3-1998

An Approximate Formula for the Intermolecular Pauli Repulsion Between Closed Shell Molecules. II. Application to the Effective Fragment Potential Method

Jan H. Jensen
Iowa State University

Mark S. Gordon
Iowa State University, mgordon@iastate.edu

An Approximate Formula for the Intermolecular Pauli Repulsion Between Closed Shell Molecules. II. Application to the Effective Fragment Potential Method

Abstract
The accuracy and efficiency of an approximate formula for the intermolecular Pauli repulsion between closed shell molecules, derived earlier [Mol. Phys. **89**, 1313 (1996)], is demonstrated for dimers of H2O, CH3OH, CH2Cl2, CH3CN, (CH3)2CO, and (CH3)2SO. The energy derivative with respect to a Cartesian coordinate and rigid rotation about the center-of-mass (torques) are presented. The Pauli repulsion energy term is then combined with the Coulomb and classical induction energy terms of the effective fragment potential method [J. Chem. Phys. **105**, 1968, 11081 (1996)] to give a general intermolecular interaction potential. This potential is applied to water and methanol clusters.

Keywords
Intermolecular forces

Disciplines
Chemistry

Comments
The following article appeared in *Journal of Chemical Physics* 108 (1998: 4772, and may be found at doi:10.1063/1.475888.

An approximate formula for the intermolecular Pauli repulsion between closed shell molecules. II. Application to the effective fragment potential method

Jan H. Jensen and Mark S. Gordon

Citation: The Journal of Chemical Physics **108**, 4772 (1998); doi: 10.1063/1.475888
View online: http://dx.doi.org/10.1063/1.475888
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/108/12?ver=pdfcov
Published by the AIP Publishing

An approximate formula for the intermolecular Pauli repulsion between closed shell molecules. II. Application to the effective fragment potential method

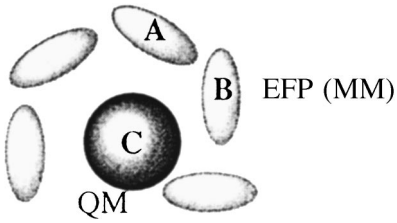
Jan H. Jensen[a] and Mark S. Gordon[b]
Department of Chemistry, Iowa State University, Ames, Iowa 50011

(Received 21 October 1997; accepted 22 December 1997)

The accuracy and efficiency of an approximate formula for the intermolecular Pauli repulsion between closed shell molecules, derived earlier [Mol. Phys. **89**, 1313 (1996)], is demonstrated for dimers of H_2O, CH_3OH, CH_2Cl_2, CH_3CN, $(CH_3)_2CO$, and $(CH_3)_2SO$. The energy derivative with respect to a Cartesian coordinate and rigid rotation about the center-of-mass (torques) are presented. The Pauli repulsion energy term is then combined with the Coulomb and classical induction energy terms of the effective fragment potential method [J. Chem. Phys. **105**, 1968, 11081 (1996)] to give a general intermolecular interaction potential. This potential is applied to water and methanol clusters. © *1998 American Institute of Physics.* [S0021-9606(98)01812-1]

I. INTRODUCTION

Most chemical reactions of interest to chemists do not take place in a vacuum. For example, the reacting molecules are frequently surrounded by solvent molecules or interacting with a solid surface, and these molecular environments can greatly influence the way in which the molecules react. However, including the molecular environment in quantum mechanical simulations is often computationally prohibitive, and it is usually not done. Fortunately, it seems that the use of quantum mechanics is often not required to model these environmental effects. Rather, the molecular environment can be replaced by semiclassical potentials, while retaining a quantum mechanical description of the molecules that undergo a chemical change. This forms the basis for the increasingly numerous ''QM/MM'' methods[1] schematically represented here.



In this cartoon, the species C is treated by *ab initio* (QM) methods, while the surrounding environment is described using effective fragment potentials (EFP) or molecular mechanics (MM).

The effective fragment potential (EFP) method[2] is one such method, that has been applied to the study of both solvent[3] and protein[4] effects. Currently, an EFP consists of three potentials; a multipole expansion of (1) the molecular electrostatic potential and (2) the classical induction energy; as well as (3) a repulsive potential. The form of the first two potentials have two very attractive features: (1) The potentials all have the form of truncated expansions and can therefore in principle be systematically improved. (2) The potentials are obtained from separate *ab initio* calculations on isolated molecules; thus there is no system-dependent fitting involved in obtaining these two potentials. The repulsive potentials used to date[3] consist of fitted functions obtained for some model system, and while they work very well, their construction is the main obstacle in constructing new EFPs.

In a previous paper[5] (hereafter referred to as Paper I) we presented a repulsive potential that shares the two attractive features outlined above. In this study we integrate this potential into the EFP method, replacing the fitted repulsive potential, to yield *a completely general intermolecular interaction potential that depends solely on the properties of the isolated molecules*, made available by separate *ab initio* calculations. In order for this new potential to be generally applicable within the EFP methodology the following issues must be considered:

(1) The evaluation of energy derivatives to facilitate geometry optimizations and molecular dynamics calculations.
(2) The accuracy of the exchange repulsion energy and the total EFP/EFP interaction energy for a variety of systems.
(3) The computational cost of evaluating the energy and gradients.
(4) The relative importance of pair and many-body contributions to the exchange repulsion energy.
(5) The derivation and implementation of an exchange repulsion operator corresponding to the exchange repulsion potential, for the *ab initio*/EFP interaction.

After a brief summary of the repulsive potential (Sec. III) we address the first three points. The last two points will be addressed in subsequent studies.

II. COMPUTATIONAL METHODOLOGY

Four basis sets were used for the calculations reported in this study; 6-31G(d,p),[6] 6-31+G(d,p),[7] 6-31+

[a]Present address: Department of Chemistry, University of Iowa, Iowa City, Iowa 52242.
[b]Author to whom correspondence should be addressed.

$+G(2d,2p)$,[8] and Sadlej's[9] polarized valence triple zeta (pVTZ) basis set. The latter basis set contains $(14s10p4d/10s6p4d/6s4p)$ primitive Gaussians contracted to $[7s5p2d/5s3p2d/3s2p]$ for elements in the third/second/first row of the Periodic Table. The geometries used for the dimer calculations in Sec. IV were all obtained at the RHF/6-31+G(d,p) level of theory, using the free monomer geometries also calculated using RHF/6-31+G(d,p).

The EFP parameters have the following attributes:[2]

(1) Distributed multipole expansions up to and including octupoles, at all atom centers and bond midpoints.[10] The multipole expansions are not corrected for charge penetration effects.
(2) Dipole polarizability tensors (calculated analytically) due to each valence localized molecular orbital[11] (LMO) at their respective LMO centroids of charge.
(3) Atomic basis set, LMO coefficients, positions of atoms and LMO centroids of charge, and LMO Fock matrix elements.

An EFP is constructed by calculating these properties *once* (and in one run) for the isolated molecule in question. It is then included in an input file for subsequent EFP calculations—much like a basis set. In order to succinctly describe the EFPs we introduce the following general notation, for example, EFP[RHF/pVTZ//RHF/6-31+G(d,p)] indicates that all the EFP parameters outlined above are calculated at the RHF/pVTZ level of theory using the RHF/6-31+G(d,p) optimized geometry of the free monomer. Since *all* EFPs in this study are constructed at the RHF level of theory, we abbreviate the previous expression as EFP[pVTZ//6-31+G(d,p)]. As an additional abbreviation we use, for example, EFP[6-31G(d,p)] for EFP[6-31G(d,p)//6-31G(d,p)]. Finally, the neglect of core MOs in the exchange repulsion potential is denoted EFP(nc)[6-31G(d,p)].

The localized molecular orbitals (LMOs) were obtained with the energy localization scheme due to Edmiston and Ruedenberg.[12] All calculations were performed with the quantum chemistry program GAMESS.[13]

III. REVIEW OF THE UNDERLYING APPROXIMATIONS TO THE EXACT EXCHANGE REPULSION ENERGY

In order to discuss the accuracy of the approximate exchange repulsion energy (see below) we present a summary of all the inherent approximations. For further details and notation refer to Paper I.

The exact zeroth order exchange repulsion energy between wave functions Ψ_A and Ψ_B (here assumed to be RHF wave functions) is extracted from the Heitler–London energy by subtracting the classical Coulomb energy as well as the energies of molecules A and B,

$$E_{\mathrm{exch}}=\frac{\langle\Psi_A\Psi_B|\mathscr{A}H_{AB}|\Psi_A\Psi_B\rangle}{\langle\Psi_A\Psi_B|\mathscr{A}\Psi_A\Psi_B\rangle}-\langle\Psi_A\Psi_A|V_{AB}|\Psi_B\Psi_B\rangle - E_A-E_B. \tag{1}$$

Here H_{AB} is the super molecular Hamiltonian for the A–B complex, given by the individual Hamiltonians for molecules A and B plus the interaction operator,

$$H_{AB}=H_A+H_B+V_{AB}. \tag{2}$$

Using this division for H_{AB} we can express the exchange repulsion energy in terms of two internal energy contributions as well as an interaction term,

$$\begin{aligned}E_{\mathrm{exch}}&=\frac{\langle\Psi_A\Psi_B|\mathscr{A}(H_A+H_B)|\Psi_A\Psi_B\rangle}{\langle\Psi_A\Psi_B|\mathscr{A}\Psi_A\Psi_B\rangle}-E_A-E_B\\&\quad+\frac{\langle\Psi_A\Psi_B|\mathscr{A}V_{AB}|\Psi_A\Psi_B\rangle}{\langle\Psi_A\Psi_B|\mathscr{A}\Psi_A\Psi_B\rangle}\\&\quad-\langle\Psi_A\Psi_A|V_{AB}|\Psi_B\Psi_B\rangle\\&=\Delta E_A+\Delta E_B+E_{\mathrm{exch}}(V).\end{aligned} \tag{3}$$

The exchange repulsion energy arises from terms in the wave function generated by the antisymmetrizer, $\mathscr{A}$, which permutes 0, 1, 2,... electron pairs,

$$\mathscr{A}=1-P_1+P_2-\cdots. \tag{4}$$

Here, only the effect of zero and one electron pair permutations are included, and this leads to an approximate exchange repulsion proportional to the square of the intermolecular overlap (S),

$$\mathscr{A}\approx 1-P_1\Rightarrow E_{\mathrm{exch}}\approx E_{\mathrm{exch}}[\mathscr{O}(S^2)], \tag{5}$$

where [cf. Eq. (3)],

$$\begin{aligned}E_{\mathrm{exch}}[\mathscr{O}(S^2)]&=\Delta E_A[\mathscr{O}(S^2)]+\Delta E_B[\mathscr{O}(S^2)]\\&\quad+E_{\mathrm{exch}}[V;\mathscr{O}(S^2)].\end{aligned} \tag{6}$$

For exact HF wave functions the internal energy contribution to the second-order exchange energy vanishes,

$$\begin{aligned}\Delta E_A[\mathscr{O}(S^2)]&=-2\sum_{i\in A}\sum_{j\in B}S_{ij}F^A_{ij}\\&\quad+2\sum_{i\in A}\sum_{j\in B}S_{ij}\sum_{k\in A}S_{kj}F^A_{ik}\end{aligned} \tag{7}$$

{and similarly for $\Delta E_B[\mathscr{O}(S^2)]$}. This is based on the observation, first made by Landshoff,[14] that if the HF equation,

$$\hat{F}^A\phi_i=\sum_k F^A_{ik}\phi_k \tag{8}$$

is solved exactly, one can obtain the following expression for Fock matrix elements connecting nonorthogonal MOs,

$$F^A_{ij}=\sum_k F^A_{ik}S_{kj}. \tag{9}$$

In practice, the HF MOs are expanded in a finite basis set, and neglecting the internal energy contributions in Eq. (6),

$$E_{\mathrm{exch}}[\mathscr{O}(S^2)]\approx E_{\mathrm{exch}}[V;\mathscr{O}(S^2)] \tag{10}$$

is another, basis set dependent, approximation. This expression for the exchange repulsion energy can be grouped into three distinct energy terms based on their *explicit* overlap dependence (they all scale as S^2),

$$E_{\mathrm{exch}}[V;\mathscr{O}(S^2)]=E_{\mathrm{exch}}(S^0)+E_{\mathrm{exch}}(S^1)+E_{\mathrm{exch}}(S^2) \tag{11}$$

and approximated separately in terms of localized molecular orbitals (LMOs),

$$E_{\mathrm{exch}}(S^0)=-2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}}\langle i|K_j|i\rangle \approx -2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} 2\sqrt{\frac{-2\ln S_{ij}}{\pi}}\frac{S_{ij}^2}{R_{ij}}, \tag{12}$$

$$E_{\mathrm{exch}}(S^1)=-2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} S_{ij}\Bigg[V_{ij,\mathrm{A}}+V_{ij,\mathrm{B}}+\sum_{k\in \mathrm{A}}\langle i|2J_k-K_k|j\rangle+\sum_{l\in \mathrm{B}}\langle i|2J_l-K_l|j\rangle\Bigg]$$

$$\Bigg\{=-2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} S_{ij}[F_{ij}^A+F_{ij}^B-2T_{ij}]\Bigg\}$$

$$\approx -2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} S_{ij}\Bigg[\sum_{k\in \mathrm{A}} F_{ik}^{\mathrm{A}}S_{kj}+\sum_{l\in \mathrm{B}} F_{jl}^{\mathrm{B}}S_{li}-2T_{ij}\Bigg], \tag{13}$$

$$E_{\mathrm{exch}}(S^2)=2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} S_{ij}\Bigg[\sum_{k\in \mathrm{A}} S_{kj}\Bigg(V_{ik,\mathrm{B}}+\sum_{l\in \mathrm{B}}\langle i|2J_l|k\rangle\Bigg)+\sum_{l\in \mathrm{B}} S_{il}\Bigg(V_{jl,\mathrm{A}}+\sum_{k\in \mathrm{A}}\langle j|2J_k|l\rangle\Bigg)-\sum_{k\in \mathrm{A}}\sum_{l\in \mathrm{B}} S_{kl}\langle ik|lj\rangle\Bigg]$$

$$\Bigg\{\approx 2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} S_{ij}^2\Bigg[V_{ii,\mathrm{B}}+\sum_{l\in \mathrm{B}}\langle i|2J_l|i\rangle+V_{jj,\mathrm{A}}+\sum_{k\in \mathrm{A}}\langle j|2J_k|j\rangle-\langle i|J_j|i\rangle\Bigg]\Bigg\}$$

$$\approx 2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} S_{ij}^2\Bigg[\sum_{J\in \mathrm{B}} -Z_JR_{iJ}^{-1}+2\sum_{l\in \mathrm{B}} R_{il}^{-1}+\sum_{I\in \mathrm{A}} -Z_IR_{Ij}^{-1}+2\sum_{k\in \mathrm{A}} R_{kj}^{-1}-R_{ij}^{-1}\Bigg]. \tag{14}$$

Here, S_{ij} and T_{ij} are, respectively, an overlap and kinetic energy integral connecting LMOs i and j. R_{ij} is the distance between the centroids of charge of LMOs i and j, and R_{Ij} is the distance between nucleus I (with nuclear charge Z_I) and the centroid of charge of LMO j. Finally, F_{ik}^{A} is the Fock matrix element connecting LMO i and k resulting from the Hamiltonian of molecule A. The approximations in Eq. (12) and Eq. (14) are based on the assumption that LMOs are used. Combining these three approximations leads to the approximate formula for exchange repulsion between closed shell molecules presented in Paper I,

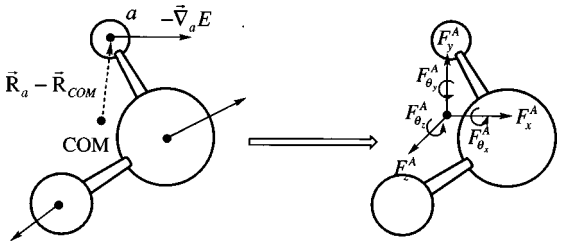


FIG. 1. Schematic representation of the transformation of the Cartesian gradient components on an EFP to coordinate components defined relative to the EFP center-of-mass (COM) [Eqs. (16)–(17)].

$$E_{\mathrm{exch}}\approx -2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} 2\sqrt{\frac{-2\ln S_{ij}}{\pi}}\frac{S_{ij}^2}{R_{ij}} -2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} S_{ij}\Bigg[\sum_{k\in \mathrm{A}} F_{ik}^{\mathrm{A}}S_{kj}+\sum_{l\in \mathrm{B}} F_{jl}^{\mathrm{B}}S_{li}-2T_{ij}\Bigg]$$

$$+2\sum_{i\in \mathrm{A}}\sum_{j\in \mathrm{B}} S_{ij}^2\Bigg[\sum_{J\in \mathrm{B}} -Z_JR_{iJ}^{-1}+2\sum_{l\in \mathrm{B}} R_{il}^{-1}+\sum_{I\in \mathrm{A}} -Z_IR_{Ij}^{-1}+2\sum_{k\in \mathrm{A}} R_{kj}^{-1}-R_{ij}^{-1}\Bigg]. \tag{15}$$

IV. RESULTS

A. Energy derivatives

The internal geometry of an EFP is always fixed, but it is allowed to rotate and translate relative to other EFPs, resulting in six degrees of freedom per EFP. The corresponding energy derivatives, the net force ($\mathbf{F}^{\mathrm{A}}$) and torque about the center-of-mass ($\mathbf{F}_\theta^{\mathrm{A}}$) of EFP A (see Fig. 1), are obtained by

$$\mathbf{F}^{\mathrm{A}}=\sum_{a\in \mathrm{A}}(-\boldsymbol{\nabla}_aE), \tag{16}$$

$$\mathbf{F}_\theta^{\mathrm{A}}=\sum_{a\in \mathrm{A}}(\mathbf{R}_a-\mathbf{R}_{\mathrm{COM}})\times(-\boldsymbol{\nabla}_aE)+\sum_{a\in \mathrm{A}}\boldsymbol{\tau}_a. \tag{17}$$

The subscript a denotes a point in EFP A, which in the case of the exchange repulsion potential can be either the centroid of a LMO (λ_a) or the center of atomic orbital χ_a used as a basis for the LMOs. $\mathbf{R}_{\mathrm{COM}}$ is the center-of-mass position vector of EFP A. The energy derivatives with respect to LMO centroid position λ_a on EFP A is given by

$$\frac{\partial E_{\mathrm{exch}}}{\partial x_a}=2\sum_{j\in \mathrm{B}} 2\sqrt{\frac{-2\ln S_{aj}}{\pi}}(x_a-x_j)\frac{S_{aj}^2}{R_{aj}^3} +2\sum_{j\in \mathrm{B}}\Bigg\{S_{aj}^2\Bigg[\sum_{J\in \mathrm{B}} Z_J(x_a-x_J)R_{aJ}^{-3} -2\sum_{l\in \mathrm{B}}(x_a-x_l)R_{al}^{-3}+(x_a-x_j)R_{aj}^{-3}\Bigg] -2\sum_{k\in \mathrm{A}} S_{kj}^2(x_a-x_j)R_{aj}^{-3}\Bigg\}. \tag{18}$$

The energy derivative with respect to the center of χ_a on EFP A is given by,

$$\frac{\partial E_{\text{exch}}}{\partial x_a} = -2\sum_{i\in \text{A}}\sum_{j\in \text{B}}\left(4\sqrt{\frac{-2\ln S_{ij}}{\pi}} - \sqrt{\frac{2}{-\pi\ln S_{ij}}}\right)$$
$$\times\frac{S_{ij}S^a_{ij}}{R_{ij}} - 2\sum_{i\in \text{A}}\sum_{j\in \text{B}}\left\{S^a_{ij}\left[2\sum_{k\in \text{A}}F^{\text{A}}_{ik}S_{kj}\right.\right.$$
$$\left.+2\sum_{l\in \text{B}}F^{\text{B}}_{jl}S_{il} - 2T_{ij}\right] - 2S_{ij}T^a_{ij}\Bigg\}$$
$$+2\sum_{i\in \text{A}}\sum_{j\in \text{B}}\left\{2S_{ij}S^a_{ij}\left[\sum_{J\in \text{B}} - Z_J R^{-1}_{iJ} + 2\sum_{l\in \text{B}} R^{-1}_{il}\right.\right.$$
$$\left.+\sum_{I\in \text{A}} - Z_I R^{-1}_{Ij} + 2\sum_{k\in \text{A}} R^{-1}_{kj} - R^{-1}_{ij}\right]$$
$$\left.+S^2_{ij}[-Z_a(x_j - x_a)R^{-3}_{aj}]\right\}. \tag{19}$$

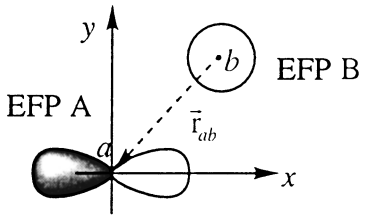
Here the superscript a is the usual[15] shorthand notation for a derivative,

$$S^a_{ij} = \sum_{\mu\in a}\sum_{\nu\in B} C_{\mu i}C_{\nu j}\left\langle\frac{\partial\chi_\mu}{\partial x_a}\middle|\chi_\nu\right\rangle, \tag{20}$$

$$T^a_{ij} = \sum_{\mu\in a}\sum_{\nu\in B} C_{\mu i}C_{\nu j}\left\langle\frac{\partial\chi_\mu}{\partial x_a}\middle|-\frac{1}{2}\nabla^2\middle|\chi_\nu\right\rangle. \tag{21}$$

Since the EFP LMOs are frozen, the derivative of $C_{\mu i}$ is zero.

The point torque $\boldsymbol{\tau}_a$ in Eq. (17) is the torque about an AO center due to the anisotropy of p, d, etc. AOs. Consider the simple case of an EFP consisting of a p_x function interacting with another EFP consisting of a s function, shown here,



The overlap, and hence the exchange repulsion energy, will clearly change if, the p function is rotated about the y- or z-axis. The resulting torque on A about point a ($\boldsymbol{\tau}_a$) for this system is given by

$$\boldsymbol{\tau}_a = -(\text{torque on B about } a)$$
$$= -(\mathbf{r}_b - \mathbf{r}_a)\times(-\boldsymbol{\nabla}_b E_{\text{exch}}) = (\mathbf{r}_a - \mathbf{r}_b)\times(\boldsymbol{\nabla}_a E_{\text{exch}})$$
$$= \mathbf{r}_{ab}\times(\boldsymbol{\nabla}_a E_{\text{exch}}). \tag{22}$$

In general, using Eq. (15), the torque is given by

TABLE I. Integrals needed to evaluate Eq. (24) ($\hat{O}=1$) and Eq. (25) ($\hat{O}=-\frac{1}{2}\nabla^2$) for $\chi_\mu = s$, p, and d functions.

	$[\mathbf{r}_{ab}\times\langle\boldsymbol{\nabla}_a\chi_\mu\|\hat{O}\|\chi_\nu\rangle]_{x,y,z}$		
χ_μ	x	y	z
s	0	0	0
p_x	0	$-\langle p_z\|\hat{O}\|\chi_\nu\rangle$	$\langle p_y\|\hat{O}\|\chi_\nu\rangle$
p_y	$\langle p_z\|\hat{O}\|\chi_\nu\rangle$	0	$-\langle p_x\|\hat{O}\|\chi_\nu\rangle$
p_z	$-\langle p_y\|\hat{O}\|\chi_\nu\rangle$	$\langle p_x\|\hat{O}\|\chi_\nu\rangle$	0
d_{x^2}	0	$-2\langle d_{xz}\|\hat{O}\|\chi_\nu\rangle$	$2\langle d_{xy}\|\hat{O}\|\chi_\nu\rangle$
d_{y^2}	$2\langle d_{yz}\|\hat{O}\|\chi_\nu\rangle$	0	$-2\langle d_{xy}\|\hat{O}\|\chi_\nu\rangle$
d_{z^2}	$-2\langle d_{yz}\|\hat{O}\|\chi_\nu\rangle$	$2\langle d_{xz}\|\hat{O}\|\chi_\nu\rangle$	0
d_{xy}	$\langle d_{xz}\|\hat{O}\|\chi_\nu\rangle$	$-\langle d_{yz}\|\hat{O}\|\chi_\nu\rangle$	$\langle d_{y^2}\|\hat{O}\|\chi_\nu\rangle - \langle d_{x^2}\|\hat{O}\|\chi_\nu\rangle$
d_{xz}	$-\langle d_{xy}\|\hat{O}\|\chi_\nu\rangle$	$\langle d_{x^2}\|\hat{O}\|\chi_\nu\rangle - \langle d_{z^2}\|\hat{O}\|\chi_\nu\rangle$	$\langle d_{yz}\|\hat{O}\|\chi_\nu\rangle$
d_{yz}	$\langle d_{z^2}\|\hat{O}\|\chi_\nu\rangle - \langle d_{y^2}\|\hat{O}\|\chi_\nu\rangle$	$\langle d_{xy}\|\hat{O}\|\chi_\nu\rangle$	$-\langle d_{xz}\|\hat{O}\|\chi_\nu\rangle$

$$\boldsymbol{\tau}_a = -2\sum_{i\in \text{A}}\sum_{j\in \text{B}}\left(4\sqrt{\frac{-2\ln S_{ij}}{\pi}} - \sqrt{\frac{2}{-\pi\ln S_{ij}}}\right)\frac{S_{ij}\mathbf{S}^{\tau a}_{ij}}{R_{ij}}$$
$$-2\sum_{i\in \text{A}}\sum_{j\in \text{B}}\left\{\mathbf{S}^{\tau a}_{ij}\left[2\sum_{k\in \text{A}}F^{\text{A}}_{ik}S_{kj} + 2\sum_{l\in \text{B}}F^{\text{B}}_{jl}S_{il} - 2T_{ij}\right]\right.$$
$$\left.-2S_{ij}\mathbf{T}^a_{ij}\right\} + 2\sum_{i\in \text{A}}\sum_{j\in \text{B}}\left\{2S_{ij}\mathbf{S}^{\tau a}_{ij}\left[\sum_{J\in \text{B}} - Z_J R^{-1}_{iJ}\right.\right.$$
$$\left.\left.+2\sum_{l\in \text{B}} R^{-1}_{il} + \sum_{I\in \text{A}} - Z_I R^{-1}_{Ij} + 2\sum_{k\in \text{A}} R^{-1}_{kj} - R^{-1}_{ij}\right]\right\}, \tag{23}$$

where

$$\mathbf{S}^{\tau a}_{ij} = \sum_{\mu\in a}\sum_{b\in \text{B}}\sum_{\nu\in b} C_{\mu i}C_{\nu j}(\mathbf{r}_{ab}\times\langle\boldsymbol{\nabla}_a\chi_\mu|\chi_\nu\rangle), \tag{24}$$

$$\mathbf{T}^{\tau a}_{ij} = \sum_{\mu\in a}\sum_{b\in \text{B}}\sum_{\nu\in b} C_{\mu i}C_{\nu j}(\mathbf{r}_{ab}\times\langle\boldsymbol{\nabla}_a\chi_\mu|-\tfrac{1}{2}\nabla^2|\chi_\nu\rangle). \tag{25}$$

In practice we simplify Eqs. (24)–(25) further so that the torque can be evaluated in terms of AO overlap and kinetic energy integrals, as shown in Table I. We demonstrate this by again considering the simple case outlined above. For a p_x function it is easily verified that

$$[\mathbf{S}^{\tau a}_{ab}]_x = (y_a - y_b)\left\langle\frac{\partial p_x}{\partial z_a}\middle|s\right\rangle - (z_a - z_b)\left\langle\frac{\partial p_x}{\partial y_a}\middle|s\right\rangle$$
$$= 2\alpha((y_a - y_b)\langle d_{xz}|s\rangle - (z_a - z_b)\langle d_{xy}|s\rangle)$$
$$= 2\alpha(\langle f_{xyz}|s\rangle - \langle f_{xyz}|s\rangle)$$
$$= 0 \tag{26}$$

(where α is the exponent of p_x), i.e., that there is no torque on A about the x-axis. Similarly,

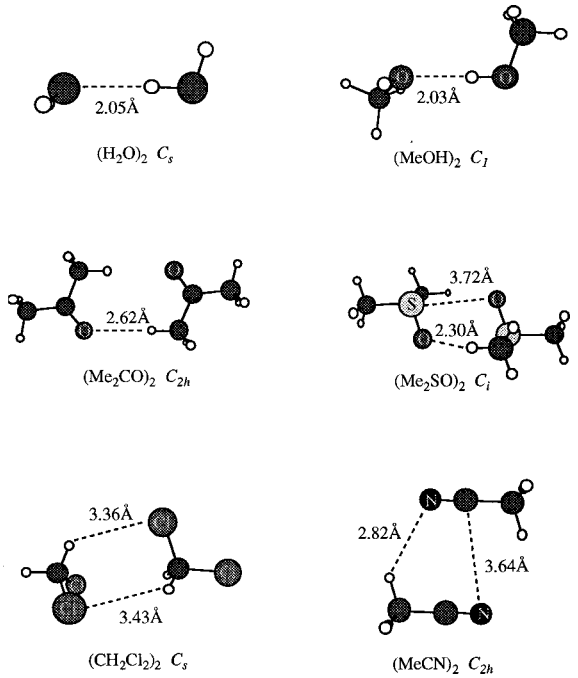


FIG. 2. Dimer geometries used in Tables II–V. See the text for an explanation of how they were obtained.

$$
\begin{aligned}
[\mathbf{S}_{ab}^{\tau a}]_y &= (z_a - z_b)\left\langle \frac{\partial p_x}{\partial x_a}\middle| s \right\rangle - (x_a - x_b)\left\langle \frac{\partial p_x}{\partial z_a}\middle| s \right\rangle \\
&= (z_a - z_b)(2\alpha\langle d_{xx}|s\rangle - \langle s|s\rangle) \\
&\quad - 2\alpha(x_a - x_b)\langle d_{xz}|s\rangle \\
&= 2\alpha(\langle f_{xxy}|s\rangle - \langle f_{xxy}|s\rangle) - \langle p_z|s\rangle \\
&= -\langle p_z|s\rangle, \qquad (27)
\end{aligned}
$$

i.e., the torque about the y-axis is proportional to the overlap between the p_z and s orbitals. Table I lists similar expressions for $\mathbf{S}_{ij}^{\tau a}$ and $\mathbf{T}_{ij}^{\tau a}$ involving s through d functions.

We note that the point torque has a classical analog in the torque on a dipole (and other multipoles) due to an electric field.[2]

B. Accuracy

1. Accuracy of the EFP/EFP exchange repulsion energy

In Paper I the severity of the various approximations outlined in Sec. III, and hence the accuracy of Eq. (15), was tested for the interaction of two water molecules. In this section we address the general applicability of Eq. (15) by applying it to dimers of methanol, dichloromethane, acetonitrile, acetone, and dimethyl sulfoxide (DMSO) in addition to the water dimer. The geometries used were obtained by first fully optimizing the dimer geometry minimum (verified by a Hessian calculation) at the RHF/6-31+G(d,p) level of theory, and then superimposing the free RHF/6-31+G(d,p) monomer geometries on the dimer structure. The latter step was done to provide a fair comparison with the EFP exchange repulsion energy, which utilizes the free monomer geometries. The geometries are displayed in Fig. 2, and the resulting exchange repulsion energies are given in Table II. The errors discussed next are obtained by subtracting the approximate values from the exact values.

TABLE II. Exact [Eq. (1)] and approximate [Eq. (15)] exchange repulsion energies, requisite CPU times for an energy and gradient evaluation at the RHF/BASIS//RHF/6-31+G(d,p) and EFP[BASIS//6-31+G(d,p)] level of theory, and % work skipped [Eq. (32)] calculated for several dimers (Fig. 2) using three basis sets. The approximate/EFP values are given in parentheses. See text for further information about the geometry and timings. Energies are in kcal/mol.

Basis=	6-31+G(d,p)	6-31++G(2d,2p)	pVTZ
		Water dimer	
E_{exch}	4.76 (4.41)	4.78 (4.90)	4.70 (4.88)
CPU seconds	52 (0.4)	251 (0.5)	839 (0.7)
Work skipped	54%	42%	23%
		Methanol dimer	
E_{exch}	5.07 (4.43)	5.19 (5.23)	5.22 (5.63)
CPU seconds	792 (0.8)	3476 (1.0)	12851 (1.6)
Work skipped	62%	52%	31%
		Dichloromethane dimer	
E_{exch}	0.80 (0.30)	0.84 (0.35)	0.84 (0.60)
CPU seconds	2226 (1.0)	14014[a] (1.1)	27222[a] (1.5)
Work skipped	72%	66%	46%
		Acetonitrile dimer	
E_{exch}	2.04 (1.60)	2.12 (1.94)	2.02 (1.96)
CPU seconds	1754 (1.0)	14608[a] (1.3)	26499[a] (2.1)
Work skipped	64%	55%	35%
		Acetone dimer	
E_{exch}	2.12 (1.53)	2.19 (1.72)	2.05 (1.67)
CPU seconds	6376 (1.6)	58519[a] (2.1)	602781[a] (3.9)
Work skipped	75%	65%	43%
		DMSO dimer	
E_{exch}	7.27 (6.59)	7.58 (7.42)	7.67 (8.65)
CPU seconds	20807[a] (2.1)	115155[a] (2.4)	671950[a] (4.5)
Work skipped	72%	63%	43%

[a]SCF calculation is run in direct mode due to size.

As noted in Paper I, the use of the 6-31+G(d,p) basis set results in a relatively small error (0.26 kcal/mol) in the water dimer exchange repulsion energy. However, for the larger systems in Table II the error increases to 0.44–0.68 kcal/mol. These errors are significantly decreased by increasing the basis set size to 6-31++G(2d,2p) with the exception of the CH_2Cl_2 and acetone dimers. The errors in Eq. (15) for these two dimers with the larger basis set are 0.49 and 0.47 kcal/mol, respectively, in contrast to the remaining dimers for which the absolute error range is 0.04–0.18 kcal/mol. A further basis set increase to Sadlej's pVTZ basis set decreases the CH_2Cl_2 error to 0.24 kcal/mol, but the error for the acetone dimer (0.38 kcal/mol) is less affected. Furthermore, the absolute errors for the methanol and DMSO dimers both increase (to 0.41 and 0.98 kcal/mol) on going to the pVTZ basis set! In order to obtain a detailed understanding of these errors it is necessary to consider the errors introduced by the individual approximations outlined in Sec. III. Table III lists the values for the exact [Eq. (1)] and approximate [Eq. (15)] exchange repulsion energies obtained using the 6-31++G(2d,2p) and pVTZ basis sets. In addition we list the exact second-order E_{exch} $\{E_{\text{exch}}[\mathscr{O}(S^2)]$, Eq. (6),$\}$ and intermolecular $\{E_{\text{exch}}[\mathscr{O}(V;S^2)]$, Eq. (11)$\}$ and internal energy $[\Delta E_{\text{A}} + \Delta E_{\text{B}}$; Eq. (7)$]$ contributions. Table IV lists the three contributions to $E_{\text{exch}}[\mathscr{O}(V;S^2)]$ as well as

TABLE III. The exact and various approximate exchange repulsion energies [approximate exchange, Eq. (15), second order, Eq. (6); intermolecular component, Eq. (11); internal energy contributions, $\Delta E_A+\Delta E_B$, Eq. (7)], calculated for several dimers [$(X)_2$, Fig. 2] using two basis sets. Energies are in kcal/mol.

X	Exact	Eq. (15)	Eq. (6)	Eq. (11)	$\Delta E_A+\Delta E_B$ [Eq. (7)]
		6-31++G(2d,2p)			
H_2O	4.78	4.90	4.73	4.76	$-0.03=0.05-0.08$
MeOH	5.19	5.23	5.12	5.05	$0.07=0.15-0.08$
CH_2Cl_2	0.84	0.35	0.84	0.60	$0.24=0.13+0.11$
MeCN	2.12	1.94	2.11	1.95	$0.16=0.08+0.08$
$(Me)_2CO$	2.19	1.72	2.18	2.08	$0.10=0.05+0.05$
$(Me)_2SO$	7.58	7.42	7.48	7.40	$0.08=0.04+0.04$
		pVTZ			
H_2O	4.70	4.88	4.64	4.72	$-0.08=0.00-0.08$
MeOH	5.22	5.63	5.15	5.28	$-0.13=-0.01-0.12$
CH_2Cl_2	0.84	0.60	0.85	0.74	$0.11=0.05+0.06$
MeCN	2.02	1.96	2.01	1.93	$0.08=0.04+0.04$
$(Me)_2CO$	2.05	1.67	2.04	2.00	$0.04=0.02+0.02$
$(Me)_2SO$	7.67	8.65	7.57	8.14	$-0.57=-0.29-0.29$

their approximate counterparts in Eq. (15). From Table III it is apparent [cf. exact vs Eq. (6)] that with a maximum error of 0.10 kcal/mol (for pVTZ DMSO), Eq. (5) is a very good approximation for all dimers. The bulk of the error is therefore introduced by the subsequent approximations.

a. CH_2Cl_2: The internal energy contributions contribute relatively little at the 6-31++G(2d,2p) level of theory, with the exception of CH_2Cl_2 where it contributes 0.24 kcal/mol. The approximation that leads to the neglect of the internal energy [Eq. (9)] is also used in Eq. (13), and so the overall basis set dependent error for CH_2Cl_2 is 0.48 kcal/mol, essentially the total error. The basis set increase to pVTZ reduces the basis set dependent error to 0.22 kcal/mol, and thus the total error to 0.24 kcal/mol.

b. Acetone: At the 6-31++G(2d,2p) level of theory the exact E_{exch} is underestimated by 0.47 kcal/mol. Only 0.20 kcal/mol of this error is basis set dependent, and the data in Table IV shows that an additional 0.25 kcal/mol comes from the spherical Gaussian overlap[16] (SGO) approximation to the intermolecular exchange energy. Thus, while a basis set increase to pVTZ significantly decreases the basis set dependent error (to 0.08 kcal/mol), the error in $E_{\text{exch}}(S^0)$ (0.31 kcal/mol), and hence in E_{exch} (0.38 kcal/mol), remains relatively large. Clearly, for this acetone dimer geometry one must go beyond the SGO approximation to obtain better accuracy.

c. Methanol: The increase in error (from -0.04 to -0.41) for methanol on going from 6-31++G(2d,2p) to pVTZ is mainly due to a loss of error cancellation. The basis set dependent error for 6-31++G(2d,2p) is 0.14 which cancels part of the -0.15 and -0.12 kcal/mol errors in $E_{\text{exch}}(S^0)$ and $E_{\text{exch}}(S^2)$, respectively. On going to pVTZ the basis set dependent error of -0.26 kcal/mol adds to the -0.10 and -0.13 kcal/mol errors in $E_{\text{exch}}(S^0)$ and $E_{\text{exch}}(S^2)$, respectively. The relatively smaller basis set dependent error for 6-31++G(2d,2p) is itself due to a cancellation of errors, since ΔE_{A} and ΔE_{B} are of opposite signs. This cancellation is lost on going to the pVTZ basis set, since only one of the internal energy contributions is re-

TABLE IV. Contributing terms to the intermolecular second order exchange [Eq. (11)] and approximate exchange repulsion [Eq. (15)] (in parentheses) calculated for several dimers [$(X)_2$, Fig. 2] using two different basis sets. Energies are in kcal/mol.

X	Eq. (12) [(12b)]	Eq. (13) [(13b)]	Eq. (14) [(14b)]	Eq. (11) [(15)]
		6-31++G(2d,2p)		
H_2O	-6.14 (-6.07)	13.37 (13.40)	-2.48 (-2.42)	4.76 (4.90)
MeOH	-6.62 (-6.47)	14.30 (14.22)	-2.63 (-2.51)	5.05 (5.23)
CH_2Cl_2	-1.33 (-1.36)	2.38 (2.14)	-0.45 (-0.42)	0.60 (0.35)
MeCN	-3.32 (-3.35)	6.43 (6.27)	-1.16 (-0.98)	1.95 (1.94)
$(Me)_2CO$	-3.39 (-3.64)	6.59 (6.49)	-1.13 (-1.11)	2.08 (1.72)
$(Me)_2SO$	-11.01 (-11.20)	22.14 (22.06)	-3.72 (-3.44)	7.40 (7.42)
		pVTZ		
H_2O	-6.03 (-6.10)	13.22 (13.30)	-2.46 (2.40)	4.72 (4.88)
MeOH	-6.75 (-6.65)	14.74 (14.87)	-2.72 (-2.59)	5.28 (5.63)
CH_2Cl_2	-1.45 (-1.51)	2.68 (2.58)	-0.50 (-0.47)	0.74 (0.60)
MeCN	-3.31 (-3.36)	6.39 (6.30)	-1.16 (-0.98)	1.93 (1.96)
$(Me)_2CO$	-3.31 (-3.62)	6.42 (6.38)	-1.11 (-1.10)	2.00 (1.67)
$(Me)_2SO$	-11.39 (-11.73)	23.45 (24.03)	-3.91 (-3.64)	8.14 (8.65)

duced. In principle, this can be remedied by using an even better basis set.

d. DMSO: Finally we address the large error in E_{exch} at the pVTZ level of theory for the DMSO dimer. This is clearly due to the 1.16 kcal/mol basis set dependent error introduced by this basis set for this system. This may be due to some inadequacy in the pVTZ basis set for sulfur, and in the following paragraphs we exclude this particular case from our discussion.

In summary we note that with the exception of the pVTZ DMSO case, all approximate exchange repulsion energies reproduce the exact value to within 0.5 kcal/mol when a 6-31++G(2*d*,2*p*) or better basis set is used. For the majority of these cases the error is below 0.25 kcal/mol; the origin of errors larger than this are discussed above. With the exception of the 6-31++G(2*d*,2*p*) water dimer case, these small errors are due to some cancellation of errors in the underlying individual approximations leading to Eq. (15). The basis set dependent error for these cases are in the range 0.06–0.32 kcal/mol, while the remaining errors are in the 0.02–0.28 kcal/mol range. Thus, basis sets like 6-31++G(2*d*,2*p*) or pVTZ result in basis set dependent errors similar in size to the nonbasis set dependent errors. In general it seems that significantly better accuracy can be obtained only through better approximations to *all* terms in Eq. (15).

e. The core MO contribution to the approximate E_{exch}: It is well known that the exchange repulsion energy falls off exponentially with distance,[17] and the exchange repulsion energy due to Pauli repulsion between valence electrons will therefore tend to dominate. In this subsection we briefly address the error introduced by summing only over valence MOs in Eq. (15) for the dimer systems discussed above. The approximate exchange repulsion energy with and without core MO contributions (note this also excludes core–valence interactions) are listed in Table V. This data indicates that the core MO contribution is roughly proportional to the total E_{exch}. Neglect of core MOs in the three most weakly interacting molecules, CH_2Cl_2, acetone, and acetonitrile lead to relatively small (0.02–0.11 kcal/mol) errors, in contrast to the remaining errors which range from 0.17 to 0.38 kcal/mol. In a few cases [the 6-31++G(2*d*,2*p*) and pVTZ water dimer case, and the pVTZ methanol case] these errors cancel other errors (discussed above) and bring the approximate E_{exch} closer to the exact value. The time savings associated with the neglect of core MO contributions will be discussed below.

TABLE V. Approximate exchange repulsion energies [Eq. (15)] calculated with and without core MO contributions, and the requisite CPU times for an energy and gradient evaluation at the EFP[BASIS//6-31+G(*d*,*p*)] and EFP(nc)[BASIS//6-31+G(*d*,*p*)] level of theory calculated for several dimers (Fig. 2) using three basis sets. The values that exclude the core MO contributions are given in parentheses. Energies are in kcal/mol.

Basis=	6-31+G(*d*,*p*)	6-31++G(2*d*,2*p*)	pVTZ
		Water dimer	
E_{exch}	4.41 (4.26)	4.90 (4.71)	4.88 (4.71)
CPU seconds	0.4 (0.3)	0.5 (0.4)	0.7 (0.5)
		Methanol dimer	
E_{exch}	4.43 (4.26)	5.24 (5.02)	5.63 (5.41)
CPU seconds	0.8 (0.6)	1.0 (0.7)	1.6 (1.3)
		Dichloromethane dimer	
E_{exch}	0.30 (0.27)	0.35 (0.32)	0.60 (0.58)
CPU seconds	1.0 (0.4)	1.1 (0.6)	1.5 (1.0)
		Acetonitrile dimer	
E_{exch}	1.60 (1.48)	1.94 (1.83)	1.96 (1.89)
CPU seconds	1.0 (0.7)	1.3 (0.9)	2.1 (1.6)
		Acetone dimer	
E_{exch}	1.53 (1.47)	1.72 (1.66)	1.67 (1.63)
CPU seconds	1.6 (1.1)	2.1 (1.7)	3.9 (3.2)
		DMSO dimer	
E_{exch}	6.59 (6.28)	7.42 (7.04)	8.65 (8.29)
CPU seconds	2.1 (1.4)	2.4 (1.8)	4.5 (3.6)

2. Accuracy of the EFP/EFP interaction energy

The EFP interaction energy is currently modeled by distributed multipole expansions of the classical Coulomb (E_{Coul}) and induction ($E_{\text{c-ind}}$) energies (truncated at octupoles and induced dipoles, respectively). In addition to the exchange repulsion energy,

$$E_{\text{int}}=E_{\text{Coul}}+E_{\text{exch}}+E_{\text{c-ind}}(+E_{\text{x-ind}}), \tag{28}$$

where $E_{\text{x-ind}}$ is the nonclassical induction energy that is not included in the current treatment. The error in this interaction energy (relative to the HF interaction energy) is therefore a sum of the errors in each term, in addition to the error introduced by the fact that the Pauli repulsion contribution to the induction energy (the exchange induction energy, $E_{\text{x-ind}}$) is not yet implemented. Good accuracy of the interaction energy is assured when each term is approximated accurately. However, as for the approximate E_{exch} itself, good accuracy may also occur due to the cancellation of errors in the individual terms. An example is presented in Table VI. Here, the water dimer interaction energy and the contributing terms [Eq. (28)] are calculated at two levels of theory and compared to the exact values.[18] At the higher level of theory, both E_{exch} and $E_{\text{c-ind}}$ are approximated quite accurately, but the overall accuracy is compromised by the neglect of $E_{\text{x-ind}}$ and the error in E_{Coul} due to charge penetration. With the smaller basis set, the only energy component that is approximated accurately is $E_{\text{c-ind}}$, while each remaining component has a large error. However, these errors largely cancel resulting in a good approximation of the total interaction energy. Clearly, such fortuitous cancellation can not be counted on in

TABLE VI. The interaction energy and its components [Eq. (28)] for the water dimer calculated at the RHF/6-31++G(2*d*,2*p*)//RHF/6-31+G(*d*,*p*), EFP(nc)[6-31++G(2*d*,2*p*)//6-31+G(*d*,*p*)], RHF6-31G(*d*,*p*), and EFP(nc)[6-31G(*d*,*p*)] level of theory. The *ab initio* values were obtained using the Kitaura-Morokuma energy decomposition scheme (Ref. 18). Energies are in kcal/mol.

	E_{int}=	E_{Coul}	$+E_{\text{exch}}$	$+E_{\text{c-ind}}$	$+E_{\text{x-nd}}$
		6-31++G(2*d*,2*p*)//6-31+G(*d*,*p*)			
Ab initio	−4.10	−7.35	4.78	−0.86	−0.67
EFP	−2.34	−6.18	4.71	−0.87	0
		6-31G(*d*,*p*)//6-31G(*d*,*p*)			
Ab initio	−5.52	−7.25	4.00	−0.46	−1.81
EFP	−5.45	−6.27	1.43	−0.60	0

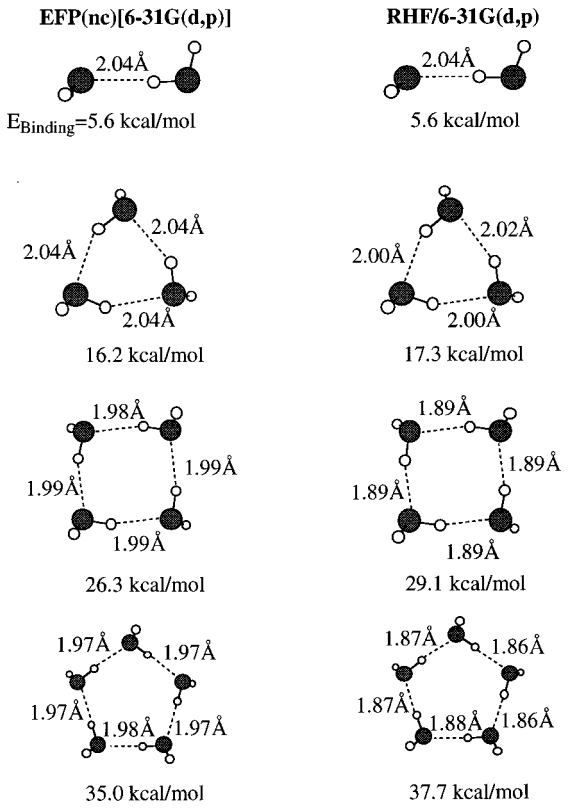


FIG. 3. Structures and binding energies of $(H_2O)_n$, $n=2-5$ at the EFP(nc)[6-31G(d,p)] and RHF/6-31G(d,p) level of theory.

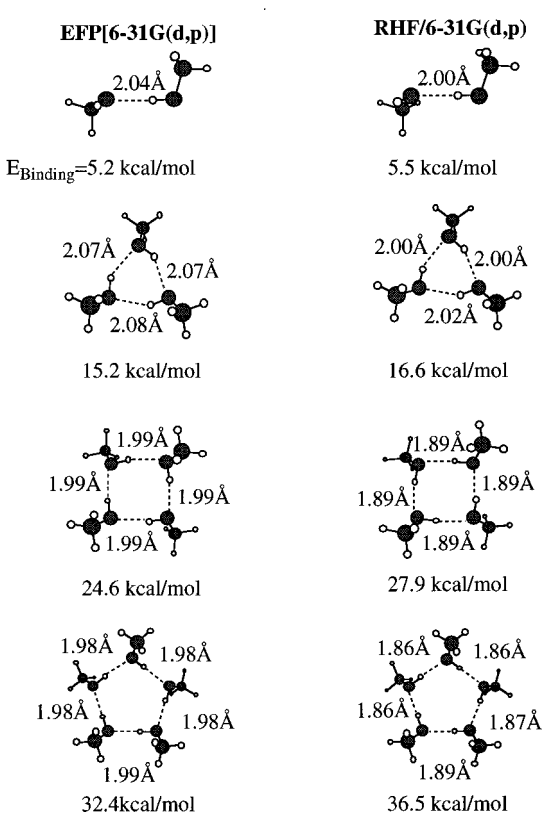


FIG. 4. Structures and binding energies of $(MeOH)_n$, $n=2-5$ at the EFP(nc)[6-31G(d,p)] and RHF/6-31G(d,p) level of theory.

general, and future studies will address the charge penetration error and the method for approximating the exchange-induction energy. However, as an initial exploration of the use of Eq. (15) within the EFP methodology we apply EFP(nc)[6-31G(d,p)] to small water and methanol clusters.

Figures 3 and 4 display the binding energies and select intermolecular bond lengths for water and methanol clusters up to pentamers, calculated at the EFP(nc)[6-31G(d,p)] and RHF/6-31G(d,p) level of theory. All structures were verified as minima by calculating and diagonalizing the Hessian (analytically and numerically for the *ab initio* and EFP structures, respectively). All structures therefore represent fully optimized minima on their respective potential energy surfaces, in contrast to the previously discussed dimers where the same geometry is used for both *ab initio* and EFP energy calculations. The errors in the binding energies are within 12% of the all *ab initio* results, with the largest relative error occurring for the methanol tetramer. All errors result from an underestimation of the binding energy. The EFP hydrogen bond lengths are generally 0.10 Å too long for the trimer and larger clusters. The largest error is 0.12 Å and occurs for the methanol pentamer. Apart from the longer hydrogen bond lengths there are no very noticeable structural differences between the EFP and RHF structures. At a minimum the EFP(nc)[6-31G(d,p)] level of theory gives a good qualitative picture of the energetic and structural changes that occur on going from water and methanol dimers to pentamers.

C. Computational cost

1. Integral screening

The most time-consuming part of evaluating Eq. (15) is the evaluation of the atomic orbital contributions to S_{ij} and T_{ij},

$$\langle\chi_\mu|\chi_\nu\rangle \text{ and } \langle\chi_\mu|-\tfrac{1}{2}\nabla^2|\chi_\nu\rangle, \tag{29}$$

where the χ's are primitive Cartesian Gaussians,

$$\chi_\mu = x^{n_\mu} y^{l_\mu} z^{m_\mu} \exp(-\alpha|\mathbf{r}_1-\mathbf{R}_\mu|^2) \text{ and}$$
$$\chi_\nu = x^{n_\nu} y^{l_\nu} z^{m_\nu} \exp(-\beta|\mathbf{r}_1-\mathbf{R}_\nu|^2). \tag{30}$$

Since both integrals in Eq. (29) are relatively short range, it is useful to define some sort of test that allows one to quickly estimate the size of the integrals and skip the computation of those that are deemed smaller than a certain threshold. The test employed here, for both integrals, is well known and involves the spherical overlap of a pair of Gaussians with arbitrary angular momentum,[19]

$$\exp\left(-\frac{\alpha\beta}{\alpha+\beta}|\mathbf{R}_\mu-\mathbf{R}_\nu|^2\right) < 10^{-\text{ITOL}}. \tag{31}$$

The very same test is applied to the gradient integrals in Eqs. (20)–(21). An ITOL value of 10 was chosen as a conservative compromise between efficiency and accuracy based on the data presented in Table VII. Table VII presents the energy and maximum gradient component for the water dimer, calculated using two different basis sets, as a function of ITOL. It is evident that the changes in energy and gradient

TABLE VII. The exchange repulsion energy [Eq. (15)] and the maximum component of the total gradient at the EFP[6-31G(d,p)] water dimer geometry, calculated using two different basis sets, as a function of ITOL [defined in Eq. (31)]. See Eq. (32) for a definition of ''worked skipped.'' Energy and gradient are given in atomic units.

ITOL	E_{exch}	Maximum gradient	Work skipped
		6-31G(d,p)	
20	0.002,270,691,2	0.000,063	33%
10	0.002,270,691,9	0.000,063	57%
5	0.002,271,550,0	0.000,061	69%
		pVTZ	
20	0.007,506,979,2	0.005,351	17%
10	0.007,506,979,3	0.005,351	23%
5	0.007,576,237,6	0.005,634	36%

are negligible in going from ITOL=20 to 10. However, a further decrease to ITOL=5 changes the fourth decimal place of the gradient for the pVTZ basis set. Since our geometry convergence criterion generally is $\leqslant 10^{-4}$ for the maximum gradient component, ITOL=5 leads to unacceptable errors.

The column labeled ''work skipped'' in Table VII is calculated by

$$\text{work skipped} = \frac{\text{skipped pairs of primitive Gaussian shells}}{\text{total pairs of primitive Gaussian shells}} \times 100\%, \tag{32}$$

where the criterion for skipping the shell-pair is Eq. (31). Since Eq. (31) is independent of angular momentum, one test is sufficient to determine whether, e.g., all nine integrals between two p shells can be neglected. It is gratifying to note that even for such a small and compact molecular system as the water dimer, it is possible to skip between 23% and 57% of the work, depending on the diffuseness of the basis set.

Equation (31), and hence the percentage of work skipped, is a complex function of the diffuseness of the atomic basis set and the geometry of the supermolecule in question. For a given basis set the amount of work skipped increases with the size of the interacting molecules. For a given geometry the amount of worked skipped decreases with the diffuseness of the basis set. These trends are evident from the data presented in Table II. In the case of the 6$-31+$G(d,p) basis set the amount of work skipped ranges from 54% for the water dimer up to 75% for the acetone dimer, reflecting the change from a system of small strongly interacting molecules to one of relatively large and weakly interacting molecules. Increasing the basis set size to 6$-31++$G($2d,2p$) decreases the amount of work skipped by 6%–12%. The decrease is largest for the water dimer due to its small size, and smallest for the CH_2Cl_2 dimer. The decrease in work drops an additional 19%–22% on going to Sadlej's pVTZ basis, indicating a relatively large increase in the diffuseness of the basis set relatively to 6$-31++$G($2d,2p$).

For relatively small molecules, such as those presented in Fig. 2, the main importance of screening is not to decrease the cost of computing the interaction energies of neighboring molecules, but to facilitate the neglect of interactions between distant molecules in a cluster. We shall return to this point in the following section.

2. Timings

Reported in this study is the average RS/6000 350 CPU time needed to compute the *total* EFP/EFP interaction energy and gradient (including the torque). We use the total energy and gradient, rather than just the exchange repulsion energy, since the total energy is the quantity of interest for molecular simulations. We note, however, that the exchange repulsion contribution is by far the most computationally demanding part of the total energy plus gradient evaluation. The EFP timings were computed by performing a double differenced numerical Hessian on the minimum energy structure in question, and dividing the total CPU time by the number of requisite energy plus gradient evaluations. The *ab initio* CPU times were obtained by performing a single RHF energy plus gradient calculation.

The EFP CPU times required for an energy plus gradient evaluation for the various dimers discussed above are presented in Table II. The CPU times for the 6$-31+$G(d,p) basis set range from 0.4 s for the water dimer to 2.1 s for the DMSO dimer. Going to the 6$-31++$G($2d,2p$) basis set increases the times by 9% (CH_2Cl_2) to 24% (acetone). A more marked increase, 27% (CH_2Cl_2) to 47% (DMSO), is observed for the transition from 6$-31++$G($2d,2p$) to pVTZ. The main cause of the time increase is the increased number of primitive Gaussians and the diffuseness of the pVTZ. The latter is evidenced by the marked drop in the relative amount of work skipped, as noted above. The CPU times for the all *ab initio* energy plus gradient evaluations are included to put the EFP times into perspective. The most time consuming EFP calculation for each molecule is two orders of magnitude faster than the least time consuming *ab initio* calculation, and in the case of the pVTZ DMSO dimer the EFP method is more than 10^5 times faster.

In Sec. IV B it was shown that the omission of core MO contributions can lead to negligible or improved changes to the approximate E_{exch}. The associated time savings are presented in Table V. The time savings range from 18% (pVTZ acetone) to 60% [6$-31+$G(d,p) CH_2Cl_2] with an average of 29%. This is a surprisingly large effect given that the neglect of core MOs does not involve any additional neglect of AO integrals (i.e., ''% work skipped'' is unchanged).

As mentioned previously, a main concern is how the computational cost scales with cluster size. Formally, the cost of computing the interactions between n monomers is the cost of computing a pair interaction (t_2) multiplied by the number of pairwise interactions,[20]

$$t_n = t_2(n^2-n)/2. \tag{33}$$

However, since the most computationally intensive energy component is subjected to the screening process outlined above, t_n should scale better than Eq. (33). This is demonstrated for the EFP(nc)[6-31G(d,p)] water clusters ranging from $n=2$ to 11. The $n=2-5$ clusters are depicted in Fig. 3. Figure 5(a) depicts the percentage of work skipped as well as

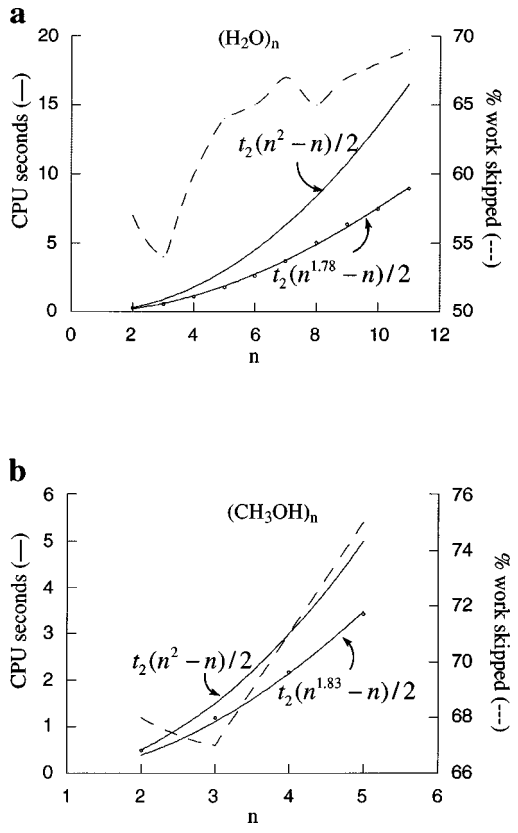


FIG. 5. Actual and predicted (see text) timings for an EFP(nc)[6-31G(d,p)] energy+gradient evaluation, as well as the % worked skipped [Eq. (32)] for water and methanol clusters.

the theoretical [based on Eq. (33)] and actual CPU time as a function of n. The percentage of work skipped is somewhat erratic, since some clusters are more compact than others. However, with the exception of $n=3$ and 8, a larger portion of the work is skipped (54%–69%) as the cluster size increases. As a result, the requisite CPU time scales better than predicted by Eq. (33) as can be seen from Fig. 5. The actual CPU time can be fitted well to an equation similar to Eq. (33) with the exponent reduced to 1.78. As a result, the CPU time increases from 0.3 s to 9.0 s, rather than the 16.5 s predicted by Eq. (33), on going from $n=2$ to 11. This scaling is expected to improve as n is increased further.

A further example of improved scaling due to Eq. (31) is presented in Fig. 5(b), which presents data similar to that of Fig. 5(a) but for the EFP(nc)[$6-31$G(d,p)] methanol clusters shown in Fig. 4. Just as for the water case, the percentage of work skipped drops for the trimer, but is then followed by a steady increase from 67% to 75%. The CPU times scale roughly with an exponent of 1.83 rather than 2, as the cluster size is increase from $n=2$ to 5. Thus the CPU time for an energy plus gradient is 3.4 rather than 4.9 seconds for $n=5$.

V. SUMMARY AND PROGNOSIS

This paper addresses three main issues related to the application of an approximate formula for the intermolecular Pauli repulsion between closed shell molecules [Eq. (15)], derived earlier,[5] to the effective fragment potential (EFP) method.[2] These issues are (1) the derivation and implementation of analytic energy derivatives, (2) the accuracy of the exchange repulsion energy and the total EFP/EFP interaction energy, and (3) the computational cost involved in evaluating the total energy and gradient.

The energy derivatives with respect to Cartesian coordinates are straightforward and presented in Eqs. (18)–(19). In order to calculate the torque on an EFP (with frozen internal geometry) it is necessary to include the effect of additional ''point torques'' on the atomic centers that use p or higher order basis functions to expand the molecular orbitals. This point torque is presented in Eq. (23).

The accuracy of the approximate exchange repulsion formula is tested against exact values for dimers of water, methanol, dichloromethane, acetonitrile, acetone, and dimethyl sulfoxide (Fig. 2). One of the underlying approximations increases in accuracy with increased basis set, and so each dimer is studied with three basis sets, $6-31+$G(d,p), $6-31++$G($2d,2p$), and Sadlej's pVTZ. The agreement with the exact exchange repulsion energy (Table II) is within 0.50 kcal/mol for the two larger basis sets with the exception of the pVTZ DMSO dimer, for which the basis set for sulfur seems suspect. This good agreement is predominantly due to a cancellation of errors (Tables III–IV) in four individual approximations ranging in size from 0.00 to 0.32 kcal/mol, and the best result is therefore not always obtained with the larger basis set. The exact exchange repulsion can be reproduced to within 0.24 kcal/mol for all but the acetone dimer if by considering only the better of the two large basis set results. The best result obtained for acetone dimer is 0.38 kcal/mol. The accuracy of the total interaction energy [Eq. (28)] can also be improved by a cancellation of errors (Table IV). We take advantage of this fact and use the $6-31+$G(d,p) basis set to study the structures and binding energies in water (Fig. 3) and methanol (Fig. 4) clusters using the EFP method. The binding energies are within 12% of the *ab initio* results, and while the bond lengths can be up to 0.12 Å too long, the overall cluster structures are very well reproduced by the EFP method.

The approximate exchange repulsion energy formula requires the evaluation of one-electron overlap and kinetic energy integrals and is therefore be the most computationally demanding part of the EFP interaction energy. However, the requisite RS/6000 Model 350 CPU time for a single energy and gradient evaluation is modest and ranges from 0.4 s [$6-31+$G(d,p) water dimer] to 4.5 s (pVTZ DMSO dimer). In contrast, the corresponding *ab initio* calculations require 52 s and 671 950 s, respectively. Furthermore, aggressive screening [Eq. (31)] (made possible by the short range nature of the overlap and kinetic energy integrals) leads to a less than N^2 scaling as the size of cluster increases. This is demonstrated for both water and methanol clusters (Fig. 5).

The application of the approximate exchange repulsion formula [Eq. (15)] to the EFP method yields *a completely general EFP/EFP interaction potential that depends solely on the properties of the isolated molecules*, made available by separate *ab initio* calculations. This already provides the opportunity to perform simulations of clusters of molecules, including mixed clusters. In order to obtain an equivalent *ab*

initio/EFP interaction potential it is necessary to derive and implement an exchange repulsion operator corresponding to the exchange repulsion energy. This will be presented in a future study. Additional studies will address the problems of charge penetration and exchange-induction.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Mike Schmidt and Dr. Grant Merrill and Mr. Simon P. Webb for making code related to the EFP methodology available as well as for many valuable discussions. This work was supported by the National Science Foundation (CHE-9313717) and the National Institute for Standards and Technology (70NANB4H1591). The computations were carried out on IBM workstations obtained from grants provided by the National Science Foundation and the Air Force Office of Scientific Research.

[1] (QM and MM stands for quantum and molecular mechanics, respectively.) (a) B. T. Thole and P. T. van Duijnen, Theor. Chim. Acta **55**, 305 (1980); (b) B. T. Thole and P. T. van Duijnen, Chem. Phys. **71**, 211 (1982); (c) K. Ohta, Y. Yoshioka, K. Morokuma, and K. Kitaura, Chem. Phys. Lett. **101**, 12 (1983); (d) U. C. Singh and P. A. Kollman, J. Comput. Chem. **5**, 129 (1984); (e) M. J. Field, P. A. Bash, and M. J. Karplus, *ibid.* **11**, 700 (1990); (f) J. Gao, J. Am. Chem. Soc. **116**, 1563 (1994); (g) M. A. Thompson, E. D. Glendening, and D. F. Feller, J. Phys. Chem. **98**, 10465 (1994); (h) F. Maseras and K. Morokuma, J. Comput. Chem. **16**, 1170 (1995); (i) A. H. de Vries, P. T. van Duijnen, A. H. Huffer, J. A. C. Rullmann, J. P. Dijkman, H. Merenga, and B. T. Thole, J. Comput. Chem. **16**, 37 (1995).

[2] (a) J. H. Jensen, P. N. Day, M. S. Gordon, H. Basch, D. Cohen, D. R. Garmer, M. Kraus, and W. J. Stevens, in *Modeling the Hydrogen Bond*, ACS Symposium Series 569, edited by D. A. Smith (American Chemical Society, Washington, D.C., 1994), Chap. 9; (b) P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Kraus, D. Garmer, H. Basch, and D. Cohen, J. Chem. Phys. **105**, 1968 (1996).

[3] (a) W. Chen and M. S. Gordon, J. Chem. Phys. **105**, 11081 (1996); (b) M. Kraus and S. P. Webb, *ibid.* **107**, 5771 (1997); (c) P. N. Day and R. Pachter, *ibid.* **107**, 2990 (1997); (d) G. Merrill and M. S. Gordon, J. Phys. Chem. (in press).

[4] B. D. Wladkowski, M. Krauss, and W. J. Stevens, J. Am. Chem. Soc. **117**, 10537 (1995).

[5] J. H. Jensen and M. S. Gordon, Mol. Phys. **89**, 1313 (1996).

[6] W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).

[7] The following diffuse sp-shell exponents were used: C=0.0438, N=0.0639, O=0.0845, S=0.0405, Cl=0.0483.

[8] (a) The diffuse hydrogen exponent used was 0.0360. (b) The two d exponents on the nonhydrogen atoms were obtained by multiplying the standard d exponent from the 6-31G(d,p) basis set by 2.0 and 0.5. Likewise for the two p exponent on the hydrogen atoms.

[9] (a) A. J. Sadlej, Coll. Czech. Chem. Commun. **53**, 1995 (1988); (b) A. J. Sadlej and M. Urban, J. Mol. Struct.: THEOCHEM **234**, 147 (1991); (c) A. J. Sadlej, Theor. Chim. Acta **79**, 123 (1991); (d) **81**, 45 (1992); (e) **81**, 339 (1992); (f) basis sets were obtained from the Extensible Computational Chemistry Environment Basis Set Database, Version 1.0, as developed and distributed by the Molecular Science Computing Facility, Environmental and Molecular Sciences Laboratory which is part of the Pacific Northwest Laboratory, P.O. Box 999, Richland, Washington 99352, USA, and funded by the U.S. Department of Energy. The Pacific Northwest Laboratory is a multiprogram laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC06-76RLO 1830. Contact David Feller, Karen Schuchardt, or Don Jones for further information.

[10] (a) Stone, Chem. Phys. Lett. **83**, 233 (1981); (b) Price and Stone, *ibid.* **98**, 419 (1983).

[11] S. P. Webb, Ph.D. thesis, Iowa State University, 1997.

[12] (a) C. Edmiston and K. Ruedenberg, Rev. Mod. Phys. **35**, 457 (1963); (b) R. C. Raffenetti, K. Ruedenberg, C. L. Janssen, and H. F. Schaefer III, Theor. Chim. Acta **86**, 149 (1993).

[13] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[14] (a) R. Landshoff, Z. Phys. **102**, 201 (1936); (b) see also A. Fröman and P.-O. Löwdin, J. Phys. Chem. Solids **23**, 75 (1962).

[15] Y. Yamaguchi, Y. Masamura, J. D. Goddard, and H. F. Schaefer, *A New Dimension, A New Dimension to Quantum Chemistry* (Oxford University Press, Oxford, 1994).

[16] J. H. Jensen, J. Chem. Phys. **104**, 7795 (1996).

[17] H.-J. Böhm and R. Ahlrichs, J. Chem. Phys. **77**, 2028 (1982).

[18] K. Kitaura and K. Morokuma, Int. J. Quantum Chem. **10**, 325 (1976).

[19] The actual test performed is $(\alpha\beta/\alpha+\beta)|\mathbf{R}_\mu-\mathbf{R}_\nu|^2>2.3025\cdot\mathrm{ITOL}$.

[20] The polarization energy, and therefore the total interaction energy, is not strictly pairwise additive. However, since the (pairwise additive) exchange repulsion energy represents the vast bulk of the computational effort the requisite CPU time can be taken to come from a pairwise additive potential for all practical purposes.