# New Angles on Standard Force Fields: Toward a General Approach for Treating Atomic-Level Anisotropy

Mary J. Van Vleet,[†] Alston J. Misquitta,[‡] and J. R. Schmidt*,[†]

[†]Theoretical Chemistry Institute and Department of Chemistry, University of Wisconsin−Madison, Madison, Wisconsin 53706, United States

[‡]Department of Physics and Astronomy, Queen Mary University of London, London E1 4NS, United Kingdom

**S** *Supporting Information*
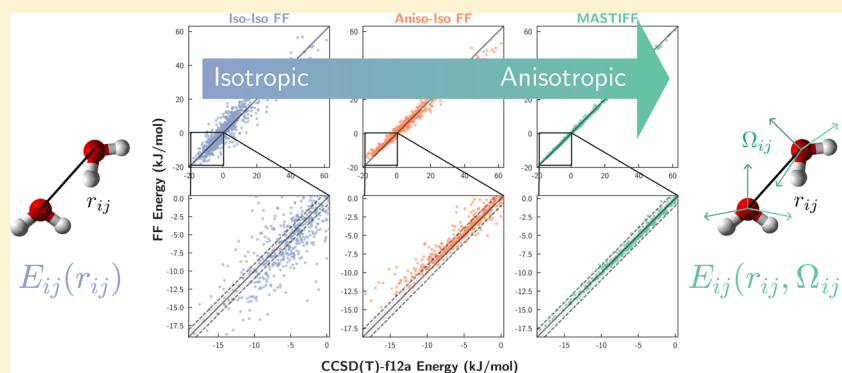
**ABSTRACT:** Nearly all standard force fields employ the "sum-of-spheres" approximation, which models intermolecular interactions purely in terms of interatomic distances. Nonetheless, atoms in molecules can have significantly nonspherical shapes, leading to interatomic interaction energies with strong orientation dependencies. Neglecting this "atomic-level anisotropy" can lead to significant errors in predicting interaction energies. Herein, we propose a simple, transferable, and computationally efficient model (MASTIFF) whereby atomic-level orientation dependence can be incorporated into ab initio intermolecular force fields. MASTIFF includes anisotropic exchange-repulsion, charge penetration, and dispersion effects, in conjunction with a standard treatment of anisotropic long-range (multipolar) electrostatics. To validate our approach, we benchmark MASTIFF against various sum-of-spheres models over a large library of intermolecular interactions between small organic molecules. MASTIFF achieves quantitative accuracy, with respect to both high-level electronic structure theory and experiment, thus showing promise as a basis for "next-generation" force field development.

## 1. INTRODUCTION

Classical molecular simulation is a standard tool for interpreting and predicting the chemistry of an incredible host of systems ranging from simple liquids to complex materials and biomolecules. Such simulations always require, as input, a mathematical description of the system's potential energy surface (PES). In principle, the PES for most chemical systems can accurately be determined from one of several high-level electronic structure methods;[1−3] nevertheless, these calculations are currently too expensive to use in simulations of large systems and/or long time scales.[4] Consequently, most routine molecular simulations are performed with the aid of force fields: computationally inexpensive, parametrized mathematical expressions that approximate the exact PES. Because the accuracy and predictive capabilities of molecular simulations are directly tied to the underlying force field, a central challenge is the development of highly accurate force fields. In contrast to the development of empirical force fields, where the typical emphasis is on generation of *effective* potentials yielding bulk

properties, for ab initio force fields, this accuracy is principally defined by a force field's fidelity to the underlying exact PES.

As of now, several common shortcomings inhibit the accuracy and predictive capabilities of standard ab initio force fields, and these limitations must be systematically addressed in order to generate improved, "next-generation" force fields.[5] One important shortcoming, and the focus of this work, is the so-called "sum-of-spheres", or "isotropic atom−atom" approximation,[6] in which it is presumed that the nonbonding interactions between molecules can be treated as a superposition of interactions between *spherically symmetric* atoms. (Note that this sum-of-spheres approximation is distinct from the commonly used pairwise additive approximation employed in force fields lacking explicit polarization;[5] challenges associated with this latter approximation are reviewed elsewhere.[7−14]) The sum-of-spheres approximation thus assumes that the pair potential, $E_2^{ij}$, between two atoms-in-molecules $i$

and $j$ (which is formally dependent on both their interatomic distance ($r_{ij}$) and relative orientation ($\Omega_{ij}$) can be modeled as

$$E_2^{ij}(r_{ij}, \Omega_{ij}) \approx f(r_{ij}) \equiv V_2(r_{ij}) \qquad (1)$$

where $f(r_{ij})$ is an arbitrary, distance-dependent function that defines the pairwise interaction. Here and throughout, we use $E$ to denote an exact PES, and we use $V$ to denote the corresponding model/force field energy. With some exceptions (vida infra), nearly all standard intermolecular force fields—ranging from the popular "Lennard-Jones plus point charges" model to more-complex and/or polarizable functional forms[15]—explicitly make use of the isotropic atom–atom approximation.

Its popularity notwithstanding, there is good experimental and theoretical evidence to suggest that the sum-of-spheres approximation does not hold in practice.[6,16,17] Importantly, and as we argue in Section 5, force fields that account for anisotropic long-range (multipolar) electrostatics, but otherwise employ the sum-of-spheres approximation, are an improved but *still incomplete* model for describing the atomic-level anisotropy of intermolecular interactions. Experimentally, it has long been known that atom-in-molecule charge densities, as determined from X-ray diffraction, can exhibit significant nonspherical features, such as with lone pair or $\pi$-electron densities.[18] Furthermore, statistical analyses of the Cambridge Structural Database have shown that the van der Waals radii of atoms-in-molecules (as measured from interatomic closest contact distances) are not isotropically distributed, but, instead, show strong orientation dependencies, particularly for halogens and other heteroatoms.[19−24] These experimental studies are corroborated by a significant body of theoretical research on both the anisotropy of the atomic van der Waals radii, as well as the nonspherical features of the atomic charge densities themselves,[23,25−29] overall suggesting that the sum-of-spheres approximation is an insufficiently flexible model for the subset of intermolecular interactions that arise from atomically nonspherical charge densities. The breakdown of the sum-of-spheres approximation may be particularly problematic for ab initio force field development, since any anisotropy cannot easily be accounted for in an average manner via empirical parametrization, and may help explain known difficulties in generating accurate atom–atom force fields for such important chemical interactions as hydrogen bonding,[30] $\pi$-interactions,[31−33] and $\sigma$-bonding[34−36] (see ref 37 and references therein).

Motivated by these observations, a small but important body of work has been devoted to addressing the limitations of the isotropic atom–atom model in the context of "next-generation" force field development. As will be discussed in detail below (see Section 2), the general conclusion from these studies is that many components of intermolecular interactions (specifically electrostatics, exchange-repulsion, induction, and dispersion) can be more accurately modeled by functional forms that go beyond the sum-of-spheres approximation.[17,38,39] While few intermolecular potentials (and virtually no standard force fields amenable to routine molecular simulation) explicitly account for atomic-level anisotropy for all aspects of intermolecular interactions, several recent standard force fields have incorporated atomic-level anisotropy into their description of long-range electrostatics.[37] Some of these potentials (notably AMOEBA[39−41] and some water potentials[30,37]) are already employed in large-scale molecular simulation, often with very

encouraging success.[37] Furthermore, others have shown that anisotropic potentials (some of which additionally model the anisotropy of exchange-repulsion and/or dispersion) lead to significant improvements in predicting molecular crystal structures.[13,37,42−46] These and other results strongly suggest that a complete incorporation of atomic anisotropy will lead to increasingly accurate and predictive molecular simulations in a wider variety of chemical interactions.[38]

Given the importance of atomic-level anisotropy in defining intermolecular interactions, and the critical role that computationally affordable standard force fields play in enabling molecular simulation, our present goal is to develop a general methodology for standard force field development that (i) can comprehensively account for atomic-level anisotropy in all components of intermolecular interactions *and* (ii) can be routinely employed in large-scale molecular simulation. Furthermore, our aim is to develop a first-principles-based model that is as accurate and transferable as possible, all while maintaining a simple, computationally tractable functional form that allows for robust parametrization and avoids overfitting/underfitting. Thus, building on prior work (both from our own group[12,15,47,48] and from other research groups[17]), here we present a general ansatz for anisotropic force field development that, at minimal computational overhead, and only where necessary, incorporates atomic-level anisotropy into all aspects of intermolecular interactions (electrostatics, exchange, induction, and dispersion), not only in the asymptotic limit of large intermolecular separations, but also in the region of non-negligible electron density overlap. After motivating and establishing the functional forms used in our anisotropic force fields, we next demonstrate, using a large library of dimer interactions between organic molecules, the accuracy and transferability of these new force fields, with respect to the reproduction of high-quality ab initio potential energy surfaces. Lastly, and using $CO_2$ as a case study, we offer an example as to how these new, "atomically anisotropic" models for dimer interactions can be used to enable highly accurate simulations of bulk properties. The theory and results presented in this manuscript should be of general utility in improving the accuracy of (specifically ab initio generated) force fields, including those amenable to large-scale molecular dynamics simulations.

## 2. BACKGROUND

Before presenting our development methodology for atomically anisotropic potentials, we provide an overview of prior approaches that go beyond the sum-of-spheres approximation. Throughout this discussion, we employ the fairly standard[49] decomposition of interaction energies into physically meaningful components of electrostatics, exchange-repulsion, induction (which includes both polarization and charge-transfer), and dispersion. Many studies on atomically anisotropic force field development have focused on incorporating anisotropy on a component-by-component basis, and, for the sake of clarity, we discuss anisotropy for each energy component individually. As in prior work,[47] we find it useful to separate the so-called "long-range"/asymptotic effects (multipolar electrostatics, polarization, and dispersion) from those "short-range" effects that arise only at smaller intermolecular separations due to the non-negligible overlap of monomer electron densities (e.g., charge penetration and exchange-repulsion).

**2.1. Prior Models for Long-Range Interactions.** The importance of atomic-level anisotropy in long-range inter-

actions, particularly as it pertains to electrostatics, is quite well-known. Several groups have found that using atomic multipoles (rather than simple point charges) greatly improves both the electrostatic potential[26,50] and the resulting electrostatic interaction energies.[10,37,39,41,51−54] Although not without additional computational cost, atomic multipoles are now routinely employed in many popular force fields.[30,39,41] As an alternate and often more computationally affordable approach, others have used off-atom point charges to effectively account for anisotropic charge densities.[35,55−57] Consistent with chemical intuition, improvements from the use of atomic multipoles/off-site charges are typically most significant when describing the electric fields generated by heteroatoms and carbons in multiple bonding environments.[58,59]

The induction and dispersion energies have also been shown to exhibit anisotropies that go beyond the sum-of-spheres model. For instance, it has been suggested that anisotropic polarizabilities (which affect both polarization and dispersion) are required to avoid an artificial overstabilization of base stacking energies in biomolecules.[32] In order to more accurately treat polarization, several molecular mechanics potentials have made use of either off-site[60] or explicitly anisotropic polarizabilities.[56,61] Similarly, the importance of anisotropic dispersion interactions has also been established,[12,62−65] particularly for $\pi$-stacking interactions,[5,32] and select potentials have incorporated directional dependence into the functional form for dispersion by expanding the dispersion coefficients in terms of $\overline{S}$-functions (see Appendix A) or, more approximately, spherical harmonics.[45,62,64,66,67]

**2.2. Prior Models for Short-Range Interactions.** At closer intermolecular separations, where overlapping electron densities between monomers leads to exchange-repulsion and charge-penetration effects, anisotropy can also be important. Exchange-repulsion has known orientation dependencies which can play a quantitative role in halogen bonding[34,68] and other chemical interactions, and many authors have developed models for describing the anisotropy of exchange-repulsion. Some potentials (albeit not those amenable to large-scale molecular simulation) employ numerically computed overlap integrals in conjunction with the density-overlap model popularized by Wheatley and Price[69−73] to quantify anisotropic exchange-repulsion, charge transfer, and/or charge penetration interactions.[51,53,54,74−76] Taking a more analytical approach, many other potentials have extended the Born−Mayer functional form[77] to allow for orientation-dependent prefactors,[6,12,13,17,43,45,76,78,79] and model short-range effects using an anisotropic functional form originally proposed by Stone and Price:[6]

$$V_{ij}^{\text{exch}} = G \exp[-\alpha_{ij}(R_{ij} - \rho_{ij}(\Omega_{ij}))] \tag{2}$$

Here, $G$ is not a parameter, but rather an energy unit,[16] $\Omega_{ij}$ describes a relative orientation, and $\alpha$ and $\rho$ represent, respectively, the hardness and shape of the pair potential. In principle, one might also allow $\alpha$ to have orientation dependence; however, this seems unnecessary in practice.[16] Similar to treatments of anisotropic electrostatics, the orientation dependence of $\rho_{ij}$ is typically expressed in terms of spherical harmonics and/or $\overline{S}$-functions.[16]

Finally, we note that, aside from exchange-repulsion, we are aware of relatively little research on the development of simple analytical expressions for the anisotropy of other overlap effects,

such as electrostatic/inductive charge penetration, charge-transfer, or short-range dispersion.

## 3. THEORY AND MOTIVATION

Building on this prior work, we now outline a methodology whereby atomic-level anisotropy can be incorporated into standard force fields amenable to large-scale molecular simulation. In particular, we present a general methodology that optimally incorporates atomically anisotropic effects, subject to the following goals:

(1) **Chemical accuracy with respect to ab initio benchmarks:** For systems that can be directly parametrized against high-quality ab initio PES, the force field should exhibit chemical accuracy (average errors smaller than 1 kJ mol$^{-1}$) with respect to the ab initio benchmark; furthermore, any errors in the force field should be random rather than systematic.

(2) **Transferability across chemical environments:** Given force fields for two different pure systems, we should be able to accurately calculate (via simple combination rules and without additional parametrization) the PES of any system that is a mixture of the pure systems.

(3) **Simplicity:** The force field should be restricted to functional forms that are already compatible with, or could be easily implemented in, existing molecular simulation packages.

(4) **Computational tractability:** The force field should impose minimal additional computational expense relative to existing polarizable multipolar force fields.[41]

Given these goals, we now outline a detailed methodology for incorporating atomic-level anisotropy into each component (electrostatic, exchange-repulsion, induction, and dispersion) of intermolecular interactions.

**3.1. Anisotropic Models for Short-Range Interactions.**
*3.1.1. Exchange-Repulsion.* We begin by considering the exchange-repulsion ($E_{ij}^{\text{exch}}$) that arises from the overlap of electron densities from two nonspherical atoms-in-molecules, $i$ and $j$. Here and throughout, we closely follow the notation and theory used by Stone.[16] Without any loss of generality, we can express the exchange-repulsion between these two atoms as a function of their interatomic distance ($r_{ij}$) and relative orientation ($\Omega_{ij}$). Furthermore, we can describe this relative orientation by assigning local coordinate axes to each $i$ and $j$, such that the exchange energy is given by

$$E_{ij}^{\text{exch}}(r_{ij}, \Omega_{ij}) \equiv E_{ij}^{\text{exch}}(r_{ij}, \theta_i, \phi_i, \theta_j, \phi_j) \tag{3}$$

where $\theta_i$ and $\phi_i$ are the polar coordinates, expressed in the local coordinate system of atom $i$, that describe the position of atom $j$. Correspondingly, $\theta_j$ and $\phi_j$ define the position of $i$ in terms of the local coordinate system of $j$. In principle, the choice of these local coordinate frames is arbitrary. However, for the models introduced below, parametrization can be dramatically simplified by exploiting the local symmetry of an atom in its molecular environment and aligning the local coordinate frame with the principal axis of this local symmetry.[16] Some examples of these local axes are shown in Figure 1.

Next, we make an ansatz that eq 3 is separable into radial- and angular-dependent contributions,
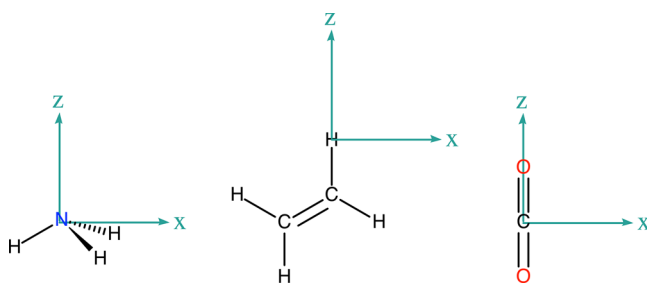
**Figure 1.** Local axis system, shown for select atoms in molecules.

$$E_{ij}^{exch}(r_{ij}, \theta_i, \varphi_i, \theta_j, \varphi_j) \approx V_{ij}^{exch}(r_{ij}, \theta_i, \varphi_i, \theta_j, \varphi_j)$$

$$= f(r_{ij})g(\theta_i, \varphi_i, \theta_j, \varphi_j) \quad (4)$$

thus subdividing the problem of finding a general functional form for $E_{ij}^{exch}$ into two more tractable tasks. First, we must find an ideal sum-of-spheres model to describe the radial (isotropic) dependence of the force field, and second, we must find a way to model the orientation dependence as a multiplicative prefactor to $f(r_{ij})$.

Given that the only requirement for $f(r_{ij})$ is that it be isotropic, how should a suitable model for $f(r_{ij})$ be chosen? Indeed, all standard isotropic force fields are of this general form, and thus might serve as a suitable starting point for anisotropic force field development. For reasons discussed below, in this work, we employ a simple and accurate model[47] for $f(r_{ij})$ that can be derived from first-principles. In particular, we employ the overlap model[12,43,69,71−73,80,81] to approximate $E_{ij}^{exch}$ as proportional to the overlap between spherically symmetric atom-in-molecule (AIM) electron densities, each with density

$$\rho_i(r) = D_i \exp^{-B_i r} \quad (5)$$

where $D_i$ and $B_i$ are both atom type-specific constants that can be parametrized from molecular electron densities and that represent, respectively, the shape and hardness of the AIM density. Using this approximation to the overlap model, the exchange energy between two atoms is then modeled by

$$E_{ij}^{exch} \approx V_{ij}^{exch} \propto S_\rho^{ij}$$

$$\approx A_{ij}^{exch} \left[ \frac{(B_{ij}r_{ij})^2}{3} + B_{ij}r_{ij} + 1 \right] \exp(-B_{ij}r_{ij}) \quad (6)$$

with combining rules

$$A_{ij}^{exch} \equiv A_i^{exch} A_j^{exch}$$

$$B_{ij} \equiv \sqrt{B_i B_j} \quad (7)$$

and where $S_\rho^{ij}$ is the electron density overlap between atoms and $A_{ij}$ is a fitted proportionality constant.

Here, and throughout, we use eq 6, referenced as the Slater-ISA formalism,[47] as our model for $f(r_{ij})$. This choice is primarily justified by the accuracy of the Slater-ISA formalism, compared to other sum-of-spheres models for repulsion.[47] Furthermore, and especially for simple test cases where one might expect the sum-of-spheres approximation to hold (e.g., argon, methane, or ethane), we have shown that the Slater-ISA FF correctly models intermolecular potential energy surfaces for a sizable library of

intermolecular interactions over the asymptotic, attractive, and repulsive regions of the PES.[47]

There is also solid theoretical motivation to utilize Slater-ISA as a model for $f(r_{ij})$. Specifically, the AIM densities used to parametrize Slater-ISA FF are partitioned using an iterated stockholder atoms (ISA) procedure, and the resulting density profiles are guaranteed to be maximally spherical.[48,82,83] This condition of "maximum sphericity" has two consequences. First, it suggests that Slater-ISA FF should be an optimal, or nearly optimal, isotropic atom−atom model. In other words, the resulting model for $f(r_{ij})$ should completely account for the radial dependence of the potential, and, consequently, $g(\theta_i, \phi_i, \theta_j, \phi_j)$ will truly represent the orientation dependence, rather than simply overfitting residual errors from the radial functional form, in turn, retaining high transferability. Second, and relatedly, having maximally spherical ISA densities suggests that anisotropic effects should be a minimal perturbation to the PES. This means that, to a first-order approximation, $g(\theta_i, \phi_i, \theta_j, \phi_j)$ is simply equal to 1. Furthermore, the nonspherical components of the ISA densities should provide us with guidance as to which atom types might require anisotropic treatment.

With the functional form for $f(r_{ij})$ determined, we now describe our model for $g(\theta_i, \phi_i, \theta_j, \phi_j)$. As motivated in Appendix A, and under the ansatz of radial and angular separability, an approximate, transferable, and orientation-dependent expression for $A_i^{exch}$ can be obtained by expanding $A_i^{exch}$ in a basis of renormalized spherical harmonics,

$$C_{lm}(\theta, \phi) = \sqrt{\frac{4\pi}{2l+1}} Y_{lm}(\theta, \phi) \quad (8)$$

thus yielding

$$A_i^{exch}(\theta_i, \phi_i) = A_{i,iso}^{exch}(1 + \xi_i^{exch}(\theta_i, \phi_i))$$

$$\xi_i^{exch}(\theta_i, \phi_i) \equiv \sum_{l>0,k} a_{i,lk}^{exch} C_{lk}(\theta_i, \phi_i) \quad (9)$$

for $A_i^{exch}$ and, subsequently,

$$V_{ij}^{exch} = A_{ij}^{exch}(\Omega_{ij}) \left[ \frac{(B_{ij}r_{ij})^2}{3} + B_{ij}r_{ij} + 1 \right] \exp(-B_{ij}r_{ij}) \quad (10)$$

with

$$A_{ij}^{exch}(\Omega_{ij}) = A_i^{exch}(\theta_i, \phi_i) A_j^{exch}(\theta_j, \phi_j) \quad (11)$$

for the exchange-repulsion potential. Note that, with the exception of the now orientation-dependent $A_i^{exch}$, the atomically anisotropic model in eq 10 is identical to our previously defined isotropic model (eq 6).

The $a_{i,lk}^{exch}$ are free parameters that must be fit to ab initio data. Still, we and other researchers have found the expansion in eq 9 to be very quickly convergent,[6,12,13,17,43,45,76,78,79] especially given a proper choice of local coordinate system that eliminates many expansion terms via symmetry. In practice, only symmetry-allowed terms up to $l = 2$ seem to be required for heteroatoms, carbons in multiple bonding environments, and select hydrogens (see equations in Section 5). Most other atom types require no anisotropic parameters whatsoever, and isotropic atom types can be easily modeled within this formalism simply by setting $\xi_i(\theta_i, \phi_i) = 0$.

*3.1.2. Other Short-Range Effects.* As in prior work,[47] we have found that other short-range effects, including charge

penetration and short-range induction, can be modeled as being proportional to exchange-repulsion. We take the same approach here, and the functional form for these two short-range effects is given by eq 10, with "exch" superscripts replaced by the appropriate short-range energy term (see Section 4).

At shorter intermolecular separations, we must also damp some of the functional forms developed for long-range interactions (vida infra), to account for charge penetration effects and avoid unphysical divergences. For induction, we use the same isotropic damping function as in the AMOEBA force field.[41] To model the dispersion energies at short-range, we damp each of the individual $C_n$ dispersion coefficients (see Section 3.2.3 and eq 18) with the Tang–Toennies[84,85] damping function,

$$f_n(x) = 1 - e^{-x} \sum_{k=0}^{n} \frac{(x)^k}{k!} \tag{12}$$

where

$$x = -\frac{d}{dr_{ij}}[\ln V_{ij}^{exch}]r_{ij}$$

By substituting our expression for $V_{ij}^{exch}$ from eq 10, we obtain

$$x = B_{ij}r_{ij} - \left(\frac{2B_{ij}^2 r_{ij} + 3B_{ij}}{B_{ij}^2 r_{ij}^2 + 3B_{ij}r_{ij} + 3}\right)r_{ij} \tag{13}$$

for our anisotropic model, which is identical to the expression derived for isotropic systems in prior work.[47]

### 3.2. Anisotropic Models for Long-Range Interactions.
*3.2.1. Electrostatics.* In the present work, we describe the asymptotic electrostatics via a distributed multipole expansion,[12,16]

$$V_{ij}^{multipole} = \sum_{tu} Q_t^i T_{tu}^{ij} Q_u^j \tag{14}$$

with multipolar interaction tensor $T$ and parametrized moments $Q$ for all multipole moments $tu$ up to rank 2. However, for increased computational efficiency, off-site point charge models could also be utilized.[37]

*3.2.2. Induction.* Just as with electrostatics, long-range induction should properly be described by a distributed multipole expansion of interacting atomic polarizabilities.[12,45] Indeed, it has been shown that inclusion of higher-order and/or anisotropic polarizabilities greatly reduces errors in the two-body induction potential, relative to commonly used isotropic dipole polarizability models.[15,41,86−88] Because the model for the two-body induction also determines the many-body polarization energy, the proper treatment of induced multipoles becomes especially important in condensed-phase simulation.[15,16,41]

Because of the increased computational cost of these higher-order and anisotropic polarizability models, and because such functional forms are not (yet) implemented in OpenMM (the molecular simulation package used in this work), we currently neglect both higher-order and anisotropic contributions to the long-range induction. As we shall show, however, errors in the induction potential limit the overall accuracy of our force fields for extremely polar molecules (notably water), and future work will likely require improved models for long-range induction.

*3.2.3. Dispersion.* Past research[16] has motivated an anisotropic atom–atom model for dispersion of the form

$$V_{ij}^{disp} = -\sum_{n=6} \frac{C_{ij,n}(\Omega_{ij})}{r_{ij}^n} \tag{15}$$

Note that, in this equation, both odd and even powers of $r$ are allowed in the dispersion expansion, where all coefficients associated with odd powers are nonzero only for anisotropic charge distributions. In order to make this model both computationally efficient and maximally compatible with our previous isotropic model for dispersion, we choose (as an ansatz) to model the dispersion anisotropy as an orientation-dependent prefactor that affects all isotropic $C_6−C_{12}$ dispersion coefficients equally:

$$V_{ij}^{disp} = -A_i^{disp}A_j^{disp}\sum_{n=3}^{6} \frac{C_{ij,2n}}{r_{ij}^{2n}} \tag{16}$$

with

$$A_i^{disp} = 1 + \xi_i^{disp}(\theta_i, \phi_i) \tag{17}$$

and $\xi_i^{disp}(\theta_i, \phi_i)$ as in eq 9. Once again, eq 16 reduces to the isotropic case by setting $\xi_i^{disp}(\theta_i, \phi_i) = 0$. We note that, although the functional form in eq 16 bears many similarities to eq 15, (unphysically) no odd powers of $r$ occur in our proposed model for dispersion. Furthermore, the model utilizes the same anisotropic expansion for each dispersion coefficient. Nonetheless, in Section 5, we will show that this simple model yields significant accuracy gains in the dispersion energy with only minimal additional parametrization and expense.

## 4. TECHNICAL DETAILS

**4.1. The 91 Dimer Test Set.** Our benchmarking procedures are the same as in prior work,[47] and we briefly summarize the relevant technical details. A full discussion of results and example calculations are presented in Section 5.

We have previously developed a large library of benchmark interaction energies involving the following 13 atomic and small organic species: acetone, argon, ammonia, carbon dioxide, chloromethane, dimethyl ether, ethane, ethanol, ethene, methane, methanol, methyl amine, and water. (As in prior work, these molecules were chosen to be broadly representative of various functional groups in organic chemistry; studies on larger and/or flexible molecules are outside of the scope of this work, but will be the subject of future work.) Using these 13 monomers, we have generated a library of dimer interaction energies for each of the 91 possible unique dimer combinations (13 homomonomeric, 78 heteromonomeric). For each of these dimer combinations, interaction energies were computed at a DFT-SAPT[89−97] level of theory for 1000 quasi-randomly chosen dimer configurations, representing 91 000 benchmark interaction energies in total. As described below, parameters for a given force field methodology are then fit on a component-by-component basis to reproduce the benchmark DFT-SAPT energies.

**4.2. Force Field Fitting.** *4.2.1. Functional Forms.* We will present three types of force field fitting methodologies in this work, termed Iso-Iso FF, Aniso-Iso FF, and Aniso-Aniso FF (also referred to as a **M**ultipolar, **A**nisotropic, **S**later-**T**ype **I**ntermolecular **F**orce **F**ield, MASTIFF). The nomenclature of each name refers to, first, the isotropic/anisotropic treatment of multipolar electrostatics and, second, the isotropic/anisotropic treatment of dispersion and short-range effects. For MASTIFF, dispersion and short-range anisotropies are only included on

heteroatoms, atoms in multiple bonding environments, and associated hydrogens (see the Supporting Information). Note that Aniso-Iso FF is virtually identical to the Slater-ISA FF model developed in our prior work, and that this *partial* treatment of anisotropy (via multipolar electrostatic terms) is very similar in spirit to the popular AMOEBA[39,41] methodology.

All force fields in this work use the following general functional form for two-body interactions:

$$V_{FF}^{2b} = \sum_{ij} V_{ij}^{exch} + V_{ij}^{elst} + V_{ij}^{ind} + V_{ij}^{\delta^{HF}} + V_{ij}^{disp} \tag{18}$$

where

$$V_{ij}^{exch} = A_{ij}^{exch} P(B_{ij}, r_{ij}) \exp(-B_{ij}r_{ij})$$

$$V_{ij}^{elst} = -A_{ij}^{elst} P(B_{ij}, r_{ij}) \exp(-B_{ij}r_{ij}) + \sum_{tu} Q_t^i T_{tu}^{ij} Q_u^j$$

$$V_{ij}^{ind} = -A_{ij}^{ind} P(B_{ij}, r_{ij}) \exp(-B_{ij}r_{ij}) + V_{pol}^{(2)}$$

$$V_{ij}^{\delta^{HF}} = -A_{ij}^{\delta^{HF}} P(B_{ij}, r_{ij}) \exp(-B_{ij}r_{ij}) + V_{pol}^{(3-\infty)}$$

$$V_{ij}^{disp} = -A_{ij}^{disp} \sum_{n=3}^{6} f_{2n}(x) \frac{C_{ij,2n}}{r_{ij}^{2n}}$$

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

$$A_{ij} = A_i A_j$$

$$B_{ij} = \sqrt{B_i B_j}$$

$$C_{ij,2n} = \sqrt{C_{i,2n}C_{j,2n}}$$

$$f_{2n}(x) = 1 - e^{-x} \sum_{k=0}^{2n} \frac{(x)^k}{k!}$$

$$x = B_{ij}r_{ij} - \left( \frac{2B_{ij}^2 r_{ij} + 3B_{ij}}{B_{ij}^2 r_{ij}^2 + 3B_{ij}r_{ij} + 3} \right) r_{ij} \tag{19}$$

$B_i$, $C_i$, and $Q_i$ coefficients are all parameters of the force field (see Section 4.2.3 for details), and $T$ is the multipolar interaction tensor given in Appendix A. For Iso-Iso FF (the completely isotropic model), the summation in $\sum_{tu} Q_t^i T_{tu}^{ij} Q_u^j$ is truncated to only include point charges, whereas Aniso-Iso FF and MASTIFF both use a multipole expansion up to quadrupoles.

Both Iso-Iso FF and Aniso-Iso FF treat each $A_i$ as a single fitting parameter, with the exception that $A_i^{disp} = 1$. By contrast, $A_i$ is modeled in our fully anisotropic model, MASTIFF, as an orientation-dependent function expressed as an expansion in terms of spherical harmonics,

$$A_i(\theta_i, \phi_i) = A_{i,iso}(1 + \xi_i(\theta_i, \phi_i))$$

$$\xi_i(\theta_i, \phi_i) \equiv \sum_{l>0,k} a_{i,lk} C_{lk}(\theta_i, \phi_i) \tag{20}$$

where $A_{i,iso}$ and $a_{i,lk}$ are fitted parameters. As with the previous two force fields, $A_{i,iso}^{disp} = 1$ for MASTIFF. For isotropic atom types in MASTIFF (listed in the Supporting Information), $\xi_i(\theta_i, \phi_i) = 0$, such that the functional form for *isotropic* atom types is identical between MASTIFF and Aniso-Iso FF, and only the

functional form for *anisotropic* atom types differ between force fields. Note, however, that the numerical values for $A_{i,iso}$ in MASTIFF can differ from that of the $A_i$ parameters used in the other models.

As in ref 98, and for the purposes of force field fitting, the polarization energy, $V_{pol} = V_{pol}^{(2)} + V_{pol}^{(3-\infty)}$, is calculated using using a Drude oscillator model. As a difference from prior work, here the Thole-damping function follows the same functional form as in the AMOEBA model,[39]

$$\rho = \frac{3a}{4\pi} \exp(-au^3) \tag{21}$$

where $a = 0.39$ is a damping parameter, and $u = r_{ij}/(\alpha_i\alpha_j)^{1/6}$ is an effective damping distance that is dependent on calculated atomic polarizabilities (vida infra) ($\alpha_i$). (The choice of damping function was selected for later compatibility with the OpenMM[99] software package; see Section 4.3 for details.) As described fully in ref 98, and for the purpose of logical consistency with the corresponding SAPT energies (see Section 4.2.2), during force field fitting, $V_{pol}$ is subdivided into second-order ($V_{pol}^{(2)}$) and higher-order $V_{pol}^{(3-\infty)}$ contributions, and each contribution to the Drude oscillator energy is then added to either $V^{ind}$ or $V^{\delta^{HF}}$, respectively.

**4.2.2. Benchmark Energies.** Because DFT-SAPT provides a physically meaningful energy decomposition into electrostatic, exchange-repulsion, induction, and dispersion terms, parameters for each term in eq 18 are directly fit to model the corresponding DFT-SAPT energy (see ref 47 and references therein for details on the DFT-SAPT terminology):

$$V^{exch} \approx E^{exch} \equiv E_{exch}^{(1)}$$

$$V^{elst} \approx E^{elst} \equiv E_{pol}^{(1)}$$

$$V^{ind} \approx E^{ind} \equiv E_{ind}^{(2)} + E_{ind-exch}^{(2)}$$

$$V^{\delta^{HF}} \approx E^{\delta^{HF}} \equiv \delta(HF)$$

$$V^{disp} \approx E^{disp} \equiv E_{disp}^{(2)} + E_{disp-exch}^{(2)} \tag{22}$$

Fitting parameters on a component-by-component basis helps ensure parameter transferability and minimizes reliance on error cancellation. Note that (i) no parameters are fit to reproduce the total energy and (ii) because the DFT-SAPT energy decomposition is only calculated to second-order, third- and higher-order terms (mostly consisting of higher-order induction) are estimated by $E^{\delta^{HF}}$.

**4.2.3. Parameters Calculated from Monomer Properties.** Of the parameters listed in eq 19, most do not need to be fit to the DFT-SAPT energies, but can instead be calculated directly on the basis of monomer electron densities. In particular, all multipolar coefficients $Q$, polarizabilities $\alpha_i$ (involved in the calculation of $V_{pol}$), dispersion coefficients $C$, and atom-in-molecule exponents $B^{ISA}$, are calculated in a manner almost identical to ref 47. Note that, for our atom-in-molecule exponents, we tested the effects of treating $B^{ISA}$ either as a hard constraint or a soft constraint in the final force field fit. While the general conclusions from this study are rather insensitive to this choice of constraint methodology, we have found that the overall force field quality is somewhat improved by relaxing the $B^{ISA}$ coefficients in the presence of a harmonic penalty function (technical details of which can be found in the Supporting Information of ref 47). The optimized $B$ coefficients in this

work are always within 5%−10% of the calculated $B^{ISA}$ coefficients, demonstrating the good accuracy of the $B^{ISA}$ calculations themselves.

*4.2.4. Parameters Fit to Dimer Properties.* In addition to the soft-constrained $B$ parameters, all other free parameters ($A$ and $a$ parameters from eqs 18 and 20 are fit to reproduce DFT-SAPT energies from the 91-dimer test set described above. For each dimer pair, 4 or 5 separate optimizations (for exchange, electrostatics, induction, $\delta^{HF}$, and, for MASTIFF, dispersion) were carried out to minimize a weighted least-squares error. with the weighting function given by a Fermi−Dirac functional form,

$$w_i = \frac{1}{\exp\left(\frac{-E_i}{5.0\,|E_{min}|}\right) + 1} \tag{23}$$

where $E_i$ is the reference energy and $E_{min}$ is an estimate of the global minimum well depth (see ref 47 for details).

*4.2.5. Local Axis Determination.* Identically to AMOEBA and other force fields that incorporate some degree of atomic-level anisotropy,[39,43,79] we use a *z*-then-*x* convention to describe the relative orientation of atomic species. By design, the *z*-axis is chosen to lie parallel to the principal symmetry axis (or approximate local symmetry axis) of an atom in its molecular environment, and the *xz*-plane is similarly chosen to correspond to a secondary symmetry axis or plane. Based on the assigned symmetry of the local reference frame, many terms in the spherical expansion of eq 9 can then be set to zero, minimizing the number of free parameters that need to be fit to a given atom type. Representative local reference frames are shown for a few atom types in Figure 1, and a complete listing of anisotropic atom types (along with their respective local reference frames and nonzero spherical harmonic expansion terms) are given in the Supporting Information.

*4.2.6. CCSD(T) Force Fields.* DFT-SAPT is known to systematically underestimate the interaction energies of hydrogen-bonding compounds, and can also exhibit small but important errors for dispersion-dominated compounds.[100] Consequently, for simulations involving $CO_2$, $CH_3Cl$, $NH_3$, and $H_2O$, we tested the effect of refitting our SAPT-based force fields to reproduce benchmark supermolecular, counterpoise-corrected CCSD(T)-F12a/aVTZ calculations on the respective dimers. All calculations were performed using the Molpro 2012 software.[101] As with the DFT-SAPT-based force fields, all fits were performed on a component-by-component basis to fit (aside from the dispersion, discussed below) the corresponding DFT-SAPT energies as calculated in prior work:[47]

$$V^{exch} \approx E^{exch} \equiv E^{(1)}_{exch}$$

$$V^{elst} \approx E^{elst} \equiv E^{(1)}_{pol}$$

$$V^{ind} \approx E^{ind} \equiv E^{(2)}_{ind} + E^{(2)}_{ind-exch}$$

$$V^{\delta^{HF}} \approx E^{\delta^{HF}} \equiv \delta(HF)$$

$$V^{disp} \approx E^{disp} \equiv E^{(2)}_{disp} + E^{(2)}_{disp-exch} + \delta(CC) \tag{24}$$

where $\delta(CC) \equiv E^{CCSD(T)\text{-}F12a}_{int} - E^{DFT\text{-}SAPT}_{int}$. In the case of dispersion, and so that the total benchmark energy corresponded to the total CCSD(T)-f12a/aVTZ interaction energy, the difference between coupled-cluster and SAPT energies was added to the SAPT dispersion energy. (This correction scheme was chosen to account for small differences

in electron correlation effects between coupled cluster and DFT-SAPT.) The dispersion model $V^{disp}$ was then parametrized to reproduce the modified $E^{disp}$ energy.

In fitting our CCSD(T)-f12a-based force fields, we somewhat relaxed the constraint that $A^{disp} = 1$ for all atom types, and instead let $0.7 \leq A^{disp} \leq 1.3$. This constraint relaxation led, in some cases, to modest improvements in the fitted potential.

*4.2.7. $CO_2$ Three-Body Potential.* For modeling bulk $CO_2$, we developed a three-body model to account for three-body dispersion effects. This three-body model is based on the three-body dispersion Axilrod−Teller−Muto (ATM)-type model developed by Oakley and Wheatley.[102] These authors fit the ATM term with the constraint that the total molecular $C_9$ coefficient be 1970 a.u. Based on our own calculations using a CCSD/AVTZ level of theory,[103] we have obtained an isotropic molecular $C_9$ coefficient of 2246 a.u.; consequently, a 1.13 universal scale factor was introduced to the Oakley potential, to obtain dispersion energies that are consistent with this new dispersion coefficient.

**4.3. Simulation Protocols.** *4.3.1. Polarization Models for Simulations.* Although we have used a Drude oscillator model in the past and during force field development, at present, Drude oscillators in the OpenMM[99] software are not compatible with use of higher-order multipoles. For this reason, here, our molecular simulations use an induced dipole model to describe polarization effects, with functional form identical to that from the AMOEBA force field.[39] Numerical differences between the Drude oscillator and induced dipole models were found to be negligible.

*4.3.2. Second Virial Calculations.* Classical second virial coefficients were calculated for $NH_3$, $H_2O$, $CO_2$, and $CH_3Cl$ using rigid monomer geometries and following the procedure described in ref 98.

*4.3.3. $\Delta H_{sub}$ for $CO_2$.* For $CO_2$, the molar enthalpy of sublimation was determined according to

$$\Delta H_{sub} = H_g - H_{crys}$$

$$= (U_g + PV_g) - (U_{el,crystal,0K} + \Delta U_{el,crystal,0K \to T_{sub}} + PV_{crys} + E_{vib,crystal})$$

$$\approx (RT) - \left(U_{el,crystal,0K} + \int_{0K}^{T_{sub}} C_p \, dT + E_{vib,crystal}\right) \tag{25}$$

which assumes ideal gas behavior and $PV_g \gg PV_{crys}$. For the crystal, an experimental measure of $C_p$ was obtained from ref 104 and numerically integrated to obtain a value $\Delta U_{el,crystal,0K \to T_{sub}} = 6.70$ kJ mol$^{-1}$. Theoretical measures of $E_{vib,crystal} \approx 2.24-2.6$ kJ mol$^{-1}$ were obtained from ref 105 and ref 106, respectively, and $U_{el,crystal,0K}$ was determined from the intermolecular force field using a unit cell geometry taken from the experiment.[107]

*4.3.4. Other $CO_2$ Simulations.* To determine the densities and enthalpies of vaporization used in this work, simulations were run in OpenMM using NPT and NVT ensembles, respectively. Bulk $CO_2$ was modeled using 780 rigid $CO_2$ molecules and periodic boundary conditions. Electrostatic interactions were described with the particle-mesh Ewald (PME) method, three-body dispersion was treated using a 9 Å cutoff, and the remainder of the potential was computed using a 14 Å cutoff and long-range energy/pressure corrections. A Langevin integrator (with a friction coefficient of 2.0 ps$^{-1}$) and Monte Carlo barostat were utilized, when required, for
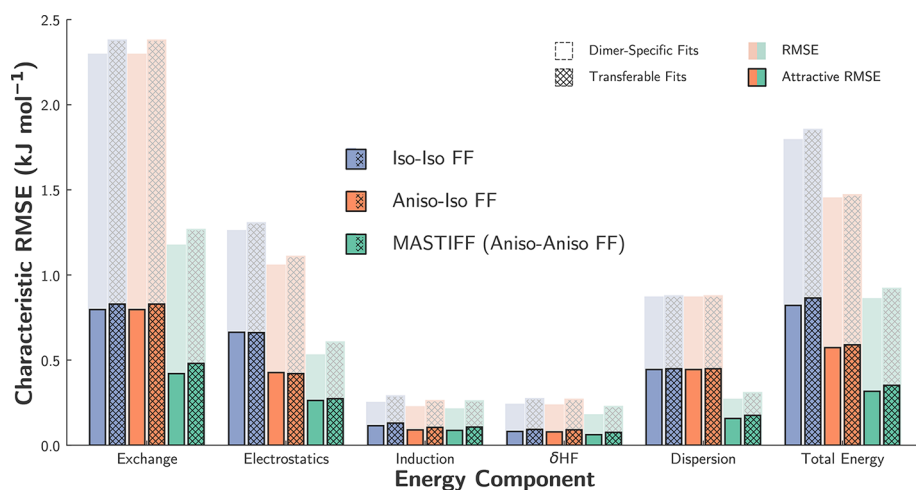
**Figure 2.** Characteristic RMSE (as described in the main text) for the Iso-Iso FF (purple), Aniso-Iso FF (orange), and MASTIFF (green) over the 91 dimer test set. The semitransparent bars represent total RMSE for each energy component, while the smaller solid bars represent "Attractive" RMSE, in which repulsive points have been excluded. For each force field, two types of fits are displayed: dimer-specific (solid) and transferable (hashed lines) (see Section 5.3 for details). Finally, note that, for Iso-Iso FF and Aniso-Iso FF, only the electrostatic and total energy RMSE values differ.

temperature and pressure coupling. A cubic box with isotropic coupling was used for NPT simulations, and a 0.5 fs time step was used for all simulations. Under these conditions, and using an unoptimized version of OpenMM (see Supporting Information for details), simulations speeds were ~2.5 ns/day (for MASTIFF) or ~3.1 ns/day (for Aniso-Iso FF). After an equilibration period of at least 100 ps, simulation data was collected for a minimum of 1 ns. Average densities were obtained directly from the simulation, and the molar enthalpy of vaporization for $CO_2$ was determined from the following formula:

$$\Delta H_{vap} = H_g - H_{liq}$$
$$= U_g - U_{liq} + P(V_g - V_{liq}) \tag{26}$$

Note that, at the state points studied, the ideal gas approximation is insufficiently accurate, and thus simulations were run for both the gas and liquid phases at experimentally determined densities and pressures.[108]

## 5. RESULTS AND DISCUSSION

**5.1. Overview.** We now benchmark our developed anisotropic force field methodology against various sum-of-spheres approximations. As is standard in ab initio force field development, we will principally rely on the following metric for force field quality: the accuracy with which a given force field functional form can reproduce high-quality ab initio benchmark energies. Furthermore, our choice of relevant benchmark energies is guided by the many-body expansion (MBE),[12,109] whereby the energy of an arbitrary $N$-particle system is expressed as a sum of $n$-body interaction potentials,

$$E_N(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N) = \sum_{i<j}^N E_2(\vec{r}_i, \vec{r}_j) + \sum_{i<j<k}^N \Delta E_3(\vec{r}_i, \vec{r}_j, \vec{r}_k) + ... \tag{28}$$

$E_2$, the "pair potential", is defined as the difference in interaction energies between a molecular dimer and the individual monomers themselves; $\Delta E_3$ corresponds to the nonadditive contributions (energy note taken into account in

$E_2$) to the interaction energies of trimers, and higher-order terms in the expansion are defined analogously. Aside from many-body polarization, for which the complete $N$-body effects can readily be calculated,[12,110] the MBE typically converges rapidly, such that only $E_2$ and occasionally $\Delta E_3$ terms are required to completely and accurately describe $E_N$.[12,16] (Notably, the combination of $E_2$ and $N$-body polarization often account for upward of 90%–95% of the total interaction energy;[16,111] as discussed in Section 5.5, any important contributions from $\Delta E_3$ can be taken into account separately and systematically using known methods.[111,112]) Thus, because the accuracy and predictive power of an ab initio force field is dependent substantially on the accuracy with which we can describe $E_2$, and because the functional forms introduced in Section 3 directly affect only this pairwise-additive portion of the intermolecular potential, *we primarily concentrate our efforts on assessing force field quality, with respect to benchmark calculations of dimer interaction energies.*

In addition to the above comparisons to ab initio benchmarks, a secondary goal of this work is to evaluate the extent to which the force field methodologies presented here can be used, not only to reproduce ab initio benchmarks, but also to accurately simulate experimental properties. Especially for ab initio force fields, accurate comparisons to the experiment are dependent not only on the quality of the two-body force field (as defined above), but also on the accuracy of the benchmark electronic structure theory, the treatment of many-body and/or quantum effects, etc. Thus, for select systems, we also compare our force fields to experimental second virial coefficients and bulk properties, with the goal of offering preliminary insight into how our anisotropic force field methodology might be utilized, in conjunction with accurate electronic structure theory and a proper treatment of many-body effects, to yield a complete $N$-body force field capable of accurately simulating experimental properties across a wide range of phase space.

**5.2. Accuracy: Comparison with DFT-SAPT.** We compare between three models in this work (see Section 4 for detailed functional forms): Iso-Iso FF, which uses a completely isotropic description of all energy components;

**Table 1. "Improvement Ratios" for Each Homomonomeric Species in the 91 Dimer Test Set**[a]

| | Exchange | Electrostatics | Induction | $\delta$HF | Dispersion | Total Energy |
|---|---|---|---|---|---|---|
| $H_2O$ (O,H) | 4.96 | 13.12 | 1.69 | 1.88 | 8.20 | 11.54 |
| $CO_2$ (C,O) | 3.83 | 9.13 | 0.99 | 0.64 | 4.91 | 8.62 |
| $NH_3$ (N,H) | 3.15 | 5.36 | 0.90 | 2.86 | 2.45 | 5.78 |
| Ethene (C,H) | 1.44 | 1.46 | 1.00 | 1.00 | 7.59 | 4.16 |
| Chloromethane (Cl) | 3.17 | 4.03 | 1.36 | 1.04 | 4.20 | 4.08 |
| Methyl Amine (N,H) | 1.70 | 2.93 | 1.05 | 2.22 | 2.95 | 2.37 |
| Methanol (O,H) | 1.81 | 3.05 | 1.11 | 2.03 | 1.00 | 2.36 |
| Dimethyl Ether (O) | 1.30 | 2.07 | 1.38 | 1.19 | 1.85 | 2.30 |
| Ethanol (O,H) | 1.29 | 3.10 | 1.04 | 1.45 | 1.79 | 2.14 |
| Acetone (O) | 1.58 | 1.98 | 1.03 | 1.34 | 1.51 | 1.08 |
| Ethane () | 1.00 | 1.26 | 1.05 | 1.01 | 1.00 | 1.08 |
| Ar () | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Methane () | 1.00 | 0.93 | 0.99 | 1.01 | 1.00 | 0.94 |

[a]For each dimer and energy component, the improvement ratio is calculated as the ratio of aRMSE between Iso-Iso FF and MASTIFF; values greater than 1 indicate decreased errors in the anisotropic model. Entries have been ordered according to the improvement ratio for the total energy.

Aniso-Iso FF, which additionally accounts for multipolar electrostatic anisotropy; and MASTIFF, which incorporates anisotropy into all energy components of the intermolecular potential. For each of the 91 dimer combinations described in Section 4, and for each model, parameters were fit to reproduce benchmark DFT-SAPT (PBE0/AC) energies calculated for 1000 different relative orientations of the constituent monomers. From these "dimer-specific" fits, and as described in our prior work,[47] we then averaged the root-mean-square error (RMSE) and mean signed error (‖MSE‖) values from each of the 91 fits to produce so-called "characteristic RMSE/‖MSE‖", metrics representative of the errors associated with a given force field methodology. Because the absolute magnitudes of the various energy components becomes large in the repulsive portion of the potential, these characteristic errors are dominated by repulsive configurations. As such, we have also calculated "attractive RMSE/‖MSE‖" (aRMSE/a‖MSE‖), defined as the characteristic errors for the subset of configurations with total interaction energies $E_{tot}$ < 0. All computed characteristic RMSE are shown in Figure 2, with ‖MSE‖ data shown in the Supporting Information. Unless otherwise stated, results in this section refer exclusively to the "Dimer-specific" fits in Figure 2, with an explanation and full discussion of so-called "Transferable" fits given in Section 5.3.

Based on the characteristic RMSE shown in Figure 2, both Aniso-Iso FF and MASTIFF offer substantial improvements over the completely isotropic model Iso-Iso FF. Though unsurprising, given the well-studied importance of higher-order electrostatic multipole moments, Aniso-Iso FF shows reduced RMSE/aRMSE that are (depending on the exact error metric used) roughly 30% smaller than Iso-Iso FF. Both RMSE and aRMSE measures show similar gains in accuracy, indicating that inclusion of higher-order multipoles (henceforth "multipolar electrostatic anisotropy") is important in both attractive and repulsive regions of the potential. Crucially, the inclusion of additional "short-range anisotropies" (anisotropic interactions arising from overlap of monomer electron densities, namely exchange-repulsion and electrostatic/inductive charge penetration) and long-range "dispersion anisotropy" yields a *further* 40% reduction in RMSE/aRMSE for MASTIFF, compared to the Aniso-Iso FF. This latter result is highly important, as it

suggests that, for the generation of highly accurate ab initio potentials, the combination of short-range and dispersion anisotropies are comparable in importance to multipolar electrostatic anisotropy. Indeed, this substantial increase in force field accuracy, arising from a *comprehensive* treatment of anisotropic effects, is one of the most important findings in the present work. In summary, and encouragingly, the combination of multipolar electrostatic, short-range, and dispersion anisotropies result in an overall 60% reduction in RMSE/aRMSE, when comparing Iso-Iso FF to MASTIFF.

Figure 2 also displays characteristic RMSE/aRMSE for each component of the force field, allowing us to account for the influence of anisotropy on a term-by-term basis. Immediately, one can see that (aside from induction, discussed below) an inclusion of atomic-level anisotropy greatly improves the description of each energy component. Unless otherwise stated, here, we report results for aRMSE and dimer-specific fits, although similar values are obtained for overall RMSE and for transferable fits. Compared to Iso-Iso FF, exchange errors in MASTIFF are reduced by 47%. Electrostatic errors are reduced by an even larger 60%. By evaluating the ratio of electrostatic errors between different models, we find that aRMSE(Aniso-Iso)/aRMSE(Iso-Iso) = 0.64 and aRMSE(MASTIFF)/aRMSE(Aniso-Iso) = 0.62, suggesting that *both* higher-order multipoles and anisotropic charge penetration terms are necessary to obtain an accurate description of the DFT-SAPT electrostatic energy. Finally, via an inclusion of dispersion anisotropy, aRMSE for dispersion are reduced by a significant 65%.

Although the trends for exchange, electrostatics, and dispersion universally suggest the importance of including atomic-level anisotropy, trends for terms describing the physics of polarization and charge-transfer (represented in DFT-SAPT by induction and $\delta^{HF}$) are less encouraging. On the one hand, including higher-order multipoles substantially lowers RMSE for induction, with aRMSE(Aniso-Iso)/aRMSE(Iso-Iso) = 0.70. Because both Iso-Iso FF and Aniso-Iso FF use isotropic polarizabilities, and because the induction energy is fundamentally dependent only on the polarizabilities and the static electric field, this result is clearly due to an improved treatment of the static electric field via anisotropy of the multipolar electrostatics. Once again, this suggests that an anisotropic

treatment of long-range electrostatics is crucial for accurate force field development. On the other hand, our functional form for anisotropic short-range induction (eqs 18 and 20) yields no improvement in the induction RMSE, with aRMSE-(MASTIFF)/aRMSE(Aniso-Iso) = 0.97. This observed lack of improvement is likely due to a combination of factors. First, and perhaps most importantly, we have chosen in this work to use isotropically averaged dipole polarizabilities, but as with electrostatics, anisotropy and higher-order terms have been shown to be important in the multipole expansion of atomic polarizabilities.[12,45,56,86,113] Second, and although probably a smaller source of error, it is also unclear how to optimally model the distance dependence of the induction energy at short intermolecular separations, where penetration and charge-transfer effects become important and the long-range polar-ization terms must be damped.[47,114−116] Given that the more elaborate short-range form of the MASTIFF induction model does not result in a tangible improvement, it is quite possible that alternative formulations are required for an accurate treatment of highly anisotropic induction.

To further analyze the effects of anisotropy on a molecule-by-molecule basis, we have calculated "improvement ratios", defined as aRMSE(Iso-Iso)/aRMSE(MASTIFF), for each energy component and for each homomonomeric species in the test set, results for which are shown in Table 1. (Improvement ratios for heteromonomeric species are given in the Supporting Information.) The most striking observation from the data presented in Table 1 is that the improvement ratios vary considerably with molecule. For example, with water, the aRMSE is improved by an order of magnitude when anisotropy is included. On the other hand, no improvement is seen for hydrocarbons such as ethane and methane (also see the Supporting Information). Consequently, anisotropy in the short-range expansions may be necessary for only some atoms types (see Section 6). Consistent with chemical intuition, we have found anisotropy to be particularly important for heteroatoms, $\pi$-bonded atoms, and all hydrogens bonded to anisotropic heavy atoms. Appealingly, this distinction between anisotropic and isotropic atom types simplifies force field parametrization and can enable more efficient molecular simulation (via a more cost-effective treatment of multipolar electrostatics) without sacrificing force field accuracy. Note that the current empirically determined definitions of anisotropic atom types match both chemical intuition and the more quantitative measures of atomic anisotropy proposed by other groups.[25,26]

Generally, the ordering of improvement ratios for exchange, electrostatics, dispersion, and the total energies (but not induction, see above) are reasonably correlated. Physically speaking, all atomically anisotropic interactions arise from the same source (atomically anisotropic electron densities), and so the observed correlation is perhaps to be expected. Never-theless, there are some exceptions to this trend. For ethene, relatively modest improvement ratios (roughly 1.4) are seen for exchange and electrostatics, whereas dispersion shows a much greater improvement ratio of 7.6. Since ethene homomono-meric interactions are dispersion-dominated, the improvement ratio for the total energy then roughly corresponds to that of dispersion. For acetone, there is strong correlation between the improvement ratios for exchange, electrostatics, and dispersion, which might lead one to suspect that the total energy improvement ratio would also be ~1.5−2.0. Nevertheless, for this molecule, the isotropic model benefits from error

cancellation between energy components, and the total energy aRMSE values between isotropic and anisotropic models are rather similar.

*Crucially, our results show that multipolar electrostatics is certainly not the exclusive, nor even always the dominant, source of atomic anisotropy.* Indeed, for molecules such as ethene, multipolar anisotropy in the electrostatic model is relatively unimportant, whereas dispersion anisotropy is essential for accurately modeling the $\pi$ interactions. Thus, generally, multipolar electrostatic, dispersion, and/or short-range aniso-tropies all must be taken into account in order to obtain accurate intermolecular models.

**5.3. Transferability: Comparison to DFT-SAPT.** From the above results, it is clear that, when explicitly parametrized, inclusion of anisotropy can greatly enhance the accuracy of an intermolecular potential. Nevertheless, for standard force field development, force field parameters must be *transferable* in order to be useful in the accurate prediction of intermolecular interactions in new chemical and/or physical environments. Indeed, in comparing simpler models to ones that introduce additional complexity, there is an ever-present danger that any accuracy gains from the more-complex functional form are simply due to overfitting or error cancellation,[117] ultimately resulting in a model with poor predictive ability and limited transferability.

We have previously shown how, with models similar to Iso-Iso FF[15,98] or Aniso-Iso FF,[47] it is possible to generate transferable potentials with applicability to a broad range of chemical and physical environments.[15] This transferability has been attributed to a combination of the physically meaningful energy decomposition of DFT-SAPT, parametrization on a component-by-component basis (rather than to the total energy), and the use of physically motivated functional forms and parameters.[15,47,98] MASTIFF largely shares this philosophy of force field development, and so we might also expect it to be transferable to heteromonomeric dimers. Indeed, the long-range multipolar electrostatic model is rigorously transferable, as are the isotropic long-range induction and dispersion coefficients used in the force field.[12,16] However, the *overall* transferability of MASTIFF cannot be taken for granted, because of the specific way in which we have incorporated non-electrostatic anisotropic effects. First, we have relied on several separability ansatzes (eqs 4 and 7), and second, in doing so, we have implicitly neglected potentially important interaction functions that are dependent on the relative orientation between monomers (see Appendix A). Both of these assumptions may affect the transferability of the resulting force field.

To assess the transferability of the MASTIFF model, we analyze the extent to which parameters developed for the homomonomeric systems can be used, without modification, to describe the interactions of the mixed dimers. Such an out-of-sample prediction, which is easily accomplished with our 91-dimer test set, is a direct measure of the extent to which our pair potentials can be applied to new chemical environments. For these transferable fits, parameters were fit to the 13 homomonomeric systems, and the combination rules shown in eq 18 were used to generate force fields for the remaining heteromonomeric systems. Thus, with these transferable fits, we have essentially generated 78 000 predictions from fits to 13 000 data points. RMSE and aRMSE values for these fits are shown in Figure 2, and we treat relative differences between these quantities for the "dimer-specific" and "transferable" fits
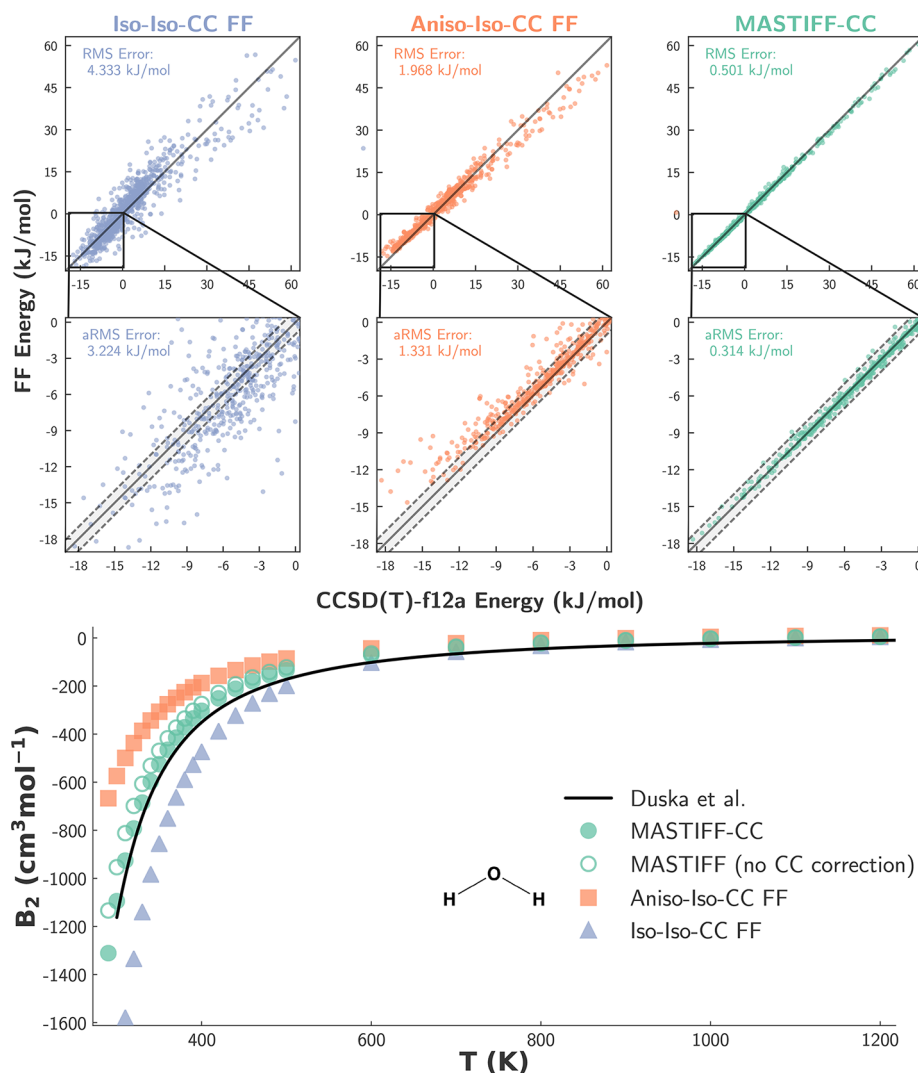
**Figure 3.** (Top) Force field fits for the water dimer using the Iso-Iso-CC FF (purple), Aniso-Iso-CC FF (orange), and MASTIFF-CC (green) methodologies. The $y = x$ line indicates perfect agreement between reference CCSD(T)-F12a energies and each force field, while shaded gray areas represent points within $\pm1$ kJ mol$^{-1}$ agreement of the benchmark. RMSE and aRMSE are as described in the main text. (Bottom) Classical second virials for water, with experimental data (black line) taken from ref 122. Note that some data points from Iso-Iso FF extend below the plot area.

as a measure of the extent of transferability for each force field methodology.

Remarkably, all three force fields—Iso-Iso, Aniso-Iso, and MASTIFF—perform similarly for the dimer-specific and transferable fits, both for the individual interaction energy components and for the total interaction energy. The degree of transferability of the MASTIFF model is very encouraging, and it indicates that the manner in which we have chosen to include the anisotropy is meaningful and does not lead to overfitting, but rather increases the accuracy of the intermolecular potentials for both in-sample and out-of-sample systems.

**5.4. Accuracy: Second Virials.** Having compared our various force fields methodologies against DFT-SAPT, we now turn our focus to our secondary goal in this work, that of evaluating the extent to which our anisotropic force field methodology can be used to more accurately simulate experimental properties. To this end, we begin by benchmarking our force fields against experimental second virial coefficients, which offer a direct experimental measure of the pair potential ($E_2$) without the complication of many-body effects (which will be discussed in Section 5.5). Notably,

comparisons to experimental second virial coefficients are dependent not only on the quality of a force field (as measured in Section 5.2), but also on the accuracy of the benchmark electronic structure theory used to fit the force field. Consequently, and to evaluate possible inaccuracies in our DFT-SAPT/aVTZ+m[47] benchmark energies, we have also parametrized models with respect to CCSD(T)-F12a/aug-cc-pVTZ+m, a level of theory which serves as a computationally affordable yet accurate prediction of the CCSD(T)/CBS limit.[118,119] We refer to these coupled cluster-based models with a -CC suffix, e.g., MASTIFF-CC; details of the refitting procedure (which minimally effect the dispersion model) can be found earlier in Section 4. Thus, aside from quantum effects (which are negligible for $CO_2$[120] and well-benchmarked for $H_2O$[121]), our second virial predictions should offer a fairly direct comparison between different models, levels of electronic structure theory, and experiment.

Using both our original and -CC potentials, we have calculated second virial coefficients for each Iso-Iso-CC FF, Aniso-Iso-CC FF, MASTIFF, and MASTIFF-CC, and for the
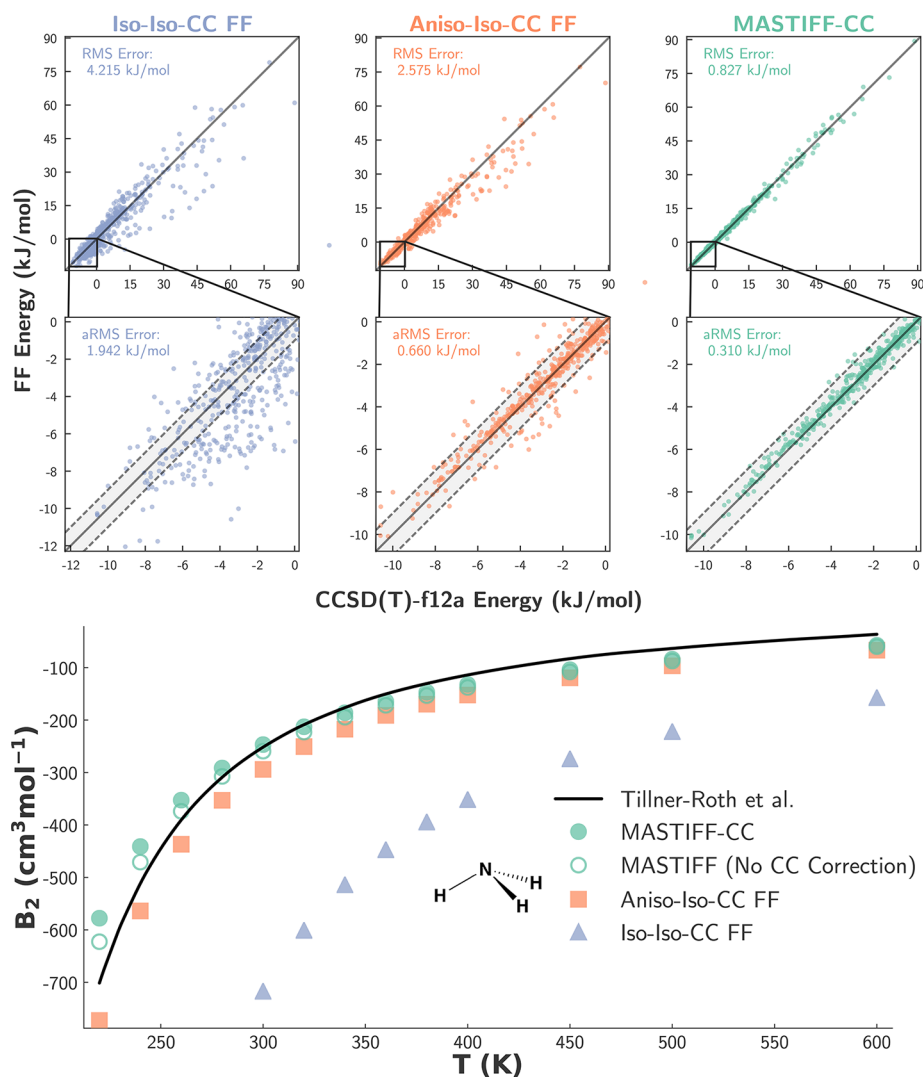
**Figure 4.** Force field fits and classical second virials for ammonia, as in Figure 3, but with experimental data taken from ref 123.

following systems: $H_2O$ (Figure 3), $NH_3$ (Figure 4), $CH_3Cl$ (Figure 5), and $CO_2$ (Figure 6).

Immediately, we observe that the effect of the coupled cluster (-CC) correction is minimal (compared to differences in force field methodologies) for most systems, with the exception of $CO_2$, where DFT-SAPT exhibits modest deficiencies with respect to CCSD(T)-f12a (see the Supporting Information and ref 119). Furthermore, we find that that the MASTIFF (and especially MASTIFF-CC) methodologies predict virial coefficients that closely correspond to experimental data. Generally, the Iso-Iso-CC FF predictions are much worse than their MASTIFF-CC or Aniso-Iso-CC FF counterparts, suggesting that an accurate treatment of long-range electrostatics is essential to obtain accurate virial coefficients. Finally, although Aniso-Iso FF-CC gives equally good predictions for some systems (notably $CH_3Cl$), compared to the MASTIFF-CC method, virial coefficients for other systems (especially $H_2O$) are less accurate, suggesting that dispersion and short-range anisotropies are also important in many systems for the accurate prediction of virial coefficients.

Generally, and given the range of systems tested ($CO_2$ dimer interactions are dispersion-dominated, while $CH_3Cl$, $NH_3$, and $H_2O$ have relatively larger electrostatic and polarization contributions), these second virial calculations suggest that,

when fit to gold-standard electronic structure theories, our anisotropic force field methodology offers an improved strategy for developing quantitatively accurate pair potentials.

**5.5. Accuracy: Condensed-Phase Properties of $CO_2$.** A major goal for standard force fields is that they be capable of accurately simulating bulk properties. To this end, we require not only an accurate pair potential, but also (in many cases) a proper treatment of polarization and other many-body effects. To provide a first example of how the MASTIFF methodology might be used as the pair potential in a complete, many-body force field useful for condensed-phase simulation, here we have developed and tested a force field for $CO_2$, which includes both pairwise additive and many-body effects. Based on its accuracy in predicting second virial coefficients, we use the MASTIFF-CC potential from Section 5.4 to describe both the pairwise potential and the many-body induction. Yet, non-inductive many-body effects have been shown to be important for $CO_2$,[102,112,125,126] and so we have additionally developed and tested a model for three-body dispersion based on the three-body dispersion potential developed by Oakley and Wheatley (see Section 4). Three-body exchange effects are not taken into account in our model; however, prior work shows they are very small under the conditions studied here.[112] Using the various $CO_2$ models described above, we have run bulk simulations for
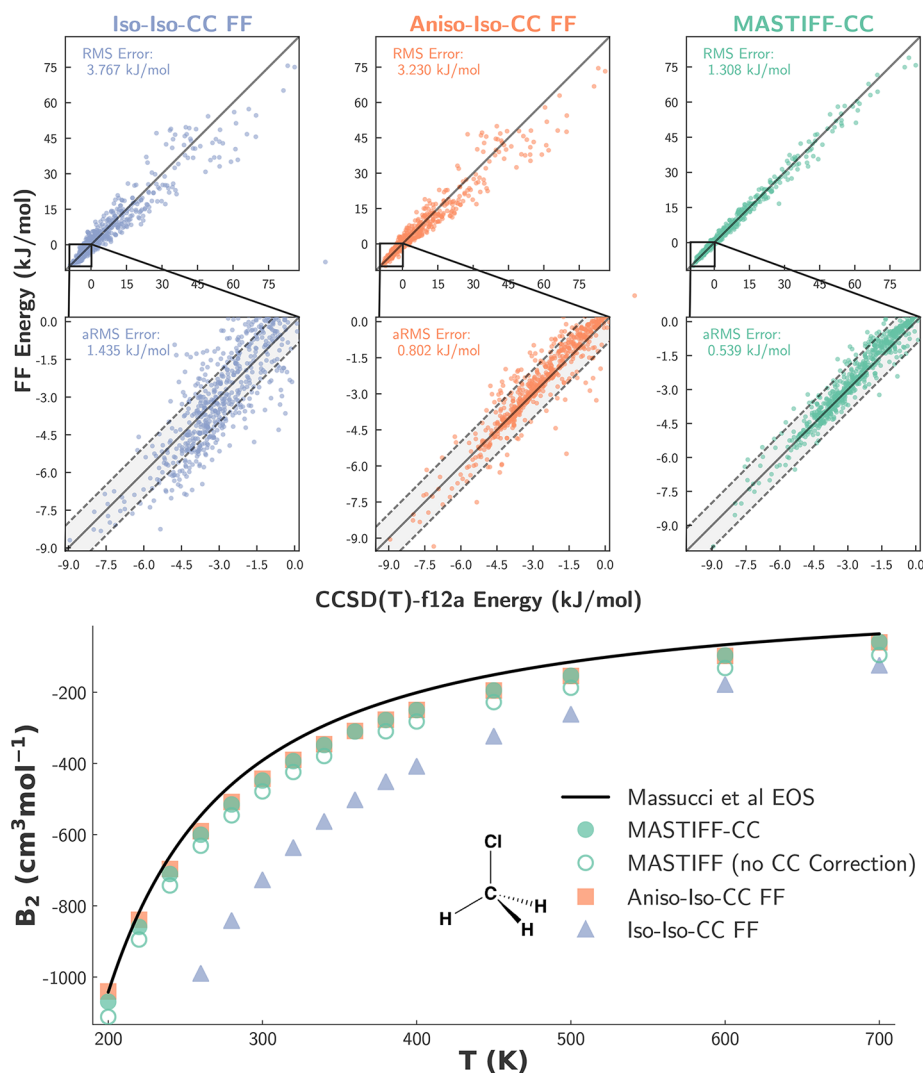
**Figure 5.** Force field fits and classical second virials for chloromethane, as in Figure 3, but with experimental data from the experimental equation of state (EOS) given in ref 124.

a rigid model of $CO_2$ over a variety of vapor, liquid, supercritical, and solid phase points. Density predictions for the vapor, liquid, and supercritical phases of $CO_2$ are shown in Table 2 and in the Supporting Information (Tables S3 and S4), and enthalpies of sublimation and vaporization are shown in Table 3 and Table S5. Simulations with a flexible $CO_2$ model yielded similar results, which are also given in the Supporting Information.

As anticipated from prior work,[127] complete neglect of three-body dispersion (Table S3) leads to an overestimation of the density at all phase points studied, particularly in the denser liquid phases. Although not surprising, this result underscores the importance of including many-body effects (at least for $CO_2$) when developing highly accurate ab initio force fields.[111,112] Upon including three-body dispersion effects, however, MASTIFF-CC succeeds in reproducing *all* studied experimental properties to within a few percent (see Tables 2 and 3). (As shown in Table S4, Aniso-Iso-CC FF reproduces some, but not all, experimental properties to within this level of accuracy, and Iso-Iso-CC FF generally has poor quantitative agreement with the experiment.) Importantly, MASTIFF-CC can correctly predict the $CO_2$ sublimation enthalpy, a quantity that is critically dependent on the lattice energy of the solid

phase. Unlike with liquid or supercritical $CO_2$, where many dimer configurations are sampled, the solid consists of only four symmetry-unique configurations. Consequently, whereas an isotropic potential might yield good property predictions for the liquid phase via averaging and/or error cancellation, it would not be expected to correctly predict the solid phase, where beneficial error cancellation is unlikely. Indeed, most theories (including Aniso-Iso-CC FF, Iso-Iso-CC FF, our previously developed SYM-3B model,[112] nearly all popular empirically developed $CO_2$ models,[128] AMOEBA,[106] and many electronic structure theories[106]) struggle to correctly predict the solid-phase properties of $CO_2$. For this reason, the enthalpy of sublimation is considered an extremely stringent test of force field quality,[128] and the fact that MASTIFF-CC can accurately reproduce this quantity is evidence for both the excellent quality of the many-body MASTIFF-CC potential (specifically) and of the importance of atomic-level anisotropy (generally). Overall, our $CO_2$ results are a preliminary indication that, provided we correctly account for many-body effects and benchmark against a gold-standard electronic structure theory, our newly developed anisotropic methodology may successfully be used as the basis for accurate, "next-generation" force fields
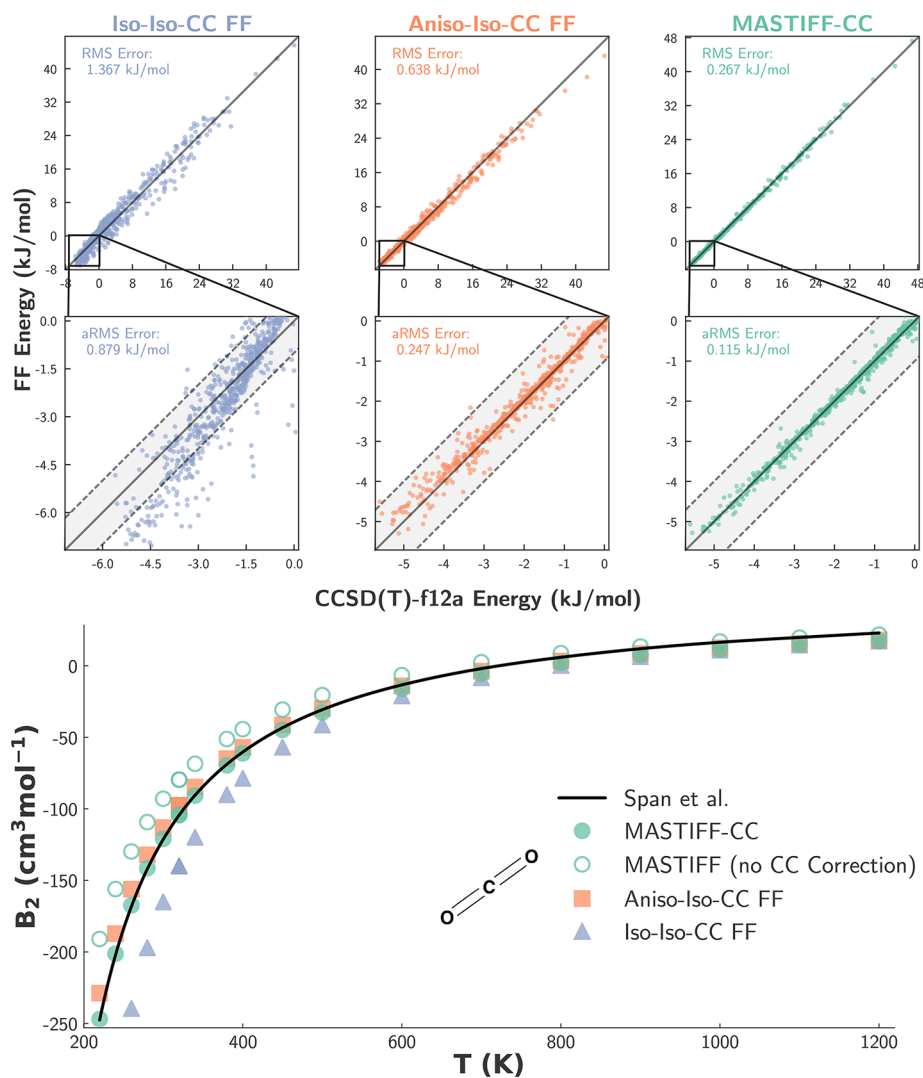
**Figure 6.** Force field fits and classical second virials for $CO_2$, as in Figure 3, but with experimental data taken from ref 108.

## Table 2. Select Densities for $CO_2$ across a Range of Experimental Conditions[a]

| phase | temp, $T$ (K) | pressure, $P$ (bar) | density (g/mL) | exp | % error |
|---|---|---|---|---|---|
| gas | 300 | 50 | 0.131 | 0.128 | 2.34 |
| supercritical | 320 | 140 | 0.728 | 0.703 | 3.56 |
| liquid | 300 | 100 | 0.825 | 0.802 | 2.87 |
| liquid | 273.15 | 100 | 1.000 | 0.974 | 2.67 |

[a]Experimental data taken from the EOS of ref 108. Entries ordered by increasing experimental density.

## Table 3. Enthalpies of Vaporization/Sublimation for $CO_2$ at Several Temperatures[a]

| phases | temp, $T$ (K) | $\Delta H$ (kJ mol$^{-1}$) | | % error |
|---|---|---|---|---|
| | | obs | exp | |
| s → g | 194.76 | 25.0 ± 0.15 | 25.2 | −0.8 |
| l → g | 288 | 7.92 | 7.80 | −1.4 |

[a]Experimental data taken from the EOS of ref 108. The uncertainty in the enthalpy of sublimation is due to ambiguity in the theoretical zero-point energy for $CO_2$ (see Section 4).

amenable to the molecular simulation of bulk properties in a variety of phases.

Despite the success of our MASTIFF-CC model for $CO_2$, it is also worthwhile to address and understand its minor shortcomings. In particular, we have studied representative two- and three-body energies taken from a snapshot of the liquid at 273.15 K and 100 bar (see Figure S5 in the Supporting Information). When benchmarked against the accurate PES developed by Hellmann,[125] the crude three-body potential utilized above is found to be systematically in error. Although some of this error may be due to inaccuracies in the benchmark potential itself, as compared to coupled-cluster,[125] most of this error is likely due to inaccuracies in our model for many-body $CO_2$ interactions. The atomically isotropic treatment of three-body dispersion, neglecting higher-order dispersion terms, and neglecting explicit three-body exchange, may all contribute to this error, and an improved model for many-body $CO_2$ interactions will be the subject of future research. Indeed, it is well-known that the density can be extremely sensitive to the treatment of many-body effects,[126] and it is highly probable that an improved many-body model would reduce the already small errors observed in our MASTIFF-CC predictions. Regardless (and despite some small residual errors arising from the simplified treatment of many-body effects), it appears that the

MASTIFF-CC methodology yields an extremely accurate two-body force field for $CO_2$, with broad applicability across a range of experimentally important phases.

## 6. CONCLUSIONS AND FUTURE WORK

We have developed a comprehensive methodology for modeling atomic-level anisotropy in standard intermolecular force fields. Via a simple extension to standard isotropic force fields,[47] we have demonstrated how a computationally efficient treatment of atomic-level anisotropy can lead to significant improvements in models for intermolecular interactions. Critically, and in contrast to popular assumption, an accurate treatment of multipolar electrostatics does not, *by itself*, account for all energetically important effects of atomic-level anisotropy. Rather, our results indicate that the combined anisotropy of dispersion, exchange, and charge penetration is of comparable importance to long-range multipolar electrostatics, and must be comprehensively taken into account in order to obtain intermolecular force fields of the highest quality. In agreement with the more quantitative metrics proposed by others,[25,26] we have found a comprehensive model of atomic-level anisotropy to be particularly important for obtaining sub-kJ $mol^{-1}$ accuracy when describing molecules with heteroatoms (particularly ones with exposed lone pairs), carbons in multiple bonding environments, and hydrogens bound to anisotropic heavy atoms. As such, our "MASTIFF" methodology show great promise with respect to both high-quality electronic structure benchmark energies and experimental property predictions, all while maintaining high transferability and ease of implementation in existing software packages for use in condensed-phase simulation.[99]

Nonetheless, several aspects of our current force field methodology require further improvement and/or study before our anisotropic MASTIFF approach can be used to develop standard force fields for arbitrary organic and/or biological systems. As an example, future work will be required to investigate how well the MASTIFF methodology can be applied to studies of large and/or nonrigid systems, although similar isotropic models have previously been shown to transferably combine with intramolecular potentials in order to describe molecular flexibility.[98] In addition, an improved description of induction effects will become essential for accurate bulk simulations of highly polarizable molecules, such as water. We are currently working to develop improved models that can describe both long-range anisotropic polarization and short-range polarization damping, as these aspects of the force field critically affect both the two- and many-body induction energies and can account for a sizable fraction of the total interaction energy in condensed phases. We anticipate that improved models for molecular flexibility and induction will, in combination with an accurate description of non-inductive many-body effects, yield a general approach to force field development that accurately models arbitrary $N$-body intermolecular interactions, in turn, enabling highly accurate, "next-generation" force field development capable of simulating a wide array of phases and chemical environments.

## ■ APPENDIX A: $\overline{S}$-FUNCTIONS AND THE MOTIVATION FOR $G(\theta_i, \phi_i, \theta_j, \phi_j)$

As shown elsewhere,[66,67] an exact model (under the ansatz of radial and angular separability) for $g(\theta_i, \phi_i, \theta_j, \phi_j)$ is given by Stone's $\overline{S}$-functions, which form a complete basis set for describing any scalar function, which is dependent on the relative orientation between molecules. These $\overline{S}$-functions are given (following Stone's notation[16]) by the formula

$$\overline{S}_{l_1 l_2 j}^{k_1 k_2} = i^{l_1 - l_2 - j} \begin{pmatrix} l_1 & l_2 & j \\ 0 & 0 & 0 \end{pmatrix}^{-1} \sum_{m_1 m_2 m} [D_{m_1 k_1}^{l_1}(\Omega_1)]^* [D_{m_2 k_2}^{l_2}(\Omega_2)]^* \\ C_{lm}(\theta, \phi) \begin{pmatrix} l_1 & l_2 & j \\ m_1 & m_2 & m \end{pmatrix} \quad \text{(A-1)}$$

The general form of these $\overline{S}$-functions can be quite complicated, involving both the Wigner $D$ rotation matrices and Wigner $3j$-symbols (quantities in parentheses) as well as the degree ($l_1$, $l_2$, and $j$) and order ($m_1$, $m_2$, and $m$ for the global coordinate system, $k_1$ and $k_2$ for the various local coordinate systems) of the spherical harmonic tensors. Here, subscripts reference either molecule 1 or molecule 2, and subscriptless quantities refer to the dimer as a whole.

In order to obtain a functional form for the exchange-repulsion that is amenable to simple combination rules (a necessary prerequisite for transferable potentials), we must somehow be able to separate $g(\theta_i, \phi_i, \theta_j, \phi_j)$ into monomer contributions. Unfortunately, many of the $\overline{S}$-functions are dependent on the relative orientation of the dimer itself, and thus must be excluded in the development of *transferable* potentials. Thus, as a second ansatz (empirically validated by us in Section 5 and by others[129]), we neglect all contributions from $\overline{S}$-functions that are dependent on both local coordinate systems. This leaves us with two sets of $\overline{S}$-functions, namely,

$$\overline{S}_{l0l}^{k0} = C_{lk}(\theta_i, \phi_i) \quad \text{(A-2)}$$

and

$$\overline{S}_{0ll}^{0k} = C_{lk}(\theta_j, \phi_j) \quad \text{(A-3)}$$

which are simply the renormalized spherical harmonics (eq 8) expressed in each of the two local coordinate systems.

Given our truncated expressions for the $\overline{S}$-functions, we now need only extend our functional form for $f(r_{ij})$ to incorporate these anisotropic contributions. We choose, in a manner analogous to the literature precedent,[6,12,13,17,43,45,76,78,79] to expand the $A_i^{exch}$ and $A_j^{exch}$ parameters of eq 7, in terms of a truncated expansion of $\overline{S}$-functions. (In principle, we could also account for anisotropy in the $B_{ij}$ parameters of our model for $f(r_{ij})$. However, previous literature suggests that, in practice, this "hardness" parameter can often be treated as constant, and we also neglect its possible anisotropy in this work.) Consequently, all short-range anisotropies are modeled in this work by the expressions given in eqs 9 and 10.

In addition to describing exchange-repulsion, $\overline{S}$-functions can also be used to accurately describe the orientation dependence of long-range electrostatic, induction, and dispersion energies. (See refs 12 and 16 for complete details.) The electrostatic interaction tensor from eq 14 can be expressed, in terms of $\overline{S}$-functions, as[16]

$$T_{tu}^{ij} \equiv T_{l_1, k_1, l_2, k_2}^{ij} = \begin{pmatrix} l_1 + l_2 \\ l_1 \end{pmatrix} \overline{S}_{l_1 l_2 l_1 + l_2}^{k_1 k_2} r_{ij}^{-l_1 - l_2 - 1} \quad \text{(A-4)}$$

where both the $tu$ or $l_1, k_1, l_2, k_2$ notations label the angular momentum of the multipole components. The long-range induction energy is also explicitly dependent on the electro-

static interaction tensor (and, hence, is implicitly dependent on the $\overline{S}$-functions),[16]

$$V^{\text{ind}} = \frac{1}{2} \sum_{I} \sum_{I \neq J} \Delta Q_t^{\,i} T_{tu}^{ij} Q_u^{\,j} \quad \text{(A-5)}$$

with $\Delta Q$ and $Q$ defining the induced and permanent multipoles, respectively, and $I$ and $J$ representing individual molecules. Lastly, the orientation dependence of the long-range dispersion is accurately described by the formula[12]

$$V^{\text{disp}} = -\frac{1}{2\pi} \sum_{iji'j'} \sum_{tut'u'} T_{tu}^{ij} T_{t'u'}^{i'j'} \int_0^{\infty} \alpha_{tt'}^{ii'}(i\nu) \alpha_{uu'}^{jj'}(i\nu) \, d\nu \quad \text{(A-6)}$$

where the primes describe the response of the local polarizability $(\alpha^{ii'})$ at site $i'$ to a perturbation at $i$, and the integration is carried out over all imaginary frequencies $i\nu$.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.7b00851.

> ‖MSE‖ values for the 91 dimer test set. Improvement ratios for all 91 dimers. Local axis definitions for all 13 monomers. MASTIFF parameters for homomonomeric systems. MASTIFF-CC parameters and OpenMM input files for $H_2O$, $CO_2$, $NH_3$, and $CH_3Cl$. Iso-Iso FF, Aniso-Iso FF, and MASTIFF fitting quality for homomonomeric systems. Representative 2- and 3-body energies taken from liquid $CO_2$. (PDF)
>
> Force field parameters and dimer energies for homomonomeric systems (ZIP)
>
> OpenMM input files for MASTIFF-CC force fields (ZIP)

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: schmidt@chem.wisc.edu.

### ORCID Ⓘ

J. R. Schmidt: 0000-0002-1438-117X

### Notes

The authors declare no competing financial interest.

## ■ REFERENCES

(1) Rezáč, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038−5071.

(2) Chalasinski, G.; Szczesniak, M. M. State of the Art and Challenges of the ab Initio Theory of Intermolecular Interactions. *Chem. Rev.* **2000**, *100*, 4227−4252.

(3) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L. The S66 × 8 benchmark for noncovalent interactions revisited: explicitly correlated ab initio methods and density functional theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20905−20925.

(4) Hassanali, A. A.; Cuny, J.; Verdolino, V.; Parrinello, M. Aqueous solutions: state of the art in ab initio molecular dynamics. *Philos. Trans. R. Soc., A* **2014**, *372*, 20120482−20120482.

(5) Zgarbová, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. Large-scale compensation of errors in pairwise-additive empirical force fields: comparison of AMBER intermolecular terms with rigorous DFT-SAPT calculations. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476−10493.

(6) Stone, A. J.; Price, S. L. Some new ideas in the theory of intermolecular forces: anisotropic atom-atom potentials. *J. Phys. Chem.* **1988**, *92*, 3325−3335.

(7) Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: Theory and applications. *Theor. Chem. Acc.* **2009**, *124*, 11−28.

(8) Panagiotopoulos, A. Z. Force Field Development for Simulations of Condensed Phases. *AIChE Symp. Ser.* **2000**, 1−10.

(9) Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y. Development of polarizable models for molecular mechanical calculations II: Induced dipole models significantly improve accuracy of intermolecular interaction energies. *J. Phys. Chem. B* **2011**, *115*, 3100−3111.

(10) Demerdash, O.; Yap, E.-H.; Head-Gordon, T. Advanced potential energy surfaces for condensed phase simulation. *Annu. Rev. Phys. Chem.* **2014**, *65*, 149−74.

(11) Lopes, P. E. M.; Guvench, O.; MacKerell, A. D. *Methods Mol. Biol.* **2015**, *1215*, 47−71.

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

(13) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Phys. Chem. Chem. Phys.* **2010**, *12*, 8478.

(14) Welch, G. W. A.; Karamertzanis, P. G.; Misquitta, A. J.; Stone, A. J.; Price, S. L. Is the Induction Energy Important for Modeling Organic Crystals? *J. Chem. Theory Comput.* **2008**, *4*, 522−532.

(15) Schmidt, J. R.; Yu, K.; McDaniel, J. G. Transferable Next-Generation Force Fields from Simple Liquids to Complex Materials. *Acc. Chem. Res.* **2015**, *48*, 548−556.

(16) Stone, A. J. *The Theory of Intermolecular Forces*, 2nd Edition; OUP: Oxford, U.K., 2013.

(17) Price, S. L. Toward More Accurate Model Intermolecular Potentials for Organic Molecules. *Rev. Comput. Chem.* **2000**, *14*, 225−289.

(18) Coppens, P.; Guru Row, T. N.; Leung, P.; Stevens, E. D.; Becker, P. J.; Yang, Y. W. Net atomic charges and molecular dipole moments from spherical-atom X-ray refinements, and the relation

between atomic charge and shape. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1979**, *35*, 63−72.

(19) Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.

(20) Nyburg, S. C.; Faerman, C. H. A revision of van der Waals atomic radii for molecular crystals: N, O, F, S, Cl, Se, Br and I bonded to carbon. *Acta Crystallogr., Sect. B: Struct. Sci.* **1985**, *41*, 274−279.

(21) Batsanov, S. S. Van der Waals Radii of Elements. *Inorg. Mater.* **2001**, *37*, 871−885.

(22) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16789−94.

(23) Lommerse, J. P. M.; Stone, A. J.; Taylor, R.; Allen, F. H. The nature and geometry of intermolecular interactions between halogens and oxygen or nitrogen. *J. Am. Chem. Soc.* **1996**, *118*, 3108−3116.

(24) Eramian, H.; Tian, Y.-H.; Fox, Z.; Beneberu, H. Z.; Kertesz, M. On the anisotropy of van der Waals atomic radii of O, S, Se, F, Cl, Br, and I. *J. Phys. Chem. A* **2013**, *117*, 14184−14190.

(25) Wheatley, R. J.; Gopal, A. A. Covalent bond orders and atomic anisotropies from iterated stockholder atoms. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2087−2091.

(26) Kramer, C.; Spinn, A.; Liedl, K. R. Charge Anisotropy: Where Atomic Multipoles Matter Most. *J. Chem. Theory Comput.* **2014**, *10*, 4488−4496.

(27) Badenhoop, J. K.; Weinhold, F. Natural steric analysis: Ab initio van der Waals radii of atoms and ions. *J. Chem. Phys.* **1997**, *107*, 5422.

(28) Kim, H.; Doan, V. D.; Cho, W. J.; Madhav, M. V.; Kim, K. S. Anisotropic Charge Distribution and Anisotropic van der Waals Radius Leading to Intriguing Anisotropic Noncovalent Interactions. *Sci. Rep.* **2015**, *4*, 5826.

(29) Bankiewicz, B.; Palusiak, M. The shape of the halogen atom — anisotropy of electron distribution and its dependence on basis set and method used. *Struct. Chem.* **2013**, *24*, 1297−1306.

(30) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chem. Rev.* **2016**, *116*, 7501−7528.

(31) Chessari, G.; Hunter, C. A.; Low, C. M. R.; Packer, M. J.; Vinter, J. G.; Zonta, C. An evaluation of force-field treatments of aromatic interactions. *Chem. - Eur. J.* **2002**, *8*, 2860−2867.

(32) Šponer, J.; Šponer, J. E.; Mládek, A.; Jurečka, P.; Banáš, P.; Otyepka, M. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics and experiment. *Biopolymers* **2013**, *99*, 978−988.

(33) Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. Assessment of standard force field models against high-quality ab initio potential curves for prototypes of π-π, CH/π, and SH/π interactions. *J. Comput. Chem.* **2009**, *30*, 2187−2193.

(34) Bartocci, A.; Belpassi, L.; Cappelletti, D.; Falcinelli, S.; Grandinetti, F.; Tarantelli, F.; Pirani, F. Catching the role of anisotropic electronic distribution and charge transfer in halogen bonded complexes of noble gases. *J. Chem. Phys.* **2015**, *142*, 184304.

(35) Rendine, S.; Pieraccini, S.; Forni, A.; Sironi, M. Halogen bonding in ligand−receptor systems in the framework of classical force fields. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19508−19516.

(36) Politzer, P.; Murray, J. S.; Concha, M. C. Sigma-hole bonding between like atoms; a fallacy of atomic charges. *J. Mol. Model.* **2008**, *14*, 659−665.

(37) Cardamone, S.; Hughes, T. J.; Popelier, P. L. a. Multipolar electrostatics. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10367.

(38) Hagler, A. T. Quantum Derivative Fitting and Biomolecular Force Fields: Functional Form, Coupling Terms, Charge Flux, Nonbond Anharmonicity, and Individual Dihedral Potentials. *J. Chem. Theory Comput.* **2015**, *11*, 5555−5572.

(39) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.

(40) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(41) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046−4063.

(42) Day, G. M.; Motherwell, W. D. S.; Jones, W. Beyond the Isotropic Atom Model in Crystal Structure Prediction of Rigid Molecules: Atomic Multipoles versus Point Charges. *Cryst. Growth Des.* **2005**, *5*, 1023−1033.

(43) Day, G. M.; Price, S. L. A Nonempirical Anisotropic Atom-Atom Model Potential for Chlorobenzene Crystals. *J. Am. Chem. Soc.* **2003**, *125*, 16434−16443.

(44) Price, S. L. Computational prediction of organic crystal structures and polymorphism. *Int. Rev. Phys. Chem.* **2008**, *27*, 541−568.

(45) Misquitta, A. J.; Stone, A. J. Ab Initio Atom-Atom Potentials Using CamCASP: Theory and Application to Many-Body Models for the Pyridine Dimer. *J. Chem. Theory Comput.* **2016**, *12*, 4184−4208.

(46) Misquitta, A.; Welch, G.; Stone, A.; Price, S. *Chem. Phys. Lett.* **2008**, *456*, 105−109.

(47) Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. Beyond Born−Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields. *J. Chem. Theory Comput.* **2016**, *12*, 3851−3870.

(48) Misquitta, A. J.; Stone, A. J.; Fazeli, F. Distributed Multipoles from a Robust Basis-Space Implementation of the Iterated Stockholder Atoms Procedure. *J. Chem. Theory Comput.* **2014**, *10*, 5405−5418.

(49) Phipps, M. J. S.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. Energy decomposition analysis approaches and their evaluation on prototypical protein-drug interaction patterns. *Chem. Soc. Rev.* **2015**, *44*, 3177−3211.

(50) Williams, D. E. Representation of the molecular electrostatic potential by atomic multipole and bond dipole models. *J. Comput. Chem.* **1988**, *9*, 745−763.

(51) Chaudret, R.; Gresh, N.; Narth, C.; Lagardere, L.; Darden, T. A.; Cisneros, G. A.; Piquemal, J.-P. S/G-1: An ab Initio Force-Field Blending Frozen Hermite Gaussian Densities and Distributed Multipoles. Proof of Concept and First Applications to Metal Cations. *J. Phys. Chem. A* **2014**, *118*, 7598−7612.

(52) Giese, T. J.; Chen, H.; Dissanayake, T.; Giambasu, G. M.; Heldenbrand, H.; Huang, M.; Kuechler, E. R.; Lee, T.-S.; Panteva, M. T.; Radak, B. K.; York, D. M. A Variational Linear-Scaling Framework to Build Practical, Efficient Next-Generation Orbital-Based Quantum Force Fields. *J. Chem. Theory Comput.* **2013**, *9*, 1417−1427.

(53) Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. Generalization of the Gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods. *J. Chem. Phys.* **2006**, *125*, 184101.

(54) Elking, D. M.; Cisneros, G. A.; Piquemal, J. P.; Darden, T. A.; Pedersen, L. G. Gaussian multipole model (GMM). *J. Chem. Theory Comput.* **2010**, *6*, 190−202.

(55) Dixon, R. W.; Kollman, P. a. Advancing beyond the atom-centered model in additive and nonadditive molecular mechanics. *J. Comput. Chem.* **1997**, *18*, 1632−1646.

(56) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B. Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2006**, *2*, 1587−1597.

(57) Chaudret, R.; Gresh, N.; Cisneros, G. A.; Scemama, A.; Piquemal, J.-P. Further refinements of next-generation force fields Nonempirical localization of off-centered points in molecules. *Can. J. Chem.* **2013**, *91*, 804−810.

(58) Mu, X.; Wang, Q.; Wang, L.-P.; Fried, S. D.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. Modeling Organochlorine Compounds and the σ-Hole Effect Using a Polarizable Multipole Force Field. *J. Phys. Chem. B* **2014**, *118*, 6456−6465.

(59) Wikfeldt, K. T.; Batista, E. R.; Vila, F. D.; Jónsson, H. A transferable $H_2O$ interaction potential based on a single center multipole expansion: SCME. *Phys. Chem. Chem. Phys.* **2013**, *15*, 16542−56.

(60) Piquemal, J. P.; Chelli, R.; Procacci, P.; Gresh, N. Key role of the polarization anisotropy of water in modeling classical polarizable force fields. *J. Phys. Chem. A* **2007**, *111*, 8170−8176.

(61) Loboda, O.; Ingrosso, F.; Ruiz-López, M. F.; Szalewicz, K.; Millot, C. Geometry-dependent distributed polarizability models for the water molecule. *J. Chem. Phys.* **2016**, *144*, 034304.

(62) Misquitta, A. J.; Stone, A. J. Dispersion energies for small organic molecules: first row atoms. *Mol. Phys.* **2008**, *106*, 1631−1643.

(63) Langhoff, P. W.; Gordon, R. G.; Karplus, M. Comparisons of Dispersion Force Bounding Methods with Applications to Anisotropic Interactions. *J. Chem. Phys.* **1971**, *55*, 2126.

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

(65) Krishtal, A.; Vannomeslaeghe, K.; Geldof, D.; Van Alsenoy, C.; Geerlings, P. Importance of anisotropy in the evaluation of dispersion interactions. *Phys. Rev. A: At., Mol., Opt. Phys.* **2011**, *83*, 024501.

(66) Stone, A. J.; Tough, R. Spherical tensor theory of long-range intermolecular forces. *Chem. Phys. Lett.* **1984**, *110*, 123−129.

(67) Stone, A. J. The description of bimolecular potentials, forces and torques: the S and V function expansions. *Mol. Phys.* **1978**, *36*, 241−256.

(68) Stone, A. J. Are halogen bonded structures electrostatically driven? *J. Am. Chem. Soc.* **2013**, *135*, 7005−7009.

(69) Wheatley, R. J.; Price, S. L. An overlap model for estimating the anisotropy of repulsion. *Mol. Phys.* **1990**, *69*, 507−533.

(70) Kita, S.; Noda, K.; Inouye, H. Repulsive potentials for ClR and BrR (R = He, Ne, and Ar) derived from beam experiments. *J. Chem. Phys.* **1976**, *64*, 3446−3449.

(71) Kim, Y. S.; Kim, S. K.; Lee, W. D. Dependence of the closed-shell repulsive interaction on the overlap of the electron densities. *Chem. Phys. Lett.* **1981**, *80*, 574−575.

(72) Nyeland, C.; Toennies, J. P. Modelling of Repulsive Potentials from Atom Charge Density Distributions: Interactions of Inert Gas Atoms. *Chem. Phys. Lett.* **1986**, *127*, 172−177.

(73) Ihm, G.; Cole, M. W.; Toigo, F.; Klein, J. R. Charge-Overlap Model of Physical Interactions and a Combining Rule for Unlike Systems. *Phys. Rev. A: At., Mol., Opt. Phys.* **1990**, *42*, 5244−5252.

(74) Duke, R. E.; Starovoytov, O. N.; Piquemal, J.-p.; Andre, G. GEM*: A Molecular Electronic Density-Based Force Field for Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 1361−1365.

(75) Gavezzotti, A. Calculation of Intermolecular Interaction Energies by Direct Numerical Integration over Electron Densities. 2. An Improved Polarization Model and the Evaluation of Dispersion and Repulsion Energies. *J. Phys. Chem. B* **2003**, *107*, 2344−2353.

(76) Torheyden, M.; Jansen, G. A new potential energy surface for the water dimer obtained from separate fits of ab initio electrostatic, induction, dispersion and exchange energy contributions. *Mol. Phys.* **2006**, *104*, 2101−2138.

(77) Born, M.; Mayer, J. E. Zur Gittertheorie der Ionenkristalle. *Eur. Phys. J. A* **1932**, *75*, 1−18.

(78) Mitchell, J. B. O.; Price, S. L.; Leslie, M.; Buttar, D.; Roberts, R. J. Anisotropic repulsion potentials for cyanuric chloride ($C_3N_3Cl_3$) and their application to modeling the crystal structures of azaaromatic chlorides. *J. Phys. Chem. A* **2001**, *105*, 9961−9971.

(79) Totton, T. S.; Misquitta, A. J.; Kraft, M. A First Principles Development of a General Anisotropic Potential for Polycyclic Aromatic Hydrocarbons. *J. Chem. Theory Comput.* **2010**, *6*, 683−695.

(80) Mitchell, J. B. O.; Price, S. L. A Systematic Nonempirical Method of Deriving Model Intermolecular Potentials for Organic Molecules: Application To Amides. *J. Phys. Chem. A* **2000**, *104*, 10958−10971.

(81) Söderhjelm, P.; Karlström, G.; Ryde, U. Comparison of overlap-based models for approximating the exchange-repulsion energy. *J. Chem. Phys.* **2006**, *124*, 244101.

(82) Lillestolen, T. C.; Wheatley, R. J. Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chem. Commun.* **2008**, *7345*, 5909−5911.

(83) Lillestolen, T. C.; Wheatley, R. J. Atomic charge densities generated using an iterative stockholder procedure. *J. Chem. Phys.* **2009**, *131*, 144101.

(84) Tang, K. T.; Toennies, J. P. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *J. Chem. Phys.* **1984**, *80*, 3726−3741.

(85) Tang, K. T.; Toennies, J. P. The damping function of the van der Waals attraction in the potential between rare gas atoms and metal surfaces. *Surf. Sci.* **1992**, *279*, L203−L206.

(86) Misquitta, A. J.; Stone, A. J.; Price, S. L. Accurate induction energies for small organic molecules. 2. Development and testing of distributed polarizability models against SAPT(DFT) energies. *J. Chem. Theory Comput.* **2008**, *4*, 19−32.

(87) Holt, A.; Karlstrom, G. Inclusion of the quadrupole moment when describing polarization. The effect of the dipole-quadrupole polarizability. *J. Comput. Chem.* **2008**, *29*, 2033−2038.

(88) Holt, A.; Boström, J.; Karlström, G.; Lindh, R. A NEMO Potential that Includes the Dipole−Quadrupole and Quadrupole−Quadrupole Polarizability. *J. Comput. Chem.* **2010**, *31*, 1583−1591.

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

(90) Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. Dispersion energy from density-functional theory description of monomers. *Phys. Rev. Lett.* **2003**, *91*, 033201.

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

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

(93) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Density-Fitting Method in Symmetry-Adapted Perturbation Theory Based on Kohn−Sham Description of Monomers. *J. Chem. Theory Comput.* **2006**, *2*, 400−412.

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

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

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

(97) Jansen, G.; Hesselmann, A.; Williams, H. L.; Chabalowski, C. F. Comment on "Using Kohn-Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions" (multiple letters). *J. Phys. Chem. A* **2001**, *105*, 11156−11158.

(98) McDaniel, J. G.; Schmidt, J. R. Physically-motivated force fields from symmetry-adapted perturbation theory. *J. Phys. Chem. A* **2013**, *117*, 2053−2066.

(99) 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. OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J. Chem. Theory Comput.* **2013**, *9*, 461−469.

(100) Parker, T. M.; Burns, L. a.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140*, 094106.

(101) Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.;

Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.; Palmieri, P.; Peng, D.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M. MOLPRO, version 2012.1, a package of ab initio programs; 2012.

(102) Oakley, M. T.; Wheatley, R. J. Additive and nonadditive models of vapor-liquid equilibrium in $CO_2$ from first principles. *J. Chem. Phys.* **2009**, *130*, 034110.

(103) Korona, T. The effect of local approximations on first-order properties from expectation-value coupled cluster theory. *Theor. Chem. Acc.* **2011**, *129*, 15−30.

(104) Giauque, W. F.; Egan, C. J. Carbon Dioxide. The Heat Capacity and Vapor Pressure of the Solid. The Heat of Sublimation. Thermodynamic and Spectroscopic Values of the Entropy. *J. Chem. Phys.* **1937**, *5*, 45.

(105) Cervinka, C.; Fulem, M. State-of-the-Art Calculations of Sublimation Enthalpies for Selected Molecular Crystals and Their Computational Uncertainty. *J. Chem. Theory Comput.* **2017**, *13*, 2840−2850.

(106) Heit, Y. N.; Nanda, K. D.; Beran, G. J. O. Predicting finite-temperature properties of crystalline carbon dioxide from first principles with quantitative accuracy. *Chem. Sci.* **2016**, *7*, 246−255.

(107) Simon, A.; Peters, K. Single-crystal refinement of the structure of carbon dioxide. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1980**, *36*, 2750−2751.

(108) Span, R.; Wagner, W. A New Equation of State for Carbon Dioxide Covering the Fluid Region from the TriplePoint Temperature to 1100 K at Pressures up to 800 MPa. *J. Phys. Chem. Ref. Data* **1996**, *25*, 1509−1596.

(109) Elrod, M. J.; Saykally, R. J. Many-Body Effects in Intermolecular Forces. *Chem. Rev.* **1994**, *94*, 1975−1997.

(110) Rick, S. W.; Stuart, S. J. Potentials and Algorithms for Incorporating Polarizability in Computer Simulations. *Rev. Comput. Chem.* **2002**, *18*, 89−146.

(111) McDaniel, J. G.; Schmidt, J. R. First-Principles Many-Body Force Fields from the Gas Phase to Liquid: A Universal Approach. *J. Phys. Chem. B* **2014**, *118*, 8042−8053.

(112) Yu, K.; Schmidt, J. R. Many-body effects are essential in a physically motivated $CO_2$ force field. *J. Chem. Phys.* **2012**, *136*, 034503.

(113) Misquitta, A. J.; Stone, A. J. Accurate Induction Energies for Small Organic Molecules: 1. Theory. *J. Chem. Theory Comput.* **2008**, *4*, 7−18.

(114) Liu, C.; Qi, R.; Wang, Q.; Piquemal, J.-P.; Ren, P. Capturing Many-Body Interactions with Classical Dipole Induction Models. *J. Chem. Theory Comput.* **2017**, *13*, 2751−2761.

(115) Misquitta, A. J. Charge Transfer from Regularized Symmetry-Adapted Perturbation Theory. *J. Chem. Theory Comput.* **2013**, *9*, 5313−5326.

(116) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341−350.

(117) Hawkins, D. M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1−12.

(118) Knizia, G.; Adler, T. B.; Werner, H.-J. Simplified CCSD(T)-F12 methods: Theory and benchmarks. *J. Chem. Phys.* **2009**, *130*, 054104.

(119) Kalugina, Y. N.; Buryak, I. A.; Ajili, Y.; Vigasin, A. A.; Jaidane, N. E.; Hochlaf, M. Explicit correlation treatment of the potential energy surface of $CO_2$ dimer. *J. Chem. Phys.* **2014**, *140*, 234310.

(120) Bukowski, R.; Sadlej, J.; Jeziorski, B.; Jankowski, P.; Szalewicz, K.; Kucharski, S. A.; Williams, H. L.; Rice, B. M. Intermolecular potential of carbon dioxide dimer from symmetry-adapted perturbation theory. *J. Chem. Phys.* **1999**, *110*, 3785−3803.

(121) Babin, V.; Leforestier, C.; Paesani, F. Development of a First Principles Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. *J. Chem. Theory Comput.* **2013**, *9*, 5395−5403.

(122) Duška, M.; Hrubý, J. Re-evaluation of experimental data on the second virial coefficient for steam and development of its analytical representation as a function of the internal energy. *EPJ Web Conf.* **2013**, *45*, 01024.

(123) Tillner-Roth, R.; Harms-Watzenberg, F.; Baehr, H. D. Eine neue fundamentalgleichung für ammoniak. In *Proceedings of the 20th DKV—Tagungsbericht*, Heidelberg, Germany, 1993; Vol. II, p 67.

(124) Massucci, M.; Wormald, C. Second virial coefficients of chloromethane and chloroethane from measurements of the excess enthalpy of $(0.5N_2+ 0.5CH_3Cl)(g)$, and $(0.5N_2+ C_2H_5Cl)(g)$. The excess enthalpy of $(0.5CH_3Cl + 0.5C_2H_5Cl)(g)$. *J. Chem. Thermodyn.* **1998**, *30*, 919−927.

(125) Hellmann, R. Nonadditive three-body potential and third to eighth virial coefficients of carbon dioxide. *J. Chem. Phys.* **2017**, *146*, 054302.

(126) Desgranges, C.; Delhommelle, J. Many-Body Effects on the Thermodynamics of Fluids, Mixtures, and Nanoconfined Fluids. *J. Chem. Theory Comput.* **2015**, *11*, 5401−5414.

(127) Yu, K.; Kiesling, K.; Schmidt, J. R. Trace Flue Gas Contaminants Poison Coordinatively Unsaturated Metal Organic Frameworks: Implications for $CO_2$ Adsorption and Separation. *J. Phys. Chem. C* **2012**, *116*, 20480−20488.

(128) Pérez-Sánchez, G.; González-Salgado, D.; Piñeiro, M. M.; Vega, C. Fluid-solid equilibrium of carbon dioxide as obtained from computer simulations of several popular potential models: The role of the quadrupole. *J. Chem. Phys.* **2013**, *138*, 084506.

(129) Millot, C.; Stone, A. Towards an accurate intermolecular potential for water. *Mol. Phys.* **1992**, *77*, 439−462.

(130) Donchev, A. G.; Ozrin, V.D.; Subbotin, M. V.; Tarasov, O. V.; Tarasov, V. I. A quantum mechanical polarizable force field for biomolecular interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7829−7834.

(131) Khoruzhii, O.; Butin, O.; Illarionov, A.; Leontyev, I.; Olevanov, M.; Ozrin, V.; Pereyaslavets, L.; Fain, B. Polarizable Force Fields for Proteins. In *Protein Modelling*; Springer International Publishing, Cham, Switzerland, 2014; pp 91−134.

(132) Donchev, A. G.; Galkin, N. G.; Illarionov, A. A.; Khoruzhii, O. V.; Olevanov, M. A.; Ozrin, V. D.; Subbotin, M. V.; Tarasov, V. I. Water properties from first principles: Simulations by a general-purpose quantum mechanical polarizable force field. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8613−8617.

(133) Donchev, A. G.; Galkin, N. G.; Pereyaslavets, L. B.; Tarasov, V. I. Quantum mechanical polarizable force field (QMPFF3): Refinement and validation of the dispersion interaction for aromatic carbon. *J. Chem. Phys.* **2006**, *125*, 244107.

(134) Kamath, G.; Kurnikov, I.; Fain, B.; Leontyev, I.; Illarionov, A.; Butin, O.; Olevanov, M.; Pereyaslavets, L. Prediction of cyclohexane-water distribution coefficient for SAMPL5 drug-like compounds with the QMPFF3 and ARROW polarizable force fields. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 977−988.

(135) Donchev, A. G.; Galkin, N. G.; Illarionov, A. A.; Khoruzhii, O. V.; Olevanov, M. A.; Ozrin, V. D.; Pereyaslavets, L. B.; Tarasov, V. I. J. Comput. Chem. *J. Comput. Chem.* **2008**, *29*, 1242−1249.

(136) Khoruzhii, O.; Donchev, A. G.; Galkin, N.; Illarionov, A.; Olevanov, M.; Ozrin, V.; Queen, C.; Tarasov, V. Application of a polarizable force field to calculations of relative protein-ligand binding affinities. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10378−10383.

## ■ NOTE ADDED IN PROOF

While in proof, the authors became aware of the QMPFF[130,131] family of atomically-anisotropic, general-purpose force fields. Importantly, QMPFF2[132] incorporates anisotropy into the description of long-range induction, and models anisotropic electrostatic and exchange interactions via multiple independent shifting spherical clouds per site. The more recent QMPFF3[133] and ARROW force fields[134] build on QMPFF2 by instead describing the anisotropy of electrostatics and exchange via a multipole expansion, somewhat similar in spirit

to the present work. The QMPFF2, QMPFF3, and ARROW force fields have all been employed in MD simulations of condensed phases, with promising results for both small molecules and proteins compared to traditional non-polarizable force fields.[132,134−136]