

# Beyond Born-Mayer: Improved models for short-range repulsion in ab initio force fields

Mary J. Van Vleet,<sup>†</sup> Alston J. Misquitta,<sup>‡</sup> Anthony J. Stone,<sup>¶</sup> and J.R. Schmidt<sup>\*,†</sup>

<sup>†</sup>*Theoretical Chemistry Institute and Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin, 53706, United States*

<sup>‡</sup>*Department of Physics and Astronomy, Queen Mary University of London, London E1 4NS, United Kingdom*

<sup>¶</sup>*Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom*

E-mail: schmidt@chem.wisc.edu

## Abstract

Short-range repulsion within inter-molecular force fields is conventionally described by either Lennard-Jones ( $A/r^{12}$ ) or Born-Mayer ( $A \exp(-Br)$ ) forms. Despite their widespread use, these simple functional forms are often unable to describe the interaction energy accurately over a broad range of inter-molecular distances, thus creating challenges in the development of ab initio force fields and potentially leading to decreased accuracy and transferability. Herein, we derive a novel short-range functional form based on a simple Slater-like model of overlapping atomic densities and an iterated stockholder atom (ISA) partitioning of the molecular electron density. We demonstrate that this Slater-ISA methodology yields a more accurate, transferable, and robust description of the short-range interactions at minimal additional computational cost compared to standard Lennard-Jones or Born-Mayer approaches. Finally,

we show how this methodology can be adapted to yield the standard Born-Mayer functional form while still retaining many of the advantages of the Slater-ISA approach.

## 1 Introduction

Molecular simulation is an essential tool for interpreting and predicting the structure, thermodynamics, and dynamics of chemical and biochemical systems. The fundamental inputs into these simulations are the intra- and intermolecular force fields, which provide simple and computationally efficient descriptions of molecular interactions. Consequently, the predictive and explanatory power of molecular simulations depends on the fidelity of the force field to the underlying (exact) potential energy surface.

In the case of intermolecular interactions, the dominant contributions for non-reactive systems can be decomposed into the following physically-meaningful energy components: electrostatic, exchange-repulsion, induction and dispersion.<sup>1–5</sup> At large intermolecular distances, where monomer electron overlap can be neglected, the physics of intermolecular interactions can be described entirely on the basis of monomer properties (e.g. multipole moments, polarizabilities), all of which can be calculated with high accuracy from first principles.<sup>6</sup> In conjunction with associated distribution schemes that decompose molecular monomer properties into atomic contributions,<sup>1,4,7–11</sup> these monomer properties lead to an accurate and computationally efficient model of ‘long-range’ intermolecular interactions as a sum of atom-atom terms, which can be straightforwardly included in common molecular simulation packages.

At shorter separations, where the molecular electron density overlap cannot be neglected, the asymptotic description of intermolecular interactions breaks down due to the influence of Pauli repulsion, charge penetration and charge transfer. These effects can be quantitatively described using modern electronic structure methods,<sup>3,12–15</sup> but are far more challenging to model accurately using computationally inexpensive force fields. For efficiency and ease of parameterization, most simple force fields use a single ‘repulsive’ term to model the cumulative influence of (chemically distinct) short-range interactions. These simple models have seen comparatively little progress over the past eighty years, and the Lennard-Jones<sup>16</sup> ( $A/r^{12}$ ) and Born-Mayer<sup>17,18</sup> ( $A \exp(-Br)$ ) forms continue as popular descriptions of short-range effects in standard force fields despite some well-known limitations (*vide infra*).

Because the prediction of physical and chemical properties depends on the choice of short-range interaction model,<sup>19–32</sup> it is essential to develop sufficiently accurate short-range force

fields. This is particularly true in the case of ab initio force field development. A principle goal of such a first-principles approach is the reproduction of a calculated potential energy surface (PES), thus (ideally) yielding accurate predictions of bulk properties.<sup>33</sup> Substantial deviations between a fitted and calculated PES lead to non-trivial challenges in the parameterization process, which in turn can often degrade the quality of property predictions. The challenge of reproducing an ab initio PES becomes particularly pronounced at short intermolecular separations, where many common force field functional forms are insufficiently accurate. For example, the popular Lennard-Jones ( $A/r^{12}$ ) functional form is well-known to be substantially too repulsive at short contacts as compared to the exact potential.<sup>27–29,34,35</sup> While the Born-Mayer ( $A \exp(-Br)$ ) functional form is more physically-justified<sup>18</sup> and fares better in this regard,<sup>34</sup> substantial deviations often persist.<sup>36</sup> In addition, parameterization of the Born-Mayer form is complicated by the strong coupling of the pre-exponential ( $A$ ) and exponent ( $B$ ) parameters, hindering the transferability of the resulting force field. These considerations, along with the observed sensitivity of structural and dynamic properties to the treatment of short-range repulsion,<sup>19</sup> highlight the need for new approaches to model short-range repulsive interactions.

Our primary goal in this work is to derive a simple and accurate description of short-range interactions in molecular systems that improves upon both the standard Lennard-Jones and Born-Mayer potentials in terms of accuracy, transferability, and ease of parameterization. Our focus in this work is on ab initio force field development, and thus we will use the fidelity of a given force field with respect to an accurate ab initio PES as a principle metric of force field quality. We note that other metrics may be more appropriate for the development of empirical potentials, where Lennard-Jones or Born-Mayer forms may yield highly accurate ‘effective’ potentials when parameterized against select bulk properties. Nonetheless, we anticipate that the models proposed in this work may prove useful for empirical force field development in cases where a more physically-motivated functional form is necessary.<sup>27–29</sup>

The outline of this paper is thus as follows: first, we derive a new functional form capable of describing short-range repulsion from first principles, and show how the standard Born-Mayer form follows as an approximation to this more exact model. Our generalization of the Born-Mayer functional form allows for an improved description of a variety of short-range effects, namely electrostatic charge penetration, exchange-repulsion, and density overlap effects on induction and dispersion. Crucially, we also demonstrate how the associated atomic exponents can be extracted from first-principles monomer charge densities via an iterated stockholder atoms (ISA) density partitioning scheme, thereby reducing the

number of required fitting parameters compared to the Born-Mayer model. Benchmarking this ‘Slater-ISA’ methodology (functional form and atomic exponents) against high-level ab initio calculations and experiment, we find that the approach exhibits increased accuracy, transferability, and robustness as compared to a typical Lennard-Jones or Born-Mayer potential. In addition, we show how the ISA-derived exponents can be adapted for use within the standard Born-Mayer form (Born-Mayer-sISA), while still retaining many of the advantages of the Slater-ISA approach. As such, our methodology also offers an opportunity to dramatically simplify the development of both empirically-parameterized and ab initio simulation potentials based upon the standard Born-Mayer form.

## 2 Theory

We begin with a formal treatment of the overlap model for the exchange-repulsion between two isolated atoms, and then extend these results to develop a generalized model for the short-range interactions in both atomic and molecular systems. Finally, we show how the conventional Born-Mayer model can be derived as an approximation to this more rigorous treatment.

### 2.1 Models for the exchange-repulsion between isolated atoms

It is well known that the exchange-repulsion interaction between two closed-shell atoms  $i$  and  $j$  is proportional, or very nearly proportional, to the overlap of their respective charge densities:<sup>37</sup>

$$E_{ij}^{\text{exch}} \approx V_{ij}^{\text{exch}} = K_{ij}(S_\rho^{ij})^\gamma \quad (1)$$

$$S_\rho^{ij} = \int \rho_i(\mathbf{r})\rho_j(\mathbf{r})d^3\mathbf{r}. \quad (2)$$

Here and throughout, we use  $E$  to denote quantum mechanical energies, and  $V$  to denote the corresponding model/force field energies. Recently two of us have provided a theoretical justification for this repulsion hypothesis (or overlap model), and have shown that  $\gamma = 1$  provided that asymptotically-correct densities are used to compute both the atomic densities and  $E_{ij}^{\text{exch}}$ .<sup>4,38</sup> As this is the case for the calculations in this paper, we assume  $\gamma = 1$  throughout this work.

The overlap model has frequently been utilized in the literature and has been found to

yield essentially quantitative accuracy for a wide variety of chemical systems.<sup>37,39,40</sup> Prior work exploiting the overlap model has generally followed one of two strategies. Striving for quantitative accuracy, several groups have developed approaches to evaluate eq. (2) via either numerical integration or density fitting of ab-initio molecular electron densities,  $\rho_i$  (e.g. SIBFA, GEM, effective fragment potentials).<sup>41–50</sup> These force fields, while often extremely accurate, lack the simple closed-form analytical expressions that define standard force fields (such as the Lennard-Jones or Born-Mayer models) and thus are often much more computationally expensive than conventional models.

In contrast, and similar to our objectives, the overlap model has also been used in the development of standard force fields. In this case, the molecular electron density as well as the overlap itself is drastically simplified in order to yield a simple closed-form expression that can be used within a conventional molecular simulation package.<sup>37,39,40</sup> As we show below, the Born-Mayer model can be ‘derived’ via such an approach. At the expense of some accuracy, the resulting overlap-based force fields exhibit high computational efficiency and employ well-known functional forms.

Building on this prior work, our present goal is to derive rigorous analytical expressions and improved approximations for both  $\rho_i$  and eq. (2), facilitating the construction of ab initio force fields that exhibit simplicity, high computational efficiency, fidelity to the underlying PES, and (with only trivial modifications) compatibility with standard simulation packages. We first start with the case of isolated atoms, where it is well-known that the atomic electron density decays asymptotically as

$$\rho_{r \rightarrow \infty}(r) \propto r^{2\beta} e^{-2\alpha r} \quad (3)$$

where the exponent  $\alpha = \sqrt{2I}$  is fixed by the vertical ionization potential  $I$ ,  $\beta = -1 + \frac{Q}{\alpha}$ , and  $Q = Z - N + 1$  for an atom with nuclear charge  $+Z$  and electronic charge  $-N$ .<sup>38,51–53</sup> The exponential term dominates the asymptotic form of the density, and the  $r$ -dependent prefactor may be neglected.<sup>11,39,40,54</sup> In this case, the density takes the even simpler form

$$\rho_{r \rightarrow \infty}(r) \approx D e^{-Br}, \quad (4)$$

where  $D$  is a constant that effectively absorbs the missing  $r$ -dependent pre-factor and  $B$  is an exponent that is now only approximately equal to  $2\alpha$ .

In the case of two identical atoms, substitution into eq. (2) yields a simple expression for

the density overlap,  $S_\rho$ ,<sup>55,56</sup>

$$S_\rho^{ii} = \frac{\pi D^2}{B^3} P(Br_{ii}) \exp(-Br_{ii}) \quad (5)$$

$$P(Br_{ii}) = \frac{1}{3}(Br_{ii})^2 + Br_{ii} + 1$$

as well as (via eq. (1)) the exchange-repulsion energy:<sup>40,57</sup>

$$V_{ii}^{\text{exch}} = A_{ii}^{\text{exch}} P(Br_{ii}) \exp(-Br_{ii}). \quad (6)$$

Here,  $r_{ii}$  represents an interatomic distance, and  $A_{ii}^{\text{exch}}$  indicates a proportionality constant that is typically fit to calculated values of the exchange-repulsion energy. The only approximations thus far are the use of the overlap model and the simplified asymptotic form of the atomic charge density.

For the general case of two hetero-atoms, substitution of eq. (4) into eq. (2) yields the more complicated expression<sup>55,56</sup>

$$S_\rho^{ij} = \frac{16\pi D_i D_j \exp(-\{B_i + B_j\}r_{ij}/2)}{(B_i^2 - B_j^2)^3 r_{ij}} \times$$

$$\left[ \left( \frac{B_i - B_j}{2} \right)^2 \left( \exp \left( \{B_i - B_j\} \frac{r_{ij}}{2} \right) - \exp \left( -\{B_i - B_j\} \frac{r_{ij}}{2} \right) \right) \right.$$

$$\times \left( \left( \frac{B_i + B_j}{2} \right)^2 r_{ij}^2 + (B_i + B_j)r_{ij} + 2 \right)$$

$$- \left( \frac{B_i + B_j}{2} \right)^2 \exp \left( \{B_i - B_j\} \frac{r_{ij}}{2} \right) \times \left( \left( \frac{B_i - B_j}{2} \right)^2 r_{ij}^2 - (B_i - B_j)r_{ij} + 2 \right)$$

$$\left. + \left( \frac{B_i + B_j}{2} \right)^2 \exp \left( -\{B_i - B_j\} \frac{r_{ij}}{2} \right) \times \left( \left( \frac{B_i - B_j}{2} \right)^2 r_{ij}^2 + (B_i - B_j)r_{ij} + 2 \right) \right], \quad (7)$$

which is too cumbersome to serve as a practical force field functional form. However, since the above expression reduces to eq. (5) in the limit  $B_i = B_j$ , and because  $|B_i - B_j|$  is small for most atom pairs, we have found that eq. (7) may be approximated using eq. (5) with an *effective* atomic exponent  $B$ . An expansion of eq. (7) about  $B_i = B_j$  suggests that this effective exponent should be given by the arithmetic mean,  $B_{ij} = \frac{1}{2}(B_i + B_j)$ . However, a Waldman-Hagler style analysis<sup>58</sup> (see Supporting Information) suggests instead that a more

suitable exponent is given by the geometric mean combination rule,

$$B = B_{ij} \equiv \sqrt{B_i B_j}. \quad (8)$$

As shown in the Supporting Information, this approximate overlap model (eq. (5) and eq. (8)) is of comparable accuracy to the exact overlap from eq. (7). Thus the density overlap and (force field) exchange energies of arbitrary hetero-atoms take the simple forms

$$S_\rho^{ij} = D_{ij} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \quad (9)$$

$$D_{ij} = \pi D_i D_j B_{ij}^{-3} \quad (10)$$

$$P(B_{ij}, r_{ij}) = \frac{1}{3}(B_{ij} r_{ij})^2 + B_{ij} r_{ij} + 1 \quad (11)$$

and

$$V_{ij}^{\text{exch}} = A_{ij}^{\text{exch}} P(B_{ij} r_{ij}) \exp(-B_{ij} r_{ij}). \quad (12)$$

Due to the connection with the overlap between two s-type Slater orbitals, we refer to eq. (12) as the Slater functional form. Note that this expression reduces to the standard Born-Mayer function by making the further approximation  $P(B_{ij} r_{ij}) = 1$ , although it is known<sup>40,59</sup> that this is a poor approximation with the  $B_{ij}$  as defined above. Instead, as we shall demonstrate in Section 4, the exponents  $B_{ij}$  need to be scaled for accurate use with a Born-Mayer functional form.

Variants of the polynomial pre-factor from eq. (9) have previously been recognized and used in intermolecular interaction models.<sup>18,57,59</sup> Early work by Buckingham<sup>18</sup> hypothesized that the functional form of eq. (12) would be more accurate than the Born-Mayer form, though no attempt was made to provide a closed-form expression for  $P$ . More recent potentials have incorporated a low-order polynomial into the exchange repulsion term, either by direct parameterization<sup>60–64</sup> or indirectly by fitting the exchange to  $S_\rho/r^2$  rather than to  $S_\rho$  itself.<sup>39,40,65</sup> Kita et al. have derived (but not tested) eq. (6) for the homoatomic case.<sup>65</sup> Recently, and most similar to the spirit of the present work, York and co-workers have derived a model based upon the overlap of Slater-type orbitals for use in QM/MM simulations, yielding an expression identical to eq. (7).<sup>66–68</sup> Those authors treated  $D_i$  and  $D_j$  as empirical fitting parameters and estimated atomic exponents ( $B_i$  and  $B_j$ ) via atomic-charge dependent functions. In contrast, we will demonstrate that utilization of the far simpler functional form from eq. (12), in conjunction with exponents calculated from analysis of the first-principles

molecular electron density, yields much higher computational efficiency and simplifies the parameterization process without significant loss of accuracy.

For an arbitrary pair of interacting atoms,  $A_{ij}^{\text{exch}}$  can be obtained by fitting to calculated exchange-repulsion energies. However, assuming that the overlap proportionality factor  $K_{ij}$  is a universal constant (or, alternatively, separable with  $K_{ij} = K_i K_j$ ), then

$$A_{ij}^{\text{exch}} = \left( K_i \sqrt{\frac{\pi}{B_i^3}} D_i \right) \left( K_j \sqrt{\frac{\pi}{B_j^3}} D_j \right) \equiv A_i^{\text{exch}} A_j^{\text{exch}}, \quad (13)$$

thus providing a combination rule for heteroatomic interaction in terms of purely atomic quantities. The universality and separability of  $K_{ij}$  are, at present, empirically rather than theoretically justified.<sup>4,69,70</sup> The  $A_i^{\text{exch}}$  can then be obtained, for example, by a straightforward fitting of calculated ab initio *homoatomic* exchange-repulsion energies.

## 2.2 Models for other short-range interactions between isolated atoms

Beyond the exchange-repulsion, the density-overlap model may also be used to model other short-range interaction components, such as the electrostatic charge penetration energy and the short-range induction and dispersion energies (that is, the portion modulated by charge overlap). Indeed, it has been demonstrated that the electrostatic charge penetration energy is approximately proportional to the exchange-repulsion energy, and consequently to the charge density overlap,<sup>4,11</sup> which has provided a successful basis for modeling the electrostatic charge penetration energy.<sup>71,72</sup> While the relation between short-range induction and charge overlap is less clear, recent results have demonstrated that the charge-transfer energy, which is the dominant short-range component of the induction energy,<sup>73</sup> is approximately proportional to the first-order exchange energy,<sup>38,74</sup> and prior work has successfully used the overlap hypothesis to describe the short-range induction.<sup>4,71,72</sup> We therefore model the electrostatic charge penetration and short-range induction interactions as

$$V_{ij}^{\text{pen}} = A_{ij}^{\text{pen}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \quad (14)$$

and

$$V_{ij}^{\text{ind,sr}} = A_{ij}^{\text{ind}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}). \quad (15)$$

Aside from the pre-factors  $A_{ij}$ , these expressions are identical to that for the exchange-repulsion term.

The behavior of the dispersion interaction at short distances poses a special challenge. In order to model the short-range dispersion and to resolve the unphysical, mathematical divergence of the  $1/r^n$  terms as  $r \rightarrow 0$ , Tang and Toennies have shown that the terms in the dispersion expansion should be damped using an appropriate incomplete gamma function

$$f_n(x) = 1 - e^{-x} \sum_{k=0}^n \frac{(x)^k}{k!} \quad (16)$$

$$x = -\frac{d}{dr} [\ln V^{\text{exch}}(r)] r \quad (17)$$

that accounts for both exchange and charge penetration effects.<sup>75,76</sup> Note that the form of this damping factor depends on the model used for exchange repulsion. For the Slater functional form (eq. (12)),

$$x_{\text{Slater}} = B_{ij}r_{ij} - \frac{2B_{ij}^2r_{ij} + 3B_{ij}}{B_{ij}^2r_{ij}^2 + 3B_{ij}r_{ij} + 3}r_{ij}. \quad (18)$$

Alternatively, if we replace the Slater functional form with the less accurate Born-Mayer expression,  $x$  simplifies to the result originally given by Tang and Toennies:

$$x_{\text{Born-Mayer}} = B_{ij}r_{ij}. \quad (19)$$

### 2.3 Models for short-range interactions between molecules

The overlap repulsion hypothesis can be extended to molecules<sup>4,69,77–79</sup> by writing the molecular density  $\rho_I$  as a superposition of atomic densities

$$\rho_I(\mathbf{r}) = \sum_{i \in I} \rho_i(\mathbf{r}) \quad (20)$$

where  $i$  represents an atom in molecule  $I$ . In this case,

$$V_{IJ}^{\text{exch}} = \sum_{i \in I} \sum_{j \in J} V_{ij}^{\text{exch}} \quad (21)$$

$$V_{ij}^{\text{exch}} = K_{ij}S_{\rho}^{ij} = \int \rho_i(\mathbf{r})\rho_j(\mathbf{r})d^3\mathbf{r}. \quad (22)$$

Note that the form of eq. (22) is identical to the corresponding expression between isolated atoms, but requires partitioning of the molecular charge density into atom-in-molecule densities,  $\rho_i$ , each decaying according to an effective atom-in-molecule density decay exponent,  $B_i$ .

In principle, such atom-in-molecule exponents could be estimated from the ionization potentials of the corresponding isolated atoms,<sup>57,71</sup> but this approach neglects the influence of the molecular environment. A more appealing possibility is to directly evaluate the atom-in-molecule densities via partitioning of the calculated monomer densities. Density partitioning has not yet (to our knowledge) been applied in the context of the overlap model to solve for eq. (22), however several successful efforts in force field development have recently relied on an atoms-in-molecule approach in order to obtain accurate scaling relationships for intermolecular force field parameters.<sup>80–82</sup> In particular, Cole et al. utilized a density-derived electrostatic and chemical (DDEC) partitioning scheme<sup>83,84</sup> to generate Lennard-Jones dispersion and short-range repulsion parameters, though the latter parameters were calculated implicitly by enforcing the coincidence of the potential minimum and the calculated atomic radius.

While no unique atom-in-molecule density partitioning scheme exists, an ideal approach should yield atom-in-molecule densities that strongly resemble those of isolated atoms, e.g. maximally spherical and asymptotically exponential.<sup>11,85–87</sup> The recently developed iterated stockholder partitioning of Lillestolen and Wheatley obeys this first important constraint of sphericity.<sup>88,89</sup> As a non-trivial extension of the original Hirshfeld method,<sup>90</sup> iterated stockholder atom (ISA) densities are defined as

$$\rho_i(\mathbf{r}) = \rho_I(\mathbf{r}) \frac{w_i(\mathbf{r})}{\sum_{a \in I} w_a(\mathbf{r})} \quad (23)$$

where the converged shape functions  $w_i(\mathbf{r})$  are spherical averages of the atomic densities  $\rho_i(\mathbf{r})$ :

$$w_i(\mathbf{r}) = \langle \rho_i(\mathbf{r}) \rangle_{\text{sph}}. \quad (24)$$

This formulation ensures, by construction, that the sum of atomic densities reproduces the overall molecular density. Furthermore, the maximally spherical nature of the atom-in-molecule densities naturally facilitates a description of short-range interactions via a simple isotropic site-site model.

Misquitta et al. have developed a rapidly convergent implementation of the ISA procedure (BS-ISA<sup>11</sup>) using a basis set expansion which, in addition to exhibiting good convergence with respect to basis set, also leads to asymptotically-exponential atomic densities. Consequently, the BS-ISA method is our preferred density partitioning scheme. As an example, the spherically-averaged atomic densities for acetone are shown in Figure 1. For simplicity, and because a full treatment of the anisotropy is beyond the scope of this paper, we subsequently refer to the spherically-averaged atomic densities (i.e. the shape functions,  $w_i(r)$ ) as atomic or atom-in-molecule densities.

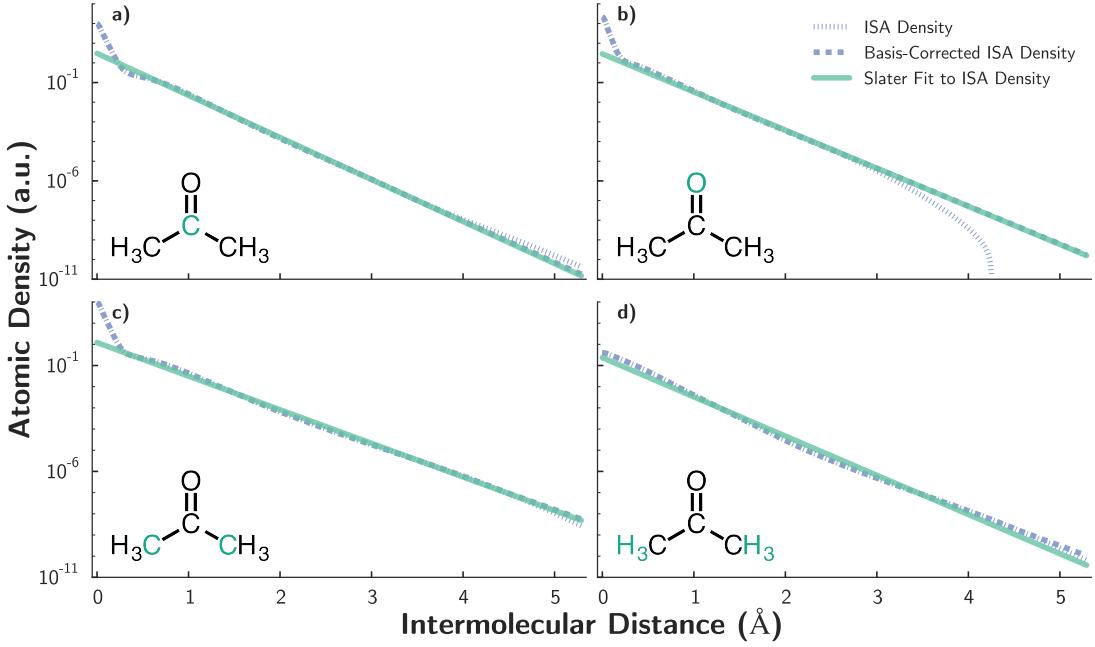


Figure 1: BS-ISA and fitted shape functions for each atom type in acetone: a) carbonyl carbon, b) oxygen, c) methyl carbon, d) hydrogen. BS-ISA shape functions (dotted line) for each atom type have been obtained at a PBE0/aug-cc-pVTZ level of theory. A modified BS-ISA shape function (dashed line) corrects the tail-region of the BS-ISA function to account for basis set deficiencies in the BS-ISA algorithm. A single Slater orbital of the form  $D_i^{\text{ISA}} \exp(-B_i^{\text{ISA}} r)$  (solid line) is fit to the basis-corrected BS-ISA shape function, and the obtained  $B_i^{\text{ISA}}$  value is used as an atomic exponent in the functional form of Slater-ISA FF. Results for acetone are typical of molecules studied in this work.

From Figure 1 we see that the ISA atomic shape functions (that is, the spherically-averaged ISA atoms-in-molecule density) decay exponentially outside the core region. However, note that the exponents governing the spherical density decay,  $B_i^{\text{ISA}}$ , differ from those of the free atoms. The ISA densities have been observed to account for electron movement

in the molecule, and the consequent density changes brought about by this movement tend to be manifested in the region of the density tails.<sup>11</sup> The ISA exponents can be obtained by a weighted least-squares fit to the BS-ISA atomic density (see the Computational Methods section for details), with the resulting fitted atomic densities shown in Figure 1. Note that even a single exponential is remarkably successful in reproducing the entirety of the valence atomic density.

Given these fitted ISA exponents, we can now apply our short-range interaction formalism to polyatomics,

$$\begin{aligned} V^{sr} &= \sum_{ij} A_{ij}^{\text{sr}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \\ P(B_{ij}, r_{ij}) &= \frac{1}{3} (B_{ij} r_{ij})^2 + B_{ij} r_{ij} + 1 \\ A_{ij}^{\text{sr}} &= A_i^{\text{sr}} A_j^{\text{sr}} \\ B_{ij} &= \sqrt{B_i^{\text{ISA}} B_j^{\text{ISA}}} \end{aligned} \quad (25)$$

where the molecular short-range energy is now a sum of atom-atom contributions. In conjunction with appropriately damped atomic dispersion (eq. (16) and eq. (18)), eq. (25) completely defines our new short-range force field. We refer to this new functional form and set of atomic exponents as the Slater-ISA FF.

### 3 Computational Methods

To evaluate the Slater-ISA FF against conventional Born-Mayer and/or Lennard-Jones models, we compare the ability of each resulting short-range force field to reproduce benchmark ab initio intermolecular interaction energies for a collection of representative dimers. Such a metric is directly relevant for ab initio force field development. Even for an empirically-parameterized force field, however, fidelity to an accurate ab initio potential should be well correlated with the highest level of accuracy and transferability achievable with a given short-range methodology.

We have developed the Slater-ISA FF, Born-Mayer, and Lennard-Jones force fields using benchmark energies calculated using the symmetry-adapted perturbation theory based on density-functional theory (DFT-SAPT or SAPT(DFT)<sup>91–99</sup>). DFT-SAPT provides interaction energies that are comparable in accuracy to those from CCSD(T) and which are rigorously free from basis set superposition error.<sup>3,100</sup> Additionally, at second-order, DFT-

SAPT also provides an explicit interaction energy decomposition into physically-meaningful contributions: the electrostatic, exchange-repulsion, induction, and dispersion energies. This decomposition is vital to the development of models as it allows the development of separate terms for each type of short-range interaction. Terms of third and higher order are estimated using the  $\delta^{\text{HF}}$  correction<sup>101</sup> which contains mainly higher-order induction contributions. Following prior work,<sup>71,85</sup> and for the purposes of fitting to the DFT-SAPT data, we keep the second-order induction term and the  $\delta^{\text{HF}}$  term separate.

Since the Slater-ISA and Born-Mayer force fields describe only short-range interactions (i.e. those terms which are modulated by the overlap of the monomer electron densities), they must both be supplemented with additional long-range terms that describe the electrostatic, polarization, and dispersion interactions. Here we have chosen a long-range potential of the form

$$V_{lr} = V_{\text{multipole}} + V_{\text{dispersion}} + V_{\text{shell}} \quad (26)$$

where

$$V_{\text{multipole}} = \sum_{ij} \sum_{tu} Q_t^i T_{tu} Q_u^j \quad (27)$$

includes distributed multipole contributions from each atom up to quadrupoles,

$$V_{\text{dispersion}} = - \sum_{ij} \sum_{n=3}^6 \frac{C_{ij,2n}}{r_{ij}^{2n}} \quad (28)$$

describes isotropic dispersion, and  $V_{\text{shell}}$  is the polarization energy modeled by Drude oscillators<sup>102,103</sup> as in ref. 71. The accuracy of each of these terms is expected to minimize errors in the long-range potential, simplifying the comparison between short-range force field functional forms. Nonetheless, we expect that our results will be qualitatively insensitive to the particular choice of long-range force field and acknowledge that simpler alternatives may be preferred for the development of highly efficient simulation potentials. In the case of the Lennard-Jones force field, we replace eq. (28) with the simple  $C_{ij,6}/r_{ij}^6$  dispersion term that is standard to the Lennard-Jones model.

We used a test set consisting of one atom (argon) and 12 small organic molecules (see Figure 2) from which dimer potentials could be generated (we will use the term ‘dimer’ to

mean two, potentially dissimilar, interacting molecules or atoms), yielding 91 dimer combinations (13 homo-monomeric, 78 hetero-monomeric). This wide range of systems allowed us to evaluate both the accuracy and transferability of the Slater-ISA model compared to conventional Born-Mayer and/or Lennard-Jones models.

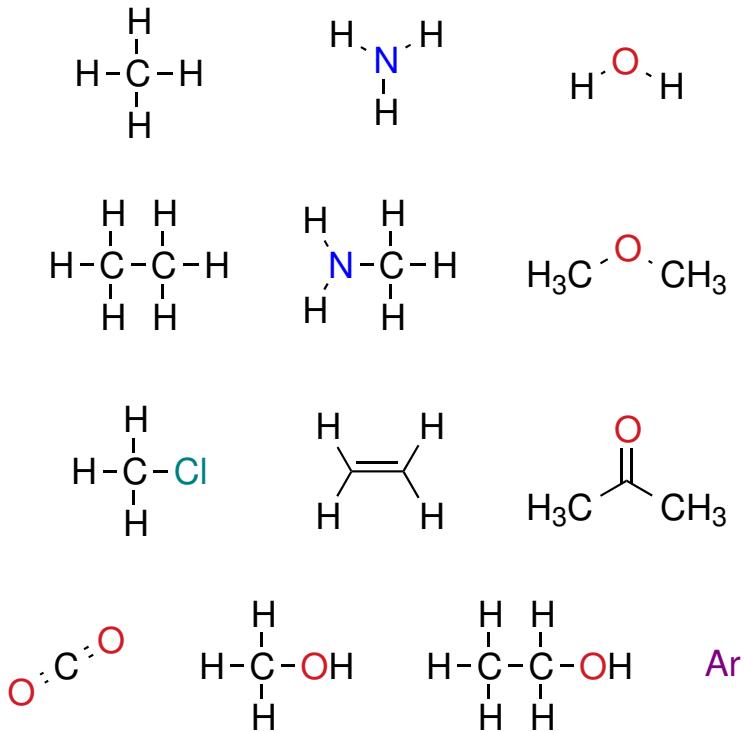


Figure 2: The 13 small molecules included in the 91 dimer (13 homomoneric, 78 heteromonomeric) test set. Cartesian geometries for all of these molecules are given in the Supporting Information.

A detailed description of this overall methodology is provided below.

### 3.1 Construction of the 91 dimer test set

Monomer geometries for each of the 13 small molecules were taken from the experimental NIST [CCCBDB] database<sup>104</sup> and can be found in the Supporting Information. For acetone and methyl amine, experimental geometries were unavailable, and thus the computational NIST [CCCBDB] database was used to obtain geometries at a high level of theory (B3LYP/aug-cc-pVTZ for acetone, CCSD(T)/6-311G\* for methyl amine). For each of the 91 dimers, a training set was constructed using DFT-SAPT (PBE0/AC) interaction energies calculated at 1000 quasi-random dimer configurations. These configurations were generated

using Shoemake's algorithm,<sup>105</sup> subject to the constraint that the nearest atom pairs be separated by between 0.75 and 1.3 of the sum of their van der Waals radii. This ensured adequate sampling of the potential energy surface in the region of the repulsive wall. The DFT-SAPT interaction energies were evaluated using an asymptotically corrected PBE0 functional (PBE0/AC) with monomer vertical (first) ionization potentials computed using the  $\Delta$ -DFT approach at a PBE0/aug-cc-pVTZ level of theory. Unless otherwise noted, all DFT-SAPT calculations used an aug-cc-pVTZ basis set in the dimer-centered form with midbond functions (the so-called DC+ form), and were performed using the MOLPRO2009 software suite.<sup>106</sup> The midbond set consisted of a 5s3p1d1f even-tempered basis set with ratios of 2.5 and centered at  $\zeta = 0.5, 0.5, 0.3, 0.3$  for the s,p,d, and f shells, respectively. This set was placed near the midpoint of the centers of mass of the two interacting monomers.

A small fraction of DFT-SAPT calculations exhibited unphysical energies, which were attributed to errors in generating the optimized effective potential used during the DFT-SAPT (PBE0/AC) calculations; these points were removed from the test set.

### 3.2 BS-ISA Calculations

BS-ISA atomic densities were obtained using CamCASP 5.8<sup>107–109</sup> following the procedure of Misquitta et al.<sup>11</sup> For the BS-ISA calculations, an auxiliary basis was constructed from an RI-MP2 aug-cc-pVTZ basis set with *s*-functions replaced by the ISA-set2 supplied with the CamCASP program; CamCASP's ISA-set2 basis was also used for the ISA basis set.<sup>11</sup> A custom ISA basis set for Ar was used (even tempered,  $n_{min} = -2, n_{max} = 8$ )<sup>11</sup> as no published basis was available. BS-ISA calculations were performed with the A+DF algorithm, which allows the ISA functional to be mixed with some fraction,  $\zeta$ , of the density-fitting functional. Following the recommendations of Misquitta et al.,<sup>11</sup> we have used  $\zeta = 0.1$  for the multipole moment calculations, and  $\zeta = 0.9$  for the density partitioning used to determine the  $B_{ij}$  coefficients.

### 3.3 Determination of $B_i^{\text{ISA}}$

The BS-ISA-derived atomic exponents,  $B_i^{\text{ISA}}$ , were obtained from a weighted least-squares fit to the spherically averaged BS-ISA atomic densities (shape functions),  $w_i(\mathbf{r})$ . In some cases, numerical instabilities and basis-set limitations of the BS-ISA procedure yielded densities that exhibited non-exponential asymptotic behavior.<sup>11</sup> To correct for these unphysical densities, we extrapolated the exponential decay of the valence region to describe the BS-ISA

tails also. Details of this procedure can be found in the Supporting Information. The ISA atom-in-molecule exponents were then derived via a log-weighted fit to the tail-corrected shape-functions  $w^a(\mathbf{r})$  for densities within the cutoff  $10^{-2} > w^a > 10^{-20}$  a.u. This region was chosen to reproduce the charge density most accurately in the valence regimes most likely to be relevant to intermolecular interactions.

### 3.4 Force Field Functional Forms and Parameterization

The general structure of the force fields  $V_{\text{FF}}$  for both the Slater-ISA FF and the Born-Mayer-type models are given by the following equations:

$$\begin{aligned}
 V_{\text{FF}} &= \sum_{ij} V_{ij}^{\text{exch}} + V_{ij}^{\text{elst}} + V_{ij}^{\text{ind}} + V_{ij}^{\delta^{\text{HF}}} + V_{ij}^{\text{disp}} \\
 V_{ij}^{\text{exch}} &= A_{ij}^{\text{exch}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) \\
 V_{ij}^{\text{elst}} &= -A_{ij}^{\text{elst}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + \sum_{tu} Q_t^i T_{tu} Q_u^j \\
 V_{ij}^{\text{ind}} &= -A_{ij}^{\text{ind}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + V_{\text{shell}}^{(2)} \\
 V_{ij}^{\delta^{\text{HF}}} &= -A_{ij}^{\delta^{\text{HF}}} P(B_{ij}, r_{ij}) \exp(-B_{ij} r_{ij}) + V_{\text{shell}}^{(3-\infty)} \\
 V_{ij}^{\text{disp}} &= - \sum_{n=3}^6 f_{2n}(x) \frac{C_{ij,2n}}{r_{ij}^{2n}} \\
 A_{ij} &= A_i A_j \\
 C_{ij,n} &= \sqrt{C_{i,n} C_{j,n}} \\
 f_{2n}(x) &= 1 - e^{-x} \sum_{k=0}^{2n} \frac{(x)^k}{k!}
 \end{aligned} \tag{29}$$

For the Slater-ISA FF:

$$\begin{aligned}
 B_i &= B_i^{\text{ISA}} \\
 B_{ij} &= \sqrt{B_i B_j} \\
 P(B_{ij}, r_{ij}) &= \frac{1}{3} (B_{ij} r_{ij})^2 + B_{ij} r_{ij} + 1 \\
 x &= B_{ij} r_{ij} - \frac{2B_{ij}^2 r_{ij} + 3B_{ij}}{B_{ij}^2 r_{ij}^2 + 3B_{ij} r_{ij} + 3} r_{ij}
 \end{aligned} \tag{30}$$

For all Born-Mayer type models:

$$\begin{aligned} P(B_{ij}, r_{ij}) &= 1 \\ x &= B_{ij}r_{ij} \end{aligned} \tag{31}$$

For the Born-Mayer-IP FF:

$$\begin{aligned} B_i &\equiv B_i^{\text{IP}} = 2\sqrt{2I_i} \\ B_{ij} &= \frac{B_i B_j (B_i + B_j)}{B_i^2 + B_j^2} \end{aligned} \tag{32}$$

For the Born-Mayer-sISA FF:

$$\begin{aligned} B_i &= 0.84 B_i^{\text{ISA}} \\ B_{ij} &= \sqrt{B_i B_j} \end{aligned} \tag{33}$$

Of the parameters in these force fields, only the coefficients  $A_i$  were fit to reproduce DFT-SAPT dimer energies (details below). All other force field parameters were derived from first-principles atom or atom-in-molecule properties. Exponents for the Slater-ISA FF and the Born-Mayer-sISA FF were derived from BS-ISA calculations, while exponents for the Born-Mayer-IP FF were determined from vertical ionization potentials of the isolated atoms. Dispersion coefficients ( $C_{ij,n}$ ) were either used directly from ref. 71 or were parameterized using analogous methods in the case of argon. Distributed multipoles  $Q_t^i$  for each system were obtained from the BS-ISA-based distributed multipoles scheme (ISA-DMA),<sup>11</sup> with the expansion truncated to rank 2 (quadrupole). Note that here,  $t = 00, 10, \dots, 22s$  denotes the rank of the multipole in the compact notation of Stone<sup>1</sup>. (In addition to rank 2 ISA-DMA multipoles, we also tested the use of DMA4 multipoles<sup>10</sup> as well as the use of rank 0 charges obtained from the rank truncation or transformation<sup>110</sup> of either ISA-DMA or DMA4 multipoles; the effect of including a Tang-Toennies damping factor<sup>71,75</sup> was studied in all cases. Each of these alternative long-range electrostatic models proved either comparably or less accurate for both the Slater-ISA FF and the Born-Mayer-IP FF in terms of their ability to reproduce the DFT-SAPT electrostatic energy, and are not discussed further.) Long-range polarization ( $V_{\text{shell}}$ ) was modeled using Drude oscillators in a manner identical to ref. 71. As in our prior work, during parameterization, the Drude energy was partitioned into 2<sup>nd</sup> ( $V_{\text{shell}}^{(2)}$ ) and higher order ( $V_{\text{shell}}^{(3-\infty)}$ ) contributions, where  $V_{\text{shell}}^{(2)}$  is the Drude oscillator

energy due to static charges (excluding intra-molecular contributions), and  $V_{\text{shell}}^{(3-\infty)}$  is the difference between the fully converged Drude energy,  $V_{\text{shell}}$ , and  $V_{\text{shell}}^{(2)}$ . Force field parameters for all homo-monomeric systems are located in the Supporting Information.

A weighted least-squares fitting procedure was used to fit  $A_i$  parameters to the benchmark DFT-SAPT (PBE0/AC) interaction energies on a component-by-component basis. That is, four separate optimizations<sup>71</sup> were performed to directly fit  $V^{\text{exch}}$ ,  $V^{\text{elst}}$ ,  $V^{\text{ind}}$ , and  $V^{\delta^{\text{HF}}}$  to, respectively, the following DFT-SAPT quantities (notation as in ref. 94):

$$\begin{aligned} E^{\text{exch}} &\equiv E_{\text{exch}}^{(1)} \\ E^{\text{elst}} &\equiv E_{\text{pol}}^{(1)} \\ E^{\text{ind}} &\equiv E_{\text{ind}}^{(2)} + E_{\text{ind-exch}}^{(2)} \\ E^{\delta^{\text{HF}}} &\equiv \delta(\text{HF}). \end{aligned} \quad (34)$$

For  $V^{\text{disp}}$ , no parameters were directly fit to the DFT-SAPT dispersion,

$$E^{\text{disp}} \equiv E_{\text{disp}}^{(2)} + E_{\text{disp-exch}}^{(2)}, \quad (35)$$

but were instead obtained solely from monomer properties as described above. Finally, note that no parameters were directly fit to the total DFT-SAPT energy,

$$E_{\text{int}} = E^{\text{exch}} + E^{\text{elst}} + E^{\text{ind}} + E^{\delta^{\text{HF}}} + E^{\text{disp}}, \quad (36)$$

for either the Slater-ISA FF or the Born-Mayer-IP FF. Rather,  $V_{\text{FF}}$  was calculated according to eq. (29).

Data points for each fit were weighted using a Fermi-Dirac functional form given by

$$w_i = \frac{1}{\exp((-E_i - \mu_{\text{eff}})/kT) + 1}, \quad (37)$$

where  $E_i$  is the reference energy and  $\mu_{\text{eff}}$  and  $kT$  were treated as adjustable parameters. The parameter  $kT$ , which sets the energy scale for the weighting function, was taken to be  $kT = \lambda|E_{\text{min}}|$ ; here  $E_{\text{min}}$  is an estimate of the global minimum well depth. Unless otherwise stated, we have used  $\lambda = 2.0$  and  $\mu_{\text{eff}} = 0.0$ . These defaults were chosen to minimize overall average attractive RMSE for all 91 dimer sets. Increases or decreases in the  $\lambda$  factor correspond to the weighting of more or fewer repulsive configurations, respectively.

In the case of Lennard-Jones, the standard Lennard-Jones functional form was used for

the van der Waals terms, with Coulomb and polarization terms modeled exactly as for the Slater-ISA FF:

$$V_{\text{FF}}^{\text{LJ}} = \sum_{ij} \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij,6}}{r_{ij}^6} + V_{\text{shell}} + \sum_{tu} Q_t^i T_{tu} Q_u^j \quad (38)$$

Lorentz-Berthelot combination rules were used to obtain heteroatomic  $A_{ij}$  and  $C_{ij}$  parameters. Unlike with the Slater-ISA FF and Born-Mayer models,  $V_{\text{FF}}^{\text{LJ}}$  was fit to the total DFT-SAPT (PBE0/AC) energy, with  $A_{ij}$  and  $C_{ij,6}$  as fitting parameters. The weighting function from eq. (37) was used in fitting.

### 3.5 Potential Energy Surface Scans

In order to visually assess fit quality, representative one-dimensional scans of the potential energy surface were calculated for several dimer pairs along low-energy dimer orientations. For each dimer pair, the minimum energy configuration of the 1000 random dimer points was selected as a starting configuration, and additional dimer configurations (not necessarily included in the original 1000 points) were generated by scanning along some bond vector. In the case of the ethane dimer, two carbon atoms (one on each monomer) were used; for acetone, the carbonyl carbon on each monomer defined the bond vector.

### 3.6 Molecular Simulations

All bulk simulations were run using OpenMM release version 7.0.<sup>111</sup> Enthalpies of vaporization were computed from

$$\Delta H_{\text{vap}} = (E_{\text{pot}}(g) + RT) - E_{\text{pot}}(l)$$

where  $E_{\text{pot}}(g)$  and  $E_{\text{pot}}(l)$  were determined from NVT simulations at the experimental gas and liquid densities, respectively. Calculated liquid densities were determined from NPT simulations. In all cases, the OPLS/AA force field was used for the intramolecular potential.<sup>112</sup> All simulations used a Langevin integrator with a 0.5 fs time step and a 1 ps<sup>-1</sup> friction coefficient; NPT simulations used a Monte Carlo barostat with a trial volume step every 5<sup>th</sup> move. Periodic boundary conditions, particle-mesh Ewald, and a non-bonding cutoff of 1.2nm with added long-range corrections were used to simulate a unit cell of 222 molecules. After an equilibration period of at least 600ps, simulation data was gathered from

production runs lasting at least 200ns.

## 4 Results and Discussion

The Slater-ISA methodology for short-range intermolecular interactions has been derived from a simple but rigorous physical model of overlapping monomer electron densities. In practice, this approach differs from the conventional Born-Mayer approach in both the choice of the short-range functional form (with the latter omitting the polynomial pre-factor) and the source of the exponents (with the former derived from ISA analysis of the monomer density). Our principal goal is to examine the influence of these modifications on the accuracy and transferability of the resulting force fields.

We initially benchmark the Slater-ISA FF against a conventional Born-Mayer potential, Born-Mayer-IP FF. The latter approach has been used extensively in prior work,<sup>33,71</sup> and both approaches use identical numbers of fitted parameters. Following prior work, combination rules for the Born-Mayer-IP FF are as in ref. 71. (We have tested the effect of using a geometric mean for the Born-Mayer-IP FF; results do not differ qualitatively from those presented below.) Owing to its popularity, we also compare the Slater-ISA FF to a Lennard-Jones functional form (LJ FF).

We first assess the accuracy of the Slater-ISA FF, Born-Mayer-IP FF, and LJ FF against benchmark ab initio intermolecular interaction energies and experimental 2<sup>nd</sup> virial coefficients, enthalpies of vaporization, and liquid densities. We next examine parameter transferability, assessing the extent to which parameters from pure homo-monomeric systems can be re-used (without further optimization) to describe mixed interactions. To assess parameter robustness, we also study the sensitivity of each methodology to changes in the weighting function (eq. (37)). Finally, we explore the application of BS-ISA-derived exponents within the Born-Mayer functional form as a straightforward method for simplifying the parameterization (and potentially increasing the accuracy) of a wide variety of standard ab initio and empirically-parameterized force fields.

### 4.1 Accuracy: Comparison with DFT-SAPT

For each of the 91 molecule pairs described in the Computational Methods section, parameters for the Slater-ISA FF, Born-Mayer-IP FF, and LJ FF were fit to reproduce DFT-SAPT (PBE0/AC) interaction energies calculated for a set of 1000 dimer configurations. These 91,000 total configurations and corresponding DFT-SAPT energies are collectively referred

to as the ‘91 dimer test set’. As a primary indication of accuracy, root-mean-square errors (RMSE) and mean signed errors (MSE), both with respect to DFT-SAPT, were computed for each methodology and for each dimer pair. Because these RMSE and MSE are dominated by repulsive contributions, and owing to the thermodynamic importance of attractive configurations, so-called ‘attractive RMSE/MSE’ were also computed by excluding net repulsive configurations (as measured by the DFT-SAPT total energy). The overall RMSE/MSE for all 91 dimers were then averaged to produce one ‘characteristic RMSE/MSE’ for the entire test set. Since these errors varied considerably in magnitude depending on the dimer in question, this overall average was taken in the geometric mean sense. (Results with an arithmetic mean do not differ qualitatively). Note that when computing the characteristic MSE, only the magnitude of each MSE,  $\|\text{MSE}\|$ , was considered.

Characteristic RMSE and  $\|\text{MSE}\|$  across the 91 dimer test set are shown in Figure 3 and Table 1. Overall, the Slater-ISA FF exhibits smaller errors compared to the Born-Mayer-IP FF. On average, the characteristic total energy RMSE for the Slater-ISA FF decrease by 33% relative to the Born-Mayer-IP FF. Even excluding repulsive configurations (dominated by short-range interactions), errors for the Slater-ISA FF are lower by 11% compared to the Born-Mayer-IP FF, demonstrating modest gains in accuracy even over the most energetically-relevant regions of the potential. A more detailed analysis of each of the 91 pairs of molecules shows that in an overwhelming 93% of such cases, force fields derived from the Slater-ISA method have smaller RMSEs compared to their Born-Mayer-IP counterparts (70% if only attractive configurations are considered). Regardless of the metric used, the Slater-ISA FF produces force fields with higher fidelity to the underlying benchmark interaction energies.

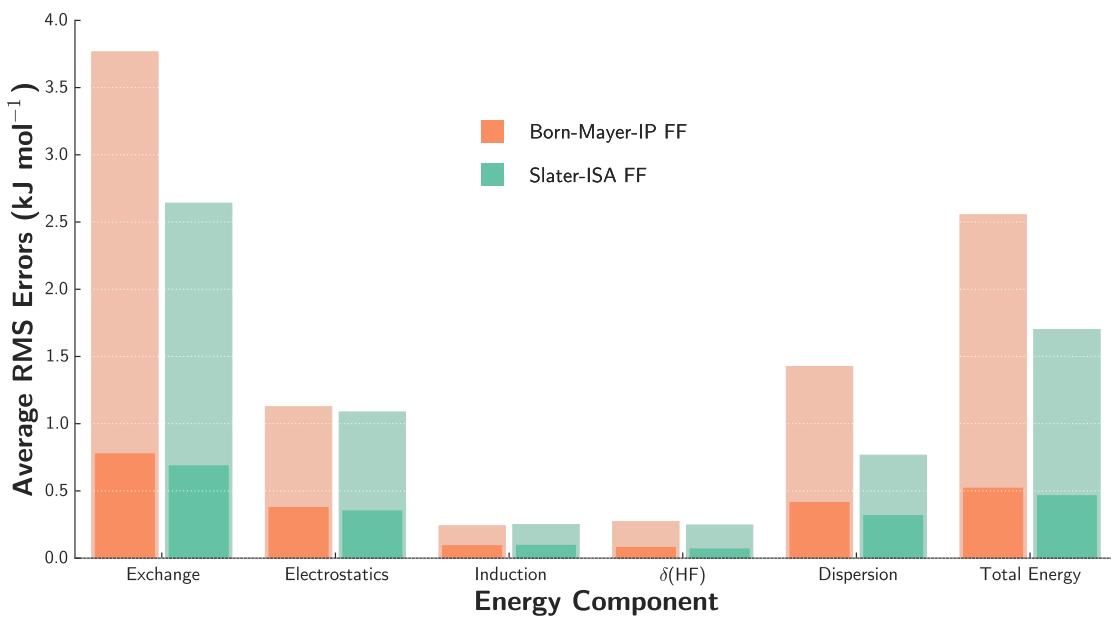


Figure 3: Characteristic RMSE (as described in the main text) for the Born-Mayer-IP FF (orange) and the Slater-ISA FF (green) over the 91 dimer test set. The translucent bars represent total RMSE for each energy component, while the smaller solid bars represent ‘Attractive’ RMSE, in which repulsive points have been excluded.

Table 1: Comparison of characteristic RMSE (as described in the main text) over the 91 dimer test set for the Slater-ISA FF, Born-Mayer-IP FF and LJ FF. For the total energy, both characteristic RMSE and MSE have been shown, with only the magnitude of the MSE,  $\|MSE\|$ , displayed. ‘Attractive’ RMSE, representing the characteristic RMSE for the subset of points whose energies are net attractive ( $E_{int} < 0$ ), are shown in parentheses to the right of the total RMS errors; ‘attractive’  $\|MSE\|$  are likewise displayed for the total energy. As discussed in Section 4.3, the ‘Dimer-Specific Fits’ refer to force fields whose parameters have been optimized for each of the 91 dimers separately, whereas the ‘Transferable Fits’ refer to force fields whose parameters have been optimized for the 13 homodimers and then applied (without further optimization) to the remaining 78 mixed systems. Unless otherwise stated, a default weighting function of  $\lambda = 2.0$  (see eq. (37)) has been used for all force fields in this work.

Component	Dimer-Specific Fits			Transferable Fits		
	Slater-ISA FF (kJ mol <sup>-1</sup> )	Born-Mayer-IP FF (kJ mol <sup>-1</sup> )	LJ FF (kJ mol <sup>-1</sup> )	Slater-ISA FF (kJ mol <sup>-1</sup> )	Born-Mayer-IP FF (kJ mol <sup>-1</sup> )	LJ FF (kJ mol <sup>-1</sup> )
Exchange	2.641 (0.686)	3.766 (0.775)	—	2.718 (0.720)	4.033 (0.836)	—
Electrostatics	1.087 (0.351)	1.126 (0.377)	—	1.134 (0.351)	1.231 (0.378)	—
Induction	0.251 (0.095)	0.241 (0.093)	—	0.278 (0.101)	0.265 (0.098)	—
$\delta^{HF}$	0.246 (0.068)	0.272 (0.079)	—	0.274 (0.076)	0.304 (0.081)	—
Dispersion	0.766 (0.317)	1.425 (0.414)	—	0.766 (0.317)	1.425 (0.414)	—
<b>Total Energy</b>						
<i>RMSE</i>	1.701 (0.464)	2.554 (0.520)	1.984 (0.603)	1.650 (0.456)	2.698 (0.555)	2.054 (0.640)
$\ MSE\ $	0.216 (0.057)	0.539 (0.127)	0.322 (0.345)	0.175 (0.051)	0.569 (0.112)	0.311 (0.368)

It is also instructive to consider each energy component individually. As might be expected, improvements in the description of  $E^{\text{exch}}$  are pronounced, with the characteristic RMSE from the Slater-ISA FF being 30% smaller than that from the Born-Mayer-IP FF. Examining each dimer pair separately (see Supporting Info for homo-monomeric fits, representative of the entire test set) we also find that, in general, the Slater-ISA FF is far better at reproducing *trends* in the exchange energy compared to the Born-Mayer-IP FF. This qualitative result is also reflected in the smaller  $\|\text{MSE}\|$  values for the Slater-ISA FF as compared to the Born-Mayer-IP FF. Nevertheless, there remains a fair amount of scatter in the exchange energies for several dimer pairs, particularly for molecules with exposed lone pairs or delocalized  $\pi$  systems. We hypothesize that this scatter is due to a breakdown of the isotropic approximation made in the Theory section, a conclusion supported by observations on the pyridine dimer system recently made by some of us.<sup>74</sup> It is therefore quite possible that the observed 30% RMSE reduction underestimates the true error reduction that might be observed if such anisotropy were accounted for.

From Figure 3, we see that the dispersion energy model from the Slater-ISA FF is also a substantial improvement; for dispersion, characteristic RMSE are 46% smaller for the Slater-ISA FF compared to the Born-Mayer model. This should not be a counter-intuitive result: while both potentials use identical dispersion coefficients, they differ in the damping model used. In the Born-Mayer-IP FF, the standard Tang–Toennies damping model is employed, and the damping parameters only depend on free atom ionization potentials; in the Slater-ISA FF, on the other hand, the damping parameters are obtained from the ISA shape functions, and thus take molecular environment effects into account. Even when only considering attractive dimer configurations (solid bar in Figure 3), errors in the dispersion energy component are reduced by 23%, demonstrating the importance of the damping function across the potential surface. From these results, and in agreement with related literature studies,<sup>113</sup> we conclude that use of the standard Tang-Toennies damping function based on atomic ionization potentials<sup>71,75,114–118</sup> lacks quantitative predictive power compared to the Slater-ISA model. Note that neither the Slater-ISA FF nor the Born-Mayer-IP FF are directly fitted to the DFT-SAPT dispersion energies (all parameters are determined from monomer properties), making this accuracy particularly striking. We hypothesize that the effect of the Slater-ISA approach is greater for dispersion than for first-order exchange because here (in contrast to the exchange energy) there are no fitted parameters to compensate for deficiencies in the exponents or functional form of the Born-Mayer-IP FF.

In contrast to the exchange and dispersion energies, the Slater-ISA FF and the Born-

Mayer-IP FF show nearly identical errors for the electrostatic and the induction ( $2^{\text{nd}}$  order induction plus  $\delta^{\text{HF}}$ ) energies. In these cases, the two models differ only in the parameters and functional form used to represent the exponentially-dependent short-range terms of these energy components, namely the penetration component for the electrostatic term and the penetration/charge-transfer term for the induction. The lack of improvement between the Slater-ISA and Born-Mayer-IP models may imply that we are not able to capture the physics of these particular short-range interactions with either the Slater-functional or Born-Mayer functional forms. Alternatively, the assumption that the short-range components of the electrostatic and induction energies are proportional to the exchange-repulsion may need to be re-examined. As discussed in Section 2.2, this proportionality is known to be approximately valid, but as yet there does not seem to be a deeper theoretical understanding of these short-range terms that may lead to a better model. Nevertheless, absolute errors in the electrostatic and induction components are relatively small for both models. Thus overall, the Slater-ISA FF functional form is promising for treating a wide variety of short-range effects.

Table 2: Comparison of characteristic RMSE and  $\|\text{MSE}\|$  over the 91 dimer test set for the various Lennard-Jones models. The LJ models are not parameterized on a component-by-component basis, thus RMSE/ $\|\text{MSE}\|$  values are only shown for the total FF energies. ‘Attractive’ errors, representing the characteristic RMSE/ $\|\text{MSE}\|$  for the subset of points whose energies are net attractive ( $E_{\text{int}} < 0$ ), are shown in parentheses to the right of the total errors. ‘Dimer-Specific Fits’ and ‘Transferable Fits’ are as in Table 1.

LJ FF Dimer-Specific Fits		LJ FF Transferable Fits	
$\lambda = 2.0$ (kJ mol $^{-1}$ )	$\lambda = 0.1$ (kJ mol $^{-1}$ )	$\lambda = 2.0$ (kJ mol $^{-1}$ )	$\lambda = 0.1$ (kJ mol $^{-1}$ )
<i>RMSE</i>	1.984 (0.603)	6.058 (0.413)	2.054 (0.640) 5.760 (0.457)
$\ \text{MSE}\ $	0.322 (0.345)	1.610 (0.041)	0.311 (0.368) 1.410 (0.060)

The comparison between the Slater-ISA FF and the LJ FF is slightly more complicated, owing to the differences in long-range potential and fitting methodology (see Section 3.4). As such, we compare the Slater-ISA FF to several versions of the LJ FF (for which characteristic RMSE and  $\|\text{MSE}\|$  are shown in Table 2). Using the same weighting function and constraining the Coulombic and polarization terms to be identical to the Slater-ISA FF, we see that the resulting Lennard-Jones force field (LJ FF,  $\lambda = 2.0$ ) is significantly worse than the Slater-ISA FF, both in terms of total RMSE and attractive RMSE. Furthermore, by comparing the  $\|\text{MSE}\|$  of both force fields, we see that errors in LJ FF are much more

*systematic* than in the Slater-ISA FF: in order to reproduce the repulsive wall correctly, the Lennard-Jones potential generally underestimates the well-depth by a considerable fraction (see the Supporting Information for ethane as a typical example).

Given the failure of the LJ FF ( $\lambda = 2.0$ ) force field to reproduce the energetically important region of the PES, we also compared the Slater-ISA FF to a ‘best-case’ scenario Lennard-Jones force field which correctly reproduces the minimum energy region at the expense of the repulsive wall. These LJ FF ( $\lambda = 0.1$ ) fits have total RMSE errors nearly 4 times that of the Slater-ISA FF; indeed, the LJ FF ( $\lambda = 0.1$ ) reproduces the repulsive wall only qualitatively. Insofar as the repulsive wall is concerned, the Slater-ISA FF is far superior to the Lennard-Jones short-range model. Nevertheless (and much more importantly for molecular simulation), the attractive region of the potential is reproduced surprisingly well by LJ FF. Characteristic attractive RMSE for the LJ FF ( $\lambda = 0.1$ ) are slightly lower than those for Slater-ISA FF, although the former has one additional free parameter per atom type and is also fit directly to reproduce the total energy. Likewise, attractive  $\|\text{MSE}\|$  between the Slater-ISA FF and the LJ FF ( $\lambda = 0.1$ ) are comparable. As we show in the Supporting Information, however, and as is well known in the literature, weighting the Lennard-Jones potential in this manner does not necessarily capture important information from the long-range attractive tail or repulsive wall of the PES, such that the LJ FF ( $\lambda = 0.1$ ) is not always expected to yield good property predictions. This latter point will be demonstrated in Section 4.2.

In order to compare the performance of the Slater-ISA FF against popular standard force fields, we also developed a ‘best case scenario’ non-polarizable point charge Lennard-Jones model, results for which are shown in the Supporting Information. Unsurprisingly, this force field is worse (in an RMSE and  $\|\text{MSE}\|$  sense) than all other force fields studied in this work, thus demonstrating how important accurate models for long-range electrostatics and polarization are to the overall accuracy of ab initio force fields.

#### 4.1.1 Argon Dimer

We now turn to several specific case studies. The Ar dimer provides an interesting test case to examine directly the impact of the polynomial pre-factor included in the Slater-ISA FF functional form. Since Ar is an atomic species, we should have  $B_{\text{Ar}}^{\text{ISA}} = B_{\text{Ar}}^{\text{IP}}$ . For numerical reasons, the Slater-ISA FF and Born-Mayer-IP FF exponents differ by 0.03 a.u.; however, this difference is insignificant, and the two FFs differ mainly in the polynomial pre-factor. Figure 4 shows the potential energy surface (PES) for the argon dimer computed using the

Slater-ISA FF and the Born-Mayer-IP FF. Here the default weighting scheme has been used so as to best reproduce the energetically attractive region. Note that, while both potentials reproduce the minimum energy configurations correctly, the Born-Mayer-IP FF considerably overestimates the exchange energy (and thus the total energy) along the repulsive wall. The Slater-ISA FF, on the other hand, maintains excellent accuracy in this region of the potential. This result is particularly notable because the repulsive wall is not heavily weighted in the fit. (A point 10 kJ mol<sup>-1</sup> along the repulsive wall, for instance, is weighted only 3% as heavily as a point near the bottom of the well). A similar, though smaller, increase in accuracy is seen in the fit to the DFT-SAPT dispersion energies, where the Slater-ISA FF is better able to model the energies for shorter interatomic separations. This increased accuracy is entirely attributable to the functional form employed, as the dispersion parameters are identical between the two FFs.

Consistent with prior literature,<sup>40,59</sup> these results suggest that neglect of the polynomial pre-factor  $P$  (as in standard Born-Mayer potentials) is *by itself* a poor approximation. However, as we show below, the Born-Mayer form can still be used as an accurate model in conjunction with appropriately scaled atomic exponents. Nonetheless, the more physically-motivated Slater form provides increased accuracy over a wider range of separations without recourse to empirical scaling.

Results for LJ FF are shown in the Supporting Information; consistent with expectations for the Lennard-Jones model, the repulsive wall is overestimated by the  $1/r_{ij}^{12}$  short-range functional form, and the magnitude of the attractive tail region is similarly overestimated by the effective  $C_{ij,6}$  dispersion parameter. Note that this  $C_{ij,6}$  coefficient has been fit to the total energy, and thus differs from the asymptotically-correct  $C_{ij,6}$  parameter used for both the Slater-ISA FF and the Born-Mayer-IP FF. An alternative parameterization strategy would have been to use the asymptotically-correct  $C_{ij,6}$  parameter in the LJ FF, but this would have worsened predictions along both the repulsive wall and the minimum energy configurations.

#### 4.1.2 Ethane Dimer

We next discuss the ethane dimer and show both a scatter plot of the 1000 dimer interactions (Figure 5) and a cut through the potential energy surface near the minimum (Figure 6) as indications of force field quality.

As with the argon dimer, for the ethane dimer the Slater-ISA FF produces more accurate exchange and dispersion energies compared to the Born-Mayer-IP FF. Here, the effects of

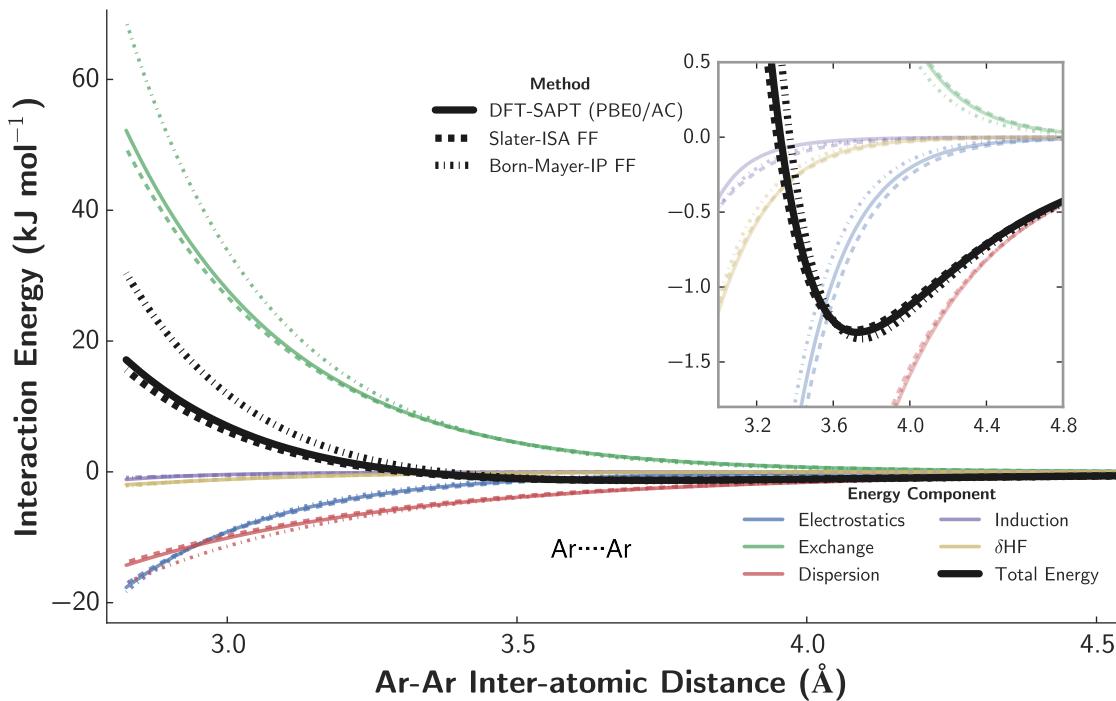


Figure 4: Potential energy surface for the argon dimer. Interaction energies for the Slater-ISA FF (dashed curves) and the Born-Mayer-IP FF (dash-dotted curves) are shown alongside benchmark DFT-SAPT (PBE0/AC) energies (solid curves). The energy decomposition for DFT-SAPT and for each force field is shown for reference.

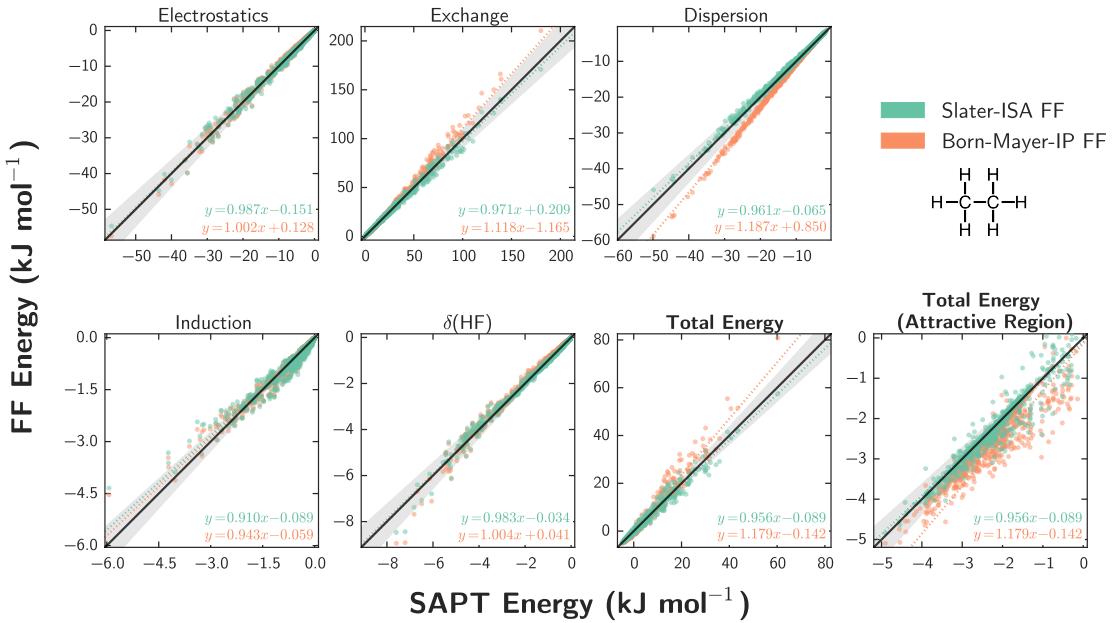


Figure 5: Force field fits for the ethane dimer using the Slater-ISA (green) and Born-Mayer-IP (orange) FFs. Fits for each energy component are displayed along with two views of the total interaction energy. The diagonal line (black) indicates perfect agreement between reference energies and each force field, while shaded grey areas represent points within  $\pm 10\%$  agreement of the benchmark. To guide the eye, a line of best fit (dotted line) has been computed for each force field and for each energy component.

the Slater-ISA FF for dispersion are even more pronounced, likely because the conventional damping of the Born-Mayer-IP FF is systematically in error due to differences in both the form of the damping function and exponents. As for the total interaction energy, we again find that the Born-Mayer-IP FF exhibits large errors for repulsive contributions, while the Slater-ISA FF naturally reproduces interactions for both attractive and strongly repulsive configurations. Even in the attractive regime, the Born-Mayer-IP FF is systematically too attractive. These systematic errors are the result of imperfect error cancellation between the exchange and dispersion components of the fit, and are discussed in more detail in Section 4.4.

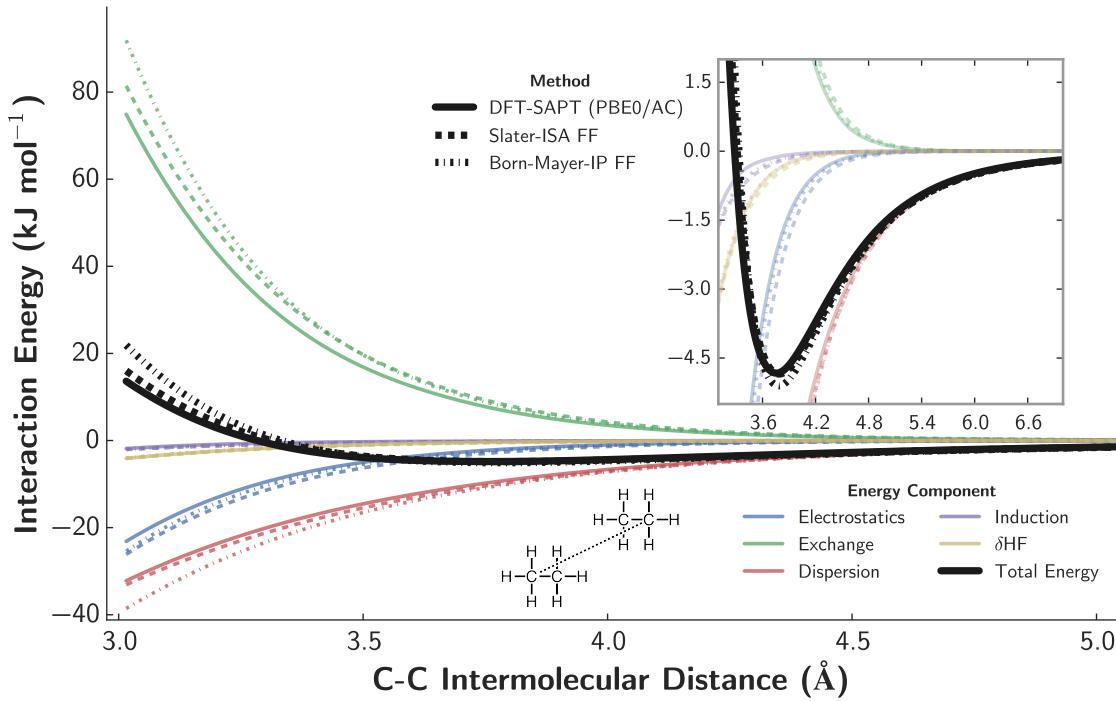


Figure 6: A representative potential energy scan near a local minimum for the ethane dimer. Interaction energies for the Slater-ISA FF (dashed curves) and the Born-Mayer-IP FF (dash-dotted curves) are shown alongside benchmark DFT-SAPT (PBE0/AC) energies (solid curves). The energy decomposition for DFT-SAPT and for each force field is shown for reference. The ethane dimer configuration in this scan corresponds to the most energetically attractive dimer included in the training set; other points along this scan are not included in the training set.

Examining a specific cut across the ethane-ethane PES (Figure 6) visually confirms these results. Both potentials do an excellent job of reproducing the benchmark DFT-SAPT energies in the minimum energy region, though the Born-Mayer-IP FF is slightly too attractive.

(Other cuts of the PES would show the Born-Mayer-IP predictions to be significantly more in error, consistent with the scatter plots). Along the repulsive wall, however, the Born-Mayer-IP FF predictions worsen in comparison to those from the Slater-ISA FF. Finally, the PES shows an increased reliance on error cancellation between the various energy components for the Born-Mayer-IP FF compared to the Slater-ISA FF.

As shown in the Supporting Information, the Lennard-Jones force field models are incapable of reproducing the entirety of the ethane PES; depending on the weighting function, either the repulsive wall or the attractive well can be reproduced, however no set of parameters can predict both regions simultaneously.

#### 4.1.3 Acetone Dimer

The acetone dimer provides a final interesting example involving a moderately sized organic molecule. From both the scatter plots (Figure 7) and the PES cross section (Figure 8), it is evident that both the Slater-ISA and Born-Mayer-IP force fields do an excellent job of reproducing DFT-SAPT energies for the low energy dimers. Along the repulsive wall, however, the Born-Mayer-IP FF shows larger systematic errors in each energy component, and seems to rely on error cancellation to achieve good agreement in the total energy. This reliance on error cancellation has two negative effects: Firstly, the additional scatter in the total energy of the Born-Mayer-IP FF fit, especially prominent for attractive configurations, indicates that this error cancellation is imperfect in certain cases. MSE for the Slater-ISA FF ( $-0.0115 \text{ kJ mol}^{-1}$ ) are an order of magnitude lower than for the Born-Mayer-IP FF ( $0.182 \text{ kJ mol}^{-1}$ ) in the attractive region of the potential. Secondly, as we shall later explore, reliance on error cancellation likely contributes to the somewhat decreased transferability of the Born-Mayer-IP FF as compared to the Slater-ISA FF.

As shown in the Supporting Information, the LJ FF predictions for acetone are reasonably good in both the tail and minimum energy regions of the potential, however the LJ FF grossly overpredicts the DFT-SAPT (PBE0/AC) energies along the repulsive wall.

## 4.2 Accuracy: Comparison with experiment

We have benchmarked the above force fields against experimental second virial coefficients and, in the case of ethane, enthalpies of vaporization and liquid densities. The classical 2<sup>nd</sup> virial coefficients were calculated for both argon and ethane using rigid monomer geometries, following the procedure described in ref. 71. Enthalpies of vaporization and liquid

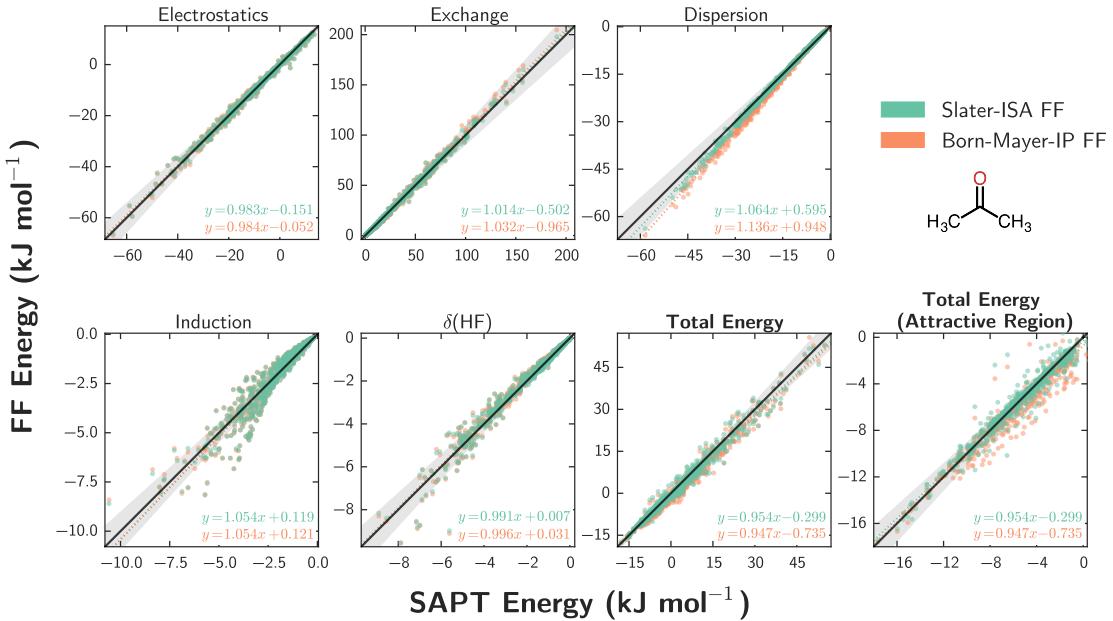


Figure 7: Force field fits for the acetone dimer using the Slater-ISA (green) and Born-Mayer-IP (orange) FFs, as in Figure 5.

densities were calculated using the OpenMM molecular simulation package<sup>111</sup> as described in Section 3. Higher-order multipole moments — which were negligible for these molecules — were neglected, and so only rank 0 terms were used in these calculations. Results are shown in Figures 9 and 10 as well as Table 4.

For argon, since both Slater-ISA FF and Born-Mayer-IP FF accurately reproduce the energetics of low-energy configurations, it is unsurprising that both force fields yield accurate virial coefficients over a wide range of temperatures. Errors in computed  $B_2$  coefficients (for both potentials) are likely attributable to small errors in the DFT-SAPT (PBE0/AC) potential itself,<sup>100</sup> and, to a much lesser extent, the neglect of nuclear quantum effects at lower temperatures.<sup>120</sup> Despite the good (in an RMSE sense) fit quality of the LJ FF ( $\lambda = 0.1$ ), this force field overpredicts the magnitude of the 2<sup>nd</sup> virial for argon, likely as a result of the effective dispersion coefficient, which overestimates the attraction in the tail region of the PES (see Supporting Information). Although it is certainly possible to parameterize a Lennard-Jones model *empirically* for argon, such a force field would rely on a subtle cancellation of errors between the minimum energy- and tail-regions of the PES. As the proper balance is impossible to predict a priori, this result highlights one of the difficulties of using the less physical LJ model in the development of ab-initio force fields.

In the case of ethane, the Slater-ISA FF is in excellent agreement with experiment,

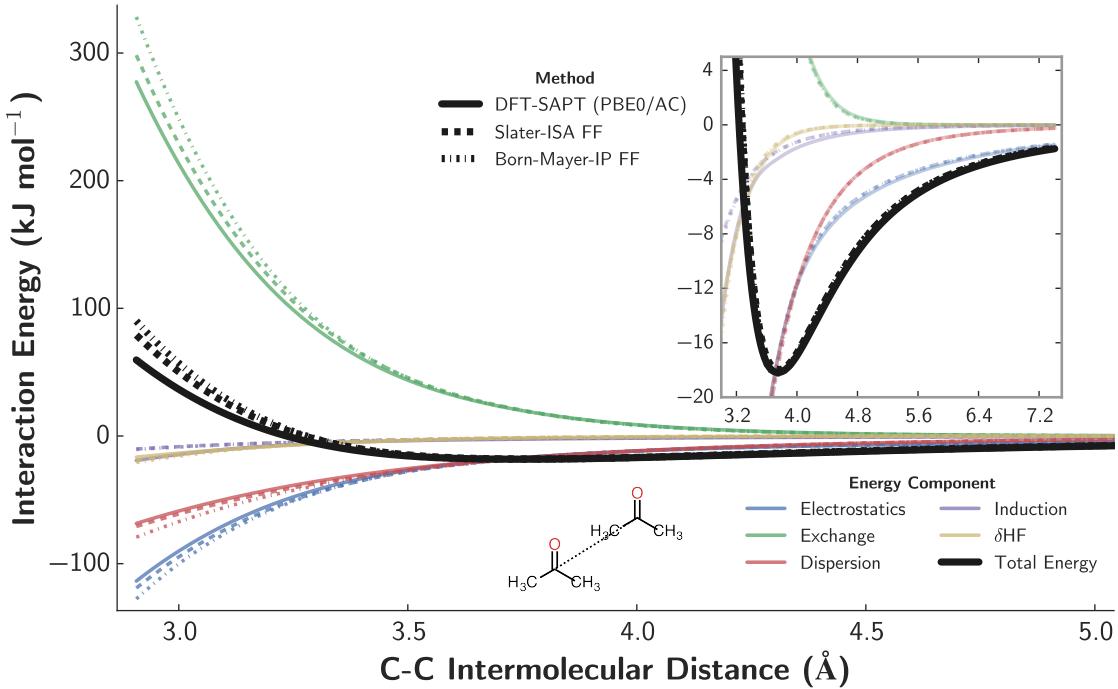


Figure 8: A representative potential energy scan near a local minimum for the acetone dimer. Interaction energies for the Slater-ISA FF (dashed curves) and the Born-Mayer-IP FF (dash-dotted curves) are shown alongside benchmark DFT-SAPT (PBE0/AC) energies (solid curves). The energy decomposition for DFT-SAPT and for each force field is shown for reference. The intermolecular distance is taken to be the internuclear distance between the two carbonyl carbons on each acetone monomer. The configuration in this scan corresponds to the most attractive dimer configuration included in the training set for the acetone dimer; other points along this scan have not explicitly been included in the training set.

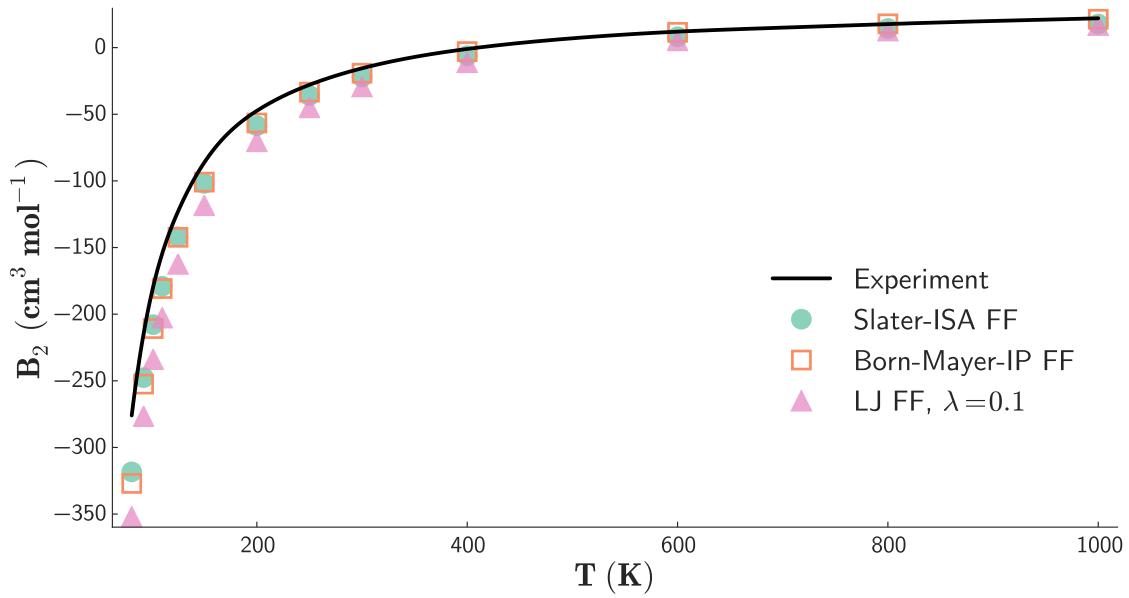


Figure 9: Second virial coefficients for argon. The Slater-ISA and the Born-Mayer-IP FFs are shown as green circles and orange squares, respectively; the black line corresponds to experiments from ref. 119.

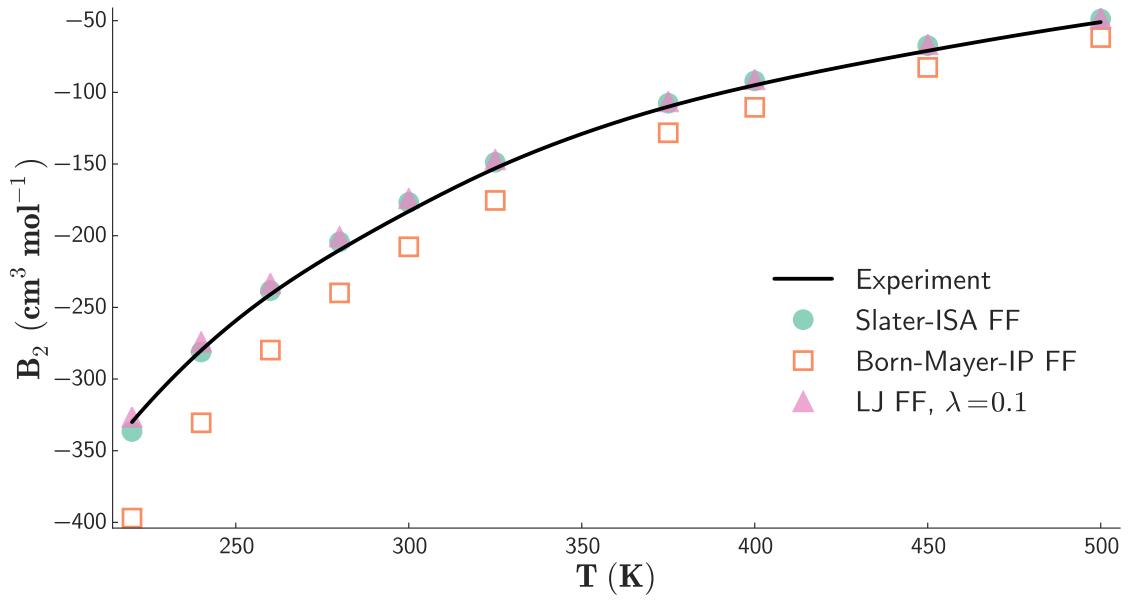


Figure 10: Second virial coefficients for ethane. The Slater-ISA and Born-Mayer-IP FFs are shown as green circles and orange squares, respectively; the black line corresponds to experiments from ref. 119.

whereas the Born-Mayer-IP FF underpredicts  $B_2$  by as much as 20%. These results are indicative, not only of the more accurate functional form and parameterization of Slater-ISA FF, but also of the high accuracy of the underlying DFT-SAPT (PBE0/AC) benchmark energies. In this case, LJ FF also correctly predicts the virial. Using weighting functions for each model that are optimal for the 91 dimer test set as a whole ( $\lambda = 2.0$  for the Slater-ISA FF and the Born-Mayer-IP FF,  $\lambda = 0.1$  for the LJ FF), all force fields produce similar results for  $\Delta H_{\text{vap}}$  and  $\rho$  (Table 4). These values are slightly overestimated by all force fields (especially in the case of the Born-Mayer-IP FF), which is to be expected given our neglect of many-body effects. McDaniel and Schmidt have calculated the 3-body correction for the Born-Mayer-IP FF; using this value as a global 3-body correction for all force fields, we see that both the Slater-ISA and the Lennard-Jones force fields compare very favorably to experiment, with the Slater-ISA FF perhaps slightly more accurate.

### 4.3 Transferability

The transferability of interaction potentials is a crucial aspect of practical molecular simulations. Here we examine ‘parameter transferability’, by which we mean the extent to which parameters from two homo-monomeric systems can combined to predict the intermolecular interactions of the resulting mixed hetero-monomeric system. As a measure of parameter transferability, we compared characteristic RMSE and  $\|\text{MSE}\|$  relative to the benchmark data for two different parameterization schemes. For the ‘Dimer-Specific Fits’,  $A_{ij}$  parameters were obtained for each of the 91 dimer pairs individually; these results are identical to those discussed in the previous two subsections. In contrast, for the ‘Transferable Fits’, the  $A_{ij}$  parameters were fit to the 13 homomeric dimer pairs and were re-used (without any further optimization) to calculate energies for the 78 mixed systems using the combination rules listed in Section 3.4. Results for each parameterization scheme are shown in Table 1. From the RMSE and  $\|\text{MSE}\|$  from the competing schemes, we see excellent parameter transferability for all force fields studied. For the Slater-ISA FF, characteristic RMSE and  $\|\text{MSE}\|$  for each component increase by a very small fraction upon constraining the fit; due to small error cancellation, errors in the total energy actually *decrease* somewhat with these constraints. (This is possible since the total energy is not directly fit.) The Born-Mayer-IP FF also displays a significant degree of transferability, though errors in the total energy increase slightly upon constraining the fit. As in prior work, the observed parameter transferability for both force fields can be attributed to our use of a term-by-term parameterization scheme (Section 3.4), which serves to minimize error cancellation between energy components and

generate a more physically-meaningful (and thus transferable) set of parameters.<sup>71,122</sup> Finally, note that for four of the five interaction energy components the relative change in RMSE on constraining the fit is smaller for the Slater-ISA FF than the Born-Mayer-IP FF. The  $\delta^{\text{HF}}$  term is the exception, but even here the relative change in errors from the two methods are comparable. This suggests that the Slater-ISA FF may be the more transferable of the force fields studied. Nevertheless, the Lennard-Jones model is surprisingly transferable, likely in part due to the same accurate and transferable ‘long-range’ electrostatics and polarization as the Slater-ISA FF. The non-polarizable, point-charge Lennard-Jones model (results for which are shown in the Supporting Information) displays the least transferability (in both an RMSE and  $\|\text{MSE}\|$  sense) of all force fields studied.

Although we do not examine it here, we expect that the previously demonstrated success<sup>59,71,121,122</sup> of the Born-Mayer-IP FF with respect to ‘environment transferability’ — the extent to which a single set of parameters can model a variety of phases and molecular environments — and ‘atom type transferability’ — the extent to which atoms in chemically similar environments can accurately be grouped together into ‘types’ and treated using one parameter set — would also apply to, or even be improved by, Slater-ISA FF. These issues are under investigation in our groups.

#### 4.4 Robustness

One of the practical challenges of ab initio force field development is the robustness of the resulting force field quality with respect to the choice of an appropriate training set and/or weighting function. To this end, the default weighting function (eq. (37),  $\lambda = 2.0$ ) was varied to produce unconstrained fits that were skewed either towards attractive ( $\lambda = 0.5$ ) or repulsive ( $\lambda = 5.0$ ) configurations, and pairwise differences in force field total energies were computed between each weighting scheme. Characteristic root-mean-square pairwise differences (RMSD) between each weighting function are shown in Table 3; as before, ‘attractive RMSD’ were calculated by excluding repulsive points from consideration. Note that, on average, the default  $\lambda = 2.0$  weighting scheme is optimal (in an RMSE sense) for both the Slater-ISA and Born-Mayer-IP FFs.

Overall, both the Born-Mayer-IP FF and the LJ FF display significant weighting function sensitivity. This sensitivity is not surprising; as both force fields are unable to reproduce the entirety of the potential energy surface, changing the weighting scheme (or equivalently, the balance of configurations in the training set) alters the parameters in the Born-Mayer-IP FF or the LJ FF models quite substantially. Even excluding repulsive configurations, RMSD of

Table 3: Characteristic RMS pairwise differences (RMSD) in force field total energies for different weighting functions with  $\lambda$  values as defined in eq. (37); values shown are the (arithmetic mean, rather than geometric) RMSD across the 91 dimer test set. Characteristic ‘Attractive’ RMSD (as defined in Table 1) are shown in parentheses to the right of each overall RMSD.

Characteristic RMSD	$\lambda = 0.5$ vs 2.0 (kJ mol $^{-1}$ )	$\lambda = 0.5$ vs 5.0 (kJ mol $^{-1}$ )	$\lambda = 2.0$ vs 5.0 (kJ mol $^{-1}$ )
Slater-ISA FF	0.742 (0.207)	0.990 (0.273)	0.306 (0.086)
Born-Mayer-IP FF	1.866 (0.409)	2.632 (0.550)	0.797 (0.153)
LJ FF	1.301 (0.216)	1.605 (0.309)	0.324 (0.099)
Born-Mayer-sISA FF	0.611 (0.178)	0.810 (0.236)	0.293 (0.081)

$\sim 0.5$  kJ mol $^{-1}$  are typical for the Born-Mayer-IP FF. RMSD are somewhat smaller for the LJ FF ( $\sim 0.3$  kJ mol $^{-1}$ ), however qualitatively we see that differences in computed force field energies are systematic: smaller weighting functions capture the minimum energy region of the potential while overestimating the magnitudes of both the repulsive and tail regions of the potential, whereas larger weighting functions tend to underestimate the minimum energy region in order to correctly reproduce the repulsive wall. Consequently, the Lennard-Jones model shows weighting-function sensitivity in a manner that is not entirely captured by the RMSD, but is instead reflected in the greater sensitivity of the LJ FF (as compared to the Born-Mayer-IP FF) in the prediction of experimental properties (*vide infra*).

Note that for practical force field development (as opposed to minimization of overall RMSE), the default weighting scheme for the Born-Mayer-IP FF and the LJ FF is suboptimal for many dimers in the test set. Because both the Born-Mayer-IP FF and the LJ FF must inherently compromise between accuracy near the minimum and along the repulsive wall, the weighting function requires system-specific fine-tuning in order to achieve proper balance. This empiricism creates significant challenges in the development of ab initio force fields.

By contrast, we find the Slater-ISA FF to be robust with respect to the choice of weighting function due to its more balanced treatment of repulsive and attractive regions of the potential energy surface. Average RMSD for the Slater-ISA FF are between two to three *times* smaller compared to the Born-Mayer-IP FF, and the Slater-ISA FF is relatively insensitive to the choice of weighting function. These conclusions hold for both attractive and overall RMSD. As a result, the Slater-ISA model largely eliminates the need for empirical fine-tuning of the weighting function, which in turn greatly simplifies the parameterization process and allows for a more robust prediction of chemical and physical properties.

For the ethane dimer, Figure 11 shows overall force field energies for both the Slater-ISA and Born-Mayer-IP FFs for three weighting functions. Results for the Lennard-Jones models are shown in the SI, and are qualitatively similar to the Born-Mayer-IP FF results. The Born-Mayer-IP FF fits vary qualitatively with  $\lambda$ , leading to a relatively large uncertainty in calculated  $B_2$  coefficients, enthalpies of vaporization, and liquid densities (see Table 4). By skewing the fits towards attractive configurations ( $\lambda = 0.5$ ), the majority of attractive configurations are predicted without systematic error, though points along the repulsive wall (including those with net negative energies) are systematically too repulsive. Using a scheme which more heavily weights repulsive configurations, the Born-Mayer-IP FF regains semi-quantitative accuracy for repulsive configurations, albeit at the expense of a systematic increase in errors for the attractive dimer configurations. Finally, we reiterate that the optimal weighting function for the ethane dimer (here  $\lambda = 0.5$  best reproduces the 2<sup>nd</sup> virial for the Born-Mayer-IP FF) is by no means universal for the molecules in the 91 dimer test set.

Table 4: Enthalpies of vaporization and liquid densities for ethane as a function of force field and weighting function. Values in parentheses include an estimation of the 3-body correction (0.628 kJ mol<sup>-1</sup> and 0.034 g mL<sup>-1</sup> for the enthalpy of vaporization and liquid density, respectively) as computed in ref. 121. Experimental data taken from ref. 123 and ref. 124.

Force Field	Weighting Function				Experiment
	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 2.0$	$\lambda = 5.0$	
<b><math>\Delta H_{\text{vap}}</math> (kJ mol<sup>-1</sup>); <math>\rho = 0.546</math> g L<sup>-1</sup>, <math>T = 184</math> K</b>					
Slater-ISA FF	15.3 (14.7)	15.3 (14.6)	15.3 (14.7)	15.2 (14.6)	
Born-Mayer-IP FF	14.3 (13.7)	15.1 (14.5)	16.6 (15.9)	18.6 (18.0)	14.7
LJ FF	15.5 (14.9)	14.6 (13.9)	11.4 (10.7)	10.1 (9.5)	
<b><math>\rho</math> (g L<sup>-1</sup>); <math>P = 1</math> atm, <math>T = 184</math> K</b>					
Slater-ISA FF	0.600 (0.566)	0.602 (0.568)	0.600 (0.566)	0.593 (0.559)	
Born-Mayer-IP FF	0.521 (0.487)	0.567 (0.533)	0.632 (0.598)	0.678 (0.644)	0.546
LJ FF	0.607 (0.573)	0.610 (0.576)	0.555 (0.521)	0.494 (0.460)	

The Slater-ISA FF fits for the ethane dimer, on the other hand, are nearly completely insensitive to the weighting function, leading to little intrinsic uncertainty in the determination of parameters or in the computation of macroscopic properties. Some other dimers, particularly those where atomic anisotropy would be anticipated (e.g., water), exhibited slightly larger sensitivity to the weighting function. Nevertheless, the vast majority of dimers in the

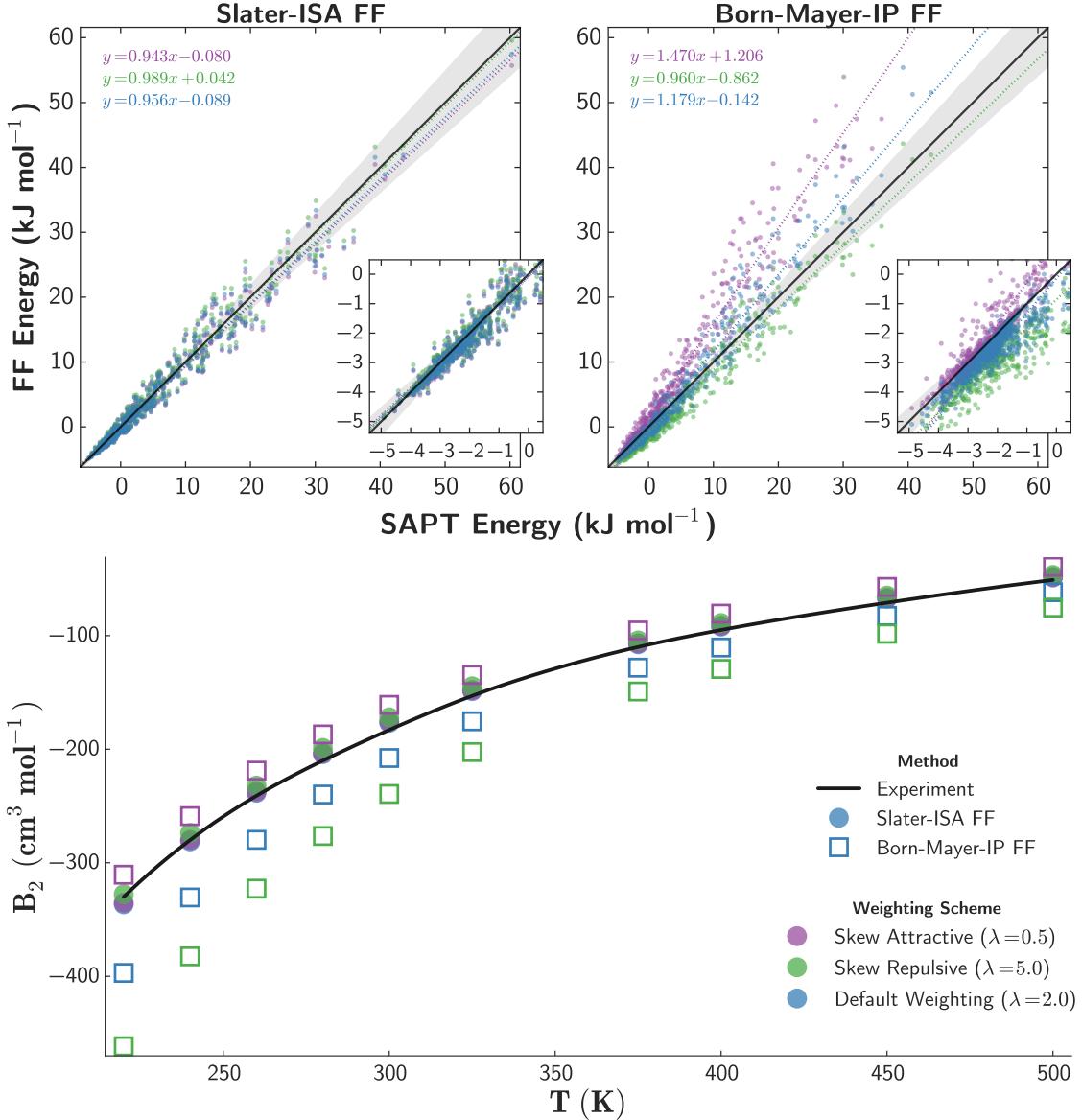


Figure 11: Comparison of the Slater-ISA FF and the Born-Mayer-IP FF in terms of sensitivity to the weighting function employed in parameter optimization for the ethane dimer. Three weighting functions,  $\lambda = 0.5$  (purple),  $\lambda = 2.0$  (blue), and  $\lambda = 5.0$  (green) are shown, with higher  $\lambda$  values indicating more weighting of repulsive configurations.

(top) Total interaction energies for the Slater-ISA FF (left) and the Born-Mayer-IP FF (right) indicating the accuracy of each force field with respect to DFT-SAPT (PBE0/AC) benchmark energies. The diagonal line (black) indicates perfect agreement between reference energies and each force field, while shaded grey areas represent points within  $\pm 10\%$  agreement of the benchmark. To guide the eye, a line of best fit (dotted line) has been computed for each force field and for each weighting function.

(bottom) Computed 2<sup>nd</sup> virial coefficients for ethane. Data for the Slater-ISA FF and the Born-Mayer-IP FF are depicted using shaded circles and open squares, respectively; colors for the different weighting functions are as above. Experimental data from ref. 119 (black line) is also shown.

test set are qualitatively insensitive to the choice of weighting function, and can be optimized with the default  $\lambda = 2.0$  weighting function without yielding undue systematic error in the attractive region of the potential, thus proving the enhanced robustness of the Slater-ISA FF model relative to conventional force fields.

## 4.5 Next-Generation Born-Mayer Models: Born-Mayer-sISA FF

We hypothesize that the increased accuracy, transferability, and robustness of the Slater-ISA FF is a direct result of its more physically-motivated functional form and its use of ISA-derived atomic exponents that directly account for the influence of the molecular environment. Nonetheless, we recognize that the standard Born-Mayer functional form remains extremely common, both in simulation software and in existing force fields. It is therefore fruitful to explore the extent to which the BS-ISA exponents themselves could be used in conjunction with a Born-Mayer functional form. These results are shown in Table 5.

Table 5: Comparison of characteristic RMSE (as described in the main text) over the 91 dimer test set for the Born-Mayer-sISA approximation compared with other methods. For the total energy, both RMSE and absolute mean signed errors (MSE) have been shown. ‘Attractive’ RMSE, representing the characteristic RMSE for the subset of points whose energies are net attractive ( $E_{\text{int}} < 0$ ), are shown in parentheses to the right of the total RMS errors; ‘attractive’  $\|MSE\|$  are likewise displayed for the total energy. Slater-ISA FF, Born-Mayer-ISA, and Born-Mayer-sISA FF are as described in the main text, and the ‘Dimer-Specific’ and ‘Transferable’ fits are as described in Table 1.

Component	Dimer-Specific Fits			Transferable Fits		
	Slater-ISA FF (kJ mol <sup>-1</sup> )	Born-Mayer-ISA (kJ mol <sup>-1</sup> )	Born-Mayer-sISA (kJ mol <sup>-1</sup> )	Slater-ISA FF (kJ mol <sup>-1</sup> )	Born-Mayer-ISA (kJ mol <sup>-1</sup> )	Born-Mayer-sISA (kJ mol <sup>-1</sup> )
Exchange	2.641 (0.686)	7.030 (1.203)	2.677 (0.686)	2.718 (0.720)	6.968 (1.228)	2.764 (0.706)
Electrostatics	1.087 (0.351)	1.406 (0.589)	1.083 (0.352)	1.134 (0.351)	1.461 (0.598)	1.141 (0.352)
Induction	0.251 (0.095)	0.229 (0.097)	0.250 (0.096)	0.278 (0.101)	0.257 (0.101)	0.275 (0.101)
$\delta^{\text{HF}}$	0.246 (0.068)	0.327 (0.120)	0.248 (0.068)	0.274 (0.076)	0.353 (0.122)	0.274 (0.076)
Dispersion	0.766 (0.317)	3.584 (0.890)	0.856 (0.336)	0.766 (0.317)	3.584 (0.890)	0.856 (0.336)
<b>Total Energy</b>						
<i>RMSE</i>	1.701 (0.464)	4.934 (1.054)	1.751 (0.453)	1.650 (0.456)	4.555 (1.035)	1.713 (0.446)
$\ MSE\ $	0.216 (0.057)	1.127 (0.505)	0.258 (0.063)	0.175 (0.051)	0.882 (0.516)	0.245 (0.057)

As expected, direct insertion of the BS-ISA exponents into the Born-Mayer functional form (Born-Mayer-ISA) does not yield promising results. Indeed, the Born-Mayer-ISA FF has significantly worse RMSE and  $\|\text{MSE}\|$  than the Born-Mayer-IP FF. We reiterate that the  $P = 1$  approximation from eq. (25), yielding the conventional Born-Mayer form, is by itself a crude model. Rather, it becomes necessary to accompany this approximation by a corresponding exponent scale factor,  $\xi$ :

$$B_i = \xi B_i^{\text{ISA}}. \quad (39)$$

Following literature precedent,<sup>40,59</sup> we hypothesized that  $\xi$  could be treated as a universal constant. To test this conjecture, we computed reference density overlaps for a variety of isolated atom pairs (details in the Supporting Information), and fitted each of these overlaps to a Born-Mayer function of the form  $S_{ij} \approx K_{ij} \exp(-\xi B_{ij}^{\text{ISA}} r_{ij})$ , where  $K_{ij} = \frac{K}{B_{ij}^3}$  in line with eq. (13). To very good approximation, both  $K$  and  $\xi$  can be treated as universal constants; that is, neither  $K$  nor  $\xi$  is sensitive to the value of  $B^{\text{ISA}}$ . However, fitted values of  $K$  and  $\xi$  do depend strongly on the range of  $r_{ij}$  values used in the optimization, yielding estimates ranging from 0.74 to 0.88.

As an alternative, we optimized  $\xi$  directly by minimizing RMSE against the 91 dimer test set. Results from various choices of  $\xi$  can be found in the Supporting Information. In agreement with prior literature and our ‘first-principles’ analysis of overlaps, we find  $\xi = 0.84$  to be optimal for minimizing characteristic overall and attractive RMSE, though in practice the errors are insensitive to  $\xi \in [0.82, 0.86]$ . We henceforth use  $\xi = 0.84$  and refer to this force field methodology (Born-Mayer functional form, ISA-derived exponents with scale factor  $\xi = 0.84$ ) as the Born-Mayer-sISA FF. Parameters and homo-monomeric fits for the Born-Mayer-sISA FF can be found in the Supporting Information.

From Table 5 we see that the Born-Mayer-sISA FF is comparable in quality to our original Slater-ISA FF methodology. For all attractive configurations, the Born-Mayer-sISA FF is equally accurate and transferable (Table 5). Furthermore, as shown in Table 3, Born-Mayer-sISA FF displays similar parameter robustness to Slater-ISA FF. These results suggest that many of the advantages of the Slater-ISA FF procedure can be captured simply by using the (scaled) ISA exponents. Note, however, that the optimal scale factor likely exhibits some system dependence, and furthermore that the enhanced Slater functional form may be important where an accurate description of highly repulsive configurations is crucial.

We also examined the Slater-ISA FF and the Born-Mayer-sISA FF against force fields where  $B_i$  values were instead treated as soft constraints, rather than fixed parameters. Using

entirely unconstrained exponents yields unphysical parameters and a severe degradation in force field transferability. Using exponents from the Slater-ISA FF and the Born-Mayer-ISA FF as Bayesian priors (in the sense used in refs. 38, 74), we generated two new force fields with optimized exponents, denoted Slater-OPT and Born-Mayer-OPT, respectively. Characteristic RMSE and  $\|\text{MSE}\|$  for these force fields can be found in the Supporting Information. We find that both methods yield only very minimal improvement, suggesting that the first-principles ISA exponents are already nearly optimal. Comparing the Born-Mayer-OPT exponents to those from Slater-ISA, we find a nearly identical average scale factor of  $\gamma = 0.83 \pm 0.07$ . Given that these optimal exponents can now be generated directly from first principles calculations of the molecular densities via the BS-ISA approach of Misquitta et al., we anticipate that the BS-ISA densities and resulting ISA exponents will be extremely useful in next-generation force field development in order to greatly simplify force field parameterization.

## 5 Conclusions and Recommendations

We have presented a new methodology for describing short-range intermolecular interactions based upon a simple model of atom-in-molecule electron density overlap. The resulting Slater-ISA FF is a simple extension of the conventional Born-Mayer functional form, supplemented with atomic exponents determined from an ISA analysis of the molecular electron density. In contrast to simple Born-Mayer or Lennard-Jones models, the Slater-ISA FF is capable of reproducing ab initio interaction energies over a wide range of inter-atomic distances, and displays extremely low sensitivity to the details of parameterization. Furthermore, the Slater-ISA FF exhibits excellent parameter transferability. We thus recommend Slater-ISA FF for use in the development of future ab initio (and possibly empirically-parameterized) potentials, particularly where accuracy across wide regions of the potential surface is paramount.

More generally, we find that analysis of the ISA densities provides an excellent first-principles procedure for the determination of atomic-density decay exponents. This analysis improves upon existing approaches (which rely upon exponents derived from atomic radii or ionization potentials)<sup>57,125–127</sup> and explicitly incorporates the influence of the molecular environment. These exponents can be used within Slater-ISA FF without further parameterization. Alternatively, in conjunction with an appropriate scale factor, the exponents can be used to enhance the accuracy of standard Born-Mayer potentials and/or Tang-Toennies

damping functions. The resulting Born-Mayer-sISA FF retains many of the advantages of Slater-ISA FF, but also maintains compatibility with existing force fields and simulations packages that do not support the Slater functional form. Given that the BS-ISA exponents appear to be essentially optimal with respect to additional empirical optimization, we strongly recommend use of these first-principles exponents in order to simplify (both ab initio and empirical) future force field development involving Born-Mayer or related functional forms.<sup>21</sup>

Overall, Slater-ISA FF enables a significantly increase in force field accuracy, particularly in describing short intermolecular contacts. Nevertheless, the neglect of atomic anisotropy remains, in some cases, a severe approximation.<sup>128–130</sup> Indeed, it has been shown by many authors<sup>1,69,77,116</sup> that quantitatively accurate  $A_{ij}$  parameters (and to a lesser extent,  $B_{ij}$  parameters) require incorporation of angular dependence for the generation of highly-accurate force fields. This anisotropy becomes crucial when describing systems containing lone pairs, hydrogen bonds, and/or  $\pi$ -interactions. Promisingly, BS-ISA densities naturally describe such anisotropy,<sup>38,74,131</sup> and a straightforward method for its inclusion (where essential) in ab initio force fields is the subject of ongoing work.

## Acknowledgement

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1256259 and by Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy, under award DE-SC0014059. J.R.S is a Camille Dreyfus Teacher-Scholar. M.V.V. thanks Dr. Jesse McDaniel for helpful discussions. Computational resources were provided in part by National Science Foundation Grant CHE-0840494 and using the compute resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. Computational resources were also provided in part by the UW Madison Chemistry Department cluster Phoenix under grant number CHE-0840494.

## Supporting Information Available

Waldman-Hagler analysis of tested  $B_{ij}$  combination rule(s). Force field fit quality for the ‘exact’ overlap model. Extrapolation algorithm for ISA exponents. Non-polarizable, point-charge Lennard-Jones Force Fields. Scale factor tests for Born-Mayer-sISA. Force field fit qualities and comparisons for Slater-OPT and Born-Mayer-OPT force fields. Geometries, ionization potentials, and HOMO values for each monomer species. Force field parameters for homomonomeric systems for each Slater-ISA FF, Born-Mayer-IP FF, and Born-Mayer-sISA FF. Force field fit quality results for homomonomeric systems. Force field accuracy tests for LJ FF. Weighting function parameter robustness tests for the argon dimer for the Slater-ISA FF and the Born-Mayer-IP FF. Weighting function parameter robustness tests for the ethane dimer for the LJ FF. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## References

- (1) Stone, A. J. *The Theory of Intermolecular Forces*, 2nd ed.; OUP Oxford, 2013.
- (2) Margenau, H.; Kestner, N. R. *Theory of Intermolecular Forces*; International series of monographs in natural philosophy; Pergamon Press: Oxford, 1969.
- (3) Riley, K. E.; Pitončák, M.; Jurecčka, P.; Hobza, P. *Chem. Rev.* **2010**, *110*, 5023–5063.
- (4) Stone, A. J.; Misquitta, A. J. *Int. Rev. Phys. Chem.*; 2007; Vol. 26; pp 193–222.
- (5) Dykstra, C. E.; Lisy, J. M. *J. Mol. Struct. THEOCHEM* **2000**, *500*, 375–390.
- (6) Stone, A. J.; Tough, R. *Chem. Phys. Lett.* **1984**, *110*, 123–129.
- (7) Williams, G. J.; Stone, A. J. *J. Chem. Phys.* **2003**, *119*, 4620–4628.
- (8) Misquitta, A. J.; Stone, A. J. *J. Chem. Phys.* **2006**, *124*, 024111.
- (9) Dehez, F.; Chipot, C.; Millot, C.; Ángyán, J. G. *Chem. Phys. Lett.* **2001**, *338*, 180–188.
- (10) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.
- (11) Misquitta, A. J.; Stone, A. J.; Fazeli, F. *J. Chem. Theory Comput.* **2014**, *10*, 5405–5418.
- (12) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887–1930.
- (13) Szalewicz, K. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 254–272.

- (14) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479–483.
- (15) Grimme, S.; Djukic, J. P. *Inorg. Chem.* **2011**, *50*, 2619–2628.
- (16) Lennard-Jones, J. *Proc. Phys. Soc.* **1931**, *43*, 461–482.
- (17) Born, M.; Mayer, J. E. *Zeitschrift für Phys.* **1932**, *75*, 1–18.
- (18) Buckingham, R. A. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1938**, *168*, 264–283.
- (19) Nezbeda, I. *Mol. Phys.* **2005**, *103*, 59–76.
- (20) Galliero, G.; Boned, C. *J. Chem. Phys.* **2008**, *129*.
- (21) Gordon, P. A. *J. Chem. Phys.* **2006**, *125*.
- (22) Ruckenstein, E.; Liu, H. *Society* **1997**, 3927–3936.
- (23) Galliero, G.; Boned, C.; Baylaucq, A.; Montel, F. *Chem. Phys.* **2007**, *333*, 219–228.
- (24) Wu, G.-W.; Sadus, R. J. *Fluid Phase Equilib.* **2000**, *170*, 269–284.
- (25) Errington, J. R.; Panagiotopoulos, A. Z. *J. Chem. Phys.* **1998**, *109*, 1093–1100.
- (26) McGrath, M. J.; Ghogomu, J. N.; Tsona, N. T.; Siepmann, J. I.; Chen, B.; Napari, I.; Vehkamaki, H. *J. Chem. Phys.* **2010**, *133*.
- (27) Parker, T. M.; Sherrill, C. D. *J. Chem. Theory Comput.* **2015**, *11*, 4197–4204.
- (28) Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. *J. Comput. Chem.* **2009**, *30*, 2187–2193.
- (29) Zgarbová, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476–10493.
- (30) Bastea, S. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* **2003**, *68*, 031204.
- (31) Errington, J. R.; Panagiotopoulos, A. Z. *J. Phys. Chem. B* **1999**, *103*, 6314–6322.
- (32) Ross, M.; Ree, F. H. *J. Chem. Phys.* **1980**, *73*, 6146.
- (33) Schmidt, J. R.; Yu, K.; McDaniel, J. G. *Acc. Chem. Res.* **2015**, *48*, 548–556.
- (34) Abrahamson, A. A. *Phys. Rev.* **1963**, *130*, 693–707.
- (35) Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
- (36) Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.

- (37) Kim, Y. S.; Kim, S. K.; Lee, W. D. *Chem. Phys. Lett.* **1981**, *80*, 574–575.
- (38) Misquitta, A. J.; Stone, A. J. Ab initio atom-atom potentials using CamCASP: Theory. 2015; <https://arxiv.org/abs/1512.06150v2>.
- (39) Nyeland, C.; Toennies, J. P. *Chem. Phys. Lett.* **1986**, *127*, 3–8.
- (40) Ihm, G.; Cole, M. W.; Toigo, F.; Klein, J. R. *Phys. Rev. A* **1990**, *42*, 5244–5252.
- (41) Duke, R. E.; Starovoytov, O. N.; Piquemal, J.-P.; Cisneros, G. A. *J. Chem. Theory Comput.* **2014**, *10*, 1361–1365.
- (42) Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G. *J. Chem. Theory Comput.* **2010**, *6*, 190–202.
- (43) Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. *J. Chem. Phys.* **2006**, *125*.
- (44) Chaudret, R.; Gresh, N.; Narth, C.; Lagardere, L.; Darden, T. A.; Cisneros, G. A.; Piquemal, J. P. *J. Phys. Chem. A* **2014**, *118*, 7598–7612.
- (45) Chaudret, R.; Gresh, N.; Cisneros, G. A.; Scemama, A.; Piquemal, J.-p. *Can. J. Chem.* **2013**, *91*, 804–810.
- (46) Öhrn, A.; Hermida-Ramon, J. M.; Karlström, G. *J. Chem. Theory Comput.* **2016**, *12*, 2298–2311.
- (47) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.
- (48) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293–307.
- (49) Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890–1900.
- (50) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459–467.
- (51) Patil, S. H.; Tang, K. T. In *Asymptotic methods in quantum mechanics*; Scäfer, F. P., Toennies, J. P., Zinth, W., Eds.; Springer, 2000.
- (52) Hoffmann-Ostenhof, M.; Hoffmann-Ostenhof, T. *Phys. Rev. A* **1977**, *16*, 1782–1785.
- (53) Amovilli, C.; March, N. H. *J. Phys. A. Math. Gen.* **2006**, *39*, 7349–7357.
- (54) Bunge, A. V.; Esquivel, R. O. *Phys. Rev. A* **1986**, *34*, 853.
- (55) Tai, H. *Phys. Rev. A* **1986**, *33*, 3657–3666.
- (56) Rosen, N. *Phys. Rev. Lett.* **1931**, *38*, 255–276.

- (57) Rappe, A. K.; Casewit, C.; Colwell, K.; Goddard, W. A. I.; Skiff, W. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
- (58) Waldman, M.; Hagler, A. T. *J. Comput. Chem.* **1993**, *14*, 1077–1084.
- (59) McDaniel, J. G.; Schmidt, J. R. *J. Phys. Chem. C* **2012**, *116*, 14031–14039.
- (60) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345–10354.
- (61) Bukowski, R.; Szalewicz, K.; Groenenboom, G.; van der Avoird, A. *J. Chem. Phys.* **2006**, *125*, 044301.
- (62) Jeziorska, M.; Cencek, W.; Patkowski, K.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2007**, *127*, 124303.
- (63) Sum, A. K.; Sandler, S. I.; Bukowski, R.; Szalewicz, K. *J. Chem. Phys.* **2002**, *116*, 7637.
- (64) Konieczny, J. K.; Sokalski, W. A. *J. Mol. Model.* **2015**, *21*, 197.
- (65) Kita, S.; Noda, K.; Inouye, H. *J. Chem. Phys.* **1976**, *64*, 3446–3449.
- (66) Kuechler, E. R.; Giese, T. J.; York, D. M. *J. Chem. Phys.* **2015**, *143*.
- (67) Giese, T. J.; York, D. M. *J. Chem. Phys.* **2007**, *127*.
- (68) Giese, T. J.; Chen, H.; Dissanayake, T.; Giambau, G. M.; Heldenbrand, H.; Huang, M.; Kuechler, E. R.; Lee, T. S.; Panteva, M. T.; Radak, B. K.; York, D. M. *J. Chem. Theory Comput.* **2013**, *9*, 1417–1427.
- (69) Day, G. M.; Price, S. L. *J. Am. Chem. Soc.* **2003**, *125*, 16434–16443.
- (70) Nobeli, I.; Price, S. L.; Wheatley, R. *J. Mol. Phys.* **1998**, *95*, 525–537.
- (71) McDaniel, J. G.; Schmidt, J. R. *J. Phys. Chem. A* **2013**, *117*, 2053–2066.
- (72) Totton, T. S.; Misquitta, A. J.; Kraft, M. *J. Chem. Theory Comput.* **2010**, *6*, 683–695.
- (73) Misquitta, A. J. *J. Chem. Theory Comput.* **2013**, *9*, 5313–5326.
- (74) Misquitta, A. J.; Stone, A. J. Ab initio atom-atom potentials using CamCASP: Application to Pyridine. 2015; <https://arxiv.org/abs/1512.06155v2>.
- (75) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726–3741.
- (76) Tang, K. T.; Toennies, J. P. *Surf. Sci.* **1992**, *279*, L203–L206.
- (77) Wheatley, R. J.; Price, S. L. *Mol. Phys.* **1990**, *69*, 507–533.

- (78) Mitchell, J. B. O.; Price, S. L. *J. Phys. Chem. A* **2000**, *104*, 10958–10971.
- (79) Söderhjelm, P.; Karlström, G.; Ryde, U. *J. Chem. Phys.* **2006**, *124*, 244101.
- (80) Tkatchenko, A.; Distasio, R. A.; Car, R.; Scheffler, M. *Phys. Rev. Lett.* **2012**, *108*, 1–5.
- (81) Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 6–9.
- (82) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2016**, *12*, 2312–2323.
- (83) Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2010**, *6*, 2455–2468.
- (84) Manz, T. A.; Sholl, D. S. *J. Chem. Theory Comput.* **2012**, *8*, 2844–2867.
- (85) Yu, K.; McDaniel, J. G.; Schmidt, J. R. *J. Phys. Chem. B* **2011**, *115*, 10054–10063.
- (86) Levy, M.; Perdew, J. P.; Sahni, V. *Phys. Rev. A* **1984**, *30*, 2745–2748.
- (87) Kitaigorodsky, A. *Molecular crystals and Molecules*; Physical Chemistry; Elsevier Science: New York, 2012.
- (88) Lillestolen, T. C.; Wheatley, R. J. *Chem. Commun.* **2008**, *7345*, 5909–5911.
- (89) Lillestolen, T. C.; Wheatley, R. J. *J. Chem. Phys.* **2009**, *131*, 144101.
- (90) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129–138.
- (91) Misquitta, A. J.; Szalewicz, K. *Chem. Phys. Lett.* **2002**, *357*, 301–306.
- (92) Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. *Phys. Rev. Lett.* **2003**, *91*, 033201.
- (93) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*.
- (94) Heßelmann, A.; Jansen, G.; Schutz, M. *J. Chem. Phys.* **2005**, *122*, 014103.
- (95) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Chem. Theory Comput.* **2006**, *2*, 400–412.
- (96) Heßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *362*, 319–325.
- (97) Heßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778–784.
- (98) Heßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464–470.
- (99) Jansen, G.; Hesselmann, A.; Williams, H. L.; Chabalowski, C. F. *J. Phys. Chem. A* **2001**, *105*, 11156–11158.

- (100) Podeszwa, R.; Szalewicz, K. Accurate interaction energies from perturbation theory based on Kohn-Sham model. 2005; <http://arxiv.org/abs/physics/0501023>.
- (101) Jeziorska, M.; Jeziorski, B.; Čížek, J. *Int. J. Quantum Chem.* **1987**, *32*, 149–164.
- (102) Drude, P.; Riborg, C.; Millikan, R. A. *The Theory of Optics... Translated from the German by CR Mann and RA Millikan*; London; New York [printed], 1902.
- (103) Lamoureux, G.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 3025.
- (104) NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101. 2015; <http://cccbdb.nist.gov/>.
- (105) Shoemaker, K. *Graph. Gems 3*; 1992; Chapter 6, pp 124–132.
- (106) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *WIREs Comput Mol Sci* **2012**, *2*, 242–253.
- (107) Misquitta, A. J.; Stone, A. J. CamCASP: a program for studying intermolecular interactions and for the calculation of molecular properties in distributed form, version 5.8. University of Cambridge, 2015.
- (108) Aidan, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansík, B.; Jensen, H. J. A.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Østdal, A.; Packer, M. J.; Pawłowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V. V.; Salek, P.; Samson, C. C. M.; de Merás, A. S.; Saué, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 269–284.
- (109) Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A. W.; Nutt, D. R.; Popelier, P. L. A.; Wales, D. J. ORIENT: a program for studying interactions between molecules, version 4.8. 2015; <http://www-stone.ch.cam.ac.uk/programs/orient.html>.
- (110) Ferenczy, G. G.; Winn, P. J.; Reynolds, C. a. *J. Phys. Chem. A* **1997**, *101*, 5446–5455.

- (111) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L. P.; Shukla, D.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S. *J. Chem. Theory Comput.* **2013**, *9*, 461–469.
- (112) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (113) Sebetci, A.; Beran, G. J. O. *J. Chem. Theory Comput.* **2010**, *6*, 155–167.
- (114) Misquitta, A.; Welch, G.; Stone, A.; Price, S. **2008**, *456*, 105–109.
- (115) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478–8490.
- (116) Totton, T. S.; Misquitta, A. J.; Kraft, M. *J. Chem. Theory Comput.* **2010**, *6*, 683–695.
- (117) Hermida-Ramón, J. M.; Ríos, M. A. *Chem. Phys.* **2000**, *262*, 423–436.
- (118) Nyeland, C. *Chem. Phys.* **1990**, *147*, 229–240.
- (119) Dymond, J. H.; Smith, E. B. *The Virial Coefficients of Pure Gases and Mixtures*, 2nd ed.; Clarendon Press: Berlin Heidelberg, 1980.
- (120) Vogel, E.; Jäger, B.; Hellmann, R.; Bich, E. *Mol. Phys.* **2010**, *108*, 3335–3352.
- (121) McDaniel, J. G.; Schmidt, J. R. *J. Phys. Chem. B* **2014**, *118*, 8042–8053.
- (122) McDaniel, J. G.; Yu, K.; Schmidt, J. R. *J. Phys. Chem. C* **2012**, *116*, 1892–1903.
- (123) Witt, R. K.; Kemp, J. D. *J. Am. Chem. Soc.* **1937**, *59*, 273–276.
- (124) Riddick, J. A.; Bunger, W. B.; Sakano, T. K. *Techniques of Chemistry, Vol. II: Organic Solvents: Physical Properties and Methods of Purification*, 4th ed.; Wiley-Interscience: New York, 1986.
- (125) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. I. *J. Phys. Chem.* **1990**, *101*, 8897–8909.
- (126) Lim, T. *Zeitschrift für Naturforschung-A* **2009**, *64*, 200–204.
- (127) Van Duin, a. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. a. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.
- (128) Eramian, H.; Tian, Y.-H.; Fox, Z.; Beneberu, H. Z.; Kertesz, M. *J. Phys. Chem. A* **2013**, *117*, 14184–14190.
- (129) Badenhoop, J. K.; Weinhold, F. *J. Chem. Phys.* **1997**, *107*, 5422.
- (130) Kim, H.; Doan, V. D.; Cho, W. J.; Madhav, M. V.; Kim, K. S. *Sci. Rep.* **2014**, *4*, 1–8.
- (131) Wheatley, R. J.; Gopal, A. A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2087–2091.

## 6 TOC Graphic

