International Reviews in Physical Chemistry

Calculating intermolecular potentials with SIMPER: the water-nitrogen and water-oxygen interactions, dispersion energy coefficients, and preliminary results for larger molecules

Richard J. Wheatley [a]; Timothy C. Lillestolen [a]
[a] School of Chemistry, The University of Nottingham, Nottingham, United Kingdom, NG7 2RD

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis
Taylor & Francis Group

Calculating intermolecular potentials with SIMPER: the water–nitrogen and water–oxygen interactions, dispersion energy coefficients, and preliminary results for larger molecules

RICHARD J. WHEATLEY* and TIMOTHY C. LILLESTOLEN

School of Chemistry, The University of Nottingham, Nottingham,
United Kingdom, NG7 2RD

Recent theoretical studies of the van der Waals complexes formed between water and the two main components of air, nitrogen and oxygen, are reviewed. Combined with previous work on the water–argon complex, the results allow non-ideal thermodynamic properties of water–air mixtures to be calculated. The intermolecular potential energy surfaces for these complexes have been calculated using a combination of supermolecule methods and perturbation theory, as described in a previous review. Theoretical techniques introduced since this previous work include the extension of the intermolecular perturbation theory to open-shell electronic species, as required for the water–oxygen complex, and new and more accurate calculations of dispersion energy coefficients. Spin-unrestricted, time-dependent coupled cluster theory has been used for calculating dispersion energy coefficients for the water–oxygen complex, and this method is described and compared with other accurate methods, and a possible alternative method is suggested for future work. The results obtained for the water–nitrogen complex are highly satisfactory. With little more computational effort than is required to produce a second-order Moller–Plesset supermolecule potential energy surface, the intermolecular potential is calculated to an accuracy which appears to be comparable to coupled-cluster calculations with perturbative triple excitations. For the water–oxygen complex, the different theoretical methods produce potential energy surfaces with larger discrepancies than for water–nitrogen, although they predict the main features of the potential energy surface better than calculations in the literature. Although few experimental measurements of water–oxygen virial coefficients are available, the agreement of theoretical predictions with these is reasonable, and the agreement with the better characterized water–air virial coefficients is very good. The review concludes with a forward look to work on larger molecules. Increasing the size of the interacting molecules creates a number of practical problems. Some problems, including the steep scaling of computation time with system size, are common to all methods. The use in the current work of a damped multipole expansion about the molecular centres also causes problems when larger molecules are considered. The review therefore considers methods that can be used to reduce the unfavourable size scaling, to reduce the size of the basis set, and to use damped atomic multipole expansions, which are centred on the nuclei of the interacting molecules.

*Corresponding author. Email: Richard.Wheatley@nottingham.ac.uk

Contents PAGE

1. Overview

Weak intermolecular potentials, such as van der Waals interactions and hydrogen bonds, determine many important physical properties of molecular substances. However, the intermolecular potential surface is difficult to calculate accurately, for several reasons: it is a small difference of much larger electronic energies, it is affected by the outer regions of the molecular electron densities, it depends strongly on electron correlation, and it is a multidimensional function which cannot easily be interpolated or fitted. In a previous review [1], the advantages and disadvantages of using super-molecule methods and perturbation theory to calculate intermolecular potentials were discussed, and a method was presented for combining the favourable qualities of both. This SIMPER method (Systematic InterMolecular Potential Extrapolation Routine) was shown to give results of comparable accuracy to the much more expensive coupled-cluster CCSD with perturbative triples (CCSD(T)) method, for a range of weakly bound molecular complexes.

The SIMPER approach to calculating interaction energies between pairs of molecules involves a (counterpoise-corrected) supermolecule calculation of the interaction energy,

E_{sup}, at each required point on the potential energy surface. Perturbation theory is then used, firstly to calculate the contributions $E_{\text{sup}}^{(k)}$ to the supermolecule energy at low orders k of the intermolecular perturbation (usually up to second or third order), and secondly to calculate analogous contributions, $E_{\text{high}}^{(k)}$, up to the same order of the intermolecular perturbation, but treating the electron correlation more accurately than was done in the supermolecule calculation. These 'higher-level' contributions $E_{\text{high}}^{(k)}$ are used instead of the 'lower-level' supermolecule contributions $E_{\text{sup}}^{(k)}$, to give the SIMPER interaction energy,

$$E_{\text{SIMPER}} = E_{\text{sup}} - \sum_k E_{\text{sup}}^{(k)} + \sum_k E_{\text{high}}^{(k)}. \tag{1}$$

In a typical implementation, the supermolecule method is second-order Moller–Plesset theory (MP2), Rayleigh–Schrödinger perturbation theory (RSPT) is used to obtain the perturbative terms, and the 'higher-level' contributions are calculated using coupled-cluster theory with single and double substitutions (CCSD). More details are given in the following sections of this review. The resulting potential energy surface is usually found to improve on the original supermolecule calculations, because the perturbative components of the supermolecule energy are replaced by components with higher-level electron correlation. Since supermolecule calculations are used, the resulting SIMPER interaction energy also contains the important short-range exchange-repulsion energy which is absent in RSPT, as well as higher-order terms in the intermolecular perturbation, which would be absent if only (truncated) perturbation theory were used. These higher-order terms in the energy are not recalculated at a higher level of electron correlation. This is reasonable, because they are usually significantly smaller than the lower-order terms, and the accurate calculation of higher-order energies using perturbation theory often requires an impractically large computational effort.

In practice, the basic scheme of equation (1) is modified in several ways. The first-order energy is treated exactly as shown, but the second-order energy, which is a sum of induction and dispersion energy contributions, is more expensive to calculate at a highly correlated level, especially the dispersion energy. The dispersion energy is therefore calculated as a damped multipole series at each point on the potential energy surface:

$$E_{\text{disp}} = - \sum_{n=6} C_n f_n(bR) R^{-n}, \tag{2}$$

where R is the separation between the molecules, C_n is a dispersion energy coefficient, and f_n is a 'universal' damping function, which depends on separation and on a length scaling parameter b, and is represented as an incomplete gamma function [2],

$$f_n(bR) = 1 - e^{-bR} \sum_{k=0}^{n} \frac{(bR)^k}{k!}. \tag{3}$$

Instead of recalculating the total dispersion energy, the SIMPER method involves recalculating only the dispersion energy coefficients C_n at a higher level of electron correlation. The damping functions are assumed to keep the same form, although the length scaling parameter is modified [3–5] by assuming that the dimensionless quantity $b\sqrt{C_8/C_6}$ is the same for the 'low-level' and 'high-level' calculations, at each point on the surface.

In order to use equation (2) to calculate the dispersion energy, it is necessary to calculate dispersion energy coefficients for the interacting pair of molecules. The methods that have been implemented in SIMPER are described in section 2, and further modifications of the SIMPER method and its application to the water–nitrogen and water–oxygen interactions are then described in section 3 and section 4. The water–oxygen potential energy surface is the first complete intermolecular potential for this interaction in the literature, and it is the first SIMPER potential energy surface for an interaction involving an open-shell molecule.

The second part of this review considers the possibility of using SIMPER for molecules larger than water and first-row diatomics. The first problem is that equation (2) cannot be used for 'large' molecules, because an expansion using the distance R between the centres of the molecules does not converge, or converges too slowly. It is likely that molecules with more than two heavy atoms (such as CO_2) cannot be treated in this way. Instead, the intermolecular potential, and the dispersion energy in particular, needs to be expanded as a sum of pair interactions between atoms. The development of an atomistic version of equation (2) requires atomic dispersion energy coefficients, and these can be calculated from atomic polarizabilities. Some recent developments in calculating atomic polarizabilities are reviewed in section 5. For even larger molecules, the supermolecule and perturbative calculations required by the SIMPER method are not feasible, and additional approximations are required. These are discussed further in section 6.

Atomic units are used in this review. The atomic unit of length is the Bohr, $a_0 = 5.291772 \times 10^{-11}$ m, and the atomic unit of energy is the Hartree, $E_h = 4.35975 \times 10^{-18}$ J.

2. Calculation of dispersion energy coefficients

The dispersion energy E_{disp} plays a leading role in van der Waals interactions between non-polar molecules, and dispersion energy coefficients are an important part of the total dispersion energy. At sufficiently large intermolecular separations R, the dispersion energy can be approximated well by the first few terms of a multipole series,

$$E_{\text{disp, mult}} = -\sum_{n \geq 6} C_n R^{-n}, \tag{4}$$

where C_n is a dispersion energy coefficient. Since the multipole series does not take charge density overlap into account, it converges to the wrong value, or diverges, at physically important separations including the equilibrium geometry. A damped

multipole series (equation (2)) is useful at intermediate and short intermolecular separations [6–9]: the damping functions $f_n(bR)$ are 1 for large R, but tend rapidly to zero as R decreases.

The dispersion energy coefficients C_n depend on the orientations of the molecules, but orientation-independent dispersion energy coefficients $C_{\mathrm{disp};l_1\kappa_1,l_2\kappa_2;l_3\kappa_3,l_4\kappa_4}$ can also be defined [10], where $n = l_1 + l_2 + l_3 + l_4 + 2$. They can be used to calculate the multipolar dispersion energy at any intermolecular geometry (R, Ω)

$$
\begin{aligned}
E_{\mathrm{disp,mult}}(R, \Omega) = -\sum_{l_1, \kappa_1} \sum_{l_2, \kappa_2} \sum_{l_3, \kappa_3} \sum_{l_4, \kappa_4} & T_{l_1, \kappa_1, l_3, \kappa_3}(R, \Omega) \\
\times\, & T_{l_2, \kappa_2, l_4, \kappa_4}(R, \Omega) C_{\mathrm{disp};l_1\kappa_1,l_2\kappa_2;l_3\kappa_3,l_4\kappa_4}.
\end{aligned}
\tag{5}
$$

The notation (l, κ) indicates that real multipoles are used; they are related to the real and imaginary parts of the complex multipoles (l, m) [11]. For example, the three real multipole operators for $l = 1$ are the expectation values of x, y and z. The function $T_{l, \kappa, l', \kappa'}(R, \Omega)$ is the interaction between a multipole $Q_{l, \kappa}$ of molecule A and a multipole $Q_{l', \kappa'}$ of molecule B [11], and consists of the product of an orientation-dependent function and an inverse power of separation $R^{-l-l'-1}$.

Accurate calculation of dispersion energy coefficients is difficult for all but the smallest atoms and molecules. For large n, high-angular-momentum polarization functions are needed in the basis set, and the calculations are sensitive to the electronic wave function remote from the nuclei, which means that large basis sets with a balance between diffuse and contracted functions are required. The contribution of electron correlation to the dispersion energy coefficients is also important.

Perturbation theory can be used to derive an expression for the dispersion energy coefficients $C_{\mathrm{disp};l_1\kappa_1,l_2\kappa_2;l_3\kappa_3,l_4\kappa_4}$, and hence C_n, for two molecules A and B in their electronic ground states $\psi_A^{(0)}$, $\psi_B^{(0)}$:

$$
\begin{aligned}
C_{\mathrm{disp};l_1\kappa_1,l_2\kappa_2;l_3\kappa_3,l_4\kappa_4} = \sum_{a>0, b>0} & \left\langle \psi_A^{(0)} \middle| \hat{Q}_{A, l_1\kappa_1} \middle| \psi_A^a \right\rangle \left\langle \psi_A^a \middle| \hat{Q}_{A, l_2\kappa_2} \middle| \psi_A^{(0)} \right\rangle \\
\times\, & \left\langle \psi_B^{(0)} \middle| \hat{Q}_{B, l_3\kappa_3} \middle| \psi_B^b \right\rangle \left\langle \psi_B^b \middle| \hat{Q}_{B, l_4\kappa_4} \middle| \psi_B^{(0)} \right\rangle \middle/ \left(E_A^a - E_A^{(0)} + E_B^b - E_B^{(0)} \right)
\end{aligned}
\tag{6}
$$

where the energies $E_A^{(0)}$, $E_B^{(0)}$, E_A^a and E_B^b correspond to the electronic wave functions $\psi_A^{(0)}$, $\psi_B^{(0)}$, ψ_A^a and ψ_B^b respectively, and $\hat{Q}_{A, l\kappa}$ is a real multipole operator centred on molecule A. The sum over electronically excited states ψ_A^a and ψ_B^b of the molecules is formally infinite, and has a significant contribution from continuum states, so this equation cannot be used directly.

There are two commonly used methods for transforming equation (6) into a form that can be used in practical calculations. One method uses frequency-dependent polarizabilities, and the other method uses pseudostates.

Frequency-dependent polarizabilities are defined using perturbation theory as

$$
\alpha_{l\kappa, l'\kappa'}^A(\omega) = 2 \sum_{a>0} \frac{\left(E_A^a - E_A^{(0)} \right) \left\langle \psi_A^{(0)} \middle| \hat{Q}_{A, l\kappa} \middle| \psi_A^a \right\rangle \left\langle \psi_A^a \middle| \hat{Q}_{A, l'\kappa'} \middle| \psi_A^{(0)} \right\rangle}{\left(E_A^a - E_A^{(0)} \right)^2 - \omega^2}.
\tag{7}
$$

The quantity ω can conveniently be taken to have dimensions of energy rather than frequency in this context, but the general usage of the term 'frequency' is followed here. Equation (7) enables the expression for the dispersion energy coefficients to be transformed to an integral over imaginary-frequency-dependent polarizabilities [12]:

$$C_{\mathrm{disp};l_1\kappa_1,l_2\kappa_2;l_3\kappa_3,l_4\kappa_4} = \frac{1}{2\pi} \int \alpha^A_{l_1\kappa_1,l_2\kappa_2}(i\omega)\alpha^B_{l_3\kappa_3,l_4\kappa_4}(i\omega)d\omega. \qquad (8)$$

The integral over frequencies is straightforward to evaluate using numerical methods, so the problem of calculating dispersion energy coefficients reduces to one of evaluating the imaginary-frequency-dependent polarizabilities of the interacting molecules. These can be obtained from the response of the molecule to a time-dependent perturbation, as discussed in section 2.1.

In the pseudostate method [13], the electronic Hamiltonian is diagonalized in a many-electron basis set to produce a ground state and a number of electronically excited states. If these are inserted into equation (6), and the basis set is chosen carefully, the result closely approximates the exact dispersion energy coefficients. Similarly, the pseudostates can be used to evaluate frequency-dependent polarizabilities using equation (7). The pseudostates do not need to include accurate excited states, with the possible exception of low-lying states, if all that is required is the polarizabilities at imaginary frequencies and the dispersion energy coefficients [14].

The computational cost of calculating the dispersion energy coefficients from pseudostates scales as the product of the number of pseudostates on the two molecules, which can become inconveniently large. To minimize the computational cost, the number of pseudostates can be reduced, often by several orders of magnitude, using a fitting procedure. The imaginary-frequency-dependent polarizabilities are first calculated using the full set of pseudostates, or using some other method such as time-dependent perturbation theory, and then a small set of pseudostates is fitted to the polarizabilities. This can be done independently for every set of angular momentum indices. Three fitted pseudostates for each set of $l\kappa, l'\kappa'$ indices is usually a sufficient number to give excellent agreement with the calculated polarizabilities over the whole range of imaginary frequencies. Unlike the electronic (pseudo-)states obtained from diagonalizing the electronic Hamiltonian, the fitted pseudostates are not many-electron functions. They simply consist of two scalar quantities: the pseudo-excitation energy, $E^a_A - E^{(0)}_A$, and the multipole interaction strength, $\langle \psi^{(0)}_A | \hat{Q}_{A,l\kappa} | \psi^a_A \rangle \langle \psi^a_A | \hat{Q}_{A,l'\kappa'} | \psi^{(0)}_A \rangle$.

Dispersion energy coefficients can therefore be calculated using either imaginary-frequency-dependent polarizabilities or pseudostates of the interacting molecules. The usual method for calculating these quantities is time-dependent perturbation theory. This is reviewed in section 2.1, and some results are given in section 2.2. Alternatively, it is possible to use a projection operator to shift the ground-state energy of the molecule, which enables pseudostates to be calculated. This 'energy shift' method was used in our calculations on the water–nitrogen dimer, see section 3, and it does not seem to have been widely reported in the literature, so it is briefly described in section 2.3.

2.1. *Time-dependent perturbation theory*

The frequency-dependent polarizability defined in equation (7) can be calculated by adding a time-dependent perturbation $\hat{Q}_A(t)$ to the unperturbed zero-order electronic Hamiltonian $\hat{H}_A^{(0)}$:

$$\hat{H}_A = \hat{H}_A^{(0)} + \lambda \hat{Q}_{A,lk}\left(e^{i\omega t} + e^{-i\omega t}\right) + \mu \hat{Q}_{Al'\kappa'},\left(e^{i\omega t} + e^{-i\omega t}\right). \tag{9}$$

The unperturbed wave function is $\psi_A^{(0)}$, which is independent of t. The perturbed wave function is expanded in powers of λ, μ and $e^{i\omega t}$:

$$\psi_A = \psi_A^{(0)} + \lambda \psi_A^{(1,0;1)} e^{i\omega t} + \lambda \psi_A^{(1,0;-1)} e^{-i\omega t} + \mu \psi_A^{(0,1;1)} e^{i\omega t} + \mu \psi_A^{(0,1;-1)} e^{-i\omega t} + \cdots. \tag{10}$$

and the energy E_A is expanded in a similar way:

$$\begin{aligned} E_A = E_A^{(0)} &+ \lambda E_A^{(1,0;1)} e^{i\omega t} + \lambda E_A^{(1,0;-1)} e^{-i\omega t} + \mu E_A^{(0,1;1)} e^{i\omega t} + \mu E_A^{(0,1;-1)} e^{-i\omega t} \\ &+ \lambda\mu E_A^{(1,1;0)} + \lambda\mu E_A^{(1,1;2)} e^{2i\omega t} + \lambda\mu E_A^{(1,1;-2)} e^{-2i\omega t} \ldots, \end{aligned} \tag{11}$$

where the quantities $\psi_A^{(p,q;s)}$ and $E_A^{(p,q;s)}$, which multiply $\lambda^p \mu^q e^{i\omega t s}$ in these equations, are time-independent. The Hamiltonian, wave function and energy are substituted into a time-dependent Lagrangian equation [15, 16]

$$\left(\hat{H}_A - i\frac{d}{dt}\right)\psi_A = E_A \psi_A. \tag{12}$$

Equating powers of λ, μ and $e^{i\omega t}$ on both sides of the resulting equation gives the following expressions for the perturbed wave function and energy:

$$\hat{H}_A^{(0)}\psi_A^{(0)} = E_A^{(0)}\psi_A^{(0)} \tag{13}$$

$$(\hat{H}_A^{(0)} + \omega)\psi_A^{(1,0;1)} + \hat{Q}_{A,lk}\psi_A^{(0)} = E_A^{(0)}\psi_A^{(1,0;1)} + E_A^{(1,0;1)}\psi_A^{(0)} \tag{14}$$

$$(\hat{H}_A^{(0)} - \omega)\psi_A^{(1,0;-1)} + \hat{Q}_{A,lk}\psi_A^{(0)} = E_A^{(0)}\psi_A^{(1,0;-1)} + E_A^{(1,0;-1)}\psi_A^{(0)} \tag{15}$$

$$(\hat{H}_A^{(0)} + \omega)\psi_A^{(0,1;1)} + \hat{Q}_{A,l'\kappa'}\psi_A^{(0)} = E_A^{(0)}\psi_A^{(0,1;1)} + E_A^{(0,1;1)}\psi_A^{(0)} \tag{16}$$

$$(\hat{H}_A^{(0)} - \omega)\psi_A^{(0,1;-1)} + \hat{Q}_{A,l'\kappa'}\psi_A^{(0)} = E_A^{(0)}\psi_A^{(0,1;-1)} + E_A^{(0,1;-1)}\psi_A^{(0)} \tag{17}$$

$$\begin{aligned} \hat{H}_A^{(0)}\psi_A^{(1,1;0)} &+ \hat{Q}_{A,lk}(\psi_A^{(0,1;1)} + \psi_A^{(0,1;-1)}) + \hat{Q}_{A,l'\kappa'}(\psi_A^{(1,0;1)} + \psi_A^{(1,0;-1)}) \\ &= E_A^{(0)}\psi_A^{(1,1;0)} + E_A^{(1,0;1)}\psi_A^{(0,1;-1)} + E_A^{(1,0;-1)}\psi_A^{(0,1;1)} \\ &+ E_A^{(0,1;1)}\psi_A^{(1,0;-1)} + E_A^{(0,1;-1)}\psi_A^{(1,0;1)} + E_A^{(1,1;0)}\psi_A^{(0)} \end{aligned} \tag{18}$$

Using the intermediate normalization convention, multiplying both sides of equation (18) by $\psi_A^{(0)}$, and integrating both sides gives

$$E_A^{(1,1;0)} = \left\langle \psi_A^{(0)} \left| \hat{Q}_{A,l\kappa} \right| \psi_A^{(0,1;1)} + \psi_A^{(0,1;-1)} \right\rangle + \left\langle \psi_A^{(0)} \left| \hat{Q}_{A,l'\kappa'} \right| \psi_A^{(1,0;1)} + \psi_A^{(1,0;-1)} \right\rangle. \quad (19)$$

The first-order perturbed wave function can be expanded using electronically excited states of molecule A:

$$\psi_A^{(1,0;1)} = \sum_{a>0} c_A^{a,(1,0;1)} \psi_A^a, \quad (20)$$

where the coefficient $c_A^{a,(1,0;1)}$ is complex, and intermediate normalization means that the ground state is not included in the expansion. Inserting equation (20) into equation (14), multiplying both sides by ψ_A^a, and integrating both sides gives

$$\left\langle \psi_A^a \left| \hat{Q}_{A,l\kappa} \right| \psi_A^{(0)} \right\rangle + c_A^{a,(1,0;1)}(E_A^a + \omega) = c_A^{a,(1,0;1)} E_A^{(0)}. \quad (21)$$

Substituting the resulting expression for $c_A^{a,(1,0;1)}$ into equation (20) gives

$$\psi_A^{(1,0;1)} = -\sum_{a>0} \psi_A^a \left\langle \psi_A^a \left| \hat{Q}_{A,l\kappa} \right| \psi_A^{(0)} \right\rangle \Big/ \left(E_A^a - E_A^{(0)} + \omega \right). \quad (22)$$

Similar expressions are found for $\psi_A^{(1,0;-1)}$, $\psi_A^{(0,1;1)}$ and $\psi_A^{(0,1;-1)}$. When these are substituted into equation (19), the resulting expression for $E_A^{(1,1;0)}$ is twice the frequency-dependent polarizability $\alpha_{l\kappa,l'\kappa'}^A(\omega)$ defined in equation (7).

   Calculation of frequency-dependent polarizabilities by this method can be implemented in practical electronic structure calculations (where the Schrödinger equation cannot be solved exactly) by using the operator $\hat{H}_A - i\frac{d}{dt}$ in place of $\hat{H}_A^{(0)}$ in the appropriate energy expression, and expanding the energy in powers of λ, μ and $e^{i\omega t}$. As shown above, the time-independent coefficient multiplying $\lambda\mu$ in the expansion of the energy is twice the required frequency-dependent polarizability. If this is done using Hartree–Fock theory, the resulting polarizabilities are time-dependent coupled Hartree–Fock (TD-CHF) polarizabilities [17]. Time-dependent second-order Moller–Plesset (TD-MP2) polarizabilities can also be calculated, and a full theoretical description has been given by Hättig and Hess [18].

   Recently, time-dependent coupled-cluster (TD-CCSD) polarizabilities have been implemented for open-shell, spin-unrestricted molecules [19]. The energy equation for the TD-CCSD method is based on the Lagrangian formulation [20]

of the CCSD equations. The TD-CCSD Lagrangian is

$$E_A = \left\langle \phi_A^{(0)} \middle| \hat{\Lambda}_A e^{-\hat{T}_A(t)} e^{-\hat{\kappa}_A(t)} \left(\hat{H}_A^{(0)} - i\frac{d}{dt} + \hat{Q}_A(t) \right) e^{\hat{\kappa}_A(t)} e^{\hat{T}_A(t)} \middle| \phi_A^{(0)} \right\rangle$$

$$+ \left\langle \phi_A^{(0)} \middle| \left[\hat{\zeta}_A, e^{-\hat{\kappa}_A(t)} \left(\hat{H}_A^{(0)} - i\frac{d}{dt} + \hat{Q}_A(t) \right) e^{\hat{\kappa}_A(t)} \right] \middle| \phi_A^{(0)} \right\rangle \tag{23}$$

where $\hat{\Lambda}_A$ and $\hat{\zeta}_A$ are de-excitation operators, and their amplitudes are Lagrange multipliers, which are introduced to enforce the usual conditions for the CCSD amplitudes, T, and for the TD-CHF orbital rotation operators, $\hat{\kappa}$, respectively. The operator $\hat{T}_A$ is a sum of single and double excitations, multiplied by CCSD amplitudes; $\hat{\Lambda}_A - 1$ is a sum of single and double de-excitations, multiplied by λ coefficients; $\phi_A^{(0)}$ is the unperturbed Hartree–Fock wave function, and the orbital rotation operators $\hat{\kappa}_A$ are fixed at their TD-CHF values.

Using the stationary properties of this Lagrangian, the required second-order energy can be expressed in a form which contains terms up to first order in amplitudes and orbital rotations, and zero-order Lagrange multipliers:

$$E_A^{(2)} = \left\langle \phi_A^{(0)} \middle| \hat{\Lambda}_A^{(0)} e^{-\hat{T}_A^{(0)}} \frac{1}{2} \left[\left[\hat{H}_A^{(0)} - i\frac{d}{dt}, \hat{T}_A^{(1)} \right], \hat{T}_A^{(1)} \right] e^{\hat{T}_A^{(0)}} \middle| \phi_A^{(0)} \right\rangle$$

$$+ \left\langle \phi_A^{(0)} \middle| \hat{\Lambda}_A^{(0)} e^{-\hat{T}_A^{(0)}} \left[\left[\hat{H}_A^{(0)} - i\frac{d}{dt}, \hat{\kappa}_A^{(1)} \right], \hat{T}_A^{(1)} \right] e^{\hat{T}_A^{(0)}} \middle| \phi_A^{(0)} \right\rangle$$

$$+ \left\langle \phi_A^{(0)} \middle| \hat{\Lambda}_A^{(0)} e^{-\hat{T}_A^{(0)}} \frac{1}{2} \left[\left[\hat{H}_A^{(0)} - i\frac{d}{dt}, \hat{\kappa}_A^{(1)} \right], \hat{\kappa}_A^{(1)} \right] e^{\hat{T}_A^{(0)}} \middle| \phi_A^{(0)} \right\rangle$$

$$+ \left\langle \phi_A^{(0)} \middle| \hat{\Lambda}_A^{(0)} e^{-\hat{T}_A^{(0)}} \left[\hat{Q}_A, \hat{T}_A^{(1)} + \hat{\kappa}_A^{(1)} \right] e^{\hat{T}_A^{(0)}} \middle| \phi_A^{(0)} \right\rangle$$

$$+ \left\langle \phi_A^{(0)} \middle| \left[\hat{\zeta}_A^{(0)}, \frac{1}{2} \left[\left[\hat{H}_A^{(0)} - i\frac{d}{dt}, \hat{\kappa}_A^{(1)} \right], \hat{\kappa}_A^{(1)} \right] \right] \middle| \phi_A^{(0)} \right\rangle$$

$$+ \left\langle \phi_A^{(0)} \middle| \left[\hat{\zeta}_A^{(0)}, \left[\hat{Q}_A, \hat{\kappa}_A^{(1)} \right] \right] \middle| \phi_A^{(0)} \right\rangle. \tag{24}$$

The zero-order quantities in this equation are evaluated [20] by performing a ground-state CCSD calculation, giving $\hat{T}_A^{(0)}$, then solving the CCSD lambda equations to obtain $\hat{\Lambda}_A^{(0)}$, and solving the CCSD zeta equations to obtain $\hat{\zeta}_A^{(0)}$. The TD-CHF equations are solved to obtain $\hat{\kappa}_A^{(1)}$.

The calculation of the first-order amplitudes in $\hat{T}_A^{(1)}$ usually dominates the overall computational expense, as they require essentially a separate CCSD calculation for each perturbation. Definitions of the first-order amplitudes are obtained by making the second-order energy in equation (24) stationary with respect to all $\hat{T}_{A,X}^{(1)}$, where X

includes all single and double excitations. The time-dependent first-order amplitudes are expanded in powers of the perturbations,

$$\hat{T}_{A,X}^{(1)} = \lambda \hat{T}_{A,X}^{(1,0;1)} e^{i\omega t} + \lambda \hat{T}_{A,X}^{(1,0;-1)} e^{-i\omega t} + \mu \hat{T}_{A,X}^{(0,1;1)} e^{i\omega t} + \mu \hat{T}_{A,X}^{(0,1;-1)} e^{-i\omega t}, \qquad (25)$$

and the time-independent second-order energy is made stationary with respect to $\hat{T}_{A,X}^{(1,0;1)}$ to obtain the first-order amplitude equation for $\hat{T}_{A}^{(0,1;-1)}$, and similarly for the other first-order amplitudes. The required integrals can then be written in terms of quantities which appear in the time-independent CCSD perturbation theory: residuals of the amplitude equations

$$\left\langle \phi_A^{(0)} \left| \hat{X}_A^{\dagger} e^{-\hat{T}_A} \hat{H}_A e^{\hat{T}_A} \right| \phi_A^{(0)} \right\rangle, \qquad (26)$$

residuals of the lambda equations

$$\left\langle \phi_A^{(0)} \left| \hat{\Lambda}_A^{(0)} e^{-\hat{T}_A} \left[\hat{H}_A, \hat{X}_A \right] e^{\hat{T}_A} \right| \phi_A^{(0)} \right\rangle, \qquad (27)$$

and the one-particle and two-particle densities

$$\left\langle \phi_A^{(0)} \left| \hat{\Lambda}_A^{(0)} e^{-\hat{T}_A} \left(\hat{X}_A + \hat{X}_A^{\dagger} \right) e^{\hat{T}_A} \right| \phi_A^{(0)} \right\rangle. \qquad (28)$$

The cluster operator $e^{\hat{T}_A}$ in these quantities appears at zero order, $e^{\hat{T}_A^{(0)}}$, and first order, $\hat{T}_A^{(1)} e^{\hat{T}_A^{(0)}}$. The operator $\hat{H}_A$ similarly appears as a zero-order Hamiltonian, $\hat{H}_A^{(0)}$, and as an effective first-order Hamiltonian, $[\hat{H}_A^{(0)}, \hat{\kappa}_A^{(1)}]$. The imaginary part of the effective first-order Hamiltonian is anti-Hermitian, so some care is needed in ordering the indices of the one-electron and two-electron integrals produced from it. As defined, the two-particle density and the effective first-order Hamiltonian both contain components with four virtual orbital indices, which would require a large amount of computer storage, but in practice all contributions of these components to the polarizability can be rewritten so that they do not require explicit calculation of the all-virtual quantities, by reordering the summation over indices.

2.2. Results

The use of the TD-CCSD method to calculate dispersion energy coefficients for the water–oxygen interaction is described in section 4. To justify the use of this method, it is demonstrated here that the TD-CCSD results for these molecules are in agreement with the best literature values.

To calculate induction and dispersion energies, polarizabilities and dispersion energy coefficients, the basis set needs to be able to describe polarization of the ground-state wavefunction, as discussed above. However, the polarization functions included in most standard basis sets optimize the intramolecular correlation, and are not diffuse enough to give converged dispersion energy coefficients C_n, especially for large n,
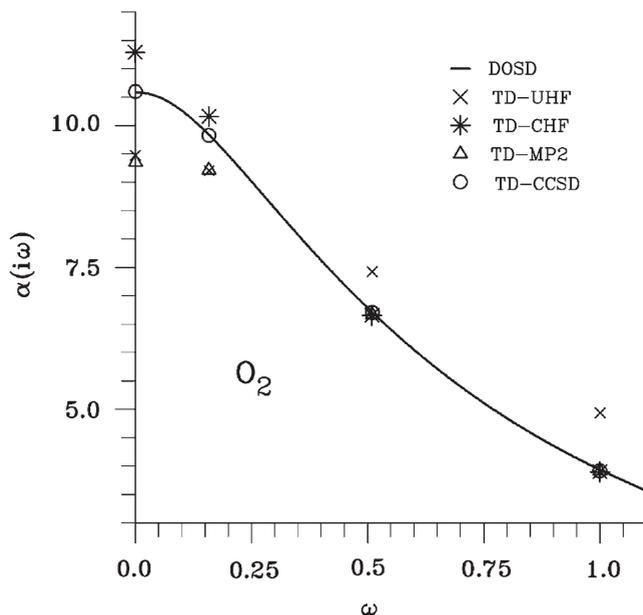
Figure 1. Calculated values of the imaginary-frequency-dependent dipole polarizability of the oxygen molecule (symbols), compared with benchmark experimental literature values (solid curve), in atomic units.

unless very extensive basis sets are used. Therefore, the SP series of basis sets has been devised [9, 21]. The SPQZ basis set for atoms B-Ar is obtained from the aug-cc-pVQZ basis set [22] by changing the exponents of the three f Gaussians to equal the exponents of the three most diffuse d Gaussians, and changing the exponents of the two g Gaussians to equal the exponents of the two most diffuse d Gaussians. Similarly, the SPTZ basis set is defined by changing the exponents of the two most diffuse f Gaussians to equal the exponents of the two most diffuse d Gaussians. For the H and He atoms, the d and f Gaussian exponents are changed so that they equal the most diffuse p Gaussian exponents.

In figure 1, the spherically averaged, imaginary-frequency-dependent dipole polarizability of the oxygen molecule is shown. The results were calculated using the SPQZ basis set, with unrestricted electron spins [19]. The agreement with the benchmark literature results, which were obtained from scaled dipole oscillator strength distributions (DOSD) [23], is excellent. Other methods based on Hartree–Fock theory and MP2 theory do not agree with the literature values for small frequencies, and the TD-UMP2 method is surprisingly poor. Similarly, the TD-CCSD method is found to give frequency-dependent polarizabilities for water in very good agreement with the DOSD results, as shown in figure 2.

Spherically averaged C_6 and C_8 dispersion energy coefficients for the dimers involving water and oxygen are shown in table 1. These were calculated from the dipole polarizabilities shown in figures 1 and 2, and from the analogous quadrupole polarizabilities. As expected from the good agreement with the experimental polarizabilities,
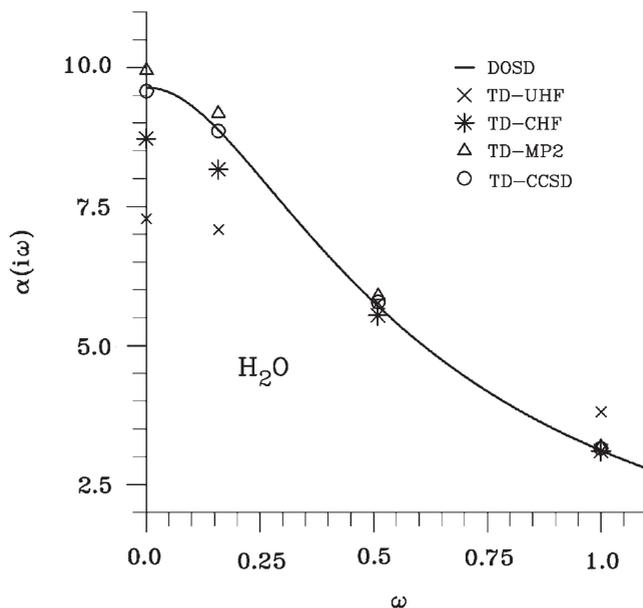
Figure 2. Calculated values of the imaginary-frequency-dependent dipole polarizability of the water molecule (symbols), compared with benchmark experimental literature values (solid curve), in atomic units.

Table 1. Spherically averaged dispersion energy coefficients (in atomic units), calculated using the TD-UHF, TD-CHF, TD-MP2 and TD-CCSD methods with the SPQZ basis set and with unrestricted electron spins. Literature values are taken from constrained dipole oscillator strength distributions.

		TD-UHF	TD-CHF	TD-MP2	TD-CCSD	Literature
O_2-O_2	C_6	71.9	62.9	57.4	61.2	61.6 [23]
	C_8	1692	1742	1762	1813	
H_2O-H_2O	C_6	42.3	40.8	48.0	45.6	45.3 [23]
	C_8	895	1016	1261	1174	
H_2O-O_2	C_6	55.1	50.6	52.2	52.7	52.7 [23]
	C_8	1232	1331	1487	1458	

the TD-CCSD method gives excellent C_6 dispersion energy coefficients for all three dimers. The other methods overestimate C_6 for one homomolecular pair and underestimate it for the other, so the C_6 results for the water–oxygen dimer, which are close to the geometric mean of the other two, are in fair agreement with the experimental value, but are still not as close to it as the TD-CCSD result. Interestingly, the differences in C_8 values do not even qualitatively resemble the differences in C_6 values between the different methods. The TD-CCSD results are expected to be the most accurate, but there are no reliable literature data for C_8 to test this.

2.3. *Energy shift method*

The aim of the energy shift method is to calculate the 'polarizability-like' quantity $\tilde{\alpha}(\omega)$, which is defined using a sum over excited states as

$$\tilde{\alpha}_{l\kappa, l'\kappa'}^{A}(\omega) = 2 \sum_a \left\langle \psi_A^{(0)} \left| \hat{Q}_{A,l\kappa} \right| \psi_A^a \right\rangle \left\langle \psi_A^a \left| \hat{Q}_{A,l'\kappa'} \right| \psi_A^{(0)} \right\rangle \Big/ \left(E_A^a - E_A^{(0)} + \omega \right). \tag{29}$$

The frequency-dependent polarizability is then given by $\alpha(\omega) = \tilde{\alpha}(\omega)/2 + \tilde{\alpha}(-\omega)/2$, and for imaginary frequencies $\alpha(i\omega) = \Re[\bar{\alpha}(i\omega)]$. However, to obtain polarizabilities at imaginary frequencies using the energy shift method, instead of calculating the quantity $\tilde{\alpha}(i\omega)$, it is more efficient to calculate $\tilde{\alpha}(\omega)$ for real, positive values of ω, and to fit the results to a set of pseudostates, which can then be inserted into equation (7) to find the frequency-dependent polarizabilities. Using positive values of the energy shift parameter ω avoids singularities in the polarizability, and using real values of ω allows only real-valued quantities to be used in the calculation of $\tilde{\alpha}(\omega)$, which increases the computational efficiency of the method compared to time-dependent perturbation theory.

The quantity $\tilde{\alpha}_A(\omega)$ can be regarded as the static polarizability of a molecule identical to A, whose ground-state energy is shifted from $E_A^{(0)}$ to $E_A^{(0)} - \omega$. This shift of the ground-state energy can be reproduced by replacing the unperturbed Hamiltonian $\hat{H}_A^{(0)}$ by $\hat{H}_A^{(0)} - |\psi_A^{(0)}\rangle \omega \langle \psi_A^{(0)}|$. The projection operator (which should not be regarded as a perturbation) shifts the ground-state energy, but not the excited-state energies, and it does not affect the ground-state or excited-state wave functions. The perturbed Hamiltonian is

$$\hat{H}'_A = \hat{H}_A^{(0)} - \left| \psi_A^{(0)} \right\rangle \omega \left\langle \psi_A^{(0)} \right| + \lambda \hat{Q}_{A,l\kappa} + \mu \hat{Q}_{A,l'\kappa'} \tag{30}$$

and $\tilde{\alpha}_{l\kappa, l'\kappa'}^{A}(\omega)$ is equal to the coefficient of $\lambda\mu$ in the resulting second-order energy expression.

The energy shift method is straightforward to apply to variational electronic structure calculations, such as Hartree–Fock and configuration interaction (CI), and it can also be applied to the non-variational MP2 and CCSD methods, using techniques similar to those used for calculating frequency-dependent polarizabilities.

The main advantages of the energy shift method are simplicity (it does not use time-dependent or anti-Hermitian operators) and speed (it does not involve complex numbers). A further advantage could be gained if use of the method with 'relaxed' orbitals avoided the TD-CHF poles that cause problems for time-dependent perturbation theory calculations of real-frequency-dependent polarizabilities. The method has not been rigorously tested using high-level electronic structure calculations, such as CCSD, with large basis sets, so its accuracy relative to the use of time-dependent perturbation theory is not known, but further investigations may be worthwhile.

3. The water–nitrogen interaction

The complex formed by water and molecular nitrogen has been a focus of several experimental and theoretical studies. Microwave spectroscopy of the H_2O–N_2 dimer [24] suggests that there are four equivalent hydrogen-bonded structures, each having a nearly linear N–N–H–O geometry with an N–H distance of 2.42 Å and an OHN angle of 169°. A semi-empirical potential of the complex of water with N_2 was proposed by Sandler *et al.* [25]. This work included calculations of the potential energy surface at the MP2 level of theory, analysis of different contributions to the interaction energy (exchange-repulsion, electrostatic, dispersion and induction), and the construction of a H_2O–N_2 potential function by fitting to *ab initio* data, followed by adjustment against experimental rotational constants and quadrupole coupling constants [24]. The minimum was predicted to have a well depth of 437 cm^{-1}.

Our study of the water–nitrogen interaction focussed on the thermodynamic properties of the mixture, for which a complete five-dimensional potential energy surface is needed. Second virial coefficients at low temperatures are sensitive to the potential well, while at higher temperatures they depend on the average of the potential over a wide range of intermolecular angles and distances. The SIMPER method was used to calculate the intermolecular potential, because the MP2 method gives inaccurate dispersion energy coefficients and is therefore likely to overestimate the depth of the potential energy. Adjusting the MP2 potential to reproduce the microwave data, as done by Sandler et al., improves the description of the potential around the minimum, but does not necessarily improve the orientation-averaged potential energy, which includes contributions from geometries far from equilibrium that are not sampled by the microwave experiment.

3.1. *Methodology*

The SIMPER method was used to obtain an intermolecular potential energy surface describing the interaction of rigid water and nitrogen molecules. The calculations used the SP series of basis sets described in section 2, in order to improve the convergence of the dispersion and induction energy to the basis set. The interaction energy was calculated for 12 288 intermolecular geometries, consisting of 12 intermolecular separations from 5 to 13 a_0, and 1024 relative orientations.

The intermolecular potential was first evaluated using counterpoise-corrected MP2 supermolecule calculations. For comparison, the intermolecular potential was also evaluated at the CCSD(T) supermolecule level of theory in the region of the global minimum; these calculations took several days per point using the SPQZ basis set. The SIMPER method was applied to the MP2 potential energy surface, using a modified version of equation (1)

$$E_{\text{SIMPER}} = \left(E_{\text{MP2}} - \sum_k E_{\text{MP2}}^{(k)}\right)(S_{\rho,\,\text{high}}/S_{\rho,\,\text{MP2}}) + \sum_k E_{\text{high}}^{(k)}, \qquad (31)$$

and the sums over orders of perturbation k were truncated at $k = 2$. The first-order energy, $E^{(1)}$, is the classical electrostatic interaction between unperturbed monomer charge densities, and at 'high' level this was calculated from CCSD charge densities. The second-order energy is a sum of induction and dispersion energies. At 'high' level, the induction energy was calculated as the sum of the MP2 response of N_2 to the electrostatic potential produced by the CCSD density of H_2O, plus the MP2 response of H_2O to the electrostatic potential produced by the CCSD density of N_2. This was the most time-consuming part of the SIMPER method, as it essentially required two monomer MP2 calculations, one for each molecule in the CCSD potential of the other, at each point on the potential energy surface. However, these calculations took less time than the MP2 supermolecule calculations on the dimer at each point, so the SIMPER method did not add greatly to the overall computational cost. The 'high-level' dispersion energy was calculated as described in section 1, by combining MP2 damping functions with high-level dispersion energy coefficients. The high-level dispersion energy coefficients were calculated using 'energy shifted' CISD theory, as outlined in section 2.3. The spherically averaged C_6 value was 59.8 atomic units, compared to 64.3 for the MP2 supermolecule method and 57.6 obtained using DOSD results from the literature [23].

The quantity $(E_{MP2} - \sum_k E_{MP2}^{(k)})$ in equation (31) is approximately equal to the exchange-repulsion energy, although it also contains higher-order 'Coulomb' perturbation contributions from RSPT. It would not be helpful to treat high orders of Coulomb and exchange energies separately, since the total Coulomb energy to infinite order is unphysically large and negative [26], the exchange energy is unphysically large and positive, and the two almost cancel each other. We investigated truncating k at 3 instead of 2, but the results were similar, and not demonstrably better, so the computationally cheaper truncation at $k = 2$ was preferred.

The exchange-repulsion energy is an important component of the intermolecular potential, and is difficult to calculate explicitly at a high level of correlation. Within SIMPER the overlap model [27, 28] is used to relate the low-level exchange-repulsion energy to the electron density overlap $S_{\rho, MP2}(R, \Omega)$

$$E_{exch, MP2}(R, \Omega) = K_{MP2}(R, \Omega) \times S_{\rho, MP2}(R, \Omega), \tag{32}$$

where $E_{exch, MP2}$ is defined to be $(E_{MP2} - \sum_k E_{MP2}^{(k)})$ and $K_{MP2}(R, \Omega)$ is a function of the intermolecular coordinates, which is found from equation (32) at each point on the surface. The value of K is assumed to be the same for the high-level exchange-repulsion energy as for the MP2 exchange-repulsion, so the high-level exchange-repulsion energy is calculated as

$$E_{exch, high}(R, \Omega) = K_{MP2}(R, \Omega) \times S_{\rho, high}(R, \Omega). \tag{33}$$

The 'high-level' charge density overlap integral was calculated from CCSD electron densities. Combining equation (32) and equation (33) produces the factor of $(S_{\rho, high}/S_{\rho, MP2})$ used to obtain the 'high-level' exchange-repulsion energy from the MP2 exchange-repulsion energy in equation (31).

To estimate the intermolecular potential in the complete basis set limit, the two-point extrapolation formula of Bak *et al.* [29] was applied to the intermolecular potentials obtained from the SPTZ and SPQZ basis sets:

$$E_{x-1,\,x} = \frac{x^3 E_x - (x-1)^3 E_{x-1}}{x^3 - (x-1)^3}, \tag{34}$$

where x is the cardinality of the basis set (here $x = 4$). The effect of basis set extrapolation was generally to shift the potential energy toward more negative values, although the difference between the SPQZ basis set and the estimated complete basis set limit was small compared to the difference between the supermolecule and SIMPER methods. Extrapolating from the SPQZ basis set to the complete basis set limit increased the well depth by no more than 2%.

The SIMPER potential energy surface, extrapolated to the complete basis set limit, was fitted to an expansion in atom-centred form, with centres on the hydrogen atoms and oxygen atom of the water molecule, and on both nitrogen atoms. A correct description of the long-range interaction energy is important for calculating the second virial coefficients. In order to ensure a proper asymptotic behaviour of the potential, a switching function

$$F(R) = \{1 + \tanh[6(R/a_0) - 78]\}/2 \tag{35}$$

was used to change from the fitted potential to a long-range multipole expansion of the potential at around $R = 13\ a_0$. Thus, the final functional form used for the intermolecular potential was

$$E(R, \Omega) = F(R)E_{\text{mult}}(R, \Omega) + (1 - F(R))E_{\text{fit}}(R, \Omega). \tag{36}$$

3.2. *Results*

A contour diagram of the fitted potential energy surface is shown in figure 3. The figure shows planar geometries of the water–nitrogen complex. The centre of the nitrogen molecule is placed at the spherical polar coordinates $(R, \theta, 0)$, and the water molecule is in the xz plane, with the O atom at the origin; the z axis points from the midpoint of the H nuclei towards the O nucleus. The orientation of the nitrogen molecule is given by the spherical polar angles $(\theta', 0)$. The intermolecular potential was minimized with respect to intermolecular separation R, and the resulting energy is shown in the figure. The figure shows a single, symmetry-distinct, planar minimum, and it can be seen that 'geared' rotation of the two molecules is a much more facile motion than rotating both molecules in the same direction from the equilibrium geometry.

A few geometries of the water–nitrogen dimer were chosen for further study [9], using CCSD(T) supermolecule calculations as a benchmark. The minimum energy was found at a 'hydrogen-bonded' structure with $\theta \approx \theta' \approx 110°$. The well depth of the fitted SIMPER potential was approximately $440\,\text{cm}^{-1}$. This entirely *ab initio* value agrees clo-
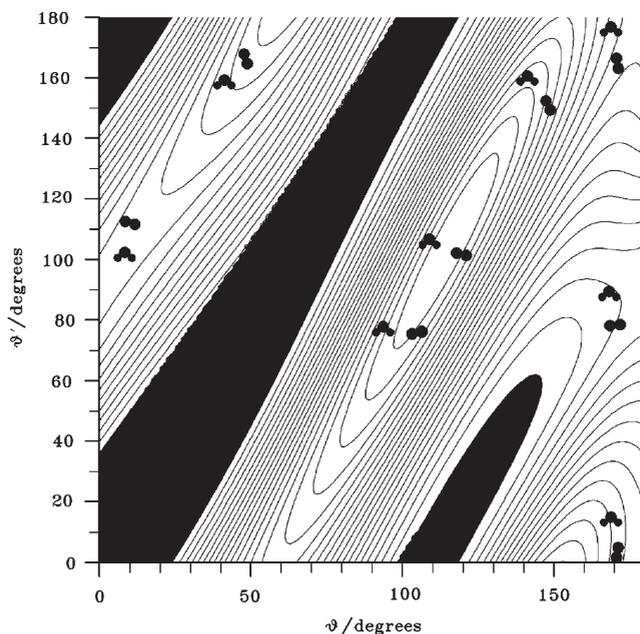
Figure 3. Contour plot of the fitted water–nitrogen potential energy surface, calculating using SIMPER and extrapolated to the complete basis set limit. The distance-optimized energy is shown for every planar orientation, where θ is the polar angle defining the centre of the nitrogen molecule, and θ' defines its orientation, as described in the text. The top and bottom of the figure are identical, by symmetry. Black regions of the plot are repulsive orientations. Contours are drawn every 20 cm^{-1}. The geometry of the complex is drawn to scale at several positions on the plot, and the global minimum appears just to the right of centre.

sely with the well depth of $437\,\mathrm{cm}^{-1}$ for the potential of Sandler *et al.* [25], which was fitted to experimental data. A repulsive C_{2v} geometry with $\theta = \theta' = 0$, where the oxygen atom of water contacts one end of the nitrogen molecule, was also considered, together with several points for intermediate values of $\theta = \theta'$. In all cases, close agreement between the SIMPER and CCSD(T) methods was found. For example, near the potential minimum, at $R = 7.3\,a_0$, $\theta = \theta' = 120°$, the MP2, CCSD(T) and SIMPER methods gave interaction energies of -433, -415 and $-412\,\mathrm{cm}^{-1}$ respectively. In a repulsive geometry with $R = 12\,a_0$, $\theta = \theta' = 0$, the results were 20.7, 18.6 and $18.8\,\mathrm{cm}^{-1}$. The results demonstrate that the SIMPER method correctly reduces the well depth of the potential, relative to MP2, and that the use of high-level components from perturbation theory gives the correct long-range behaviour.

Table 2 shows the components of the MP2 and SIMPER potentials at two planar geometries, one close to the global minimum ('H') at $\theta = \theta' = 120°$, and the other ('O') at $\theta = 0$, $\theta' = 90°$, where the oxygen atom is equidistant from the N atoms. At both orientations the first-order (electrostatic) energy is a major attractive component of the potential energy. The interaction between the dipole moment of water and the quadrupole moment of nitrogen qualitatively determines the sign of the electrostatic energy, but higher multipoles and non-multipolar charge overlap effects also have

Table 2. Components of the water–nitrogen intermolecular potential (cm^{-1}) in the vicinity of the hydrogen-bonded global minimum (H), and in a C_{2v} geometry (O) defined in the text. Results were obtained with the SPQZ basis set, and the total potential was distance-optimized.

Geom	Level	Electrostatic	Induction	Dispersion	Exch-Rep	Total
H	MP2	−550.7	−190.2	−374.4	688.3	−427.0
H	SIMPER	−517.8	−190.8	−351.9	650.8	−409.7
O	MP2	−367.9	−98.9	−411.8	629.1	−249.5
O	SIMPER	−335.7	−99.1	−374.4	601.3	−207.9
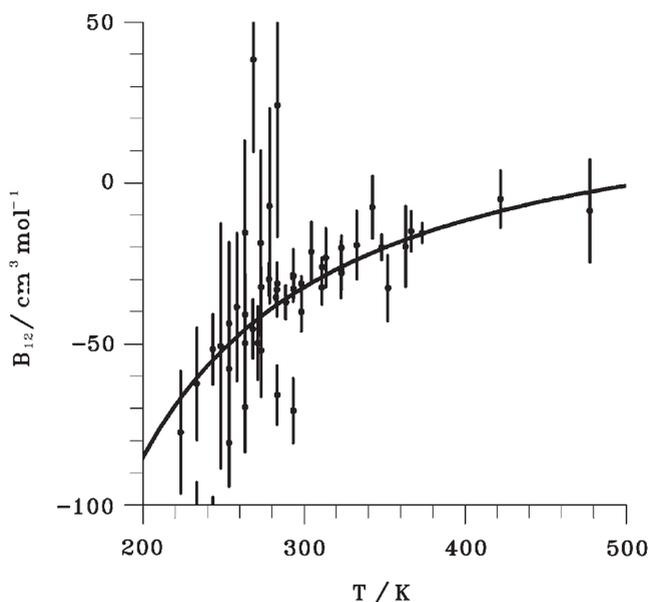
Figure 4. Calculated water–nitrogen mixed second virial coefficients (curve) and experimental data with quoted uncertainties in the virial coefficients (error bars). The calculations and experimental data sources are described in the text.

an important effect. The dispersion energy is less negative for the SIMPER potential than for the supermolecule MP2 potential, which results partly from the difference in the C_6 coefficients. The MP2 multipoles are larger than the CCSD multipoles, and the MP2 method therefore overestimates the magnitude of the electrostatic energy relative to the calculation based on CCSD densities. The MP2 density is also more spatially extended than the CCSD density, which produces a larger electron density overlap, so the MP2 exchange-repulsion energy is larger than the SIMPER exchange-repulsion energy. Overall, the changes in the dispersion and electrostatic energy outweigh the change in the exchange-repulsion, so the SIMPER potential is shallower than the MP2 potential.

The second virial coefficient B_{12} for the water–nitrogen pair was calculated from the SIMPER potential as a function of temperature, including translational and rotational quantum effects to first order, as described in [30]. The results are shown in figure 4.

Experimental water–nitrogen data from a number of different sources [31–39] have been collated and analysed to produce experimental second virial coefficients [9, 40–42]. These are also shown in the figure. The experimental data are clearly inconsistent within their mutual uncertainties. However, the SIMPER calculations are consistent with the majority of sources. The theoretical results cover temperatures from 100 K and 3000 K. The lower temperature of 100 K is estimated to be the maximum temperature at which higher quantum corrections may become important, and second virial coefficients at temperatures much higher than 3000 K will be significantly affected by thermal excitation of molecular vibrations, especially the water bending mode, and by points high on the repulsive wall that may not be represented well by the fitted potential. Nevertheless, the theoretical work provides values of B_{12} under extreme temperatures where no experimental data exist, and the uncertainty in the calculations is estimated to be less than, or similar to, the uncertainty in the best experimental data.

In conclusion, the SIMPER method, which scales as the fifth power of system size per point on the surface (plus a single sixth-power scaling CCSD calculation per monomer), gives a water–nitrogen potential energy surface in excellent agreement with the seventh-power scaling supermolecule CCSD(T) method, while the supermolecule MP2 method, which requires a similar computational effort to SIMPER, overestimates the magnitude of the potential well depth. Second virial coefficients calculated using the SIMPER method are more reliable than the experimental data, and cover a larger temperature range, so they will be useful for calculating the thermodynamic properties of humid gases in a number of applications, including humidity standards and combustion gases.

4. The water–oxygen interaction

The weak interaction between water and oxygen molecules is important in chemistry and biochemistry, in combustion processes, and in understanding the properties of humid air for metrology. The absorption of light by oxygen gas is affected by collisions between oxygen and water molecules, and this may have a significant effect on transmission of light through the Earth's atmosphere, influencing the climate [43, 44]. To calculate the second virial coefficient of the water–oxygen mixture, a potential energy surface is needed which covers all intermolecular orientations and a range of intermolecular distances from the low repulsive wall out to the long-range asymptotic limit, as was the case for the water–nitrogen interaction.

There appears to be no definitive experimental measurement of the equilibrium structure of the water–oxygen dimer, and early theoretical studies of the water–oxygen dimer [45] did not consider the planar, van der Waals bonded intermolecular geometry that is now believed to be the potential minimum [46–48]. This was first suggested in 2002 by Kjaergaard and co-workers [49], who used QCISD calculations. However, several different equilibrium geometries were proposed in the following three years, including an out-of-plane geometry [50] and hydrogen-bonded geometries [51, 52].

4.1. *Methodology*

Our calculations of the water–oxygen potential energy surface [53] considered several different theoretical methods employing both restricted (R) and unrestricted (U) electron spins: the supermolecule methods RMP2, UMP2, RCCSD, UCCSD, RCCSD(T), and the SIMPER method with unrestricted spins. The SIMPER method has not yet been implemented with restricted spins for open-shell molecules, and UCCSD(T) calculations proved rather troublesome, although some (unpublished) results were eventually obtained, with a triple-zeta basis set, which were in close agreement with the RCCSD(T) calculations.

The length of computer time required for CCSD(T) calculations (eight hours per point for the SPTZ basis set) made this method impractical for obtaining a complete potential energy surface, and the RMP2, RCCSD and UCCSD supermolecule methods were not in good agreement with the RCCSD(T) results, so the water–oxygen intermolecular potential energy surface was calculated using the supermolecule UMP2 method, and using SIMPER calculations based on the UMP2 results. In the unrestricted-spin calculations, the expectation value of the total spin quantum number S^2 was 2.048 ± 0.001 for the oxygen monomer and for the water–oxygen dimer, so spin contamination of the wave function was small.

The intramolecular water and oxygen geometries were fixed at their ground-state expectation values, and a set of 465 orientations was chosen for the calculation. Ten different water–oxygen distances R between $4.5\,a_0$ and $10\,a_0$ were used in the calculations, so the final potential energy surface was calculated at 4650 separate points. The calculations used the SPTZ and SPQZ basis sets described in section 2, with extrapolation to the complete basis set limit [54].

The SIMPER method was applied as described for water–nitrogen in section 3.1, with some small modifications. The first-order electrostatic interaction energy was recalculated in the usual way: the UMP2 electrostatic energy was removed from the supermolecule energy and replaced by the electrostatic interaction energy between monomer UCCSD charge densities. The UMP2 'exchange-repulsion' energy was defined in the same way as for water–nitrogen, as the difference between the UMP2 supermolecule energy and the RSPT perturbative energy to second order, and it was multiplied by the ratio of the UCCSD to UMP2 values of the electron density overlap integral S_ρ to give the SIMPER exchange-repulsion energy. The UMP2 second-order dispersion energy was expressed as a sum of damped multipole terms $C_n f_n(bR)R^{-n}$, using the UMP2 dispersion energy coefficients $n \geq 6$, multiplied by damping functions with the damping length scale parameter b calculated at each point on the surface from the UMP2 dispersion energy and the supermolecule UMP2 dispersion energy coefficients. To obtain the SIMPER dispersion energy, the UMP2 dispersion energy coefficients were replaced by time-dependent coupled-cluster (TD-UCCSD) dispersion energy coefficients [19], and the damping length scale parameter was multiplied by the ratio of the TD-UCCSD to supermolecule UMP2 values of the quantity $\sqrt{C_6/C_8}$ at the same orientation. It was shown in section 2 that the spherically averaged C_6 value from TD-UCCSD calculations (52.7 atomic units) is in close agreement with the best literature value, and improves on the supermolecule UMP2 value of C_6, which is the TD-UHF result in table 1 (55.1 atomic units).

Three methods for recalculating the second-order induction energy were investigated in preliminary calculations. First, the UMP2 induction energy was used without modification as the SIMPER induction energy. Second, the same method was used as for water–nitrogen: the CCSD density of water was used to calculate the potential for the oxygen molecule, and the second-order energy response of oxygen to that potential was calculated using UMP2 theory; and similarly for the polarization of water by oxygen. Third, UCCSD polarizabilities and multipoles of both molecules were used to calculate long-range C_n multipole coefficients for the second-order induction energy, and these were used to replace the analogous supermolecule UMP2 quantities in a damped multipole series, using the same damping functions as for the dispersion energy. In practice, the induction energy was found to be relatively small at the potential energy minimum, and all three methods for modifying the induction energy gave similar results. Therefore, in the calculation of the SIMPER potential energy surface, the UMP2 induction energy was not modified. The total SIMPER interaction energy was calculated as the sum of the UCCSD electrostatic energy, the UMP2 induction energy, and the SIMPER dispersion and exchange-repulsion energies.

$$E_{\text{int}}(\text{SIMPER}) = E_{\text{elec}}(\text{UCCSD}) + E_{\text{ind}}(\text{UMP2}) + E_{\text{disp}}(\text{SIMPER})$$
$$+ E_{\text{rep}}(\text{UMP2})\frac{S_\rho(\text{UCCSD})}{S_\rho(\text{UMP2})}. \tag{37}$$

4.2. *Results*

The equilibrium geometry of the water–oxygen dimer having lowest energy was found to be planar in both the UMP2 and SIMPER calculations. In the equilibrium 'O-bonded' structure, the O_2 molecule is located toward the oxygen end of the water molecule, and oriented roughly parallel to the nearest OH bond. A planar 'H-bonded' local minimum was also found. The symmetry of the water and oxygen molecules means that there are four equivalent versions of each equilibrium structure.

Table 3 shows the separate components of the energy (electrostatic, induction, dispersion and exchange-repulsion). In the O-bonded geometry, the electrostatic and dispersion energies both became less negative when the relevant MP2 quantities were replaced by their CCSD counterparts. In the H-bonded geometry, the dispersion energy became more negative, and this roughly cancelled the change in the electrostatic

Table 3. Components of the water–oxygen intermolecular potential (cm^{-1}) near the global minimum (O), and the hydrogen-bonded local minimum (H). Results were extrapolated to the complete basis set limit from SPTZ and SPQZ calculations.

Geom	Level	Electrostatic	Induction	Dispersion	Exch-Rep	Total
O	UMP2	−154	−49	−301	293	−211
O	SIMPER	−140	−49	−285	287	−187
H	UMP2	−87	−95	−263	317	−128
H	SIMPER	−73	−95	−277	293	−152

energy, but the largest effect was a reduction in exchange-repulsion energy, arising from the reduced overlap of the CCSD ground-state electron densities relative to the MP2 densities.

Supermolecule interaction energies were calculated for two geometries close to the 'O-bonded' and 'H-bonded' equilibrium structures, using different basis sets. As for water–nitrogen, the change in the potential energy produced by enlarging the basis set was found to be smaller than the difference between the supermolecule methods. The binding energy changed by about 4% from SPTZ to SPQZ, and basis set extrapolation increased the SPQZ binding energy by a further 2% or so. This contrasts with a difference of 29% at the O-bonded minimum between the most negative interaction energy ($-200 \, cm^{-1}$, UMP2), and the least negative energy ($-155 \, cm^{-1}$, UCCSD/RCCSD) found with the SPTZ basis set. The RCCSD(T) and SIMPER interaction energies were both less negative than the UMP2 energy: -192 and $-180 \, cm^{-1}$ respectively. For the H-bonded local minimum, there was an even larger percentage difference between the most negative energy ($-137 \, cm^{-1}$, SIMPER) and the least negative ($-86 \, cm^{-1}$, RCCSD), and the SIMPER binding energy was 11% more than the RCCSD(T) binding energy ($-123 \, cm^{-1}$).

To obtain a complete potential energy surface, the calculated points were fitted using a function of the O$\rightarrow$B and H$\rightarrow$B vectors, where O is the oxygen atom of water and B is the bond centre of the oxygen molecule. A contour diagram of the fitted surface is shown in figure 5.

The mixed second virial coefficient B_{12} for the water–oxygen mixture was calculated from the SIMPER potential energy surface extrapolated to the complete basis set limit. The calculation was essentially the same as for water–nitrogen, and included translational and rotational quantum effects to first order; higher-order quantum effects were expected to be negligible above $100 \, K$ [30]. In figure 6 these calculated second virial coefficients are compared with experimental measurements. Second virial coefficients obtained from the UMP2 potential energy surface, extrapolated to the complete basis set limit, are shown for comparison.

The experimental B_{12} data for water–oxygen shown on the plot come from the study of Wylie and Fisher [55], and these are the only such data for this mixture. The SIMPER values for $B_{12}(T)$ are systematically less negative than the experimental data, although the uncertainties in the SIMPER potential were estimated [53] to be large enough that the results cannot definitively be said to be inconsistent. On the other hand, the calculations agree well with the values obtained indirectly from water–air and water–nitrogen data by Hall and Iglesias-Silva [56]. This agreement is consistent with recent work by Harvey and Huang [57], which shows that combining the water–oxygen and water–nitrogen results reported in this review with previous calculations for water–argon [58] produces excellent agreement with available high-quality data for $B(T)$ for water with air.

The UMP2 potential energy surface produces somewhat less negative virial coefficients than SIMPER. Although the SIMPER virial coefficients are in better agreement with the experimental data, the small amount of data, and the difficulty of the experiments, mean that this is not definitive evidence in favour of the SIMPER method. Since the RCCSD(T) and SIMPER calculations differ significantly, especially near the H-bonded minimum, it was important to determine whether this overall
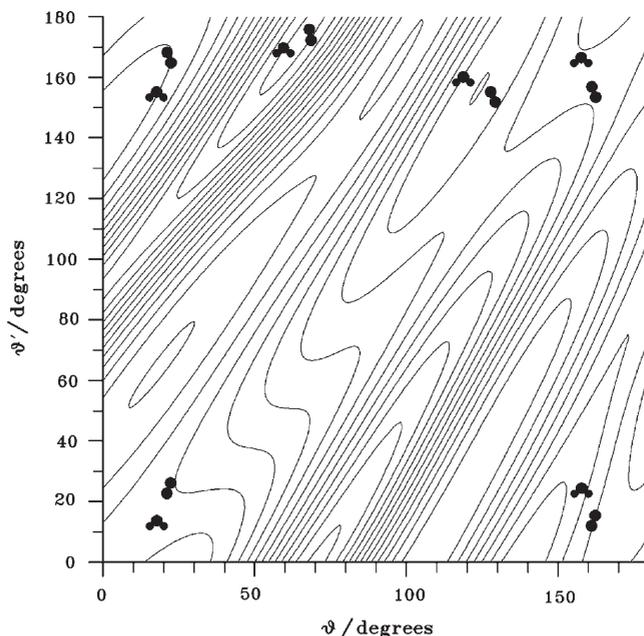
Figure 5. Contour plot of the fitted water–oxygen potential energy surface, calculating using SIMPER and extrapolated to the complete basis set limit. See the caption to for details. Contours are drawn every 10 cm^{-1}. The global 'O' minimum is near the top and to the left of centre; the local 'H' minimum is near the top and to the right of centre. The lowest-energy pathway between the two involves moving straight downwards on the figure from the 'H' minimum, over the saddle point (which is about 19 cm^{-1} above the H minimum in energy), to $\theta' \approx 80°$, then turning to descend the narrow valley (down and left), and finally jumping back to the top of the diagram near the global 'O' minimum.

deepening of the potential by the SIMPER procedure was realistic or just an artifact of the method. To this end, 17 further intermolecular orientations were chosen, and UMP2, RCCSD(T) and SIMPER calculations were performed at these orientations with the SPTZ basis set, at intermolecular separations $R = 5.5\,a_0$, $R = 6.5\,a_0$, and $R = 7\,a_0$. At $R = 5.5\,a_0$, the intermolecular potential averaged over all 17 points was +1245 (UMP2), +1224 (RCCSD(T)) and +1189 (SIMPER). At $R = 6.5\,a_0$, the average was −14.3 (UMP2), −30.7 (RCCSD(T)), and −37.0 (SIMPER), and at $R = 7\,a_0$ it was −83.6 (UMP2), −95.2 (RCCSD(T)) and −96.4 (SIMPER). The main difference between the SIMPER and RCCSD(T) calculations for these 17 orientations was found to be near to the H-bonded geometry. It is interesting to note that further (unpublished) calculations have shown that the difference between the spherically averaged UMP2 and SIMPER potentials arises almost entirely from the exchange-repulsion energy.

Based on the comparison with RCCSD(T) calculations, the deepening of the potential well by SIMPER appears to be qualitatively correct, and the deeper H-bonded minimum predicted by the SIMPER method, relative to RCCSD(T), is offset to some extent by the shallower O-bonded minimum. The spherically averaged RCCSD(T) potential is significantly deeper than the UMP2 potential. It is not as
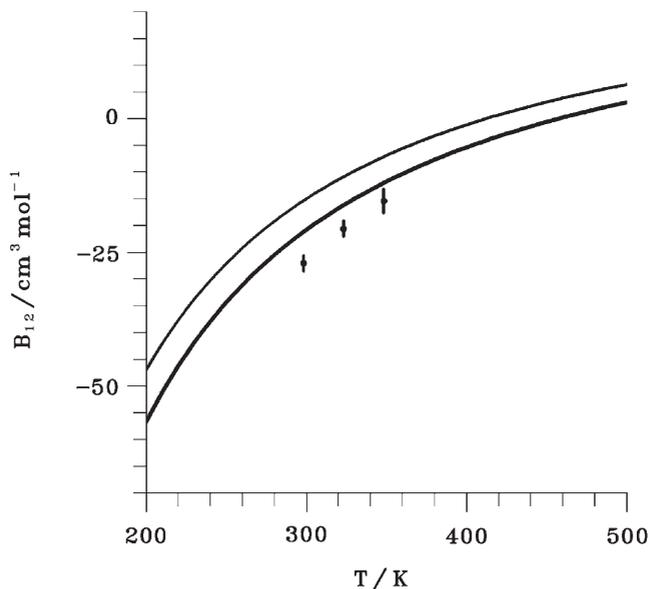
Figure 6. Calculated water–oxygen mixed second virial coefficients (upper curve, MP2, lower curve, SIMPER) and experimental data with quoted uncertainties in the virial coefficients (error bars). The calculations and experimental data sources are described in the text.

deep as the SIMPER potential, but it is closer to the SIMPER potential than it is to the UMP2 potential. It is noted that agreement with the limited experimental second virial coefficient data [55] on the water/oxygen mixture would require an even deeper potential than the SIMPER potential. This seems unlikely, especially in light of the experimental water–air data, but more calculations and/or measurements are required to be certain that the experimental water–oxygen second virial coefficients are too low.

In summary, the water–oxygen complex shows more 'van der Waals' character than the water–nitrogen complex: water–nitrogen is more strongly bound at the equilibrium geometry, and its potential energy surface is more anisotropic, as a result of the larger nitrogen quadrupole moment. Some water–nitrogen orientations are repulsive for all R, which is not the case for water–oxygen. Furthermore, enlarging the basis set for water–oxygen increases the binding energy for most of the geometries considered, including all the geometries with $R < 8$ a_0; this is also typical of a van der Waals complex.

Second virial coefficients are affected by a large region of the potential energy surface, so it is desirable to avoid methods that systematically under- or over-estimate the interaction energy for most geometries. Based on the RCCSD(T) method as a benchmark, SIMPER appears to suffer less from this problem than the supermolecule UMP2 method, which seems to be overall too shallow for water–oxygen, and the supermolecule CCSD methods, which significantly underestimate the binding energy.

The SIMPER method can be used for smaller basis sets, and no 'counterpoise' correction seems to be required. It was originally thought that this might be necessary because the supermolecule energy is calculated in the dimer basis set, whereas the RSPT

calculations and calculations of the dispersion energy coefficients are done in the mono-
mer basis sets. It is fortunate that this does not lead to a significant basis set error (such
as would be obtained by neglecting the counterpoise correction in a supermolecule
calculation), because calculating the dispersion energy coefficients in the dimer basis
set at every point would not be practicable. The SIMPER method also seems to be
reliable at short intermolecular separations, even though RSPT perturbation theory,
by itself, does not converge to a physically reasonable result. Scaling the 'exchange-
repulsion' energy in the SIMPER method also scales the higher-order RSPT terms
that are included in it, and this scaling is not justified, so the method can only
work if higher-order terms in the perturbation expansion (including their exchange
counterparts) are small, and/or the scaling factor is close to 1. This may not be true
for ion-molecule interactions or strong hydrogen bonds, for which third-order RSPT
may be needed.

5. Atomic polarizabilities and atomic dispersion energy coefficients

We now turn our attention to larger molecules, for which the SIMPER method,
as described in the first part of this review, cannot be applied. As the size of the
system increases, first the use of a damped intermolecular multipole series for the dis-
persion energy becomes unsuitable, then MP2 supermolecule calculations and RSPT
perturbative calculations become infeasible. For still larger molecules, accurate calcula-
tion of monomer properties using the CCSD method also becomes impossible. This
section considers the case of 'medium-sized' molecules, where MP2 supermolecule
calculations are still possible, but where the damped intermolecular multipole series
cannot be used.

There are two reasonable methods for calculating accurate dispersion energies for
interactions involving medium-sized molecules. One method involves calculating the
dispersion energy directly, without any multipole expansion, using a perturbative
method such as Symmetry-Adapted Perturbation Theory (SAPT) [59]. However,
direct calculation of the dispersion energy is computationally expensive, so the level
of electron correlation that can be used is lower than monomer TD-CCSD. The alter-
native is to expand the dispersion energy as a sum of atomic dispersion energies, using
damped multipole series between individual atoms a and b of the interacting molecules:

$$E_{\mathrm{disp}} = - \sum_{a,b} \sum_{n=6} C_{n,ab} f_{n,ab}(bR_{ab}) R_{ab}^{-n}. \qquad (38)$$

If a supermolecule MP2 calculation is possible, as we are assuming, then it must also be
possible to do the less time-consuming calculation of the second-order MP2 dispersion
energy E_{disp}. Then, if the MP2 atomic dispersion energy coefficients $C_{n,ab}$ can also be
calculated, the MP2 atomic damping functions f can be obtained in an exactly analo-
gous way to the current SIMPER method: the damping functions are assumed to
have a universal functional form, and the quantity $b\sqrt{C_8/C_6}$ is assumed to be the
same for all atomic pairs, at each point on the surface. If the atomic dispersion

energy coefficients can also be calculated using a higher level of electron correlation, these higher-level dispersion energy coefficients can be used in equation (38), with the same atomic damping functions, to give the SIMPER dispersion energy. The main requirements for the SIMPER dispersion energy are, therefore, atomic dispersion energy coefficients at MP2 supermolecule level and at a 'high' level of electron correlation.

Atomic dispersion energy coefficients can be calculated from equation (8) using the imaginary-frequency-dependent polarizabilities of the atoms in the interacting molecules. Atomic polarizabilities are not physically measurable or uniquely definable properties, unlike the total polarizabilities of the molecules, but because they have a crucial role to play in developing intermolecular potentials, a considerable amount of work has been devoted to finding methods for defining and calculating them [60–65]. Our recent work in this area is described here [66]. For simplicity, the rest of this section considers atomic polarizabilities at zero frequency, $\alpha^a_{l\kappa, l'\kappa'}(0) = \alpha^a_{l\kappa, l'\kappa'}$, but the methodology and discussion apply equally to frequency-dependent polarizabilities.

It is generally agreed that atomic polarizabilities should be symmetrical, $\alpha^a_{l\kappa, l'\kappa'} = \alpha^a_{l'\kappa', l\kappa}$, and that they should not include multipole components with $l = 0$ or $l' = 0$, which would violate charge conservation. Non-local polarizabilities $\alpha^{a, a'}_{00, 00}$, which include charge flow between different atoms under the influence of an external potential, can be defined and used in theory, but they give unwieldy expressions for the dispersion energy which suffer from cancellation of large terms of opposite sign, so they are not considered further here. If atomic polarizabilities are to be useful, combining all the atomic polarizabilities of a molecule should reproduce the molecular polarizabilities, and atomic polarizabilities should increase more slowly in magnitude with increasing l, l' than the molecular polarizabilities, so that the dispersion (or induction) energy converges more quickly when it is expressed in terms of atomic contributions. It is also desirable for atomic polarizabilities to be similar for atoms in similar chemical environments, and to be correctly related by symmetry for symmetry-equivalent atoms in a molecule. Physical intuition also suggests that the diagonal polarizabilities with $l\kappa = l'\kappa'$ should be positive, and that they should be larger for larger, 'softer' atoms.

5.1. *Methodology*

Our method [66] for calculating atomic polarizabilities α^a is based on the physical definition of polarizability as the change in multipole Q per unit multipole component of the external potential V,

$$\alpha^a_{l\kappa, l'\kappa'} = -\left(\frac{\partial Q^a_{l\kappa}}{\partial V^a_{l'\kappa'}}\right)_{V^a_{l'\kappa'}=0}. \tag{39}$$

Given a method for calculating atomic multipoles $Q^a_{l\kappa}$, the atomic polarizabilities can be calculated from this definition by placing the molecule in a series of external potentials, and calculating the change in the atomic multipoles per unit applied potential $V^a_{l'\kappa'}$ at the nucleus of atom a.

For example, consider the diatomic molecule CO, with a bond length of 2.132 a_0, and let the z axis coincide with the molecular axis. When a constant electric field is applied in the x direction, the atomic multipoles of the C and O atoms change. Using the Hartree–Fock method with the aug-cc-pVDZ basis set, the first-order response of the atomic x dipoles to the field is found using distributed multipole analysis (DMA) [67] to be −7.48 on C and −3.50 on O. The atomic polarizabilities are therefore $\alpha_{xx}^C = 7.48$ and $\alpha_{xx}^O = 3.50$.

However, when a constant electric field (V_{10} potential) is applied in the z direction, parallel to the bond, the first-order response involves changes in the DMA charges of the atoms as well as the dipoles. The charge of the C atom (which is located at negative z relative to the O atom) changes by 3.73 per unit applied field, and the charge of the O atom changes by −3.73. This change in charge cannot be represented using local atomic polarizabilities.

Assuming that local atomic polarizabilities exist, and that they can be interpreted as the first-order change in atomic multipoles with respect to a local applied potential, there is only one way to resolve this problem: the DMA method used to obtain the atomic multipoles must be changed. This is done in a way which involves no change in the field-free DMA, and by making the minimum changes necessary to the DMA in the applied field to make the atomic polarizabilities local and symmetrical.

The response of the DMA atomic multipoles of the CO molecule to the field in the z direction is therefore modified such that no change in atomic charges occurs. This is done in a symmetrical fashion. Half of the 'unwanted' charge of 3.73 on the C atom is re-expanded as a multipole series on the O atom, giving an additional charge of 1.87, an additional z dipole of −3.98, and additional higher multipoles on O, and reducing the charge on C by 1.87. Half of the 'unwanted' charge of −3.73 on the O atom is similarly re-expanded about the C nucleus. The resulting change in charge is now zero at each nucleus, as required, and the total change in dipole of the C atom is the sum of the change in the atomic dipole per unit applied V_{10} potential, namely −3.16, plus −3.98 obtained from re-expanding the oxygen charge, so the total atomic polarizability is $\alpha_{zz}^C = 7.14$. Similarly, $\alpha_{zz}^O = 7.38$.

The dipole–quadrupole polarizabilities of the atoms are also obtained from this calculation. The quadrupole Q_{20} of the C atom changes by 6.63 per unit applied V_{10} potential, and re-expanding the oxygen charge at the C atom adds a further −8.48, so the dipole–quadrupole polarizability is $\alpha_{10,20}^C = \alpha_{20,10}^C = 1.85$. Note that the atomic polarizability is required to be symmetrical. A similar calculation for the oxygen atom gives $\alpha_{10,20}^O = \alpha_{20,10}^O = -9.84$.

Next, a constant field gradient (V_{20} potential) is applied to the CO molecule. If the origin is taken to be midway between the nuclei, this corresponds to applying the same V_{20} potential to the two nuclei, plus a potential component $V_{10}^C = -2.132V_{20}$ at the C nucleus, and $V_{10}^O = 2.132V_{20}$ at the O nucleus. The change in charge of the two nuclei should be zero in response to the finite field, and the change in dipole of the C atom should be given by $-\alpha_{10,20}^C V_{20}^C - \alpha_{10,10}^C V_{10}^C$, that is, 13.36 per unit applied field gradient, given the previously calculated values of the atomic polarizabilities of C. In practice, the change in the charge of C is not zero, but −6.94 per unit applied potential, so half of this is re-expanded about the O nucleus, and half of the corresponding O charge polarization is re-expanded about C. The resulting change in

dipole of the C atom is 9.83 (from the applied field gradient), plus 14.79 (from moving the oxygen charge). This is 11.26 more than the required value of 13.36, and the change in dipole of the O atom is similarly 11.26 less than its required value. These 'unwanted' dipoles are therefore re-expanded about the other nuclei: a dipole of 5.63 is removed from C, producing an additional dipole and quadrupole on O, and a dipole of -5.63 is removed from O, producing an additional dipole and quadrupole on C. The total change in quadrupole on the C atom, resulting from the applied field gradient, plus the result of re-expanding the unwanted O charge, plus the result of re-expanding the unwanted O dipole, is -45.95 atomic units. Of this, $+3.95$ is the result of polarization by the field $V_{10}^C = -2.132 V_{20}$ at the C nucleus, using the previously calculated value of $\alpha_{10,20}^C$, so the atomic quadrupole polarizability of C must be $\alpha_{20,20}^C = 49.90$. The atomic quadrupole polarizability of O is similarly found to be $\alpha_{20,20}^O = 49.11$.

This method is continued to arbitrarily high multipole rank in the applied potential and the multipole response, and all the corresponding atomic polarizabilities are obtained. For larger molecules, and molecules with lower symmetry, two additional points need to be considered. The first is that polarizabilities of the same multipole rank obtained from the calculation are not necessarily symmetrical, for example, $\alpha_{x,z}^a \neq \alpha_{z,x}^a$. The solution to this is to force the values of $\alpha_{x,z}^a$ and $\alpha_{z,x}^a$ to equal the mean of the calculated values. Polarization of the atom in the x direction, in a field applied in the z direction, is then adjusted to agree with this mean value, by moving the 'unwanted' part of the x dipole to other atoms, and similarly with $x \leftrightarrow z$.

The second important consideration is that, in a molecule with more than two atoms, the 'unwanted' multipoles on one atom can be divided between the other atoms in more than one unique way. To choose a suitable method, we attempt to optimize the convergence of the multipole series by re-expressing multipoles over atoms as near to the original atom as possible. This is related to (but not the same as) the localization procedure introduced by Le Sueur and Stone [68], which was recently used by Misquitta and Stone [64] to localize atomic polarizabilities obtained from a constrained electron density fitting method.

For covalent molecules, where bonds between atoms are well defined, the 'unwanted' multipoles of an atom can be moved to other atoms bonded to it. If the molecule is 1-connected, that is, if there is only one way to reach any atom from any other by traversing bonds without backtracking, the amount of unwanted multipole to move across each bond in this way is unique. If the molecule contains rings, or is not connected (for example, a hydrogen-bonded dimer, or a transition state in a chemical reaction) then a more sophisticated method for redistributing the 'unwanted' multipoles is needed. A method has been introduced [66] for molecules with rings by analogy with the flow or diffusion of a substance towards equilibrium. A set of first-order 'rate equations' is used, in which the rate of transfer of a multipole across a bond is proportional to the difference in unwanted multipoles between the two atoms. At large 'times', the amount of unwanted multipole is reduced to zero for all the atoms. The total amount of multipole moved across each bond is then obtained by integrating the solution of the rate equations over time.

If the bonds in a molecule are not well-defined, or, more generally, to produce a first-principles method which does not rely on intuitive concepts such as chemical bonds,

then the diffusion analogy can be extended to allow for different 'rates of flow' between different pairs of atoms. Atoms that are closer together, and/or that share a larger amount of charge density, have a larger rate of 'flow' of multipole between them. This proposal, although apparently reasonable, has not yet been tested in practice, so it is not considered further in this review.

5.2. Results

The atomic polarizability method was tested using coupled Hartree–Fock (CHF) calculations with an aug-cc-pVDZ basis set, and the systems chosen for study were carbon dioxide, water, ethylene, acetylene, and the alkane series (C_nH_{2n+2}, $n = 1-6$). Two sets of atomic polarizabilities were calculated for the hydrocarbons: all-atom polarizabilities, where each atom constitutes a separate region, and united-atom polarizabilities, where each region is composed of a CH_n group and is centred at the carbon nucleus.

The results for the isotropic and anisotropic atomic polarizabilities of CO, CO_2, H_2O, C_2H_2, and C_2H_4 are presented in table 4. They are defined by

$$\bar{\alpha}_1 = \frac{1}{3}\sum_\kappa \alpha_{1\kappa,1\kappa}$$

$$\Delta\alpha_1 = \left[\frac{1}{2}((\alpha_{z,z}-\alpha_{x,x})^2 + (\alpha_{z,z}-\alpha_{y,y})^2 + (\alpha_{x,x}-\alpha_{y,y})^2) + 3\left(\alpha_{z,x}^2 + \alpha_{z,y}^2 + \alpha_{x,y}^2\right)\right]^{1/2}$$

$$\bar{\alpha}_2 = \frac{1}{5}\sum_\kappa \alpha_{2\kappa,2\kappa}. \tag{40}$$

Table 4. Atomic, united-atom and molecular polarizabilities, in atomic units.

	$\bar{\alpha}_1$	$\Delta\alpha_1$	$\bar{\alpha}_2$
CO			
C	7.37	−0.34	37.75
O	4.79	3.88	19.00
CO	12.15	3.54	86.17
CO_2			
C	3.97	3.48	13.11
O	5.53	4.21	15.66
CO_2	15.02	11.90	118.06
H_2O			
O	5.69	0.96	17.67
H	1.14	1.79	1.78
H_2O	7.97	1.34	27.43
C_2H_2			
C	9.46	4.07	47.80
H	1.80	2.22	12.08
CH	11.26	6.28	64.31
C_2H_2	22.53	12.57	170.57
C_2H_4			
C	7.54	6.10	42.68
H	3.10	4.09	16.78
CH_2	13.74	6.66	98.72
C_4H_4	27.48	13.31	269.93

Since $\Delta\alpha_1$ equals $|\alpha_{z,z} - \alpha_{x,x}|$ for the linear molecules CO, CO_2 and C_2H_2, the more informative (signed) quantity $\alpha_{z,z} - \alpha_{x,x}$ is given in table 4 for these molecules; it is positive in all cases except for the C atom of CO.

The results are in reasonable agreement with physical intuition. The diagonal polarizabilities are positive, the hydrogen atom has the smallest polarizability, and the carbon atom usually has the largest polarizability, but in CO_2, where the carbon atom transfers some of its valence electron density to the electronegative oxygen atoms, its polarizability is lower. It is encouraging that even for these small molecules, the ratio $\bar{\alpha}_2/\bar{\alpha}_1$ is smaller for the individual atoms than it is for the molecules containing them, which suggests that the convergence properties of the atomic multipole series will be better than they are for the molecular series.

Very promising results were also obtained for the alkane series. For the largest molecule considered, $n\text{-}C_6H_{14}$, the atomic dipole polarizabilities varied from 3.75 to 5.23 for C and from 3.05 to 3.34 for H. These values are in reasonable agreement with the atomic dipole polarizabilities obtained by Ferraro *et al.* [60] and by Williams and Stone [69]. The largest quadrupole polarizability for any atom was 23.37 atomic units, compared to 3537 atomic units for the entire molecule. However, the quadrupole polarizabilities of two carbon atoms (related by symmetry, at positions 2 and 5 in the chain) were negative.

In order to test the usefulness of the atomic polarizabilities in practical applications, point to point induction energies were calculated using the atomic polarizabilities and compared with CHF calculations. The point to point induction energy is defined as the second-order change in energy when two non-polarizable point charges are located near the molecule. The *ab initio* CHF induction energies were computed using 12 500 randomly selected points from the van der Waals envelope for each molecule, and the deviations between these and the corresponding energies obtained from atomic polarizabilities were calculated for pairs of these points. The maximum multipole rank used in the polarizabilities was varied from 1 to 3, to assess the convergence of the multipole expansion. The results are presented in tables 5 and 6. The 'Atomic'

Table 5.   Comparison of induction energies calculated using the CHF method and using atomic polarizabilities. See the text for details.

	Atomic			United Atom			Molecular		
	$l=1$	$l=2$	$l=3$	$l=1$	$l=2$	$l=3$	$l=1$	$l=2$	$l=3$
H_2O									
rms %	0.21	0.07	0.02	–	–	–	0.46	0.07	0.03
max %	−5.84	−2.36	1.23	–	–	–	−13.94	−2.78	1.22
CO_2									
rms %	0.23	0.14	0.02	–	–	–	0.76	0.45	0.07
max %	−5.36	−2.30	0.82	–	–	–	−14.09	−7.59	−1.77
C_2H_2									
rms %	0.29	0.12	0.02	0.32	0.11	0.02	0.50	0.20	0.05
max %	−7.13	3.40	2.42	−10.55	−5.47	−1.53	−16.14	−8.49	−3.42
C_2H_4									
rms %	0.39	0.12	0.07	0.51	0.14	0.09	0.68	0.27	0.07
max %	11.13	3.06	−2.25	−17.98	−6.41	−3.82	−21.14	−8.91	−3.69

results use polarizabilities on every nucleus, 'United Atom' results do not have polarizabilities on H atoms, and the 'Molecular' results use a single set of polarizabilities located at the centre of the molecule. The root mean square (RMS) error and the maximum error are expressed as a percentage of the total range of energies.

The results indicate that the atomic and united-atom polarizabilities perform well, especially as the size of the molecule increases. The atomic polarizabilities generally perform better than the the united-atom polarizabilities, although they appear to be converging to the same result as the maximum angular momentum l is increased. When atomic or united-atom polarizabilities up to $l = 2$ are used, the RMS difference from the CHF results is only a few hundredths of a percent. The improved convergence of the atomic polarizability models is particularly noticeable for the larger alkanes, where the molecular convergence is very slow, or non-existent, as l is increased.

From these results, it is concluded that using atomic polarizabilities up to a maximum rank of $l = 2$ is sufficient to reproduce the *ab initio* second order induction energies closely. While using the atomic polarizabilities is more accurate than the united-atom models, the added simplicity of the united-atom model is appealing, and the united-atom polarizabilities seem to be more transferable.

6. Towards intermolecular potentials for larger molecules

If two interacting molecules are too large for supermolecule or second-order RSPT calculations of their interaction energy to be performed, then the quantities that can be

Table 6. Comparison of induction energies calculated using the CHF method and using atomic polarizabilities. See the text for details.

	Atomic			United Atom			Molecular		
	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$	$l = 1$	$l = 2$	$l = 3$
CH_4									
rms %	0.23	0.15	0.04	–	–	–	0.59	0.14	0.08
max %	6.50	4.15	4.04	–	–	–	−20.25	−6.68	−2.63
C_2H_6									
rms %	0.17	0.12	0.04	0.58	0.16	0.09	0.89	0.41	0.10
max %	7.14	3.47	4.41	−22.78	−8.06	−4.67	−25.22	−11.73	−4.77
C_3H_8									
rms %	0.18	0.09	0.05	0.46	0.12	0.07	0.91	0.45	0.17
max %	5.95	2.91	5.49	−16.81	−6.01	−3.27	−25.11	−12.86	9.10
C_4H_{10}									
rms %	0.16	0.10	0.05	0.44	0.12	0.07	1.24	0.72	0.31
max %	7.32	2.33	3.28	−16.47	−5.97	−3.50	−29.82	−18.36	−10.32
C_5H_{12}									
rms %	0.25	0.12	0.07	0.49	0.14	0.08	1.77	1.18	0.71
max %	10.09	-3.35	4.90	−18.13	−6.26	−4.13	−37.36	37.11	37.37
C_6H_{14}									
rms %	0.23	0.12	0.07	0.46	0.14	0.08	2.15	1.70	1.33
max %	6.06	−2.69	5.80	−17.47	−7.38	4.31	−50.75	−62.33	−100.36

Table 7. Errors produced by extrapolating the separate components of the HCl dimer energy to the basis set limit. Mean errors and RMS errors are expressed in microHartree.

		DZ-TZ	DZ-SPTZ	TZ-QZ	SPTZ-SPQZ
MP2	Mean	20.0	3.7	3.3	14.1
	RMS	50.2	48.3	17.9	30.3
Electrostatic	Mean	15.4	−37.8	−3.8	31.5
	RMS	47.9	102.6	45.4	79.7
Dispersion	Mean	153.5	51.4	42.6	−27.3
	RMS	258.3	139.6	68.9	38.1
Induction	Mean	301.7	28.8	117.6	−172.8
	RMS	784.5	246.2	284.8	334.5
Exch-Rep	Mean	−450.6	−38.6	−153.1	182.7
	RMS	1045.5	283.2	313.6	379.2
Damped Induction	Mean	29.7	21.5	0.3	−7.4
	RMS	57.1	49.5	3.2	15.1
Modified Exch-Rep	Mean	−178.6	−31.3	−35.8	17.3
	RMS	328.9	118.5	62.9	86.1

calculated in equation (2) and equation (37) are the first-order electrostatic energy, the charge density overlap, the dispersion energy coefficients and the polarizabilities. The quantities that cannot be calculated are the MP2 exchange-repulsion energy (and the parameter K that is obtained from it), the second-order dispersion energy (and the dispersion damping length scale parameter b that is obtained from it), and the second-order induction energy.

As a first attempt to find a computationally cheaper method of calculating these quantities, we recently investigated the possibility of calculating the 'cheap' components of the SIMPER energy, such as the first-order electrostatic energy, using a large basis set, and the 'expensive' components, such as the parameter K, using a smaller basis set [70]. Five different systems were studied, and typical results are shown in tables 7 and 8. The complete basis set (CBS) limit was estimated using a two-point extrapolation based on the aug-cc-pVQZ and aug-cc-pV5Z basis sets, and compared with two-point extrapolation using smaller basis sets. The aug-cc-pVnZ basis set is abbreviated as nZ in the tables. The rows labelled 'MP2' show the mean and RMS difference between the MP2 supermolecule energy extrapolated using the basis sets at the top of the columns, and the estimated CBS limit. Similarly, rows labelled 'Electrostatic', 'Dispersion', 'Induction' and 'Exch-Rep' show the difference between extrapolating these components of the energy using the basis sets shown, and their estimated values in the CBS limit. The sign of the 'Mean' errors is defined as (extrapolated result)−(CBS limit), and the averages were calculated over a number of attractive and repulsive points on the potential energy surface.

These initial results were not encouraging. They show that constructing a potential energy surface from different components calculated using small basis sets, and then extrapolating these components to the CBS limit, can produce substantial errors. The errors obtained by extrapolating the components of the energy are larger than the

Table 8. Errors produced by extrapolating the separate components of the Ar-N_2 energy to the basis set limit. Mean errors and RMS errors are expressed in microHartree.

		DZ-TZ	DZ-SPTZ	TZ-QZ	SPTZ-SPQZ
MP2	Mean	52.4	25.1	16.1	17.6
	RMS	75.2	61.7	19.0	23.0
Electrostatic	Mean	19.6	13.2	−1.0	−12.4
	RMS	33.9	31.0	21.5	31.9
Dispersion	Mean	91.4	−5.0	36.2	−32.5
	RMS	125.7	18.8	47.8	39.4
Induction	Mean	42.2	22.7	12.2	−84.5
	RMS	89.4	40.0	20.5	205.0
Exch-Rep	Mean	−100.8	−5.8	−31.3	147.0
	RMS	144.4	35.5	40.2	286.1
Damped Induction	Mean	1.5	−0.1	0.7	−0.5
	RMS	2.2	0.6	1.1	0.6
Modified Exch-Rep	Mean	−60.1	17.0	−19.9	62.9
	RMS	70.9	58.9	24.8	88.2

errors obtained by extrapolating the total supermolecule energy. This suggests that there is a substantial cancellation of errors when the components are added together.

Extrapolating the induction energy gave worse results than extrapolating the electrostatic and dispersion energies, for most of the systems considered. It was suggested [70] that the reason for this is that the induction energy includes a 'non-physical' negative part, which describes charge transfer from one molecule to an occupied orbital on the other molecule. This does not occur in the supermolecule calculation, because it violates the Pauli Principle, but it can occur in RSPT calculations, where the Pauli Principle is not enforced for intermolecular electron exchange. This non-physical part of the induction energy converges very slowly with increasing basis set size, so attempted extrapolation to the CBS limit gives an inaccurate result.

Since the induction energy extrapolates poorly, the exchange-repulsion energy suffers from the same problem, because the induction energy is used in calculating the exchange-repulsion energy. The 'non-physical' negative part of the induction energy produces a corresponding positive contribution to the exchange-repulsion energy; this positive contribution is part of the exchange-induction energy in SAPT [59].

The simplest solution to the problem of extrapolating the induction energy therefore seems to be to calculate the induction and exchange-induction energies, add them together, and extrapolate their sum, which is free from non-physical contributions. However, calculating these components of the energy is computationally expensive. It takes a similar amount of effort to calculating the MP2 supermolecule energy itself, and if the MP2 supermolecule energy is available then it can be extrapolated stably, without the need to consider its separate components.

Our solution was to write the induction energy, like the dispersion energy, as a damped multipole series in the form of equation (2) and equation (3). Ideally, the multipole series should involve atomic multipole terms, especially for large molecules, but

our test calculations only involved only small molecules so a multipole expansion about the centres of the molecules was used. Unlike the dispersion energy, the damping length scale parameter b in the induction energy was not fixed by requiring the sum of the damped multipole series to equal the calculated induction energy. Instead, b was simply set equal to the value of b used for the dispersion energy. This produces a 'damped induction energy', which is not equal to the RSPT induction energy except at long range, and is much quicker to calculate. We also expect the damped induction energy to be less affected by non-physical charge transfer than the RSPT induction energy, since charge transfer is a short-range effect which decreases exponentially with increasing separation, and should therefore not appear in the multipole series.

In tables 7 and 8, the errors obtained by extrapolating this damped induction energy are shown in the rows labelled 'Damped Induction'. The results show that the errors are greatly reduced, and that they are now smaller than the errors obtained by extrapolating any of the other components of the energy. When the induction energy is redefined in this manner, the exchange-repulsion energy also changes, because it is obtained by subtracting the electrostatic, dispersion and induction energies from the supermolecule energy. The errors obtained by extrapolating this modified exchange-repulsion energy are shown in the tables as 'Modified Exch-Rep', and they are smaller than the errors obtained from the unmodified exchange-repulsion energy. They are larger than the errors produced by extrapolating the damped induction energy, but this is not unexpected, because the exchange-repulsion energy is larger than the induction energy.

Since the multipole coefficients C_n in the damped induction energy series can be calculated quickly, the remaining 'expensive' components of the potential energy surface are the parameters K and b. As previously mentioned, obtaining the parameter K requires expensive supermolecule calculations. The damping functions, defined by the length scaling parameter b, require second-order RSPT calculations, which are nearly as expensive as the supermolecule calculations. We therefore tested [70] whether the parameters K and b could be predicted using calculations with smaller basis sets. The variation of these parameters with basis set size was not smooth, so no attempt was made to extrapolate them; the values calculated for smaller basis sets were simply used directly in the expression for the energy. To test the basis set dependence of b, the dispersion and damped induction energies were calculated in the CBS limit, and expressed as damped multipole series, with CBS multipole coefficients C_n, to obtain the CBS value of b. This was replaced by a value of b calculated from a smaller basis set, which changed the induction plus dispersion energy calculated from the series by a small amount. The RMS errors produced for the HCl dimer were 73 microHartree (when b was calculated from the TZ basis set) and 16 microHartree (QZ basis set), and for Ar-N_2 the RMS errors were 8 microHartree (TZ) and 5 microHartree (QZ). Using the value of b obtained from these basis sets therefore produced smaller errors, for the same basis set size, than the errors produced using basis set extrapolation (shown in the tables). However, the value of K was found to change more significantly with basis set. The exchange-repulsion energy was estimated using the value of K calculated from a small basis set, multiplied by the CBS charge density overlap integral, and the result did not compare well with the CBS exchange-repulsion energy. The errors

were larger than the errors produced by basis set extrapolation, and this was true for both the 'modified' and unmodified exchange energies.

It will also be useful to investigate whether the CBS extrapolation of the dispersion energy can be improved. As discussed for the exchange-induction and induction energies, there is an exchange-dispersion energy which cancels some of the dispersion energy at short range, although the cancellation is usually much less than it is for the induction energy. It is reasonable to consider treating the dispersion and exchange-dispersion energies together as a single term. This would be done by modifying the dispersion damping length scale parameter b, so that the sum of the damped multipole series equals the sum of the dispersion and exchange-dispersion energies. Work along these lines is in progress.

Finally, we note that for even larger molecules, the electrostatic energy and charge density overlap become too expensive to evaluate directly, because the number of intermolecular two-electron integrals becomes too large. This can be solved relatively straightforwardly, using charge density fitting methods [71] to reduce the number of integrals. Highly correlated *ab initio* calculations of polarizabilities and dispersion energy coefficients also become impossible when the molecules are large, and it is necessary either to use computationally cheaper methods, such as time-dependent Kohn-Sham theory [72], or to assume that atomic polarizabilities are transferable between chemically similar atoms.

7. Concluding remarks

The calculation of force fields for molecular simulations using mainly or entirely *ab initio* methods is an exciting prospect. Some progress has been made recently along these lines [73–75], but significant advances are needed and are likely to be made in the near future. For example, the next generation of force fields should rely less on the assumption of transferability, and more on separate calculations for each molecule of interest. While the complexity of some molecules (e.g., proteins or DNA) makes an assumption of transferability inevitable, more sophisticated transferability schemes should be devised, such as making the atomic C_6 dispersion energy coefficients decrease as the atoms become more positively charged.

Since the early days of modelling intermolecular potentials, the calculation of long-range multipole coefficients has played an important role, and this should also be true in the future as models for the short-range damping and exchange-repulsion energy become more soundly based. The development of computationally inexpensive density functional theory for calculating molecular properties will be very helpful in constructing intermolecular potential energy surfaces. The choice of an appropriate functional is important for calculating molecular electron densities and polarizabilities, and high-accuracy calculations on smaller molecules will serve as an invaluable benchmark for these more approximate methods. The use of density functional theory to calculate short-range components of the intermolecular potential has increased greatly in recent years [65, 76–79], and this trend will no doubt continue.

As computer power increases, as well as simply increasing the size of the molecules that can be studied, it will become possible to tackle other important

problems in molecular simulation. These will include chemical reactions in solution [80], intramolecular flexibility [81], and the development of models for non-pairwise-additive potentials [82, 83], which could be incorporated into effective pair potentials.

Acknowledgements

References

[1] R. J. Wheatley, A. S. Tulegenov, and E. N. Bichoutskaia, Int. Rev. Phys. Chem. **23**, 151 (2004)
[2] K. T. Tang and J. P. Toennies, J. Chem. Phys. **80**, 3726 (1984).
[3] R. R. Fuchs, F. R. W. McCourt, A. J. Thakkar, and F. Grein, J. Phys. Chem. **88**, 2036 (1984).
[4] R. J. Wheatley and W. J. Meath, Mol. Phys. **80**, 25 (1993).
[5] M. Hodges, E. N. Bichoutskaia, A. S. Tulegenov, and . J. Wheatley, Int. J. Quant. Chem. **96**, 537 (2004).
[6] A. K. Dham, A. R. Allnatt, A. Koide, and W. J. Meath, Chem. Phys. **196**, 81 (1995).
[7] K. T. Tang and J. P. Toennies, J. Chem. Phys. **118**, 4976 (2003).
[8] H. Wei, R. J. Le Roy, R. Wheatley, and W. J. Meath, J. Chem. Phys. **122** (2005).
[9] A. S. Tulegenov, R. J. Wheatley, M. P. Hodges, and A. H. Harvey, J. Chem. Phys. **126**, 094305 (2007).
[10] A. J. Stone, *The Theory of Intermolecular Forces*, Clarendon Press, Oxford, 2002.
[11] S. L. Price, A. J. Stone, and M. Alderton, Mol. Phys. **52**, 987 (1984).
[12] P. W. Langhoff, Chem. Phys. Lett. **20**, 33 (1973).
[13] H. J. Kolker and H. H. Michels, J. Chem. Phys. **43**, 1027 (1965).
[14] J. N. Murrell and G. Shaw, J. Chem. Phys. **49**, 4731 (1968).
[15] H. Koch *et al.*, J. Chem. Phys. **92**, 4924 (1990).
[16] O. Christiansen, P. Jorgensen, and C. Hattig, Int. J. Quant. Chem. **68**, 1 (1998)
[17] P. W. Langhoff, S. T. Epstein, and M. Karplus, Rev. Mod. Phys. **44**, 602 (1972).
[18] C. Hattig and B. A. Hess, Chem. Phys. Lett. **233**, 359 (1995).
[19] R. J. Wheatley, J. Comput. Chem. (2007), accepted for publication.
[20] T. B. Pedersen and H. Koch, J. Chem. Phys. **108**, 5194 (1998).
[21] E. N. Bichoutskaia, M. P. Hodges, and R. J. Wheatley, J. Comput. Meth. Sci. Eng. **2**, 391 (2002).
[22] T. H. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[23] D. J. Margoliash and W. J. Meath, J. Chem. Phys. **68**, 1426 (1978).
[24] H. O. Leung, M. D. Marshall, R. D. Suenram, and F. J. Lovas, J. Chem. Phys. **90**, 700 (1989).
[25] P. Sandler, J. oh Jung, M. M. Szczesniak, and V. Buch, J. Chem. Phys. **101**, 1378 (1994).
[26] W. H. Adams, J. Math. Chem. **10**, 1 (1992).
[27] R. J. Wheatley and S. L. Price, Mol. Phys. **69**, 507 (1990).
[28] P. Soderhjelm, G. Karlstrom, and U. Ryde, J. Chem. Phys. **124**, 244101 (2006).
[29] K. L. Bak, P. Jörgensen, J. Olsen, T. Helgaker, and W. Klopper, J. Chem. Phys. **112**, 9229 (2000).
[30] M. P. Hodges, R. J. Wheatley, G. K. Schenter, and A. H. Harvey, J. Chem. Phys. **120**, 710 (2004).
[31] N. E. Kosyakov, B. I. Ivchenko, and P. P. Krishtopa, Vopr. Khim. i Khim. Tekhnol. **68**, 33 (1982).
[32] N. E. Kosyakov, B. I. Ivchenko, and P. P. Krishtopa, J. Appl. Chem. USSR **50**, 2436 (1977).
[33] M. B. Iomtev, V. G. Piskunov, and V. L. Kadzhayev, *Tables of the National Standard Reference Data Service of the USSR R88-84*, All-Union Research Institute of Technical Information, Archival Deposit No. 220, Moscow, 1984, cited in V. A. Rabinovich and V. G. Beketov, *Moist Gases. Thermodynamic Properties*. (Begell House, New York, 1995).
[34] L. R. Oellrich and K. Althaus, GERG – Water Correlation: Relationship Between Water Content and Water Dew Point Keeping in Consideration the Gas Composition in the Field of Natural Gas, in *GERG Technical Monograph 14, Fortschr.-Ber. VDI, Series 3, No. 679*, VDI Verlag, Düsseldorf, 2001.
[35] S. T. Blanco, I. Velasco, and S. Otín, Phys. Chem. Liq. **40**, 167 (2002).
[36] M. Rigby and J. M. Prausnitz, J. Phys. Chem. **72**, 330 (1968).
[37] P. C. Gillespie and G. M. Wilson, GPA Research Report, Technical Report RR-41, Gas Processors Association, Tulsa, OK, 1980.

[38] O. A. Podmurnaya, O. Gudkov, and N. Dubovikov, Russ. J. Phys. Chem. **78**, 300 (2004).
[39] A. H. Mohammadi, A. Chapoy, B. Tohidi, and D. Richon, J. Chem. Eng. Data **50**, 541 (2005).
[40] A. V. Plyasunov and E. L. Shock, J. Chem. Eng. Data **48**, 808 (2003).
[41] A. H. Harvey and E. W. Lemmon, J. Phys. Chem. Ref. Data **33**, 369 (2004).
[42] R. Span, E. W. Lemmon, R. T. Jacobsen, W. Wagner, and A. Yokozeki, J. Phys. Chem. Ref. Data **29**, 1361 (2000).
[43] R. H. Tipping, A. Brown, Q. Ma, and C. Boulet, J. Mol. Spectrosc. **209**, 88 (2001).
[44] H. G. Kjaergaard *et al.*, J. Phys. Chem. A **107**, 10680 (2003).
[45] I. M. Svishchev and R. J. Boyd, J. Phys. Chem. A **102**, 7294 (1998).
[46] T. W. Robinson and H. G. Kjaergaard, J. Chem. Phys. **119**, 3717 (2003).
[47] A. Sabu, S. Kondo, N. Miura, and K. Hashimoto, Chem. Phys. Lett. **391**, 101 (2004).
[48] A. Sabu, S. Kondo, R. Saito, Y. Kasai, and K. Hashimoto, J. Phys. Chem. A **109**, 1836 (2005).
[49] H. G. Kjaergaard, G. R. Low, T. W. Robinson, and D. L. Howard, J. Phys. Chem. A **106**, 8955 (2002).
[50] J. A. G. Gomes, J. L. Gossage, H. Balu, M. Kesmeg, F. Bowen, R. S. Lumpkin, and D. L. Cocke, Spectrochim. Acta **61**, 3082 (2005).
[51] A. J. Bell and T. G. Wright, Phys. Chem. Chem. Phys. **6**, 4385 (2004).
[52] D. M. Upadhyay and P. C. Mishra, J. Molec. Struct. (Theochem) **624**, 201 (2003).
[53] R. J. Wheatley and A. H. Harvey, J. Chem. Phys. (2007), accepted for publication.
[54] J. M. L. Martin, Chem. Phys. Lett. **259**, 669 (1996).
[55] R. G. Wylie and R. S. Fisher, J. Chem. Eng. Data **41**, 175 (1996).
[56] K. R. Hall and G. A. Iglesias-Silva, J. Chem. Eng. Data **39**, 873 (1994).
[57] A. H. Harvey and P. H. Huang, Int. J. Thermophys. (2007), in preparation.
[58] M. P. Hodges, R. J. Wheatley, and A. H. Harvey, J. Chem. Phys. **117**, 7169 (2002).
[59] B. Jeziorski, R. Moszynski, and K. Szalewicz, Chem. Rev. **94**, 1887 (1994).
[60] M. B. Ferraro, M. C. Caputo, and P. Lazzeretti, J. Chem. Phys. **109**, 2987 (1998).
[61] J. R. Maple and C. S. Ewig, J. Chem. Phys. **115**, 4981 (2001).
[62] F. Dehez, C. Chipot, C. Millot, and J. G. Angyan, Chem. Phys. Lett. **338**, 180 (2001).
[63] G. J. Williams and A. J. Stone, J. Chem. Phys. **119**, 4620 (2003).
[64] A. J. Misquitta and A. J. Stone, J. Chem. Phys. **124**, 024111, (2006).
[65] A. J. Stone and A. J. Misquitta, Int. Rev. Phys. Chem. **26**, 193 (2007).
[66] T. C. Lillestolen and R. J. Wheatley, J. Phys. Chem. A (2007), submitted for publication.
[67] A. J. Stone, Chem. Phys. Lett. **83**, 233 (1981).
[68] C. R. Le Sueur and A. J. Stone, Mol. Phys. **83**, 293 (1994).
[69] G. J. Williams and A. J. Stone, Mol. Phys. **102**, 985 (2004).
[70] T. C. Lillestolen and R. J. Wheatley, Phys. Chem. Chem. Phys. (2007), in preparation.
[71] G. A. Cisneros, J. P. Piquemal, and T. A. Darden, J. Chem. Phys. **123**, 044109, (2005).
[72] L. J. Bartolotti, J. Chem. Phys. **80**, 5687 (1984).
[73] G. Tabacchi, C. J. Mundy, J. Hutter, and M. Parrinello, J. Chem. Phys. **117**, 1416 (2002).
[74] G. A. Kaminski, H. A. Stern, B. J. Berne, and R. A. Friesner, J. Phys. Chem. A **108**, 621 (2004).
[75] J. P. Piquemal, G. A. Cisneros, P. Reinhardt, N. Gresh, and T. A. Darden, J. Chem. Phys. **124**, (2006).
[76] A. Hesselmann and G. Jansen, Chem. Phys. Lett. **367**, 778 (2003).
[77] A. Hesselmann, G. Jansen, and M. Schutz, J. Chem. Phys. **122**, 014103, (2005).
[78] A. J. Misquitta and K. Szalewicz, J. Chem. Phys. **122**, 214109, (2005).
[79] A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **123**, 214103, (2005).
[80] S. T. Arroyo, J. A. S. Martin, and A. H. Garcia, J. Phys. Chem. A **111**, 339 (2007).
[81] G. Murdachaew and K. Szalewicz, Faraday Discussions **118**, 121 (2001).
[82] E. M. Mas, R. Bukowski, and K. Szalewicz, J. Chem. Phys. **118**, 4386 (2003).
[83] R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van der Avoird, Science **315**, 1249 (2007).