JCTC Journal of Chemical Theory and Computation

Accurate Induction Energies for Small Organic Molecules: 1. Theory

Alston J. Misquitta[†,‡] and Anthony J. Stone*[,†]

*University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K., and
University College London, 20 Gordon Street, London WC1H 0AJ, U.K.*

Abstract: The induction energy often plays a very important role in determining the structure and properties of clusters of organic molecules, but only in recent years has an effort been made to include this energy in such calculations, notably in the field of organic crystal structure prediction. In this paper and the following one in this issue we provide ab initio methods suitable for the accurate inclusion of the induction energy for molecules containing as many as 30 atoms or so. These techniques are based on Symmetry-Adapted Perturbation Theory using Density Functional Theory [SAPT(DFT)] and use distributed polarizabilities computed using the recently developed density-fitting algorithm with constrained refinement. With this approach we are able to obtain induction models of varying complexity and study the effects of overlap and related numerical issues. Basis set effects on the exact and asymptotic induction energies are investigated, and the roles of higher-order induction energies and many-body effects are explored.

I. Introduction

The induction energy plays an important role in determining the structures of clusters of polar molecules. The cooperative nature of the induction means that, for polar molecules, induction effects dominate the many-body contributions to the interaction energy. These many-body effects can be very important in determining the structures of clusters of molecules. For example, three- and four-body effects have been shown to be responsible for the tetrahedral structure of liquid water.[1,2] However, this interaction energy component is often neglected or treated incorrectly. This is not only because it is hard to calculate accurately but also because important aspects of the induction energy are still poorly understood.

There are many features that make the induction energy hard to handle. First, it is not pair-additive. The induction energy of a particular molecule, $-(1/2)\alpha F^2$ in its simplest form, depends on the square of the total electric field F due to its neighbors, and the fields of different neighbors may interfere constructively or destructively.[3] Second, it is cooperative: the charge distribution of each molecule is polarized by the electric field of its neighbors, and it is the modified charge distribution of each molecule that is the source of the field that polarizes the others. This cooperative behavior is important in clusters and condensed phases of polar molecules and favors the hydrogen-bonded networks that are seen for example in water. Since each molecule is polarized by all the others, it is necessary to solve coupled equations for the modified charge distributions. This can usually be carried out by a simple iterative procedure, but it is a time-consuming additional step in a simulation.

The most general way to calculate the induction[4] uses the frequency-dependent density susceptibility,[5] or FDDS, $\alpha(\mathbf{r},\mathbf{r}'|\omega)$, which describes the change in charge density at $\mathbf{r}$ due to a delta-function change in electrostatic potential at $\mathbf{r}'$ oscillating at frequency ω. To describe induction we need only the static FDDS, $\alpha(\mathbf{r},\mathbf{r}'|0)$, and this can be calculated efficiently and accurately by modern methods. The resulting description is however much too cumbersome for practical use; it would be necessary to solve a set of coupled integral equations for the changes to the electron density of each molecule and then to carry out integrals over each molecule to determine its induction energy. Consequently we need to

* Corresponding author phone: +44 1223 336375; fax: +44 1223 336362; e-mail: ajs1@cam.ac.uk.
† University Chemical Laboratory.
‡ University College London.

extract a description in terms of polarizabilities. For all but the smallest of molecules, single-site descriptions, using the overall molecular polarizabilities, are inadequate, and the polarizability needs to be described in a distributed way, as the charge distribution does.

This does not however solve all the problems. At short range, when the molecular charge densities overlap, the distributed-polarizability description is subject to penetration error, just like the distributed-multipole description of the electrostatic interaction. Using the FDDS we could describe the penetration effects correctly, but unfortunately that is usually impracticable.

This is by no means the end of the story. The polarizability, even in the accurate form of the FDDS, describes the *linear* response of the molecule to external fields and gives the induction energy only to second order in perturbation theory.[4] Moreover, as usually formulated, it ignores effects arising from electron exchange between molecules at short range. For a more complete description, we need to include higher-order terms in the perturbation series and to include the effects of exchange.[6,7]

In fact a large part of the higher-order effects for clusters can be recovered if the polarization problem is solved self-consistently,[3] by the iterative procedure mentioned above, but this still treats the response of each individual molecule as linear in the field. Higher-order effects, described by hyperpolarizabilities, are not included in such a treatment and become increasingly significant at short distances. As we shall see in this paper, these neglected effects can result in a significant error in the dimer energy and geometry.

As for the exchange effects, it has become clear from recent calculations using Symmetry-Adapted Perturbation Theory (SAPT) that they make a very significant contribution to the total induction energy.[7,8]

In this paper and the following one in this issue[1] we attempt to obtain a practical procedure for calculating accurate induction energies for assemblies of molecules in clusters or in the condensed phase. Here we discuss the theoretical issues associated with accurate calculations of the nonexpanded and expanded induction energies. The numerical issues associated with basis sets and model building will be discussed in part 2.

II. General Overview

For the interaction between two molecules, the second-order induction energy can be accurately computed using symmetry-adapted perturbation theory (SAPT)[7] or the more recent version of SAPT based on a density-functional treatment of the monomers (SAPT(DFT)[9−12] or the very similar and independently derived DFT-SAPT[13−16]). In the following we refer to SAPT(DFT) for brevity, but it should be understood that the DFT-SAPT method is essentially the same. The superior computational scalings and accuracies of the SAPT(DFT) expressions make this theory the method of choice, particularly for organic molecules, for which the SAPT expressions are usually too expensive computationally to be evaluated. The SAPT(DFT) expression for the second-order induction involves the frequency-dependent density susceptibility (FDDS) at zero frequency,[11] which can be

evaluated quite efficiently using coupled Kohn−Sham (CKS) theory (also known as Kohn−Sham linear response theory). This expression does not include exchange effects; the second-order exchange-induction energy cannot be written in terms of the FDDS and is calculated using scaling rules[11,12] which have been demonstrated to result in rather accurate energies.

For polar molecules with large polarizabilities, the higher-order induction and exchange-induction energies (in which we include terms of third order and above) can make significant contributions to the two-body interaction energy. (We include in the category of polar molecules any molecule that gives rise to large electric fields in its neighborhood, whether or not it has a significant dipole moment. The interaction between such molecules is dominated by the electrostatic energy.) These higher-order effects are strongest in hydrogen-bonded complexes, where they can account for as much as $10−15\%$ of the total two-body interaction energy. In most SAPT and SAPT(DFT) calculations of the intermolecular energy the higher-order energies have been estimated using the so-called $\delta_{\text{int,resp}}^{\text{HF}}$ term, defined as the difference between the supermolecular Hartree−Fock interaction energy of the dimer and certain low-order SAPT energy terms.[7,17,18] This procedure has been shown to result in interaction energy potentials of high accuracy for hydrogen-bonded complexes like the water dimer,[2,19] but recent evidence seems to suggest that the $\delta_{\text{int,resp}}^{\text{HF}}$ term may not be suitable for non-hydrogen-bonded complexes.[20] We will return to this issue below. In any case, the $\delta_{\text{int,resp}}^{\text{HF}}$ term is cumbersome to calculate as it requires a supermolecular Hartree−Fock calculation in the dimer basis (so as to avoid the basis-set superposition error) and a low-order SAPT calculation, in addition to the SAPT(DFT) calculation. Consequently an alternative means of estimating the higher-order contributions to the interaction energy is needed.

Recently, the third-order SAPT interaction-energy components have been derived and implemented, though without the inclusion of intramonomer correlation effects.[20] It has often been assumed that the third-order energies would account for most of the higher-order contributions to the interaction energy. However, Patkowski et al.[20] have demonstrated that while this is true for non-hydrogen-bonded complexes, the third-order terms may account for less than half of the higher-order energies for hydrogen-bonded complexes. It is believed that this is the case for two reasons: first, the higher-order energies are dominated by induction and exchange-induction effects,[20] and second, the induction series is known to be divergent due to the presence of Coulomb singularities in the interaction operator.[21]

In spite of the problems associated with the higher-order interaction energy components, we shall demonstrate that the third-order induction and exchange-induction energies can form a good and computationally convenient approximation to the true higher-order energies if evaluated within the SAPT(KS) theory,[9,11] that is, using Kohn−Sham orbital energies and eigenvalues. This procedure has many advantages: (1) The higher-order energies for non-hydrogen-bonded complexes are recovered very accurately. (2) While there will be non-negligible errors made in hydrogen-bonded

Induction Energies for Small Organic Molecules: 1

J. Chem. Theory Comput., Vol. 4, No. 1, 2008 **9**

geometries, these will be much smaller than the errors incurred if the higher-order corrections were ignored altogether. (3) The interaction energies are obtained in one calculation. (4) And finally, the third-order induction and exchange-induction energies are the least computationally demanding of the third-order energy components and so do not add significantly to the overall computational cost of the SAPT(DFT) method.

The many-body contribution to the interaction energy can be very important for polar clusters, in which many-body effects can account for as much as 15−30% of the interaction energy.[22,23] This is not surprising, as the induction energy, which is very important for such systems, is strongly nonadditive in nature.[3] The dominant contribution to the many-body energy arises from the three-body energy which can be computed using the three-body formulation of SAPT (see ref 7 for a review) or using supermolecular techniques,[22,24] but the computational expense is so large as to make these methods applicable to systems of a few atoms only. Fortunately, for polar systems, which include most organic molecules, the many-body contributions are dominated by many-body induction effects which can be well approximated using damped classical polarizable models if accurate molecular polarizabilities and multipole moments are known.

This paper is organized as follows: In section IV.1 we outline the theoretical details of the SAPT(DFT) expression for the second-order induction energy. In subsection IV.1.2 we explore and assess ways of including the higher-order induction and exchange-induction energies using several examples. In section IV.2 we briefly describe ways of including the many-body contributions to the induction energy using the damped classical polarizable model. The damped classical polarizable model is also used to calculate the asymptotic induction energies. This needs the molecular polarizabilities in distributed form. In section IV.3 we describe the distribution method based on the constrained density-fitting procedure and present a method for optimizing the resulting distributed polarizabilities. In section V we conclude with a summary of the main results of this paper.

III. Notation

If electron exchange between molecules is ignored, which is a good approximation at large separations, the interaction energy can be obtained using standard perturbation theory. This is conventionally described as the polarization approximation, though this is an unsatisfactory terminology, particularly in the context of the induction energy. The contribution that is usually called the induction energy, and denoted E_{ind}, appears at second order, but we denote it here as $E_{ind,pol}^{(2)}$ to distinguish it explicitly from the exchange-induction $E_{ind,exch}^{(2)}$. $E_{ind,pol}^{(2)}$ is the term defined as the induction and denoted E_{ind}, in SAPT and SAPT(DFT).[12] However there are also induction energy contributions at higher orders, $E_{ind,pol}^{(n)}$. The second-order term $E_{ind,pol}^{(2)}$ can be expressed in a 'nonexpanded' form which remains valid at any intermolecular separation, however small, but it is conventionally expressed as a power series in $1/R$. This power-series form is often referred to as the classical model, and we denote

the damped version of this model by $E_{ind,d-class}^{(2)}$. Higher-order contributions can be expanded in the same way. We should have $E_{ind,pol}^{(2)} \sim E_{ind,d-class}^{(2)}$ for medium to large separations, but it turns out that the basis-set converged value of $E_{ind,pol}^{(2)}$ can be an order of magnitude larger (i.e., more negative) than $E_{ind,d-class}^{(2)}$ at equilibrium geometry and even larger at shorter distances. It has recently been shown[21] that $E_{ind,pol}^{(2)}$ is too large in magnitude because of the Coulomb singularities in the interaction operator. These singularities are absent in the expanded form of the operator, so it is not surprising that $E_{ind,d-class}^{(2)}$ is much smaller in magnitude than $E_{ind,pol}^{(2)}$. The Coulomb singularities also result in a very large exchange-induction energy, $E_{ind,exch}^{(2)}$, which is positive and significantly quenches $E_{ind,pol}^{(2)}$. Therefore neither energy is meaningful on its own.

Because of these complications, we believe that rather than referring to the conventional definitions[12] of the induction and exchange-induction energies separately, it is much better to define the induction energy as the sum. That is, the *n*th order induction energy is

$$E_{ind,tot}^{(n)} = E_{ind,pol}^{(n)} + E_{ind,exch}^{(n)} \qquad (1)$$

Since $E_{ind,exch}^{(n)}$ decays exponentially with increasing R, $E_{ind,tot}^{(n)}$ and $E_{ind,pol}^{(n)}$ both tend to $E_{ind,d-class}^{(n)}$ asymptotically. However, as will be demonstrated in sections V and VI of part 2, $E_{ind,tot}^{(n)}$ agrees far better with $E_{ind,d-class}^{(n)}$ at all distances than $E_{ind,pol}^{(n)}$ does.

The total interaction energy including terms up to order n is denoted by $U^{(n)}$ rather than the more conventional $E_{int}^{(n)}$. This has been done so as to avoid possible confusion arising from the similarity of the subscripts 'ind' and 'int'.

IV. Theory

IV.1. Induction Contributions to the Two-Body Energy.
In the two-body energy, the induction contributes to terms of second and higher orders in the interaction operator. The second-order induction is the most important, constituting between 85% and 96% of the total two-body induction energy.

IV.1.1. At Second Order: $E_{ind,pol}^{(2)}$ and $E_{ind,exch}^{(2)}$. From the polarization expansion,[6,7] $E_{ind,pol}^{(2)}$ for molecule X is

$$E_{ind,pol}^{(2)}(X) = \sum_{r \neq 0} \frac{|\langle \Phi_0^X | \hat{V} | \Phi_r^X \rangle|^2}{E_0^X - E_r^X} \qquad (2)$$

where Φ_r^X and E_r^X are the eigenstates and energy eigenvalues of the monomer Hamiltonian H_X, and $\hat{V}$ is the perturbation due to the electrostatic potential arising from the rest of the system. $E_{ind,pol}^{(2)}(X)$ can be interpreted as the second-order response of monomer X to the static field $\hat{V}$. One can show that eq 2 can be rewritten in terms of the frequency-dependent density susceptibility (FDDS) $\alpha_X(\mathbf{r}, \mathbf{r}'|\omega)$ evaluated at zero frequency[4]

$$E_{ind,pol}^{(2)}(X) = -\frac{1}{2} \int \int \alpha_X(\mathbf{r}, \mathbf{r}'|0) V(\mathbf{r}) V(\mathbf{r}') d\mathbf{r} d\mathbf{r}' \qquad (3)$$

where

$$\alpha_X(\mathbf{r}, \mathbf{r'}|\omega) =$$

$$2\sum_{r\neq0}\frac{E_r^X - E_0^X}{(E_r^X - E_0^X)^2 - \omega^2}\langle\Phi_0^X|\hat{\rho}_X(\mathbf{r})|\Phi_r^X\rangle\langle\Phi_r^X|\hat{\rho}_X(\mathbf{r'})|\Phi_0^X\rangle \quad (4)$$

and $V = \int\rho_Y^{tot}(\mathbf{r'})/|\mathbf{r} - \mathbf{r'}|d\mathbf{r'}$ is the electrostatic potential of the rest of the system. In the equation above, $\hat{\rho}_X(\mathbf{r}) = -\sum_{i\in X}\delta(\mathbf{r} - \mathbf{r}_i)$ is the electron density operator. The FDDS describes the linear response of the electron density to a frequency-dependent perturbation.

To calculate $E_{ind,pol}^{(2)}$ for an interacting pair of molecules within SAPT(DFT), the electrostatic potentials of the unperturbed monomers are evaluated using the Kohn–Sham orbitals, and the FDDS is evaluated using coupled Kohn–Sham theory (CKS) [25-27] (also known as linear-response DFT). In CKS theory, the FDDS takes the form

$$\alpha(\mathbf{r},\mathbf{r'}|\omega) = \sum_{iv,i'v'}C_{iv,i'v'}(\omega)\phi_i(\mathbf{r})\phi_v(\mathbf{r})\phi_{i'}(\mathbf{r'})\phi_{v'}(\mathbf{r'}) \quad (5)$$

where the subscripts i and i' (v and v') denote occupied (virtual) molecular orbitals, and ϕ_i is a molecular orbital. In CKS theory (and coupled Hartree–Fock theory (CHF)) the coefficients $C_{iv,i'v'}(\omega)$ can be written as[26]

$$C_{iv,i'v'}(\omega) = 4[(\mathbf{H}^{(2)}\mathbf{H}^{(1)} - \hbar^2\omega^2\mathbf{I})^{-1}\mathbf{H}^{(2)}]_{iv,i'v'} \quad (6)$$

where $\mathbf{I}$ is the unit matrix, and the $\mathbf{H}^{(1)}$ and $\mathbf{H}^{(2)}$ matrices, called the electric and magnetic Hessians, respectively, are defined in the CKS theory (in the adiabatic approximation[25,26]) as follows

$$\mathbf{H}_{iv,i'v'}^{(1)} = (e_v - e_i)\delta_{iv,i'v'} + 4(iv|i'v') - c_x[(ii'|vv') +$$
$$(iv'|i'v)] + 4\int\phi_i\phi_v\phi_{i'}\phi_{v'}\frac{\delta(v_{xc} - c_x v_x)}{\delta\rho}d^3\mathbf{r} \quad (7)$$

and

$$\mathbf{H}_{iv,i'v'}^{(2)} = (e_v - e_i)\delta_{iv,i'v'} - c_x[(ii'|vv') - (iv'|i'v)], \quad (8)$$

where e_i is the Kohn–Sham energy eigenvalue of molecular orbital ϕ_i, c_x is the fraction of the Hartree–Fock exchange included in the exchange-correlation (XC) functional ($c_x = 0$ for a nonhybrid functional), v_x is the exchange part of v_{xc}, and the two-electron coulomb integral $(ij|kl) = \langle\phi_i(1)\phi_j(1)|1/r_{12}|\phi_k(2)\phi_l(2)\rangle$. CKS theory is, in principle, exact if the exact exchange-correlation functional is used. In practice, only approximations are known, and from extensive numerical experiments it has been concluded[11,12,28] that the asymptotically corrected[29,30] PBE0[31] exchange-correlation functional[12,32] is the most suitable for accurate interaction energies. For large molecules however, it is too expensive computationally to evaluate the last integral in eq 7 using v_{xc} and v_x from the PBE functional. A more practical approach is to use the exchange-only LDA functional in the last term of eq 7. This approximation results in a small (less than 1%) loss in accuracy which is more than compensated by an order of magnitude reduction in computational expense.[12]

In order to implement the asymptotic correction, accurate vertical ionization potentials (IPs) are needed for the monomers. When they are not available experimentally, good estimates may be obtained from the difference between the energies of the N and $N - 1$ electron systems. The PBE0 functional is best suited for this calculation too as tests on atoms, diatoms, and small organic molecules have shown that it gives IPs with mean errors centered about 0.0 au with a standard deviation of only 0.007 au.[33]

As with the dispersion energy,[10,12] density-fitting techniques can be used to make the evaluation of the $E_{ind,pol}^{(2)}$ more efficient. Using density-fitting, the molecular orbital products $\phi_i(\mathbf{r})\phi_v(\mathbf{r})$ that appear in eq 5 can be expanded as

$$\phi_i(\mathbf{r})\phi_v(\mathbf{r}) = \sum_p D_{iv,p}\chi_p(\mathbf{r}) \quad (9)$$

where $\{\chi\}$ is an auxiliary basis set, and the coefficients $D_{iv,p}$ are determined by least-squares. The density-fitted FDDS takes the form

$$\alpha(\mathbf{r},\mathbf{r'}|\omega) = \sum_{p,q}\tilde{C}_{pq}(\omega)\chi_p(\mathbf{r})\chi_q(\mathbf{r'}) \quad (10)$$

where the $\tilde{C}_{pq}(\omega)$ are the transformed coefficients given by $\tilde{C}_{pq}(\omega) = \sum_{iv,i'v'}D_{iv,p}C_{iv,i'v'}(\omega)D_{i'v',q}$. Within Kohn–Sham theory, the total charge density ρ_Y^{tot} of closed-shell systems is given by

$$\rho_Y^{tot}(\mathbf{r}) = \sum_\beta Z_\beta\delta(\mathbf{r} - \mathbf{R}_\beta) - 2\sum_j|\phi_j(\mathbf{r})|^2 \quad (11)$$

where Z_β and $\mathbf{R}_\beta$ are the nuclear charge and position, and j labels the occupied orbitals. Consequently the electrostatic potential V of the rest of the system can be written as

$$V(\mathbf{r}) = \sum_\beta\frac{Z_\beta}{|\mathbf{r} - \mathbf{R}_\beta|} - 2\sum_j\int\frac{|\phi_j(\mathbf{r'})|^2}{|\mathbf{r} - \mathbf{r'}|}d\mathbf{r'} \quad (12)$$

where the sums run over the nuclei and orbitals of the rest of the system, which in this case is just monomer Y. Using the density-fitted FDDS and the above expression for V in eq 3 we obtain the density-fitted form of $E_{ind,pol}^{(2)}$ for monomer X

$$E_{ind,pol}^{(2)}(X) = -\frac{1}{2}\sum_{pq}M_p^Y\tilde{C}_{pq}^X(0)M_q^Y \quad (13)$$

where M_p^Y is defined as

$$M_p^Y = L_p^Y - 2\sum_{q'}\sum_j J_{pq'}D_{jj,q'}^Y \quad (14)$$

and we have used the definitions

$$J_{pq'} = \int\int\frac{\chi_p(\mathbf{r})\chi_{q'}(\mathbf{r'})}{|\mathbf{r} - \mathbf{r'}|}d\mathbf{r}d\mathbf{r'} \quad (15)$$

and

$$L_p^Y = \sum_\beta\int\frac{Z_\beta}{|\mathbf{r} - \mathbf{R}_\beta|}\chi_p(\mathbf{r})d\mathbf{r} \quad (16)$$

Induction Energies for Small Organic Molecules: 1

J. Chem. Theory Comput., Vol. 4, No. 1, 2008 **11**

The evaluation of $E_{\text{ind,pol}}^{(2)}$ using eq 13 involves a computational cost that scales as $O(m^2)$, where m is the size of the auxiliary basis. This is smaller than the $O(n_o^2 n_v^2)$ without density-fitting, but, much more importantly, the computationally demanding 4-center 2-electron integral transformation is avoided in the evaluation of eq 13, having been replaced by the 2-center 2-electron integrals $J_{pq'}$.

Futhermore, when monomer basis sets are used, only J_{pq} and L_p^Y need to be recomputed for each dimer geometry. This can result in a considerable savings in computational effort as only $O(m^2)$ operations are needed to evaluate the induction energy at each dimer geometry. However, see section VI in part 2 for some of the numerical issues associated with using such basis sets.

In part 2 we will need to evaluate the second-order induction energy of a molecule and a point charge. For such a system, the electrostatic potential is $V = Q/|\mathbf{r} - \mathbf{R}_Q|$ where $\mathbf{R}_Q$ is the location of the point charge and Q its value. The resulting induction energy is simply

$$E_{\text{ind,pol}}^{(2)} = -\frac{1}{2}\sum_{pq}L_p^Q \tilde{C}_{pq}(0)L_q^Q \tag{17}$$

where $L_p^Q = \int[Q/|\mathbf{r} - \mathbf{R}_Q|]\chi_p(\mathbf{r})d\mathbf{r}$.

The exchange-induction energy at second order, $E_{\text{ind,exch}}^{(2)}$, quenches $E_{\text{ind,pol}}^{(2)}$ significantly. $E_{\text{ind,exch}}^{(2)}$ cannot be expressed in terms of the FDDS and electron densities of the monomers, so it is estimated from the SAPT(KS) energies (denoted by 'KS') using a scaling relation[12]

$$E_{\text{ind,exch}}^{(2)} \approx E_{\text{ind,exch}}^{(2)}(\text{KS}) \times \frac{E_{\text{ind,pol}}^{(2)}}{E_{\text{ind,pol}}^{(2)}(\text{KS})} \tag{18}$$

The large quenching of $E_{\text{ind,pol}}^{(2)}$ by $E_{\text{ind,exch}}^{(2)}$ is believed to be an effect of excessive electron tunneling due to the electron−nuclear Coulomb singularities in the interaction operator.[21] These singularities are also responsible for the divergence of the perturbation theory. It has been shown that a convergent perturbation theory can be built using a regularized form of the interaction operator, that is, one in which the singularities arising from the electron−nuclear terms are removed.[21] It is also possible that for a regularized version of SAPT(DFT), the exchange-induction terms would not be needed. Preliminary evidence from our group suggests that this may well be the case.[34]

In summary, then, we recommend that the second-order induction energy be expressed according to eq 1, that is, as

$$E_{\text{ind,tot}}^{(2)} = E_{\text{ind,pol}}^{(2)} + E_{\text{ind,exch}}^{(2)} \tag{19}$$

with $E_{\text{ind,pol}}^{(2)}$ calculated by coupled Kohn−Sham theory as described above, and $E_{\text{ind,exch}}^{(2)}$ given by eq 18.

IV.1.2. Higher-Order Two-Body Energies. Contributions to the two-body interaction energy from terms beyond the second order in perturbation theory, denoted by $U^{(3-\infty)}$, are often large and cannot be neglected. For polar molecules like water, induction energies dominate these higher-order terms, which can constitute as much as 15% of the equilibrium binding energy of the dimer. On the other hand, for nonpolar molecules, the higher-order energies contribute only about 3−5% of the total interaction energy and could even be ignored if high accuracies are not needed.

$U^{(3-\infty)}$ is often approximated by the $\delta_{\text{int,resp}}^{\text{HF}}$ correction, defined as[17,18]

$$\delta_{\text{int,resp}}^{\text{HF}} = U^{\text{HF}} - (E_{\text{elst}}^{(10)} + E_{\text{exch}}^{(10)} + E_{\text{ind,tot,resp}}^{(20)}) \tag{20}$$

where $E_{\text{elst}}^{(10)}$, $E_{\text{exch}}^{(10)}$, and $E_{\text{ind,tot,resp}}^{(20)}$ are the SAPT corrections with no intramonomer correlation effects included (The subscript 'resp' indicates that the induction energy is calculated with response effects included.[7]), and U^{HF} is the supermolecule Hartree−Fock interaction energy computed with the counterpoise correction. Bear in mind that $E_{\text{ind,tot,resp}}^{(20)}$ has been defined using eq 1 and includes the exchange-induction contribution. The δ^{HF} term approximates the third- and higher-order contributions to the interaction energy but must be defined within SAPT as it has no counterpart in SAPT(DFT). Unfortunately, this term is too cumbersome to calculate on a routine basis, as its evaluation using eq 20 involves a supermolecule Hartree−Fock calculation in the dimer basis in addition to a low-order SAPT calculation. Furthermore, there is evidence that it may be a poor approximation to the higher-order energies for nonpolar systems[20] (but see the discussion in section V).

Another means of approximating $U^{(3-\infty)}$ is through the damped classical polarizable model. The derivation of the induction energy in a classical polarizable model is given in ref 3. Here we reproduce the final expressions in their general form. The damped classical induction energy of molecule A in a cluster is

$$E_{\text{ind,d−class}}(A) = \frac{1}{2}\sum_{a \in A}\sum_{B \neq A}\sum_{b \in B}\sum_{tu}\Delta Q_t^a f_{(tu)}(\beta R_{ab})T_{tu}^{ab}Q_u^b \tag{21}$$

where Q_t^a is the multipole moment operator for moment t at site a, and T_{tu}^{ab} is the interaction tensor[3] which describes the interaction between a multipole Q_u at b and a multipole Q_t at a. $f_{(tu)}(\beta R_{ab})$ is a damping function, which is conventionally assumed, in the absence of evidence to the contrary, to depend only on the distance R_{ab} between sites a and b and not on their relative orientation; the parameter β specifies the strength of the damping and may depend on the nature of the sites. ΔQ_t^a is the change in multipole moment t at a due to the self-consistent polarization of site a in the field of all sites on other molecules and is given by

$$\Delta Q_t^a = -\sum_{a' \in A}\sum_{B \neq A}\sum_{b \in B}\sum_{t'v}\alpha_{tt'}^{aa'}f_{(t'v)}(\beta R_{a'b})T_{t'v}^{a'b}(Q_v^b + \Delta Q_v^b) \tag{22}$$

where $\alpha_{tt'}^{aa'}$ is the distributed polarizability for sites (a, a') which describes the response of the multipole moment component Q_t^a at site a to the t'-component of the field at site a'. Notice that eq 22 must be solved iteratively for all molecules in the system, as the ΔQ occur on both sides of the equation. In the case of a dimer the sum over B just includes the other member of the dimer. If ΔQ is omitted from the right-hand side, we recover the damped classical approximation of the second-order induction energy, $E_{\text{ind,d−class}}^{(2)}$. At the mth iteration, we additionally obtain the
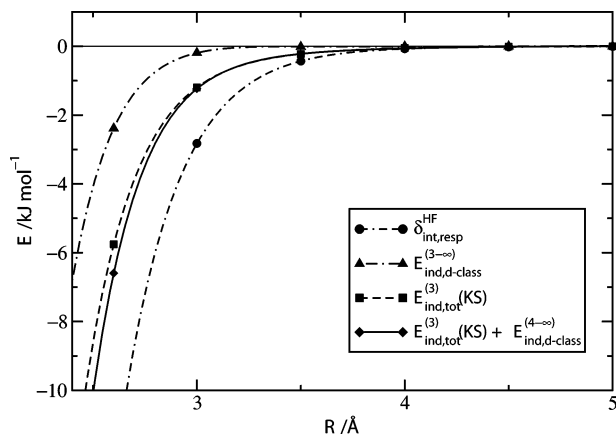
Figure 1. Approximations to the higher-order induction and exchange-induction energies for the water dimer. The relative orientations of the water molecules are fixed at their minimum geometry,[35] and the center-of-mass separation is varied. The minimum is located close to $R = 3.0$ Å. The classical models used a damping coefficient $\beta = 1.93$ au (see section V in part 2).

damped classical approximation to $E_{ind,tot}^{(m+2)}$. Call this approximation $E_{ind,d-class}^{(m+2)}$. The quantum expressions for the higher-order induction energies involve, besides the FDDS, which is a linear response function, the quadratic and higher-order response functions. The classical polarization model completely neglects these higher-order response functions. Furthermore, orbital overlap effects are neglected by the classical model, and these effects become increasingly important with increasing order in perturbation theory. The incorporation of the damping function attempts to correct for this neglect, but little is known about the form that it should take. Therefore, even though the damped classical polarizable model works reasonably well for the second-order induction energy, it is not reliable for the higher-order two-body induction effects.

The failure of the classical polarization models to recover the higher-order energies is clearly illustrated in Figure 1 for the water dimer. For this system, comparisons with interaction energies calculated using CCSD(T) suggest that that $\delta_{int,resp}^{HF}$ is a good estimate of the higher-order energies. The damped classical estimate of these energies, $E_{ind,d-class}^{(3-\infty)}$, is clearly inadequate, being an order of magnitude smaller than the $\delta_{int,resp}^{HF}$ correction for energetically relevant center-of-mass separations.

An alternative to the above methods is to approximate $U^{(3-\infty)}$ by the third-order induction energy and include the missing terms of fourth and higher orders using the damped classical polarizable model, that is

$$U^{(3-\infty)} \approx E_{ind,tot}^{(3)} + E_{ind,d-class}^{(4-\infty)} \qquad (23)$$

Here $E_{ind,d-class}^{(4-\infty)}$ is the damped classical induction energy summed from the second iteration onward (i.e., to convergence). The third-order energies are expected to dominate $U^{(3-\infty)}$, so this approximation would include most of the overlap effects that are missing from the classical polarizable model. The SAPT expressions for $E_{ind,pol}^{(3)}$ and $E_{ind,exch}^{(3)}$ have been derived[20,38] but without the inclusion of orbital relax-

ation effects. That is, when used with Kohn–Sham orbitals and eigenvalues, these energies are obtained at the SAPT-(KS) level of theory. We will denote the resulting approximation by $U^{(3-\infty)}(KS)$ which, using eq 1, is defined as

$$U^{(3-\infty)}(KS) = E_{ind,tot}^{(3)}(KS) + E_{ind,d-class}^{(4-\infty)} \qquad (24)$$

Orbital relaxation effects have been demonstrated to play a relatively minor role for the second-order induction.[11] In ref 11 this was argued to be at least in part due to a cancellation of errors. As there is no reason to expect the same to happen for the third-order energies, one might question the use of the SAPT(KS) expressions here. However, $U^{(3-\infty)}$ is relatively small in magnitude compared with the second-order energy, so any errors incurred by the use of the SAPT(KS) expressions are probably less important.

From Figure 1 we see that $U^{(3-\infty)}(KS)$ is a far better approximation to the higher-order energy contributions than the damped classical polarizable model alone, but the higher-order terms are probably still underestimated by this approximation. In Table 1 we report second-order interaction energies and the various estimates of the third-order terms discussed above for a few dimers. For the water dimer, taking $U^{CCSD(T)}$ as a reference, we see that while adding $\delta_{int,resp}^{HF}$ to $U^{(2)}$ from SAPT may make sense, adding it to $U^{(2)}$ from SAPT(DFT) leads to an overestimate of the interaction energy by about 4%. On the other hand, adding $U^{(3-\infty)}(KS)$ to $U^{(2)}$ from SAPT(DFT) leads to an interaction energy *under*estimated by about 4%. Therefore it is possible that the $\delta_{int,resp}^{HF}$ correction is an overestimate of the higher-order terms.

The effects of these approximations on the total interaction energy for the water dimer are more clearly represented in Figure 2. The SAPT(DFT) interaction energy at second order, $U^{(2)}[SAPT(DFT)]$, results in a potential that is clearly too shallow, with the repulsive wall and minimum both moved out toward larger R. Adding $\delta_{int,resp}^{HF}$ to $U^{(2)}[SAPT(DFT)]$ results in a potential curve that is apparently too deep with both the repulsive wall and minimum moved inward. This potential curve is quite similar to the SAPT curve. The best agreement with the CCSD(T) potential is obtained with the $U^{(2)}[SAPT(DFT)]+U^{(3-\infty)}(KS)$ approximation, but this could be due to the slow convergence of the CCSD(T) interaction energy with respect to basis set.

Also reported in Table 1 are interaction energies for the strongly polar hydrogen fluoride dimer and the $H_2O\cdots H_3N$ dimer in a weakly polar geometry. The hydrogen fluoride dimer is probably the worst case for perturbation theory. For SAPT(DFT), $U^{(2)}$ constitutes only 83% of the reference CCSD(T) interaction energy and including $U^{(3-\infty)}(KS)$ results in an improvement but still recovers only 91% of the reference. On the other hand, using the $\delta_{int,resp}^{HF}$ estimate of higher-order energies results in a near perfect agreement with CCSD(T). In contrast, for the $H_2O\cdots H_3N$ dimer, perturbation theory is rather rapidly convergent with $U^{(2)}[SAPT(DFT)]$ and CCSD(T) differing by about 1% only. This good agreement is made slightly worse by the addition of higher-order energy estimates as $U^{(2)}[SAPT(DFT)]+\delta_{int,resp}^{HF}$ and $U^{(2)}[SAPT(DFT)] + U^{(3-\infty)}(KS)$ differ from CCSD(T)

Table 1. Contribution of Third- and Higher-Order Corrections to the Interaction Energy for the Water, Hydrogen Fluoride, Carbon Dioxide, and Benzene Dimers and the $H_2O\cdots H_3N$ and $H_2\cdots CO$ Complexes[e]

method	energy component	$(H_2O)_2$	$(HF)_2$	$H_2O\cdots H_3N$	$(CO_2)_2$	$(C_6H_6)_2$	$H_2\cdots CO$
SAPT	$U^{(2)}$	−18.08	−15.59	−5.803	−7.05		−1.123
	$E^{(30)}_{ind,tot}$	−0.80	−0.95	−0.017	−0.12		−0.018
	$\delta^{HF}_{int,resp}$	−2.82	−3.14	−0.098	−0.23		−0.187
	$U^{(2)}+\delta^{HF}_{int,resp}$	−20.90	−18.73	−5.901	−7.28		−1.310
SAPT(DFT)	$U^{(2)}$	−18.44	−15.38	−5.776	−5.72	−8.11[c]	−1.066
	$E^{(3)}_{ind,tot}(KS)$	−1.20	−1.41	−0.027	−0.17	+0.29[d]	−0.023
	$E^{(4-\infty)}_{ind,d-class}$	−0.04	−0.02	0.00	0.00	0.00	0.00
	$U^{(2)}+E^{(3)}_{ind,tot}(KS)+E^{(4-\infty)}_{ind,d-class}$	−19.67	−16.81	−5.803	−5.90	−7.82	−1.090
	$U^{(2)}+\delta^{HF}_{int,resp}$	−21.26	−18.52	−5.874	−5.95		−1.253
CCSD(T)	$U^{CCSD(T)}$	−20.45	−18.50	−5.694	−5.94[a]	−7.57[b]	−1.063

[a] Misquitta et al.[12] Dispersion optimized basis augmented with mid-bond functions. [b] Sinnokrot et al.[37] Estimate of the complete basis set CCSD(T) energy. [c] $E^{(2)}_{disp}$ was calculated in a TZ/MC$^+$ basis set , the rest of the interaction energy components in the Sadlej/MC$^+$ basis set. [d] Computed using a Sadlej/MC$^+$ basis set. [e] The first three dimers are at their equilibrium geometries, the benzene dimer is in the parallel stacked geometry with a center-of-mass separation of 3.8 Å. $H_2\cdots CO$ is in the linear geometry with C toward H_2 and a center-of-mass separation of 7.8 au, and $H_2O\cdots H_3N$ is in geometry (a) of Figure 5 from ref 36. As stated in section III, the induction energy, $E^{(n)}_{ind,tot}$, is defined as the *sum*: $E^{(n)}_{ind,pol}+E^{(n)}_{ind,exch}$. Unless otherwise specified, interaction energies were calculated using the aug-cc-PVTZ/MC$^+$ basis, and molecular properties needed for the damped classical model were obtained using the aug-cc-pVTZ/MC basis. All energies are reported in kJ mol^{-1}.
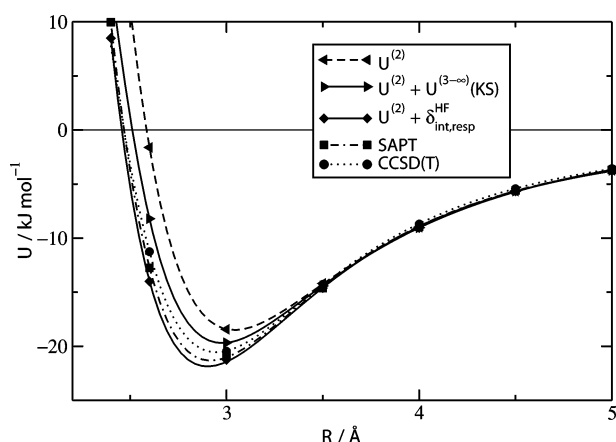

Figure 2. Total interaction energies of the water dimer obtained with SAPT(DFT) using different approximations to the higher-order energies. SAPT and CCSD(T) interaction energies are also shown. We have used the standard SAPT definition of the interaction energy that includes the $\delta^{HF}_{int,resp}$ correction (see, for example, eqs 4 and 5 in ref 12). All calculations were performed using the aug-cc-pVTZ basis in the MC$^+$ format. The dimer geometry is the same as used in Figure 1.

by 3% and 2%, respectively. Since these differences are small enough to be ascribed to basis incompleteness effects, we conclude that either estimate of the higher-order energies is suitable for this system.

We now turn to dimers for which the induction energy plays a relatively unimportant role. In Table 1 we report interaction energies for three such dimers: the carbon dioxide dimer, the benzene dimer, and the $H_2\cdots CO$ dimer. Results are far more encouraging for the nonpolar dimers. For the carbon dioxide dimer, all estimates of $U^{(3-\infty)}$ are very small but still constitute about 3% of the interaction energy. $\delta^{HF}_{int,resp}$ and $U^{(3-\infty)}(KS)$ are both about −0.2 kJ mol^{-1}, and SAPT(DFT) interaction energies obtained using either estimate of the higher-order energies results in an interaction energy almost identical with the CCSD(T) reference. For this particular system, $U^{(2)}$ from SAPT is too large in magnitude

due to a severe overestimation of the dispersion energy made by the current implementation of SAPT.[12]

The benzene dimer is a rather unusual system as the higher-order induction effects are positive in the stacked geometry. Due to its size, no SAPT calculation was possible for this system. Indeed, accurate reference energies for this system are hard to obtain. The best reference energies currently available are those from Sinnokrot et al.,[37] obtained using R12-MP2 energies together with a CCSD(T) correction obtained using an aug-cc-pVDZ basis. Unfortunately, since the MP2 interaction energy is particularly poor for this system, being an overestimation by nearly 100%, the estimate of Sinnokrot et al. is probably in error by a few percent, and we should keep this in mind in the following comparisons. $U^{(2)}$ from SAPT(DFT) differs from the estimated CCSD(T) interaction energy differ by −7%. On adding the $U^{(3-\infty)}(KS)$, this difference is reduced to only −3%. The higher-order contributions constitute about 4% of the total interaction energy of the dimer. Basis set incompleteness effects, estimated by Podeszwa et al.,[39] are of the same magnitude but of opposite sign. Therefore, the higher-order contributions to the interaction energy cannot be neglected in high-accuracy calculations, even for nonpolar systems, particularly as it is becoming usual to estimate the complete basis set limit of the interaction energies.

The $H_2\cdots CO$ dimer is the most weakly bound system included in Table 1. SAPT(DFT) converges very quickly to the CCSD(T) reference energy with $U^{(2)}$ [SAPT(DFT)] and CCSD(T) being essentially equal and $U^{(2)}[SAPT(DFT)]+U^{(3-\infty)}(KS)$ differing from CCSD(T) by only 3%. In contrast, $U^{(2)}$-[SAPT(DFT)]$+\delta^{HF}_{int,resp}$ overshoots the CCSD(T) reference by 18%. The situation is worse for SAPT energies: $U^{(2)}$-[SAPT] is already too negative by about 6%, and the inclusion of the $\delta^{HF}_{int,resp}$ term makes the SAPT energy about 23% too negative. Clearly then, in agreement with Patkowski et al.,[20] the $\delta^{HF}_{int,resp}$ term is a very poor estimate of the higher-order energies for this dimer. This is the only system for which we have found this to be the case though Patkowski

et al.[20] have found that $\delta^{\mathrm{HF}}_{\mathrm{int,resp}}$ behaves likewise for the argon dimer.

From Table 1 and Figure 1 we see that for all of these dimers, the damped classical estimate of the induction effects above third order, $E^{(4-\infty)}_{\mathrm{ind,d-class}}$, is very small and could be neglected without a significant loss in accuracy. Thus our recommended approximation for the higher-order contributions to the interaction energy of a dimer is

$$U^{(3-\infty)} \approx U^{(3-\infty)}(\mathrm{KS}) \approx E^{(3)}_{\mathrm{ind,tot}}(\mathrm{KS}) \qquad (25)$$

IV.2. Many-Body Induction. In the condensed phase of water, the two-body interaction energies have been found to account for only about 85% of the total interaction energy per molecule.[23] The remaining 15% arises from many-body nonadditive effects, that is, that part of the interaction energy that cannot be represented by the sum of pairwise interactions. This nonadditivity is responsible for some of the important structural properties of water and, in particular, has a large role to play in hydrogen bonding. The nonadditive effects are even larger in small water clusters where they constitute between 17% and 30% of the total interaction energy.[22,24] These effects are expected to be equally important for polar molecules other than water and must be included in atom−atom potentials for organic molecules, which are commonly very polar and often form hydrogen-bonded networks.

The perturbative treatment of nonadditivity is a complex field of research, and while there is a version of SAPT that includes the three-body nonadditivity (see ref 7 for a review), the computational demands are so high as to preclude its applications to organic molecules. However, one of the major conclusions of accurate studies on water clusters[22−24] has been that the bulk of the nonadditivity for polar systems can be recovered using the relatively simple damped classical polarization model (see ref 3 for a description). While the exchange nonadditivity is not negligible for small clusters,[22,24] it is less important in relative terms for large clusters and the condensed phase, because additional coordination shells around any given molecule increase the dispersion, induction, and electrostatic energies but not the short-range contributions like the exchange nonadditivity.[23]

The damped classical polarizable model for a cluster of molecules is again given by eqs 21 and 22, but we now take into account all the molecules in the cluster. While the effect of the iterations in eq 22 is quite small for the two-body energy, iterations have a much larger role to play in larger clusters of polar molecules. For example, in clusters of water molecules optimized using the ASP-W4 potential,[22] iterations in eq 22 stabilize the dimer, trimer, tetramer, and pentamer by 2.5%, 6.0%, 10.4%, and 12.3% of the total interaction energy, respectively. These effects are clearly considerable. Furthermore, without iterations, many-body contributions to the interaction energy above the 3-body energies are absent (the $(3 + m)$-body terms arise at the mth iteration of eq 22). Work on organic crystals also indicates that iterations contribute strongly to the lattice energy.[40]

IV.3. Asymptotic Induction Energies. The asymptotic expansion of the two-body, second-order induction energy involves polarizability tensors and multipole moments of the unperturbed monomers (see ref 7 for a review). For large molecules, these molecular properties must be distributed, that is, expressed in terms of multiple sites—usually chosen to be the atomic centers—so as to improve the convergence of the multipole expansion. The distribution of the multipole moments has been the subject of many decades of research (see refs 3 and 41 for reviews). The Distributed Multipole Analysis (DMA) of Stone[42] is widely used, and a recent modification[43] overcomes a shortcoming of the earlier method that arose when diffuse functions were present in the basis. The problem of distributing the polarizabilities has proved harder, and only fairly recently have methods become available that are suitable for molecules of 20−30 atoms and modern basis sets, while being general enough to be applicable to frequency-dependent polarizabilities.[44,45]

The frequency-dependent polarizability, $\alpha_{lm,l'm'}(\omega)$, can be defined in terms of the FDDS as

$$\alpha_{lm,l'm'}(\omega) = \int\int \alpha(\mathbf{r},\mathbf{r}'|\omega)\hat{Q}_{lm}(\mathbf{r})\hat{Q}_{l'm'}(\mathbf{r}')\mathrm{d}^3\mathbf{r}\,\mathrm{d}^3\mathbf{r}' \quad (26)$$

Real-space partitioning schemes have been based on ways of defining the multipole moment operators in the above expression so that they act on finite regions of space, defined, for example, by using integration grids[46] or by Bader's theory of atoms in molecules.[47] Both of these methods have shortcomings[45] that make them unsuitable for practical use. In contrast, the distribution scheme of Misquitta and Stone[45] focuses on a partitioning of the FDDS. The FDDS, as defined by eq 5, cannot be directly partitioned as the molecular orbitals that appear in this expression are generally delocalized. Rather, density-fitting[48,49] is used to simplify the form of the FDDS and then achieve the necessary site−site partitioning.

If the auxiliary basis set used to obtain the density-fitted FDDS in eq 10 is partitioned into contributions from individual sites, that is, $\{\chi\} = \{\chi^{(1)}, \chi^{(2)}, ...\}$, then the FDDS can be written as

$$\alpha(\mathbf{r},\mathbf{r}'|\omega) \approx \sum_{ab} \alpha^{a,b}(\mathbf{r},\mathbf{r}'|\omega) \qquad (27)$$

where

$$\alpha^{a,b}(\mathbf{r},\mathbf{r}'|\omega) = \sum_{p\in a,q\in b} \tilde{C}_{pq}(\omega)\chi_p(\mathbf{r})\chi_q(\mathbf{r}') \qquad (28)$$

Finally, using eqs 27 and 26 the distributed polarizability for sites (a, b) is defined as

$$\alpha_{lm,l'm'}{}^{a,b}(\omega) = \sum_{p\in a,q\in b} \tilde{C}_{pq}(\omega)N^p_{lm}N^q_{l'm'} \qquad (29)$$

where $N^p_{lm} = \int\hat{Q}_{lm}(\mathbf{r} - \mathbf{a})\chi_p(\mathbf{r})\mathrm{d}^3\mathbf{r}$, where $\mathbf{a}$ is a suitable reference origin for site a that will typically be taken to be the nucleus.

The standard density-fitting procedure[48,49] involves the minimization of the function

$$\Delta_{iv} = \int\int [\rho_{iv}(\mathbf{r}_1) - \tilde{\rho}_{iv}(\mathbf{r}_1)]\frac{1}{r_{12}}[\rho_{iv}(\mathbf{r}_2) - \tilde{\rho}_{iv}(\mathbf{r}_2)]\mathrm{d}^3\mathbf{r}_1\mathrm{d}^3\mathbf{r}_2 \quad (30)$$

where the transition density $\rho_{iv} = \phi_i\phi_v$ is approximated by $\tilde{\rho}_{iv} = \sum_k D_{iv,k}\chi_k$. For the method of distribution based on

Induction Energies for Small Organic Molecules: 1

J. Chem. Theory Comput., Vol. 4, No. 1, 2008 **15**

eq 29 to work, it has been shown[45] that the function to be minimized must be replaced by one involving additional constraints

$$\Xi_{iv} = \Delta_{iv} - \eta \sum_{a, b \neq a} E_{iv}^{ab} + \lambda \left(\sum_a \sum_{k \in a} D_{iv,k} I_k \right)^2 \quad (31)$$

where η and λ are constants empirically determined to be about 0.0005 and 1000.0, respectively, $I_k = \int \chi_k(\mathbf{r}) d^3\mathbf{r}$, so that the last term imposes the orthogonality of the occupied and virtual orbitals (cf. eq 9), and E_{iv}^{ab} is the Coulomb interaction between the contributions of the basis functions of sites a and b to the transition density ρ_{iv} and is defined as

$$E_{iv}^{ab} = \int \int \frac{\tilde{\rho}_{iv}^a(\mathbf{r}_1) \tilde{\rho}_{iv}^b(\mathbf{r}_2)}{r_{12}} d^3\mathbf{r}_1 d^3\mathbf{r}_2 \quad (32)$$

The distributed polarizabilities obtained using this constrained density-fitting procedure contain nonlocal terms, that is, terms involving pairs of distinct sites. In contrast to other distribution methods, the nonlocal terms describing flow of charge from site to site are very small (around 10^{-2} and 10^{-3} in magnitude) for all systems, irrespective of the type of bonding involved. Nevertheless, nonlocal terms are best avoided as they complicate the description unnecessarily. In ref 45, the Le Sueur and Stone localization method[50] was used to transform the nonlocal terms into local polarizabilities and remove the charge-flow terms altogether. The localization by this procedure causes a deterioration of the convergence properties of the model, because multipole expansions are used to move the polarizabilities around. In principle, this can increase the radius of divergence of the description to be equal to the size of the molecule, thereby causing significant losses in accuracy for large molecules. Thus, while good results have been obtained for molecules like formamide and urea, there is already an appreciable loss in accuracy for *N*-methyl propanamide.[45]

In ref 45 it was suggested that the Williams and Stone method[44] of obtaining local polarizabilities using a fit to the point-to-point polarizabilities could be used to refine the polarizability model obtained from the constrained density-fitting procedure. This can be done as follows. The point-to-point polarizability α_{PQ} describes the response of the electrostatic potential at a point Q to the frequency-dependent potential produced by a unit oscillating point charge at point P and is given by the expression

$$\alpha_{PQ}(\omega) = \int \int \alpha(\mathbf{r}, \mathbf{r}'|\omega) \hat{O}^P(\mathbf{r}) \hat{O}^Q(\mathbf{r}') d^3\mathbf{r} d^3\mathbf{r}' \quad (33)$$

where $\hat{O}^P(\mathbf{r}) = -q_P/(4\pi\epsilon_0|\mathbf{P}-\mathbf{r}|)$ and $\hat{O}^Q(\mathbf{r}) = -q_Q/(4\pi\epsilon_0|\mathbf{Q}-\mathbf{r}|)$. These polarizabilities are evaluated on a random grid of points typically between the vdW $\times$ 2 and vdW $\times$ 4 surfaces. These responses can be evaluated very efficiently using the density-fitted form of eq 33 given in ref 45. For a grid of N points, where N is typically a few thousand, there are $\frac{1}{2}N(N+1)$ responses, which can be obtained in a single calculation. In the original Williams and Stone method, a model is postulated comprising polarizabilities α_{tu}^{ab}, where $t = 00, 10, 10c, 10s, ...,$. In terms of this model the responses are given by

$$\tilde{\alpha}_{PQ} = \sum_{ab} \sum_{tu} T_{0t}^{Pa} \alpha_{tu}^{ab} T_{u0}^{bQ} \quad (34)$$

where T_{0t}^{Pa} is the interaction tensor[3] which describes the interaction between a point charge at P and a multipole Q_t at point a. The model polarizabilities α_{tu}^{ab} are then obtained by minimizing the squared difference

$$S = \sum_{PQ} (\tilde{\alpha}_{PQ} - \alpha_{PQ})^2 \quad (35)$$

This procedure is very accurate and leads to a compact description of the polarizability, but the resulting polarizabilities are not always positive definite. This tends to happen for 'buried' atoms, i.e., atoms hidden under the van der Waals spheres of neighboring atoms, and could, in principle, lead to positive induction energies, which is physically impossible.

These unphysical terms can be avoided by using the models obtained using the localized polarizabilities from the constrained density-fitting method as 'anchor' values and minimizing

$$S = \sum_{PQ} (\tilde{\alpha}_{PQ} - \alpha_{PQ})^2 + \sum_{kk'} g_{kk'}(p_k - p_k^0)(p_{k'} - p_{k'}^0) \quad (36)$$

where the p_k are the parameters in the model, i.e., the α_{tu}^{ab} defined above, p_k^0 are the 'anchor' values, and $g_{kk'}$ are elements of a positive definite matrix that could be taken to be diagonal.

This combination of the Williams–Stone and Misquitta–Stone procedures will be called the Williams–Stone–Misquitta (WSM) procedure. Initial results using the WSM procedure have been very encouraging and have been presented in a recent review article.[41] More extensive results will be presented in part 2.

V. Summary

In this first part of our investigation, we have laid down the theoretical framework for the accurate calculation of the induction energy of clusters of organic molecules.

We have broken away from convention by identifying the induction energy with the *sum* of the second-order induction energy as defined through the polarization expansion,[6,7] termed $E_{\text{ind,pol}}$, and its exchange counterpart, the exchange-induction energy $E_{\text{ind,exch}}$. Thus, we define the *n*th order induction energy as $E_{\text{ind,tot}}^{(n)} = E_{\text{ind,pol}}^{(n)} + E_{\text{ind,exch}}^{(n)}$. This definition was motivated by both theoretical and numerical considerations.

The two-body induction energy at second order in the interaction operator, $E_{\text{ind,pol}}^{(2)}$, is the most important contribution of the induction energy to the interaction energy of a cluster of molecules. We have presented a density-fitted form of the SAPT(DFT) expression for $E_{\text{ind,pol}}^{(2)}$ that is both accurate and computationally efficient. This is a natural extension of the density-fitting technique that one of us has used for the dispersion energy[10,12] and that has already been proposed—in a different form—by Hesselmann et al.[16] As well as a reduction in the computational cost from $O(n_o^2 n_v^2)$ to $O(m^2)$ where n_o and n_v are the number of occupied and virtual orbitals, respectively, and m is the number of auxiliary basis functions, we gain by avoiding the computationally

expensive 4-index Coulomb integrals needed to evaluate the original SAPT(DFT) expression. The new formulation has been implemented in the CamCASP program[51] which was used for all calculations of $E_{\text{ind,pol}}^{(2)}$ reported in this paper.

The two-body interaction energy has major contributions from energies of third and higher order in the interaction operator. These higher-order energies are predominantly induction in nature and can contribute as much as 17% of the two-body interaction energy for hydrogen-bonded complexes. They have usually been estimated using the $\delta_{\text{int,resp}}^{\text{HF}}$ correction[7,17,18] which is cumbersome to calculate as it involves a supermolecular Hartree−Fock calculation of the induction energy and a low-order SAPT calculation. We have proposed that these energies be approximated using the third-order induction energy calculated using SAPT(KS), i.e., $E_{\text{ind,tot}}^{(3)}$(KS). Since SAPT(KS) is the first step in a SAPT(DFT) calculation of the interaction energy, this entails little additional effort.

$E_{\text{ind,tot}}^{(3)}$(KS) has been shown to approximate the higher-order energies rather well for non-hydrogen-bonded dimers, where we get almost perfect agreement with the reference interaction energies, but it underestimates them for hydrogen-bonded dimers. For example, the interaction energy calculated using this approximation is underestimated by around 4% for the water dimer and 9% for the hydrogen fluoride dimer, both at their global minimum geometries. The hydrogen fluoride dimer is probably the worst case for methods based on perturbation theory as higher-order energies are estimated to constitute about 17% of the interaction energy at the global minimum geometry. For the non-hydrogen-bonded dimers, the higher-order energies are smaller but still constitute about 4% of the interaction energy. Patkowski et al.[20] were led to similar conclusions in an investigation of the SAPT interaction energy.

With the exception of the very weakly bound $H_2\cdots CO$ dimer, for the polar and nonpolar dimers studied here, we found that the $\delta_{\text{int,resp}}^{\text{HF}}$ correction provides a reasonably accurate estimate of higher-order energies. This conclusion complements that of Patkowski et al.[20] It is quite possible that $\delta_{\text{int,resp}}^{\text{HF}}$ is indeed a poor estimate of the higher-order energies for very weakly bound dimers but might be more reasonable for the more strongly bound dimers, whether polar of not. More data from a larger variety of systems will be needed to test this conjecture.

Yet another way of estimating the higher-order energies is through the damped classical polarizable model. The classical estimate for the higher-order energy is obtained by iterating the fields and induced multipoles self-consistently to convergence. It is necessary to use a distributed-polarizability description for all but the smallest molecules, and the Williams−Stone−Misquitta procedure[44,45] provides an efficient and accurate route to such descriptions. (See results provided in part 2.) We have found that the damped classical polarizable model severely underestimates the higher-order energies for the two-body interaction. Recall that it is incomplete because it assumes linear response of each molecule to external fields and neglects orbital overlap effects. For example, it recovers less than 10% of the interaction energy contribution from third and higher orders

for the water dimer at its global minimum geometry. Therefore, using the damped classical model to estimate these energies would result in an error of about 9%, or about 1.8 kJ mol^{-1}, in the total interaction energy of the water dimer.

However, the effect of the iterations can be quite large in clusters of polar molecules. For example, the additional stabilization is about 12% of the total interaction energy for the water pentamer.[22] Iterations have also been shown to make major contributions to the lattice energy of organic crystals.[40] Therefore, we do recommend that the iterated form of the damped classical polarizable model be used in calculations involving polar clusters.

We should perhaps emphasize that the higher-order energies are *larger* in magnitude than the basis set incompleteness errors in the SAPT(DFT) calculations. Therefore these energies cannot be ignored in accurate calculations, especially those attempting to estimate the complete-basis-set energy.

Our recommended expression for the total interaction energy of a dimer calculated using SAPT(DFT)[12] is

$$U \approx E_{\text{elst}}^{(1)}(\text{KS}) + E_{\text{exch}}^{(1)}(\text{KS}) + E_{\text{ind,tot}}^{(2)} +$$
$$E_{\text{disp}}^{(2)} + E_{\text{disp,exch}}^{(2)} + U^{(3-\infty)} \quad (37)$$

where $E_{\text{elst}}^{(1)}(\text{KS})$ and $E_{\text{exch}}^{(1)}(\text{KS})$ are the first-order electrostatic and exchange energies, respectively, $E_{\text{disp}}^{(2)}$ and $E_{\text{disp,exch}}^{(2)}$ are the second-order dispersion and exchange-dispersion energies, respectively, $E_{\text{ind,tot}}^{(2)}$ is given by eq 19, and $U^{(3-\infty)}$ is approximated as in eq 25. This approximation does not include higher-order dispersion terms or nonlinear induction effects. In Part 2 we will explore further approximations and also investigate the numerical issues associated with calculations of the induction energy of dimers and clusters of organic molecules.

References

(1) Misquitta, A. J.; Stone, A. J.; Price, S. L. Accurate induction energies for small organic molecules: 2. Models and numerical details. *J. Chem. Theory Comput.* **2007**, *3*, 19−32.

(2) Bukowski, R.; Szalewicz, K.; Groenenboom, G.; van der Avoird, A. Interaction potential for water dimer from symmetry-adapted perturbation theory based on density functional description of monomers. *J. Chem. Phys.* **2006**, *125*, 044301.

(3) Stone, A. J. *The Theory of Intermolecular Forces*; Clarendon Press: Oxford, 1996.

(4) Magnasco, V.; McWeeny, R. Weak interaction between molecules and their physical interpretations. In *Theoretical Models of Chemical Bonding*; Maksić, Z. B., Ed.; Springer: New York, 1991; Vol. 4, pp 133−169.

(5) McWeeny, R. *Methods of Molecular Quantum Mechanics*, 2nd ed.; Academic Press: New York, 1992.

(6) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation theory approach to intermolecular potential energy surfaces of Van der Waals complexes. *Chem. Rev.* **1994**, *94*, 1887−1930.

(7) Jeziorski, B.; Szalewicz, K. Symmetry-adapted perturbation theory. In *Handbook of Molecular Physics and Quantum Chemistry*; Wilson, S., Ed.; Wiley: 2002; Vol. 8, pp 37−83.

(8) Szalewicz, K.; Patkowski, K.; Jeziorski, B. Intermolecular interactions via perturbation theory: From diatoms to biomolecules. In *Intermolecular Forces and Clusters II*; Wales, D. J., Ed.; Springer-Verlag: Berlin, Heidelberg, 2005; Vol. 116 of *Structure and Bonding*, pp 43−117.

(9) Misquitta, A. J.; Szalewicz, K. Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem. Phys. Lett.* **2002**, *357*, 301−306.

(10) Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. Dispersion energy from density-functional theory description of monomers. *Phys. Rev. Lett.* **2003**, *91*, 33201.

(11) Misquitta, A. J.; Szalewicz, K. Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *J. Chem. Phys.* **2005**, *122*, 214109.

(12) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional theory. *J. Chem. Phys.* **2005**, *123*, 214103.

(13) Hesselmann, A.; Jansen, G. First-order intermolecular interaction energies from Kohn−Sham orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464−470.

(14) Hesselmann, A.; Jansen, G. Intermolecular induction and exchange-induction energies from coupled-perturbed Kohn−Sham density functional theory. *Chem. Phys. Lett.* **2002**, *362*, 319−325.

(15) Hesselmann, A.; Jansen, G. Intermolecular dispersion energies from time-dependent density functional theory. *Chem. Phys. Lett.* **2003**, *367*, 778−784.

(16) Hesselmann, A.; Jansen, G.; Schutz, M. Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J. Chem. Phys.* **2005**, *122*, 014103.

(17) Jeziorska, M.; Jeziorski, B.; Cizek, J. Direct calculation of the Hartree−Fock interaction energy via exchange perturbation expansion−the He···He interaction. *Int. J. Quantum Chem.* **1987**, *32*, 149−164.

(18) Moszynski, R.; Heijmen, T. G. A.; Jeziorski, B. Symmetry-adapted perturbation theory for the calculation of Hartree−Fock interaction energies. *Mol. Phys.* **1996**, *88*, 741−758.

(19) Mas, E. M.; Bukowski, R.; Szalewicz, K. Ab initio three-body interactions for water. I. Potential and structure of water trimer. *J. Chem. Phys.* **2003**, *118*, 4386−4403.

(20) Patkowski, K.; Szalewicz, K.; Jeziorski, B. Third-order interactions in symmetry-adapted perturbation theory. *J. Chem. Phys.* **2006**, *125*, 154107.

(21) Patkowski, K.; Jeziorski, B.; Szalewicz, K. Symmetry-adapted perturbation theory with regularized coulomb potential. *J. Mol. Struct. (THEOCHEM)* **2001**, *547*, 293−307.

(22) Hodges, M. P.; Stone, A. J.; Xantheas, S. S. Contribution of many-body terms to the energy for small water clusters: A comparison of ab initio calculations and accurate model potentials. *J. Phys. Chem. A* **1997**, *101*, 9163−9168.

(23) Mas, E. M.; Bukowski, R.; Szalewicz, K. Ab initio three-body interactions for water. II. Effects on structure and energetics of liquid. *J. Chem. Phys.* **2003**, *118*, 4404−4413.

(24) Milet, A.; Moszynski, R.; Wormer, P. E. S.; van der Avoird, A. Hydrogen bonding in water clusters: Pair and many-body interactions from symmetry-adapted perturbation theory. *J. Phys. Chem. A* **1999**, *103*, 6811−6819.

(25) Casida, M. E. Time-dependent density-functional response theory for molecules. In *Recent Advances in Density-Functional Theory*; Chong, D. P., Ed.; World Scientific: 1995; p 155.

(26) Colwell, S. M.; Handy, N. C.; Lee, A. M. Determination of frequency-dependent polarizabilities using current density-functional theory. *Phys. Rev. A* **1996**, *53*, 1316−1322.

(27) Petersilka, M.; Gossmann, U. J.; Gross, E. K. U. Excitation energies from time-dependent density-functional theory. *Phys. Rev. Lett.* **1996**, *76*, 1212−1215.

(28) Hesselmann, A.; Jansen, G. First-order intermolecular interaction energies from Kohn−Sham orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464−470.

(29) Tozer, D. J.; Handy, N. C. Improving virtual Kohn−Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities. *J. Chem. Phys.* **1998**, *109*, 10180−10189.

(30) Tozer, D. J. The asymptotic exchange potential in Kohn−Sham theory. *J. Chem. Phys.* **2000**, *112*, 3507−3515.

(31) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158−6170.

(32) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Density-fitting method in symmetry-adapted perturbation theory based on Kohn−Sham description of monomers. *J. Chem. Theory Comput.* **2006**, *2*, 400−412.

(33) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew−Burke−Ernzerhof exchange-correlation functional. *J. Chem. Phys.* **1999**, *110*, 5029−5036.

(34) Misquitta, A. J.; Stone, A. J. Regularized SAPT(DFT) **2007**, manuscript in preparation.

(35) Mas, E. M.; Szalewicz, K.; Bukowski, R.; Jeziorski, B. Pair potential for water from symmetry-adapted perturbation theory. *J. Chem. Phys.* **1997**, *107*, 4207−4218.

(36) Langlet, J.; Caillet, J.; Bergès, J.; Reinhardt, P. Comparison of two ways to decompose intermolecular interactions for hydrogen-bonded dimer systems. *J. Chem. Phys.* **2003**, *118*, 6157−6166.

(37) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. Estimates of the ab initio limit for $\pi-\pi$ interactions: The benzene dimer. *J. Am. Chem. Soc.* **2002**, *124*, 10887−10893.

(38) Moszyński, R.; Cybulski, S. M.; Chałasiński, G. Many-body theory of intermolecular induction interactions. *J. Chem. Phys.* **1994**, *100*, 4998−5010.

(39) Podeszwa, P.; Bukowski, R.; Szalewicz, K. Potential energy surface for the benzene dimer and perturbational analysis of $\pi-\pi$ interactions. *J. Phys. Chem. A* **2006**, *110*, 10345−10354.

(40) Welch, G. W. A.; Karamertzanis, P. G.; Misquitta, A. J.; Stone, A. J.; Price, S. L. Is the induction energy important for modelling organic crystals? *J. Chem. Theory Comput.* **2007**, manuscript in preparation.

(41) Stone, A. J.; Misquitta, A. J. Atom−atom potentials from *ab initio* calculations. *Int. Rev. Phys. Chem.* **2007**, *26*, 193−222.

(42) Stone, A. J.; Alderton, M. Distributed multipole analysis−methods and applications. *Mol. Phys.* **1985**, *56*, 1047−1064.

(43) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128−1132.

(44) Williams, G. J.; Stone, A. J. Distributed dispersion: a new approach. *J. Chem. Phys.* **2003**, *119*, 4620−4628.

(45) Misquitta, A. J.; Stone, A. J. Distributed polarizabilities obtained using a constrained density-fitting algorithm. *J. Chem. Phys.* **2006**, *124*, 024111.

(46) Le Sueur, C. R.; Stone, A. J. Practical schemes for distributed polarizabilities. *Mol. Phys.* **1993**, *78*, 1267−1291.

(47) Angyan, J. G.; Jansen, G.; Loos, M.; Hattig, C.; Hess, B. A. Distributed polarizabilities using the topological theory of atoms in molecules. *Chem. Phys. Lett.* **1994**, *219*, 267−273.

(48) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On first-row diatomic molecules and local density models. *J. Chem. Phys.* **1979**, *71*, 4993−4999.

(49) Dunlap, B. I. Robust and variational fitting. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2113−2116.

(50) Le Sueur, C. R.; Stone, A. J. Localization methods for distributed polarizabilities. *Mol. Phys.* **1994**, *83*, 293−308.

(51) Misquitta, A. J.; Stone, A. J. *CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form*; University of Cambridge: 2006. Inquiries to A. J. Misquitta, am592@cam.ac.uk.