Accuracy of typical approximations in classical models of intermolecular polarization

Söderhjelm, Pär; Öhrn, Anders; Ryde, Ulf; Karlström, Gunnar

Published in:
Journal of Chemical Physics

DOI:
10.1063/1.2814240

2008

Document Version:
Peer reviewed version (aka post-print)

Accuracy of typical approximations in classical models of intermolecular polarization

Pär Söderhjelm,* Anders Öhrn, Ulf Ryde, and Gunnar Karlström

Department of Theoretical Chemistry,

Chemical Center, Lund University

P.O. Box 124,

S-22100 Lund

Sweden

Abstract

One of the largest limitations of standard molecular-mechanics force fields is the neglect of intermolecular polarization. Several attempts to cure this problem have been made, but the results have not always been fully satisfactory. In this paper, we present a quantitative study of the fundamental approximations that underlie polarization models for classical force fields. The induced density of a large set of molecular dimers is compared to supermolecular calculations for a hierarchy of simplified models. We study the effect of the Pauli principle on the intermolecular polarization, non-linear response and intramolecular coupling of the induced dipoles. We show that point-polarizability models work rather well, despite their drastic approximations, because of a systematic error cancellation of the neglect of Pauli effects and non-linear effects. However, the cancellation is not perfect and therefore, polarizable force-fields could be improved if both Pauli effects and non-linear effects are explicitly treated.

PACS numbers:

*Corresponding author: `par.soderhjelm@teokem.lu.se`

I. INTRODUCTION

During the recent decades, computational methods have become an important tool in chemistry and a versatile complement to experimental investigations. The most accurate results are obtained with quantum chemical calculations, but they can only be used for relatively small systems. Therefore, classical molecular-mechanics methods have become popular for, among others, biomolecular investigations. These methods replace the quantum mechanical description with an empirical force field. Naturally, the accuracy of such methods is determined by the accuracy of this force field. In the first generation of force fields, the inductive polarization effects were omitted and electrostatics were typically modelled by atom-centered partial charges. However, it is well-known that polarization often accounts for as much as $\sim$20% of the interaction energy. Therefore, many attempts have been made to include such effects in the simulations [1]. They often gave marked improvements in simulations of simple condensed systems, but for more complex systems, like proteins, the inclusion of polarization in the force fields often led to a too small improvement (if any) to justify the increase of the computational load.

For this reason, much effort has been directed to improve the treatment of polarization, e.g. by developing new methods to compute the distributed polarizabilities [2–7]. Moreover, more fundamental work on polarization as such has been performed [8, 9], partially motivated by the fact that no individual term in a force field is an observable; in fact, the decomposition of the total energy is ambiguous, which has led to many different methods to decompose quantum-chemical interaction energies [10–12], of which the SAPT approaches may be the most rigorous [13].

In this paper, we present a fundamental study of the polarization in molecular dimers, with the purpose to investigate the approximations introduced in a standard polarization description. We employ a hierachy of theoretical methods, ranging from quantum-chemical supermolecular calculations to point-polarizability models. These allow us to examine in detail the effects of the various approximations implicit in standard polarizability models. We do not intend to develop a practical polarization model for force fields; instead, we aim at a detailed understanding of the physical process of polarization, which we believe is an important foundation for the development of sound and transferable polarizable force fields.

As a starting point, we consider the point-polarizability model, which describes the re-

sponse of each monomer in terms of distributed dipoles induced by the inhomogeneous electric field from the other molecule. This model involves a number of approximations:

- The induced density, i.e. the change in electron density upon polarization, is expanded in terms of dipoles, whereas in reality, the charge density has a finite extension, i.e. it is a *non-expanded density*.

- The induced dipoles within the same molecule do not influence each other directly, i.e. there is no explicit *intramolecular coupling*.

- Only the electric field in each point influences the response, i.e. not the field gradient or higher-order terms. Since the field is the linear term in the expansion of the electric potential, we call the missing effects *non-linear response*. This should not be confused with the non-linear effects that follow from the saturation of the response at high field strengths. The latter effects are of course also missing in the point-polarizability model (they require dipole hyperpolarizabilities to be described properly), but they are expected to be negligible for weakly charged molecules and will not be further discussed.

- The electrons of the two charge distributions are fermions and should therefore obey the Pauli principle, but the *Pauli effects* the electrons of one monomer has on the electrons of the other are not included.

- Finally, no charge can flow between the molecules, i.e. *charge transfer* has not been accounted for.

We investigate to what extent and in what way these five approximations contribute to the total error in a point-polarizability model.

These questions are not completely new. Pauli effects have been addressed in the discussions of the well-know Kitaura–Morokuma decomposition (KMD) [10] and its relation to the Restricted Virtual Space (RVS) model [11, 12, 14, 15]. The polarization energy in the KMD scheme is computed without any account of the Pauli effects, and it has been found that as the basis set increases in size, the polarization energy becomes more and more attractive and fails to converge, although the total energy converges. The reason for this is that a large basis set enables electrons on one monomer to penetrate into the electron cloud

of the other monomer, a space they are forbidden to occupy according to the Pauli principle. The RVS method includes Pauli effects and has a polarization term with better convergence properties. We will extend this discussion to see how the induced density is affected by the Pauli effects. To do this, we introduce a model, called CPOL, that is a generalization of the polarization term in the KMD scheme.

The CPOL model uses the true inhomogeneous electric field to polarize the monomers, so that the non-linear response is included, but the Pauli effects are absent. Furthermore, we have developed a model called LinCPOL, which has all features in common with CPOL, except that the electric potential is locally expanded up to the linear term. In contrast to the point-polarizability model, LinCPOL includes the intramolecular coupling in an essentially exact way. Hence, the difference between these two models enables us to analyze the effect of the intramolecular coupling and the non-expanded density. Note that the LinCPOL model is not intended to be a complete polarizability model for force fields, in contrast to other polarizability models that take both the coupling and charge flow into account, e.g. [16].

For all considered models, we analyze the errors in both the induction energy and the induced density. In fact, we will concentrate on the latter property, because it is an observable and therefore can be compared for all models. The induction energy, on the other hand, is not an observable, and therefore depends on the energy decomposition, especially at short intermolecular distances where coupling with the exchange–repulsion is inevitable. Another reason to focus on the induced density is its importance for the many-body contribution. It is the charge redistribution in a dimer of interacting molecules that causes the major change in the interaction with a third molecule. Thus, an accurate modeling of the induced density is a key requirement for a reliable polarizable force field. In principle, the pair-interaction term may be fine-tuned by empirical parameters to reproduce a given potential curve, but if the induced density is incorrect, three-body effects will be incorrect as well. Moreover, in our opinion, an accurate induced density is an indication of a physically sound polarization model, which will increase the likelihood to obtain transferable force-field parameters.

We present the various models and their relations in detail in the next section. There, we also describe how we analyze the induced densities. The results are presented in the subsequent section, where we give a detailed analysis of two particular systems (the water dimer and the formamide dimer), before a large data set is analyzed. Finally, we interpret the results in the context of force-field development.

II. METHODS

We consider the interaction between two molecules (monomers) A and B in vacuum. The geometries of the isolated monomers are kept fixed as the dimer is formed. Hence, no nuclear relaxation is considered, only electronic polarization. The wave functions, Ψ_A^0 and Ψ_B^0, of the isolated monomers are obtained by two separate quantum-chemical calculations at a given level of theory and with a specific basis set. When we in the following refer to *exact* quantities, we mean exact within the approximations inherent in this quantum-chemical calculation.

Let us take the polarization of monomer B by monomer A as an example (the polarization of monomer A is analogously defined). A key quantity in all models is the electrostatic potential ϕ_A^0 from Ψ_A^0 (or the charge distribution implied by it, ρ_A^0) in the region of monomer B. We call a model *classical* if this is the only information about monomer A, which is used in the calculation of the response of monomer B. All considered models are classical, except for the Restricted Virtual Space (RVS) method and the supermolecular quantum-chemical calculations. The electric potential (or field, defined as $\mathbf{F}_A^0 = -\nabla\phi_A^0$) from monomer A is the same in all models. Specifically, the potential used comes from the exact charge distribution implied by Ψ_A^0. What differs between the various models is the way in which the potential is applied, and how the response of monomer B is modeled. This construction warrants that there is no confusion in the analysis between the modeling of the response and the modeling of the charge distributions. The effects included in each model are also summarized in Table I. As a visual guide, the main idea of each model is depicted in Fig. 1.

Before the actual models are presented in detail, we introduce a common notation. The complete Hamiltonian for monomer B will in all models be of the form

$$H^{B(i)} = H^B_{intra} + V(\rho_i^A + n^A) \tag{1}$$

where H^B_{intra} contains all the intramolecular contributions to the energy and $V(\rho_i^A + n^A)$ is the perturbation from monomer A, which depends on the electron density ρ_i^A and the nuclear density n^A. For a classical model, the operator $V(\rho_i^A + n^A)$ simply stands for the operation "multiply with ϕ_A^i". If the charge distribution on monomer A is optimal for the unperturbed monomer, then index i on ρ_i^A is zero; if the charge distribution is optimal for the polarized state of A in the presence of B, then $i = \infty$. Given an electron density for

monomer B, this Hamiltonian gives the following energy:

$$H^{B(i)}[\rho_j^B] = H^B_{intra}[\rho_j^B] + V(\rho_i^A + n^A)[\rho_j^B] = E^B_{intra,j} - \int \phi^i_{A,nuc}(\vec{x})\rho_j^B(\vec{x})d\vec{x} - \int \phi^i_{A,el}(\vec{x})\rho_j^B(\vec{x})d\vec{x} \tag{2}$$

If a variational method is used to evaluate the energy, the following relation also holds for the intramolecular contributions:

$$E^B_{intra,0} \leq E^B_{intra,\infty} \tag{3}$$

The difference between $E^B_{intra,\infty}$ and $E^B_{intra,0}$, called ΔE^B_{intra}, is the energy cost (sometimes called self-energy) of polarization. We should finally note that the nuclear repulsion energy is not included in this notation. For the total energy, it is needed, but since we are interested in the induction energy, which is the additional energy that follows from electronic relaxation, the nuclear repulsion will not appear in these energy expressions.

A. Classical polarization

In the *exact classical polarization* (CPOL) model, the exact electric (Coulomb) potential from monomer A is applied as a perturbation in the quantum-chemical calculation of monomer B. Thus, Eq. 2 reduces to

$$H^{B(i)}[\rho_j^B] = E^B_{intra,j} - \sum_l Z_l^A \int \frac{\rho_j^B(\vec{x})}{|R_l^A - \vec{x}|}d\vec{x} + \int\int \frac{\rho_j^B(\vec{x})\rho_i^A(\vec{y})}{|\vec{x} - \vec{y}|}d\vec{x}d\vec{y} \tag{4}$$

where nuclear attraction (with the nuclear coordinates $\vec{R_l^A}$ and their respective valencies Z_l^A) and electron repulsion integrals are easily identified. From Eq. 4 it follows trivially that

$$V(\rho_i^A)[\rho_j^B] = V(\rho_j^B)[\rho_i^A] \tag{5}$$

Using the Hamiltonian $H^{B(0)}$, a polarized density for monomer B, ρ_1^B, is generated when the Schrödinger equation is solved with some quantum-chemical method. The corresponding equation is solved for monomer A in its basis set and ρ_1^A is obtained. This procedure is then iterated to self-consistency, since the density of monomer B will give rise to a new density on monomer A. From the expectation values of the Hamiltonians of monomers A and B, the energy in a given iteration m is defined as

$$E_m = H^{B(m)}[\rho_m^B] - E^A_{intra,0} + H^{A(m)}[\rho_m^A] - E^B_{intra,0} - V(\rho_m^A)[\rho_m^B] \tag{6}$$

and the induction energy as

$$E_{ind} = E_\infty - E_0 = \Delta E^A_{intra} + \Delta E^B_{intra} + V(n^A)[\rho^B_\infty - \rho^B_0] + V(n^B)[\rho^A_\infty - \rho^A_0] + V(\rho^B_\infty)[\rho^A_\infty] - V(\rho^B_0)[\rho^A_0] \quad (7)$$

Observe that the last term in Eq. 6 is needed since otherwise the electron–electron repulsion will be double-counted.

The CPOL model has been used extensively before, although rarely in the present formulation. For example, the polarization energy term in the KMD scheme can be shown to be equivalent to the induction energy obtained with CPOL if the Hartree–Fock (HF) method is used [10, 15, 17, 18]. The Fragment Molecular Orbital (FMO) method, which is designed to divide a large quantum-chemical calculation into many smaller approximately coupled calculations, also uses a form of CPOL to estimate the electronic coupling between the various fragments [19]. This leads to a natural generalization of KMD for large systems [20]. The present formulation of CPOL has the advantage that it is independent of the quantum-chemical method, and can easily be extended beyond HF. The MP2 and the CASSCF method are two examples that we have implemented. In the present study, however, all calculations are done with HF to enable comparisons with the results with the RVS method (*vide infra*).

Finally, with reference to Table I, we observe that (1) the CPOL method contains no charge transfer between the two monomers, because their charge densities are described with separate basis sets, (2) no Pauli effects are enforced between the monomers, (3) the potential on B from A can be non-linear, (4) the polarization of B is solved fully coupled with all other terms for monomer B, i.e. any intramolecular coupling in the polarization is included, and (5) the charge density is described in terms of Gaussian basis functions, i.e. no expansion in discrete points is made.

B. Linearized CPOL

In the *linearized classical polarization* (LinCPOL) model, the potential ϕ^i_A is Taylor-expanded to first order at particular points (centers) in monomer B, here chosen as the coordinates of each nucleus and the midpoint between each pair of nuclei (i.e. not only at

covalent bonds). For each center v at $\mathbf{r}_v$, this defines an approximate linearized potential:

$$\tilde{\phi}^i_{A,v}(\mathbf{r}) = \phi^i_A(\mathbf{r}_v) - \mathbf{F}^i_A(\mathbf{r}_v) \cdot (\mathbf{r} - \mathbf{r}_v) \tag{8}$$

It is not straight-forward to apply these local perturbations to monomer B, however. One natural choice would be to define a spatial region belonging to each center, construct a continuous linearized potential, and then numerically integrate the potential over all pair-products of basis functions. However, the selection of regions introduces an ambiguity into the method. Therefore, we have chosen another approach, namely to partition the pair-products into groups depending on which atom each basis function is centered on, and for a given group apply the same linear perturbation (Eq. 8) throughout the whole space. The ambiguity now resides only in the selection of suitable basis functions. The ordinary atomic orbital (AO)-basis set is local, but non-orthogonal. Therefore the local perturbations would mix in the optimization of the wave function, so that the overall perturbation would not be well-defined. To avoid this, we instead use the LoProp basis [7]. In the LoProp method, a basis is obtained as a transformation of the ordinary AO-basis, $\chi = \mathbf{T}\theta$, where the transformation matrix is constructed to make the functions χ both orthogonal and local. In the original application of LoProp, this basis is used to distribute molecular properties, whereas in the present study we use it to properly apply the locally linearized perturbation. The LinCPOL perturbation from monomer A to monomer B is written in bracket notation as

$$V(\rho^A_i + n^A) = \sum_{r,s} |\chi_r\rangle\langle\chi_r|\tilde{\phi}^i_{A,[r,s]}|\chi_s\rangle\langle\chi_s| \tag{9}$$

where $\tilde{\phi}^i$ is defined as in Eq. 8 and $[r, s]$ denotes the center of the basis function pair (χ_r, χ_s), which is either an atom center if χ_r and χ_s are centered on the same atom, or the bond center if χ_r and χ_s are centered on different atoms. The construction of the LinCPOL perturbation matrix in LoProp basis becomes a summation of matrices for all the various linearized perturbations, where all elements that do not correspond to the right basis function pair for that particular perturbation are set to zero. To use the LinCPOL perturbation in an ordinary quantum chemical calculation, the perturbation matrix in LoProp basis is transformed back to the ordinary AO-basis, $V' = (T^{-1})^t V T^{-1}$. Finally, the same type of iterative procedure has to be used as in CPOL, with the difference that the electric potential the monomers "feel", is of the locally linearized type described above.

Since the final potential in LinCPOL is not exactly the Coulomb potential from the other charge distribution, the proper definition of the induction energy is not as easily obtained as in CPOL, where it was motivated by physical principles. The problem lies in that the energy evaluation has to be consistent with the potential in the perturbation. We use the following expression:

$$\begin{aligned} E_{ind} &= \frac{1}{2}\Big(H^{A(\infty)}[\rho_\infty^A] + H^{A(0)}[\rho_\infty^A] - H^{A(\infty)}[\rho_0^A] - H^{A(0)}[\rho_0^A] \\ &+ H^{B(\infty)}[\rho_\infty^B] + H^{B(0)}[\rho_\infty^B] - H^{B(\infty)}[\rho_0^B] - H^{B(0)}[\rho_0^B]\Big) \end{aligned} \tag{10}$$

which in fact is equivalent to Eq. 7 in the special case that the perturbation is the Coulomb potential. The equality is easily shown by inserting Eq. 2 into Eq. 10. This gives:

$$\begin{aligned} E_{ind} &= \Delta E_{intra}^A + \Delta E_{intra}^B + (V(n^A)[\rho_\infty^B - \rho_0^B] + V(n^B)[\rho_\infty^A - \rho_0^A]) \\ &+ \frac{1}{2}\Big(V(\rho_\infty^B)[\rho_\infty^A] + V(\rho_0^B)[\rho_\infty^A] - V(\rho_\infty^B)[\rho_0^A] - V(\rho_0^B)[\rho_0^A] \\ &+ V(\rho_\infty^A)[\rho_\infty^B] + V(\rho_0^A)[\rho_\infty^B] - V(\rho_\infty^A)[\rho_0^B] - V(\rho_0^A)[\rho_0^B]\Big) \end{aligned} \tag{11}$$

which reduces to Eq. 7 because of the symmetry relation (Eq. 5) valid for CPOL. For an approximate perturbation (such as LinCPOL), where Eq. 5 does not hold, Eq. 10 may be seen as a way of taking half of the energy contribution from one form of the perturbation (e.g. A to B) and the other half from the other form (B to A). In other words, Eq. 10 is more general and can be used for both CPOL and LinCPOL, but it is less transparent than Eq. 7 for the CPOL method.

Finally, with reference to Table I, we observe that in LinCPOL, neither (1) charge transfer nor (2) Pauli effects are included for the same reason as in CPOL. Moreover, the local Taylor expansions is truncated at the linear term, therefore locally, (3) there is no non-linear response in LinCPOL, unlike in CPOL. Since the electronic wave function is obtained through the same type of coupled equations and the Gaussian basis functions are used, both (4) intramolecular coupling is accounted for and (5) the density is non-expanded in the same sense as in CPOL.

C. Polarizabilities

The *polarizability* model is commonly used in polarizable molecular mechanics force fields (e.g. [21]). Monomer B is modeled as a set of distributed anisotropic dipole polarizabilities.

The response is obtained as

$$\boldsymbol{\mu}_{B,v}^{i+1} = \boldsymbol{\alpha}_{B,v} \cdot \mathbf{F}_A^i(\mathbf{r}_v) \tag{12}$$

where $\boldsymbol{\alpha}_{B,v}$ is the polarizability tensor for center v in monomer B, $\boldsymbol{\mu}_{B,v}^i$ is the induced dipole and $\mathbf{F}_A^i(\mathbf{r}_v)$ is as before the electric field from monomer A at center v. Just as in the two previous models, the equations are iterated to self-consistency, because the induced dipoles themselves contribute to the electric field. Observe that the field $\mathbf{F}_A^0$ from the static charge distribution of monomer A is not expanded in a set of multipoles, instead it is the exact electric field (see the discussion of Fig. 1).

In this study, standard LoProp [7] polarizabilities are used, located at the atomic nuclei and at the midpoints of covalent bonds. The polarizabilities are obtained by distributing the molecular response to homogeneous electric fields using the localized basis described in connection with the LinCPOL model. The contributions to the polarizability from centers between atoms not covalently bonded, are dispersed onto the two relevant atoms following the standard procedure [7]. We have found that such an approach leads to an insignificant difference compared to a non-dispersed approach.

There is no explicit coupling between the polarizabilities within the same monomer. Instead, as seen from the equation above, a given polarizability in monomer B is only directly coupled to the electric field from monomer A (and vice versa). This usage of the polarizabilities is dictated by their derivation using a homogenous electric field, because a given polarizability is strictly correct only under the assumption that the rest of the molecule is responding to the same electric field. Thus, there is an implicit coupling in the model, which is correct when a homogenous field is applied but only approximative when the field is inhomogenous. The approximation is expected to be good for very small molecules but crude for e.g. macromolecules, where the field in different regions of the molecule does not need to be similar at all. For medium-sized molecules, the accuracy of the approximation has not been well tested.

The polarizability model can be seen as an approximation to the LinCPOL model. The applied electrostatic potential is in fact identical in the two models. Observe that the constant term $\phi_A^i(\mathbf{r}_v)$, however, is not relevant for the polarizability model, since the charge can not flow between different centers. In the LinCPOL model, the perturbation is applied in a quantum-chemical calculation, whereas in the polarizability model the response is simply computed by Eq. 12. This may appear to be a dramatic difference, but in the latter model the

linear perturbations have been applied "before-hand" in the computation of $\alpha_{B,v}$. Therefore, with reference to Table I, neither model have (1) charge-transfer, (2) Pauli effects, or (3) non-linear response terms, but in the polarizability model, (4) the polarizabilities are not intramolecularly coupled. Also, the response in the polarizability model is (5) not obtained as an induced density, but as a collection of induced dipoles, which may be seen as a multipole expansion of some induced density. Observe also that with this perspective, the multipole expansion is truncated at dipolar order.

D. Restricted virtual space

In the *Restricted Virtual Space* (RVS) model [11], the exact electrostatic potential is applied, but in contrast to the CPOL method, Pauli effects are included. This is done by restricting the space in which the perturbed wave function of monomer B is optimized, forcing it to be orthogonal to the occupied space of monomer A. The optimization is consequently done in the dimer basis set, but with the occupied space of monomer A frozen, and the virtual space of monomer A removed from the variational space. If the virtual space of monomer A is included, charge transfer can be described with RVS, but in the present application, this term is not considered. As far as our analysis is concerned, RVS is completely equivalent to Constrained Space Orbital Variation (CSOV) [12]. To the best of our knowledge, RVS is limited to single-configurational methods (in principle only HF but tested also for density functional theory [22]), which sets it apart from the other models in the present study.

In the other models, the polarization is solved iteratively, to include *higher-order effects* of the polarization. This procedure is necessary to be able to directly compare to the supermolecular results. RVS, on the other hand, do not include these higher-order effects. In RVS, the occupied orbitals of the perturbing molecule are frozen at their value in the isolated monomer. Thus, the perturbing electric potential is always equal to the electric potential from an isolated molecule and not from the self-consistently polarized one. This corresponds to only performing one iteration in the CPOL or LinCPOL method. Higher-order effects are not negligible (for example, they give ≈ 14 % of the induced dipole moment in the water dimer); therefore, the pure RVS results are of limited interest for quantitative comparisons. We made an attempt to include higher-order by simply updating the orbitals of

the perturbing molecule. However, we experienced convergence problems, probably reflecting a deeper difficulty with the actual iterative solution to the problem. To avoid the problem, we instead use an empirical strategy: To the RVS induction energy, the difference between the corresponding fully iterated LinCPOL quantity and the same quantity in the first iteration of LinCPOL, is added. The LinCPOL model was chosen because the CPOL model tends to overestimate the magnitude of the polarization and thereby possibly also the iterative effects, as will be shown below.

The induced density is not obtained in the standard RVS method and needs to be properly defined. The first step in an RVS calculation on monomer B is an orthogonalization of the orbitals of monomer B to the occupied orbitals of monomer A. The HF energy thus obtained is the standard Heitler–London energy, which includes the electrostatic and zeroth-order exchange–repulsion energies. In the RVS energy decomposition, this energy is reported as a separate quantity and the polarization energy of monomer B is defined relative to this energy. However, in this study we compare the induced densities obtained with several methods, some of which include Pauli effects and some of which do not. Therefore, the correspondence between the induced densities is highest if the density change caused by the orthogonalization is included in the induced density ρ_{RVS}, i.e.

$$\rho_{RVS} = \rho_{ORTH} + \rho_{POLA} + \rho_{POLB} + \rho_{\infty}^{LinCPOL} - \rho_{1}^{LinCPOL} \tag{13}$$

Likewise, we define the induction energy E_{RVS} as

$$E_{RVS} = E_{POLA} + E_{POLB} + E_{\infty}^{LinCPOL} - E_{1}^{LinCPOL} \tag{14}$$

There is always a risk that mixing different models may confuse causes and effects. However, from our calculations, these definitions seem to work well, and indeed provide us with a model that includes Pauli effects, but still has much in common with the other methods.

E. Computed quantities

Once all response models are defined, we also need well-defined methods to compare them. There are several ways to quantify how similar an induced density is to that of the reference method (supermolecular). The most basic approach is to compare the total induced dipole moment of the dimer. It is always uniquely defined, because even if the molecules

are charged, the induced density always has total charge equal to zero. To reduce this to a scalar quantity, we only discuss the component of the dipole moment in the direction of an interaction axis, which is defined for each dimer. For hydrogen-bonded dimers, it is the line connecting the proton and the proton acceptor.

Although the induced dipole is well-defined, it does not provide any detailed information about the charge distribution. For more details, three other quantities are defined below, which essentially give the same information, but all have their limitations.

The second-moment tensor is the expectation value of $\mathbf{r}\otimes\mathbf{r}$. It differs from the quadrupole moment in that the spherical component, related to the size of the induced charge cloud, has not been removed. Thus, the second moment describes more directly the induced charge distribution. As in the dipole moment case, we only discuss the component of the second moment related to the interaction direction $\hat{\mathbf{r}}_z$, defined as

$$Q_{ind,zz} = -\int \rho_{ind}(\mathbf{r}) \left[(\mathbf{r}-\mathbf{r}_0)\cdot\hat{\mathbf{r}}_z\right]^2 d\mathbf{r} \tag{15}$$

This quantity gives a simple picture of the polarization. For example, if $Q_{ind,zz}$ for a model is larger than that of the reference, the induced density between the monomers has been stretched too far, or in other words, too much polarization has been obtained. However, unlike the induced dipole, $Q_{ind,zz}$ depends on the choice of origin $\mathbf{r}_0$. This implies that, even if the origin is fixed, $Q_{ind,zz}$ obtained for two different models are strictly comparable only if the induced dipole moments are equal. We have noticed, however, that the dependence on the origin is weak, as long as it is placed in a reasonable region on the interaction axis between the monomers. We conclude that the error introduced by the origin-dependence in $Q_{ind,zz}$ is small, and therefore this quantity mirrors the real effect in the induced density sufficiently well to be useful in the analysis.

Another way to describe the charge distribution in more detail is to decompose the total induced dipole moment into monomer contributions. This requires a method to partition the induced density into monomer contributions. For the models that work with the monomer basis set, this is easy. But for the supermolecular calculations, such a partitioning is not straight-forward. This problem is an instance of the more general problem of partitioning the total electron density into local contributions, although it is greatly simplified here by the fact that the density should be partitioned over non-covalent bonds only. Several methods have been proposed for solving the general problem (see e.g. [23–29]). Again, we use the LoProp

method [7] for the partitioning. It should be noted that the LoProp method guarantees that no charge is allocated in the bonds. In the Mulliken population analysis, such charge density has to be subsequently divided among the contributing atoms. Recently, it was shown that the LoProp method performs better than Mulliken-based methods when diffuse basis sets are used [30]. The monomer dipole moments are determined in the following manner. First, LoProp is used to expand the total induced density into distributed multipole moments: charges and dipoles at the nuclei, and dipoles at the bond midpoints. The multipoles that belong to each monomer are used to calculate the dipole moment of each monomer. In general, the two monomer dipole moments do not add up to the total induced moment, so a rest term, $\boldsymbol{\mu}_{ind}^{AB}$ is obtained.

$$\boldsymbol{\mu}_{ind} = \boldsymbol{\mu}_{ind}^{A} + \boldsymbol{\mu}_{ind}^{B} + \boldsymbol{\mu}_{ind}^{AB} \tag{16}$$

It is easy to show that $\boldsymbol{\mu}_{ind}^{AB}$ vanishes if the induced charges of each monomer add up to zero. Thus, $\boldsymbol{\mu}_{ind}^{AB}$ is related to the transfer of charge from one monomer to the other. For simplicity, this term is not explicitly reported in the results, rather it is absorbed into the monomer dipole moments, using the expression

$$\tilde{\boldsymbol{\mu}}_{ind}^{A} = \boldsymbol{\mu}_{ind}^{A} + \frac{1}{2}\boldsymbol{\mu}_{ind}^{AB} \tag{17}$$

and similarly for $\tilde{\boldsymbol{\mu}}_{ind}^{B}$. The justification for this is that the rest term is small and that the decomposition is in any case somewhat arbitrary. However, in the discussion of the basis-set superposition error (BSSE) below, we explicitly consider all three terms.

Finally, for more quantitative conclusions, the induced electrostatic potential for each model is compared to the supermolecular calculation. The comparison is done by selecting $10,000$ points around the dimer and calculating the mean absolute difference between the potential to the induced density for the given model and for the supermolecular calculation. The points are randomly distributed within 8 Å of *some* nucleus but not within R_i of *any* nucleus i, where R_i is an approximative element-dependent interaction distance [44]. A new set of points is generated for each intermolecular separation. The mean absolute difference is renormalized for each separation by division with the mean absolute value of the induced supermolecular potential. This gives the mean error of the model m relative to a model with zero induced potential (e.g. a non-polarizable force field) as

$$ERR_m = \frac{\sum_i |\phi_m(x_i) - \phi_{sup}(x_i)|}{\sum_i |\phi_{sup}(x_i)|} \tag{18}$$

where ϕ is the electrostatic potential caused by the induced density ρ_{ind}.

F. Basis set superposition errors

All supermolecular results are BSSE-corrected. For energies, we use the normal counterpoise correction. For induced densities, we use the analogous expression

$$\rho_{ind} = \rho_{AB} - \rho_{A(B)} - \rho_{B(A)} \tag{19}$$

where ρ_{AB} denotes the dimer density and $\rho_{A(B)}$ denotes the charge density of monomer A calculated in the dimer basis set [31]. There have been previous studies on the effect of BSSE on the charge density [31–35].

When the induced dipole moment is decomposed into the two monomer dipole moments and a rest term (as described above), a further BSSE issue arises, which is related to the LoProp decomposition. Consider the following three quantities.

1 Isolated monomer density expanded in the monomer LoProp basis

2 Isolated monomer density expanded in the dimer LoProp basis

3 Dimer-basis-calculated monomer density, i.e. $\rho_{A(B)}$, expanded in the dimer LoProp basis

If the LoProp basis was perfectly localized, then quantities 1 and 2 would have been the same and the difference between 3 and 1 would have been the BSSE. However, the LoProp basis is orthogonal and thus the localization is compromised. This implies that even if the monomer is described completely within its own basis set, a small part of the density is attributed to the other monomer. Fortunately, this problem fully disappears if a proper counterpoise correction is applied. As an example of the effect, numerical values are given for the BSSE in Table II for a water dimer at O$\cdots$H distance 2.01 Å. The BSSE is divided into a decomposition BSSE (difference between 2 and 1, denoted D-BSSE) and an ordinary BSSE (difference between 3 and 2). As can be seen, the D-BSSE is the largest in magnitude. Note that, by definition, only the ordinary BSSE affects the total induced dipole moment.

A comparison with the situation in a Mulliken-type of analysis may be helpful. Because the standard AO basis is localized, there would be nothing corresponding to the D-BSSE

in a Mulliken decomposition. However, the non-orthogonality of the basis would make the decomposition of the induced dipole moment into monomer contributions more ambiguous and more dependent on the basis set. For example, one would probably obtain a rather large charge between the monomers (whereas it is identically zero in the LoProp method), and the further division of that charge among the two monomers would significantly affect the results. Therefore, we believe that with a careful treatment of the D-BSSE, LoProp is one of the best possible intermolecular decomposition methods [45].

A final note on BSSE regards the RVS method. Because the virtual space of the perturbing molecule is not included in the calculation of RVS polarization, there is no (or extremely little) ordinary BSSE in the calculations, and thus no counterpoise correction has been made. However, the D-BSSE is present in exactly the same way as in the supermolecular calculation, and it has been properly accounted for.

G. Computational details

The HF method was employed throughout this study. However, we stress that except for RVS, the models are not limited to HF, and we do not expect that the conclusions would change if a different method had been used. Two different basis sets were used: ANO-L [36] for the small dimers studied in detail and cc-pVTZ [37, 38] for the larger dimers studied in a more statistical manner. The two basis sets contain the same number of basis functions (3s2p1d for hydrogen, 4s3p2d1f for carbon, nitrogen, and oxygen, and 5s4p2d1f for sulfur). The cc-pVTZ is used for the large data sets because it has been used previously with these data sets [39]. We tested it also for the small dimer calculations and observed no qualitative differences.

All monomer geometries were optimized using the B3LYP/6-311+G(2d,2p) method, except for water in the water dimer, which was kept at an experimental structure with bond distance $\mathrm{O-H} = 0.97$ Å and $\wedge\mathrm{HOH} = 103°$. The geometry of the formamide monomer had bond distances $\mathrm{C-O} = 1.21$ Å, $\mathrm{C-N} = 1.36$ Å, and $\mathrm{N-H} = 1.00$ Å.

The geometry of the water dimer was chosen so that the hydrogen bond $\mathrm{O-H\cdots O}$ forms a straight line (in the z direction) and these atoms together with the hydrogens of the proton acceptor lie in the same plane. The formamide dimer was optimized at the B3LYP/6-311+G(2d,2p) level, giving a doubly hydrogen-bonded complex with C_{2h} symmetry. The

interaction axis z is defined as the vector between the nitrogen atom of one molecule and the oxygen atom of the other. The interaction distance was varied by displacing one monomer along that axis, keeping the $\mathrm{C}-\mathrm{O}\cdots\mathrm{N}$ angle fixed at 120.3°. This gives a hydrogen bond angle ($\mathrm{O}\cdots\mathrm{H}-\mathrm{N}$) between 173° and 174° for the considered distances. The geometries included in the large data set are optimized at the same level. All details about the large data set, including coordinates, may be found in ref. [39], from which the molecule–water (with the seven halogen-containing interactions removed) and molecule–molecule data sets were selected.

All calculations were done with the program package MOLCAS [40, 41].

III. RESULTS

A. Water dimer

For the water dimer, we studied a non-optimized hydrogen bond. The proton donor is denoted A and the acceptor is denoted B. The induced monomer dipole moments (in the direction of the interaction axis) obtained with the various models are shown in Fig. 2. The most striking feature is the qualitatively different behavior of the CPOL model for monomer A. Since this behavior is typical for all studied systems, we will explain this deviation in detail for the water dimer. The electrostatic potential from monomer B is shown in Fig. 3 with a typical position of monomer A superimposed on the curve. In the linear models, the linearized potential (shown with a green line) is applied to the hydrogen atom. This is an electric field in the same direction as if monomer B was modeled as an ideal dipole, which pushes the electrons of the hydrogen atom of momoner A towards the oxygen atom of the same monomer. In the CPOL model, on the other hand, part of the density of monomer A experiences a steep rise in electrostatic potential at around 1 Å caused by charge penetration. Consequently, there is an electron-attracting effect at short separations, causing the decrease in induced dipole moment. Comparison to the supermolecular results show that this effect becomes artificially large because of the lack of Pauli effects.

Another point to note from Fig. 2 is that the LinCPOL and polarizability models give similar results, despite the many additional approximations done in the polarizability model. Moreover, the results with these linear models are more similar to the supermolecular result

than the CPOL result is. This indicates that the error from the lack of non-linear response to some extent cancels the error from the neglect of Pauli effects.

The total induced dipole moment and second moment, which are alternative quantities to analyze the induced density, are shown in Fig. 4. The differences in the total induced dipole moment between the various models are much less pronounced than the corresponding differences in induced monomer dipole moments. On the other hand, larger differences are seen again in the total second moment. The classical models overestimate the second moment at closer separations, i.e. give a too contracted induced density, probably owing to their lack of Pauli effects. The RVS model underestimates the second moment slightly, probably because of the lack of charge transfer.

As a quantitative test of the significance of the induced density deviations for the interaction with other molecules, the induced electrostatic potential from each model was compared to the supermolecular calculation as described in the Methods section. The relative mean error for various models as a function of the interaction distance is shown in Fig. 5. At long distances, the error in the CPOL model approaches zero, which follows from the exact nature of CPOL when Pauli effects can be neglected. For the linear models, a constant relative error is reached, reflecting the approximations involved (of course the *absolute* error goes towards zero). As the Pauli effects grow larger when the molecules approach each other, the CPOL error increases steeply. However, the error of the LinCPOL model increases more slowly, and thus this model gives lower error at distances closer than 2.8 Å. The RVS model gives the lowest error across the whole interaction range, which is expected because of its inclusion of Pauli effects. The increase of the RVS error at ~ 3 Å is attributed to the lack of charge transfer and the retardation of this increase at short distances is probably due to fortuitous error cancellation with the inaccurate iterative correction. Without the iterative correction, however, the RVS model gives an error as high as the CPOL model.

The difference between LinCPOL and the polarizability model has been analyzed further. In principle, the larger error for the polarizability model can be caused by the lack of intramolecular coupling (apart from the implicit coupling assuming homogenous fields), the lack of local induced charges, the truncation of the multipole expansion of the induced density, the overlap error from the multipole expansion, or any combination of these four factors. To test the influence of the truncation of the multipole expansion, the final density from the LinCPOL model was expanded with LoProp, in one case up to dipoles and in another case

up to quadrupoles. The relative errors for these expansions are also shown in Fig. 5. Interestingly, the dipole expansion almost perfectly reproduces the results with polarizabilities, whereas the quadrupole expansion gives essentially the same results as the non-expanded induced density. Thus, we conclude that the whole difference comes from the truncation of the multipole expansion (or more specifically from the lack of local quadrupoles) and that the other factors have a negligible effect. This would also explain the difference between the polarizability model and the LinCPOL model in Fig. 4.

Finally, we turn to the estimation of the induction energy. Induction energies obtained with the various models are shown in Fig. 6. It can be seen that the CPOL model gives substantially more attraction than the LinCPOL model, which is expected from the above discussed difference in the density. The RVS polarization behaves more like the linear models. This suggests that also for energies, the non-linear response (difference between CPOL and LinCPOL) cancels the Pauli effects (difference between RVS and CPOL).

B. Formamide dimer

Next, we study the optimal, doubly hydrogen-bonded structure of the formamide dimer. By symmetry, the first nonzero induced multipole moment is the induced second moment, which is therefore independent of the origin. The zz component of the second moment is shown in Fig. 7. Clearly, the linear models give better agreement with the supermolecular moment than CPOL, but they erroneously predict an increase in second moment, i.e. an overall charge contraction, when the distance is shortened. The induced monomer dipole moment in the direction of the interaction axis follows the same pattern (not shown).

The induced electrostatic potential around the dimer was analyzed as described in the Methods section, and the result is shown in Fig. 8. All models give significantly larger relative errors than for the water dimer. For example, the error in the LinCPOL model for interaction distances close to the optimal dimer geometry approaches 100 %, whereas it is less than 10 % for the water dimer. The absolute errors are more similar, but there is still ~ 3 times larger error for the formamide dimer. Another interesting result is that the conclusion from the water dimer, that the difference between the LinCPOL and Polarizability models is caused by the lack of local quadrupoles in the latter, holds also for the formamide dimer. Thus, intramolecular coupling is negligible also for this system.

Scatter plots for the correlation between the electric potential from the supermolecular calculations and CPOL as well as LinCPOL are shown in Fig. 9. For CPOL, there is no significant correlation. In fact, a model that always gives zero induced potential (e.g. a non-polarizable force field) would be a better description. With the LinCPOL model, a correlation is seen, but the spread is large and there is a systematic error at negative values of the electric potential. The polarizability model has a similar correlation, but a slightly larger deviation (result not shown). It should be noted that the large errors observed with the linear models are not caused by the approximations inherit in the localization method for the polarizabilities. A related study [30], which used homogeneous fields, showed that the errors due to the localization is at least one order of magnitude smaller than those observed in this study. Therefore, we attribute the large errors to the neglect of Pauli effects, which were not tested in the previous study. In situations with weaker electric fields, such as in the water dimer case, the cancellation between the neglect of Pauli effects and the lack of nonlinear effects works reasonably well. However, when the fields are strong, as in this case, the Pauli effects are more important and the cancellation becomes poor, although the final result is still significantly better than the CPOL model.

Finally, induction energies obtained with the various models are shown in Fig. 10. Again, the LinCPOL and polarizability models give significantly less attractive energies than the CPOL model. Building on the results for the electrostatic potential, it is reasonable to assign the difference between the LinCPOL and the polarizability models to the inclusion of higher-order induced multipoles in the LinCPOL model. The RVS and polarizability models incidentally give almost identical energies.

C. Large data set

In the previous sections, two special cases were analyzed in detail. For the water dimer, LinCPOL is much better than CPOL, because the cancellation between the neglect of Pauli effects and lack of non-linear effects is rather good, whereas for the formamide dimer, this is not the case. In this section, we consider a large set of interactions to try to determine which behavior is most common.

We have chosen a set of interactions relevant to simulations of proteins, which is an important usage of molecular mechanics force fields and therefore an interesting test case.

The set consists of 64 optimized geometries of amino acid side chains (and a few other organic molecules) interacting with water, primarily by regular hydrogen bonds, but also by $\mathrm{O-H\cdots\pi}$ and $\mathrm{C-H\cdots O}$ interactions, as well as 12 optimized geometries of the same molecules interacting with each other through regular hydrogen bonds.

For each structure and each polarization model, we compute four quantities: the error in the induced monomer dipole moment relative to the supermolecular calculation for the two monomers, the relative error in the induced electrostatic potential around the dimer (sampled in the same way as for the formamide and water dimers), and finally the estimate of the induction energy obtained with the model. The full results are listed as supplementary material [46]. For computational reasons, only the LinCPOL and polarizability models are tested for the geometries not including water. The values for the water dimer (at $\mathrm{O\cdots H}$ separation 1.96 Å) and formamide dimer (at $\mathrm{O\cdots H}$ separation 1.87 Å) are also given in the supplemental material for comparison.

A graphical summary of the results is given in Fig. 11, in which the errors in induced monomer dipole moments for each model (relative to the reference supermolecular calculation) are plotted against the reference induced dipole moments. For the CPOL model, the errors in induced monomer dipole moments always have the same sign, namely negative for monomer A and positive for monomer B. This means that the CPOL density is always over-polarized, or that too much density of monomer A is displaced towards monomer B and vice versa (the z direction is defined as the interaction direction from monomer A to monomer B).

The errors are smaller in the LinCPOL and polarizability models, which among themselves give similar results, thus confirming that the two models are physically similar. The lower errors for the linear models can be attributed to a systematic cancellation of error between the lack of Pauli effects and the neglect of non-linear effects. However, with a few exceptions, the errors in the induced monomer dipole moments for the linear models have the same sign as in the CPOL model. In addition, the few monomer dipole moments that deviate from this rule are all close to zero and thus within the uncertainty of the dipole moment decomposition method. This indicates that the cancellation is in almost all cases biased such that the lack of Pauli effects has a greater impact than the neglect of non-linear effects. Hence, over-polarization is observed also for the LinCPOL model. This is in line with earlier research [42], which has suggested that diffuse functions should be avoided when computing

polarizabilities in order to capture some of the Pauli effects present in the condensed state (for a more thorough discussion, see ref. [43]). However, we prefer to see the problem as a limitation of the polarization model as such. Finally, the RVS model gives monomer dipole moments that do not follow the same trends, thus indicating that a fundamentally different physics is obtained when Pauli effects are explicitly included.

The distribution of the error in the induced electrostatic potential for the whole set of interactions is shown in Fig. 12 as a histogram for each considered model. It can be seen that the error with the CPOL model is rather evenly distributed in the range up to 25 %, whereas with the LinCPOL model, only a few complexes give errors larger than 10 %, although the maximum error is approximately the same. This shows that the LinCPOL model works well in most cases, but that the performance is not guaranteed, which is of course an expected feature for all instances of error cancellation. The polarizability model gives similar results, with slightly more failures than the LinCPOL model, whereas the RVS model gives significantly better results, but also has a few problematic cases. The average errors in induced electrostatic potential in Fig. 12 are 16.0 %, 9.4 %, 10.1 % and 6.5 % for the CPOL, LinCPOL, Polarizability, and RVS model, respectively. It is interesting to note that the errors with the linear models are around 10 %, i.e. similar to the case of the water dimer. This means that the much larger errors observed for the formamide dimer are exceptional among the cases we have considered.

As we mentioned previously, there is no unambiguous definition of the induction energy. In particular, RVS uses a different definition, since its starting point is the Heitler–London complex, which already has been perturbed by the orthogonalization. We would therefore expect that, at short distances, the stronger coupling of repulsion and polarization will exaggerate the difference between RVS and the other models. The most noticeable example of this is a complex between acetate and water in the test set, with an O$\cdots$H distance of only 1.67 Å, which actually has a more attractive induction energy in RVS than in CPOL, despite the over-polarization of the latter model. There are other interactions for which CPOL is more attractive than RVS, but most energies are quite similar. This clearly shows that the decomposition of the total energy in individual terms can be done in many ways, with different relations between the induction term and the observable induced density.

Owing to the similar behavior of the LinCPOL and polarizability models in the previous results, we analyze this pair in greater detail. In Fig. 13, the distribution of the relative

potential difference as well as the energy difference between the two models are shown. The relative difference in the induced electrostatic potential is usually within 10 % and the difference in energy is usually within 0.5 kJ/mol. Among the nine structures that show larger errors in the electric potential, only two are regular hydrogen bonds, whereas the rest are $H \cdots \pi$ interactions. On the other hand, among the seven structures that show an absolute energy difference greater than 0.5 kJ/mol, most are strong hydrogen bonds, with the HIE–HIE complex giving the largest difference (1.2 kJ/mol). Further investigation of the potential difference for this complex shows that only a fifth of the relative error comes from truncation of the multipole expansion. Thus, we conclude that for this complex, there is most likely an effect of the neglect of intramolecular coupling of polarizabilities. Similar results are obtained for the other complexes, with a general trend that larger molecules (especially aromatic ones) show a larger effect of intramolecular polarization. Still, this effect is rather small in all considered cases. Thus, our results suggest that the different behavior of the LinCPOL and polarizability model is mostly of theoretical interest; a maximum error in the induction energy of 1.2 kJ/mol is most often negligible compared to the error in other terms of the total interaction energy.

IV. CONCLUSIONS

We have compared the induced charge density obtained by a range of models with the density from supermolecular calculations. The evaluation has been performed for a large data set of dimers relevant for biomolecular simulations. The purpose was to investigate on a fundamental level how different approximations of the electronic response to the presence of another molecule influence the final result.

We clearly establish that the simple point-polarizability model, despite its drastic approximations, works remarkably well for most applications because of a cancellation of errors from the neglect of both Pauli effects and non-linear response. If the non-linear response is included, but the Pauli effects still are neglected, the results deteriorate significantly. However, the cancellation is not perfect, and the Pauli effects seem to be quantitatively more important than the non-linear response. Moreover, in a few cases the cancellation is rather poor. Therefore, there are clearly limitations in the accuracy and transferability of a point-polarizability model.

In the analysis of the magnitude of the different errors, we find that the two approximations already mentioned are the most important. The lack of intramolecular coupling of polarizabilities seems to be unimportant for most molecules in the data set, whereas the truncation of the response at the dipole level has a noticeable effect. For some of the largest molecules in the data set, the lack of intramolecular coupling has a larger effect, but it is still small compared to the effect of the two major approximations. If significantly larger molecules are considered, however, we expect the effect to be important.

In conclusion, the point-polarizability model could be slightly improved by including locally induced quadrupoles and intramolecular coupling between the induced dipoles. However, to significantly improve polarization models systematically, special attention should be paid to (1) the Pauli effects, which restrict the relaxation of the charge distribution under the influence of the other molecule, and (2) the non-linear effects that basically account for the fact that the charge distribution extend over an interval of the inhomogeneous electric potential. It is important to note that these two effects must be improved simultaneously. A possible way to include the Pauli effects is with a pseudo-potential that models the Pauli effect as an energy penalty depending on the overlap of the wave functions and thus couples the polarization to the repulsion. Such an approach is currently investigated in our group.

V. ACKNOWLEDGMENTS

We thank LUNARC for generously providing computational resources and AstraZeneca and the Swedish Science Research Council for financial support.

---

[1] S. W. Rick and S. J. Stuart, in *Reviews in Computational Chemistry*, edited by K. B. Lipowitz and D. B. Boyd (Wiley-VCH, New York, 2002), vol. 18, p. 89.

[2] A. J. Stone, Mol. Phys. **56**, 1065 (1985).

[3] G. J. Williams and A. J. Stone, J. Chem. Phys. **119**, 4620 (2003).

[4] A. J. Misquitta and A. J. Stone, J. Chem. Phys. **124**, 024111 (2006).

[5] J. G. Ángyán, G. Jansen, M. Loos, C. Hättig, and B. A. Heß, Chem. Phys. Letters **219**, 267 (1994).

[6] F. Dehez, C. Chipot, C. Millot, and J. G. Ángyán, Chem. Phys. Letters **338**, 180 (2001).

[7] L. Gagliardi, R. Lindh, and G. Karlström, J. Chem. Phys. **121**, 4494 (2004).

[8] T. J. Giese and D. M. York, J. Chem. Phys. **120**, 9903 (2004).

[9] A. Krishtal, P. Senet, M. Yang, and C. Van Alsenoy, J. Chem. Phys. **125**, 034312 (2006).

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

[11] W. J. Stevens and W. H. Fink, Chem. Phys. Letters **139**, 15 (1987).

[12] P. S. Bagus, K. Hermann, and C. W. J. Bauschlicher, J. Chem. Phys. **80**, 4378 (1984).

[13] B. Jeziorski, R. Moszynski, and K. Szalewicz, Chem. Rev. **94**, 1887 (1994).

[14] W. Chen and M. S. Gordon, J. Phys. Chem. **100**, 14316 (1996).

[15] M. S. Gordon and J. H. Jensen, in *Encyclopedia of Computational Chemistry*, edited by H. F. Schaefer and P. v. R. Schleyer (Wiley, Chichester, 1998), p. 3198.

[16] J. G. Ángyán, C. Chipot, F. Dehez, C. Hättig, G. Jansen, and C. Millot, J. Comput. Chem. **24**, 997 (2003).

[17] K. Morokuma, Acc. Chem. Res. **10**, 294 (1977).

[18] R. F. Frey and E. R. Davidson, J. Chem. Phys. **90**, 5555 (1989).

[19] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Letters **313**, 701 (1999).

[20] D. G. Fedorov and K. Kitaura, J. Comput. Chem. **28**, 222 (2007).

[21] P. Ren and J. W. Ponder, J. Phys. Chem. B **107**, 5933 (2003).

[22] J.-P. Piquemal, A. Marquez, O. Parisel, and C. Giessner-Prettre, J. Comput. Chem. **26**, 1052 (2005).

[23] R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).

[24] P.-O. Löwdin, J. Chem. Phys. **18**, 365 (1950).

[25] A. J. Stone and M. Alderton, Mol. Phys. **56**, 1047 (1985).

[26] A. Reed, L. Curtiss, and F. Weinhold, Chem. Rev. **88**, 899 (1988).

[27] R. F. W. Bader, *Atoms in Molecules* (Clarendon Press, Oxford, 1990).

[28] J. Cioslowski, J. Am. Chem. Soc. **111**, 8333 (1989).

[29] M. Ferraro, M. Caputo, and P. Lazzeretti, J. Chem. Phys. **109**, 2987 (1988).

[30] P. Söderhjelm, J. W. Krogh, G. Karlström, U. Ryde, and R. Lindh, J. Comput. Chem. **28**, 1083 (2007).

[31] F. B. van Duijneveldt, in *Molecular Interactions*, edited by S. Scheiner (Wiley, Chichester, 1997), p. 81.

[32] G. Karlström and A. Sadlej, Theor. Chim. Acta **61**, 1 (1982).

[33] J. G. C. M. van Duijneveldt-van de Rijdt and F. B. van Duijneveldt, J. Chem. Phys. **97**, 5019 (1992).

[34] P. Salvador, X. Fradera, and M. Duran, J. Chem. Phys. **112**, 10106 (2000).

[35] J. Bentley, J. Phys. Chem. A **102**, 6043 (1998).

[36] P.-O. Widmark, P.-Å. Malmqvist, and B. O. Roos, Theor. Chim. Acta **77**, 291 (1990).

[37] T. H. Dunning Jr, J. Chem. Phys. **90**, 1007 (1989).

[38] D. E. Woon and T. H. Dunning Jr, J. Chem. Phys. **98**, 1358 (1993).

[39] P. Söderhjelm, G. Karlström, and U. Ryde, J. Chem. Phys. **124**, 244101 (2006).

[40] G. Karlström, R. Lindh, P.-Å. Malmqvist, B. O. Roos, U. Ryde, V. Veryazov, P.-O. Widmark, M. Cossi, B. Schimmelpfennig, P. Neogrady, et al., Computational Materials Science **28**, 222 (2003).

[41] V. Veryazov, P.-O. Widmark, L. Serrano-Andrés, R. Lindh, and B. O. Roos, Int. J. Quantum Chem. **100**, 626 (2004).

[42] G. A. Kaminski, H. A. Stern, B. J. Berne, and R. A. Friesner, J. Phys. Chem. A **108**, 621 (2004).

[43] A. Serr and R. R. Netz, Int. J. Quantum Chem. **106**, 2960 (2006).

[44] R_i equals 2.04 Å for oxygen, 2.19 Å for nitrogen, 2.28 Å for carbon, 2.37 Å for sulfur, 2.86 Å for non-polar hydrogen, and 1.97 Å for polar hydrogen

[45] There is one additional technical note on LoProp which we should mention. In CPOL and LinCPOL, densities for each monomer are computed in their monomer basis set. The monomer dipole moments are thus taken from each calculation separately, as one would expect. There is an alternative, however, namely to merge the two monomer densities and analyze them by LoProp in the full dimer basis set. After correcting for D-BSSE, there is still a slight difference between the two approaches. The reason for this is again the imperfect locality of the LoProp basis. The best correspondence between the CPOL (or LinCPOL) monomer dipole moments and the supermolecular ones would actually be obtained if CPOL and LinCPOL were analyzed in the dimer basis, since the analysis of the supermolecular results obviously has to be done in the dimer basis. From test calculations, we conclude, however, that this difference is for LinCPOL no larger than the error from ordinary BSSE. The error is somewhat larger for CPOL, but still sufficiently small to make comparisons of the intuitive CPOL and LinCPOL

monomer dipole moments with the supermolecular results meaningful.

[46] See EPAPS Document No. XXXXX. This document can be reached through a direct link in the online article HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).

Tables

TABLE I: Summary of the considered models.

Model	Non-expanded induced density	Intramolecular coupling	Non-linear response	Pauli effects	Charge transfer
Supermolecular	yes	yes	yes	yes	yes
RVS	yes	yes	yes	yes	no
CPOL	yes	yes	yes	no	no
LinCPOL	yes	yes	no	no	no
Polarizabilities	no	no	no	no	no

TABLE II: Typical values for the two types of basis set superposition error (BSSE) in the induced dipole moment, divided into monomer A, monomer B, and rest term (AB) contributions. The system is the water dimer at O$\cdots$H distance 2.01 Å and all values are in thousandths of a.u. The total BSSE-corrected induced dipole moments are given for comparison.

	A	B	AB	Total
D-BSSE A	10	1	-11	0
D-BSSE B	-25	-22	47	0
Tot D-BSSE	-15	-21	35	0
Ordinary BSSE A	0	3	0	3
Ordinary BSSE B	0	3	-1	2
Tot Ordinary BSSE	0	6	-1	4
Total ind. dip.	123	87	-1	209

Figure captions

FIG. 1: Visual guide to the (a) CPOL, (b) LinCPOL, (c) Polarizabilities and the (d) RVS models. The density of monomer A is the same in all models, but how the resulting electric potential (dashed line) is applied to monomer B differs. The additional potential shown with dot-dashed line for RVS is an informal representation of the Pauli effects. Since these formally include non-local terms, a potential like that shown can not really be constructed.

FIG. 2: Induced dipole in the z direction (i.e. along the $\mathrm{H}\cdots\mathrm{O}$ axis) for monomer A (left) and B (right) of the water dimer as a function of the intermolecular separation. In the left plot, the supermolecular curve coincides with the RVS curve; in both plots, the polarizabilities curve is slightly above the LinCPOL curve.

FIG. 3: Illustration of the polarization behavior of monomer A. The electrostatic potential from monomer B is shown as a dashed curve. Its first-order Taylor expansion at the hydrogen nucleus, used in the linear models, is shown as a straight line.

FIG. 4: Total induced dipole moment and second moment in the z direction for the water dimer as a function of the intermolecular separation. The origin for the calculation of the second moment is kept fixed at the position of the interacting hydrogen atom of monomer A. In the left plot, the order of the three uppermost curves is, from above: polarizabilities, LinCPOL, and supermolecular.

FIG. 5: Relative mean error of the induced electrostatic potential around the water dimer as a function of the intermolecular separation.

FIG. 6: Induction energy of the water dimer as a function of the intermolecular separation. The polarizabilities curve coincides with the LinCPOL curve.

FIG. 7: Total induced second moment in the z direction for the formamide dimer as a function of the intermolecular separation.

FIG. 8: Relative mean error of the induced electrostatic potential around the formamide dimer as a function of the intermolecular separation.

FIG. 9: Scatter plots of the induced potential from the CPOL model (left) and the LinCPOL model (right) against the corresponding induced potential from the supermolecular calculation. The O $\cdots$ H distance is 1.87 Å.

FIG. 10: Induction energy of the formamide dimer as a function of the intermolecular separation. The polarizabilities curve coincides with the RVS curve.

FIG. 11: Error in the induced dipole moment of monomer A (left) and monomer B (right), plotted against the corresponding supermolecular result. Each point represents one dimer in the large data set.

FIG. 12: Statistical distribution of the relative error in the induced electrostatic potential for the large data set.

FIG. 13: Difference between the polarizability model and the LinCPOL model. Left: Statistical distribution of the relative difference in the induced electrostatic potential for the large data set. Right: Statistical distribution of the difference in the induction energy for the large data set.

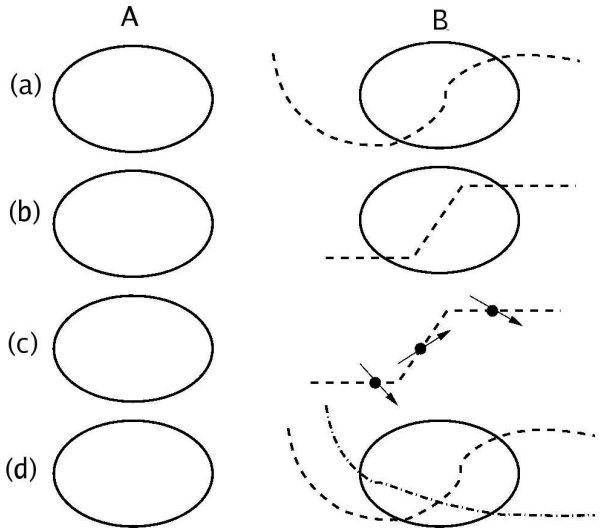

FIG. 1: P. Soderhjelm et al.

Visual guide to the (a) CPOL, (b) LinCPOL, (c) Polarizabilities and the (d) RVS models. The density of monomer A is the same in all models, but how the resulting electric potential (dashed line) is applied to monomer B differs. The additional potential shown with dot-dashed line for RVS is an informal representation of the Pauli effects. Since these formally include non-local terms, a potential like that shown can not really be constructed.

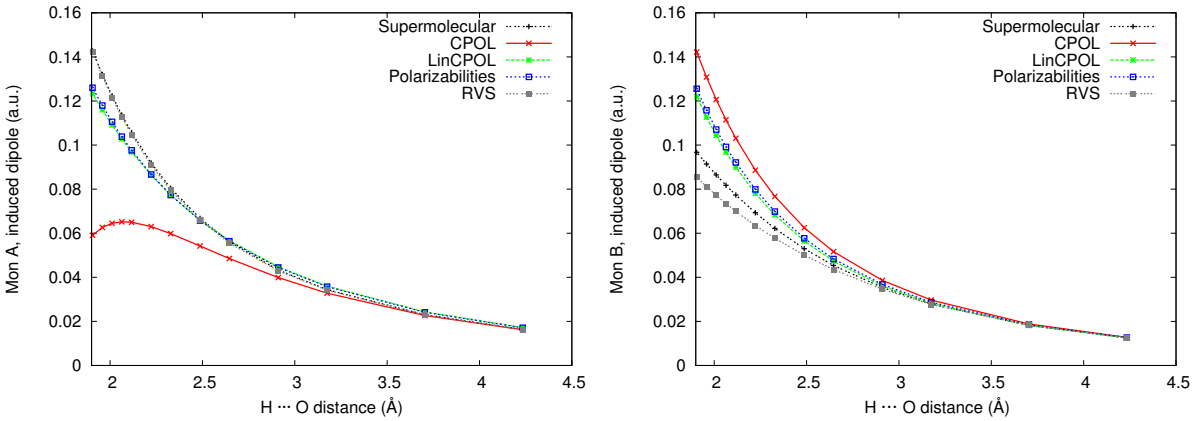

FIG. 2: P. Soderhjelm et al.

Induced dipole in the z direction (i.e. along the $\mathrm{H}\cdots\mathrm{O}$ axis) for monomer A (left) and B (right) of the water dimer as a function of the intermolecular separation. In the left plot, the supermolecular curve coincides with the RVS curve; in both plots, the polarizabilities curve is slightly above the LinCPOL curve.

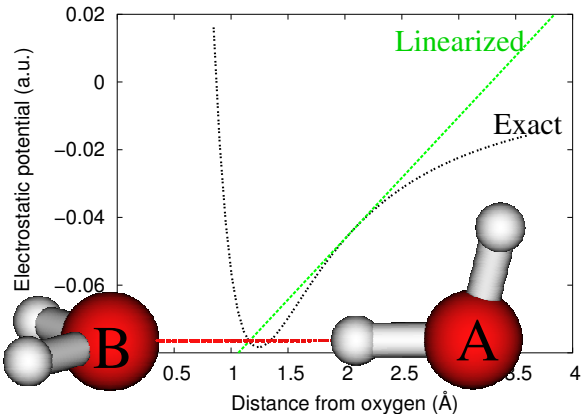

FIG. 3: P. Soderhjelm et al.

Illustration of the polarization behavior of monomer A. The electrostatic potential from monomer B is shown as a dashed curve. Its first-order Taylor expansion at the hydrogen nucleus, used in the linear models, is shown as a straight line.

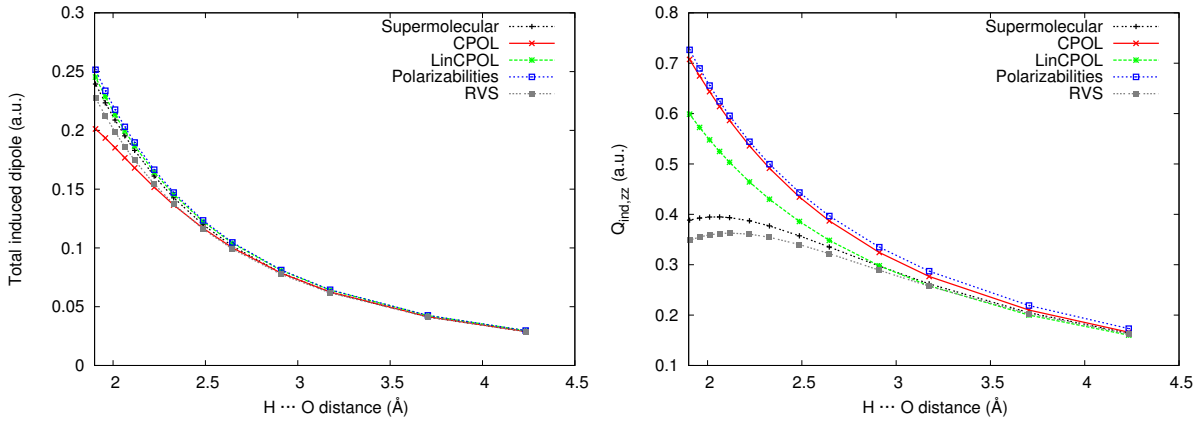

FIG. 4: P. Soderhjelm et al.

Total induced dipole moment and second moment in the z direction for the water dimer as a function of the intermolecular separation. The origin for the calculation of the second moment is kept fixed at the position of the interacting hydrogen atom of monomer A. In the left plot, the order of the three uppermost curves is, from above: polarizabilities, LinCPOL, and supermolecular.

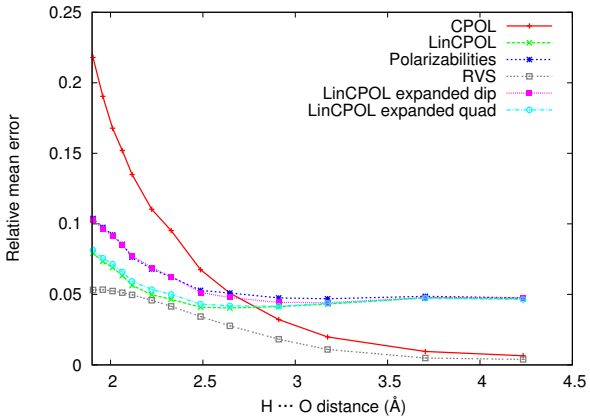

FIG. 5: P. Soderhjelm et al.

Relative mean error of the induced electrostatic potential around the water dimer as a function of the intermolecular separation.

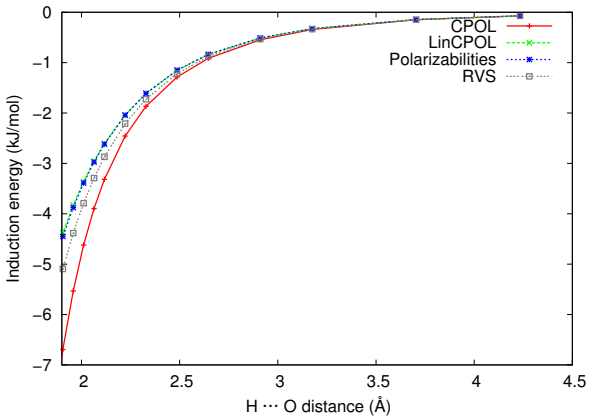

FIG. 6: P. Soderhjelm et al.

Induction energy of the water dimer as a function of the intermolecular separation. The polarizabilities curve coincides with the LinCPOL curve.

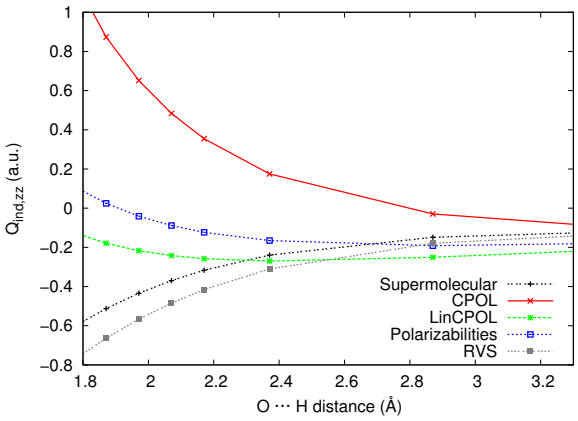

FIG. 7: P. Soderhjelm et al.

Total induced second moment in the z direction for the formamide dimer as a function of the intermolecular separation.

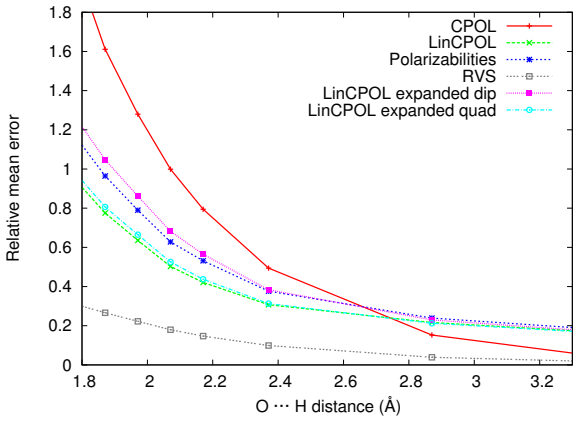

FIG. 8: P. Soderhjelm et al.

Relative mean error of the induced electrostatic potential around the formamide dimer as a function of the intermolecular separation.

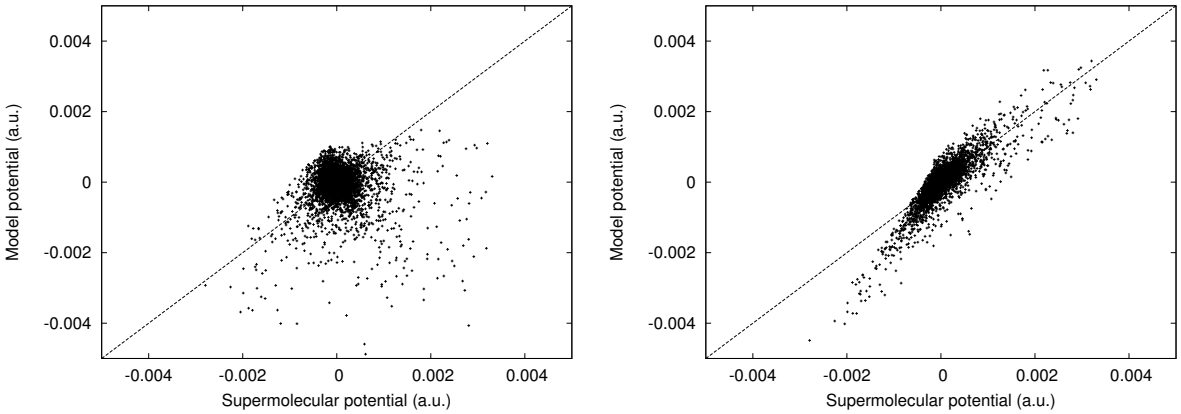

FIG. 9: P. Soderhjelm et al.

Scatter plots of the induced potential from the CPOL model (left) and the LinCPOL model (right) against the corresponding induced potential from the supermolecular calculation. The O $\cdots$ H distance is 1.87 Å.

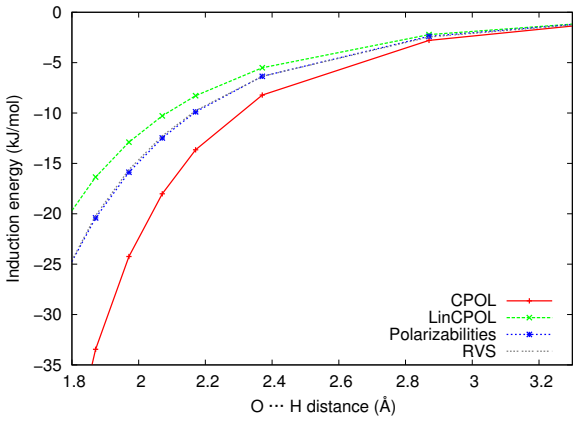

FIG. 10: P. Soderhjelm et al.

Induction energy of the formamide dimer as a function of the intermolecular separation. The polarizabilities curve coincides with the RVS curve.

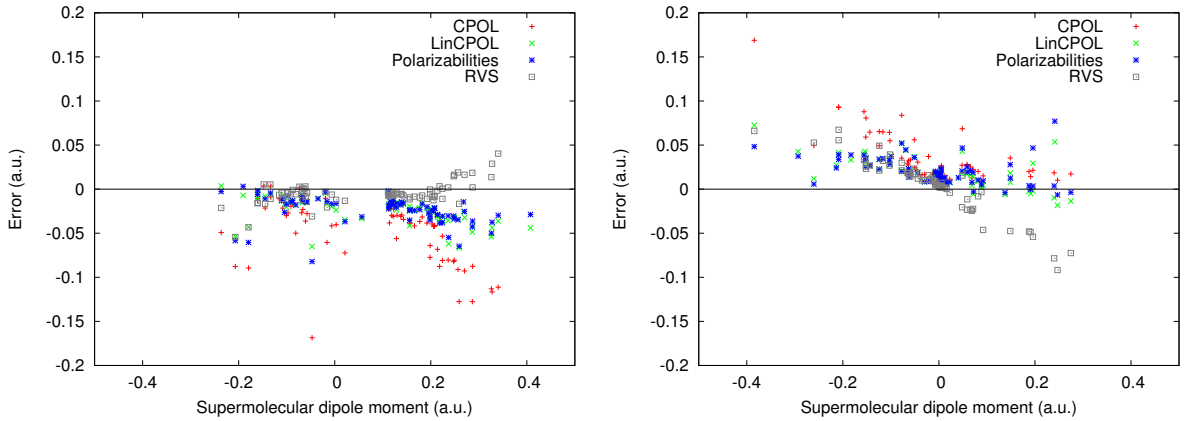

FIG. 11: P. Soderhjelm et al.

Error in the induced dipole moment of monomer A (left) and monomer B (right), plotted against the corresponding supermolecular result. Each point represents one dimer in the large data set.

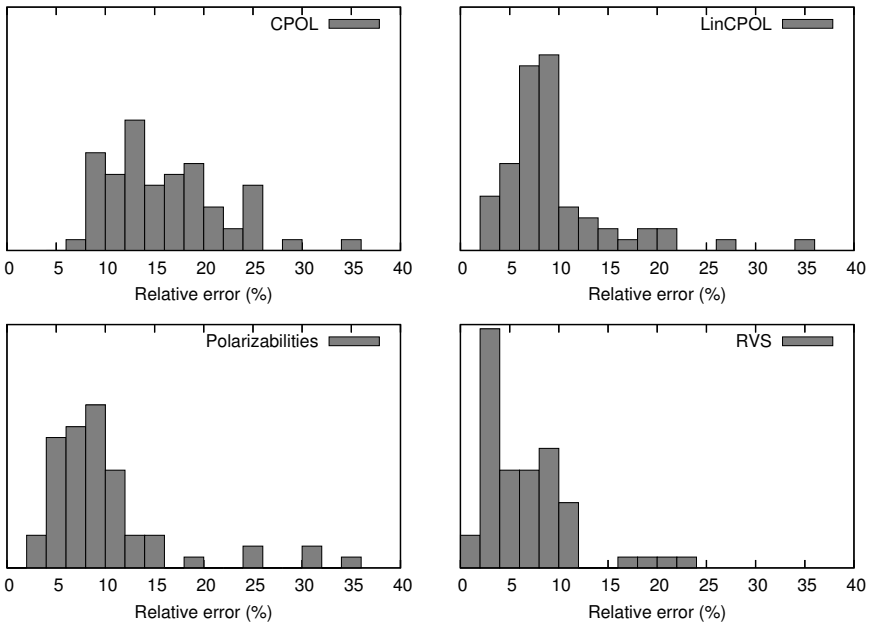

FIG. 12: P. Soderhjelm et al.

Statistical distribution of the relative error in the induced electrostatic potential for the large data set.

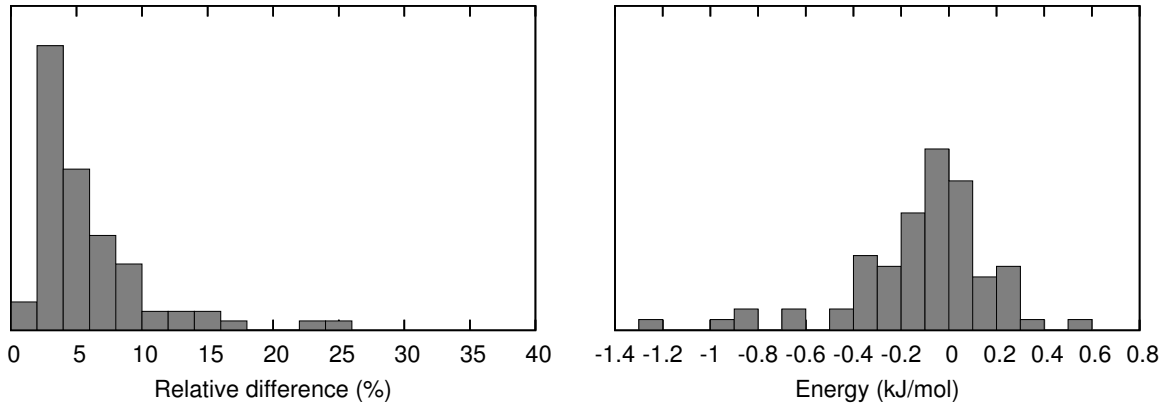

FIG. 13: P. Soderhjelm et al.

Difference between the polarizability model and the LinCPOL model. Left: Statistical distribution of the relative difference in the induced electrostatic potential for the large data set. Right: Statistical distribution of the difference in the induction energy for the large data set.