the buffer regions. DFI is the simplest version of DAC without including any buffer regions. For this reason, the charge-transfer (CT) effect between the subsystems or fragments³⁰ has been neglected in the DFI method. FMO^{22,23} employs the fragment pairs to replace the buffer treatments in DAC and captures the CT effects between fragment pairs. However, to the best of our knowledge, for CT between fragments separated by long distance or several intermediate fragments, the efficiency of FMO remains to be tested. X-POL^{21,26} is based on the wave function theory and uses the Mulliken charges and classical van der Waals terms to model the fragment-fragment interactions. The FFP approach in DFI is efficient to obtain the correct binding curve for water dimers, more importantly, the accurate electron densities. Generally, for any linear-scaling algorithms by dividing the entire system into several subsystems, they all share the same physical intuition, the nearsightedness of electron interaction, as discussed above. The differences among these approaches reside in how the interactions between subsystems are modeled. The balance between efficiency and accuracy should be considered for the method to be suitable for molecular dynamics (MD) simulations.

IV. Conclusion

In this paper, the new DFI scheme was proposed and tested for liquid water simulations. The electrostatic interactions between fragments are calculated by simple ESP charges to improve the speed of simulations with moderate accuracy. As a result of the fragmentation, the effects of non-additivity of kinetic and exchange-correlation functionals are approximated by the FFP method. The FFP parameters have been fitted according to the full QM calculations on simple model systems. For instance, we used a set of water dimer conformations to fit the FFP parameters for hydrogen and oxygen atoms. Our DFI scheme was also parallelized using MPI, with excellent parallel performance. Finally, we performed the liquid water simulations of a system with 256 water molecules using PBC and parallel DFI/6-31G*/SG1/B3LYP. The computed structures and properties of liquid water are in good agreements with the experimental results, demonstrating that *ab initio* QM MD simulations can be feasible with the combination of our DFI method and parallel techniques.

It is possible to improve further the accuracy of the electrostatic calculations between fragments. Other FFP forms need to be explored with larger training set to obtain more transferable and accurate FFP parameters for different atomic types. The FFP parameters in DFI can also be fitted to reproduce the high-level binding energy curve, for instance, computed by MP2 or CC method. Subsequently, the accurate dispersion interactions can be included naturally in DFI by FFP. In addition, for the systems with the covalently linked fragments or highly delocalized electrons, DFI may be problematic and require new methodology developments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial supports from the National Institute of Health (R01-GM061870 to W.Y.) and the Research Grant Council of Hong Kong (to H. H.) are greatly appreciated. We thank the Brigham Young University Fulton Supercomputing Lab, the Research Computing for UNC Chapel Hill, and the computer center at the University of Hong Kong for providing computational resources. X.H. thanks the 2010 HPC Advisory Council University Award and personally thanks the helps from Pak Lui.

References

1. Wang ML, Hu XQ, Beratan DN, Yang WT. J Am Chem Soc. 2006; 128:3228. [PubMed: 16522103]
2. Carter EA. Science. 2008; 321:800. [PubMed: 18687955]
3. Cohen AJ, Mori-Sanchez P, Yang WT. Science. 2008; 321:792. [PubMed: 18687952]
4. Hu H, Yang WT. Annu Rev Phys Chem. 2008; 59:573. [PubMed: 18393679]
5. Hu XQ, Beratan DN, Yang WT. J Chem Phys. 2008; 129
6. Hu XQ, Beratan DN, Yang WT. Sci China, Ser B: Chem. 2009; 52:1769.
7. Chen ZF, Concepcion JJ, Hu XQ, Yang WT, Hoertz PG, Meyer TJ. Proc Natl Acad Sci U S A. 2010; 107:7225. [PubMed: 20360565]
8. Soper AK. Chem Phys. 2000; 258:121.
9. Parr RG, Yang WT. Annu Rev Phys Chem. 1995; 46:701.
10. Yang WT. Phys Rev Lett. 1991; 66:1438. [PubMed: 10043209]
11. Yang WT. Phys Rev A. 1991; 44:7823. [PubMed: 9905920]
12. York D, Lu JP, Yang WT. Phys Rev B. 1994; 49:8526.
13. Yang W, Lee TS. J Chem Phys. 1995; 103:5674.
14. Lee TS, York DM, Yang WT. J Chem Phys. 1996; 105:2744.
15. Zhao QS, Yang WT. J Chem Phys. 1995; 102:9598.
16. Gogonea V, Merz KM. J Phys Chem A. 1999; 103:5171.
17. Burger SK, Weitao Y. J Phys : Condens Matter. 2008; 20:294209–4.
18. Kobayashi M, Kunisada T, Akama T, Sakura D, Nakai H. J Chem Phys. 2011; 134:034105. [PubMed: 21261328]
19. Guerra CF, Snijders JG, te Velde G, Baerends EJ. Theor Chem Acc. 1998; 99:391.
20. Goedecker S. Rev Mod Phys. 1999; 71:1085.
21. Gao JL. J Chem Phys. 1998; 109:2346.
22. Kitaura K, Sugiki SI, Nakano T, Komeiji Y, Uebayasi M. Chem Phys Lett. 2001; 336:163.
23. Nakano T, Kaminuma T, Sato T, Fukuzawa K, Akiyama Y, Uebayasi M, Kitaura K. Chem Phys Lett. 2002; 351:475.
24. Nagata T, Fedorov DG, Sawada T, Kitaura K, Gordon MS. J Chem Phys. 2011; 134:034110. [PubMed: 21261333]
25. Zhang DW, Zhang JZH. J Chem Phys. 2003; 119:3599.
26. Xie WS, Song LC, Truhlar DG, Gao JL. J Chem Phys. 2008; 128
27. Gordon MS, Mullin JM, Pruitt SR, Roskop LB, Slipchenko LV, Boatz JA. J Phys Chem B. 2009; 113:9646. [PubMed: 19368406]
28. Mullin JM, Roskop LB, Pruitt SR, Collins MA, Gordon MS. J Phys Chem A. 2009; 113:10040. [PubMed: 19739681]
29. Hirata S. Phys Chem Chem Phys. 2009; 11:8397. [PubMed: 19774268]
30. Fujimoto K, Yang WT. J Chem Phys. 2008; 129
31. Narten AH, Levy HA. J Chem Phys. 1971; 55:2263.
32. Soper AK, Phillips MG. Chem Phys. 1986; 107:47.

33. Soper AK, Bruni F, Ricci MA. J Chem Phys. 1997; 106:247.
34. Hura G, Sorenson JM, Glaeser RM, Head-Gordon T. J Chem Phys. 2000; 113:9140.
35. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. J Chem Phys. 1983; 79:926.
36. Dang LX, Chang TM. J Chem Phys. 1997; 106:8149.
37. Burnham CJ, Xantheas SS. J Chem Phys. 2002; 116:1500.
38. Ren PY, Ponder JW. J Phys Chem B. 2003; 107:5933.
39. Kaminski GA, Stern HA, Berne BJ, Friesner RA. J Phys Chem A. 2004; 108:621.
40. Defusco A, Schofield DP, Jordan KD. Mol Phys. 2007; 105:2681.
41. Burnham CJ, Anick DJ, Mankoo PK, Reiter GF. J Chem Phys. 2008; 128
42. Kumar R, Wang FF, Jenness GR, Jordan KD. J Chem Phys. 2010; 132
43. Chialvo AA, Horita J. J Chem Phys. 2010; 133
44. Hu H, Lu ZY, Elstner M, Hermans J, Yang WT. J Phys Chem A. 2007; 111:5685. [PubMed: 17474727]
45. Laasonen K, Sprik M, Parrinello M, Car R. J Chem Phys. 1993; 99:9080.
46. Sprik M, Hutter J, Parrinello M. J Chem Phys. 1996; 105:1142.
47. Silvestrelli PL, Parrinello M. J Chem Phys. 1999; 111:3572.
48. Silvestrelli PL, Parrinello M. Phys Rev Lett. 1999; 82:3308.
49. Izvekov S, Voth GA. J Chem Phys. 2002; 116:10372.
50. Berendsen, HJC.; Postma, JPM.; van Gunsteren, WF.; Hermans, J. In Intermolecular Forces. Pullman, B., editor. Reidel: Dordrecht; 1981. p. 331
51. Delley B. J Chem Phys. 1990; 92:508.
52. Hu H, Lu ZY, Yang WT. J Chem Theory Comput. 2007; 3:1004.
53. Cortona P. Phys Rev B. 1991; 44:8454.
54. Cortona P. Phys Rev B. 1992; 46:2008.
55. Cortona P, Monteleone AV. Int J Quantum Chem. 1994; 52:987.
56. Cui Q, Guo H, Karplus M. J Chem Phys. 2002; 117:5617.
57. Govind N, Wang YA, Carter EA. J Chem Phys. 1999; 110:7677.
58. Wesolowski TA, Warshel A. J Phys Chem. 1993; 97:8050.
59. Kahn LR, Baybutt P, Truhlar DG. J Chem Phys. 1976; 65:3826.
60. Kahn LR, Goddard WA. J Chem Phys. 1972; 56:2685.
61. Melius CF, Goddard WA. Phys Rev A. 1974; 10:1528.
62. Schwarz WHE. Theor Chim Acta. 1968; 11:307.
63. Schwarz WHE. Theor Chim Acta. 1968; 11:377.
64. Parks JM, Hu H, Cohen AJ, Yang WT. J Chem Phys. 2008; 129:6.
65. Zhang YK, Lee TS, Yang WT. J Chem Phys. 1999; 110:46.
66. Flores-Moreno R, Alvarez-Mendez RJ, Vela A, Koster AM. J Comput Chem. 2006; 27:1009. [PubMed: 16628539]
67. QM⁴D: an integrated and versatile quantum mechanical/molecular mechanical simulation package (<http://www.qm4d.info/>)
68. Goffe WL, Ferrier GD, Rogers J. J Econometrics. 1994; 60:65.
69. Gill PMW, Johnson BG, Pople JA. Chem Phys Lett. 1993; 209:506.
70. von Lilienfeld OA, Tavernelli I, Rothlisberger U, Sebastiani D. Phys Rev Lett. 2004; 93:153004. [PubMed: 15524874]
71. <http://www.mcs.anl.gov/research/projects/mpi/index.html>
72. <http://www.intel.com/technology/infiniband/>
73. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. J Chem Phys. 1984; 81:3684.
74. Caldwell J, Dang LX, Kollman PA. J Am Chem Soc. 1990; 112:9144.
75. Coulson CA, Eisenber D. P Roy Soc Lond A Mat. 1966; 291:445.

Figure 1.
The OH binding curves for two water molecules by full QM and DFI calculations.

Figure 2.

The oxygen-oxygen RDF from DFI simulations of a system of 256 water molecules (solid black line) compared to the TIP3P (blue dashed dot line) and experiment data in Ref. 8 (red dashed line) at $T=298\text{K}$.

Figure 3.

The oxygen-hydrogen RDF from DFI simulations of a system of 256 water molecules (solid black line) compared to the TIP3P (blue dashed dot line) and experiment data in Ref. 8 (red dashed line) at $T=298\text{K}$.

Figure 4. The hydrogen-hydrogen RDF from DFI simulations of a system of 256 water molecules (solid black line) compared to the TIP3P (blue dashed dot line) and experiment data in Ref. 8 (red dashed line) at T=298K.

Table 1

The fitted FFP parameters for H and O atoms using B3LYP/6-31G*/SG1 in Gaussian 94 format for ECP.

Hydrogen	H 0		
	S 0 0		
	1		
	2	3.4629	0.54392
Oxygen	O 0		
	D 2 0		
	1		
	1	0.57423	1.2785
	SD_Projection		
	2		
	0	836.38	0.30840
	2	560.22	390.00
	PD_Projection		
	2		
	0	2.6310	17.317
	2	793.12	987.21

Table 2

The average time per MD step in second for fragmentations versus number of CPUs. 256 water molecules are included. The “Speed up” values are computed for each individual fragmentation.

Fragmentation\#CPU	32	64	128	256	512
32	Time (s)	309.1	159.2	85.9	48.7
	Speed up	1.0	1.9	3.6	6.3
64	Time (s)	N/A	43.2	22.7	13.1
	Speed up	N/A	1.0	1.9	3.3
128	Time (s)	N/A	N/A	10.9	6.4
	Speed up	N/A	N/A	1.0	1.7
256	Time (s)	N/A	N/A	N/A	5.3
	Speed up	N/A	N/A	N/A	1.0