Fundamenteel inzicht in niet-covalente interacties beschreven door krachtvelden: toepassing op gasadsorptie in metaal-organische roosters

Understanding Noncovalent Interactions in Force Fields through Quantum Mechanics: Application to Gas Adsorption in Metal-Organic Frameworks

Steven Vandenbrande

Promotoren: prof. dr. ir. V. Van Speybroeck, prof. dr. ir. T. Verstraelen
Proefschrift ingediend tot het behalen van de graad van
Doctor in de ingenieurswetenschappen: toegepaste natuurkunde

UNIVERSITEIT
GENT

Vakgroep Toegepaste Fysica
Voorzitter: prof. dr. ir. C. Leys
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2018 - 2019

ISBN 978-94-6355-224-0
NUR 928
Wettelijk depot: D/2019/10.500/32

## Members of the Examination Committee

**Chair**

prof. dr. ir. Filip De Turck (Ghent University)

**Examination Board**

prof. dr. Alexandre Tkatchenko (University of Luxembourg)
dr. David Dubbeldam (Universiteit van Amsterdam)
prof. dr. ir. Jutho Haegeman (Ghent University)
dr. ir. Louis Vanduyfhuys (Ghent University)
prof. dr. ir. Christophe Leys (Ghent University)
em. prof. dr. Michel Waroquier (Ghent University)

**Supervisors**

prof. dr. ir. Veronique Van Speybroeck (Ghent University)
prof. dr. ir. Toon Verstraelen (Ghent University)

This research has been conducted at the **Center for Molecular Modeling**

# Preface

I would like to thank Veronique for providing me the opportunity to start a Ph.D., for pulling me through a difficult patch in the middle, and for being persistent to ensure that this dissertation could come about as the end product.

I would like to thank Toon for giving me insight into his perspective on science, for his endless stream of ideas and his practical guidance.

I would like to thank all the people who made my days at the CMM more cheerful. In particular Sam, An, Sven, Jelle, and Arthur: it was a pleasure sharing an office with you.

Steven Vandenbrande
Ghent, April 2019

# Contents

# List of Symbols

| | |
|---|---|
| $\alpha$ | Adsorption selectivity |
| $B_2$ | Second virial coefficient |
| $\beta$ | Inverse of thermodynamic temperature ($\beta = \frac{1}{k_B T}$) |
| $E_{\mathrm{disp}}$ | Dispersion energy |
| $E_{\mathrm{elst}}$ | Electrostatic energy |
| $E_{\mathrm{exch-rep}}$ | Exchange-repulsion energy |
| $E_{\mathrm{ind,ct}}$ | Charge-transfer contribution to induction energy |
| $k_B$ | Boltzmann constant |
| $K_H$ | Henry coefficient |
| $\Delta H_{\mathrm{ads}}$ | Adsorption enthalpy |
| $N_{A_i}$ | MBIS electron population of shell $i$ of atom $A$ |
| $N_v$ | MBIS electron population of valence (outer) shell |
| $P$ | External gas pressure |
| $q$ | Electric charge |
| $\boldsymbol{r}, \boldsymbol{R}$ | Position vectors in Cartesian space |
| $r_{ij}, R_{AB}$ | Distances between particles |
| $\rho$ | Mass density |
| $\rho(\mathbf{r})$ | Electron density at point $\mathbf{r}$ |
| $\rho^0(\mathbf{r})$ | Pro-molecular density at point $\mathbf{r}$ |
| $\rho_A(\mathbf{r})$ | Electron density attributed to nucleus A at point $\mathbf{r}$ |
| $\rho_A^0(\mathbf{r})$ | Pro-atomic density of nucleus A at point $\mathbf{r}$ |
| $S$ | Density overlap |
| $\sigma_{A_i}$ | MBIS Slater width of shell $i$ of atom $A$ |
| $\sigma_v$ | MBIS Slater width of valence (outer) shell |
| $T$ | Temperature |
| $U$ | Potential energy |

# List of Abbreviations

| | |
|---|---|
| **AIM** | Atoms in molecules |
| **ATM** | Axilrod-Teller-Muto |
| **CBS** | Complete basis set |
| **CCSD(T)** | Coupled-cluster with single, double and (perturbatively) triple excitations |
| **CSOV** | Constrained space orbital variation |
| **DFT** | Density functional theory |
| **ESP** | Electrostatic potential |
| **GCMC** | Grand canonical Monte Carlo |
| **HF** | Hartree-Fock |
| **HI** | Iterative Hirsfeld |
| **ISA** | Iterated stockholder atoms |
| **MBD** | Many-body dispersion |
| **MBIS** | Minimal basis iterative stockholder |
| **MC** | Monte Carlo |
| **MD** | Molecular dynamics |
| **MEDFF** | Monomer electron density force field |
| **MOF** | Metal–organic framework |
| **NQEs** | Nuclear quantum effects |
| **OPLS** | Optimized Potentials for Liquid Simulations |
| **PES** | Potential energy surface |
| **QTAIM** | Quantum theory of atoms in molecules |
| **SAPT** | Symmetry-adapted perturbation theory |
| **TraPPE** | Transferable potentials for phase equilibria |
| **UFF** | Universal force field |

# Samenvatting

Moleculaire modellering is een relatief recente tak van de wetenschap. Simulaties op de moleculaire schaal werden immers pas mogelijk na de ontwikkeling van digitale computers en nog steeds is het vakgebied nauw verbonden met ontwikkelingen in de computerwetenschappen. De theoretische fundamenten worden gevormd door een ontzettend belangrijk stukje kennis dat we als mensheid vergaard hebben: de atomaire hypothese, die stelt dat alle materie rondom ons bestaat uit onderling interagerende deeltjes die constant in beweging zijn. Het basisidee van moleculaire modellering is dat door het gedrag van deze atomen te simuleren in virtuele experimenten, we de eigenschappen van materialen op een fundamenteel niveau kunnen bestuderen. Eén van de grote uitdagingen hierbij is het vinden van een evenwicht tussen de precisie waarmee de atomaire interacties beschreven worden enerzijds, en de bijhorende computationale kostprijs anderzijds. In zeker opzicht is het verrassend dat, ondanks de exponentiële groei van computerkracht, het momenteel nog steeds onmogelijk is om de Schrödingervergelijking exact op te lossen voor chemisch interessante systemen. De verklaring hiervoor is dat ook de complexiteit van kwantummechanische problemen exponentieel toeneemt naarmate meer en meer deeltjes in rekening worden gebracht.

De zoektocht naar nieuwe methodes die een accurate beschrijving van atomen geven én tegelijkertijd snel berekend kunnen worden door een computer, is volop bezig. Hierbij is het belangrijk om een onderscheid te maken tussen zogenaamde ab initio methoden en krachtvelden. De ab initio aanpak houdt in dat er vertrokken wordt van de fundamentele vergelijkingen van de fysica. Meestal worden zo nauwkeurige resultaten bekomen worden, weliswaar aan een hoge computationale kostprijs. Diagonaal tegenover de ab initio aanpak, staan methoden gebaseerd op krachtvelden. Krachtvelden beschrijven elektronen niet expliciet, wat ertoe leidt dat alle berekeningen zeer snel uitgevoerd kunnen worden. Spijtig genoeg houdt dit meestal in dat er slechts een zeer benaderende beschrijving van de interacties gegeven wordt. In dit proefschrift wordt een aanpak onderzocht die het beste van de

twee werelden probeert te verenigen: de ontwikkeling van een krachtveld dat gebaseerd is op ab initio methoden.

Het vertrekpunt voor dit nieuwe krachtveld is de ontwikkeling van een methode die de elektronendichtheid partitioneert in atomaire bijdragen. Er zijn gelijkenissen met bestaande partitioneringsmethoden zoals Iterative Stockholder, maar nu worden de pro-atomen geschreven als een lineaire combinatie van eenvoudige Slater functies. Het nieuwe partitioneringsschema wordt aangeduid met MBIS, afkorting van de Engelstalige benaming Minimal Basis Iterative Stockholder. MBIS geeft een beknopte maar betrouwbare beschrijving van de elektronendichtheid. Het is precies de complexiteit van de volledige beschrijving van de elektronendichtheid die ertoe leidt dat ab initio methoden een hoge computationale kostprijs hebben. De eenvoudige MBIS representatie van de elektronendichtheid, die kan samengevat worden met slechts enkele parameters voor elk atoom, is dus een goede basis voor het krachtveld dat verder ontwikkeld zal worden.

Een hele waaier aan krachtvelden is ontwikkeld in de voorbije decennia. De keuze voor een bepaald krachtveld dat geschikt is voor een bepaald toepassing is helemaal niet voor de hand liggend. Een krachtveld is meestal ontwikkeld en getest voor één specifieke toepassing. Hetzelfde krachtveld gebruiken voor een andere toepassing houdt meestal in dat er een hele reeks aanpassingen moeten gebeuren, die allemaal grondig getest (zouden) moeten worden. De reden hiervoor is dat veel krachtvelden afgestemd worden op het reproduceren van enkele specifieke eigenschappen, zonder dat het onderliggende mechanisme correct beschreven wordt. Deze waarneming heeft geleid tot een aantal basisprincipes die we opleggen aan de nieuwe methode voor de ontwikkeling van een krachtveld. Deze basisprincipes (onder andere een fysische onderbouw voor alle termen voorzien, het vermijden van overfitting en het gebruik van de meest accurate ab initio methoden in de parametrisatie) worden later uitgebreid besproken.

In deze thesis werd een nieuwe methode ontwikkeld voor de beschrijving van niet-covalente interacties met een krachtveld. Dit krachtveld kreeg de naam MEDFF, de afkorting van Monomer Electron Density Force Field. Om de vooropgestelde uitdrukking voor de energie te rechtvaardigen en om het (kleine aantal) parameters vast te leggen, wordt de ab initio energie eerst ontbonden in componenten met een fysische betekenis. Als eindresultaat bekomen we een krachtveld dat een correcte beschrijving geeft van dispersiegedomineerde complexen in de gasfase, zonder MEDFF expliciet af te stemmen op de molecules die onderzocht worden. Simulaties van dezelfde moleculen in de vloeistoffase levert niet steeds correcte resultaten op, wat kan verklaard worden door onder andere het ontbreken van veeldeeltjesdispersieinteracties en het kwantumgedrag van atoomkernen. Deze twee

effecten worden standaard verwaarloosd in krachtveldsimulaties, en op die manier helpt MEDFF in het verkrijgen van inzicht in de redenen waarom een krachtveld soms wel en soms niet het gewenste resultaat oplevert.

Daarna wordt het nieuwe niet-covalent krachtveld MEDFF gebruikt in de context van een recente, praktische toepassing: de adsorptie van gastmoleculen in poreuze materialen. Specifiek worden interacties tussen $CH_4$, $CO_2$, $N_2$ en verschillende metaal-organische roosters bestudeerd. Hierbij wordt aangetoond dat MEDFF een betere overeenkomst vertoont met de ab initio adsorptie-energie dan andere krachtvelden die momenteel beschikbaar zijn in de literatuur, opnieuw zonder dat MEDFF specifiek moet aangepast worden voor deze systemen. Daarnaast wordt ook duidelijk dat een systematische overeenkomst tussen gesimuleerde en experimentele adsorptie-isothermen zeer uitdagend blijft.

MEDFF wordt tenslotte ook gebruikt om simulaties van gasadsorptie in metaal-organische roosters te bekijken vanuit een ander perspectief. Een variant van het bestaande importance sampling schema wordt ontwikkeld. Hierbij wordt een krachtveld gebruikt als hulpmiddel voor de ab initio bepaling van de Henry coëfficiënt en de adsorptie-enthalpie. In een eerste stap van dit schema wordt een krachtveld gebruikt om de meest interessante adsorptiesites te identificeren. Daarna worden ab initio berekeningen uitgevoerd voor deze sites, wat toelaat om accurate waarden voor de Henry coëfficiënt en de adsorptie-enthalpie te vinden. Dit onderzoek onthult niet alleen verschillen tussen generieke krachtvelden en ab initio methoden, maar werpt ook licht op het belang van dispersiecorrecties in dichtheidsfunctionaaltheorie. Op deze manier kan dit onderzoek opnieuw beschouwd worden als een voorbeeld van hoe MEDFF toelaat om fundamenteel inzicht in niet-covalente interacties te bekomen, ook voor toepassingen die relevant zijn voor hedendaagse problematieken.

Samenvattend kan gesteld worden dat het centrale onderdeel van dit proefschrift de ontwikkeling van een nieuwe methode voor niet-covalente krachtvelden is: MEDFF. Enkele belangrijke kenmerken van MEDFF zijn het gebruik van een decompositieschema voor de ab initio energie, een inzichtelijke parametrisatiemethode, en het gebruik van de meest accurate ab initio methoden voor het genereren van referentiedata. Er wordt aangetoond dat het resulterende krachtveld robuust is en gebruikt kan worden voor de beschrijving van gasadsorptie in nanoporeuze materialen, een toepassingsgebied dat niet gebruikt hoefde te worden bij de constructie van MEDFF. Er kan dus besloten worden dat MEDFF een handig hulpmiddel is om inzicht in niet-covalente interacties te verkrijgen, ook voor industrieel relevante toepassingen zoals de opslag van $CO_2$ in metaal-organische roosters, en dat MEDFF toekomstige ontwikkelingen in dit gebied kan ondersteunen.

# Summary

Molecular modeling is a fairly young branch of science, owing to its reliance on and close connection to developments in digital computers. Its theoretical foundation is one of the most important pieces of human knowledge: the atomic hypothesis, stating that everything around us consists of perpetually moving particles interacting with each other. Molecular modeling enables the study of the behavior of materials at a very fundamental level by performing virtual experiments on model systems. One of the greatest challenges, present ever since the first simulations, is striking the balance between the accuracy of the atomic interactions on the one hand and the associated computational cost on the other hand. In fact it is in a way very surprising that, despite the exponential growth in computing power, the exact solution of the Schrödinger equation for chemically relevant systems is still far out of reach. This is ultimately due to the complexity of quantum mechanical problems, which also rises exponentially as the number of particles is increased.

The quest for methods that allow for an accurate yet computationally feasible description of interatomic interactions is still ongoing. A rough distinction can be made between electronic structure approaches and force-field based methods. The former are developed from theoretical considerations and usually deemed accurate but involving a high computational cost. Force fields, which generally lack an explicit description of the electronic structure, are comparatively very simple from a computational point of view but considered to provide only a very approximate description of the interactions. In this dissertation, an approach is investigated which aims to combine the best of both worlds: the development of a force field that is rooted in ab initio methods.

The first cornerstone of the newly developed force-field methodology is a novel procedure that allows a partitioning of the electron density into atomic contributions. It shows similarities to the existing Iterative Stockholder approach, but here the pro-atoms are expanded in a small basis of Slater functions. The resulting partitioning procedure is therefore named Minimal

Basis Iterative Stockholder. The end result is a concise but faithful representation of the electron density, whose complexity plays a determining role in the computational cost associated with ab initio methods. The concise representation, involving only a few parameters for each atom, will be the foundation on which the force field will be built.

Many different force fields have been developed over the years, yet the selection of the appropriate method for the application of interest is far from trivial. When an existing force field is employed for simulations other than those for which it was specifically designed, it is common that a daunting and laborious cycle of testing and reparametrization has to be iterated. The reason for this is that many existing methodologies are fitted to reproduce a certain quantity, rather than providing a proper description of the underlying mechanisms. This served as inspiration for the guiding principles of the novel method to derived force fields that will be explained later. Aiming to ensure robustness, these guiding principles include a clear physical interpretation of each energy contribution, avoiding ill-conditioned fitting of parameters and the ability to use high-level ab initio methods to generate reference data, while at the same time keeping the approach as simple and transparent as possible.

In this dissertation, a new force field methodology for the description of noncovalent interactions was developed. It will be referred to as MEDFF, the abbreviation of Monomer Electron Density Force Field. MEDFF employs an ab initio energy decomposition scheme, both to justify the proposed potential energy expression and to calibrate the (very limited number of) associated parameters. The end result is a potential that succeeds in properly describing dispersion-dominated complexes in the gas phase, without the need to explicitly fit to the complexes of interest. Simulations performed for similar systems in the condensed phase are less satisfying, but discrepancies can be explained by incorporating effects such as many-body dispersion and the quantum nature of nuclei, which are commonly not explicitly accounted for in current force fields. Ultimately, this leads to a better comprehension of the reasons why a force field performs well or badly.

Turning to more practical applications, the noncovalent force field MEDFF is then used to investigate a currently highly relevant topic in materials science: the adsorption of small guest molecules in porous materials. More specifically, interactions between $CH_4$, $CO_2$, $N_2$ and various metal–organic frameworks, which are often porous, are considered. These interactions are crucial for the potential use of metal–organic frameworks for carbon capture and storage or for the purification of natural gas. It turns out that the newly developed force field surpasses most other force fields when it comes to predicting ab initio adsorption energies for these systems. At the same time

it is shown that a consistent agreement between simulated and experimental isotherms remains very challenging for the systems under consideration.

Finally the same subject, namely simulation of gas adsorption in metal–organic frameworks, is considered from a different point of view. A variant of importance sampling is developed which employs force fields to facilitate the ab initio calculation of properties such as the Henry coefficient or heat of adsorption. Initially, a force field is used as a screening tool in order to sample favorable adsorption configurations. After that, more accurate ab initio calculations are performed for these configurations to obtain refined values for the Henry coefficient and heat of adsorption. This research reveals interesting differences not only between generic force fields and ab initio methods, but also highlights for instance the importance of the choice of dispersion corrections in density functional theory methods. In this way, it provides another example of how the MEDFF force field developed in this dissertation is a tool that allows gaining fundamental insight into noncovalent interactions, also for real-life applications of interest.

In conclusion, this Ph.D. dissertation is centered on the development of a novel method to construct noncovalent force fields, MEDFF. Some characteristics that distinguish the MEDFF approach from existing force-field development methodologies, are the employment of an ab initio decomposition scheme, fitting a minimal number of parameters and usage of high-level ab initio methods for parametrization. It is shown that this leads to a robust force field, which can be used to describe gas adsorption in nanoporous materials, a domain outside of the data set for which MEDFF was originally conceived. The current work thus provides additional insight into noncovalent interactions for systems which hold great promise for industrial applications (for instance involving gas separation by metal–organic frameworks), and might be a valuable tool for future developments in this area.

# Part I

# Understanding Noncovalent Interactions in Force Fields through Quantum Mechanics: Application to Gas Adsorption in Metal–Organic Frameworks

# 1

# Introduction

*One does not become a theorist by buying chalk,*
*an experimentalist by buying a microscope*
*or a computational scientist by downloading software.*
Mikko Karttunen, 2016

It goes without saying that the discovery of new materials has been of paramount importance in the evolution of mankind. Ranging from polymers, over semiconductors to quantum dots, the recent successes of materials science are all around. A crucial milestone of the past decades was the advent of High Performance Supercomputers, which allowed unseen progress in the field of computational molecular modeling. This is precisely the field of the current thesis, which starts from the fundamental equations governing the behavior of atoms and ends with computer simulations at the atomic scale of real-life applications. Therefore, a general introduction to computational molecular modeling is provided in Section 1.1. The main message is that even though the fundamental laws are well-known, exact solutions of the resulting equations are out of reach. This implies that appropriate approximations need to be introduced, striking a balance between correctness and computational tractability. The balance of these apparent antipodes will be a common thread throughout all upcoming chapters. Some approximations that are currently used are briefly discussed, one of them being force fields, which is an approach that enables simulations on the microscale while still providing atomistic detail.

If Section 1.1 is concerned with setting the scene, Section 1.2 is about meeting the players: an overview of currently available force fields is given. They are classified according to their relevant properties, such as functional form, parameter determination and coverage of chemical space. The strengths and weaknesses of each class are discussed, and this allows a demonstration of the opportunities for further force-field development.

Finally, in Section 1.3 those opportunities are used to draw the blueprint for a novel ab initio derived noncovalent force field, which is the main methodological innovation presented in further chapters of this thesis. The introductory chapter ends with an outline of this dissertation.

## 1.1 Molecular modeling

Physics is the branch of science that tries to explain the fundamental mechanisms governing the behavior of matter and energy. A long-standing dream of many physicists is the discovery of a simple set of laws that would allow a description of the entire universe. Such a Theory of Everything is very unlikely to be available in the near future, let alone its application to practical problems on earth. All current physical models describing the world around us are thus approximations, and this thesis is rooted in one of the most successful of those approximate theories: quantum mechanics. Quantum mechanics provides a description at the atomic level and an important assumption is that all matter is built up of electrons and nuclei as fundamental building blocks. Choosing an appropriate level of detail is critical in any physical model of the world, and the atomic level is a splendid choice for the understanding and design of novel materials. This subbranch of materials science can be termed *molecular modeling*.

Section 1.1.1 reviews the basic concepts underlying quantum mechanics, as well as some of its applications, in particular those belonging to quantum chemistry or molecular quantum mechanics. The resulting theoretical framework has very solid foundations, but has a limited practical applicability because of the complexity of the encountered equations. Luckily, many derived models are available which allow exploiting the computational power of current computers in order to apply the theory to real-world problems. A broad overview of such models aiming to approximately satisfy quantum mechanical equations is given in Section 1.1.2.

### 1.1.1 From the Schrödinger equation to molecular mechanics

Quantum mechanics is a branch of physics describing nature at the microscopic level, at the atomic or subatomic scale. Its origins can be traced back to the beginning of the twentieth century, when some famous physicists explained experimental phenomena using models that were incompatible with theories existing at that point in time. Examples include the ultraviolet catastrophe (Max Planck, 1900), the photoelectric effect (Albert Einstein, 1905) and the particle nature of electromagnetic radiation (Arthur Compton, 1923). In the mid-1920s, these preliminary efforts were formalized into a theory, nowadays referred to as *standard* quantum mechanics, thanks to contributions of Niels Bohr, Erwin Schrödinger, Werner Heisenberg, Paul Dirac, and others. Afterwards, the theory was extended to treat relativistic effects, which resulted in the Dirac equation, the Klein-Gordon equation and ultimately in the development of quantum electrodynamics, which gives a complete description of the interactions between matter and light. This paved the way for quantum field theory, of which the Standard Model is an example that provides the most complete picture of fundamental forces (eletromagnetic, weak and strong interactions) to date (the exclusion of gravitational force however prevents the Standard Model from being a candidate Theory of Everything). The latter developments are only relevant at the subatomic scale and not considered in this work. Neither do relativistic effects come into play for the systems studied here. This is why the following theoretical introduction is limited to *standard* quantum mechanics, considering electrons and nuclei as basic building blocks without further fine structure.

#### The Schrödinger equation

One of the radical paradigm shifts in quantum mechanics is the abolishment of the deterministic universe. In classical physics, given the initial state of a system, one can compute all measurable properties of this system with certainty for any point in time. In quantum mechanics, measurements disturb the system (this is related to Heisenberg's uncertainty principle) and in general only the *probability* of a certain measurement outcome can be predicted. To represent this mathematically, a quantum system is described by its wave function $|\Psi(t)\rangle$, which is a complex-valued vector in a Hilbert space that encodes all information about the quantum system at a time $t$. A single, spinless particle is a simple example of a quantum system and its wave function can be projected to coordinate space. The wave function is then written as a complex-valued function in Cartesian coordinates $\Psi(\mathbf{r};t)$, with

$|\Psi(\mathbf{r};t)|^2$ (which is real-valued) acting as the probability density function of finding the particle at position $\mathbf{r}$ at a certain time $t$.

Perhaps the most important equation in quantum mechanics is the Schrödinger equation:[1]

$$\hat{H}|\Psi(t)\rangle = i\hbar\frac{\partial}{\partial t}|\Psi(t)\rangle \tag{1.1}$$

It is a partial differential equation that describes the time evolution of the wave function and can be considered the quantum-mechanical counterpart of Newton's second law of motion in classical mechanics. Despite its appealing simplicity, the Schrödinger equation is in general very hard to solve. Much of the complexity lays hidden in the expression for the Hamiltonian operator $\hat{H}$, which is the operator corresponding to the total energy of the system. In case the Hamiltonian does not depend explicitly on the time $t$, a separation of variables can be applied to the Schrödinger equation 1.1 with $E_n$ as the constant of integration:

$$\hat{H}|\varphi_n\rangle = E_n|\varphi_n\rangle \tag{1.2}$$

$$i\hbar\frac{d}{dt}f(t) = E_n f(t) \tag{1.3}$$

The first equation 1.2 is called the time-independent or stationary Schrödinger equation, which is a differential eigenvalue equation. By applying boundary conditions on the time-independent part of the wave function $|\varphi\rangle$, this equation will only have bound solutions for discrete values $E_n$, which is indicated by the discrete label $n$ in subscript. The time-dependent equation is easily solved to yield for the total wave function:

$$|\Psi(t)\rangle = \sum_n c_n e^{-i\frac{E_n t}{\hbar}}|\varphi_n\rangle \tag{1.4}$$

where the coefficients $c_n$ are determined based on the initial conditions.

The wave function for a set of indistinguishable particles should also satisfy symmetry constraints. For fermions (half-integer spin particles, such as electrons) the wave function has to be anti-symmetric with respect to swapping the position of two particles. This leads to the Pauli exclusion principle, which says that two fermions can never be in the same state simultaneously.

### The molecular Hamiltonian and the Born-Oppenheimer approximation

A molecular system is a collection of $N_e$ electrons (fermions with spin $1/2$ and negative unit charge) and $N_N$ atomic nuclei (spin and charge depend on

the number of protons and neutrons in the nucleus). Without external perturbations, the Hamiltonian for this system consists of five contributions: the nuclear kinetic energy $\hat{T}_N$, the electronic kinetic energy $\hat{T}_e$ and the Coulomb interaction between nuclei-nuclei ($\hat{V}_{NN}$), between electrons-electrons ($\hat{V}_{ee}$) and between nuclei-electrons ($\hat{V}_{Ne}$). These contributions are expressed as follows (in atomic units):[2]

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{NN} + \hat{V}_{ee} + \hat{V}_{Ne} \tag{1.5}$$

$$\hat{T}_N = -\sum_A \frac{1}{2M_A}\nabla_A^2 \tag{1.6}$$

$$\hat{T}_e = -\sum_i \frac{1}{2}\nabla_i^2 \tag{1.7}$$

$$\hat{V}_{NN} = +\sum_{A>B} \frac{Z_A Z_B}{R_{AB}} \tag{1.8}$$

$$\hat{V}_{ee} = +\sum_{i>j} \frac{1}{r_{ij}} \tag{1.9}$$

$$\hat{V}_{Ne} = -\sum_A \sum_i \frac{Z_A}{r_{Ai}} \tag{1.10}$$

where $M_A$ and $Z_A$ represent respectively the mass and charge of nucleus $A$ and $R_{AB}$, $r_{ij}$ and $r_{Ai}$ denote the distance between the different particles. Note that this Hamiltonian does not show any explicit time dependence, so it suffices to solve the stationary Schrödinger equation 1.2.

The molecular Hamiltonian appears to show symmetry in the appearance of terms related to electrons and terms related to nuclei, but there is an important "imbalance" between the two. Because nuclei are much heavier than electrons (the lightest element, hydrogen, has a nuclear mass that is $1836$ times the electron mass), the dynamics of both particles happen on different time scales. In the Born-Oppenheimer approximation[3] it is therefore assumed that the nuclei remain fixed with respect to the motion of the electrons. This assumption allows treating the nuclear coordinates (represented by the vector $\mathbf{R}$) at first instance as parameters, and by additionally assuming that excited electronic states show a much higher energy than nuclear excitations, the total wave function can be written as a product of a nuclear wave function $|\chi\rangle$ and an electronic wave function $|\Psi_e\rangle$. The latter obeys the following equation:

$$\hat{H}^e|\Psi^e\rangle = E_n^e|\Psi^e\rangle \tag{1.11}$$

$$\hat{H}^e = \hat{T}_e + \hat{V}_{NN} + \hat{V}_{ee} + \hat{V}_{Ne} \tag{1.12}$$

Note that here $\hat{V}_{NN}$ is simply a constant term because the nuclei are assumed to be clamped to their positions. Next, still within the Born-Oppenheimer or adiabatic approximation, the nuclear part of the wave function is found by solving:

$$\left[\hat{T}_N + E_n^e\left(\mathbf{R}\right)\right]|\chi_{nm}\rangle = E_{nm}|\chi_{nm}\rangle \tag{1.13}$$

Note that $E_n^e\left(\mathbf{R}\right)$ now has to be considered as a function of the nuclear positions.

### Molecular dynamics

Equation 1.13 shows that, within the Born-Oppenheimer approximation, the nuclei obey a Schrödinger equation where the potential energy is generated by the electrons as described in Equation 1.11. For the study of extended molecular systems, necessary in order to obtain thermodynamic properties, the model is further simplified by neglecting the quantum mechanical nature of the nuclei. In this case the nuclei move on the Born-Oppenheimer potential energy surface (PES) $E_n^e\left(\mathbf{R}\right)$ and their positions satisfy Newton's second law of motion:

$$M_A\frac{d^2\mathbf{R}_A}{dt^2} = -\nabla_A E_n^e\left(\mathbf{R}\right) \tag{1.14}$$

Simulations that solve the resulting set of equations are called Molecular Dynamics (MD). Using statistical mechanics,[4] values for macroscopic observables can be extracted by taking appropriate averages during MD simulations. Because the thermodynamic limit is out of reach for any computer (the number of nuclei considered in the simulation would have to be on the order of Avogadro's number), the set of equations 1.14 has to be extended in order to sample the appropriate ensemble. For instance, the temperature of the system can be controlled by thermostatting algorithms and the external pressure by barostatting algorithms. Alternatively to MD simulations, Monte Carlo (MC) methods can also be used to obtain thermodynamic information. In MC simulations, the same PES $E_n^e\left(\mathbf{R}\right)$ is used to sample the phase space according to a specific distribution, but now this is done employing randomly generated nuclear configurations. As a final note, we mention that nuclear quantum effects (NQEs) can be incorporated into MD simulations, by for instance using the path-integral approach.[5]

### 1.1.2 Bridging theory and experiment with ab initio based force fields

The theory exposed in the previous section in principle allows the prediction of macroscopic observables, starting from nothing but the Schrödinger equation, basic properties of electrons and nuclei, and the expression for the molecular Hamiltonian. The direct connection between fundamental postulates from quantum mechanics and experiment is however hampered by the complexity of the electronic Schrödinger equation 1.11. Concerning molecules, only for one-electron hydrogen-like atoms, a complete analytical solution is available. For any chemically relevant system, a direct numerical solution (for instance using a finite element method) would require an enormous computational effort that is out of reach for current hardware. This is why several methods have been developed to compute approximations to the electronic energy $E_n^e(\mathbf{R})$. The most important distinction is the one between so-called ab initio methods and empirical potentials, which is further explained below.

#### Ab initio methods

Ab initio (Latin for "from the beginning") or first-principles methods use only fundamental physical equations and constants as input. Examples include Quantum Monte Carlo, Configuration Interaction, Complete Active Space perturbation theory and Density Functional Theory (DFT). Although some of these approaches (for instance DFT) usually do include some empirical parameters, they will be considered ab initio in this work, because at least conceptually they start from the fundaments of quantum mechanics.

The basis for many first-principles methods is Hartree-Fock (HF),[6] where electrons move in a mean-field potential generated by the other electrons. This self-consistent field approach neglects electron correlation and has therefore been largely superseded by more accurate methods. An example of such a method is coupled-cluster theory,[7] which starts from the HF solution and then computes corrections based on perturbation theory by considering excitations of HF orbitals. In practice, the spatial wave function is expanded in a finite basis set, which for molecular (non-periodic) systems typically consists of Gaussian functions centered on the nuclei. This is also an approximation, which however offers very accurate results if an extrapolation from increasingly larger basis sets is performed in order to estimate the complete basis set (CBS) limit. CCSD(T)/CBS (coupled-cluster with single, double and (perturbative) triple excitations in the complete basis set limit) is often considered the golden standard of quantum chemistry for the description

of closed-shell molecules near equilibrium.[6] It is interesting to note that treating the triple excitations perturbatively often works better than an exact treatment: the reason is that the difference between the two is close to the contribution that would be included by considering quadruple excitations. The computational cost of CCSD(T) formally scales as $\mathcal{O}\left(N_{\text{basis}}^7\right)$ with $N_{\text{basis}}$ the number of basis functions, although this can be improved by techniques such as density fitting.[8] CCSD(T)/CBS can currently be used for complexes containing several tens of atoms.

There is a one-to-one mapping between the ground-state wave function and density of a system of electrons subjected to an external potential.[9] Therefore in DFT, the electron density, which is much less complicated to work with than the wave function, is used as the central quantity.[10] For instance, the total energy of the system is a functional of the electron density. Although DFT provides in principle an exact formalism to solve the Schrödinger equation, practical calculations always involve approximations because an exact representation of the functional is not known. Many functionals exist,[11] which can be determined from physical considerations or parametrized with experimental data. A weakness of many DFT functionals is the lack of long-range dispersion interactions. In recent years, many methods have been developed, aiming to solve this deficiency either by altering the functional directly[12–15] or by adding a dispersion correction on top of the DFT energy.[16–18] The formal scaling of DFT is $\mathcal{O}\left(N_{\text{basis}}^4\right)$ but this can be improved by various techniques. Using aggressive optimizations, a linear scaling algorithm for condensed phase systems was developed enabling the simulation of one million atoms with DFT.[19] This is however an extreme example requiring massive computational resources. Currently DFT simulations of systems containing several hundreds of atoms for thousands of configurations have become routine.

### Force fields

Force fields or empirical potentials take a radically different approach: the electrons are not considered explicitly. Instead, the energy is expressed as a rather simple analytic function of the positions of the nuclei. For instance the chemical bond between nuclei $A$ and $B$ can be modeled as a harmonic spring, so the contribution to the energy is given by

$$E_{\text{bond}} = \frac{1}{2} k_{AB} \left(R_{AB} - R_{AB}^0\right)^2 \tag{1.15}$$

Note that two parameters, the force constant $k_{AB}$ and rest value $R_{AB}^0$, are included which are not fundamental constants from quantum mechanics and

need to be determined in another way. Similarly simple expressions are often included to describe for instance dispersion and electrostatic interactions between atoms. A straightforward implementation of the typical force-field energy expression scales as $\mathcal{O}\left(N_{\text{atoms}}^2\right)$, which can be improved easily to $\mathcal{O}\left(N_{\text{atoms}} \log N_{\text{atoms}}\right)$. Not only is the complexity lower than for ab initio methods, the scaling is also in the number of atoms (not in the size of the basis set) and prefactors are typically a lot smaller. This makes force fields orders of magnitude faster, which allows them to be routinely used in MD simulations spanning millions of time steps for systems containing thousands of atoms.

### Ab initio derived force fields

Three different computational methods to compute the energy of a molecular system were briefly described: CCSD(T)/CBS, DFT and force fields. They are arranged in order of decreasing complexity: from providing results close to the solution of the Schrödinger equation (CCSD(T)/CBS) to a functional form that is not rooted in quantum mechanics (force fields). Accuracy of course comes at a price: while force fields can easily sample the phase space to compute macroscopic observables, CCSD(T)/CBS (an example of a post HF method) is limited to simulations at much smaller length and time scales as schematically shown in Figure 1.1.

The fact that force fields are not based on first principles can be seen as a weakness, but it also provides many opportunities because there is a lot flexibility in their design. Both the expression for the force-field energy and the accompanying force-field parameters can be chosen freely. This immediately sparks the idea of combining the best of both worlds: deriving force fields with input from quantum mechanics. This would allow performing simulations where the energetics are close to those predicted by the Schrödinger equation, at time and length scales that allow comparison with real-life experiments. This idea has inspired many force-field developers, as evidenced by a recent review on this very topic.[21] Still, current ab initio derived force fields can be improved tremendously. The most stringent attention points are limiting the overfitting of parameters, the ability to use highly accurate wave function methods to generate reference data and a physically grounded description of atomic interactions. The main goal of this dissertation is the development of a noncovalent force field that satisfies these requirements and then apply this methodology to simulations concerning the adsorption of guest molecules inside metal–organic frameworks (MOFs).

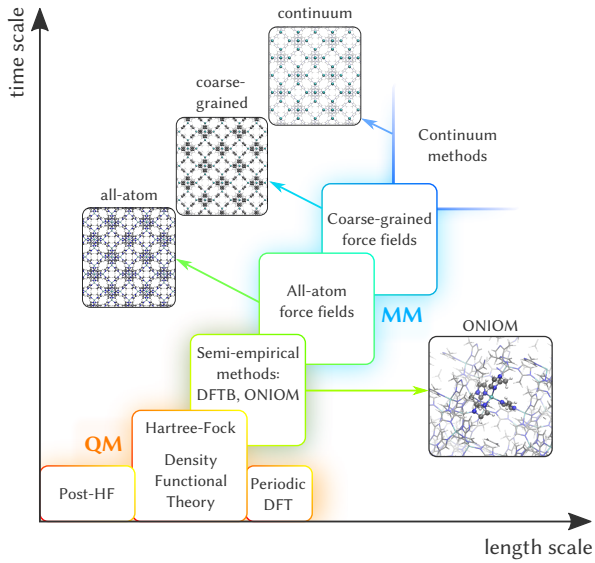



**Figure 1.1:** Schematic comparison of the length and time scales achievable with various levels of theory. Reproduced with permission from Ref. 20. Copyright (2017) The Royal Society of Chemistry.

## 1.2 Overview of the force-field landscape

Before starting the development of a novel force field, it is instructive to review efforts from the literature in order to indicate the strengths and weaknesses of current methods. This will allow a better understanding of the design choices that have been made in the novel methodology that will be presented later on. The number of available force fields is staggering and a complete list is simply impossible to compile. The following discussion only includes a selection of force fields, based on their relevance in the context of this Ph.D. The force fields are divided into different categories based on characteristics such as the complexity of the functional form, the input data used to construct the force field and the range of systems that can be treated. This classification is debatable: force fields often show characteristics of multiple classes. The guiding line for the classificaction presented here is to give a comprehensible overview of existing force fields, that helps to understand the challenges in the development of the new force field, which is the subject of Chapter 2. A schematic overview of how the different categories are positioned with respect to each other is provided in Figure 1.2.

### 1.2.1 General considerations

As mentioned earlier, the expression for the force-field energy can be freely chosen. Yet there are some terms which appear in the vast majority of force fields, except for instance in some machine-learned force fields. Therefore some general considerations on possible functional forms are discussed before moving on to actual examples from the literature.

Covalent interactions (interactions between chemically connected atoms) are usually described in terms of internal coordinates such as bond lengths, bending angles, dihedral angles... Next to automatically respecting translational and rotational invariance for molecules, the choice to use internal coordinates also allows interpretation of force-field terms similar to a spectroscopic analysis. Often bond lengths and bending angles remain close to their equilibrium value, which means a harmonic potential offers a good description. For dihedral angles, a Fourier expansion is more common in order to more easily satisfy symmetry constraints. More advanced covalent energy expressions can be constructed by including anharmonic terms, by considering additional internal coordinates such as out-of-plane terms, and by coupling different internal coordinates.

Also concerning noncovalent interactions, many force fields share similar potentials. For convenience, the following discussion is initially focused on the noncovalent interactions between two molecules situated at a large separation distance, meaning that the overlap of the wave functions is neglegible. Because molecules consist of charged nuclei and charged electrons, they will interact electrostatically. In principle, it is possible to perform a multipole expansion of each molecular charge density, centered for instance on the center of mass of the molecule. Aside from possible convergence issues in the expansion series, this also requires the implementation of electrostatic interactions between high-order multipoles. It is therefore very common to work only with monopoles (charges, for which $\ell = 0$) but on each atomic site. A first term that is present in nearly all noncovalent force fields, is therefore the electrostatic or Coulomb interaction between atomic point charges. Extensions of this basic model include higher-order multipoles, distributed charge densities, and additional centers next to the nuclei.

Noncovalent interactions are generally much weaker than covalent interactions, and perturbation theory is thus a proper tool to investigate noncovalent interactions for the situation where two molecules are situated far apart.[22] Concerning dispersion interactions, this approach leads to the well known $E_{\text{disp}} = -\frac{C_6}{R^6}$ expression, where $R$ is the separation distance, which is also nearly universally present in all classical potentials. More complex dispersion models can include higher order terms, many-body terms, atomic

anisotropy, and short-range damping. Next to dispersion interactions, also induction can be treated using second-order perturbation theory to provide a leading term which can be interpreted as stemming from classical inducible dipoles. This term is however less common, as it increases the computational complexity of the force-field energy calculation.

There is a wide consensus on the treatment of noncovalent long-range interactions between molecules, because imposing antisymmetry on the dimer wave function (starting from monomer wave functions) is trivial and there is no charge transfer between the monomers. At shorter ranges around the van der Waals separation distance, the monomer wave functions start to overlap and the situation is much more complicated. In this case, the antisymmetry of the dimer wave function can be imposed in different ways. Especially when higher orders of perturbation theory are considered, different terms become increasingly difficult interpret, as different physical effects (electrostatics, dispersion, exchange-repulsion, induction) are all intertwined. A more elaborate treatment of the above discussion can be found in Chapter 6 of the book "The Theory of Intermolecular Forces" by Stone.[23]

A common workaround to this problem is to simply ignore all considerations mentioned above. Instead, a simple term that prohibits atoms from colliding is included (for instance a hard-spheres model, the $R^{-12}$ repulsion of the Lennard-Jones potential, or the exponentially decaying wall of the Buckingham potential) and include all short-range effects effectively in the parameters of the model, without explicitly providing a model for them. Indeed, for instance the Dreiding force field describes noncovalent interactions by a combination of atomic point charge interactions and a Lennard-Jones potential. This particular combination is in fact a very popular choice and common to many force fields, as discussed in the following section.

### 1.2.2 Generic force fields

Although the first MC simulation for a system of rigid spheres dates back to 1953,[24] the foundations of many modern force fields were only laid in the 1980s and 1990s. The first of these efforts that will be discussed is the Dreiding[25] force field, developed by Goddard and coworkers, as it allows the introduction of some basic concepts shared by many force fields. It is a *generic* force field in the sense that one general set of parameters is available, from which the parameters specific to a system are extracted based on the topology and simple hybridization considerations. The nonbonded or noncovalent contributions $E_{\mathrm{nb}}$ of the standard Dreiding force field include point-charge electrostatics ($E_{\mathrm{Q}}$) and the Lennard-Jones model for van der

Waals interactions ($E_{\mathrm{vdW}}$):

$$E_{\mathrm{nb}} = E_{\mathrm{Q}} + E_{\mathrm{vdW}} \tag{1.16}$$

$$E_{\mathrm{Q}} = \sum_{i>j} f_{ij} \frac{q_i q_j}{r_{ij}} \tag{1.17}$$

$$E_{\mathrm{vdW}} = \sum_{i>j} f_{ij} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{1.18}$$

The sum runs over all pairs of atoms $i$ and $j$ and $r_{ij}$ indicates the distance between those atoms. The van der Waals parameters are only defined for cases where $i$ and $j$ denote atoms of the same type. In other cases, mixing rules are used to obtain $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$ and $\sigma_{ij} = \frac{\sigma_{ii}+\sigma_{jj}}{2}$. For each pair of atoms, a so-called fudge factor $f_{ij}$ is included. In the Dreiding force field, it is equal to zero if $i$ and $j$ are neighbors (1-2 interactions) or second nearest neighbors (1-3 interactions), and equal to unity in all other case. The idea behind this is that chemical bonds are much stronger than noncovalent interactions, so the latter will be contained in the bonded energy terms (discussed later). The original Dreiding paper acknowledges difficulties in determining atomic charges $q_i$ (a topic which will be extensively discussed in Section 2.1), and suggests to use either no charges or Gasteiger charges.[26] The van der Waals parameters $\epsilon_{ii}$ and $\sigma_{ii}$ are provided for $45$ different atom types. Atom types are used to distinguish chemically different atoms (according to the specific force field), and for Dreiding depend on the atomic number, the hybridization or geometry and possibly the oxidation state. The atom types are focused on the elements H, C, N and O, but additionally elements such as B, Na, Ca, Zn and Fe are also parametrized. Numerical values for the Dreiding van der Waals parameters are partly adopted from the force field by Williams and Cox,[27] which in turn has been fitted to reproduce experimental crystal structures and sublimation energies, and partly by interpolating, extrapolating or fitting new values. Note that sometimes hydrogen atoms bonded to carbon atoms are treated implicitly, meaning that the hydrogen nucleus is not explicitly present in the simulation, but its effect is included in the modified parameters for the effective carbon atom. This is also referred to as the united-atom approach, in contrast to the all-atom approach. As a final note, it is mentioned that Dreiding sometimes includes additional nonbonded terms to describe hydrogen bonds.

The valence interactions consist of bond stretching ($E_B$), angle bending ($E_A$), dihedral angle torsion ($E_T$) and inversion terms ($E_I$). Expressions for

individual instances of these internal coordinates are given by:

$$E_B = \frac{1}{2} K_B \left(R - R_e\right)^2 \tag{1.19}$$

$$E_A = \frac{1}{2} K_A \left(\theta - \theta_e\right)^2 \tag{1.20}$$

$$E_T = \frac{1}{2} V \left\{1 - \cos\left[n\left(\varphi - \varphi_0\right)\right]\right\} \tag{1.21}$$

$$E_I = \frac{1}{2} C \left(\cos\Psi - \cos\Psi_0\right)^2 \tag{1.22}$$

where $R$, $\theta$, $\varphi$ and $\Psi$ denote bond lengths, bending angles, dihedral angles, and angles between a bond and the plane formed by three atoms respectively. Other variables are force-field parameters that depend on the atom types of the atoms in question.

Other generic force fields share much similarities with Dreiding and these will be discussed more concisely. The Universal Force Field (UFF)[28] extends Dreiding in the sense that parameters are provided for the entire periodic table up to Lw (Z=103). Several functions for modeling van der Waals interactions are discussed, but finally the Lennard-Jones potential is retained. The corresponding parameters are based on a combination of literature values, fitting to experimental crystal structures, ab initio calculations, and extrapolation schemes. To compute atomic charges, the QEq method is advised.[29]

Assisted Model Building with Energy Refinement (Amber) is rather a family of force fields, originally developed at the Kollman group, sharing a common functional form.[30] The original work again uses the Lennard-Jones potential in the nonbonded energy, with parameters based on work by Hagler et al. [31] but with *ad hoc* adjustments.

The final example that is discussed is the MM3[32–34] force field. The MM3 energy expression is more complex than for the previous cases. The van der Waals interactions for example are described by the Buckingham potential:

$$E_{\mathrm{vdW}} = \sum_{i>j} f_{ij} \epsilon_{ij} \left[-2.25 \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 + 1.84\,(10)^5 \exp\left\{-12 \frac{r_{ij}}{\sigma_{ij}}\right\}\right] \tag{1.23}$$

The exponential function used to model repulsion at short interatomic distances, is physically more justified than the term proportional to $r^{-12}$ used in the Lennard-Jones description.[35–37] The latter has been introduced because it is computationally very cheap to compute once $r^{-6}$ has been calculated, which needs to be done for the long-range dispersion anyway. The Buckingham parameters are obtained starting from the values used in the preceding MM2 force field,[38] but *ad hoc* corrections are introduced. The charge

distribution in MM3 is represented using bond dipoles, instead of the more frequently used atomic point charges.

Some more examples of generic force fields such as the Transferable Potentials for Phase Equilibria (TraPPE)[39] or Chemistry at Harvard Macromolecular Mechanics (CHARMM)[40] could be discussed, but by now the general characteristics of this class should be clear. Generic force fields feature relatively simple expressions for the potential energy, provide a general set of parameters that can be used for a wide range of systems, and are fitted at least partly to reproduce experiments. The original papers that were discussed here acknowledge that the construction of the noncovalent part of the force field is problematic. Mostly very simple models (Lennard-Jones or Buckingham potential and optionally point-charge electrostatics) are employed, with parameters whose origin is not always clear and are often determined *ad hoc*. It is therefore striking that these force fields are still used today (sometimes modified), and indeed often provide sensible results. Yet, generic force fields have been criticized many times because of their poor accuracy.[41] All following categories of potentials aim to improve this in one way or another.

### 1.2.3 Case-specific alterations to existing parameter sets

There are countless papers where some generic force field is used as a starting point, but because it provides poor agreement with experiment for the system at hand, the parameters are altered in an attempt to fix this. Hereafter some examples of this procedure are provided, mostly related to nanoporous materials. Pérez-Pellitero et al. rescaled UFF Lennard-Jones parameters to obtain better agreement with experimental isotherms for $N_2$ and $CO_2$ in zeolitic imidazolate frameworks (ZIFs).[42] Wu et al. rescaled the UFF $\epsilon$ parameters by $0.635$ so the force field reproduces the experimental adsorption isotherm of $CH_4$ in ZIF-8.[43] Boulanger et al. found that rescaling GAFF molecule-water dispersion energies by $1.115$ resulted in optimal hydration free energies.[44]

A more systematic approach to adapting an existing generic force field is UFF4MOF.[45, 46] As the name suggests, UFF4MOF extends UFF with additional atom types for transition metal elements and oxygen, in the environment in which they typically occur in MOFs. The functional expression for the energy as well as the noncovalent parameters are unchanged, so only covalent parameters for the new atom types need to be determined. Of these covalent parameters, bond angles and coordination numbers are determined "by observation" from structures (combined experimental and DFT computed structures), while the new bond radii are fitted so the new force field reproduces the same MOF structures.

*Ad hoc* alterations to existing parameter sets seem attractive as they allow obtaining the wanted results quickly, without the need to go through the time-consuming development of a new force field from scratch. It is however questionable whether for exampling rescaling all van der Waals parameters is a physically meaningful way to mend deficiencies in the existing PES. There is also the obvious risk of overfitting, rendering the *predictive* capabilities questionable despite the capability of these force fields to *reproduce* reference results.

### 1.2.4 Parametrization methodologies

Instead of directly providing a list of parameter values, as generic force fields do, it is also possible to specify the method that is used to determine those parameter values. Such parametrization methodologies are also applicable to a wide range of systems, but contrary to generic force fields which can be considered "Plug and Play", a simulation has to be preceded by a setup procedure executed by the user. This setup can include collecting experimental data, performing ab initio calculations to obtain reference results or running (short) force-field simulations.

A first example of a parametrization methodology is potfit,[47–49] which is a code (primarily aimed at solid-state systems) that implements the force matching approach.[50] The basic idea is that force-field parameters are determined so they minimize a cost function, which measures the deviation from ab initio computed atomic forces, energies and/or stresses obtained for a number of configurations relevant for the application of interest. Potfit can handle different interaction models, including the Embedded Atom Method[51] which is often used for metallic systems, but can also work with tabulated potentials for which no *a priori* analytical expression is assumed. Finding the global minimum of the force-matching cost function is often problematic as it is likely to be ill-conditioned. Several advanced algorithms to perform the search for the best parameters are therefore implemented in potfit. Because any new force field should be tested and validated, the potfit authors state that it typically takes a few weeks before for the new force field can actually be used in simulations. ForceFit[52] and ForceBalance[53] are other codes that implement the force matching methodology, but are designed rather for applications in organic chemistry.

The Joyce[54, 55] protocol determines all parameters of the covalent force field, using information about energies, forces and the Hessian, which are computed for the target molecule using a first-principles method. Next to the often encountered bond, angle and dihedral terms, Joyce also allows the

use of off-diagonal terms, which include an explicit coupling between different internal coordinates. Some of the Joyce authors have also developed an automated protocol to construct noncovalent force fields, again using quantum mechanical data.[56] In this approach, named Picky, intermolecular interactions are expressed as a Lennard-Jones potential and point-charge electrostatics. The Picky protocol optimizes parameters iteratively. First MD simulations are performed using the current parameters, from which a number of dimer configurations are extracted. The interaction energy of the dimers is calculated using a quantum mechanical method of choice and the force-field parameters are further optimized (by minimizing a Boltzmann-weighted deviation of interaction energies) to better fit those results. Using the updated parameters, new MD simulations are performed after which the cycle is repeated until the parameters change negligibly. The original paper used Picky to derive a pyridine ($C_5H_5N$) force field with the quantum mechanical calculations performed at the MP2/cc-pVDZ(mod) level of theory, resulting in good agreement with experiment for the liquid density and vaporization enthalpy. It is interesting to note that the initial parameter estimates were taken from the Optimized Potentials for Liquid Simulations (OPLS) force field, but the final parameters deviated significantly. Still, both parameter sets predict very similar bulk properties, highlighting the redundancy in the parameter set.

During this Ph.D. a novel in-house parametrization methodology for covalent force fields named QuickFF was developed.[57, 58] One of the distinguishing features of QuickFF is that correlations among the parameters are dealt with using so-called perturbation trajectories, in order to avoid the earlier mentioned problems with ill-conditioned cost functions and redundant parameters sets. It permits the use of both periodic and cluster ab initio calculations as reference data and can describe the coupling between internal coordinates by using cross terms. The usefulness of QuickFF has been mainly demonstrated by its application in the context of force-field simulations of (flexible) MOFs.[59–63]

### 1.2.5 Reactive force fields

One of the common features of the force fields discussed up to now is that the topology is part of the system specification and remains unchanged in any simulation. In other words, bonds can never be broken and thus chemical reactions can not be modeled. Chemical reactions involve important redistributions of the electron cloud, and as the latter is not explicitly modeled, it is perhaps no surprise that force fields are not a proper starting point to describe for instance bond formation or breaking. Yet there are certain so-

called reactive force fields which aim to do exactly this, the most famous one being ReaxFF developed by van Duin, Goddard and coworkers.[64] Instead of using explicit bonds between atoms, ReaxFF assigns bond orders which are updated during a simulation, which allows bonds to form or break in a continuous manner. Parameter sets are available for many systems, including hydrocarbon reactions and oxygen interactions with silica surfaces. One of the weak points is that ReaxFF requires an enormous number of parameters, which additionally are not straightforward to determine.

### 1.2.6 Polarizable force fields

The computational effort to compute the potential energy scales quadratically with the number of atoms, $\mathcal{O}\left(N_{\mathrm{atoms}}^2\right)$, for most force fields. The number of terms contributing to the covalent energy is proportional to the system size, and for large systems most of the computational time is spent on the noncovalent energy, where the number of terms is proportional to the number of atom pairs. This explains the formal quadratic complexity, although it should be mentioned that algorithms such as domain decomposition and Fast Fourier transform allow a reduction to $\mathcal{O}\left(N_{\mathrm{atoms}} \log N_{\mathrm{atoms}}\right)$ in practical applications. This is all under the assumption that the force-field model of the charge density is not influenced by the presence of other charge densities, i.e. no polarization effects are included (most often this means that fixed point charges are used).

This section discusses force fields that explicitly model microscopic polarization, for instance by considering inducible dipoles, which should give a physically more correct description at the cost of increasing the formal complexity to $\mathcal{O}\left(N_{\mathrm{atoms}}^3\right)$. As a prototypical case the atomic multipole optimized energetics for biomolecular applications (AMOEBA) force field will be discussed.[65–68] Polarization is modeled by attaching an inducible dipole $\boldsymbol{\mu}_i$ at site $i$ which is assumed to be proportional to the electric field $\mathbf{E}_i$ at the position of site $i$:

$$\boldsymbol{\mu}_i = \overline{\overline{\alpha}}_i \boldsymbol{E}_i \tag{1.24}$$

Note that this induced dipole will change the electric field at the positions of the other sites, in turn generating induced dipoles at the other sites which then again influence the electric field at site $i$ et cetera. This shows that the above set of equations can be solved iteratively (alternatives are direct matrix inversion or extended Lagrangian schemes), explaining the additional complexity for such a force field that includes many-body effects (compared to the more conventional pairwise-additive models for the noncovalent energy). Much effort has been devoted to enable large-scale MD simulations using

the AMOEBA force field. This resulted in the Tinker-HP[69] software package which, thanks to its good scaling over up to thousands of processors, allows AMOEBA simulations of large water boxes and biosystems.

A fundamental difficulty with inducible point dipoles is the so-called polarization catastrophe, which arises when these dipoles are too close together. This deficiency has to be mended, for instance by damping the electrostatic interactions at short range.[70] Also the determination of the additional force-field parameters (the atomic polarizability tensors $\overline{\overline{\alpha_i}}$) can be problematic. The atom-condensed Kohn-Sham DFT approximated to second order (ACKS2) polarizable force field is an alternative approach (rooted in quantum mechanical equations instead of derived from classical electrostatics) which helps to deal with some of these problems.[71, 72]

### 1.2.7 Machine learning potentials

Machine learning is a branch of computer science where a computer system learns to perform a certain task based on input data delivered by the user, but without being explicitly programmed to execute the task. The progress in machine learning in other fields has also incited new research in molecular modeling and more specifically in the development of force fields. The basic idea that is often used is the following. Quantum mechanical data (energies, forces, but also for example atomic charges) are generated for a range of configurations. This is the only input that is needed to construct the force field: the machine learning algorithm does not make any assumptions about the physical model for the quantum mechanical data, but simply makes implicit connections between configuration geometry and resulting values of interest. Later, these implicit connections can then be used to predict estimates for quantum mechanical data for new configurations not present in the training set.

The gradient-domain machine learning (GDML)[73] and its successor incorporating spatial and temporal symmetries (sGDML)[74] are examples of models for intramolecular forces, applicable to small- and medium-sized molecules. A distinctive feature of GDML is that atomic forces are used as input data rather than the energy. The sGDML offers an important improvement by exploiting physical symmetries present in the input data, which allows a significant reduction of the number of simulations required to generate the reference data. This makes it feasible to use the accurate CCSD(T) level of theory for this task. Some of the same authors also developed a classical intermolecular potential with the help of machine learning, coined IPML. The energy expression is based on the MEDFF[75] potential, which was developed

in the context of this Ph.D., but extended with electrostatic multipoles, Thole-damped inducible dipoles and many-body dispersion. The spirit of MEDFF is also retained, by working with only 8 global parameters that have to be fitted once and are further used across compounds. The crucial new element in the IPML approach is that environment-dependent local, atomic properties (electrostatic multipoles, description of valence electron densities and polarizabilities) are predicted by machine learned models. These models have been shown to be transferable across small neutral organic and biologically relevant molecules.

FFLUX[76, 77] (formerly known as QCTFF[78]) is a protein force field in development based on atomistic kriging models. These machine learning models allow predicting energies and multipole moments given the geometry, with atomic partitioning achieved using the Quantum Theory of Atoms in Molecules (QTAIM).[79] The development of FFLUX has been going on for many years, focusing on the individual components separately. Only recently, efforts to place the different pieces in their place have emerged holding promise for a truly innovative force field that can actually perform meaningful simulations in coming years.

### 1.2.8 Force fields directly incorporating ab initio quantities

The ab initio force fields discussed before usually only incorporate quantum mechanical results *indirectly*. Typically, ab initio energies (and their derivatives towards nuclear coordinates) are used to construct a cost function, which is then minimized in order to obtain the many parameters present in the predefined force-field energy expression. In contrast, this section features force fields that incorporate quantum mechanics in a more direct way.

The first example is the Gaussian Electrostatic Model (GEM),[80–82] which explicitly retains the electron density. The ab initio computed electron density is expanded in an auxiliary basis of Hermite Gaussian functions at specific sites (not limited to the positions of the nuclei). The coefficients of the expansion are determined by density fitting, an approach in which the Coulomb integral of the difference in ab initio and force-field density is minimized. The resulting model density in terms of Hermite Gaussian functions can obviously be used to evaluate the electrostatic interaction between fragment densities at a low computational cost, including the often-neglected penetration effect. In GEM, the model density is also used to compute the exchange-repulsion contribution to the force-field energy. More precisely, the exchange-repulsion energy is assumed to be proportional to the overlap of fragment densities,[83] with the proportionality factor a fitting parameter. The GEM force field differs from previously discussed potentials be-

cause its development is based on a decomposition scheme, which partitions ab initio interaction energies into physically meaningful components. The employed decomposition scheme is called constrained space orbital variation (CSOV),[84] and it provides the electrostatic, exchange-repulsion, polarization and charge-transfer contributions to the total interaction energy. The latter two terms are modeled in GEM similar as is done in the sum of interacting fragments ab initio computed (SIBFA) force field.[85]

Another ab initio decomposition scheme is symmetry-adapted perturbation theory (SAPT),[86] which gives rise to exchange, electrostatic, induction and dispersion contributions to the total interaction energy. A recent force field where the separate terms correspond precisely to this decomposition is SAPTFF by McDaniel et al. [87] The starting point is again a monomer electron density, from which monomer properties (atomic charges, static polarizabilities and dispersion coefficients) are derived based on a distributed multipole analysis (DMA).[88] These asymptotic interaction parameters are sufficient to construct the "long-range" force field, which works well if the closest contact distance between monomers is significantly larger than their approximate van der Waals contact distance. At shorter ranges however, the remaining parameters are derived based on SAPT calculations of dimers and ionization potentials of isolated atoms. The fitting of parameters is done separately for the different contributions of the force field (electrostatic, exchange, polarization and induction), with the corresponding SAPT contribution(s) as reference data. This approach guarantees that the resulting force field is physically grounded and should provide better transferability to systems not included in the reference data. SAPTFF was later extended by adding explicit terms that describe three-body exchange and dispersion interactions.[89] The contribution of these terms turned out to be significant for the simulation of liquids and can therefore not be neglected in an ab initio derived force field that works both for gas and condensed phase.

A similar approach has been presented by Misquitta et al. A first effort concerns the development of an ab initio derived force field for the pyridine dimer.[90] As many parameters as possible are "computed", rather than fitted, from monomer calculations using a basis-space algorithm for the iterated stockholder atom procedure.[91, 92] This fixes the parameters for the long-range terms, and for the short-range terms, first-order SAPT(DFT) energies are first partitioned into contributions from pairs of atoms in order to avoid difficulties encountered when fitting the sum of many exponentials together. Further investigation of the short-range repulsion resulted in the Slater-ISA force field,[93] which is based on overlapping atomic densities modeled by a Slater like function. Another distinctive feature of these potentials is the inclusion of atomic anisotropy in most force field terms. The effects of

anisotropic terms has been further investigated and lead to the Multipolar, Anisotropic, Slater-Type Intermolecular Force Field, MASTIFF.[94] The basis-space algorithm for the iterated stockholder atom procedure can also be used to obtain non-local distributed frequency-dependent polarizabilities.[95] This method, termed ISA-Pol, can be used as the basis for polarization and dispersion models that are directly derived from ab initio calculations.

## 1.3 Goals and outline

### 1.3.1 A wish list for a novel noncovalent force field

The discussion in the previous section provided an overview of different classes of force fields. Considering that only a small subset of potentials from the literature was considered, it should be clear that a myriad of force fields is currently at ones disposal. This raises the question whether there is actually a need for yet another new force field, which is the methodological contribution of this thesis. More precisely, a novel methodology to construct noncovalent force fields completely from ab initio input will be presented. The development of this force field started after deficiencies were noted during simulations of MOFs with a widely used noncovalent potential, which had to be mended by *ad hoc* modifications to the parameters.[96] Because such modifications to empirical potentials are not a satisfactory long-term solution, and existing more advanced noncovalent force fields were not applicable to the materials of interest, it was justified to begin the construction of a novel methodology from the bottom up. A number of principles that the novel force field ideally would adhere to are listed below. This list can be considered a blueprint for the methodology that will be developed in the remainder of this work.

- Each contribution to the force-field energy should have a clear physical meaning. This can be achieved by fitting term by term to results from an ab initio energy decomposition scheme.
- Parameter fitting is often an ill-conditioned optimization problem. The condition number should be kept low, for instance by limiting the number of parameters and thus the dimensionality of the parameter space.
- Reference data should be generated by highly accurate quantum mechanical methods. In other words, a modest amount of reference calculations on relatively small systems should suffice.

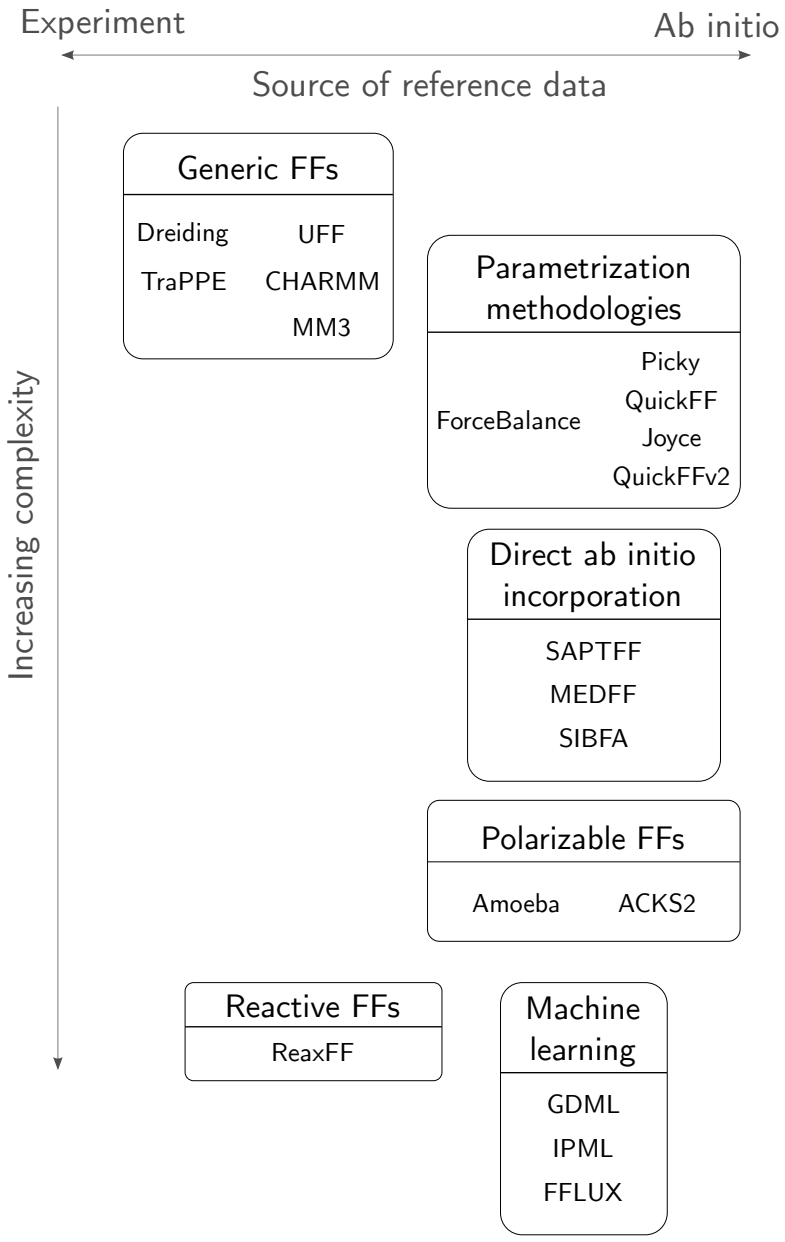



**Figure 1.2:** Schematic overview of the classification of some selected force fields, sorted by increasing complexity from top to bottom. The presented scheme is necessarily debatable because of the many degrees of freedom existing in the construction of force fields and a more elaborate discussion is provided in the main text.

- The method should be robust and applicable to a large portion of chemical space. This means that the force field should be constructed for an entire database, rather than adjusted to suit a specific case.

- Simplicity should be favored over complexity so the approach remains conceptually transparent. On a related note, the computational cost should be comparable to conventional force fields.

It is clear that these points have received attention in the literature before (for instance in the works discussed in Section 1.2.8 on force fields that directly incorporate ab initio quantities), but never all in one encompassing effort.

The last two points are especially important to apply such a noncovalent potential to a challenging problem such as gas adsorption in MOFs. As mentioned earlier, conventional force fields are often not suited for the description of noncovalent interactions in MOFs, which can be demonstrated by considering some recently constructed force fields, specifically aimed at improving the description of gas adsorption in MOFs.[97–99] With applications such as carbon capture and sequestration or natural gas vehicles in mind, gas adsorption in MOFs holds promise for industrial applications. Still, further developments in molecular simulation methodologies are required and therefore the last part of this dissertation is devoted to an investigation of how the noncovalent force field derived in this work can contribute to this topic.

### 1.3.2 Outline of this dissertation

The subject of this dissertation has been extensively introduced in this first chapter and the goals have been made clear. Now is an appropriate time to provide an overview of the remainder of this work. Naturally, a division into two parts emerges: a methodological section, the development of a novel noncovalent force field, and an application section, where this force field is applied to simulate the adsorption of small gas molecules in MOFs Frameworks.

In Chapter 2, a novel noncovalent force field methodology is developed. The first step is the introduction of a new atoms in molecule scheme, which partitions the molecular electron density into atomic contributions. This results in a simplified model for that density, which in turn leads to a direct computation of many force-field parameters. Next, the functional form of the force field is motivated and a robust procedure to obtain physical values for the remaining parameters is introduced. After calibration to a database

of dimers, the performance of the new method for the prediction of second virial coefficients and liquid properties of small hydrocarbons is investigated.

The following Chapter 3 applies the newly developed force field to a technologically relevant problem in molecular modeling: simulating the adsorption of small guest molecules in MOFs. First, different potentials (both ab initio and force field) are compared with respect to the adsorption of methane in Zr-based MOFs. Second, an importance sampling method is introduced which allows computing ab initio Henry constants using a minimal amount of quantum mechanical calculations.

The final Chapter 4 provides concluding remarks and prospects on future work. The extent to which the proposed new force field fulfills the criteria put forward in the introduction is discussed, along with a personal reflection on the impact of this dissertation on further development of noncovalent force fields.

# 2

# Methodology: Development of a Noncovalent Force Field

In Chapter 1 a broad overview of force fields available in the literature was presented. Despite the wealth of potentials at one's disposal, the need for the development of new force fields was extensively motivated resulting in a list of desirable properties those force fields should satisfy. These requirements form the blueprint for the new noncovalent force field presented in the current chapter, whose development is the main methodological advancement of this Ph.D. dissertation.

In Section 2.1 a new methodology to condense information from a quantum mechanical calculation into atomic parameters is presented. These atomic parameters are, together with an ab initio energy decomposition scheme, vital ingredients for the novel noncovalent force field, whose recipe is discussed in Section 2.2. Results obtained for some seemingly simple test cases are provided and contrasted with experiment in Section 2.3, which leads to a re-evaluation of the originally proposed requirements and pointers for future research on the generation of noncovalent potentials in Section 2.4.

## 2.1 Concise representation of the electron density: the MBIS partitioning scheme

Most ab initio derived force fields employ only the end results of quantum mechanical calculations, such as energies or forces on the nuclei. Here,

an approach that uses quantum mechanical data in a more direct way is pursued. As explained in Chapter 1, the electronic part of the wave function $|\Psi^e\rangle$ contains all information given a set of nuclear coordinates $\{\mathbf{R_1}, \mathbf{R_2} \ldots \mathbf{R_N}\}$. The complexity of the wave function is in first instance reduced by only considering the electron density $\rho(\mathbf{r})$. Evaluating integrals involving the electron density, for instance to calculate electrostatic interactions, is still computationally expensive. Furthermore, in force fields it is desirable to express quantum mechanical information in terms of atomic properties. Therefore, the density $\rho(\mathbf{r})$ is written as a sum of (analytically simple) densities with each term corresponding to a nucleus, which allows a further reduction to a limited number of force-field parameters for each nucleus. The partitioning of the electron density into atomic contributions can be done using an atoms in molecules (AIM) scheme. The term AIM is also used for the partitioning of the electron density of extended systems such as liquids or crystals, as this is conceptually the same as for molecules. Despite the fact that considering a molecule as being composed of atoms offers valuable chemical insight, this picture does not arise directly from quantum mechanics: the wave function is ignorant about atoms in a molecule. Still, many *useful* AIM schemes are available in the literature that use different assumptions to achieve a quantitative partitioning of a molecule into atomic contributions. In the context of this Ph.D. dissertation, another AIM scheme is developed that is especially suited as the basis for noncovalent force fields.

Below, the need for this novel partitioning scheme, referred to as Minimal Basis Iterative Stockholder (MBIS), is briefly motivated and its main methodological features are explained. Next, a comparison with other schemes is presented, focusing on their further use in force fields. The detailed derivation and results of the MBIS partitioning scheme have been published in **Paper I [Ref. 100]**.

### 2.1.1 The need for a robust atoms in molecules method

Existing AIM methods can be classified into different categories. The first category consists of methods that fit atomic charges in order to reproduce the electrostatic potential (ESP), a function that can be computed from the electron density, as closely as possible. Because atomic point charges are not adequate to model the ESP generated by nuclei and the electron cloud *inside* a molecule, a cost function has to be constructed that only takes deviations with the reference ESP into account in regions outside the molecule with relatively low electron density. The detailed choice of this cost function is to a certain extent arbitrary and this has led to a multitude of ESP fitting variants such as the Merz-Singh-Kollman scheme,[101, 102] the ChelpG scheme[103], the

Hu-Lu-Yang fitting method,[104] or the REPEAT method[105] which all differ in the choice of the cost function. ESP fitting methods in general can show some drawbacks such as limited transferability of charges, excessive conformational dependency, sensitivity to the level of theory of the reference ESP and ill-conditioned charges for buried (highly coordinated and thus shielded) atoms, but workarounds to most of these problems have been presented in the literature.[101, 102, 104, 106] By construction, ESP fitted charges will reproduce the ESP outside the molecule and this seems to guarantee a faithful reproduction of molecular multipole moments (at least the lowest orders). This in turn seems to lead to good predictions of intermolecular electrostatic interactions, as long as molecules are very far apart and their electron clouds do not overlap. This reasoning is often used to justify the use of ESP fitted atomic charges in force fields. However, atomic point charges can not capture the electrostatic penetration effect, which is much more important at equilibrium dimer geometries than commonly assumed. This is the major reason why fitting to the ESP is not a proper starting point for the construction of a physically motivated force field.

The second class of partitioning methods act in Hilbert space, which means the density matrix is partitioned directly. The Mulliken[107] scheme is the most widespread of this class, despite showing some deficiencies such as unphysical negative populations, unduly basis set sensitivity and unreasonable results for complexes with ionic character. By using an orthogonalization procedure due to Löwdin, problems with negative populations can be avoided.[108] A further improvement is made by the construction of so-called Natural Atomic Orbitals, from which Natural populations can be calculated.[109] An important assumption in these methods is that each orbital basis function is centered on one of the nuclei, which is the case for most localized basis sets. For calculations performed on a grid or using plane waves, these Hilbert-partitioning schemes are not applicable unless the wave function is first transformed into a set of spatially localised quasiatomic orbitals.[110] Together with the finding that Mulliken, Löwdin and Natural charges all provide a rather poor description of the ESP, this raises objection to the use of these methods in force-field development.

The third and final considered class of AIM approaches are density partitioning/fitting methods. Density fitting can be done in a conceptually straightforward way by introducing an auxiliary basis set and fitting the expansion coefficients of those basis functions to minimize some metric, typically the Coulomb self-interaction energy of the difference between the fitted density and the reference density.[80] When both the reference wave function and the auxiliary basis set consist of Gaussian-type functions, the coefficients are determined by a set of linear equations. This approach is

frequently used in quantum mechanical calculations in order to simplify four-center integrals, but has also been used as the basis for the GEM-0[81] and GEM*[111] force fields as discussed in Chapter 1. The auxiliary basis sets used for density fitting are typically rather large, leading to a non-negligible additional computational cost when used in a force field afterwards. They also lack a clear physical interpretation, as it is not trivial to relate the expansion coefficients to properties of the atoms. In this respect, density-based partitioning methods are an interesting alternative. A first example of such a method is QTAIM, which divides a quantum mechanical system in non-overlapping basins that are delimited by surfaces through which the flux of the gradient of the electron density is zero,[79] which ensures that the atoms have a well-defined kinetic energy. The resulting basins (generally representing atoms) are *oddly* shaped and do not resemble isolated atoms at all. The Hirshfeld method on the contrary enforces this by considering the density as a weighted sum of so-called pro-atoms, which are densities computed for isolated, neutral atoms.[112] Hirshfeld partial charges are however rather small in absolute value (i.e. not in agreement with chemical intuition, unable to reproduce the reference ESP, ...), an issue that can be solved by also including charged pro-atoms as is done in variants of the Hirshfeld method such as Iterative Hirshfeld[113] (HI), Extended Hirshfeld[114] and the DDEC methods.[115–117] All aforementioned Hirshfeld approaches require a precomputed database of pro-atoms, possibly including nonexistent highly charged anions. The Iterated Stockholder Atoms[91] (ISA) approach alleviates this requirement by constructing pro-atoms on the fly. This however leads to ill-defined density tails of such pro-atoms[92], which in turn leads to limited conformational robustness.[114]

As an illustration, the atomic densities obtained using some selected density-based partitioning methods are shown in Figure 2.1 for a hydrogen fluoride (HF) molecule. Note that Hirshfeld (left) and MBIS (center) atomic densities appear similar, but do show an important difference in the tail of the hydrogen atom. The QTAIM atomic densities (right) are disjoint and sometimes also called atomic basins. This leads to rather *oddly* shaped atoms.

It is clear that numerous AIM methods are available in the literature, with many variants that all try to improve certain aspects of their predecessor. This has led to increasingly complex algorithms or the introduction of many tuned (empirical) parameters. The novel MBIS approach presented below aims to be elegant, straightforwardly applicable to large systems and free from empirical input without losing positive aspects of previously developed methodologies. This makes MBIS the only scheme that ticks all required boxes to be used as the basis for a noncovalent force field that fulfills the criteria explained earlier.

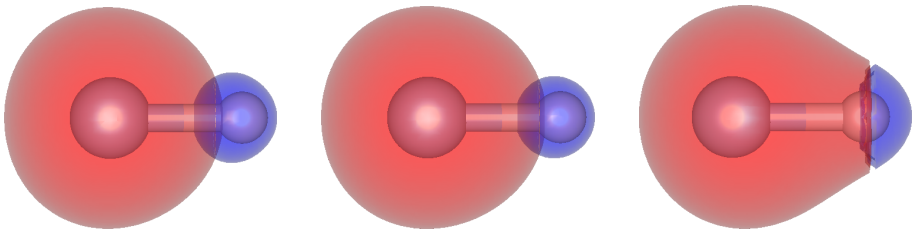


**Figure 2.1:** The isosurface of atomic densities at $\rho_A = 0.1\,\mathrm{bohr}^{-3}$ for Hirshfeld (left), MBIS (center) and QTAIM (right). The hydrogen fluoride molecular density is calculated using the PBE functional with a 6-311+G(2df,p) basis set. Fluoride densities are shown in red and hydrogen densities in blue. The fluoride atomic charges are $-0.21$, $-0.49$ and $-0.69$ for the different partitioning schemes respectively.

### 2.1.2 Theoretical derivation

Density-based AIM methods partition the given total electron density $\rho(\mathbf{r})$ into atomic contributions $\rho_A(\mathbf{r})$, $A = 1 \ldots N_{\mathrm{atoms}}$ that are to be determined. In the Hirsfeld approach (and variants) the atomic contributions are determined by the stockholder formula

$$\rho_A(\mathbf{r}) = \rho(\mathbf{r}) \frac{\rho_A^0(\mathbf{r})}{\rho^0(\mathbf{r})} \quad \text{with} \quad \rho^0(\mathbf{r}) = \sum_{B=1}^{N_{\mathrm{atoms}}} \rho_B^0(\mathbf{r}) \tag{2.1}$$

where $\rho_A^0(\mathbf{r})$ are pro-atomic densities. The factor $\frac{\rho_A^0(\mathbf{r})}{\rho^0(\mathbf{r})}$ can be interpreted as a weight function that assigns part of the total electron density to atom $A$, in other words it represents the *share* of atom $A$ in the total density (hence the name *stockholder*). The stockholder formula is supported by arguments from information theory. The Kullback-Leibler divergence $\int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) \mathrm{d}\mathbf{r}$ is a measure for the amount of information that is lost when the AIM density is approximated by the pro-atomic density. By minimizing the total information loss

$$\Delta S\left[\{\rho_A(\mathbf{r})\}; \{\rho_A^0(\mathbf{r})\}\right] = \sum_{A=1}^{N_{\mathrm{atoms}}} \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) \mathrm{d}\mathbf{r} \tag{2.2}$$

under the constraint that the sum of AIM densities equals the total density

$$\sum_{A=1}^{N_{\mathrm{atoms}}} \rho_A(\mathbf{r}) = \rho(\mathbf{r}) \tag{2.3}$$

one obtains Equation 2.1.

In the original Hirshfeld method, the pro-atomic densities $\rho_A(\mathbf{r})$ are fixed and obtained as spherically averaged isolated neutral atoms. The ISA on the other hand places no restrictions on the pro-atoms (except that they are spherically symmetric) and instead determines them self-consistently from the density when iterating the stockholder procedure. The MBIS method can be situated somewhere in between: the pro-density is expanded in a minimal basis of atom-centered s-type Slater density functions, whose parameters are determined as part of an iterative procedure. As explained above, the final parameters minimize the Kullback-Leibler divergence of the pro-density from the reference electron density. The MBIS procedure can in this sense also be understood as a density fitting technique employing a Slater-type basis (instead of the more common Gaussian basis functions) and the Kullback-Leibler divergence as the cost function (instead of the more commonly used Coulomb metric). The pro-density of atom $A$, with nucleus at $\mathbf{R}_A$, consists of a sum of 1s Slater-type density functions with $m_A$ terms:

$$\rho_A^0(\mathbf{r}) = \sum_{i=1}^{m_A} \rho_{Ai}^0(\mathbf{r}) \tag{2.4}$$

$$\rho_{Ai}^0(\mathbf{r}) = \frac{N_{Ai}}{\sigma_{Ai}^3 8\pi} \exp\left(-\frac{|\mathbf{r}-\mathbf{R}_A|}{\sigma_{Ai}}\right) \tag{2.5}$$

where $m_A$ is the number of shells of atom $A$ or its row in the periodic table. The prefactor for each term is chosen so that $\int \rho_{Ai}^0(\mathbf{r})d\mathbf{r} = N_{Ai}$ and thus the electron population of pro-atom A is given by $N_A^0 = \sum_{i=1}^{m_A} N_{Ai}$. Atom-centered Slater-type densities are chosen because these are the simplest functions that display a cusp at the nucleus as well as the correct asymptotic behavior far from the nucleus.

The parameters $\{N_{Ai}\}$ and $\{\sigma_{Ai}\}$ can be determined using a self-consistent algorithm, which can be derived starting from the following Lagrangian:

$$\begin{aligned} L[\{\rho_A(\mathbf{r})\}, \lambda(\mathbf{r}), \{N_{Ai}\}, \{\sigma_{Ai}\}, \{\mu_A\}] = & \sum_{A=1}^{N_{\text{atoms}}} \int \rho_A(\mathbf{r}) \ln \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \mathrm{d}\mathbf{r} \\ & + \int \lambda(\mathbf{r}) \left( \sum_{A=1}^{N_{\text{atoms}}} \rho_A(\mathbf{r}) - \rho(\mathbf{r}) \right) \mathrm{d}\mathbf{r} \\ & + \sum_{A=1}^{N_{\text{atoms}}} \mu_A \int \left( \rho_A(\mathbf{r}) - \rho_A^0(\mathbf{r}) \right) \mathrm{d}\mathbf{r} \end{aligned} \tag{2.6}$$

The first term of this Lagrangian gives the Kullback-Leibler divergence discussed before, the Lagrange multiplier $\lambda(\mathbf{r})$ in the second term constrains

the sum of AIM densities to the total density, and the Lagrange multipliers $\{\mu_A\}$ in the third term are used to ensure that the population of each pro-atom equals the corresponding AIM population. The latter is enforced in order to avoid ambiguities in the statistical interpretation of the Kullback-Leibler divergence.[118] The resulting Lagrange equations can be solved to obtain following expressions for the parameters of the Slater-type density functions:

$$N_{Ai} = \int \rho(\mathbf{r}) \frac{\rho^0_{Ai}(\mathbf{r})}{\rho^0(\mathbf{r})} \mathrm{d}\mathbf{r} \tag{2.7}$$

$$\sigma_{Ai} = \frac{1}{3N_{ai}} \int \rho(\mathbf{r}) \frac{\rho^0_{Ai}(\mathbf{r})}{\rho^0(\mathbf{r})} |\mathbf{r} - \mathbf{R}_A| \mathrm{d}\mathbf{r} \tag{2.8}$$

The same expressions can also be obtained by minimizing Equation 2.2 under the condition that $\int \rho\left(\mathbf{r}\right) d\mathbf{r} = \int \rho^0\left(\mathbf{r}\right) d\mathbf{r}$. The right-hand sides of the above equations contain the pro-density $\rho^0(\mathbf{r})$, which in turn contains the parameters that are present in the left-hand sides. This warrants a self-consistent solution: first Equations 2.4 and 2.5 are used to determine $\{\rho^0_{Ai}(\mathbf{r})\}$ with current values for $\{N_{Ai}\}$ and $\{\sigma_{Ai}\}$, and from this the pro-density $\rho^0(\mathbf{r})$ is obtained. These results are used to update the parameters $\{N_{Ai}\}$ and $\{\sigma_{Ai}\}$ according to Equations 2.7 and 2.8. The procedure is bootstrapped by generating initial guesses for the parameters (equal to the number of electrons in each shell of the corresponding neutral isolated atom for $\{N_{Ai}\}$ and inspired by hydrogenic s-type orbitals for $\{\sigma_{Ai}\}$) and repeated until the pro-atom parameters no longer change significantly from one self-consistent cycle to the next.

To illustrate the outcome of an MBIS partitioning, the results for a $CO_2$ molecule are shown in Figure 2.2. The gray line depicts the reference molecular density $\rho(\mathbf{r})$ along the O-C-O axis, while the black line shows the MBIS pro-molecular density $\rho^0(\mathbf{r})$. The dotted lines show the breakdown of the latter into its contributions: because oxygen and carbon are second-row elements, there are two basis functions for each atom. The 1s Slater-type densities show up as straight lines because of the log scale for the y axis. The lower panel of Figure 2.2 depicts similar results, but with the core Slater functions condensed into an effective point charge (visible as straight lines). This allows one to simplify the calculation of Coulomb or overlap integrals for pro-molecular densities, because the overlap between such core Slater functions is negligible.

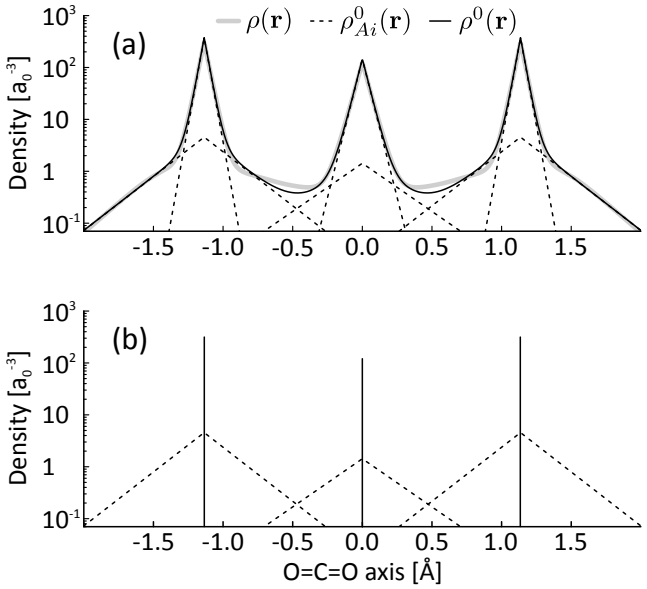



**Figure 2.2:** The molecular electron density of $CO_2$ (gray line) is approximated as a sum 1s Slater-type density functions. Reprinted with permission from Ref. 100. Copyright (2016) American Chemical Society.

### 2.1.3 MBIS as the basis for force-field development

The MBIS partitioning scheme outlined in the previous section was applied to various molecular data sets and compared to other methods regarding reproduction of the ESP, accuracy of electrostatic interactions and the robustness of atomic charges. First, a qualitative description of the results is provided for each metric separately. Afterwards, the performance of the different methods is assessed by performing a Pareto analysis involving combinations of the previously mentioned metrics.

#### Reproduction of ESP

It is sensible to require atomic charges used in a force field to be able to reproduce the ESP, computed using an ab initio level of theory, around a molecule. A good reproduction of this ESP is indeed a necessary condition for the force field to make good predictions of electrostatic interactions for the right reasons. The reference molecular electron densities were calculated using DFT (B3LYP[119–121] functional and 6-311+G(2df,p)[122] basis set) for the monomers contained in four dimer data sets:

- S66[123] containing a diverse set of neutral organic molecules
- IHB15[124] featuring ionic hydrogen-bonded complexes
- X40[125] which contains dimers with halogen bonds
- ZG237[100] consisting of dimers of neutral and anionic silica clusters and guest/template molecules typically adsorbed in porous media

as well as three data sets of fairly large molecules

- PENTA103[126] with multiple penta-alanine conformers
- SILICA245[114] which contains topologically difference silica clusters
- MIL53(M)10[100] containing organometallic clusters with the same structure (based on the MIL-53(Al) MOF) but with a variety of metals.

To quantify the quality of the ESP produced by various sets of point charges, the Hu-Lu-Yang cost function is used, which gives a weighted quadratic deviation between the reference ESP and the ESP generated by the point charges. The corresponding RMSE value is computed as

$$\mathrm{RMSE}_{\mathrm{ESP}} = \sqrt{\frac{\int w_{\mathrm{HLY}}\left(\mathbf{r}\right)\left(V_{\mathrm{DFT}}\left(\mathbf{r}\right) - \sum_A \frac{q_A}{|\mathbf{r}-\mathbf{R}_A|}\right)^2 \mathrm{d}\mathbf{r}}{\int w_{\mathrm{HLY}}\left(\mathbf{r}\right)\mathrm{d}\mathbf{r}}} \tag{2.9}$$

where $w_{\mathrm{HLY}}\left(\mathbf{r}\right)$ is a weight function that goes to zero both inside the molecule and at larger distances, ensuring that the most relevant part of space around the molecule is sampled. Furthermore, $V_{\mathrm{DFT}}$ is the DFT electrostatic potential.

By construction, Hu-Lu-Yang ESP fitted charges will show the lowest $\mathrm{RMSE}_{\mathrm{ESP}}$ as they are determined precisely to minimize this very cost function. Other ESP fitting methods such as the Merz-Singh-Kollman and the ChelpG scheme show slightly higher errors across the various data sets of molecules. RESP charges on the other hand perform considerably worse: the charges are restrained in order to increase the robustness of the method, but this comes at the prize of poor reproduction of the ESP. The original Hirshfeld density-based partitioning schemes provides charges that do not reproduce the reference ESP very well, compared to the unrestrained ESP fitting methods. The follow-up methods (HI, DDEC, ISA and the novel MBIS) all provide a considerable improvement and are in some cases close to the ESP fitting methods. The QTAIM scheme is consistently the worst, but can be improved tremendously if higher order multipoles are considered, which

is usually the case in force fields based on this method. In correspondence with earlier work, the Hilbert-space partitioning methods (Mulliken, Löwdin and Natural charges) provide a poor description of the ESP.

### Electrostatic interactions

A more direct test for the accuracy of a physically-motivated force field is its ability to predict electrostatic interactions. This can be done using the frozen-density approach: the ab initio densities of isolated monomers are used to compute the Coulomb integral of those densities of the monomers in a dimer configuration. This gives a reference value against which the electrostatic interaction predicted by the force field (for which the charges are derived from the ab initio monomer densities) can be compared. Polarization and charge-transfer effects are not present in the frozen-density approximation, which allows one to focus entirely on the performance of the electrostatic component of force fields.

The sets of molecular dimers used for this investigation have been mentioned before (S66, IHB15, X40 and ZG237). The reference energies range from $-0.06\ \mathrm{kJ\,mol^{-1}}$ to $-664\ \mathrm{kJ\,mol^{-1}}$, indicating that a wide variety of electrostatic interactions are considered. Surprisingly, and perhaps contradicting common belief, ESP-fitted charges do not provide the most accurate electrostatic interaction energies. MBIS point charges for example perform in general better than all of the considered ESP-fitting schemes, although the improvement is rather small. Next to point charges, the MBIS partitioning method also provides a simple expression for the electron density in terms of 1s Slater functions, allowing a simple evaluation of electrostatic interactions including the penetration effect. This offers a consistent and big improvement compared to any set of point charges, resulting in a RMSE that drops by more than $50\,\%$. Only for complexes with very strong electrostatic interactions (more attractive than $-50\ \mathrm{kJ\,mol^{-1}}$), the MBIS model including a valence Slater function does not offer a more accurate description than the MBIS point charges. The reason for this is due to neglecting atomic dipole moments, which is the main source of error in these cases. Still, it can be concluded that a force field aiming to faithfully reproduce the individual components of dimer interaction energies, of which electrostatics is one, has to include a proper model for the penetration effect. MBIS allows this in a natural way by providing an explicit model of the monomer electron density in a minimal basis of Slater functions, balancing accuracy and computational efficiency.

### Robustness

A third criterion to evaluate AIM methods is the robustness of the atomic charges. It is desirable that atomic charges are not overly sensitive to details of the ab initio calculation, such as small changes in geometry or the level of theory. This is required to obtain transferable charges, i.e. charges that are the same for atoms in different molecules but with a similar chemical environment. The robustness with respect to conformational changes was quantified by computing the standard deviation of atomic charges across 103 conformers of the same molecule (PENTA103 data set). It should be stressed that some fluctuation on the charges can be expected, as internal polarization will cause changes in the electron density for the different conformers. However, the standard deviation of the charges from ESP fitting methods is significantly larger than it is for all density-based partitioning schemes that were tested. This confirms previous work from the literature that ESP charges are typically ill-defined because the cost function is rank deficient,[126] which leads to large changes in the charges for small changes in molecular geometry. Mulliken charges are about as sensitive as ESP-fitted charges, while Löwdin and Natural charges are roughly as robust as the density-based partitioning schemes. Another robustness test concerns the basis-set dependence of the metal atom charge across 10 MIL53(M) clusters. Again ESP fitting methods are in general less robust than density-partitioning methods. Mulliken and Löwdin charges are very sensitive to changes in the basis set.

### Pareto analysis

The newly developed MBIS partitioning scheme was compared to other AIM approaches considering three criteria: reproducing the reference ESP, accuracy of dimer electrostatic interactions and robustness of the charges. The ideal AIM method should score well on all criteria (especially the latter two) and therefore the trade-offs between the desired properties are visualized in the Pareto plots in Figure 2.3. Panel (a) shows the error on the interaction energies on the X-axis and the error on the ESP on the Y-axis, with red dots indicating the point charge models on the Pareto front. Point charge models in black are not on the Pareto front and thus suboptimal, in the sense that there are other methods available which perform better for both criteria simultaneously. The MBIS-S model (where the final S indicates that Slater-type functions are used to model the electron density) goes beyond point charges which are used in all other methods, and is therefore not included in the Pareto front. It is clear that there is some correlation between the error on the interaction energies and the error on the ESP. The correlation

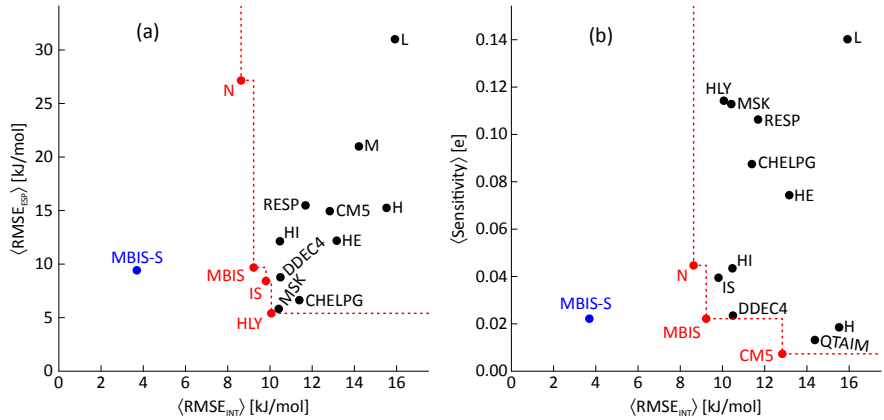



**Figure 2.3:** Pareto plots showing the trade-offs between different desirable properties of AIM methods. Reprinted with permission from Ref. 100. Copyright (2016) American Chemical Society.

is however far from perfect, as for example IS, MBIS and Natural charges predict interaction energies more accurately than Hu-Lu-Yang ESP fitted charges, despite the latter showing the smallest deviation with respect to the reference ESP. As discussed earlier, including the Slater-type functions in the MBIS density results in a tremendously lower error on the interaction energies, which explains why the MBIS-S method is situated left of the Pareto front.

Panel (b) of Figure 2.3 shows the Pareto plot with the error on the interaction energies on the X-axis and the sensitivity (the inverse of robustness) on the Y-axis. Again MBIS is on the Pareto front, which means that there is no other method included in the test which can improve one desired property of an AIM scheme without deteriorating the other. Together with the success of the MBIS density to model the electrostatic penetration effect, this provides evidence that MBIS is a proper starting point to construct a physically-motivated noncovalent force field, which will be done in the following section.

## 2.2 Development of the monomer electron density based noncovalent force field

The introductory chapter concluded with a list of desired properties of a noncovalent force field. This wish list will now be filled in more concretely in order to explain the novel methodology for the construction of a noncovalent force field that was developed during this thesis. As a basic ingredient,

the monomer electron density (partitioned with MBIS) will be used and hence the force field is named Monomer Electron Density based Force Field (MEDFF). In first instance (Section 2.2.1), the MEDFF energy expression will be presented, which consists of four components, each one corresponding to a component of the SAPT decomposition of interaction energies. This ensures that each term of the force field, which features at most one so-called interaction parameter, can be physically motivated. Additionally, the interaction parameter for each component can be fitted separately which avoids a difficult minimization in a high-dimensional parameter space, as demonstrated in Section 2.2.2. In second instance, all interaction parameters (three in total) are recalibrated using very accurate CCSD(T)/CBS interaction energies for a database containing $66$ dimers and a first internal and external validation concerning dimer interaction energies is presented.

The following discussion of the MEDFF methodology is largely based on **Paper II [Ref. 75]**, which features additional details and supplementary results.

### 2.2.1 Motivating the functional form of the MEDFF energy expression

A proper starting point to develop and test a noncovalent force field, is the study of dimer interaction energies in the frozen-monomer approach. This means the internal monomer geometry is not altered in the dimer, so the covalent energy is constant and this allows complete focus on noncovalent interactions. Note that the reverse is not possible: when developing a covalent force field, variations of monomer geometries need to be considered which will at the same time lead to changes in the noncovalent energy. The logical sequence is thus to first develop a noncovalent force field and afterwards derive a covalent part on top of the previously derived noncovalent terms. This idea is also incorporated in the QuickFF methodology,[57, 58] which is a prime candidate to derive covalent force fields from ab initio input complementing MEDFF. The mentioned approach has been followed to derive covalent force fields for alkanes and alkenes on top of MEDFF, which was required for the simulations discussed in Section 2.3.

Below, the MEDFF expression for the interaction energy between two monomers will be provided, with the ab initio molecular ground-state electron density of the separated, isolated monomers as input. The MBIS partitioning scheme is applied to these monomer electron densities, which provides a compact approximation $\tilde{\rho}_1(\mathbf{r})$ of the ab initio density $\rho_1(\mathbf{r})$ of

monomer 1 with $N^1_{\text{atoms}}$ as follows:

$$\tilde{\rho}_1(\mathbf{r}) = \sum_{A=1}^{N^1_{\text{atoms}}} \frac{N_{A,v}}{8\pi\sigma^3_{A,v}} \exp\left(-\frac{|\mathbf{r}-\mathbf{R}_A|}{\sigma_{A,v}}\right) \tag{2.10}$$

In this approximation, the original MBIS density has been simplified by condensing the core Slater functions onto the nucleus as illustrated in Figure 2.2 for the case of $CO_2$. For force-field applications, only one Slater function (with valence population $N_{A,v}$ and valence width $\sigma_{A,v}$) is retained for every atom and the effective core charge $q_{A,c}$ is given by:

$$q_{A,c} = Z_A - \sum_{i=1}^{m_A-1} N_{Ai} \tag{2.11}$$

It has been shown that empirical force fields rely to a large extent on compensation of errors between different energy components.[127] For example, if dispersion interactions are underestimated (compared to an ab initio decomposition scheme) while electrostatics are overestimated by the same amount, the total energy will still be in good agreement with the reference. Such a cancellation of errors is however most often fortuitous rather than physically inspired, limiting the transferability of these force fields. An approach to avoid this problem is to construct a force field in close correspondence with an ab initio energy decomposition scheme, in order to guarantee that each force-field component has a physically sound foundation. There is a large number of available decomposition schemes, many of them rooted in either the supermolecular, variational approach by Kitaura and Morokuma[128, 129] or in the direct, perturbative approach named Symmetry-Adapted Perturbation Theory (SAPT).[86] As there are no quantum mechanical operators corresponding to components of the interaction energy, such a decomposition is always ambiguous and schemes need to be judged for instance on their correspondence to chemical intuition, their accuracy for total interaction energies, and their computational cost. Based on these considerations, Hartree-Fock SAPT2+(3) (where 2+(3) refers to the order of the perturbation expansion that is included) as implemented in PSI4[130, 131] is used below for the development of a new noncovalent force field. The SAPT intermolecular energy is expressed as the following sum of components:

$$E_{\text{inter}} = E_{\text{elst}} + E_{\text{exch-rep}} + E_{\text{disp}} + E_{\text{ind,ct}} \tag{2.12}$$

The behavior of these different components as well as the total interaction energy along the dissociation curve is shown in Figure 2.4 for two complexes: a hydrogen-bonded water dimer on the left and the $\pi - \pi$ stacked benzene dimer on the right.

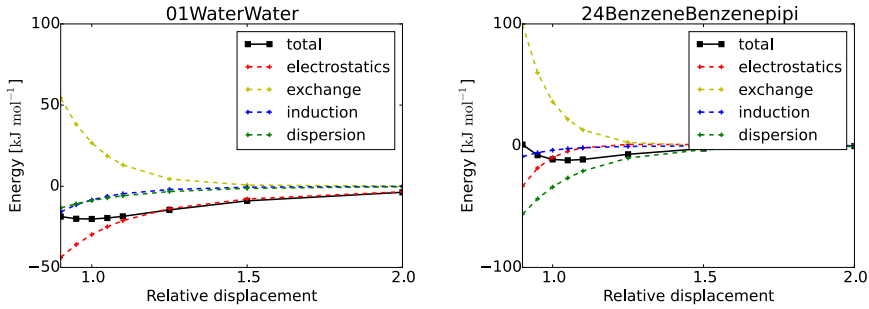



**Figure 2.4:** Dissociation curves showing components of the SAPT energy for the water dimer (left, hydrogen-bonded) and $\pi - \pi$ benzene dimer (right, dispersion-dominated). Reprinted with permission from Ref. 75. Copyright (2017) American Chemical Society.

The components of the noncovalent force field which will be developed show a one-to-one correspondence to the SAPT components. These force-field terms, electrostatics, exchange-repulsion, dispersion, and induction, will be discussed subsequently in the following subsections.

### Electrostatics

The first component describes the frozen-density electrostatic interaction and is calculated as the Coulomb integral of the monomer charge distributions. Given the force-field model of the electron density (eq 2.10) and the effective-core point charges located at the nuclei (eq 2.11), the force-field electrostatic energy is computed as:

$$E_{\text{elst}} = \sum_{A=1}^{N_{\text{atoms}}^1} \sum_{B=1}^{N_{\text{atoms}}^2} E_{\text{elst}}^{AB} \tag{2.13}$$

$$\begin{aligned} E_{\text{elst}}^{AB} = & \frac{q_{A,c} q_{B,c}}{R_{AB}} \\ & - \frac{q_{A,c} N_{B,v}}{R_{AB}} \left[1 - g\left(\sigma_{B,v}, R_{AB}\right)\right] \\ & - \frac{N_{A,v} q_{B,c}}{R_{AB}} \left[1 - g\left(\sigma_{A,v}, R_{AB}\right)\right] \\ & + \frac{N_{A,v} N_{B,v}}{R_{AB}} \left[1 - f\left(\sigma_{A,v}, \sigma_{B,v}, R_{AB}\right) - f\left(\sigma_{B,v}, \sigma_{A,v}, R_{AB}\right)\right] \end{aligned} \tag{2.14}$$

where the functions $g$ and $f$ are

$$g(\sigma, r) = \left(1 + \frac{r}{2\sigma}\right) \exp\left(-\frac{r}{\sigma}\right) \tag{2.15}$$

$$f(\sigma_i, \sigma_j, r) = \frac{\sigma_i^4}{\left(\sigma_i^2 - \sigma_j^2\right)^2} \left[1 + \frac{r}{2\sigma_i} - \frac{2\sigma_j^2}{\sigma_i^2 - \sigma_j^2}\right] \exp\left(-\frac{r}{\sigma_i}\right) \tag{2.16}$$

This expression for the electrostatic energy between atoms $A$ and $B$ can be simplified by introducing the net atomic charges , $q_A = q_{A,c} - N_{A,v}$ and $q_B = q_{B,c} - N_{B,v}$:

$$E_{\text{elst}}^{AB} = \frac{q_A q_B}{R_{AB}} + E_{\text{penetration}}^{AB} \tag{2.17}$$

The first term simply provides the electrostatic interaction between point charges, so the second term (which decays exponentially as the atoms are separated) describes the penetration effect due to the use of a distributed model for the electron density. The above expression, as well as all other terms of MEDFF, are implemented in the in-house Yaff force-field code.

### Exchange-repulsion

The second term represents exchange-repulsion, which is a quantum mechanical effect mostly arising from the Pauli exclusion principle and is therefore also referred to as the Pauli repulsion energy. In the context of dimers, it is present when the electron densities of the monomers overlap and can thus also be termed the steric repulsion energy. The exchange-repulsion energy cannot be simply expressed as a functional of the isolated monomer densities, however many approximate models have been proposed in the literature.[83, 90, 132, 133] Figure 2.5 illustrates two such models for a rather simple case: the helium dimer. The first model ($K_0 S_\rho$) assumes that the exchange-repulsion energy is proportional to the density overlap of isolated helium atoms. A least squares fit to the data points for this specific case results in $K_0 = 9.7$. The second model ($K_1 S_\Psi^2$) proposes an expression proportional to the square of the wave function overlap of the isolated helium atoms, with $K_1 = 1.2$ in this case. It is clear that while both models correlate with the exchange-repulsion energy, the first model allows a much better correspondence to the reference data.

In MEDFF it is assumed that the exchange-repulsion energy is simply proportional to the overlap of the monomer electron densities:

$$E_{\text{exch-rep}} = U_{\text{exch-rep}} \int d\mathbf{r} \tilde{\rho}_1(\mathbf{r}) \tilde{\rho}_2(\mathbf{r}) \tag{2.18}$$

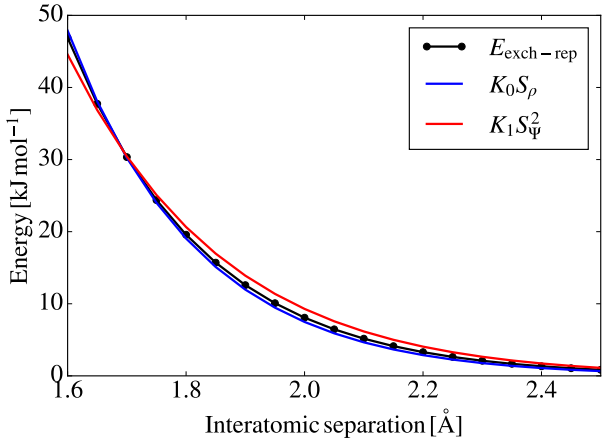



**Figure 2.5:** The exchange-repulsion energy (computed using SAPT2+(3)/aug-cc-pvtz) of the helium dimer is compared to two overlap models. $S_\rho$ denotes the overlap of densities, $S_\Psi^2$ denotes the square of the wave function overlap, where in both cases the isolated atoms are computed using HF/aug-cc-pvtz.

This energy can be expressed as a sum over pairs of atoms by making use of the force-field model for the electron density in eq 2.10.

$$E_{\text{exch-rep}} = \sum_{A=1}^{N^1_{\text{atoms}}} \sum_{B=1}^{N^2_{\text{atoms}}} E^{AB}_{\text{exch-rep}} \tag{2.19}$$

$$E^{AB}_{\text{exch-rep}} = U_{\text{exch-rep}} S^{AB} \tag{2.20}$$

$$S^{AB} = \frac{N_{A,v} N_{B,v}}{8\pi R_{AB}} \left[ h(\sigma_{A,v}, \sigma_{B,v}, R_{AB}) + h(\sigma_{B,v}, \sigma_{A,v}, R_{AB}) \right] \tag{2.21}$$

where the function $h$ equals

$$h(\sigma_i, \sigma_j, r) = \left[ \frac{4\sigma_i^2 \sigma_j^2}{\left(\sigma_j^2 - \sigma_i^2\right)^3} + \frac{\sigma_i}{\left(\sigma_j^2 - \sigma_i^2\right)^2} r \right] \exp\left(-\frac{r}{\sigma_i}\right) \tag{2.22}$$

A crucial difference with most existing models, is that the prefactor $U_{\text{exch-rep}}$ is independent of the atoms and even the molecules involved. To clearly distinguish such a parameter from atomic or atom-pair specific parameters, it is called an *interaction* parameter of the MEDFF model. The interaction parameter $U_{\text{exch-rep}}$ is the proportionality factor that relates the total overlap of the monomers to the total exchange-repulsion interaction energy. In Section

2.2.2 it will be shown that a universal value of $U_{\text{exch-rep}}$ can be determined which predicts fairly accurate exchange-repulsion interaction energies for many dimers included in a database.

### Dispersion

London dispersion forces or simply dispersion interactions are another example of a quantum phenomenon without a direct classical analogue. It arises from the correlation between the spatial parts of electronic wave functions, due to Coulombic repulsion. When two molecules are far apart (the overlap between their wave functions is negligible), the dispersion interaction energy can be obtained using perturbation theory and can be written as the following function of the intermolecular separation distance $R$:[22]

$$E_{\text{disp}} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - \dots \tag{2.23}$$

This expansion can be interpreted semi-classically as the electrostatic interaction between instantaneous multipoles induced by fluctuations in the electron densities. Such an interpretation is however not fundamentally correct and additionally rather complex as the basis for a force-field model. A simpler model is adopted in MEDFF, by only retaining the leading two terms of the perturbation expansion and introducing a damping function to modify the behavior at short intermolecular separation distances:

$$E_{\text{disp}} = \sum_{A=1}^{N^1_{\text{atoms}}} \sum_{B=1}^{N^2_{\text{atoms}}} E^{AB}_{\text{disp}} \tag{2.24}$$

$$E^{AB}_{\text{disp}} = -f_6\left(x_{AB}\right) \frac{C^6_{AB}}{R^6_{AB}} - f_8\left(x_{AB}\right) \frac{U_{s8} C^8_{AB}}{R^8_{AB}} \tag{2.25}$$

$$f_n\left(x_{AB}\right) = 1 - \left( \sum_{k=0}^{n} \frac{x^k_{AB}}{k!} \right) \exp\left(-x_{AB}\right) \tag{2.26}$$

The Tang–Toennies damping function[134] (with $x_{AB} = \frac{R_{AB}}{\left(\sigma_{A,v}+\sigma_{B,v}\right)/2}$) is chosen and another interaction parameter $U_{s8}$ is introduced. This parameter attempts to fix some deficiencies of the proposed energy expression, such as the truncation of the perturbation expansion, short-range effects that are not properly described by the damping function, and approximations used to determine $C^8$ coefficients. The value for $U_{s8}$ will be determined in Section 2.2.2 by parametrizing the model based on SAPT dispersion energies.

In order to complete the MEDFF dispersion model, the coefficients $C^6_{AB}$ and $C^8_{AB}$ need to be determined for each pair of atoms $A$ and $B$. The Tkatchenko-

Scheffler[16] method allows computing values for the $C^6$ coefficients by modifying free-atom values (which are computed using Time-Dependent DFT) based on the AIM partitioning of the isolated monomer densities.

$$C_{6AA}^{\text{eff}} = \left(\frac{V_A^{\text{eff}}}{V_A^{\text{free}}}\right)^2 C_{6AA}^{\text{free}} \tag{2.27}$$

$$C_{6AB} = \frac{2C_{6AA}C_{6BB}}{\left[\frac{\alpha_B^0}{\alpha_A^0}C_{6AA} + \frac{\alpha_A^0}{\alpha_B^0}C_{6BB}\right]} \tag{2.28}$$

where $\alpha^0$ are static atomic polarizabilities, $V$ are atomic volumes (computed as the expectation value of $r^3$ times the atomic density), and the superscripts *eff* and *free* refer to atoms in molecules and isolated atoms respectively. The MBIS partitioning is used to determine the AIM densities, as this has been shown to be a reliable procedure to produce molecular $C^6$ coefficients.[100] The $C^8$ coefficients finally are determined based on the $C^6$ coefficients using a recursive relation proposed by Starkschall et al.[135]

$$C_{AB}^8 = \frac{3}{2}C_{AB}^6\sqrt{\frac{\langle r_A^4\rangle}{\langle r_A^2\rangle} + \frac{\langle r_B^4\rangle}{\langle r_B^2\rangle}} \tag{2.29}$$

In this case, the expectation values of powers of $r_A$ and $r_B$ are evaluated using the free-atom densities.

### Induction

The final contribution to the MEDFF energy (and also the final component of the SAPT energy decomposition scheme) is the induction energy. The interpretation of the SAPT induction energy is not straightforward, and more specifically the separation into contributions from polarization and charge-transfer effects to this component is debatable.[136, 137] Polarization refers to the classical effect where charge densities are deformed in the presence of other charge densities, resulting in electrostatic interactions that are not included when considering unperturbed charge densities (as done in the frozen-density approach used to compute the electrostatic component). This term should in principle be modeled using a polarizable force field, employing for instance Drude oscillators or inducible dipoles, and many such methods are available in the literature.[70, 138–147] Further investigation of some polarizable force-field models revealed that none of them could be incorporated in MEDFF without altering the design philosophy. Because polarization additionally increases the algorithmic complexity of the force-field energy

calculation, a drastic simplification is employed and the MEDFF induction energy is assumed to be proportional to the overlap of the monomer densities:

$$E_{\text{ind,ct}} = \sum_{A=1}^{N^1_{\text{atoms}}} \sum_{B=1}^{N^2_{\text{atoms}}} E^{AB}_{\text{ind,ct}} \tag{2.30}$$

$$E^{AB}_{\text{ind,ct}} = -U_{\text{ind,ct}} S^{AB} \tag{2.31}$$

This can in part be motivated by the observation that charge-transfer energies, which most likely make an important contribution to the SAPT induction energy, decay exponentially with increasing separation distance,[85, 87, 148–152] just as the overlap of the monomer densities falls off exponentially. It can be noted that the MEDFF induction model is exactly the same as the exchange-repulsion model (apart from a minus sign which is merely introduced to obtain positive values for all interaction parameters), in line with an earlier observation that the exchange part of the charge-transfer contribution to SAPT is approximately proportional to the SAPT exchange-repulsion energy.[153]

### 2.2.2 Calibration of the interaction parameters

The MEDFF expression for the total intermolecular energy consists of four components, representing electrostatic, exchange-repulsion, dispersion, and induction contributions. The analytic function for each component was introduced in the previous section. In order to complete the model, values for the three interaction parameters need to be determined, which is the subject of this section.

The aim of MEDFF is to reproduce dimer dissociation curves from high-level ab initio calculations, and therefore the MEDFF interaction parameters will be determined using CCSD(T)/CBS interaction energies as reference in a later paragraph. First however, each of the four components of MEDFF will be compared separately with the corresponding SAPT contribution in order to validate the proposed energy expressions and obtain more insight into strengths and weaknesses of the proposed force field. In the current section, only an internal validation is performed, meaning that the reproduction of data that are used in the construction of MEDFF is investigated. In later Sections 2.2.2 and 2.3, an external validation will be performed. Finally in Sections 3.1 and 3.2 MEDFF will also be used to investigate an interesting application (gas adsorption in MOFs) which lies outside of the scope of systems that are used in the current chapter for the construction of the MEDFF methodology.

Next to picking an appropriate level of theory used to generate reference data, the choice of the data set employed for the parameter calibration is a crucial component of the force-field derivation. The S66x8[123] database was chosen from the literature, which contains 66 dimer complexes consisting of monomers such as acetic acid, benzene, neopentane, uracil, and water. For each of the 66 complexes, 8 displacements along the dissociation curve are considered. The geometries of the monomers are kept rigid in order to maintain a clear separation between intermolecular and intramolecular interactions. The S66x8 database was constructed in such a way that it provides a balanced representation of the noncovalent interactions commonly found in biomolecules. More specifically a classification into three categories is made: 23 hydrogen-bonded complexes, 23 dispersion-dominated complexes, and 20 "other" complexes (defined as those that do not pertain to either of the two first categories) are included in the S66x8 database. Finally it should be mentioned that all complexes consist of light elements (hydrogen, carbon, nitrogen and oxygen) and contain at most 34 atoms. This makes it computationally feasible to perform the necessary SAPT and CCSD(T)/CBS calculations which are used as reference data in the following two subsections.

### Separate fitting to SAPT components

As mentioned earlier, the Hartree-Fock SAPT2+(3) variant of symmetry-adapted perturbation theory is chosen as it provides, for the current purposes, the correct balance between computational tractability and correspondence of total interaction energies with CCSD(T)/CBS values. The first component to be discussed is the electrostatic component.

The MEDFF electrostatic interactions are simply computed as the Coulomb integral of the MBIS representation of monomer electron densities (which are computed using B3LYP/aug-cc-pVTZ for the monomers of the S66x8 database), without the need to introduce any fitting parameters. A parity plot comparing SAPT and MEDFF electrostatic energies is provided in Figure 2.6 (a). As the S66x8 database features 8 displacements for each of the 66 complexes, a total of $8 \times 66 = 528$ dimer configurations is considered and each of these configurations is represented as one datapoint in the parity plot. Overall, the performance of MEDFF seems satisfactory over a wide range of interaction strengths encountered in the database and irrespective of the class of complexes (hydrogen-bonded, dispersion-dominated or other complexes, which are indicated with a color code) that is considered. This confirms earlier findings reported in the validation of the MBIS methodology in Section 2.1.3. When the MEDFF electrostatic model is replaced with more

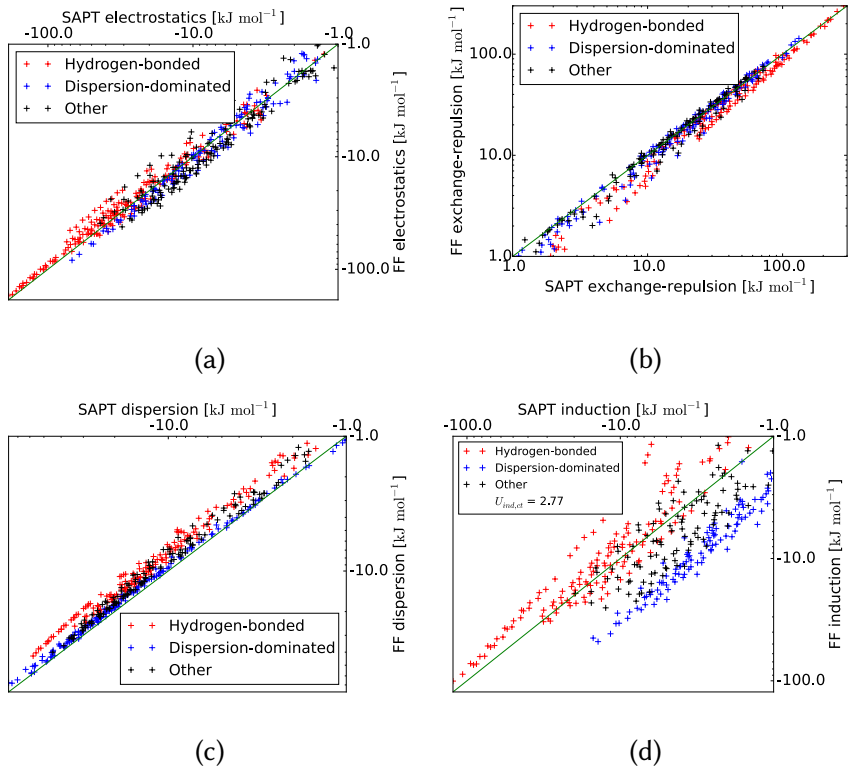



**Figure 2.6:** Correlation plots of SAPT and force-field energy components for the complete S66x8 database. Note the use of logarithmic scales. Reprinted with permission from Ref. 75. Copyright (2017) American Chemical Society.

conventional methods such as point-charge interactions between RESP fitted partial charges, deviations with the SAPT values are considerably larger. This is mostly notable for very strong electrostatic interactions (for which the penetration effect is crucial) and for the dispersion-dominated dimers.

The proposed MEDFF model to describe the exchange-repulsion energy is a term proportional to the overlap of the monomer densities. The overlap itself only depends on monomer properties: it can be calculated using the MBIS parameters resulting from the partitioning of the monomer electron densities. The proportionality factor, $U_{\text{exch-rep}}$ as defined in eq 2.18, however is related specifically to the interaction of two monomers and has to be derived from dimer calculations. Here it is determined such that the mean-square deviation between force-field and SAPT exchange-repulsion energies is minimized (only dispersion-dominated complexes were considered, as

motivated in the next paragraphs). As the MEDFF model is linear in the parameter $U_{\text{exch-rep}}$, its value can be computed by linear regression which gives $U_{\text{exch-rep}} = 8.13$ in atomic units. The resulting scatter plot in Figure 2.6 (b) shows that the proposed model is adequate for dispersion-dominated and "other" complexes, employing only one parameter that is fitted to the reference data. The performance for hydrogen-bonded complexes is somewhat worse, but it turns out that this is partially compensated, through fortuitous cancellation of an error with the opposite sign encountered in the electrostatic component. Obviously the correspondence between SAPT and MEDFF exchange-repulsion energies could be improved tremendously by determining separate values for $U_{\text{exch-rep}}$ for each complex individually. Such an approach would however lead to many fitting parameters, introduce the need for additional dimer calculations for each new system, and violate the design principles of MEDFF. As a final note on the exchange-repulsion contribution, it can be mentioned that Söderhjelm et al. found a proportionality factor of $8.6$ between exchange-repulsion energy and density overlap.[154] Given the many differences in computational details (a different database of dimers representing functional groups in proteins is used, reference exchange-repulsion energies are computed differently, and the density overlap is calculated by numerical integration), this is not too far from the MEDFF value $U_{\text{exch-rep}} = 8.13$.

Similar to the exchange-repulsion model, the MEDFF expression for the dispersion interaction contains one interaction parameter, $U_{s8}$, which can not be obtained solely from the properties of the monomers involved. This parameter scales the term of eq 2.25 that is proportional to $R^{-8}$ and can thus also be interpreted as a rescaling of all interatomic $C^8_{AB}$ coefficients. As the dispersion energy is a linear function of the parameter $U_{s8}$ it can once again be determined from a linear regression, this time with reference energies obtained as the SAPT dispersion component. If this linear regression is performed for each complex of the S66x8 database separately, the obtained interaction parameters vary over a rather large range which means that a universal value for $U_{s8}$ applicable for all complexes would not provide proper results. Within the class of dispersion-dominated complexes however, the value $U_{s8} = 0.57$ is an acceptable estimate which can be used for all dimers of this class. The scatter plot obtained using this value is shown in 2.6 (c) and clearly for the dispersion-dominated dimers (blue symbols) there is a good agreement between SAPT and MEDFF dispersion energies. The dispersion energy for the hydrogen-bonded complexes is systematically underestimated (in absolute value) by MEDFF.

The MEDFF induction features the same expression as the MEDFF exchange-repulsion. The corresponding interaction parameter, $U_{\text{ind,ct}}$, can therefore be

determined in the same way as earlier by fitting to the SAPT induction component instead of the SAPT exchange-repulsion component. Figure 2.6 (d) reveals that this is not an appropriate approach to cover the entire database of dimers. Apparently, induction is a much more complex phenomenon than the simple overlap-based model assumed in MEDFF. A more thorough investigation is required, probably including polarization effects explicitly. An important consideration in this respect is that the SAPT induction energies might need to be further decomposed. It has been argued that charge-transfer and (classical) polarization effects, which both are part of the SAPT induction energy, show a fundamentally different behavior and methods aiming to separate these contributions have been developed.[136, 137] Such a further decomposition of the reference interaction energies and the introduction of corresponding additional MEDFF terms could indeed lead to a better induction model. Another approach would be to focus on some particular dimers which show up as outliers. For instance, it turns out that especially double hydrogen-bonded complexes show large deviations between SAPT and MEDFF induction energies, which can be seen as a line of red markers above the diagonal on the lower left corner of Figure 2.6 (d). This suggests that including an additional term in MEDFF that explicitly models hydrogen bonds could offer a better correspondence with SAPT induction energies.

During the development of MEDFF, several more complicated models for the induction energy were tested for the S66x8 database, such as inducible dipoles, a hydrogen bond term similar to the one used in QMDFF,[155] more complicated functions of the overlap (instead of the currently used proportionality), prefactors dependent on the atomic number, … Overall, these modifications resulted in an improved correspondence to the reference data. It was however judged that, within the spirit of MEDFF, the gains in accuracy were not outweighed by the additional complexity of the models and the increased number of parameters. This is the reason why the rather simple induction model was retained. More accurate refinements of MEDFF will certainly need to improve the current induction model, and the previous discussion provides some pointers about how this might be achieved.

Fortunately, the MEDFF procedure adopted thus far can be completed by considering only dispersion-dominated complexes. Indeed, Figure 2.7 shows that if $U_{\text{ind,ct}} = 0.87$, an acceptable agreement between MEDFF and SAPT induction energies is obtained. This finding is the reason why the fitting of all interaction parameters was done based on reference data only for the dispersion-dominated complexes.

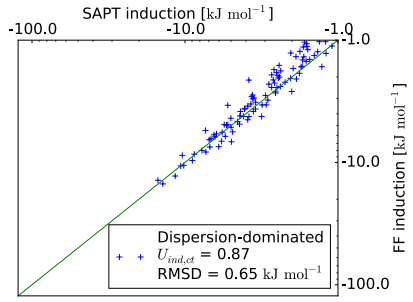



**Figure 2.7:** Correlation plot of SAPT and MEDFF induction energies for the dispersion-dominated complexes of the S66x8 database.

### Constrained recalibration with respect to CCSD(T)/CBS energies

In the previous two sections, the MEDFF expression for the interaction energy was presented and values for the three interaction parameters ($U_{\text{exch-rep}}$ appearing in the exchange-repulsion energy, $U_{s8}$ appearing in the dispersion energy, and $U_{\text{ind,ct}}$ appearing in the induction energy) were determined by linear regression using the corresponding SAPT energy contributions for a database of $23$ dispersion-dominated complexes, considering $8$ displacements for each complex. The advantage of fitting the interaction parameters term by term is that the proposed energy expressions could be validated and the force field can be motivated from a physical point of view. This advantage is however at the same time also a weak point for the following reason. The different SAPT components cancel each other to a large extent, leading to a situation where the sum (which is the total interaction energy) is much smaller in absolute value than each of the contributions for configurations near the equilibrium. This behavior is illustrated in Figure 2.4 for the dissociation curves of a water dimer (left) and a benzene dimer (right). The downside is thus that a small relative error on one of the components, will result in a large deviation for the total interaction.

In order to tackle this problem, the three MEDFF interaction parameters are recalibrated simultaneously in order to obtain a better correspondence with ab initio total interaction energies. The prediction of total interaction energies will improve by making use of a cancellation of errors present in the different energy terms. In order to do this in a controlled fashion and still retain some information from the individual fitting of interaction parameters, the recalibration will be done using ridge regression. The basic idea of ridge regression, also termed Tikhonov regularization, is the addition of a term to the cost function that restrains the fitting parameters to predefined

values. The use of such prior information about the parameters is similar to techniques used in Bayesian statistics.[156–158] In the case presented here, the cost function will contain on the one hand the sum of quadratic deviations of total interaction energies and on the other hand the sum of quadratic relative deviations of the interaction parameters from their values determined from fitting individual components to SAPT reference data.

It is not a priori clear what the relative contribution of these two terms to the total cost function should be, and therefore the contribution of the term that restrains the interaction parameters to their SAPT values was varied from very large to very small. It turns out that the estimated prediction error of the interaction energies initially decreases significantly by allowing the interaction parameters to deviate slightly from their SAPT values. Further relaxation of this restraint however does not result in a further decrease of the estimated prediction error, which signifies the presence of soft or floppy modes in the parameter space: it is possible to change a combination of interaction parameter values over a wide range without having a serious impact on the resulting total interaction energies. The best choice of the final parameters is obtained at the point where the parameter values are closest to their SAPT values, but still provide a low estimated prediction error. The final values that will be used further on are $U_{\text{exch-rep}} = 8.43$, $U_{s8} = 0.57$, and $U_{\text{ind,ct}} = 0.86$ (all in atomic units), which seems rather close to the values obtained from fitting to individual components of the SAPT interaction energy, which resulted in $U_{\text{exch-rep}} = 8.13$, $U_{s8} = 0.57$, and $U_{\text{ind,ct}} = 0.87$. This completes the construction of the MEDFF force field, and before proceeding to a validation on systems outside of the database used in the parameter fitting, it is mentioned that the root-mean square deviations between MEDFF and CCSD(T)/CBS interaction energies for the dispersion-dominated complexes in the S66x8 database range from $3.7\,\text{kJ}\,\text{mol}^{-1}$ at closest intermolecular distances, over $1.5\,\text{kJ}\,\text{mol}^{-1}$ at equilibrium dimer geometries, to $0.2\,\text{kJ}\,\text{mol}^{-1}$ at long-range separations.

### External validation on the hsg data set

One of the guiding principles in the design of MEDFF, was that it should be a robust method that is transferable to chemical systems that were not present in the database used to determine the interaction parameters. In order to test whether this objective has been fulfilled, the performance of MEDFF for the dispersion-dominated complexes of the hsg data set[159] is investigated. This means 13 dimers extracted from a system where indinavir is bound to an HIV-II protease crystal structure are considered. The values of the interaction parameters are not reoptimized, but simply take on the values

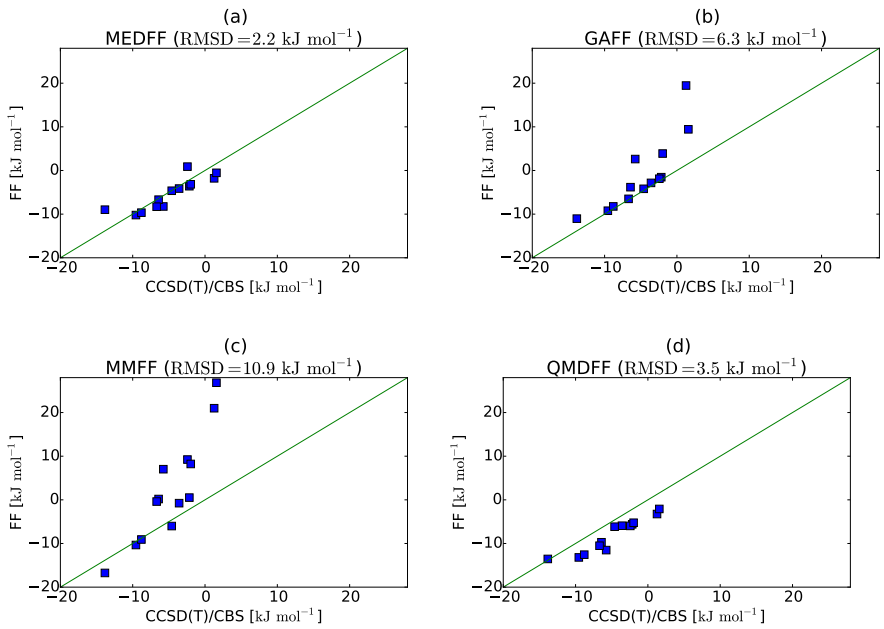



**Figure 2.8:** Performance of several force fields in reproducing CCSD(T)/CBS interaction energies for the dispersion-dominated complexes of the hsg data set. Reprinted with permission from Ref. 75. Copyright (2017) American Chemical Society.

as determined in the previous section. This means that the only ab initio calculations necessary to compute MEDFF interaction energies for these new dimers are DFT calculations performed on the isolated monomers, which poses a computational cost that is similar to what is required to construct force fields that use partial charges fitted to ab initio ESPs. The performance of MEDFF for the considered part of the hsg data set is shown in Figure 2.8 and compared to other force fields from the literature: the Generalized Amber Force Field[160] (GAFF) and the Merck Molecular Force Field[161] (MMFF), which can be considered generic force fields, and the Quantum Mechanically Derived Force Field[155] (QMDFF), which is based on ab initio calculations and shares some features with MEDFF. None of these force fields was fitted explicitly to the systems considered, so this approach should provide a somewhat fair comparison. MEDFF clearly shows the best correspondence to the CCSD(T)/CBS interaction energies, but QMDFF is only very slightly worse. The correlation with the reference energies is markedly worse for GAFF and MMFF. This first external validation provides some evidence that the MEDFF approach is indeed a proper way to construct robust noncovalent

force fields, using only a small amount of fittable parameters and featuring energy expressions that do have a clear physical interpretation, at a total computational cost that allows scaling to larger systems.

## 2.3 Application of MEDFF to alkanes and alkenes in gas and condensed phase

Further tests of MEDFF pertain to the reproduction of macroscopic experimental properties of small alkanes and alkenes, first in the gas phase and later in the condensed phase. As MEDFF was constructed without any empirical input from such experiments, this can be considered a challenging procedure to establish whether or not MEDFF is suited to predict experimental properties. Alkanes and alkenes are chosen as these can be considered to form dispersion-dominated complexes, the class for which the MEDFF interaction parameters were calibrated. It could be argued that single-component alkane or alkene simulations are not the most "exciting" ones nowadays. Indeed, force fields that provide good agreement with experimental liquid densities of methane, ethane, propane, ... were established in the 1980s.[162] Yet, these seemingly simple systems can provide fundamental insight into differences between ab initio derived and empirical force fields, as will become clear in the following Sections 2.3.1 (gas-phase simulations) and 2.3.2 (condensed-phase simulations).

### 2.3.1 Reproduction of second virial coefficients

The deviation of a real gas from an ideal gas can be expressed using the virial equation of state, which relates the pressure $P$ to a power series in the density $\rho$: $P = k_B T \left[\rho + B_2(T)\rho^2 + B_3(T)\rho^3 + \ldots\right]$. The second virial coefficient, $B_2(T)$, represents the initial deviation from ideal-gas behavior and only depends on the interaction between two molecules:[163]

$$B_2(T) = 2\pi \int \left(1 - \langle \exp\left[-\beta U_{\text{inter}}\left(r\right)\right]\rangle\right) r^2 dr \tag{2.32}$$

As demonstrated in the previous section, MEDFF provides $U_{\text{inter}}$ values for dispersion-dominated dimers in reasonable agreement with CCSD(T)/CBS values. Figure 2.9 shows that the MEDFF estimation for $B_2(T)$ is comparable to experimental values for the investigated small alkanes (methane, ethane, propane, and butane) and alkenes (ethene and propene). For the larger molecules propene, propane, and butane the experimental values are slightly underestimated (the absolute value of the MEDFF second virial coefficient is

larger than experiment) which means that MEDFF is in general overbinding in these cases.

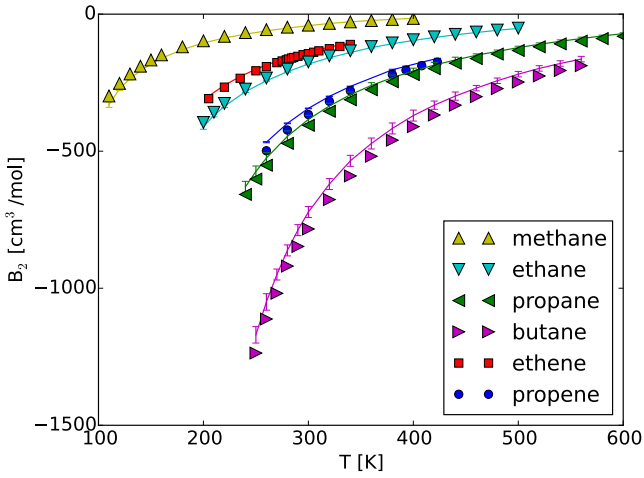



**Figure 2.9:** Comparison of second virial coefficients from our force field (symbols) with experiment (solid line). Experimental data for alkanes from Ref. 164, for propene from Ref. 165, 166 and for ethene from Ref. 167. Reprinted with permission from Ref. 75. Copyright (2017) American Chemical Society.

It should be stressed that these MEDFF results are obtained without empirical input and without fitting specifically to dimers of the molecules involved (except for ethene which happens to be in the S66x8 database). The necessary atomic parameters are obtained from monomer calculations of the alkanes and alkenes, while the three interaction parameters retain the values that were determined earlier by calibration with respect to CCSD(T)/CBS dimer interaction energies. This should be contrasted to approaches typically used in the literature, where parameters for ab initio force fields can only be obtained by performing expensive calculations of interaction energies for a large set of comlexes of the molecules of interest. The good performance of MEDFF for the test performed here can therefore be considered remarkable and further proves that the suggested approach can be used to generate useful noncovalent force fields.

### 2.3.2 Reproduction of liquid densities

Inspired by the good reproduction of experimental second virial coefficients, MEDFF is further used to simulate the same molecules in the condensed

phase. The liquid density and enthalpy of vaporization are calculated at a pressure of $1$ atm and temperatures just below the respective boiling points of methane, ethane, propane, butane, ethene and propene. These properties are obtained from MD simulations of the liquids in the NPT ensemble and of a single molecule in the NVT ensemble (the latter are only necessary to obtain enthalpies of vaporization). The results are compared to experimental values in Table 2.1, and the agreement is in general rather poor. The densities are systematically overestimated with errors ranging from $+11\%$ for ethane up to $+20\%$ for ethene. The enthalpies of vaporization are also overestimated for all molecules that were studied, with errors ranging from $+12\%$ for ethane up to $+23\%$ for butane. This means that MEDFF predicts interactions that are overly attractive, when compared to experiment, in the condensed phase. This was however not observed in the previous tests, indicating that some complications arise in condensed-phase simulations using an ab initio derived force field, warranting a more extensive discussion as provided in the next Section 2.4.

**Table 2.1:** Comparison of simulated and experimental liquid density $\rho$ and enthalpy of vaporization $\Delta H_{\text{vap}}$. Experimental data for methane, ethane, propane and butane are taken from Ref. 168, for ethene from Ref. 169 and for propene from Ref. 170. Reprinted with permission from Ref. 75. Copyright (2017) American Chemical Society.

| Molecule | T [K] | $\rho$ [$\mathrm{g\,l^{-1}}$] | | $\Delta H_{\text{vap}}$ [$\mathrm{kJ\,mol^{-1}}$] | |
|---|---|---|---|---|---|
| | | Exp. | Sim. | Exp. | Sim. |
| methane | 112 | 423 | 503 | 8.2 | 9.2 |
| ethane | 185 | 545 | 618 | 14.7 | 16.4 |
| propane | 231 | 581 | 684 | 18.8 | 22.8 |
| butane | 273 | 601 | 714 | 22.4 | 28.3 |
| ethene | 169 | 568 | 673 | 13.5 | 16.1 |
| propene | 225 | 609 | 674 | 18.5 | 18.8 |

## 2.4 Discussion of condensed-phase simulations

In the previous section it was concluded that MEDFF predicted the liquid density of some small alkanes and alkenes in rather poor agreement with experiment, with simulated densities more than $10\%$ above the reference value. This can be considered a rather poor performance when taking into account that force fields derived specifically to reproduce these quantities

typically obtain errors smaller than $1\%$. An example of a such force field is the all atom optimized potential for liquid simulations (OPLS-AA), which for instance results in a liquid density of butane at the boiling point deviating only $-0.5\%$ from the experimental value.[171] However, the same force field has been reported to fail in correctly reproducing the second virial coefficient of the same substance,[172] as graphically represented in Figure 2.10. At a temperature $T = 280\,\mathrm{K}$, only slightly above the boiling point, the deviation between the experimental and simulated second virial coefficient is $25\%$. It is therefore inevitable to conclude that OPLS-AA does not provide a value for the liquid density of butane close to experiment because it offers a fundamentally correct description of the interaction between butane molecules, but rather because it relies on a compensation of errors. It is true that some

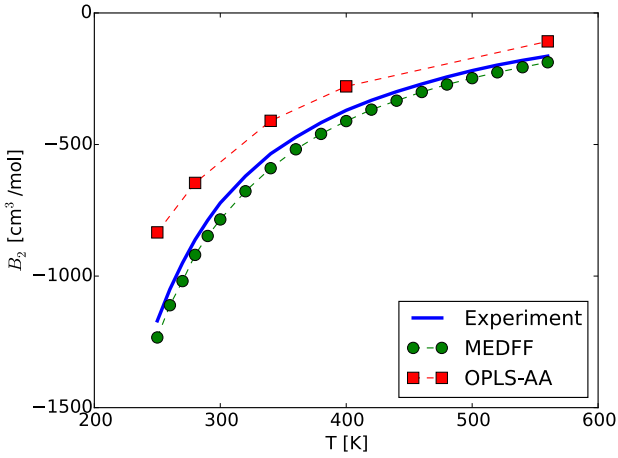



**Figure 2.10:** Second virial coefficient of butane. Experimental data from Ref. 164, OPLS-AA results from Ref. 172. Reprinted with permission from Ref. 75. Copyright (2017) American Chemical Society.

other force fields might provide appropriate results both for the condensed and gas phase at the same time. Still, it has been recognized some time ago that empirical force fields in general do not provide *the right answers for the right reasons*,[39] a topic which is not always fully appreciated in the force-field development community, although it recently gained some more attention.[127, 173]

A more elaborate discussion on whether it is desirable that force fields rely on compensation of errors will be provided in the concluding Section 4.3 of this work. In the next sections, more insight into differences between gas-phase and condensed-phase simulations is gained, using the MEDFF description of

methane as a test case. The topics that will be considered subsequently in the following three sections are changes in atomic parameters, many-body effects and nuclear quantum effects coupled with an anharmonic covalent force field.

### 2.4.1 MBIS parameters from the condensed phase

One of the basic assumptions in the derivation of MEDFF was that the atomic parameters could be obtained from ab initio calculated densities of the monomers of interest, by applying the MBIS partitioning scheme to these densities. When many monomers are placed together in a configuration representative of for instance a liquid, the ab initio densities of the monomers will deform and this will have an effect on the MBIS partitioning and thus on the resulting atomic parameters that serve as input for MEDFF.

A configuration of $100$ methane molecules in a periodic box was obtained from an NVT simulation at $111.6\,\mathrm{K}$ and the experimental density employing the OPLS-AA force field. For this configuration, the PBE electron density was calculated and partitioned using the MBIS scheme. The resulting MBIS parameters of each atom change from molecule to molecule, and for further simulations the averaged value over the $100$ methane molecules was taken. A comparison between some MBIS parameters (core charge $q_c$, valence population $N_v$ and valence width $\sigma_v$) obtained as the average value from a condensed-phase electron density with the corresponding parameters as derived from monomer calculations is provided in Table 2.2. Despite the seemingly subtle changes in the atomic parameters, there is a considerable influence on the calculated liquid density of methane at the boiling point: the error with respect to experiment drops from about $19\%$ to less than $12\%$. Interestingly, the correspondence with CCSD(T)/CBS interaction energies for a set of methane dimers is adversely affected. This seems to confirm that many-body effects play an important role in condensed-phase simulations, even for the apparently simple system that is liquid methane. The procedure presented here allows the *implicit* capture of such effects in a pairwise-additive potential, similar to what is done when a force field is directly fitted to experimental condensed-phase data. The next section includes a further discussion on the *explicit* inclusion of many-body effects.

### 2.4.2 The importance of many-body (dispersion) effects

Another basic premise in the construction of MEDFF was that the interaction parameters could be determined based on reference dimer interaction energies. This does not allow investigating many-body effects (here, a molecule is

**Table 2.2:** Comparison of MBIS parameters for C and H atoms in methane, derived from monomer electron densities or condensed-phase electron densities.

| Atom | Parameter | Monomer | Condensed |
|---|---|---|---|
| C | $q_c$ | 4.347 | 4.346 |
| | $N_v$ | 4.965 | 5.043 |
| | $q$ | −0.617 | −0.697 |
| | $\sigma_v$ [Å] | 0.279 | 0.281 |
| H | $q_c$ | 1.000 | 1.000 |
| | $N_v$ | 0.846 | 0.826 |
| | $q$ | 0.154 | 0.174 |
| | $\sigma_v$ [Å] | 0.202 | 0.200 |

considered to be one body, obviously many-atom and many-electron effects are included in the reference interaction energies). The most common way to include many-body contributions in a force field is by describing polarization effects employing inducible dipoles, fluctuating charges or Drude oscillators. For alkanes and alkenes in the liquid phase, it is most likely crucial to also include many-body effects arising from dispersion interactions.[174]

Force fields that provide an explicit description of many-body interactions are challenging for several reasons. A first concern is the selection of an appropriate model. Inducible dipoles (to describe electrostatic polarization) or the Many-Body Dispersion[18, 175] (MBD) method provide at first sight an elegant inclusion of many-body effects up to infinite order. Both suffer however from the so-called *polarization catastrophe*, which means these methods break down for close contacts between atoms and need *ad hoc* fixes. A second major problem is the determination and validation of parameters. Because many-body effects are rather small for complexes of a few alkane/alkene molecules, it is hard to obtain appropriate reference data using a high-level ab initio method, as computationally expensive calculations on large complexes have to be performed. A final issue is the computational complexity of the resulting force field. While the time to compute the energy of a pairwise-additive force field formally scales as $\mathcal{O}\left(N_{\mathrm{atoms}}^2\right)$, for a force field including many-body terms the complexity increases to $\mathcal{O}\left(N_{\mathrm{atoms}}^3\right)$. Obviously, this makes simulations on long timescales and large length scales challenging and requires a substantial coding effort.

The combination of these difficulties is the reason why no simulations with a many-body force field were conducted within the scope of this work. The

influence of three-body dispersion and exchange interactions on the liquid density and heat of vaporization of some small molecules has been investigated by McDaniel et al.[89] In short, a two-body force field based on SAPT is complemented with an Axilrod-Teller-Muto (ATM) term damped at short interatomic distances. *Effective* values for $C_9$ coefficients are used, which supposedly mimic higher order terms, and are obtained by comparison with SAPT computations of methane trimers. Note that the three-body contributions for these trimer configurations are generally smaller than $0.1\ \mathrm{kJ\,mol^{-1}}$ in magnitude. The density of liquid methane at the boiling point (the same conditions used in the test for MEDFF) reduces from $481\,\mathrm{g\,cm^{-3}}$ for the two-body force field to $453\,\mathrm{g\,cm^{-3}}$ for the three-body force field. The error with respect to experiment reduces from $14\%$ to $8\%$, which is certainly a substantial effect. As three-body dispersion interactions are usually repulsive, it is expected that adding such terms increases the average intermolecular distance and thus lowers the liquid density. The amount by which the density decreases is perhaps larger than anticipated based on the gas-phase trimer calculations, but this only confirms the crucial importance of many-body effects in the condensed phase.

### 2.4.3 Nuclear quantum effects and anharmonicity of the covalent force field

All calculations performed with MEDFF thus far assumed that the nuclei behave as classical point particles. This is an approximation because the quantum nature of the nuclei is not taken into account. Nuclear quantum effects (NQEs) are mostly important for light atoms and at low temperatures, conditions that are satisfied by the simulation of liquid methane at the boiling point. The influence of NQEs can be modeled by performing imaginary time path-integral MD (PIMD) simulations, where the actual physical system is replicated $N_{\mathrm{beads}}$ times with the different replicas or beads connected by harmonic springs. This also means that unfortunately the computational cost increases by a factor $N_{\mathrm{beads}}$. Using the i-PI interface,[176, 177] PIMD simulations were performed for liquid methane at the boiling point (additional computational details are provided in Appendix B.1). The resulting densities are plotted as a function of the number of replicas or beads in Figure 2.11 and compared with both the MEDFF density obtained without NQEs (classical MD) and the experimental density. Although complete convergence is not achieved at $N_{\mathrm{beads}} = 64$, the density obtained for this number of beads is a good estimate of the influence of NQEs, which is substantial for the case investigated here.

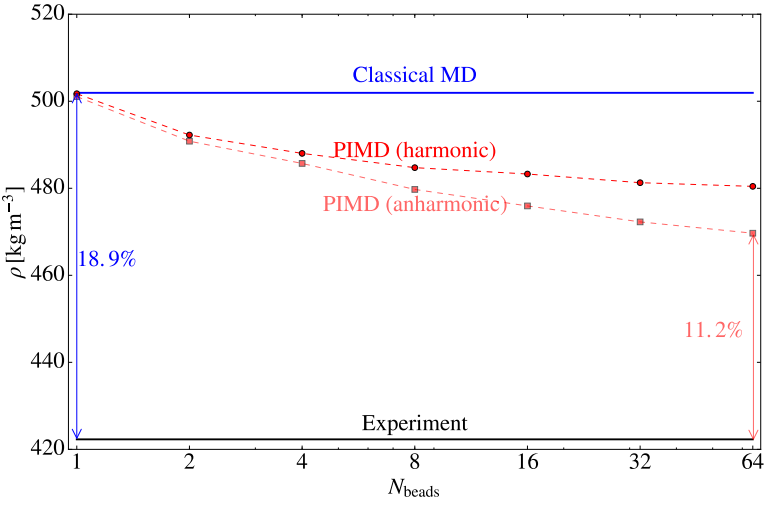



**Figure 2.11:** The liquid density of methane at $112\,\mathrm{K}$ and atmospheric pressure, as simulated using MEDFF with and without the inclusion of nuclear quantum effects

Another aspect that has not been highlighted thus far is the description of the intramolecular interactions within one methane molecule. For methane, this is often deemed unimportant and some popular force fields such as OPLS or TraPPE even represent methane as a single particle, completely neglecting any internal structure of the monomer. In the simulations presented earlier, the noncovalent terms were coupled with a simple harmonic covalent force field (featuring harmonic bonds and bends) derived using QuickFF.[57, 58] For the simulations performed with PIMD, more complicated covalent terms (anharmonic bonds and bends) were introduced, with the necessary parameters again determined using QuickFF. Both cases are compared in Figure 2.11, labeled as PIMD (harmonic) and PIMD (anharmonic) respectively. When $N_{\text{beads}} = 1$, the density is independent of the covalent description and both coincide with the density from classical MD (when $N_{\text{beads}} = 1$, no NQEs are taken into account). It is however striking that as NQEs are included, the difference in the covalent force field does lead to different liquid densities, with deviations of a few percent. The reason for this behavior is that higher nuclear vibrational levels (far from equilibrium, where the largest differences between harmonic and anharmonic potentials are observed) are only sampled in PIMD simulations with a large number of beads.[178] Finally note that the error of MEDFF with respect to experiment reduces from $19\%$ for classical MD, to about $11\%$ when NQEs are fully included and the covalent force field

**Table 2.3:** Comparison of OPLS and MEDFF with respect to the description of methane.

| | OPLS | MEDFF |
|---|---|---|
| Fitted to | Experimental data | Ab initio data |
| Complexity | Very simple | More complicated |
| Computational cost | Extremely low | Moderate |
| Dimer interaction energies | Moderate | Good |
| Second virial coefficients | Good | Good |
| Liquid density | Very good | Overestimated |
| Inclusion of many-body effects | Degrades predictions | Improves description |
| Inclusion of NQEs | Degrades predictions | Improves description |
| Accurate covalent interactions | No influence | Improves description |

is anharmonic. Again, this is a sizable correction that was not accounted for in original MEDFF simulations.

As a concluding note, the recent work by Pereyaslavets et al. should be mentioned.[179] In this study, ab initio force fields for alkanes were derived and it was concluded that NQEs were crucial for the prediction of liquid densities. For alkanes near their boiling point, correction of roughly $6\%$ to $12\%$ were observed. The results provided in Figure 2.11 lie within the same range.

## 2.5 Intermediate conclusions

The methodological advancement presented in this chapter is a new approach to obtain noncovalent force fields from quantum mechanical calculations: MEDFF. The main points that distinguish MEDFF from existing force fields are that it is physically grounded in symmetry-adapted perturbation theory and that overfitting is avoided by employing a minimal amount of parameter tuning. Despite good results for dimers, agreement with experiment for condensed phase properties was under par and several explanations were explored. At this point, a summary of the differences between the generic OPLS force field and MEDFF is provided in Table 2.3 in order to properly understand this behavior. The question which of the two force fields performs "best" is far from trivial, as the definition of a "good" force field is ambiguous. If one is only interested in reproducing the experimental density of methane, OPLS is the obvious choice as it is computationally very cheap (one site per molecule and only Lennard-Jones interactions) and the parameters are fitted precisely to reproduce the experimental density using classical simulations. It is however clear that OPLS accomplishes this good correspondence partially for the wrong reasons. For instance when NQEs are included, which offers a more accurate simulation methodology

than classical simulations, the correspondence with experiment degrades because such effects are already *implicitly* included in the parameterization. In contrast, MEDFF systematically improves as more advanced simulation methodologies are considered and based on the previous sections, it can be assumed that MEDFF with inclusion of many-body effects, NQEs and an anharmonic covalent force field would predict the liquid density of methane in acceptable agreement with experiment. Of course, this approach would be computationally very costly compared to classical simulation of OPLS.

To clarify the statements about the computational cost of both force fields, the actual simulation time is now discussed for the MD simulations in the NPT ensemble of liquid methane. For a system of $100$ molecules described with OPLS, LAMMPS[180] performs $9\,351$ timesteps per second when using $8$ Intel Xeon cores. The same simulation with MEDFF as the PES, performs only $270$ timesteps per second, which is about $35$ times slower than OPLS. There are two important remarks that need to be made in order to properly judge these numbers. First of all, the density predicted by MEDFF is substantially higher than for OPLS. MEDFF thus simulates a denser system which means more pairs of particles need to be taken into account. More importantly, OPLS represents methane as a single site which means that during an MEDFF simulation the number of particles is $5$ times larger for this system (as each methane molecule features $5$ atoms). It it therefore also instructive to consider the all-atom variant of OPLS, which is called OPLS-AA,[181] where each atom is explicitly simulated, as is the case for MEDFF. The same simulation with OPLS-AA performs $1722$ timesteps per second, which means that MEDFF is only $6$ times slower that this generic force field.

This topic is further discussed in Section 4.2, where an outlook on the future of ab initio derived force fields is provided. For now, the current chapter detailing the MEDFF methodology, can be concluded by recapitulating that MEDFF appears to be a robust method that provides gas phase interaction energies in good correspondence with reference values and deviations for the condensed phase results can be (at least partially) explained by considering more advanced extensions. This suffices to proceed to the application of MEDFF to systems that currently pose challenges for standard noncovalent force fields, which is the subject of the next Chapter 3

# 3

# Application: Gas Adsorption in MOFs

In the previous Chapter 2, a new approach to derive noncovalent force fields from ab initio calculations was presented. The resulting MEDFF methodology was extensively tested and shown to provide a reliable description of interactions between dispersion-dominated molecules for some model systems. In the current chapter, the application of MEDFF to technologically relevant problems is investigated focussing on the adsorption of small gas molecules in Metal–Organic Frameworks (MOFs). After a very brief introduction to MOFs at the beginning of Section 3.1, the simulation procedure for computing adsorption isotherms is critically evaluated by comparing different force fields for the case of $CH_4$ adsorption in zirconium (Zr) based MOFs in Section 3.1.1. Because a consistent agreement between simulated and experimental uptakes is shown to be lacking, the same systems are subjected to ab initio calculations which help to understand the high sensitivity of Grand Canonical Monte Carlo (GCMC) results to the underlying PES and in this way explain the previous apparent inconsistencies (Section 3.1.2).

A second investigation, presented in Section 3.2, deals with a novel procedure to facilitate the calculation of Henry coefficients, which characterize adsorption in the low-pressure regime. This procedure allows computing ab initio properties using a limited number of calculations, by using a force field as a biasing potential. Henry coefficients of $CO_2$, $CH_4$ and $N_2$ in various MOFs, including frameworks featuring open metal sites, are calculated and it turns out that MEDFF is the best choice as the biasing potential for subsequent

ab initio calculations. This research shows that MEDFF has its merits even outside of its original intention to predict interaction energies of dispersion-dominated complexes. Owing to the robust way MEDFF was constructed, from the choice of the functional form to the calibration of the interaction parameters, MEDFF appears to provide semi-quantitative agreement with ab initio calculations for a broad range of systems.

## 3.1 Methane adsorption in Zr-based MOFs

MOFs are coordination polymers consisting of metal ions coordinated by organic ligands and often show nano- or micropores. One of the main points that triggered scientific interest, is the sheer amount of imaginable MOFs. Different MOF building units (of which some examples are shown in Figure 3.1) allow much freedom when it comes to coordination, which results in a multitude of possible topological nets. Together with the availability of many different building units, this has led to a huge increase of reported MOFs thanks to the concept of reticular or modular synthesis.[182, 183]

A first possible application is the development of single-site catalysts.[20] Catalytic sites can be included by considering open metal sites or coordinatively unsaturated sites (CUSs), but among many other possibilities catalytic activity can also be achieved by incorporating nanosized clusters within the pores. Another practical use exploits the ability of some MOFs to undergo large structural transformations. More generally, crystalline materials that show multistable behavior are named soft porous crystals,[184] and can for instance behave as shock adsorbers and be used for energy storage. The stimuli that can trigger structural transformation, including mechanical pressure, temperature, and guest adsorption, have been studied from a computational point of view.[58] In this work, the focus will be on MOFs as promising materials for applications involving gas storage or separation.[185–187] When compared with other porous materials such as zeolites, some MOFs show exceptional structural features. Thanks to their modifiable structure and many functionalization posibilities, materials with very high porosity and large surface area have been synthesized. These properties are of course vital when it comes to adsorption of small gas molecules.

One of the limiting factors for the use of MOFs in real-life applications, is the limited mechanical stability of many frameworks.[188] A remarkable exception is the UiO-66 framework, which remains crystalline under rather high external pressures, to some extent even when defects are present.[60] This is the reason why the work presented in the next sections (a summary of the

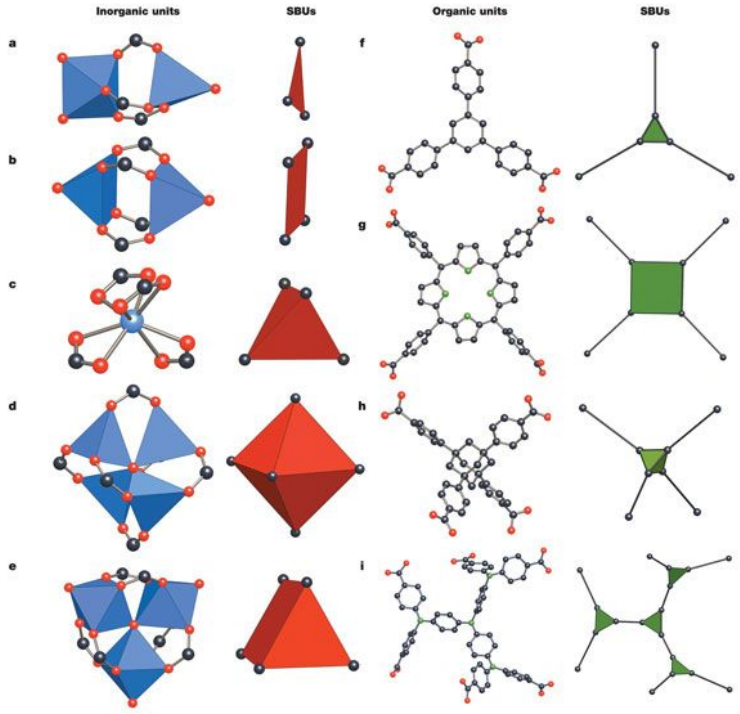



**Figure 3.1:** Example of different building blocks that can be used to construct carboxylate MOFs. Reprinted with permission from Ref. 183. Copyright (2003) Springer Nature.

findings from **Paper III [Ref. 189]**) will focus on Zr-based MOFs, of which the very stable UiO-66 framework is a prototypical example.

The considered materials are UiO-66,[190] UiO-67,[190] DUT-52,[191] NU-1000[192] and MOF-808[193] which all feature $Zr_6(\mu_3\text{-}O)_4(\mu_3\text{-}OH)_4$ bricks. The former three share the same **fcu** topology as depicted in Figure 3.2 and differ only in the ditopic organic ligand that connects the metal centers. Crystals of NU-1000 and MOF-808 on the other hand display respectively a **csq** topology (8 tetratopic linkers coordinating the Zr node) and a **spn** topology (6 tritopic linkers coordinating the Zr node), which is shown in Figure 3.3. Experimental results for methane adsorption in UiO-66, UiO-67 and DUT-52 are available in the literature and a comparison between experimental and simulated isotherms will be the subject of the following section.

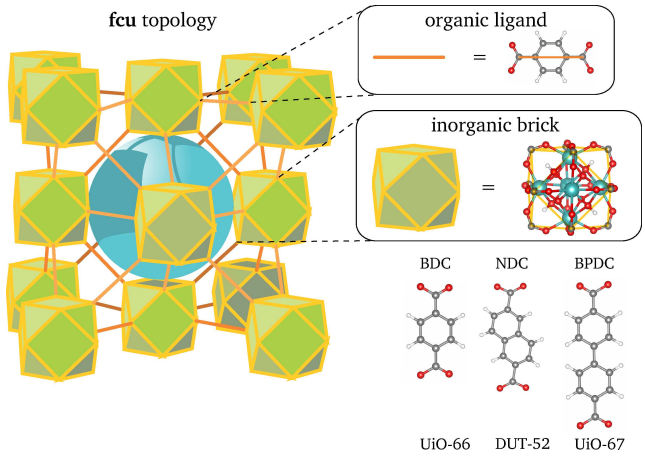



**Figure 3.2:** Depiction of the **fcu** topology which is shared by UiO-66, DUT-52 and UiO-67. Also the inorganic brick is common between these materials, which differ only in the organic linker. Zirconium atoms are shown in cyan, oxygen atoms in red, carbon atoms in gray, and hydrogen atoms in white. Reprinted with permission from Ref. 60. Copyright (2016) American Chemical Society.

### 3.1.1 Comparison with experimental isotherms: quantitative and qualitative investigation

In order to computationally screen a database of MOFs for materials apt for a certain gas-adsorption related application, ideally there should be quantitative agreement between experimentally measured and simulated gas uptakes. To compute the uptake or number of adsorbed guest molecules, most often a Monte Carlo simulation in the grand canonical ensemble is performed, referred to as a GCMC simulation. It is obvious that the simulated gas uptake will depend on the PES that is used to describe interactions within the framework, among guest molecules and between the guests and the framework. The first question is therefore whether one or more force fields from the literature can provide a systematic agreement with experimental adsorption isotherms.

Five force fields are considered and an overview of the properties of each is provided below. The **UFF/TraPPE**[29, 39] model considers methane as a single site. This means there are no covalent terms for methane. Electrostatic interactions involving methane are also absent. Additionally, as the frameworks are considered rigid in this work (which is conventional in GCMC simulations and appropriate for the rigid MOFs studied here), no interactions

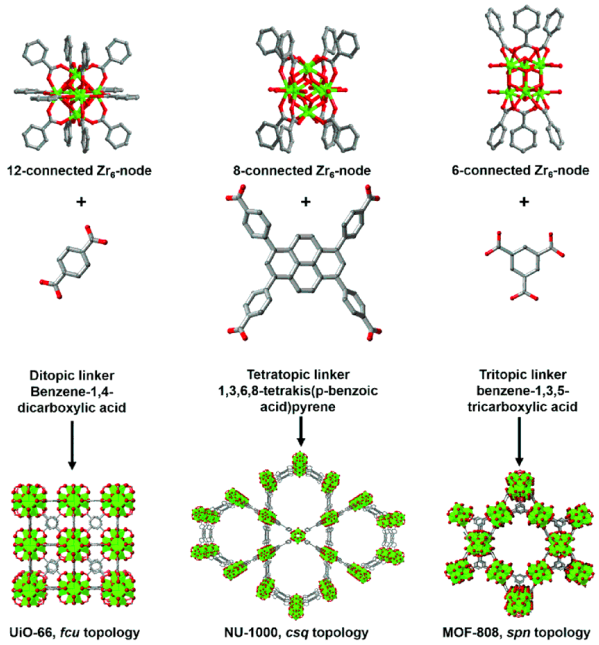



**Figure 3.3:** Node connectivity, linker and topology for UiO-66, NU-1000 and MOF-808. Reproduced from Ref. 194 with permission from The Royal Society of Chemistry.

between framework atoms need to be considered. In the end, the UFF/TraPPE model thus only features Lennard-Jones interactions between pairs of guest-guest and guest-framework atoms. The Lennard-Jones $\epsilon$ and $\sigma$ parameters for framework atoms are taken from UFF, which were determined using a mixture of calculations, experiments and empirical rules. The corresponding methane parameters from TraPPE were determined in order to reproduce the critical temperature and saturated liquid density of bulk methane. The second force field will be referred to as **Dreiding-UFF/TraPPE** and is very similar to the first considered model. The only difference is that the framework Lennard-Jones parameters are now taken from the Dreiding[25] force field, which is based on experimental crystal structures and sublimation enthalpies (except for zirconium atoms, for which no Dreiding parameters are available and thus the UFF values are used). The third model, **MM3-MBIS**,[32–34, 195] replaces the Lennard-Jones potential with a Buckingham potential which should provide a better description of the repulsive wall for close interatomic contacts. In this case, all atoms from the methane guest molecule are considered explicitly and thus electrostatic interactions involving methane molecules are present and are modeled by including MBIS point charges (although it should be mentioned that the influence of the electrostatic

interactions is minor for the simulations performed here). Another important difference with the previous force fields, is that the noncovalent parameters now also depend on the chemical environment of the atoms and no longer are uniquely determined by the atomic number of the atoms. The number of parameters and complexity of the model is further increased in the next force field, **SAPTFF**.[87, 196] This is an ab initio based force fields where all necessary parameters are obtained from a combination of monomer properties and SAPT calculations of dimer interactions. It shares some similarities with the final force field that will be considered, **MEDFF**,[75] which was discussed at length in the previous chapter of this dissertation. All GCMC simulations discussed in this section have been performed using the RASPA software package.[197]

The simulated and experimental adsorption isotherms at room temperature for methane in UiO-66, DUT-52 and UiO-67 are plotted in Figure 3.4, considering pressures up to $80\,\mathrm{bar}$, with the plots in the right column zooming in on the low-pressure region. For UiO-66, **Dreiding-UFF/TraPPE** offers a good agreement with the experimental isotherm while **UFF/TraPPE** overestimates the uptake at relatively high pressure, a finding which was published in earlier work.[198] The ab initio based force fields **MEDFF** and **SAPTFF** both overestimate the adsorption while **MM3-MBIS** provides an underestimation. At pressures below $1\,\mathrm{bar}$ however, rather different conclusions can be drawn as shown in Figure 3.4 (b). Now **MM3-MBIS** and **MEDFF** are closest to experiment, while Dreiding-UFF/TraPPE overestimates the uptake by a factor $2$. This is particularly surprising given that at low pressures (and thus low loadings), adsorption is completely determined by host-guest interactions while at higher pressures (and thus high loadings) additionally the description of guest-guest interactions will have a sigificant impact on the simulated adsorption. The fact that **Dreiding-UFF/TraPPE** is deficient at low pressures must mean that it does not provide a proper description of the host-guest interactions. The correspondence with experiment at high pressure must then be explained as a deficient description of the guest-guest interactions which cancels the previously mentioned error in the host-guest interactions in the case of UiO-66. This seems to be confirmed by the poor performance of the same model for the frameworks DUT-52 and UiO-67, which are from a chemical point of view rather similar to UiO-66. All in all, no single force field achieves a consistent agreement with experiment for UiO-66, DUT-52 and UiO-67 over a wide pressure range. The **MM3-MBIS** model shows the best overall performance, with **MEDFF** as a close second in the low-pressure regime.

When simulating isotherms using a force field, one is not always interested in the absolute loading but rather in qualitative trends such as the ranking

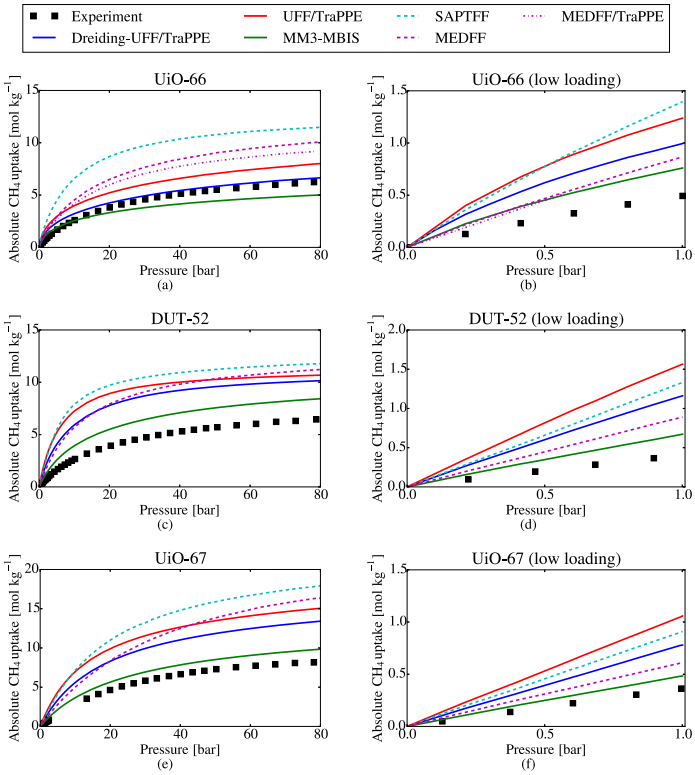



**Figure 3.4:** Comparison of experimental and simulated gravimetric adsorption isotherms of methane in a series of isoreticular MOFs. Experimental data for (a), (c), (d), (e) and (f) from Ref. 199, experimental data for (b) from Ref. 200. Reprinted with permission from Ref. 189. Copyright (2017) American Chemical Society.

of different frameworks. A second assessment of the force fields considered here is about the ranking of the Zr-based MOFs discussed earlier with respect to gravimetric methane uptake at $P = 30\,\mathrm{bar}$ and room temperature. Both **Dreiding-UFF/TraPPE** and **MM3-MBIS** predict

$$\text{UiO-66} < \text{MOF-808} < \text{DUT-52} < \text{UiO-67} < \text{NU-1000}$$

while the other models, **UFF/TraPPE**, **SAPTFF** and **MEDFF** reverse the position of MOF-808 and UiO-66:

$$\text{MOF-808} < \text{UiO-66} < \text{DUT-52} < \text{UiO-67} < \text{NU-1000}$$

Despite large differences in uptake predicted by the different models, there is a quite good agreement on the ranking, and if the uptake at $P = 100\,\mathrm{bar}$

is considered all rankings are in perfect agreement. There are however two critical remarks about this apparent consistency between results obtained with the different force fields. First, it should be mentioned that methane uptakes at high pressure correlate well with the surface area and pore volume of the framework,[201] quantities that can be computed using a simple geometric calculation instead of complicated GCMC simulations. A second remark is that **UFF/TraPPE**, **Dreiding-UFF/TraPPE** and **MM3-MBIS** predict that the uptake at $P = 30\,\mathrm{bar}$ is *much* higher in DUT-52 than in UiO-66, while experimentally they are nearly the same. Concluding this section, it can be stated that different force fields will generally predict very dissimilar uptakes, but do agree to a wider extent on qualitative results such as the ranking of a set of materials.

### 3.1.2 Comparison with ab initio adsorption energies: sensitivity of GCMC simulations to the potential energy surface

In order to gain more insight at the atomic level on differences between the force fields employed in the previous section, interactions between a single methane molecule and the UiO-66 framework are studied, aided by the use of periodic ab initio calculations. First, an appropriate level of theory that can serve to provide reference interaction energies has to be selected. Here, the PBE[202] functional with modified D3 dispersion corrections with Becke-Johnson damping (D3MBJ)[203], including three-body ATM contributions[17] is chosen. This is justified by the good agreement with interaction energies for a set of dimers consisting of methane and terepthtalic acid molecules (the latter is chosen as it represents the UiO-66 BDC organic linkers) computed using CCSD(T)/CBS. At the same time, the PBE-D3MBJ-ATM is computationally tractable in order to study many configurations of methane in UiO-66.

A second consideration which turns out to be very important, is the selection of those configurations. Ideally, all relevant adsorption sites should be taken into account, and an attractive approach is to simply extract snapshots from GCMC simulations of the system under consideration. This however introduces a certain bias, as the resulting configurations will obviously depend on the PES that was used for those GCMC simulations.[204] In an attempt to reduce this bias, $100$ configurations of a single methane molecule in UiO-66 were extracted from a GCMC simulation (at room temperature and $1$ bar) for *each* of the five force fields, resulting in a set of $500$ configurations in total. The scatter plots of force-field versus ab initio adsorption energies for two subsets of all configurations are shown in Figure 3.5, where blue dots indicate configurations extracted from a **Dreiding-UFF/TraPPE** GCMC simulation and the red dots indicate extracted from a **SAPTFF** GCMC simulation. In

general, the ab initio derived force fields **SAPTFF** and **MEDFF** are closer to the reference adsorption energies than the generic **UFF/TraPPE** and **UFF-Dreiding/TraPPE** models. More surprisingly perhaps, the RMSD for the latter two models is significantly higher when **SAPTFF** configurations are considered than when **Dreiding-UFF/TraPPE** configurations are considered. Note that the **SAPTFF** configurations (red dots) are indeed relevant as the ab initio computed adsorption energies are well below zero. It has been claimed that generic force fields such **UFF/TraPPE** provide good agreement with quantum chemical predictions for the adsorption of methane in MOFs.[205] The results shown here contradict this: the generic force fields clearly are overly repulsive in many cases, at least if the complete portion of the PES that is relevant for adsorption is considered. If the generic force field itself is used to sample the PES, some favorable adsorption sites are overlooked, biasing the results. Finally note that the results obtained using the **MM3-MBIS** model are markedly better than those from **UFF/TraPPE** and **UFF-Dreiding/TraPPE**. This suggests that the overly steep repulsive wall (proportional to $r^{-12}$ in the latter cases versus exponentially decaying for **MM3-MBIS**) plays an important part in the previous observations.

Finally, the results concerning single-molecule adsorption energies are connected to the isotherms presented in the previous section. For **UFF/TraPPE**, **Dreiding-UFF/TraPPE** and **SAPTFF**, the most favorable configurations are overstabilized, i.e. below the diagonal in the corresponding scatter plots of Figure 3.5. This explains the high uptakes at low pressures as shown in Figure 3.4 (b). As shown in the scatter plots, **UFF/TraPPE** and **Dreiding-UFF/TraPPE** are overly repulsive for many configurations that are predicted to be favorable by the ab initio level of theory. A detailed analysis of the PESs confirms that the volume of accesible pore space is relatively small for these models and this explains why the corresponding isotherms saturate at a relatively low loading. This offers an additional explanation for the earlier statement that **Dreiding-UFF/TraPPE** achieves agreement with experiment by a cancellation of errors.

Although there are clear differences between the force fields concerning the prediction of ab initio adsorption energies, the errors are generally below $1\,\mathrm{kcal\,mol^{-1}}$ which is often considered the threshold for "chemical accuracy". The errors with respect to the experimental Henry coefficient (which is the slope of the isotherm at low pressure), however range from $80\%$ for **MEDFF** up to $300\%$ for **UFF/TraPPE**. A further investigation revealed that by changing parameters of the **Dreiding-UFF/TraPPE** model such that adsorption energies vary by $2\,\mathrm{kJ\,mol^{-1}}$, changes in the predicted uptake at low pressure are as high as $80\%$. This high sensitivity to small changes in the PES, can be explained by considering that the uptake at low pressure is proportional to

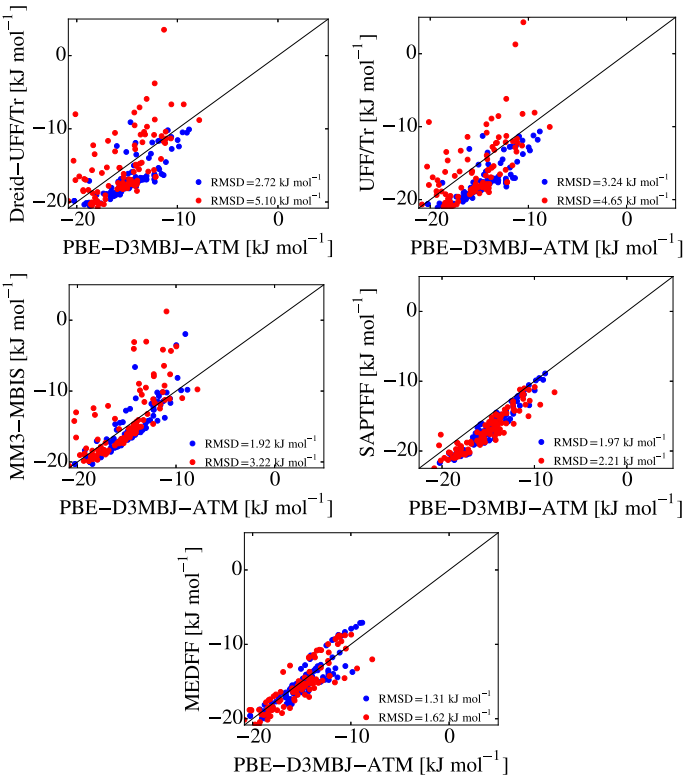



**Figure 3.5:** Scatter plots of adsorption energies for 200 configurations of methane in UiO-66 (blue dots: sampled from Dreiding-UFF/TraPPE GCMC, red dots: sampled from SAPTFF GCMC). Force-field energies are compared with PBE-D3MBJ-ATM results. Reprinted with permission from Ref. 189. Copyright (2017) American Chemical Society.

the Boltzmann factor $e^{-\beta E_{\mathrm{ads}}}$. From this it can be concluded that precise quantitative agreement between experimental and simulated isotherms is extremely challenging.

## 3.2 Ab initio determination of Henry coefficients

In the previous section, a critical assessment of different force fields was presented with respect to their description of methane adsorption in Zr-based MOFs. One of the added values of ab initio derived force fields such as SAPTFF and MEDFF, was that these allowed to correctly sample all relevant regions of the PES. This information was then used to gain fundamental

insight into force-field predicted isotherms, starting from a comparison between ab initio and force-field single-molecule adsorption energies.

In this section, which summarizes the methodology and findings presented in **Paper IV [Ref. 206]**, another attractive feature of MEDFF is highlighted, namely as a tool that can facilitate the computation of ab initio Henry coefficients.

### 3.2.1 Importance sampling with MEDFF as biasing potential

As stated before, the slope of the isotherm at low pressures is called the Henry coefficient. For many materials at typical operating conditions, the uptake at low pressures is indeed proportional to the gas pressure and therefore the Henry coefficient completely characterizes the low-pressure adsorption regime. Assuming that both the host geometry and guest geometry are rigid, the Henry coefficient at temperature $T$ can be computed as:

$$K_H = \beta \frac{\int d\mathbf{s} \exp\left(-\beta \Delta U\right)}{\int d\mathbf{s}} \tag{3.1}$$

where $\beta = \frac{1}{k_B T}$, $\Delta U = U_{\text{host+guest}} - U_{\text{host}} - U_{\text{guest}}$ is the adsorption energy and $\int d\mathbf{s}$ indicates an integration over all configurations of the guest molecule in the host. The integration is usually performed using a Monte Carlo approach, more specifically the Widom insertion method.[207] The Widom insertion method consists of generating a set of $N$ random configurations without any bias, after which the estimated value $\hat{K}_H$ is determined as:

$$\hat{K}_H = \beta \frac{1}{N} \sum_{i=1}^{N} \exp\left(-\beta \Delta U_i\right) \tag{3.2}$$

As the number of Widom insertions $N$ increases, the estimate gets better and for the cases considered here (small molecules adsorbing inside MOFs) a typical order of magnitude is $1$ to $10$ million insertions. When a force field is used to evaluate the adsorption energies $\Delta U_i$, the associated computational cost is acceptable but for many ab initio methods it will present an unsurmountable barrier. The goal of the following derivation is therefore to reduce the number of calculations required to estimate the Henry coefficient for a given PES, to the point where ab initio methods become feasible.

The main problem when evaluating the integral in eq 3.1 is the large variance on the integrand: $\exp\left(-\beta \Delta U\right)$ ranges from nearly zero for repulsive configurations ($\Delta U$ positive and much larger than $k_B T$) to a very large number ($\Delta U$

negative and several times as large in magnitude as $k_BT$). A solution that has been used many times before for such a problem is importance sampling. In this context, it will prove advantageous to introduce a biasing PES for which the adsorption energies are represented as $\Delta\tilde{U}$. The Henry coefficient can be rewritten as:

$$K_H = \frac{\int d\mathbf{s} \exp\left(-\beta\left[\Delta U - \Delta\tilde{U}\right]\right)\exp\left(-\beta\Delta\tilde{U}\right)}{\int d\mathbf{s}\exp\left(-\beta\Delta\tilde{U}\right)}\beta\frac{\int d\mathbf{s}\exp\left(-\beta\Delta\tilde{U}\right)}{\int d\mathbf{s}} \tag{3.3}$$

$$= \left\langle \exp\left(-\beta\left[\Delta U - \Delta\tilde{U}\right]\right)\right\rangle_{e^{-\beta\Delta\tilde{U}}} \tilde{K}_H \tag{3.4}$$

where $\tilde{K}_H$ is the Henry coefficient for the PES $\Delta\tilde{U}$. This expression might not seem to be very useful at first sight, because in order to compute $K_H$, another Henry coefficient $\tilde{K}_H$ has to be computed as well as the integral $\left\langle \exp\left(-\beta\left[\Delta U - \Delta\tilde{U}\right]\right)\right\rangle_{e^{-\beta\Delta\tilde{U}}}$ (where the subscript indicates that configurations need to be selected with probability proportional to $e^{-\beta\Delta\tilde{U}}$. The crucial idea behind eq 3.4 is the assumption that the biasing PES $\Delta\tilde{U}$ is computationally cheap (typically a force field), which allows an accurate estimate of $\tilde{K}_H$ using Widom insertion. Calculations with the original PES (typically an ab initio method) only need to be performed to evaluate $\left\langle \exp\left(-\beta\left[\Delta U - \Delta\tilde{U}\right]\right)\right\rangle_{e^{-\beta\Delta\tilde{U}}}$. The latter integral however only requires a limited number of energy evaluations, on condition that the biasing PES and the original PES are sufficiently similar. Indeed, as $\Delta U \approx \Delta\tilde{U}$, the integrand will always be close to unity and show a rather small variance. The importance sampling method introduced here can be summarized by the following steps. First, the Henry coefficient of a biasing PES $\Delta\tilde{U}$ (typically a force field) is computed using Widom insertion, using a large number of $N$ samples. Next, a limited number of $n$ configurations is selected from the Widom insertion simulation, with the probability to include a certain configuration proportional to $e^{-\beta\Delta\tilde{U}}$. Finally, the adsorption energy $\Delta U$ is computed for these $n$ configurations which allows the evaluation of eq 3.4. Before proceeding to a validation of the proposed procedure, it should be mentioned that the previous steps also provide all necessary information in order to evaluate quantities such as the average adsorption energy for the $\Delta U$ PES.

### Validation

A crucial condition for the practical application of the method outlined above, is the availability of a biasing force field that provides adsorption energies

*close* to the corresponding ab initio values. The term *close* is rather vague at this point: the validation performed in this section will show what it means by considering a relevant use case. For $CH_4$ in UiO-66, the integral in eq 3.1 was calculated on a regular grid using the PBE+D3(BJ)[202, 208] ab initio level of theory. This required a substantial computational effort and was meant to be performed once in order to obtain a benchmark value against which the results from the proposed importance sampling method could be compared. This comparison is shown in Figure 3.6, where the convergence of the estimated Henry coefficient is plotted as a function of $n$, the number of ab initio calculations performed when estimating $\left\langle \exp\left(-\beta\left[\Delta U - \Delta\tilde{U}\right]\right)\right\rangle_{e^{-\beta\Delta\tilde{U}}}$. The black line indicates the benchmark result from the numerical integration on a grid and different subplots are provided for different biasing PESs (MEDFF, MM3-MBIS, Dreiding-UFF/TraPPE and UFF/TraPPE). It seems that

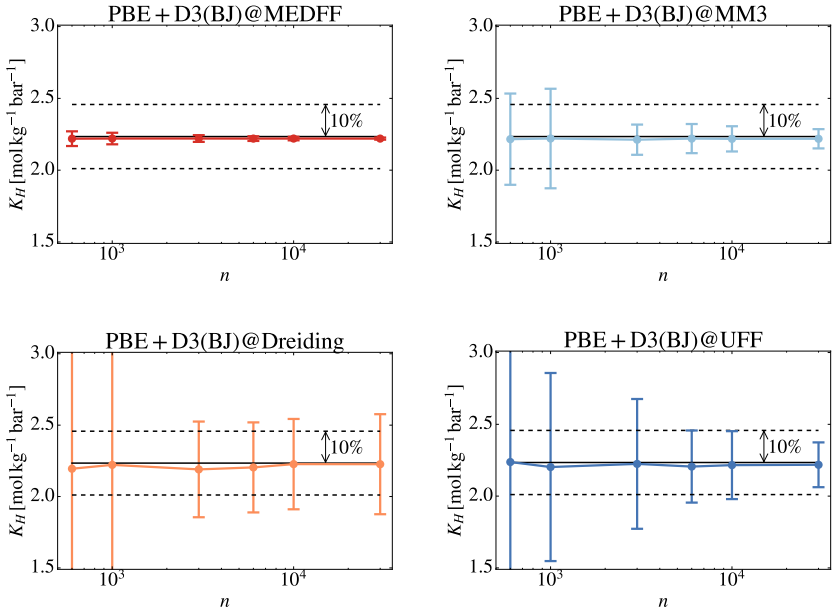



**Figure 3.6:** Comparison of the importance sampling method with a grid-based approach for the Henry coefficient $K_H$ of $CH_4$ in UiO-66 at room temperature. Notations such PBE+D3(BJ)@MEDFF indicate that PBE+D3(BJ) plays the role of the original PES $\Delta U$, while MEDFF is used as the biasing potential $\Delta\tilde{U}$. Reprinted with permission from Ref. 206. Copyright (2017) American Chemical Society.

the Henry coefficient estimated using importance sampling converges to

the correct value, regardless of which force field is used as the biasing potential. The rate of convergence however varies considerably. When using MEDFF as the biasing potential, the least amount of ab initio calculations $n$ needs to be performed. MM3-MBIS requires considerably more samples and Dreiding-UFF/TraPPE and UFF/TraPPE in turn require even more. This can be explained completely by the discussion in **Paper III [Ref. 189]**, where the same order was observed when ranking these force fields for the reproduction of ab initio adsorption energies for the same system, $CH_4$ in UiO-66. It is concluded that by using MEDFF as the biasing potential, the number of required ab initio calculations can be reduced by roughly two orders of magnitude compared to the conventional Widom insertion approach. This reveals another possible use of MEDFF, namely as a tool to facilitate the ab initio prediction of adsorption properties and it will therefore be employed in the following applications of the importance sampling methodology for the calculation of Henry coefficients.

### 3.2.2 Adsorption on open metal sites: $CO_2$ in Mg-MOF-74

One of the cases where generic force fields are known to fail in describing adsorption, is when polar guest molecules adsorb at open metal sites present in some MOFs.[209] A prototypical case is $CO_2$ in Mg-MOF-74, where it would thus be interesting to compute Henry coefficients using an ab initio PES. The convergence of the Henry coefficient and adsorption enthalpy as function of the number of importance samples $n$ was investigated using MEDFF as the biasing PES, and $n = 20\,000$ was considered appropriate. It is not trivial that MEDFF provides adsorption energies in acceptable correspondence with ab initio values without any reparametrization, because in this example the interaction between $CO_2$ and the open metal sites is not dispersion dominated (and only for those cases it was shown that MEDFF predicted accurate interaction energies). The explanation why MEDFF does work as the biasing potential is twofold. First, its derivation and parametrization are much more robust than the way empirical force fields are constructed. Second, the correspondence between biasing potential and reference PES does not need to be incredibly good for the importance sampling method to work well, as long as the relevant parts of the reference PES are explored by the biasing potential. This explains why MEDFF is an appropriate choice for the task at hand.

Henry coefficients $K_H$ and adsorption enthalpies $\Delta H_{\mathrm{ads}}$ were calculated for $CO_2$ in Mg-MOF-74 for different ab initio levels of theory using the importance sampling approach. The results are summarized in Table 3.1, together with force-field results (obtained using a Widom insertion simulation) and

experimental values from different sources. It turns out that the simulation results are rather sensitive to the framework geometry, which was considered rigid. Therefore each simulation was performed for two geometries: the first one is a DFT optimized structure of the empty framework (labeled "geometry empty", on the left-hand side of Table 3.1) and the second one a DFT optimized structure loaded with $CO_2$ molecules located at the primary adsorption sites (labeled "geometry loaded", on the right-hand side of Table 3.1). Although both geometries are visually very similar, the difference in results is substantial: adsorption enthalpies are roughly $4\ \mathrm{kJ\,mol^{-1}}$ lower in the loaded geometry and the Henry coefficients consequently a factor $3$ to $4$ larger. This can be seen as an indication that the rigid framework assumption is not valid in this case, and a hybrid MC/MD extension of the importance sampling approach should be pursued.

| | Geometry Empty | | Geometry Loaded | |
|---|---|---|---|---|
| Level of theory | $K_H$ | $\Delta H_{\mathrm{ads}}$ | $K_H$ | $\Delta H_{\mathrm{ads}}$ |
| PBE+D2 | 34 | -34.2 | 124 | -38.1 |
| PBE+D3(BJ) | 48 | -34.1 | 168 | -38.2 |
| PBE+D3M(BJ) | 112 | -37.5 | 434 | -41.4 |
| vdW-DF2 | 130 | -37.8 | 522 | -42.1 |
| MEDFF | 62 | -37.8 | 488 | -44.8 |
| MM3 | 28 | -31.9 | 71 | -35.1 |
| Dreiding | 27 | -30.6 | 45 | -32.5 |
| UFF | 31 | -31.2 | 47 | -32.8 |

| Experimental source | $K_H$ | $\Delta H_{\mathrm{ads}}$ |
|---|---|---|
| Britt et al. [210] | - | -39 |
| Mason et al. [211] | 384 | -42 |
| Yu et al. [212] | 414 | -42 |
| Queen et al. [213] | 407 | -44 |

**Table 3.1:** Henry coefficients in $\mathrm{mol\,kg^{-1}\,bar^{-1}}$ and enthalpies of adsorption in $\mathrm{kJ\,mol^{-1}}$ of $CO_2$ in Mg-MOF-74 at $298\ \mathrm{K}$. Reprinted with permission from Ref. 206. Copyright (2018) American Chemical Society.

Still, some interesting conclusions can be drawn from the presented results. First of all, the generic UFF, Dreiding and MM3 force fields all provide rather low Henry coefficients compared both to ab initio and experimental results. This is related to their inability to provide a proper description of adsorption at open metal sites. MEDFF systematically predicts more attractive adsorption energies and thus higher Henry coefficients, for the loaded geometry even providing values in semiquantitative agreement with

experiment. Because differences between MEDFF and ab initio adsorption energies are observed to be non-negligible, this agreement is likely to be at least in part due to a cancellation of errors. Finally the large spread on results from different ab initio levels of theory should be mentioned. For instance the PBE+D2[13] Henry coefficient is four times smaller that the prediction from vdW-DF2[12, 214–216], confirming a benchmark study of van der Waals corrected functionals.[217] Also note the sizable difference between PBE+D3M(BJ)[203] and PBE+D3(BJ),[208] two levels of theory that only differ in the data set that was used to fit the damping parameters of the dispersion correction scheme. Yet, the former Henry coefficient is more than twice the value of the latter while the adsorption enthalpies differ by about $3\,\mathrm{kJ\,mol^{-1}}$, indicating that even in a situation where dispersion is not seen as the dominant adsorption mechanism the choice of the description of dispersion in DFT should be carefully considered.

### 3.2.3 Comparison of ab initio levels of theory for adsorption in UiO-66

In the previous section it was shown that different variants of two-body dispersions corrections (for instance D3M(BJ) versus D3BJ) could result in sizable differences in Henry coefficients. This topic is further explored here, now focusing on the role of many-body dispersion interactions. To this end, the Henry importance sampling method is used to investigate the adsorption of $CH_4$ in UiO-66, with the PBE functional coupled with the many-body dispersion (MBD)[18] method as the ab initio level of theory and MEDFF as the biasing potential. MBD adds a van der Waals contribution to the electronic energy, computed using the coupled fluctuating dipole model, which includes many-body dispersion effects up to arbitrary order. The obtained Henry coefficients and adsorption energies are plotted in Figure 3.7 as a function of the order of many-body effects that are taken into account. Note that in this context, many-body dispersion effects have to be interpreted as many-atom dispersion effects. In other words, 2-body means only pairs of atoms are considered when computing the dispersion correction energy, 3-body means that also triplets of atoms are considered, *etc.* The Henry coefficient $K_H$ decreases each time an additional odd order $(3, 5, \ldots)$ is included indicating that these are generally repulsive terms, while the Henry coefficient increases each time an additional even order $(4, 6, \ldots)$ is included which shows that these terms provide additional stabilization. Note that the Henry coefficient obtained when including only two-body dispersion interactions is twice as large as the value obtained when fully treating many-body effects. Also note that the latter is closer to the experimental value of

$K_H = 0.6\ \mathrm{mol\,kg^{-1}\,bar^{-1}}$.[200] A similar reasoning explains the behavior of the adsorption enthalpy $\Delta H_{\mathrm{ads}}$. Adding odd orders of many-body dispersion leads to more repulsive interactions, so the adsorption enthalpy increases (its absolute value decreases). Because even orders of many-body dispersion lead to more attractive interactions, this means the adsorption enthalpy decreases (its absolute value increases).

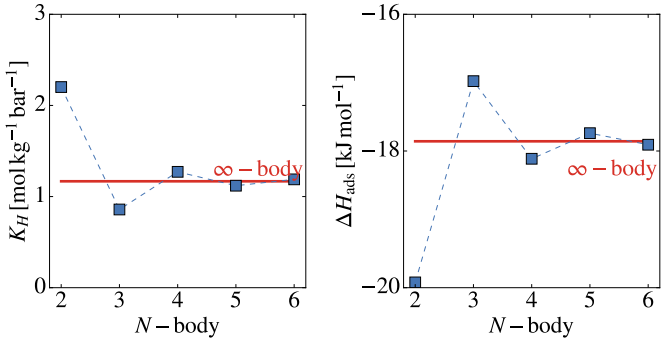



**Figure 3.7:** Henry coefficient (left) and adsorption enthalpy (right) of $CH_4$ in UiO-66 computed with PBE with MBD corrections, as a function of the order of included many-body effects (dotted lines are only present to guide the eye). Reprinted with permission from Ref. 206. Copyright (2018) American Chemical Society.

Finally, a qualitative comparison of the influence of linker functionalizations on UiO-66 is presented. The pristine UiO-66 framework features benzenedicarboxylate as organic linkers. Four variants are considered by respectively functionalizing each linker with amino (–$NH_2$), nitro (–$NO_2$), methoxy (–2,5-$OMe_2$) and naphthyl (–1,4-Naph) groups. For each of the resulting MOFs, the Henry coefficients for $CO_2$, $CH_4$ and $N_2$ were calculated at room temperature and from this the selectivity of equimolar mixtures of those guest molecules was calculated. The influence of the functionalization on these selectivities is shown in Figure 3.8 for some selected PESs: the generic force field UFF, the ab initio derived force field MEDFF, and the PBE functional with D3(BJ) and MBD dispersion corrections. Note that this is a proper example of a *qualitative* investigation: firstly because trends among functionalized versions of the same framework are considered and secondly because only ratios of Henry coefficients (i.e. selectivities at low pressure) for different guest molecules are considered. Still there are some differences between the curves obtained with distinct PESs. For instance, UiO-66-$NH_2$ and UiO-66-$NO_2$ show a higher $CO_2/CH_4$ selectivity than UiO-66-2,5-$(OMe)_2$ and UiO-

66-1,4-Naph according to PBE+D3(BJ), PBE+MBD and MEDFF. The opposite is observed when using the generic UFF potential, providing an example where one should be very careful when dealing with force-field results, even when simply extracting qualitative adsorption properties.

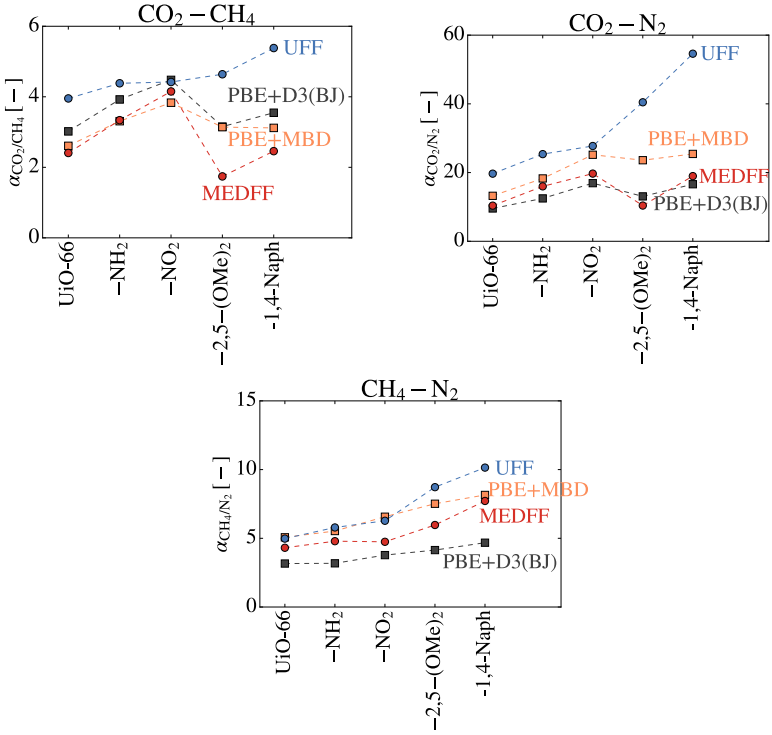



**Figure 3.8:** Selectivities of equimolar mixtures at room temperature in functionalized versions of the UiO-66 MOF. Reprinted with permission from Ref. 206. Copyright (2018) American Chemical Society.

It can be noted that the $CO_2/CH_4$ selectivities predicted by PBE+D3(BJ) and PBE+MBD are rather close to each other, but much larger deviations are observed when $N_2$ is involved. This was further investigated by considering dimers composed of a guest molecule and a functionalized linker. For each combination of guest molecule and functionalized linker, 100 configurations were considered and the mean deviations between PBE+MBD and PBE+D3(BJ) interaction energies are reported in Table 3.2. It is not very surprising that all mean deviations are positive: as mentioned earlier, including many-body dispersion generally leads to more repulsive interactions. Perhaps more surprisingly, the largest deviations are observed for dimers involving $N_2$, despite the fact that absolute interaction energies are much smaller in this case than for the other guest molecules. This also

means that PBE+MBD will show a much lower Henry coefficient for $N_2$ than PBE+D3(BJ), resulting in much higher $CO_2/N_2$ and $CH_4/N_2$ selectivities for the former level of theory. It was attempted to further clarify the observed differences between MBD and D3(BJ) dispersion corrections by investigating molecular $C_6$ coefficients. Again, the largest differences between both methods were observed for the case of $N_2$. Because many other effects (such as the short-range damping or the different functional expressions) come into play, it is however not possible to fully explain the differences in adsorption properties solely based on differences in molecular $C_6$ coefficients.

**Table 3.2:** Mean deviation between interaction energies computed using PBE+MBD and PBE+D3(BJ): $\left\langle \Delta U_{\mathrm{PBE+MBD}} - \Delta U_{\mathrm{PBE+D3(BJ)}} \right\rangle$ in $\mathrm{kJ\,mol^{-1}}$ where the average is over $100$ dimers. Note that the differences between the two levels of theory appear to be systematic for all configurations of a particular dimer, so RMSD values are very close to the (absolute value of) the reported mean deviations.

| Framework | $CO_2$ | $CH_4$ | $N_2$ |
|---|---|---|---|
| UiO-66 | +0.85 | +0.61 | +1.15 |
| UiO-66–$NH_2$ | +0.94 | +0.70 | +1.20 |
| UiO-66–$NO_2$ | +0.88 | +0.66 | +1.29 |
| UiO-66–2,5-$OMe_2$ | +0.91 | +0.68 | +1.44 |
| UiO-66–1,4-Naph | +0.98 | +0.74 | +1.35 |

# 4

# Conclusions and Outlook

## 4.1 Conclusions

The challenge of finding a proper representation of the potential energy surface has been present ever since the early days of molecular modeling. It requires a careful consideration because of the delicate balance between accuracy and computational cost. In many instances, there has been a dividing line between two approaches that have radically different starting points. Ab initio methods simplify the Schrödinger equation to the point where an approximate solution becomes computationally manageable. Force fields on the contrary are oblivious to the quantum nature of molecular systems and try to implicitly incorporate these effects by fitting parameters. In this dissertation it was attempted to bridge the two approaches, by developing a new methodology to derive noncovalent force fields rooted in an ab initio energy decomposition method.

The derivation of a novel AIM method was described in Section 2.1 as a first fundamental step towards ab initio based noncovalent force fields. The MBIS partitioning approach provides a concise representation of an ab initio electron density as a sum of atom-centered Slater-like functions. This way, the electron density is summarized by a few parameters for each atom in a force-field like fashion. It was demonstrated that MBIS is robust, i.e. the resulting parameters are not very sensitive to detailed changes in the electron density, such as small conformational changes or variations of the basis set. MBIS densities were also shown to accurately reproduce reference electrostatic

interactions as well as molecular dispersion coefficients. These properties make it an ideal starting point for the construction of an ab initio derived noncovalent force field.

The methodology of this noncovalent force field, named MEDFF, was presented in Section 2.2. Similar to the SAPT ab initio energy decomposition scheme, the MEDFF energy consists of four distinct components: electrostatic, exchange-repulsion, dispersion and induction interactions. Each of these components was validated and calibrated separately by applying SAPT to a database of molecular dimers. This approach avoided the ill-conditioned fitting in a high-dimensional parameters space, ensured that the proposed energy expression could be justified from a physical perspective and contributed to the overall robustness and transferability. The final MEDFF interaction parameters were refined in order to better reproduce CCSD(T)/CBS interaction energies, especially for dispersion-dominated complexes and an external validation on another database of dimers confirmed the good performance compared to other force fields.

The next part of the validation concerned the reproduction of experimental properties of small alkanes and alkenes, but of course without adapting the MEDFF parameters to these specific cases. In the gas phase (second virial coefficients) encouraging results were obtained. In the condensed phase however (liquid density and enthalpy of vaporization) the agreement with experiment was less satisfying. Several reasons for this discrepancy were explored and it was concluded that effects such as many-body dispersion, nuclear quantum effects, but also an anharmonic description of the intramolecular interactions all have an appreciable impact on the simulation results. A correct description of all these effects would tremendously improve results from ab initio derived force fields, but unfortunately such models are not always trivial to construct as well as incurring a serious computational penalty. Still, this study provided more insight into the apparent success of some generic force fields, which include the previously mentioned effects implicitly in their parameters by fitting directly to experimental data.

After the development and validation of MEDFF, the methodology was used to derive force fields describing the interactions between MOFs and small guest molecules. In Section 3.1, the performance of several force fields, including MEDFF, was compared concerning the adsorption of $CH_4$ in Zr-based MOFs such as UiO-66. Important quantitative differences in the predicted uptake were established, whereas all tested force fields shared a lot of (but not all) qualitative features such as the ranking of frameworks by their high-pressure uptake. It was also found that ab initio derived force fields predict adsorption energies in better correspondence with dispersion-corrected DFT values, but this does not necessarily lead to a better agreement with exper-

iment at low pressures. By performing a sensitivity analysis, it was shown that small changes in the PES can have a dramatic impact on the predicted uptake which explains why a consistent agreement between experimental and simulated isotherms for the right reasons is hard to achieve.

A final application of MEDFF explored the role force fields can play in assisting the ab initio calculation of observables, as detailed in Section 3.2. A variant of importance sampling was introduced which allowed the computation of Henry coefficients where adsorption energies need to be calculated only for a fraction of configurations that is necessary for the conventional Widom insertion approach. This was achieved by introducing a biasing potential which picks out the relevant adsorption sites. After calculating the ab initio adsorption energy for these selected sites, the ab initio Henry coefficient could be determined using appropriate weights. In principle, any PES can be used as the biasing potential but a faster convergence is reached if the biasing potential and ab initio PES are as similar as possible. It turns out that MEDFF is the best choice among the considered force fields, and in this way MEDFF can also be considered as a tool that supports the fully ab initio calculation of quantities of interest.

Applying this importance sampling scheme to the calculation of ab initio Henry coefficients for small gas molecules in MOFs revealed some interesting results. First, it was shown that small changes in the framework geometry can have an important impact when $CO_2$ adsorbs at an open metal site, raising questions about the often invoked rigid-framework approximation for this case. Secondly, the influence of dispersion correction schemes and more specifically many-body dispersion was investigated. It turns out that, perhaps surprisingly, even for $CH_4$ in UiO-66 the three-body dispersion terms cannot be neglected when predicting gas uptake. Finally it was also shown that the generic UFF force field does not always lead to the same trends as ab initio methods, as evidenced by studying the influence of linker functionalization on selectivities of $CO_2$, $CH_4$ and $N_2$ adsorption.

An important consideration when proposing a new force field, is its practical usability for other members of the community. A first attention point is the computational cost of running simulations with this force field. In this respect, the calculation of the MEDFF energy and derived quantities is of the same computational complexity as for instance a standard Lennard-Jones potential with point-charge electrostatics. Admittedly, the pair interactions given by Equations 2.14, 2.20, and 2.25 are considerably more involved, but this can be alleviated by tabulating them at the start of a simulation. Tabulating the noncovalent potentials is the standard approach in for instance the popular DL_POLY[218] molecular dynamics simulation package and is supported by many force-field codes. Next to computational

efficiency, this approach also ensures that MEDFF can be used in many other codes without requiring additional programming. A user can simply employ Yaff, the in-house code developed at the CMM, in which the reference MEDFF implementation is available, to generate tables and use his or her preferred software package for further simulations. A second aspect is the amount of work required to obtain MEDFF parameters for the system of interest. In principle, only the electron density of the monomers and the MBIS partitioning of this density is required. Again, this is comparable to what is required for many force fields in the literature, where atomic charges need to be determined from an ab initio calculation. The MBIS partitioning method is implemented in the publicly available HORTON[219] program, which can process the electron density output of a few popular ab initio packages, so this should not present a significant obstacle. It should be noted that in case the interaction parameters ($U_{\text{exch-rep}}$, $U_{\text{ind,ct}}$, and $U_{s8}$) as determined in the original MEDFF paper are not deemed adequate for the system at hand, the fitting process to obtain more appropriate values is cumbersome and will require a significant computational cost as well as some "manual" labour. Thirdly, and finally, it is worthwhile to discuss the coverage of chemical space. While many force fields only provide parameters for a limited number of systems, the necessary MEDFF input can be obtained irrespective of which chemical elements are present, as long as an ab initio density and corresponding MBIS parameters can be computed, which is likely to be the case except for some "exotic" molecules. Of course, whether the subsequent MEDFF simulations will actually produce sensible results is not guaranteed. For the systems to which MEDFF has been applied (alkanes and gas adsorption in MOFs), mixed results have been obtained, however always observing acceptable agreement with ab initio interaction energies. Further investigation is required to find out if similar conclusions hold when different molecules and crystals are investigated using MEDFF.

A final note about the interplay of MEDFF with other force fields is in order. In the force-field community, it is not uncommon to mix energy contributions from different sources. For instance, the van der Waals model of force field A is combined with the electrostatic model from force field B, which is superimposed on the covalent model of yet another force field C. Assuming that each component faithfully describes only the interactions it is supposed to model and that the different components are additive, this is an acceptable approach. It has however been shown that for many common force fields, these conditions are not satisfied.[127] Even for MEDFF, it was shown that there is a strong interplay between the energy contributions (Section 2.2.2), as small modifications of one term needed to be compensated with changes in other terms. The mixing of certain MEDFF components with other force

fields, or the mixing of terms from different force fields in general, is therefore not advocated, except possibly when the MEDFF interaction parameters would be recalibrated accordingly.

## 4.2 Future developments

Based on the previous section with conclusions, some considerations with respect to related future research are presented.

One of the main shortcomings of MEDFF that was already encountered during the initial stages of its development, were the problems related to the induction component which prohibited a good description of for instance hydrogen-bonded complexes. The inclusion of a proper polarization model is therefore the primary candidate to make MEDFF more widely applicable. Earlier it was argued that such a model will increase the computational complexity, but recently there have been major advances in algorithms for point dipole polarizable models (such as implemented in Tinker-HP[69]). Combined with the ever increasing availability of computational power, it does not seem unlikely that many-body induction will become commonplace in new force fields. As many-body dispersion effects can be included at a similar computational cost and appear to have an important impact on condensed-phase simulations, the extension of MEDFF with such terms is also an interesting pathway. It should however be mentioned that constructing such many-body models is far from trivial. Some polarization models from the literature have been tried, but seriously underestimated the absolute value of SAPT induction energies for the S66 database. The results of this investigation are detailed in the Supporting Information of Ref. 75.

Next to modifying the MEDFF energy expression, also the determination of the interaction parameters could be revisited. At the moment, those parameters have been assigned "universal" values which are used throughout. It is however also possible to refine these values specifically for the system at hand to obtain more accurate interaction energies. While this is not completely in line with the MEDFF design philosophy, it could be argued that the reparametrization of only the three interaction parameters requires much less reference data and much less effort than constructing a complete new force field from scratch. Taking this point even further, it could also be interesting to introduce separate interaction parameters for different atoms within one molecule to test how appropriate the current expression for the MEDFF energy is.

Some of the proposed modifications and extensions of the MEDFF methodology have been investigated by Bereau et al.[220] They recently proposed

the Intermolecular Potential with Machine Learning (IPML), which uses MEDFF as a starting point, mainly aiming to make the resulting force field applicable to a wider range of chemical systems but also to obtain an even better correspondence with reference interaction energies. A first crucial difference is that the atomic parameters are determined using a machine learning model trained on partitioned ab initio densities. This means that the monomer electron density is no longer required to obtain atomic MEDFF parameters, as these can now be computed using the machine learning model. An important extension is the inclusion of additional energy terms: point multipole electrostatics, the Thole model of smeared inducible dipoles[66, 70] and the many-body dispersion energy formalism by Tkatchenko et al.[18, 221] Finally, the parameter of the repulsion model now depends on the interacting atoms. This brings the total number of interaction parameters in IPML to $4+4N$ with $N$ the number of different atom types that are considered. IPML achieves low errors on gas-phase dimers of diverse data sets, also including hydrogen-bonded complexes. Additionally, denser systems such as water clusters and the benzene crystal are considered. For the latter case, predicted cohesive energies are not far off experiment, but the potential is still slightly overstabilizing.

MEDFF was successfully applied to the simulation of gas adsorption in MOFs, showing better correspondence to ab initio adsorption energies than force fields from the literature. Because this was the case for a variety of frameworks (including MOFs with CUSs), it seems that MEDFF can be confidently used in many cases. One of the limitations of those simulations is however that only rigid frameworks were considered. While for the investigated MOFs this was a good approximation, the flexible nature of some frameworks is of paramount importance for their adsorption properties.[222] One of the future challenges is therefore the construction of covalent force fields (preferably using the in-house QuickFF approach) in conjunction with MEDFF. Combined MEDFF/QuickFF force fields would include both covalent and noncovalent interactions derived using nothing but ab initio input, which would open up possibilities for many areas of molecular modeling.

One such area is the study of NQEs. When force fields are fitted to experimental data, NQEs will be implicitly included in the parameters. If NQEs are then also included explicitly in simulation, double counting will occur. Therefore simulations explicitly including these effects should be done with force fields without empirical input (or with a direct ab initio evaluation of the PES). As shown in the simulations of liquid methane with NQEs, anharmonic effects in the covalent force field turn out to have an appreciable influence. A further investigation on how to correctly parameterize these anharmonic effects is therefore required.

## 4.3 Perspective

*This final section contains a personal reflection on the possible impact of my doctoral dissertation on future work related to the development of noncovalent force fields, but also considering molecular modeling in general. I will state some of my impressions formed by reading many papers, attending conferences, discussions with other scientists, or in short dealing with the topic of my Ph.D. day in day out for a few years. I hope it serves as an appropriate conclusion of this work and provides some insights that are more valuable than just numbers rolling out of a computer.*

The point of departure for this outlook is the concept of a "Pauling point", attributed to one of the founders of quantum chemistry and overall brilliant scientist Linus Pauling. There are at least two distinct interpretations of the Pauling point concept. The first one is the notion that a problem should be approached using the broadest possible picture that does not distort any crucial aspects. In other words, one should avoid looking so closely at a problem that irrelevant details increase the complexity to a point where the crucial features become blurred. This vision of course comes close to Occam's razor. A second interpretation, which I find more appealing, shows the other side of the coin and is explained by quoting Per-Olov Löwdin, another important name in the history of quantum chemistry:[223]

> [...]a characteristic feature of quantum chemistry is that even a fairly simple theory could sometimes give excellent agreement with experimental experience, but that this agreement may disappear whenever one tries to improve the theory. The point of excellent agreement was coined the "Pauling point" in honor of one of the great pioneers in our field[...]

There are many examples of Pauling points in the field of molecular modeling, some clearly visible but many covered by the accumulation of approximations and assumptions present in the comparison between experiment and theory. An illustrative example is found in coupled-cluster theory, which introduces a particular perturbation expansion to solve the many-electron Schrödinger equation. Different variants such as CCSD, CCSDT, ... are distinguished by which excitations from the reference orbitals are allowed (S-for single excitations, D-for double excitations, T-for triple excitations, ...). A particular feature is that CCSD(T) and CCSD[T], variants which only provide an approximation for contributions from triple excitations, are actually more accurate than the correct treatment of those triple excitations.[224] CCSD(T) or CCSD[T] can thus be considered Pauling points, as a better theoretical

description given by CCSDT actually produces worse results. The origin of this surprising behavior is a cancellation of errors arising from approximate treatment of triple excitations and neglect of higher excitations in the CCSD(T) and CCSD[T] method. As alluded to a few times in this dissertation, many results from the literature are based on such a compensation of errors, often without a general awareness of this behavior.

Löwdin extended the concept of a Pauling point, probably only partly joking, by introducing the related "Ph.D. point", illustrated in Figure 4.1, which is an adaptation from Ref. 225. The Ph.D. point refers to the fact that many Ph.D. proposals aim at solving a particular, seemingly small inconsistency in the current methodology. The Ph.D. student then gets his/her teeth into the topic and quickly discovers that current methods are less robust than proclaimed. The attempt at an improved approach can offer better insights, but unfortunately can also lead to worse agreement with experiment, which unfortunately is rather *hard to sell* in the scientific community.

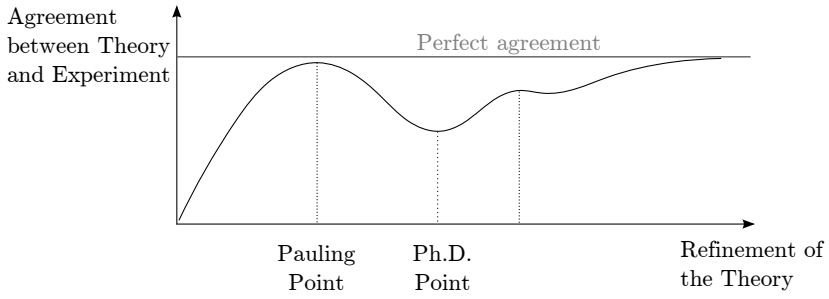



**Figure 4.1:** Graphical representation of a Pauling point, which shows that a more detailed theoretical model does not necessarily lead to a better agreement with experiment. Figure adapted from Ref. 225. Copyright (2009) John Wiley & Sons, Inc.

It is my conviction that the construction of MEDFF and the accompanying comparison with the liquid density of methane is a perfect example of the Ph.D. point, with a generic force field such as OPLS playing the role of the Pauling point, as illustrated in Figure 4.2. MEDFF obviously provides a more complex description of the interactions between methane molecules than OPLS, yet the agreement with experiments in the condensed phase is worse for the former. It was however shown that by additionally including effects such as many-body dispersion, nuclear quantum effects and anharmonic intramolecular interactions into MEDFF, results that are closer to experiment are obtained. In OPLS on the contrary, taking these effects into account worsens the results because they are already included *implicitly* when fitting the OPLS parameters.

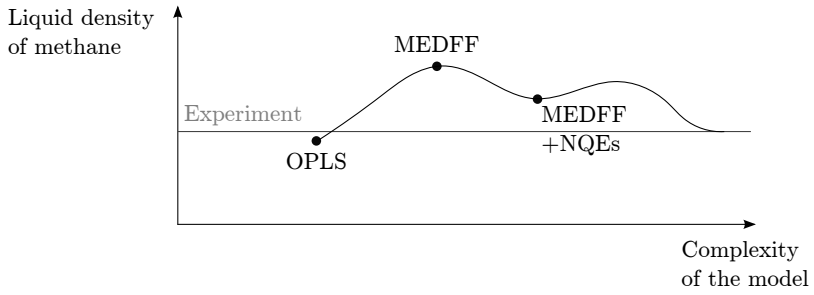



**Figure 4.2:** Pictorial representation of the liquid density of methane as predicted by various methodologies discussed in this work (not to scale). The lines between the data points are merely present to guide the eye.

A question that force-field end users like to see answered by force-field developers is: which force field provides the most accurate results for my simulations? By now it should be obvious that a clear-cut answer is not possible. It should be easy to see that MEDFF is not "the force field to end all other force fields", in the sense that it would surpass existing methods in all aspects. Even tough the following phrase is an example of bad self-marketing, I must say that it seems very unlikely to me that MEDFF will become the *de facto* standard concerning the choice of PES. The reason being that its agreement with experiment is often unsatisfactory and although ways for further improvement were explored, these will lead to an additional complexity and computational cost making it unsuitable for practical situations.

Based on these considerations, it seems that generic force fields are the way to go then. This is supported by the first interpretation of the Pauling point, advising to use the simplest model that captures the relevant features. However, because parameters are fitted directly to reproduce experiments, the predictive power of such force fields is highly questionable and they possess, at most, some "reproducing power". It also needs to be stressed that many details are brushed under the carpet in a non-transparent way. When moving away from the systems to which parameters were originally fitted, this will impact simulations without providing any real insight. I think this is where the added value of ab initio derived force fields such as MEDFF is mainly situated. They allow scientists to investigate what is happening and why, as has for example been demonstrated in the Applications chapter of this work. Even though they will often not match empirically fitted models when it comes to reproducing experiment, ab initio derived potentials can help to provide a better understanding of both limitations and success stories of generic force fields. The future of molecular simulations performed with

classical potentials thus needs empirical, generic, simple force fields as well as ab initio derived more complex models: the challenge lies in using the appropriate model for the task at hand.

# Part II

# Published Papers

# A

# Publications in International Peer-Reviewed Journals

**Paper I** **Minimal Basis Iterative Stockholder: Atoms-in-Molecules for Force-Field Development**

T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, P.W. Ayers

*J. Chem. Theory Comput.*, **2016**, *12* (8), 3894–3912

**Paper II** **The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Non-Covalent Interactions**

S. Vandenbrande, M. Waroquier, V. Van Speybroeck and T. Verstraelen

*J. Chem. Theory Comput.*, **2017**, *13* (1), 161-179

**Paper III** **Methane Adsorption in Zr-Based MOFs: Comparison and Critical Evaluation of Force Fields**

S. Vandenbrande, T. Verstraelen, J.J. Gutiérrez-Sevillano, M. Waroquier, V. Van Speybroeck

*J. Phys. Chem. C*, **2017**, *121* (45), 25309-25322

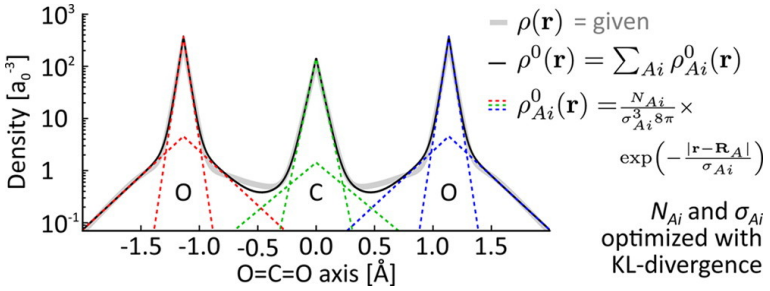
**Paper IV** **Ab initio Evaluation of Henry Coefficients Using Importance Sampling**

S. Vandenbrande, M. Waroquier, T. Verstraelen, V. Van Speybroeck

*J. Chem. Theory Comput.*, **2018**, *14* (12), 6359-6369

Paper I

# Minimal Basis Iterative Stockholder: Atoms-in-Molecules for Force-Field Development




T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, P.W. Ayers

*J. Chem. Theory Comput.*, **2016**, *12* (8), 3894–3912

S. Vandenbrande performed most simulations, interpreted the resulting data, and made contributions to the manuscript.

JCTC Journal of Chemical Theory and Computation
Article
pubs.acs.org/JCTC

# Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development

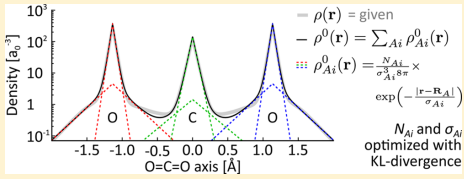
Toon Verstraelen,*,† Steven Vandenbrande,† Farnaz Heidar-Zadeh,‡ Louis Vanduyfhuys,† Veronique Van Speybroeck,† Michel Waroquier,† and Paul W. Ayers‡

†Center for Molecular Modeling (CMM), Member of the QCMM Ghent−Brussels Alliance, Ghent University, Technologiepark 903, B9000 Ghent, Belgium

‡Department of Chemistry and Chemical Biology, McMaster University, 1280 West Main Street, Hamilton, Ontario L8S 4M1, Canada

Supporting Information

**ABSTRACT:** Atomic partial charges appear in the Coulomb term of many force-field models and can be derived from electronic structure calculations with a myriad of atoms-in-molecules (AIM) methods. More advanced models have also been proposed, using the distributed nature of the electron cloud and atomic multipoles. In this work, an electrostatic force field is defined through a concise approximation of the electron density, for which the Coulomb interaction is trivially evaluated. This approximate "pro-density" is expanded in a minimal basis of atom-centered s-type Slater density functions, whose parameters are optimized by minimizing the Kullback−Leibler divergence of the pro-density from a reference electron density, e.g., obtained from an electronic structure calculation. The proposed method, Minimal Basis Iterative Stockholder (MBIS), is a variant of the Hirshfeld AIM method, but it can also be used as a density-fitting technique. An iterative algorithm to refine the pro-density is easily implemented with a linear-scaling computational cost, enabling applications to supramolecular systems. The benefits of the MBIS method are demonstrated with systematic applications to molecular databases and extended models of condensed phases. A comparison to 14 other AIM methods shows its effectiveness when modeling electrostatic interactions. MBIS is also suitable for rescaling atomic polarizabilities in the Tkatchenko−Scheffler scheme for dispersion interactions.




## 1. INTRODUCTION

The importance of force-field models is evident from recent hallmarks of atomistic force-field simulations in biology, such as the full characterization of β2 adrenergic receptor with a Markov state model[1] and the Anton 2 computer that can perform 10 μs molecular dynamics simulations per day on a system containing one million atoms.[2] In many other domains, impressive scientific breakthroughs were also realized with atomistic force-field simulations, such as the virtual screening of 87 000 zeolites for the selective adsorption of $CO_2$.[3]

An efficient and reliable model for electrostatic interactions is a fundamental component of a force-field model. For example, molecular recognition in proteins can be driven by electrostatic interactions.[4] Several authors studied partial charges derived from electronic wave functions to model electrostatic interactions in metal−organic frameworks.[5−8] Depending on the framework type, the values of the partial charges can strongly affect the predicted adsorption isotherms and self-diffusion coefficients.[9] Energy decomposition methods have also shown that electrostatic interactions are one of the main driving forces in the formation of hydrogen bonds.[10]

In this work, we propose a new and transparent method to derive, from an electronic wave function, a robust, compact, and reliable model for electrostatic interactions that is easily included in a force-field model. The goal is thus an efficient computation of electrostatic interactions between the molecules in the frozen density approximation,[10−12] i.e., without accounting for induction or polarization effects. Although it is important and challenging to account for polarization in force fields,[13−15] the development of polarizable force fields goes beyond the scope of this paper. We also do not consider so-called "polarized force fields", where polarization is described effectively by computing the charges from an electronic structure calculation with a polarizable continuum model.[16]

The most basic and widespread electrostatic force-field model consists of interacting atomic point charges placed at the positions of the nuclei. In older works, e.g., the TraPPE force field for $CO_2$,[17] the partial charges are fitted to experimental thermodynamic reference data. More recently, e.g., as in the TraPPE-EH models,[18] charges are often derived from electronic wave functions. Plenty of methods exist to compute such partial charges but usually, for force-field purposes, they are fitted to the electrostatic potential around model compounds of interest,[19] e.g., extensions of the AMBER force field often use the RESP method for partial charges.[20] The

Received: May 4, 2016
Published: July 6, 2016





DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

point-charge model is only a very crude representation of the molecular charge distribution; it does not account for finer details such as atomic multipoles[21] and the spatial distribution of the electron cloud. The spatial distribution becomes important when electron densities of two atoms or molecules begin to overlap: in that regime, point−multipole models neglect a relatively large attractive electrostatic force, which is known as the penetration effect.[12,22−26]

In principle, atomic multipoles are easily computed with an atoms-in-molecules (AIM) method. In some works, the acronym AIM is used exclusively for Quantum Theory of Atoms in Molecules[27] (QTAIM). Here, it is used more generally, to refer to any method that partitions the molecular electron density, $\rho(\mathbf{r})$, into atomic contributions, $\rho_A(\mathbf{r})$, from which, e.g., atomic multipole moments can be derived. The spatial distribution of the electron density is sometimes also modeled with density-fitting techniques, e.g., as in the Gaussian Electrostatic Model[28,29] or related methods.[30−32] This leads to very accurate models of the electronic density but the use of such advanced charge distributions in force-field simulations poses some difficulties: the conformational dependence of atomic multipoles can be very complex and it is far from trivial to include torques acting on higher moments in a force-field model. Several authors have proposed methods to overcome these challenges, e.g., with rigid molecules[33] or with machine learning methods.[34] Such advanced techniques are not always feasible for large-scale simulations. In this work, we propose a mathematically elegant and compact approximation of the electron distribution that results in relatively accurate electrostatic interactions in force-field models, without compromising computational efficiency. Only spherically symmetric models for atoms are considered and generalizations toward non-spherical atoms will be studied in future work.

Our new method minimizes the Kullback−Leibler (KL) divergence of a pro-density, a minimal expansion in atom-centered s-type Slater functions, from a given molecular electron density.[35,36] This approach is closely related to the Iterative Stockholder (IS) method,[37] where the pro-molecule density is a sum of spherical non-negative pro-atom densities, without any restrictions on their radial dependence. As will be discussed in section 2, the algorithm to optimize our pro-density parameters is also very similar to IS. Hence, we refer to our new method as Minimal Basis Iterative Stockholder (MBIS).

MBIS can be perceived in two different ways. In the first place, it is a variant of the Hirshfeld method:[38] a partitioning of the molecular electron density inspired by information theory.[39] Second, it can also be seen as a density fitting technique that uses the KL divergence,[36] instead of the more common least-squares approach with a Coulombic metric,[40] to optimize the model density. This duality also permits many applications, beyond the scope of modeling electrostatic interactions. For example, the Hirshfeld method is extensively used in different dispersion corrections for density functional theory (DFT) computations.[41−43] Furthermore, AIM populations are widely used in conceptual density functional theory to compute condensed reactivity indicators.[44]

Several related AIM methods were proposed in the literature, each trying to improve certain properties of their predecessors. The original Hirshfeld method[38] has some well-known weaknesses, such as the relatively low partial charges[45] and some deficiencies in its motivation from information theory.[46] These issues were mostly fixed in the Iterative Hirshfeld (HI) method:[46] charges computed with this method reproduce well the electrostatic potential around a molecule.[47−49] Compared to ESP-fitted charges, HI charges are also relatively robust, with respect to conformational changes, choice of basis set, etc.[49,50] Unfortunately, also the HI method has its deficiencies. For example, when the method is applied to highly polar oxides, it requires the spherically averaged density of the nonexisting oxygen dianion as input.[51] When this dianion density is computed with a localized basis set, HI charges severely overestimate electrostatic potentials of metal oxides.[51,52] The Iterative Stockholder (IS) analysis was developed independently from the HI method and it addresses most of the issues mentioned so far.[51,53] However, IS charges are not very robust, with respect to conformational changes, similar to ESP-fitted charges.[54] A recent analysis revealed that the lack of robustness is strongly related to the ill-defined density tails of the IS pro-atoms, while the core region of the IS pro-atom is usually well-defined.[30] Several authors have presented solutions to overcome the weaknesses of the HI and IS methods.[30,51,54−57] A general difficulty with these recent efforts is that they all significantly increase the algorithmic complexity and/or introduce many tuned parameters that are needed as extra input for the partitioning. In this work, we will reverse this trend and propose a method that is mathematically elegant, straightforward to implement for large systems, and free from empirical input (like atomic radii) or precomputed pro-atoms.

In the development of the MBIS method, we paid special attention to its applicability to condensed phases and extended systems. One of the applications of interest is the automatic derivation of environment specific force-field parameters for supramolecular systems[58] and porous materials.[6,59] In such applications, density partitioning is applied to DFT calculations of large atomistic models, from which force-field parameters are derived. Besides the obvious requirement that an accurate model for electrostatics must be obtained, it is also essential that the AIM method is computationally feasible for large systems. In practice, this means that the computational cost must scale linearly with the system size. This is achieved in MBIS by using only well-behaved integrals over atomic regions whose cost is independent of the system size.

The paper is organized as follows. In section 2, the MBIS method is derived using arguments from information theory, followed by more practical aspects such as numerical algorithms and software implementations. Section 3 showcases typical MBIS results with two brief applications. Section 4 compares MBIS to 14 other AIM methods, assessing the robustness of charges and the accuracy of electrostatic potentials and electrostatic interactions. Some specific advantages of MBIS over HI are presented in section 5, by testing different variants of the Tkatchenko−Scheffler dispersion model.[42] Finally, our conclusions and an outlook on future work are given in section 6.

## 2. MINIMAL BASIS ITERATIVE STOCKHOLDER METHOD

**2.1. Information Theory Approach to Hirshfeld Partitioning.** It is instructive to review the information theory arguments[39] that support the Hirshfeld method.[38] The amount of information lost when atoms-in-molecules (AIM) densities are approximated by pro-atoms, can be expressed as the sum of the KL divergence for every atom:



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

Journal of Chemical Theory and Computation Article

$$\Delta S[\{\rho_A\};\{\rho_A^0\}] = \sum_{A=1}^{N_{atoms}} \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) d\mathbf{r} \quad (1)$$

Traditionally, the pro-atom densities ($\rho_A^0$) are fixed and the AIM densities) ($\rho_A$) are the unknowns to be determined. In the original Hirshfeld method, spherically averaged isolated neutral atoms are used as pro-atoms. To obtain AIM densities that are maximally similar to the pro-atoms, one minimizes the information loss with the constraint that the AIM densities must add up to the total density: $\sum_{A=1}^{N_{atoms}} \rho_A(\mathbf{r}) = \rho(\mathbf{r})$. Hence, the optimal AIM densities are a stationary point of the following Lagrangian:

$$L_0[\{\rho_A\}, \lambda(\mathbf{r}); \{\rho_A^0\}] = \sum_{A=1}^{N_{atoms}} \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) d\mathbf{r} + \int \lambda(\mathbf{r}) \left(\sum_{A=1}^{N_{atoms}} \rho_A(\mathbf{r}) - \rho(\mathbf{r})\right) d\mathbf{r} \quad (2)$$

where $\lambda(\mathbf{r})$ is the Lagrange multiplier and $\rho_A^0(\mathbf{r})$ represents the fixed pro-atom densities. The Lagrange equations take the following form:

$$0 = \frac{\delta L_0}{\delta \rho_A(\mathbf{r})} = \ln \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} + 1 + \lambda(\mathbf{r}) \qquad \forall\ A \quad (3)$$

The solution is

$$\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} = \frac{\rho_B(\mathbf{r})}{\rho_B^0(\mathbf{r})} \qquad \forall\ A \neq B \quad (4)$$

After multiplication by $\rho_A^0(\mathbf{r})\rho_B^0(\mathbf{r})$ and summing over all atoms B, one obtains the well-known stockholder partitioning:[39]

$$\rho_A(\mathbf{r}) = \rho(\mathbf{r}) \frac{\rho_A^0(\mathbf{r})}{\rho^0(\mathbf{r})} \quad \text{with } \rho^0(\mathbf{r}) = \sum_B \rho_B^0(\mathbf{r}) \quad (5)$$

which corresponds to the definition originally given by Hirshfeld.[38] The name *stockholder* comes from the ratio $\rho_A^0(\mathbf{r})/\rho^0(\mathbf{r})$: at every point in space, it represents the *share* of pro-atom A in the total pro-density. It can be interpreted as an atomic weight function that assigns part of the total electron density to atom A. In most Hirshfeld variants,[38,46,51,53,54] the weight function varies smoothly over the range [0,1]. In QTAIM,[27] a similar atomic weight function, derived from the topology of $\rho(\mathbf{r})$, is either 1 inside the atomic basin or 0 elsewhere. Because of the minimization of the KL divergence, the Hirshfeld AIM densities are maximally similar to the pro-atoms, ensuring some degree of transferability between AIM densities in different molecules.[60]

The use of fixed pro-atoms has some important disadvantages. Results obtained with the Hirshfeld partitioning method are largely dependent on the choice of the fixed pro-atoms, which is essentially arbitrary.[46] Furthermore, $\rho_A^0(\mathbf{r})$ and $\rho_A(\mathbf{r})$ do not necessarily have the same norm ($N_A^0 \neq N_A$), such that the KL divergence cannot be used as a proper measure for information loss.[61] This shortcoming was one of the motivations to develop the HI method.[46] In HI, the pro-atoms are not fully fixed *a priori* but rather updated iteratively to achieve consistency between the charge of the pro-atom and the AIM density.

**2.2. Definition of the MBIS Partitioning.** In this paper, we will make use of the information theory concepts reviewed in the previous subsection, yet with a different model for the pro-atomic density:

$$\rho_A^0(\mathbf{r}) = \sum_{i=1}^{m_A} \rho_{Ai}^0(\mathbf{r}) \quad (6)$$

with

$$\rho_{Ai}^0(\mathbf{r}) = N_{Ai} f_{Ai}(\mathbf{r}) = \frac{N_{Ai}}{\sigma_{Ai}^3 8\pi} \exp\left(-\frac{|\mathbf{r} - \mathbf{R}_A|}{\sigma_{Ai}}\right) \quad (7)$$

where the number of Slater functions, $m_A$, is the number of shells of atom A (i.e., its row in the periodic table). Both the population ($N_{Ai}$) and the width ($\sigma_{Ai}$) of each atomic shell are free variables. The shape functions ($f_{Ai}(\mathbf{r})$) are normalized 1s Slater-type density functions ($\int f_{Ai}(\mathbf{r})\, d\mathbf{r} = 1$) and, hence, the population of a pro-atom is simply $N_A^0 = \sum_{i=1}^{m_A} N_{Ai}$. Figure 1a illustrates the expansion of the density in Slater functions, for the case of a carbon dioxide molecule.

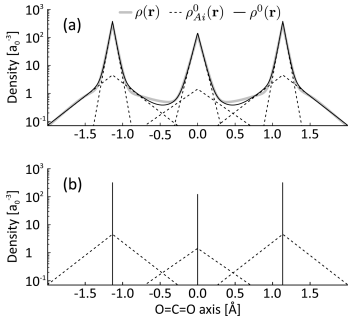



**Figure 1.** (a) An expansion of the molecular electron density of carbon dioxide (solid gray curve, $\rho$) in a minimal number of 1s Slater-type density functions (black dashed curves, $\rho_{Ai}^0$). The sum of all Slater functions is the pro-molecular density (solid black curve, $\rho^0$). (b) A reduction suitable for force-field models: for every atom A, the nuclear charge and the core Slater functions are condensed into an effective core charge (solid vertical line, $q_{A,c}$), while the valence Slater function (dashed black curve, parameters $N_{A,v}$ and $\sigma_{A,v}$) is retained.

It is clear that the pro-atom parametrization with s-type Slater functions is only applicable to (reconstructed) all-electron densities. Regardless of this requirement, the MBIS method has many advantages over existing methods, as will be extensively shown in the remainder of the paper. Future work will focus on more advanced pro-atom models, e.g., to make them also suitable for pseudo-densities, while still maintaining a numerically robust algorithm. In this work, only the most minimal, yet very effective, parametrization of the pro-atoms is considered.

All the pro-atom parameters, $\{N_{Ai}\}$ and $\{\sigma_{Ai}\}$, and the AIM densities ($\rho_A(\mathbf{r})$) will be optimized by minimizing the information loss. The main difference with the conventional Hirshfeld method is that a set of pro-atom parameters also is varied, such that the pro-atom densities become a good approximation of the AIM densities. These additional degrees of freedom also allow us to constrain the population of each pro-atom and corresponding AIM to be equal, avoiding any ambiguity in the statistical interpretation of eq 1.[61] The



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

Lagrangian for this problem is an extension of eq 2 with additional variables and constraints:

$$L_1[\{\rho_A\}, \lambda(\mathbf{r}), \{N_{Ai}\}, \{\sigma_{Ai}\}, \{\mu_A\}] = \sum_{A=1}^{N_{atoms}} \int \rho_A(\mathbf{r}) \ln\left(\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})}\right) d\mathbf{r} + \int \lambda(\mathbf{r})\left(\sum_{A=1}^{N_{atoms}} \rho_A(\mathbf{r}) - \rho(\mathbf{r})\right) d\mathbf{r} + \sum_{A=1}^{N_{atoms}} \mu_A \int (\rho_A^0(\mathbf{r}) - \rho_A(\mathbf{r}))\, d\mathbf{r} \quad (8)$$

where $\mu_A$ are new Lagrange multipliers associated with the consistency of the pro-atom and AIM populations.

Independent variation of the Lagrangian $L_1$ with respect to each variable ($\rho_A(\mathbf{r})$, $N_{Ai}$, or $\sigma_{Ai}$) leads to a set of Lagrange equations, which, together with the constraints, determine the MBIS AIM and pro-atom densities and the Lagrange multipliers $\lambda(\mathbf{r})$ and $\{\mu_A\}$.

We first consider the derivative of $L_1$ toward $N_{Ai}$:

$$0 = \frac{\partial L_1}{\partial N_{Ai}} = \int \frac{\delta L_1}{\delta \rho_A^0(\mathbf{r})} \frac{\partial \rho_A^0(\mathbf{r})}{\partial N_{A,i}} d\mathbf{r} \quad (9)$$

$$= \int \left(-\frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} + \mu_A\right) f_{Ai}(\mathbf{r})\, d\mathbf{r} \quad (10)$$

$$= \mu_A - \int \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} f_{Ai}(\mathbf{r})\, d\mathbf{r} \quad (11)$$

When we multiply by $N_{Ai}$ and sum over the shells $i$ of atom A, we get

$$0 = \sum_i^{m_A} N_{Ai} \frac{\partial L_1}{\partial N_{Ai}} = \mu_A N_A^0 - \int \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \sum_i^{m_A} N_{Ai} f_{Ai}(\mathbf{r}) \quad (12)$$

$$= \mu_A N_A^0 - N_A \quad (13)$$

Because of the constraint $N_A = N_A^0$, we have $\mu_A = 1$ for each atom. Next, we take the functional derivative of $L_1$ toward $\rho_A(\mathbf{r})$ and make use of $\mu_A = 1$:

$$0 = \frac{\delta L_1}{\delta \rho_A(\mathbf{r})} = \ln \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} + \lambda(\mathbf{r}) \quad (14)$$

whose solution is the stockholder partitioning formula in eq 5. Finally, we consider the derivative of $L_1$ toward $\sigma_{Ai}$:

$$0 = \frac{\partial L_1}{\partial \sigma_{Ai}} = \int \frac{\delta L_1}{\delta \rho_A^0(\mathbf{r})} \frac{\partial \rho_A^0(\mathbf{r})}{\partial \sigma_{A,i}} d\mathbf{r} \quad (15)$$

$$= -\int \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \frac{\partial \rho_A^0(\mathbf{r})}{\partial \sigma_{A,i}} + \mu_A N_{Ai} \frac{\partial}{\partial \sigma_{Ai}} \int f_{Ai}(\mathbf{r})\, d\mathbf{r} \quad (16)$$

$$= \int \frac{\rho_A(\mathbf{r})}{\rho_A^0(\mathbf{r})} \left(\frac{3}{\sigma_{Ai}} - \frac{|\mathbf{r} - \mathbf{R}_A|}{\sigma_{Ai}^2}\right) \rho_{Ai}^0(\mathbf{r})\, d\mathbf{r} \quad (17)$$

where we made use of $\int f_{Ai}(\mathbf{r})\, d\mathbf{r} = 1$.

The AIM densities can be eliminated from the Lagrange equations (eqs 11 and 17) by making use of the expression $\rho_A(\mathbf{r})/\rho_A^0(\mathbf{r}) = \rho(\mathbf{r})/\rho^0(\mathbf{r})$. They can be rewritten in the following form:

$$N_{Ai} = \int \rho(\mathbf{r}) \frac{\rho_{Ai}^0(\mathbf{r})}{\rho_0(\mathbf{r})} d\mathbf{r} \quad (18)$$

$$\sigma_{Ai} = \frac{1}{3N_{Ai}} \int \rho(\mathbf{r}) \frac{\rho_{Ai}^0(\mathbf{r})}{\rho_0(\mathbf{r})} |\mathbf{r} - \mathbf{R}_A|\, d\mathbf{r} \quad (19)$$

These identities form the basis for the self-consistent algorithm that will be explained in the next subsection.

**2.3. Self-Consistent Algorithm.** Figure 2 depicts a flowchart of the self-consistent algorithm discussed in this subsection. The individual steps are described in more detail below.

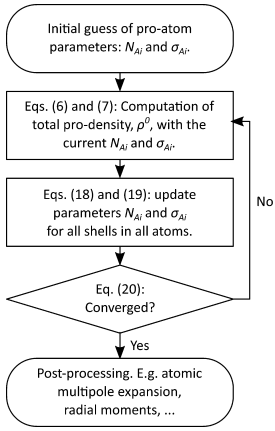



**Figure 2.** Flowchart of the self-consistent algorithm for the refinement of the MBIS parameters.

In order to determine all of the pro-atom parameters, $\{N_{Ai}\}$ and $\{\sigma_{Ai}\}$, an initial guess is generated first, which will be refined later. Because the parameters $\{\sigma_{Ai}\}$ are nonlinear, it is not guaranteed that $L_1$ is convex or has a unique minimum. Hence, multiple stationary points may exist and a reasonable initial guess is needed to find the solution of interest. The initial values of the parameters $\{N_{Ai}\}$ of atom A are set to the number of electrons in each shell of the corresponding neutral isolated atom. The initial guess of $\{\sigma_{Ai}\}$ is inspired by hydrogenic s-type orbitals. For the innermost and outermost shell of atom A, we take $a_0/(2Z_A)$ and $a_0/2$, respectively, where $Z_A$ is the atomic number. Initial values of $\sigma_{Ai}$ for the intermediate shells are fixed by geometric interpolation:

$$\sigma_{Ai} = \frac{a_0}{2Z_A^{1-[(i-1)/(m_A-1)]}}$$

.

Given the initial guess, the parameters are refined iteratively with a self-consistent update. In a single iteration, eqs 18 and 19 are evaluated, using the "old" parameters in the right-hand side, yielding the "new" parameters in the left-hand side. These iterations are repeated until the pro-atom parameters no longer change significantly. In this work, the iterative algorithm (of MBIS and other HI flavors) is stopped after the root-mean-square (rms) deviation between the pro-atom densities of the



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

last and the previous iteration drops below a threshold of $10^{-8}$ a.u.:

$$\max_{\mathrm{A}} \sqrt{\int \mathrm{d}\mathbf{r}[\rho^{0}_{\mathrm{A,last}}(\mathbf{r}) - \rho^{0}_{\mathrm{A,previous}}(\mathbf{r})]^{2}} < 10^{-8} \tag{20}$$

Other convergence criteria could be used as well, e.g., based on the gradient of the Lagrangian $L_1$.

It should also be possible to optimize the pro-atom parameters with a quasi-Newton optimizer. However, robust quasi-Newton optimizers that can handle various types of equality and inequality constraints (to fix the total population and to keep all parameters positive) are nontrivial. The algorithm sketched above satisfies all constraints at every iteration, is much easier to implement, and converges smoothly, even with very tight convergence settings. The Levenberg−Marquardt algorithm is not applicable, because it is specifically designed for least-squares objective functions, while MBIS uses the KL divergence as the objective function.

The integrals in eqs 18 and 19 must be evaluated numerically. Because each integrand is well-localized on one atom, it is possible to implement the self-consistent update with a cost that scales linearly with the number of atoms.

**2.4. Relevant Pro-Atom Parameters for Modeling Electrostatic Interactions with Force Fields.** After the optimization of pro-atom parameters, one may reduce the pro-density to a simpler picture, which is suitable for force-field models. The nuclear charge and the Slater functions associated with core electrons can be condensed into a single effective *core charge*, $q_{\mathrm{A,c}}$. This is illustrated in Figure 1b. The remaining *valence* Slater function is characterized by two parameters: its *valence population*, denoted as $N_{\mathrm{A,v}}$, and its *valence width*, which is denoted as $\sigma_{\mathrm{A,v}}$. The net atomic charge ($q_{\mathrm{A}} = q_{\mathrm{A,c}} - N_{\mathrm{A,v}}$) can be used to approximate long-range electrostatic interactions. The two remaining degrees of freedom can be used to model the penetration effect[12,22−26] (i.e., the deviation of the electrostatic interactions from the simple point-charge model when the electronic densities begin to overlap). It can be computed efficiently with analytic expressions for the Coulomb interaction between Slater densities.[25,32]

**2.5. MBIS Implementation.** In the remainder of this work, MBIS will be tested extensively with applications to theoretical electron densities of molecules and condensed phases. In these applications, the all-electron density is first computed on an integration grid suitable for the numerical evaluation of eqs 18 and 19. The implementation of these numerical integrals differs significantly between isolated molecules and periodic systems.

For isolated molecules, all-electron densities are computed with Gaussian09,[62] using density functional theory (DFT). Different functionals and Gaussian basis sets were used, as will be explained in the following sections. The MBIS partitioning of isolated molecule densities is carried out with HORTON 2.0.0,[63] which uses a standard atom-centered Becke−Lebedev integration grid.[64] This implementation can be combined with any level of theory in Gaussian09 that produces an all-electron one-particle reduced density matrix (1RDM) with the "density = current" option. Because Gaussian09 does not write out the 1RDM when relativistic corrections are used, our tests on isolated molecules are limited to molecules that contain no elements heavier than krypton.

Electron densities of periodic models are computed with the Projector Augmented Wave (PAW) method,[65] as implemented in GPAW-0.11.0.[66−68] Integrals involving the all-electron density of periodic systems are carried out as follows. In the PAW formalism, the total electron density is separated in a smoothly varying part, denoted as $\tilde{\rho}(\mathbf{r})$, and a correction for every atom A in the so-called augmentation sphere.[65] In GPAW, the smooth density is represented on an equidistant real-space grid and the corrections are evaluated on atom-centered grids in spherical coordinates. This combination of integration grids makes it possible to perform very accurate integrations involving (reconstructed) all-electron densities of periodic systems. Our second MBIS implementation can handle any type of integration grid; therefore, we used the same grid structure as in GPAW for periodic calculations. The advanced numerical techniques in this implementation, such as linear-scaling computational cost[69] and convergence acceleration, will be discussed in future work.

**2.6. Relation to Other Partitioning Methods.** The MBIS pro-atom model has been used previously, but not yet in the context of Hirshfeld partitioning. For example, a similar pro-atom model (with fixed parameters) was also used in an ESP fitting scheme.[70] A similar density model is also used in the Stewart−Slater method.[71] Although our pro-atom model is obviously inspired by Slater's work on atomic shielding constants,[72] the typical polynomial prefactors are omitted. This omission is inspired by the piecewise exponential ansatz from statistical models for atomic densities.[73−75] The reduced model for force-field applications in subsection 2.4 has also been previously used in the development of force-field models.[31,32,76]

Our approach is comparable to the IS method.[53] In IS, spherical pro-atoms are defined by generic radial functions without further restrictions in terms of density basis functions; in practice, they are represented by function values on a radial grid. A self-consistent update, in the same spirit as eqs 18 and 19, guarantees that the optimal IS atoms minimize the KL divergence over all possible spherically symmetric pro-atoms.[53,77] Even though this is a convex problem, a well-documented weakness of IS is that the density tails of the pro-atoms are ill-defined, which leads to numerical instabilities and poorly defined atomic charges.[30,54,77] In MBIS, this is resolved by modeling the density tail of each atom with only a single Slater function, which is comparable to the BS-ISA+DF method.[30] The MBIS self-consistent update algorithm is also very similar to the HI algorithm.[46] The main difference with HI is that MBIS makes use of an analytic ansatz for each pro-atom with several parameters per atom (i.e., the populations and widths of all shells in each atom), while HI varies just one population parameter per atom and makes use of precomputed isolated atom densities. Furthermore, HI cannot be derived by replacing in Lagrangian L1 the MBIS pro-atom by its HI counterpart.[35]

It is also important to realize that density fitting[78−80] is closely related to MBIS. This connection becomes clear by considering the following Lagrangian:

$$L_2[\{N_{\mathrm{Ai}}\}, \{\sigma_{\mathrm{Ai}}\}, \mu] = \int \rho(\mathbf{r}) \ln\left(\frac{\rho(\mathbf{r})}{\rho^{0}(\mathbf{r})}\right) \mathrm{d}\mathbf{r} + \mu \int \rho^{0}(\mathbf{r}) - \rho(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{21}$$

The self-consistent update eqs 18 and 19 can also be derived from $L_2$. This shows that the optimal MBIS pro-atom parameters can also be found by minimizing the KL divergence of the pro-molecule density, $\rho^{0}(\mathbf{r})$, from a given molecular density, $\rho(\mathbf{r})$. This interpretation is similar to density fitting in force-field development,[28,29] except for the following two



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

Journal of Chemical Theory and Computation Article

**Table 1. Comparison of HI and MBIS Charges for a Selection of Molecules and Solids (See Text)[a]**

| | | | Charge, $q$ [e] | | | | |
|---|---|---|---|---|---|---|---|
| molecule or solid | $atom_{type}$ | oxidation number, ON | iterative Hirshfeld (HI) method | minimal basis iterative stockholder (MBIS) method | core charge, $q_{A,c}$ [e] | valence charge, $q_{A,v}$ [e] | valence width, $\sigma_{A,v}$ [Å] |
| MIL-53(Al) | $H_{ph}$ | 0 | 0.132 | 0.155 | 1.0 | −0.845 | 0.198 |
| $H_2O_2$ | H | 1 | 0.387 | 0.414 | 1.0 | −0.586 | 0.186 |
| $H_2O$ | H | 1 | 0.436 | 0.443 | 1.0 | −0.557 | 0.187 |
| MIL-53(Al) | $H_{hy}$ | 1 | 0.565 | 0.519 | 1.0 | −0.481 | 0.175 |
| $LiO_2^{\cdot}$ | Li | 1 | 0.912 | 0.825 | 1.076 | −0.251 | 0.387 |
| MIL-53(Al) | $C_{pc}$ | 0 | −0.132 | −0.148 | 4.359 | −4.507 | 0.266 |
| MIL-53(Al) | $C_{ph}$ | 0 | −0.087 | −0.110 | 4.354 | −4.463 | 0.266 |
| CO | C | 3 | 0.144 | 0.108 | 4.327 | −4.218 | 0.270 |
| MIL-53(Al) | $C_{ca}$ | 4 | 0.910 | 0.849 | 4.340 | −3.491 | 0.246 |
| $CO_2$ | C | 4 | 0.847 | 0.863 | 4.340 | −3.477 | 0.243 |
| MgO | O | −2 | −2.220 | −1.934 | 6.243 | −8.177 | 0.247 |
| chabazite | $O_b$ | −2 | −1.497 | −1.261 | 6.335 | −7.596 | 0.222 |
| chabazite | $O_r$ | −2 | −1.480 | −1.250 | 6.337 | −7.588 | 0.222 |
| MIL-53(Al) | $O_{hy}$ | −2 | −1.952 | −1.220 | 6.329 | −7.548 | 0.223 |
| chabazite | $O_x$ | −2 | −1.431 | −1.219 | 6.344 | −7.563 | 0.220 |
| quartz | O | −2 | −1.473 | −1.213 | 6.343 | −7.555 | 0.221 |
| $O^-$ | O | −1 | −1.0 | −1.0 | 6.194 | −7.194 | 0.246 |
| $H_2O$ | O | −2 | −0.872 | −0.885 | 6.333 | −7.219 | 0.220 |
| MIL-53(Al) | $O_{ca}$ | −2 | −0.781 | −0.748 | 6.354 | −7.103 | 0.214 |
| $CO_2$ | O | −2 | −0.424 | −0.431 | 6.380 | −6.811 | 0.208 |
| $H_2O_2$ | O | −1 | −0.387 | −0.414 | 6.351 | −6.765 | 0.212 |
| $LiO_2^{\cdot}$ | O | −1/2 | −0.456 | −0.412 | 6.330 | −6.743 | 0.215 |
| $O_3$ | $O_t$ | 0 | −0.194 | −0.177 | 6.364 | −6.541 | 0.207 |
| CO | O | −3 | −0.144 | −0.108 | 6.398 | −6.506 | 0.201 |
| O | O | 0 | 0.0 | 0.0 | 6.348 | −6.348 | 0.207 |
| $O_2$ | O | 0 | 0.0 | 0.0 | 6.371 | −6.371 | 0.203 |
| $OF_2$ | O | 2 | 0.156 | 0.121 | 6.367 | −6.247 | 0.203 |
| $O_3$ | $O_c$ | 0 | 0.389 | 0.354 | 6.386 | −6.032 | 0.197 |
| $O^+$ | O | 1 | 1.0 | 1.0 | 6.431 | −5.431 | 0.182 |
| $OF_2$ | F | −1 | −0.078 | −0.060 | 7.381 | −7.441 | 0.182 |
| MgO | Mg | 2 | 2.220 | 1.934 | 10.551 | −8.617 | 0.117 |
| MIL-53(Al) | Al | 3 | 2.780 | 2.111 | 3.215 | −1.104 | 0.352 |
| quartz | Si | 4 | 2.946 | 2.425 | 4.425 | −2.000 | 0.314 |
| chabazite | Si | 4 | 2.944 | 2.490 | 4.395 | −1.905 | 0.316 |

[a]Results are grouped per element and sorted by the MBIS charge within each group. For the MBIS method, also the quantities from subsection 2.4 also are reported.

points. First, the MBIS pro-density is expanded in Slater functions while density-fitting techniques usually rely on contracted Gaussian functions, also with higher multipoles.[40] Second, MBIS uses the KL divergence as a cost function to fit the pro-atom parameters, while conventional density-fitting makes use of a least-squares cost, often with a Coulomb metric. The least-squares cost function was also used in other related works, e.g., the least-squares analogue of IS is known as Stewart atoms[81] and the least-squares analogue of MBIS is very similar to Stewart–Slater atoms.[71] Hybrid approaches, combining least-squares and KL-divergence cost functions, were also proposed, such as Hirshfeld-E,[51] Gaussian ISA,[54] and BS-ISA +DF.[30] In the development of the MBIS method, a least-squares cost function was avoided, because it was recently found to lead to nonlocal AIM densities.[36] Finally, note that the pro-atom parameters are sufficient to construct monopolar electrostatic force fields. Therefore, we expect that the direct optimization of the pro-atoms with a Lagrangian similar to eq 21, i.e., without constructing AIM densities, can be an attractive alternative to conventional AIM methods.

## 3. EXAMPLE MBIS APPLICATIONS

This section provides two illustrative applications of the MBIS method. Their main purpose is to show the applicability of MBIS in very different scenarios and to provide the reader with some typical results. The first example discusses the robustness of MBIS and its compatibility with chemical intuition, when applied to rather extreme variations of the oxidation states of oxygen. The second example shows that MBIS is also sufficiently robust when studying subtle variations of the



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

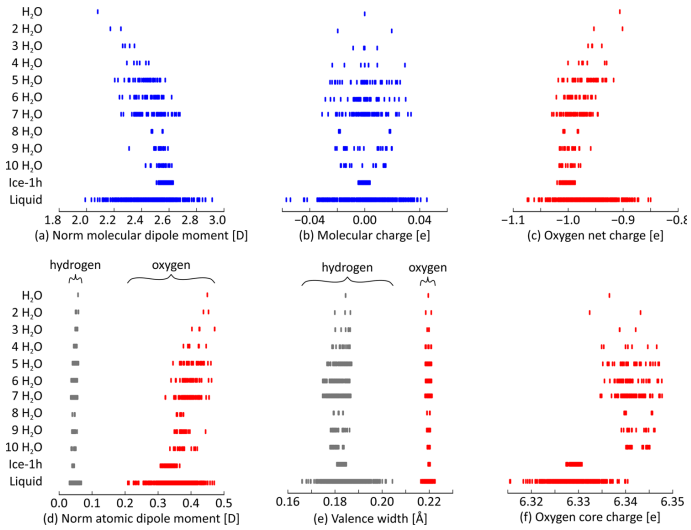



**Figure 3.** MBIS results for different phases of water (isolated water molecule, water clusters, ice-1h, and liquid water) derived from all-electron densities computed at the BLYP level of theory: (a) the norm of the molecular dipole moment of each water molecule, (b) the net charge of each water molecule, (c) the oxygen charge of each water molecule, (d) the norm of each atomic dipole moment, (e) the valence width of each atom ($\sigma_{A,v}$; see subsection 2.4), and (f) the atomic core charge of each oxygen atom ($q_{A,c}$; see subsection 2.4).

electron density of water between the gas, liquid, and solid phases.

The examples below only illustrate the usefulness of the MBIS method. A more systematic assessment can be found in section 4.

**3.1. Oxygen in Different Oxidation States.** In previous studies, Hirshfeld-I (HI) partitioning was criticized for its poor applicability to oxides.[51,52,55] During the iterative convergence of the charges, HI requires reference densities for the oxygen dianion (or sometimes even trianion) in vacuum, which does not exist.[55] When the oxygen dianion is computed with a finite basis, one obtains a very diffuse density that is not representative for the oxygen atom in a molecule or crystal. This mismatch results in very large absolute values for the atomic charges in oxides, overestimating the polarity of oxide clusters or the electrostatic potential in solid oxides.[52,55] Many modifications of HI were proposed to surmount this limitation,[51,55−57,82−84] often using different (somewhat arbitrary) techniques for the computation of unstable anions. The MBIS method does not need (unstable) ion densities as input and one would therefore expect that it does not suffer from the same overpolarization issues as HI.

Table 1 compares MBIS and HI results for the oxygen element, in a series of systems where the oxidation state of oxygen varies from −2 to +3: MgO, chabazite, MIL-53(Al), quartz, $H_2O$, $CO_2$, $H_2O_2$, $LiO_2$, $O_3$, CO, $O_2$, and $OF_2$. Results for other elements in these systems also are given, for the sake of completeness. In addition, the isolated oxygen cation, atom, and anion are included because these have different valence electron densities that result in different parameters for the outer shell in the MBIS pro-density. Atom types, which are used to differentiate all nonequivalent atoms, are defined in section S1 in the Supporting Information. All electron densities are computed at the PBE level of theory. For isolated molecules, Gaussian09[62] was used with the 6-311+G(2df,p) basis and charges were computed with HORTON.[63] Electron densities of crystal unit cells were computed with GPAW,[66−68] using a grid spacing of 0.1 Å and charges were derived from the periodic densities with a second implementation of MBIS. (See subsection 2.5.)

The main trend in Table 1 is the strong correlation between HI and MBIS charges. For systems where HI was found to be useful for force-field development, MBIS gives very comparable results. However, when oxygen has an oxidation state of −2 and has (semi-ionic) bonds to cations with a high oxidation number, MBIS charges for oxygen are less negative, making them more suitable for force-field development.

A reasonable correlation between atomic charges and oxidation numbers is found. Such correlations are not expected to be perfect, because the oxidation number is based on simple counting rules that do not account for the (partial) covalent character of chemical bonds. It may be surprising that the core charge, $q_{A,c}$, is systematically larger than the integer value one would get by combining the nuclear charge and an integer number of core electrons. Because the valence Slater function does not decay toward the nucleus, it also contributes to the core region, which is compensated by a slightly more positive core charge. All the variations in the net charge are reflected in the valence charge ($q_{A,v}$). The valence width ($\sigma_{A,v}$) linearly correlates with the net charge: within each group of a given element, more negative atoms tend to have a slightly larger valence width.

**3.2. Application of MBIS to the Three Phases of Water.** The MBIS method will first be illustrated with an application to an isolated water molecule, 38 clusters of water molecules, a model for the hexagonal phase of ice, and 10



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

snapshots of a liquid water MD simulation. The electron densities of the isolated systems were computed with Gaussian09[62] at the BLYP/6-311+G(2df,p) level of theory.[85,86] The 38 clusters, ranging from 2 to 10 water molecules in size, were taken from the work of Temelso et al.[87] We used the 3 × 3 × 2 model for the ice-1h phase of water from the work of Hayward and Reimers.[88] This model contains 96 water molecules and is tuned for computational applications: the water molecules have realistic randomized orientations, yet the net dipole moment of the unit cell is constrained to zero. The geometry of the ice-1h model is refined with CP2K-2.6.0[89−92] at the BLYP-D3 level of theory,[64,85,93,94] using the MOLOPT-DZVP-SR-GTH[95] basis set and GTH pseudo-potentials.[96,97] CP2K was also used to generate periodic structures of liquid water (32 molecules per unit cell). These geometries were sampled every 10 ps from a 100 ps NVT[98] molecular dynamics run at 300 K and at the experimental density, using the same level of theory. The electron densities of all periodic structures (ice and liquid water snapshots) were computed with GPAW[66−68] using the BLYP functional and a grid spacing of 0.1 Å, as explained above. (CP2K was not used for this purpose, because it cannot print out a reconstructed all-electron density on suitable integration grids.)

Figure 3 displays the key MBIS results for the water systems in this section. The partitioning of the density into atomic contributions, $\rho_A(\mathbf{r})$, is first used to construct electron densities of separate water molecules from which multipoles can be derived (relative to the molecular center of mass). The most obvious result is the increase of the molecular dipole moment as water forms hydrogen bonds with surrounding molecules (Figure 3a). Similar trends are usually found in simulations of water with polarizable force fields.[99−102] This increase is seen throughout all water clusters and the solid ice-1h phase. Liquid water exhibits relatively large random fluctuations in the molecular dipole moment, due to variations in the water geometry and its local environment. Another clear trend is that water molecules have a tendency to exchange a small fraction of an electron with their surrounding, leading to nonzero molecular charges (Figure 3b).

Atomic charges and dipole moments are directly derived from the AIM densities, $\rho_A(\mathbf{r})$. The increased polarization of water in larger clusters (Figure 3c) is due to the decrease of the (negative) oxygen charge, while the norm of the oxygen dipole moment follows the opposite trend (Figure 3d), slightly reducing the overall polarization. Hydrogen atoms have a small and constant dipole moment, showing that they are only weakly polarizable.

Because the MBIS pro-molecular density is a sum of spherical atoms, it was to be expected that AIM dipole fluctuations play a minor role, compared to atomic and molecular charge fluctuations. In general, the partitioning of the total polarization into contributions from atomic charges and/or dipoles is inherently ambiguous and can be strongly dependent on the AIM method. However, Mei et al. observed, for a large set of molecules and for all AIM methods tested in their work, that the overall polarization always involves a significant amount of charge fluctuations.[103]

MBIS pro-atom parameters reveal additional trends that are not easily observed with other methods. The *valence width*, $\sigma_{A,v}$ as introduced in subsection 2.4, is also sensitive to the molecular environment, which is most notable for the hydrogen atoms, while oxygen has a more constant valence width (Figure 3e). Polarizable force fields usually consider fluctuating atomic charges and/or dipoles, but fluctuations in the width of the atomic electron distribution are rarely included. Our results indicate that these may also be relevant to model electronic polarization.

Finally, Figure 3f shows the *atomic core charge* ($q_{A,c}$), defined in subsection 2.4. This quantity varies relatively little, which is consistent with the expectation that the properties of core electrons should be transferable. There is a small but notable difference between the core charge for Gaussian09 (1−10 $H_2O$) and GPAW (ice-1h, liquid) calculations. GPAW calculations on the isolated clusters confirm that this is due to the different treatment of the core electrons in both programs (results not shown).

Generally, the MBIS results for different phases of water show that the method is robust enough to uncover several subtle trends in the electronic polarization, which is very helpful for the interpretation of these trends and the construction of polarizable force fields.

## 4. SYSTEMATIC COMPARISON TO OTHER AIM METHODS

In this section, the MBIS method is compared to a series of other AIM methods, using several molecular datasets. This assessment focuses on properties that are relevant for modeling electrostatic interactions in force fields: the quality of the electrostatic potential (ESP), the accuracy of electrostatic interactions, and the robustness of the charges.

**4.1. Molecular Datasets.** Several datasets of molecular dimers and isolated molecules are considered in this section: five were taken from the literature and two new datasets are introduced below. All molecular electron densities were computed at the B3LYP/6-311+G(2df,p) level of theory with Gaussian09.[62]

Three sets of molecular dimers were taken from the work of Hobza et al., namely, S66 (diverse noncovalent interactions between neutral organic molecules),[104] IHB15 (ionic hydrogen bonds),[105] and X40 (halogen bonds).[106] From the X40 set, dimers containing iodine were omitted, because proper all-electron densities for such heavy elements can only be computed with relativistic corrections. (See subsection 2.5.) Also, a new set of molecular dimers is introduced, i.e., ZG237, which consisted of a set of 237 dimers of silica clusters and typical guest/template molecules for porous media. Neutral and anionic silica clusters are present in ZG237 and the guest molecules include noble gases, as well as neutral and cationic organic molecules. (More details are provided in section S2 in the Supporting Information.) The goal of this assessment with molecular dimers is to test how well atomic charges obtained with different AIM methods can reproduce the electrostatic interaction.

Three datasets of larger isolated molecules are also used in the tests below, two of which were taken from earlier work: PENTA103 (103 random penta-alanine conformers)[49] and SILICA245 (topologically different hydrogen-terminated silica clusters containing up to 8 Si atoms).[51] One new set, MIL53(M)10, was created based on our experience with the development of a flexible force field for the metal−organic framework MIL-53(AL).[6] This set contains 10 organometallic clusters with the same structure (see Figure 4), but with different metals in a trivalent (+3) oxidation state: Al, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, and Ga. The spin multiplicity of each cluster was fixed by coupling the spins of the transition metals to obtain a maximal $\langle \hat{S}_z \rangle$ value. The $\mu$-OH group is located at



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

Journal of Chemical Theory and Computation Article

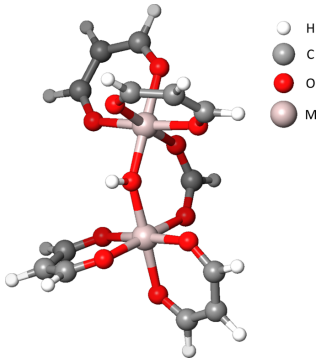



**Figure 4.** Depiction of a metal-oxide cluster, where M is a metal atom in a trivalent (+3) oxidation state (Al, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, and Ga).

the center and the cluster is carefully terminated by four malondialdehyde anions and one formic acid anion. This neutral configuration is stable for many first-row transition metals (excluding Cu and Zn), and it resembles well the metal-oxide structure found in the MIL-53 framework.[107]

**4.2. Selection of AIM Methods.** Three categories of AIM methods are used for comparison: ESP-fitted charges, density partitioning methods (Hirshfeld variants and QTAIM), and Hilbert-space partitioning methods.

Atomic charges fitted to the electrostatic potential (ESP) are among the most ubiquitous for the development of force-field models. We selected four such variants: Merz−Singh−Kollman (MSK),[108] CHELPG,[109] Restrained ESP (RESP),[20] and Hu-Lu-Yang (HLY).[70] MSK, CHELPG, and HLY differ in the way the volume around the molecule is sampled in the fitting procedure, but they all make use of a standard least-squares procedure. The RESP method extends the MSK cost function with hyperbolic restraints to penalize large absolute atomic charges.

A large selection of Hirshfeld variants is used in our comparison, starting with the original Hirshfeld method (H).[38] CM5 is a popular empirical correction to the Hirshfeld method to better reproduce experimental dipole moments of small molecules.[110] Another popular related method is Hirshfeld-I (HI).[46] While the Hirshfeld method simply uses spherically averaged neutral atoms as pro-atoms, HI iteratively updates the pro-atoms to enforce consistency between the AIM charges and the charges of the pro-atoms. Charged pro-atoms are constructed by a linear interpolation between spherically averaged densities of isolated neutral atoms and ions. A particular improvement of HI over Hirshfeld is that HI charges make a good estimate of the electrostatic potential of organic molecules.[47] This is no longer the case for metal oxides, e.g., the ESP in the pores of zeolites, which inspired several groups to further improve the method, leading to variants such as Hirshfeld-E[51] (HE) and DDEC4.[56] The IS analysis is another variant of the Hirshfeld method, proposed independently of HI but with many similarities.[37] IS also iteratively updates its pro-atoms but just uses the spherical averages of the AIM densities from the previous iteration as the new pro-atoms. Besides all the Hirshfeld variants mentioned so far (H, CM5, HI, HE, DDEC4, and IS), one more density-based method, QTAIM,[27] is also included in the comparison.

The third group of methods are Hilbert-space methods, which partition the density matrix instead of the density. Mulliken (M) is the oldest AIM method[111] and two popular improvements of this scheme are also widely used: Löwdin (L)[112] and Natural (N) charges.[113] These three Hilbert-space methods assume that each orbital basis function is centered on one of the atoms, which is always the case when using standard Gaussian basis sets. However, in several popular periodic DFT codes, e.g., VASP, CPMD, and GPAW, such information is not available, because they use delocalized basis sets.

Results for methods MSK, CHELPG, CM5, M, L and N were obtained with Gaussian09.[62] The MSK or CHELPG methods make use of van der Waals radii but do not define them for all elements, in which case UFF radii were used instead. Gaussian-formatted checkpoint files were used to post-process the densities with HORTON-2.0.0[63] to compute the HLY, H, HI, HE, IS, and MBIS charges. The RESP program from the Antechamber program[114] was used to compute the RESP charges. DDEC4 charges were computed with Chargemol-09.15.2014,[115] and QTAIM charges were computed with AIMAll-11.06.19.[116] Practically all MBIS results below are

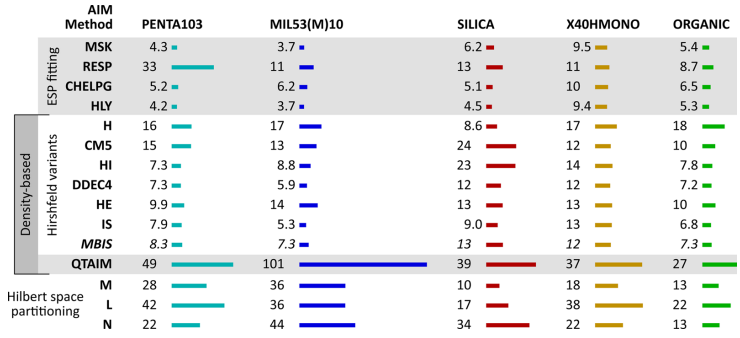



**Figure 5.** $RMSE_{ESP}$ (in kJ $mol^{-1}$) computed for all AIM methods in this work, averaged over groups of isolated molecules. (See text for definition of groups.)



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

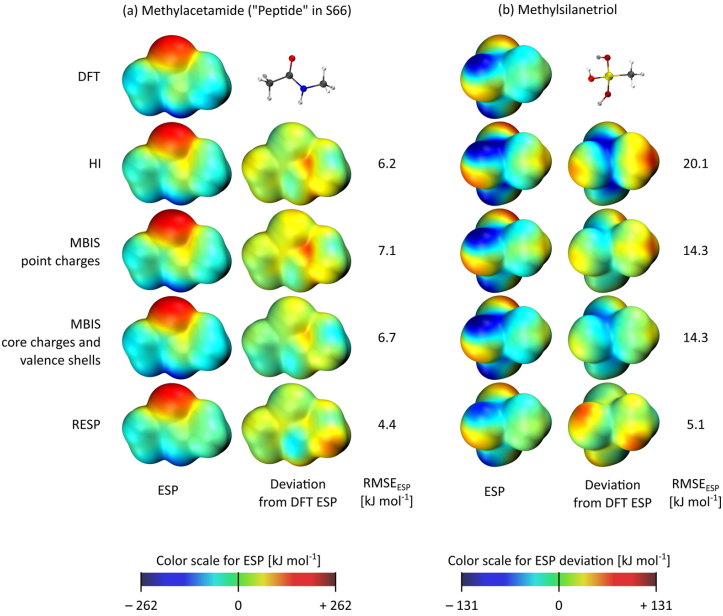



**Figure 6.** ESP maps plotted on the $\rho = 0.002$ a.u. isosurface of (a) methylacetamide and (b) methylsilanetriol. In the left column for each molecule, the DFT ESP and the ESP due to selected point-charge methods (HI, MBIS, and RESP) and the core charge + valence shell (MBIS) are shown. For each approximate ESP, the deviation from the DFT reference is shown in the right column, for each molecule.

obtained from the pro-atom parameters discussed in section 2.4. MBIS AIM densities, as such, are not used unless noted otherwise.

**4.3. Quality of the Electrostatic Potential (ESP).** In the context of force-field development, it is assumed that whenever atomic charges accurately reproduce the electrostatic potential (ESP) around a molecule, they also make good predictions of the electrostatic interactions.[19] Hence, one of the desirable properties of atomic charges is their ability to reproduce the ESP as well as possible, which is the topic of this subsection. A direct assessment of the quality of electrostatic interactions is discussed in subsection 4.4.

We have tested the quality of the ESP for all sets of isolated molecules discussed in section 4.1 and also for the monomers present in all dimer datasets. The HLY ESP cost function is used to measure the quality of the ESP and the results would not change much if we had used an MSK or CHELPG cost function instead. The HLY cost functions is an integration over a volume surrounding the molecule,[70] which we converted to an RMSE value as follows:

$$\mathrm{RMSE}_{\mathrm{ESP}} = \sqrt{\frac{\int w_{\mathrm{HLY}}(\mathbf{r})\left(V_{\mathrm{DFT}}(\mathbf{r}) - \sum_{\mathrm{A}} \frac{q_{\mathrm{A}}}{4\pi\epsilon_0 |\mathbf{r} - \mathbf{R}_{\mathrm{A}}|}\right)^2 \mathrm{d}\mathbf{r}}{\int w_{\mathrm{HLY}}(\mathbf{r})\,\mathrm{d}\mathbf{r}}} \tag{22}$$

where $w_{\mathrm{HLY}}(\mathbf{r})$ is the weight function designed by Hu, Lu, and Yang: it becomes 1 in the region surrounding a molecule and goes smoothly to zero inside the molecule and at larger distances.[70] $V_{\mathrm{DFT}}(\mathbf{r})$ is the reference ESP from the DFT calculation and $q_{\mathrm{A}}$ are the atomic charges. The smoothness of the weight function guarantees that the ESP cost is not sensitive to the exact position of the grid points, which is a clear advantage over other ESP fitting methods. $\mathrm{RMSE}_{\mathrm{ESP}}$, which we computed for every molecule and every AIM method, is a measure for the error on the frozen-density interaction energy[10−12] of a unit charge with the molecule when it is placed near its van der Waals surface.

Figure 5 compares, for every AIM method, the average of $\mathrm{RMSE}_{\mathrm{ESP}}$ within five groups of isolated molecules: PENTA103 and MIL53(M)10 are those discussed in section 4.1. SILICA contains all those of the SILICA245 set plus all silica clusters from the ZG237 set. X40HMONO contains all halogenide molecules present in the X40 set of dimers. Finally, ORGANIC contains all other monomers from the dimer sets S66, IHB15, X40, and ZG237. (The noble-gas atoms from ZG237 are not included.)

Obviously, the ESP-fitted charges (MSK, RESP, CHELPG, and HLY) perform well in this test, because they are optimized to reproduce the ESP surrounding each molecule. RESP charges are not as optimal as the other three, because of the hyperbolic restraints, which becomes very pronounced for large molecules as in the PENTA set. Reducing the strength of the restraints could relieve this issue, but it would also result in less-robust charges. The halogenides in X40HMONO have an ESP that is relatively difficult to reproduce with point charges: the sigma-hole of the halogen atom corresponds to a large and local



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

dipole moment whose effect on the ESP cannot be explained in terms of atomic monopoles.[58,117]

The original Hirshfeld method (H) usually predicts poor ESPs, because the absolute values of the atomic charges are too low.[45] All variants of the Hirshfeld method (HI, CM5, DDEC4, HE, IS, and MBIS) produce more-accurate ESPs, except for CM5 and HI when tested with the SILICA set. The good performance of the IS method is not surprising: it partitions the electron density in AIMs that are as spherical as possible, sometimes by introducing an unreasonable radial dependence,[48] thus having small atomic dipole and higher multipole moments. In fact, any method that performs better than IS likely biases the charges to mimic effects of atomic multipoles. Such overfitting clearly occurs in the ESP-fitting methods. In the case of MBIS, we only considered MBIS point charges and not the more-advanced model with valence Slater functions (see Figure 1b), simply because eq 22 only tests the ESP outside the molecule, where the density is very low. In this region, the ESP generated by the Slater functions is very well-approximated by that of point charges. Of all Hirshfeld variants, DDEC4, IS, and MBIS are comparably good.

Consistent with previous observations, QTAIM charges are inadequate for the purpose of modeling ESPs.[51] This can only be fixed by including higher atomic QTAIM multipoles, as is often done in QTAIM-based force fields.[34] Mulliken (M), Löwdin (L), and Natural (N) charges rarely produce useful ESPs in our tests.

Figure 6 shows ESP maps plotted on the $\rho = 0.002$ a.u. isosurface of two representative molecules: methylacetamide and methylsilanetriol, which have electrostatic potentials that are, respectively, easy and difficult to reproduce with point charges. The isosurface approximates the molecular van der Waals surface,[118] which is convenient for visualizing non-covalent interactions. In addition to the ESP of the DFT calculation, the ESPs obtained with a subset of atomic charge methods, and their deviation from the DFT result, are shown. The isosurfaces sample the ESP at a higher density than the HLY cost function, which has two important consequences. First, the scale of the ESP deviations is large, compared to the reported $RMSE_{ESP}$ values. Second, the ESP maps are different for MBIS point charges and MBIS core charges with delocalized valence shells, the latter accounting for the penetration effect.

The main observation is that all model ESPs qualitatively agree with the DFT result. (See the left-hand column in Figures 6a and 6b.) Even though RESP charges have a relatively low $RMSE_{ESP}$, the deviations from the DFT ESP are not significantly smaller than those for the other methods. When MBIS core charges and valence shells are used to estimate the ESP, a better visual agreement is found, because the penetration effect is already significant at the selected isodensity surface. Also note that the ESP obtained with HI charges for methylsilanetriol deviates the most from the DFT reference, which is consistent with the limitations of HI for oxides, which were also discussed in subsection 3.1 and also are seen in Figure 5. Finally, note that this visualization of two representative molecules merely serves as an illustration. Solid conclusions can only be drawn from a thorough statistical analysis involving many molecules, such as the one presented in Figure 5.

**4.4. Accuracy of Electrostatic Interactions.** A common assumption in force-field development is that ESP-fitted charges also reproduce electrostatic interaction energies in general. Here, we assess the validity of this assumption for molecular dimers: the electrostatic interaction in the frozen-density approximation[10−12] ($E_{FD}$) will be used as a reference to test approximate electrostatic interactions obtained with atomic point charges from different AIM methods. The frozen-density approximation does not include any effects from polarization or charge transfer. Such effects should be modeled with a polarizable (or polarized) force field, which is beyond the scope of this test.

The four sets of molecular dimers described in section 4.1 (S66, IHB15, X40, and ZG237) cover a large variety of electrostatic interactions, from as little as −0.06 kJ mol$^{-1}$ to rather extreme values of −664 kJ mol$^{-1}$. Especially in the ZG237 set, it is often hard to classify dimers into specific interaction types, such as hydrogen bonding, salt bridge, etc. To facilitate the interpretation of the results, we have classified the dimers more conveniently, just using thresholds on the strength of the electrostatic interaction in the frozen-density approximation: "weak" ($E_{FD} > -10$ kJ mol$^{-1}$), "medium" ($-10$ kJ mol$^{-1} \geq E_{FD} > -50$ kJ mol$^{-1}$), and "strong" ($E_{FD} \leq -50$ kJ mol$^{-1}$). Figure 7a shows the numbers of dimers from each dataset in each class. The IHB15 set contributes exclusively to the "strong" class, while all other sets have dimers in each class of interaction strength. Figure 7b shows the root-mean-square error (RMSE) on the electrostatic interaction energy for each

(a) Number of dimers from each dataset per category

| Dataset | Weak | Medium | Strong |
|---|---|---|---|
| S66 | 16 | 39 | 11 |
| IHB15 | 0 | 0 | 15 |
| X40 | 13 | 13 | 5 |
| ZG237 | 63 | 83 | 91 |

(b) RMSE on the electrostatic interaction

| AIM Method | Weak | Medium | Strong |
|---|---|---|---|
| MSK | 5.9 | 15 | 41 |
| RESP | 5.9 | 17 | 51 |
| CHELPG | 5.9 | 17 | 43 |
| HLY | 5.8 | 14 | 40 |
| H | 6.1 | 25 | 87 |
| CM5 | 6.0 | 20 | 54 |
| HI | 5.5 | 15 | 35 |
| DDEC4 | 5.8 | 15 | 36 |
| HE | 5.9 | 20 | 47 |
| IS | 5.9 | 14 | 31 |
| *MBIS* | *5.7* | *13* | *30* |
| *MBIS-S* | *2.5* | *4.9* | *30* |
| QTAIM | 7.2 | 22 | 39 |
| M | 5.8 | 23 | 61 |
| L | 6.2 | 26 | 89 |
| N | 5.9 | 11 | 29 |

**Figure 7.** Accuracy of the electrostatic interactions for three classes of electrostatic interactions: "weak" ($E_{FD} > -10$ kJ mol$^{-1}$), "medium" ($-10$ kJ mol$^{-1} \geq E_{FD} > -50$ kJ mol$^{-1}$), and "strong" ($E_{FD} \leq -50$ kJ mol$^{-1}$). Panel (a) shows the number of dimers that each dataset contributes to each class. Panel (b) shows the RMSE of the electrostatic interaction (in units of kJ mol$^{-1}$), obtained with atomic charges from different AIM methods, for each group. MBIS-S goes beyond the simple point-charge model and uses spherical valence Slater functions to model the penetration effect.



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

Journal of Chemical Theory and Computation Article

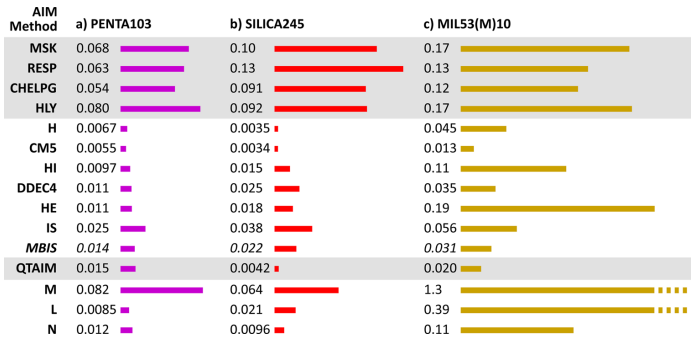



**Figure 8.** Standard deviations on atomic charges (in e), within different sets of molecules: (a) the fluctuation on the atomic charges of 103 penta-alanine conformers, (b) the fluctuation on three different types of Si charges in a set if 245 silica clusters (see text for definition of groups of Si atoms), and (c) the fluctuation on the metal charge in the MIL53(M)10 clusters, due to changes in the basis set.

of the three classes and for each AIM method. The label "MBIS-S" refers to the interaction energy computed using effective core charges and valence Slater density functions, as shown schematically in Figure 1b.

A surprising result in Figure 7 is that electrostatic interactions computed with ESP-fitted charges are not the most accurate. MBIS point charges perform better in all three classes ("weak", "medium", and "strong") than the best ESP-fitting method (HLY). This can be understood as follows: as explained in subsection 4.3, ESP-fitted charges are biased to reproduce effects of atomic multipoles on the ESP. Although this may improve the accuracy with which the ESP is reproduced, it is a form of overfitting that may deteriorate other results obtained with ESP-fitted charges, as we observe here.

The second important result is that the MBIS method is a very effective model for the penetration effect. The MBIS-S results in Figure 7 are obtained by describing every atom with an effective core charge and a valence Slater function. Even though this is a very simple (and thus computationally efficient) approach, it already reduces the RMSE by more than 50% in the "weak" and "medium" classes. Only for "strong" electrostatic interactions, there are no apparent benefits from using such Slater density functions. A more detailed analysis, in which we computed electrostatic interactions with multipole expansions of MBIS AIM densities, showed that the greatest error in the "strong" electrostatic interactions is due to the neglect of atomic dipole moments.

**4.5. Robustness of the Atomic Charges.** For the development of an electrostatic force-field model or for the chemical interpretation of atomic charges, it is desirable that the charges are robust, i.e., not too sensitive to small details in the electronic structure calculations from which they are derived. Robustness is a prerequisite for transferability, i.e., the assumption that parameters derived from a molecule remain valid when that molecule is embedded (non)covalently in a molecular environment. Even for environment-specific force-field parameters,[58,119] a robust partitioning is of interest to ensure that such parameters remain valid as far as possible from the reference point for which they were computed. In this subsection, three types of sensitivity (the inverse of robustness) of atomic charges are investigated: sensitivity to conformational changes, sensitivity to chemical changes in the environment, and sensitivity to changes in the basis set.

Figure 8a shows the sensitivity of the atomic charges to conformational changes of the penta-alanine chain, for all AIM methods. For a given AIM method, the standard deviation of the atomic charges in the PENTA103 set are computed with respect to the average charge of each atom over all 103 conformations. These 103 conformers are randomly generated metastable structures.[49] Although some fluctuation of the charges may be expected due to internal polarization, some methodological artifacts will cause larger fluctuations without physical origin.

Figure 8b compares an averaged standard deviation of the Si charges in the SILICA245 set. The Si atoms are divided into three groups, based on the number of terminating hydrogen atoms they are bonded to (ranging from 0 to 3). Within each group, the standard deviation on the Si charge is computed and the average over the four groups is shown in Figure 8b. This standard deviation should be small, because the Si atoms within one group have a very comparable chemical environment.

Finally, Figure 8c shows the sensitivity of the metal atom charge in the MIL53(M)10 clusters to the basis set. Their electron densities were computed with 6-311+G(2dp,f), 6-311+G*, 6-31+G*, 6-31+G, or 6-31G*. The standard deviation is computed relative to the average charge of each transition metal over all basis sets. Note that the sensitivity values for Mulliken and Löwdin fall literally off the chart and the corresponding bars in the bar plot were truncated for the sake of clarity.

The ESP-fitted and Mulliken charges have a very high sensitivity in all three cases, which is consistent with earlier work.[49] This is problematic, because it is almost impossible to provide definitive charges with such methods. The basis set sensitivity seems to be the most difficult to control: the standard deviation is larger than 0.1 e for the MSK, RESP, CHELPG, HLY, HI, and HE methods, as well as the M, L, and N methods.

The basis set robustness of the H, HI, and HE methods can be improved as follows. Currently, we have used consistent levels of theory for pro-atom and molecular electron densities. If the pro-atoms were computed with a single level of theory and basis set, independent of the settings of the molecular calculation, the robustness would significantly improve. This is noticeable in the low sensitivity of the DDEC4 and especially



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

the CM5 charges. Both CM5 and DDEC are implemented with a unique set of pro-atoms.

The MBIS charges are more robust than the IS charges. This is simply because MBIS pro-atoms have fewer degrees of freedom than IS pro-atoms, which is consistent with previous observations.[54]

The restraints in the RESP method only have a marginal impact on the robustness, compared to MSK (the same method without restraints), showing that the restraints do not meet their purpose while they may cause a very poor fit to the ESP (see Figure 5). The HLY method, a rather recent ESP-fitting method with a more carefully constructed cost function, does not guarantee robust results either.

QTAIM charges are among the most robust in our test, which is consistent with previous studies assessing the transferability of QTAIM results.[34,120] Still, several Hirshfeld variants, such as CM5, MBIS, and DDEC4, are comparably robust.

**4.6. Pareto Analysis.** In our comparative analysis, we have considered three main criteria that atomic charges should meet for the development of force fields: accuracy of the ESP, accuracy of electrostatic interactions, and robustness. Ideally, an AIM method should combine all these features, especially the last two. The Pareto plots in Figure 9 visualize the tradeoffs between different criteria discussed in the previous subsections. It is unavoidable that some subjective choices slightly affect the Pareto analysis, such as the selected molecules in the datasets, their classification into groups, the ESP cost function, etc. Nevertheless, some clear trends can be observed.

Figure 9a compares the average $RMSE_{ESP}$ over all groups in Figure 5 (*Y*-axis) to the average RMSE of the "weak" and "medium" electrostatic interactions from Figure 7 (*X*-axis). The "strong" electrostatic interactions are not included, because we found that these can never be reproduced reliably with any model using just atomic monopoles. The data point for QTAIM was omitted, because of its excessively large average $RMSE_{ESP}$. The Pareto front only considers genuine point-charge models. Obviously, the MBIS-S method performs far better for electrostatic interactions, because it goes beyond the simple point-charge model. This figure mainly shows that an accurate electrostatic potential does not guarantee accurate electrostatic interactions, and vice versa. If both qualities are of interest, the N, MBIS, IS, and HLY methods are Pareto optimal. Obviously, HLY is Pareto optimal, because its cost function was used to compute $\langle RMSE_{ESP} \rangle$, which is the reason for its advantage over MSK and CHELPG.

Figure 9b uses the same *X*-axis as Figure 9a but has the average of the three sensitivity values from Figure 8 on the *Y*-axis. The data point for the Mulliken method was omitted because of its excessively large average sensitivity value. The Pareto-optimal point-charge models are N, MBIS, and CM5. Again, when going beyond point charges, MBIS-S has very attractive performance. The poor performance of ESP-fitted charges in Figure 9b is striking. The RESP method has been the method of choice in the development of many force-field models, most notably in the AMBER community. Our results indicate that RESP and other ESP fitting methods are relatively poor methods for modeling electrostatic interactions in force fields.

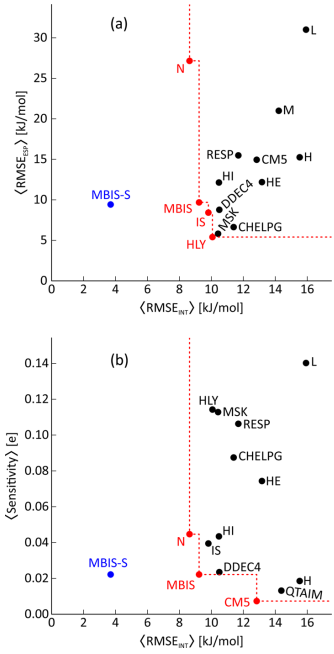



**Figure 9.** Pareto plots showing the tradeoffs between different desirable properties of AIM methods: (a) the quality of the ESP versus the accuracy of electrostatic interactions and (b) the sensitivity of the atomic charges versus the accuracy of the electrostatic interactions. Red data points are on the Pareto front for all models using just point charges. The MBIS-S method goes beyond point charges by introducing a model for the distributed valence electron density.

## 5. APPLICATION TO DENSITY-DEPENDENT DISPERSION MODELS

Several dispersion models, typically used to correct DFT calculations, apply the Hirshfeld partitioning method to estimate AIM polarizabilities. The polarizability of atom A in a molecule ($\alpha_A$) is obtained by rescaling the experimental value of the free neutral atom using the third radial moment of AIM density, $\rho_A(\mathbf{r})$:

$$\alpha_A = \alpha_{A,free} \frac{\langle r^3 \rangle_A}{\langle r^3 \rangle_{A,free}} \tag{23}$$

where

$$\langle r^3 \rangle_A = \int d\mathbf{r} \, |\mathbf{r} - \mathbf{R}_A|^3 \rho_A(\mathbf{r}) \tag{24}$$

and similarly for $\langle r^3 \rangle_{A,free}$. These rescaled polarizabilities are used in various methods[41−43] to obtain environment-specific atomic $C_6$ coefficients. The original Hirshfeld method is most often used in this context. The HI method is sometimes used instead[121] and found to improve dispersion-corrected DFT calculations for ionic systems.[82,83] Density-based dispersion models are not only used for correcting DFT calculations but were recently also employed in force-field development.[58] In



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

this section, we will directly compare the accuracy of molecular $C_6$ coefficients when the Tkatchenko−Scheffler method is used, in combination with different Hirshfeld variants.[42] We expect that similar results can be obtained with the exchange−hole dipole model (XDM)[41] and related approaches such as dDsC.[43]

The expectation value $\langle r^3\rangle_{A,free}$ can be computed in two different ways. One may use the symmetry-broken ground-state density of the free atom or one may constrain the atom to be spherically symmetric and closed-shell. (In both cases, the same level of theory as that for the molecule is used.) The second choice is the most common in the context of dispersion models. For some elements, however, this results in higher-energy states with fractionally occupied orbitals. For the sake of consistency, compatible choices are made when computing the reference atoms for the (Iterative) Hirshfeld method. In this section, we consider in total six variants of the TS dispersion model, using three different partitioning methods: H, HI, and MBIS. For each partitioning method, ground-state (GS) reference atoms or spherical closed-shell (SCS) atoms are used. The molecular $C_6$ coefficients are also computed with Grimme's D3 model.[93]

The dispersion models are tested with the database of $C_6$ coefficients for dimers of neutral molecules by Tkatchenko and Scheffler,[42] which are derived from experimental dipole oscillator strengths. B3LYP/6-311+G(2df,p) densities[86,122] are computed for all molecules in this database and the atoms that they contain. Gaussian09[62] is used for the B3LYP calculations, except for the SCS atoms, for which a new SCF program was written. Dimers with the Xe atom are omitted because their all-electron density can only be computed properly with relativistic corrections. As explained in subsection 2.5, the density cannot be written out by Gaussian09 when relativistic corrections are used.

Figure 10 shows the scatter plots of the model $C_6$ values versus the experimental reference data, including the mean percentage errors (MPE) and the root-mean-square percentage errors (RMSPE). The original TS model corresponds to TS-H-SCS, which is one of the better variants. When using the Iterative Hirshfeld method instead (TS-HI-SCS), the RMSPE increases, indicating that the model becomes less accurate. In both H and HI variants, the use of ground-state atoms, i.e., TS-H-GS or TS-HI-GS, leads to a systematic overestimation of the reference $C_6$ coefficients. The situation is reversed when using MBIS partitioning, i.e., the dispersion model is most accurate when using ground-state reference atoms (TS-MBIS-GS), while the use of spherical reference atoms (TS-MBIS-SCS) is clearly inferior. Finally, it is worth noting that Grimme's D3 model performs slightly better than any TS variant, which is impressive, given that it only makes use of the nuclear coordinates and not the electron density.

We will now analyze why the TS model only works well for certain methodological combinations. First, the spherical closed-shell atoms have slightly larger (or equal) $\langle r^3\rangle_{A,free}$ values compared to ground state atoms, as shown in Table 2. It turns out that, for completely different reasons, the Hirshfeld method exhibits some artifacts in the partitioning that also result in increased values of $\langle r^3\rangle_A$, as explained below. For most molecules, these two effects balance out, except for the $H_2$ molecule, which causes some outliers in Figure 10 for TS-H-SCS and TS-HI-SCS at low $C_6$ values.

The artifact of the H method is very clear in Figure 11: it shows the AIM density ($\rho_A$) of the hydrogen atom in hydrogen fluoride, computed with the H and MBIS methods. The

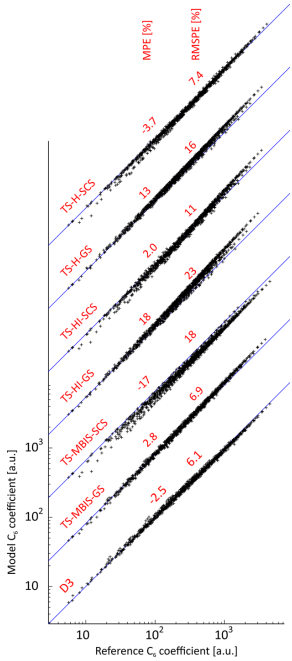



**Figure 10.** Scatter plots of model molecular $C_6$ coefficients versus experimental reference values. Except for the D3 results, all data points are shifted upward, to avoid overlap between results from different models. The mean percentage error (MPE) and the root-mean-square percentage error (RMSPE) between model and reference molecular $C_6$ coefficients are printed just above the corresponding data points.

**Table 2. Value of $\langle r^3\rangle_{A,free}$ (in Units of $a_0^3$) for the Elements Present in the Molecules in the Tkatchenko−Scheffler Set,[42] Computed in Two Ways: Using Ground-State (GS) Electron Densities and Using Densities of Atoms That Are Constrained to Be Spherical and Closed-Shell (SCS)**

| element | $\langle r^3\rangle_{A,free,GS}$ $[a_0^3]$ | $\langle r^3\rangle_{A,free,SCS}$ $[a_0^3]$ | $\frac{\langle r^3\rangle_{A,free,SCS}}{\langle r^3\rangle_{A,free,GS}}$ |
|---|---|---|---|
| H | 7.9 | 9.7 | 1.23 |
| Li | 89.0 | 105.6 | 1.19 |
| C | 35.7 | 41.4 | 1.16 |
| N | 27.0 | 30.9 | 1.14 |
| O | 22.7 | 23.8 | 1.05 |
| F | 18.6 | 18.9 | 1.01 |
| Ne | 15.4 | 15.4 | 1.00 |
| Si | 103.0 | 114.3 | 1.11 |
| S | 77.0 | 79.7 | 1.04 |
| Cl | 66.7 | 67.4 | 1.01 |
| Ar | 57.2 | 57.2 | 1.00 |
| Br | 97.0 | 97.8 | 1.01 |
| Kr | 89.1 | 89.1 | 1.00 |

Hirshfeld AIM density (solid blue line) is asymmetric, with more electron density toward the fluoride. This can be understood as follows. Any variant of the H method exhibits



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

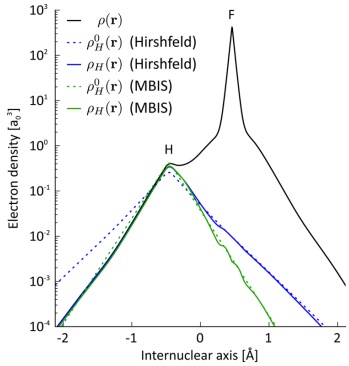



**Figure 11.** Total electron density (black) of hydrogen fluoride along the internuclear axis, including the AIM density (solid) and the pro-atomic density (dashed) of hydrogen for the H method (blue) and the MBIS method (green).

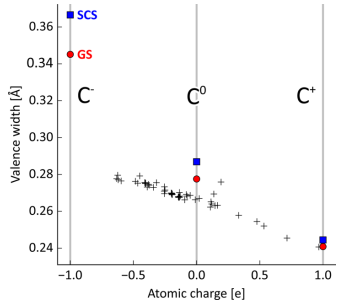



**Figure 12.** MBIS *valence width* ($\sigma_{A,v}$, see subsection 2.4) of carbon atoms in the Tkatchenko–Scheffler set of molecules[42] versus their MBIS charge (black plus signs, +). The valence widths of isolated carbon atoms and ions are also included: ground-state atoms (red circles, ●) and spherical closed-shell atoms (blue squares, ■).

the same similarity principle,[60] because of eq 5: the Hirshfeld AIM density will be as close as possible to that of the pro-atom. Because the density tail of isolated hydrogen (dashed blue line) is so different from that of hydrogen in HF, the Hirshfeld AIM density is aspherical with too much density in the bonding region. This accumulation of density, relatively far away from the hydrogen nucleus, leads to larger values of high radial moments, such as $\langle r^3\rangle_A$. This is a general feature of the H method, also seen in other molecules. In the MBIS method, no such asymmetries are found because the parameters $\sigma_{A,i}$ in eq 7, i.e., the widths of the Slater functions, are also optimized to match the molecular electron density. By consequence, the MBIS AIM density (solid green curve) almost coincides with the corresponding pro-atom (dashed green curve) and is therefore close to symmetric.

Figure 11 also shows a side effect of the MBIS method: where the MBIS AIM density (solid green curve) passes through the nucleus of fluoride, some ripples can be seen, because some details in the fluoride core electron density cannot be reproduced by the Slater functions. These ripples are very local and, therefore, have a negligible effect on $\langle r^3\rangle_A$.

Figure 12 further illustrates the mismatch between the density tails of the HI pro-atoms and the molecular densities in the TS set. It shows the MBIS *valence width* of the carbon atoms of all molecules in the TS set[42] versus their MBIS charge. This figure also includes the data points for free carbon atoms and ions, which are used as pro-atoms in the HI method. The trends in this figure are general: neutral atoms and anions have density tails that decay slower (i.e., higher valence width) than molecular electron densities. This leads to asymmetric HI AIM densities, such as the one observed in Figure 11. Especially when computing higher radial or multipole moments with the HI method, this may result in undesirable artifacts. In previous work, the erroneous density tails of unstable anions were corrected, e.g., by computing these atoms in a Watson sphere.[123] For example, such corrections were used by Bučko et al.[82,83] in their tests of the TS-HI variant. The results in Figure 12 suggest that it is also advantageous to reduce the density tails of stable anions and neutral atoms.

Finally, note that a few other Hirshfeld variants were proposed, in which the density tails of the pro-atoms are optimized to match the molecular electron density, most notably the IS method[37,53] and some of its variants including the Gaussian ISA[54] and BS-ISA+DF.[30]

## 6. CONCLUSIONS AND OUTLOOK

The MBIS method is a new density-based AIM method that is particularly suitable for the development of efficient and relatively accurate electrostatic force-field models. MBIS belongs to the family of Hirshfeld methods. Its pro-density is expanded in a minimal set of atom-centered s-type Slater density functions, whose parameters are fitted to a given molecular electron density by minimizing the Kullback–Leibler divergence. In that sense, it can also be interpreted as an information theory density-fitting method, where the Slater functions as such are used in applications, rather than the atoms-in-molecules densities.

The MBIS method is extensively tested for the development of electrostatic force-field models. When it is just used for the purpose of deriving atomic charges, it is one of the best methods available to date, in terms of robustness and accuracy of the electrostatic interactions. When the MBIS Slater functions are used to describe the valence electron density in a force field, the error on the electrostatic interactions can be reduced by 50%, if the electrostatic interaction is not too strong. This is a computationally efficient approach to describe the so-called penetration effect, i.e., the deviation of interatomic electrostatic interactions from that of point charges, when the atomic densities begin to overlap. MBIS is also useful beyond the scope of frozen-density electrostatics, e.g., when modeling dispersion interactions, or to analyze density fluctuations that a polarizable force field should reproduce.

In future work, we will focus on improving our method, its implementation, and more applications in different areas. The obvious methodological improvement is a better model for the pro-molecular density, e.g., by including atomic multipoles[30,32] or by making it compatible with pseudodensities. An improved pro-molecule model should not merely result in a better fit to a given electron density; one should also avoid too many degrees of freedom for the sake of robustness. Moreover, the use of an



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894–3912

improved pro-density model in force fields should remain computationally efficient.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00456.

Cartesian coordinates of all molecular dimers in the ZG237 set (ZIP)

Atom types in MIL-53(Al), chabazite, and ozone. Description of the ZG237 set of molecular dimers and Cartesian coordinates of the MP2/cc-pVTZ+CP optimized dimer geometries in this set (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: toon.verstraelen@ugent.be.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors would like to thank F. De Proft and P. Bultinck for inspiring discussions and useful suggestions. T.V., S.V., and L.V. acknowledge the Foundation of Scientific Research—Flanders (FWO). T.V., L.V., M.W., and V.V.S. thank the Research Board of Ghent University (BOF), and BELSPO in the frame of IAP/7/05 for their financial support. V.V.S. acknowledges the European Research Council for funding the European Union's Horizon 2020 research and innovation programme [consolidator ERC Grant Agreement No. 647755—DYNPOR (2015—2020)]. The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure). F.H.-Z. was supported by a Vanier-CGS fellowship and a Michael Smith Foreign Study Supplement from the National Sciences and Engineering Research Council of Canada (NSERC); P.W.A. was supported by a Discovery Grant and a E.W.R. Steacie Fellowship form NSERC.

## ■ REFERENCES

(1) Kohlhoff, K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* **2014**, *6*, 15—21.

(2) Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L.-S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y.-H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Schafer, U. B.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. *SC14 Proc.* **2014**, 41—53.

(3) Kim, J.; Abouelnasr, M.; Lin, L.-C.; Smit, B. Large-Scale Screening of Zeolite Structures for $CO_2$ Membrane Separations. *J. Am. Chem. Soc.* **2013**, *135*, 7545—7552.

(4) Macchiarulo, A.; Costantino, G.; Sbaglia, R.; Aiello, S.; Meniconi, M.; Pellicciari, R. The role of electrostatic interaction in the molecular recognition of selective agonists to metabotropic glutamate receptors. *Proteins: Struct., Funct., Genet.* **2003**, *50*, 609—619.

(5) Ramsahye, N. A.; Maurin, G.; Bourrelly, S.; Llewellyn, P.; Loiseau, T.; Ferey, G. Charge distribution in metal organic framework materials: transferability to a preliminary molecular simulation study of the $CO_2$ adsorption in the MIL-53 (Al) system. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1059—1063.

(6) Vanduyfhuys, L.; Verstraelen, T.; Vandichel, M.; Waroquier, M.; Van Speybroeck, V. *Ab initio* parametrized force field for the flexible metal-organic framework MIL-53(Al). *J. Chem. Theory Comput.* **2012**, *8*, 3217—3231.

(7) Kadantsev, E. S.; Boyd, P. G.; Daff, T. D.; Woo, T. K. Fast and Accurate Electrostatics in Metal Organic Frameworks with a Robust Charge Equilibration Parameterization for High-Throughput Virtual Screening of Gas Adsorption. *J. Phys. Chem. Lett.* **2013**, *4*, 3056—3061.

(8) Gabrieli, A.; Sant, M.; Demontis, P.; Suffritti, G. B. Partial Charges in Periodic Systems: Improving Electrostatic Potential (ESP) Fitting via Total Dipole Fluctuations and Multiframe Approaches. *J. Chem. Theory Comput.* **2015**, *11*, 3829—3843.

(9) Hamad, S.; Balestra, S. R. G.; Bueno-Perez, R.; Calero, S.; Ruiz-Salvador, A. R. Atomic charges for modeling metal-organic frameworks: Why and how. *J. Solid State Chem.* **2015**, *223*, 144—151.

(10) Wu, Q.; Ayers, P. W.; Zhang, Y. Density-based energy decomposition analysis for intermolecular interactions with variationally determined intermediate state energies. *J. Chem. Phys.* **2009**, *131*, 164112.

(11) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for *ab initio* calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050—8053.

(12) Tafipolsky, M.; Engels, B. Accurate Intermolecular Potentials with Physically Grounded Electrostatics. *J. Chem. Theory Comput.* **2011**, *7*, 1791—1803.

(13) Warshel, A.; Kato, M.; Pisliakov, A. V. Polarizable Force Fields: History, Test Cases, and Prospects. *J. Chem. Theory Comput.* **2007**, *3*, 2034—2045.

(14) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. ACKS2: Atom-condensed Kohn—Sham DFT approximated to second order. *J. Chem. Phys.* **2013**, *138*, 074108.

(15) Verstraelen, T.; Vandenbrande, S.; Ayers, P. Direct computation of parameters for accurate polarizable force fields. *J. Chem. Phys.* **2014**, *141*, 194114.

(16) Ji, C.; Mei, Y. Some Practical Approaches to Treating Electrostatic Polarization of Proteins. *Acc. Chem. Res.* **2014**, *47*, 2795—2803.

(17) Potoff, J. J.; Siepmann, J. I. Vapor—liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE J.* **2001**, *47*, 1676—1682.

(18) Rai, N.; Siepmann, J. I. Transferable potentials for phase equilibria. 10. Explicit-hydrogen description of substituted benzenes and polycyclic aromatic compounds. *J. Phys. Chem. B* **2013**, *117*, 273—288.

(19) Fox, T.; Kollman, P. A. Application of the RESP Methodology in the Parametrization of Organic Solvents. *J. Phys. Chem. B* **1998**, *102*, 8070—8079.

(20) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269—10280.

(21) Stone, A. J.; Alderton, M. Distributed multipole analysis. *Mol. Phys.* **1985**, *56*, 1047—1064.

(22) Kairys, V.; Jensen, J. H. Evaluation of the charge penetration energy between non-orthogonal molecular orbitals using the Spherical Gaussian Overlap approximation. *Chem. Phys. Lett.* **1999**, *315*, 140—144.

(23) Krapp, A.; Bickelhaupt, F. M.; Frenking, G. Orbital Overlap and Chemical Bonding. *Chem.—Eur. J.* **2006**, *12*, 9196—9216.

(24) Spackman, M. The use of the promolecular charge density to approximate the penetration contribution to intermolecular electrostatic energies. *Chem. Phys. Lett.* **2006**, *418*, 158—162.

(25) Lu, Z.; Zhou, N.; Wu, Q.; Zhang, Y. Directional Dependence of Hydrogen Bonds: A Density-Based Energy Decomposition Analysis and Its Implications on Force Field Development. *J. Chem. Theory Comput.* **2011**, *7*, 4038—4049.



DOI: 10.1021/acs.jctc.6b00456
*J. Chem. Theory Comput.* 2016, 12, 3894—3912

(26) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609−2618.
(27) Bader, R. F. W. A quantum theory of molecular structure and its applications. *Chem. Rev.* **1991**, *91*, 893−928.
(28) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. Towards a force field based on density fitting. *J. Chem. Phys.* **2006**, *124*, 104101.
(29) 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.
(30) 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.
(31) Wang, B.; Truhlar, D. G. Screened Electrostatic Interactions in Molecular Mechanics. *J. Chem. Theory Comput.* **2014**, *10*, 4480−4487.
(32) Öhrn, A.; Hermida-Ramon, J. M.; Karlström, G. Method for Slater-Type Density Fitting for Intermolecular Electrostatic Interactions with Charge Overlap. I. The Model. *J. Chem. Theory Comput.* **2016**, *12*, 2298.
(33) Leslie, M. DL_MULTI-A molecular dynamics program to use distributed multipole electrostatic models to simulate the dynamics of organic crystals. *Mol. Phys.* **2008**, *106*, 1567−1578.
(34) Popelier, P. L. A. QCTFF: On the construction of a novel protein force field. *Int. J. Quantum Chem.* **2015**, *115*, 1005−1011.
(35) Ghillemijn, D.; Bultinck, P.; Van Neck, D.; Ayers, P. W. A self-consistent Hirshfeld method for the atom in the molecule based on minimization of information loss. *J. Comput. Chem.* **2011**, *32*, 1561−1567.
(36) Heidar-Zadeh, F.; Ayers, P. W. How pervasive is the Hirshfeld partitioning? *J. Chem. Phys.* **2015**, *142*, 044107.
(37) Lillestolen, T. C.; Wheatley, R. J. Redefining the atom: atomic charge densities produced by an iterative stockholder approach. *Chem. Commun.* **2008**, 5909−5911.
(38) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chem. Acc.* **1977**, *44*, 129−138.
(39) Nalewajski, R. F.; Parr, R. G. Information theory, atoms in molecules, and molecular similarity. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8879−8882.
(40) 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.
(41) Becke, A. D.; Johnson, E. R. Exchange-hole dipole moment and the dispersion interaction revisited. *J. Chem. Phys.* **2007**, *127*, 154108.
(42) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.
(43) Steinmann, S. N.; Corminboeuf, C. Comprehensive Benchmarking of a Density-Dependent Dispersion Correction. *J. Chem. Theory Comput.* **2011**, *7*, 3567−3577.
(44) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103*, 1793−1874.
(45) Davidson, E. R.; Chakravorty, S. A test of the Hirshfeld definition of atomic charges and moments. *Theor. Chem. Acc.* **1992**, *83*, 319−330.
(46) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007**, *126*, 144111.
(47) Van Damme, S.; Bultinck, P.; Fias, S. Electrostatic Potentials from Self-Consistent Hirshfeld Atomic Charges. *J. Chem. Theory Comput.* **2009**, *5*, 334−340.
(48) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. The electronegativity equalization method and the split charge equilibration applied to organic systems: Parametrization, validation, and comparison. *J. Chem. Phys.* **2009**, *131*, 044127.
(49) Verstraelen, T.; Pauwels, E.; De Proft, F.; Van Speybroeck, V.; Geerlings, P.; Waroquier, M. Assessment of Atomic Charge Models for Gas-Phase Computations on Polypeptides. *J. Chem. Theory Comput.* **2012**, *8*, 661−676.
(50) Bultinck, P.; Ayers, P.; Fias, S.; Tiels, K.; Van Alsenoy, C. Uniqueness and basis set dependence of iterative Hirshfeld charges. *Chem. Phys. Lett.* **2007**, *444*, 205−208.
(51) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. Hirshfeld-E partitioning: AIM charges with an improved trade-off between robustness and accurate electrostatics. *J. Chem. Theory Comput.* **2013**, *9*, 2221−2225.
(52) Verstraelen, T.; Sukhomlinov, S. V.; Van Speybroeck, V.; Waroquier, M.; Smirnov, K. S. Computation of charge distribution and electrostatic potential in silicates with the use of chemical potential equalization models. *J. Phys. Chem. C* **2012**, *116*, 490−504.
(53) Lillestolen, T. C.; Wheatley, R. J. Atomic charge densities generated using an iterative stockholder procedure. *J. Chem. Phys.* **2009**, *131*, 144101.
(54) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. The conformational sensitivity of iterative stockholder partitioning schemes. *Chem. Phys. Lett.* **2012**, *545*, 138−143.
(55) Manz, T. A.; Sholl, D. S. Chemically Meaningful Atomic Charges That Reproduce the Electrostatic Potential in Periodic and Nonperiodic Materials. *J. Chem. Theory Comput.* **2010**, *6*, 2455−2468.
(56) Manz, T. A.; Sholl, D. S. Improved Atoms-in-Molecule Charge Partitioning Functional for Simultaneously Reproducing the Electrostatic Potential and Chemical States in Periodic and Nonperiodic Materials. *J. Chem. Theory Comput.* **2012**, *8*, 2844−2867.
(57) Vanpoucke, D. E. P.; Bultinck, P.; Van Driessche, I. Extending Hirshfeld-I to bulk and periodic materials. *J. Comput. Chem.* **2013**, *34*, 405−417.
(58) Cole, D. J.; Vilseck, J. Z.; Tirado-Rives, J.; Payne, M. C.; Jorgensen, W. L. Biomolecular Force Field Parameterization via Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2016**, *12*, 2312.
(59) Haldoupis, E.; Nair, S.; Sholl, D. S. Finding MOFs for Highly Selective $CO_2/N_2$ Adsorption Using Materials Screening Based on Efficient Assignment of Atomic Point Charges. *J. Am. Chem. Soc.* **2012**, *134*, 4313−4323.
(60) Ayers, P. Atoms in molecules, an axiomatic approach. I. Maximum transferability. *J. Chem. Phys.* **2000**, *113*, 10886−10898.
(61) Parr, R. G.; Ayers, P. W.; Nalewajski, R. F. What Is an Atom in a Molecule? *J. Phys. Chem. A* **2005**, *109*, 3957−3959.
(62) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision D.01*; Gaussian, Inc.: Wallingford, CT, 2013.
(63) Verstraelen, T.; Boguslawski, K.; Tecmer, P.; Heidar-Zadeh, F.; Chan, M.; Kim, T. D.; Zhao, Y.; Vandenbrande, S.; Yang, D.; González-Espinoza, C. E.; Limacher, P. A.; Berrocal, D.; Malek, A.; Ayers, P. W. *HORTON 2.0.0*; 2015; Available via the Internet at: http://theochem.github.com/horton (accessed June 14, 2016).
(64) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *J. Chem. Phys.* **1988**, *88*, 2547−2553.
(65) Blöchl, P. E.; Först, C. J.; Schimpl, J. Projector augmented wave method: *Ab initio* molecular dynamics with full wave functions. *Bull. Mater. Sci.* **2003**, *26*, 33−41.



DOI: 10.1021/acs.jctc.6b00456
J. Chem. Theory Comput. 2016, 12, 3894−3912

(66) Mortensen, J. J.; Hansen, L. B.; Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *71*, 035109.
(67) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dułak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. a.; Kristoffersen, H. H.; Kuisma, M.; Larsen, a. H.; Lehtovaara, L.; Ljungberg, M.; Lopez-Acevedo, O.; Moses, P. G.; Ojanen, J.; Olsen, T.; Petzold, V.; Romero, N. a.; Stausholm-Møller, J.; Strange, M.; Tritsaris, G. a.; Vanin, M.; Walter, M.; Hammer, B.; Häkkinen, H.; Madsen, G. K. H.; Nieminen, R. M.; Nørskov, J. K.; Puska, M.; Rantala, T. T.; Schiøtz, J.; Thygesen, K. S.; Jacobsen, K. W. Electronic structure calculations with GPAW: A real-space implementation of the projector augmented-wave method. *J. Phys.: Condens. Matter* **2010**, *22*, 253202.
(68) Bahn, S. R.; Jacobsen, K. W. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.* **2002**, *4*, 56−66.
(69) Lee, L. P.; Cole, D. J.; Skylaris, C.-K.; Jorgensen, W. L.; Payne, M. C. Polarized Protein-Specific Charges from Atoms-in-Molecule Electron Density Partitioning. *J. Chem. Theory Comput.* **2013**, *9*, 2981−2991.
(70) Hu, H.; Lu, Z.; Yang, W. Fitting Molecular Electrostatic Potentials from Quantum Mechanical Calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1004−1013.
(71) Gill, P. M. W. Extraction of Stewart atoms from electron densities. *J. Phys. Chem.* **1996**, *100*, 15421−15427.
(72) Slater, J. C. Atomic shielding constants. *Phys. Rev.* **1930**, *36*, 57−64.
(73) Wang, W.-P.; Parr, R. G. Statistical atomic models with piecewise exponentially decaying electron densities. *Phys. Rev. A: At., Mol., Opt. Phys.* **1977**, *16*, 891−902.
(74) Wang, W.-P. Fixed-shell statistical atomic models with piecewise exponentially decaying electron densities. *Phys. Rev. A: At., Mol., Opt. Phys.* **1982**, *25*, 2901−2912.
(75) Fernandez Pacios, L. A simplified representation of atomic electron densities and electrostatic potentials. *J. Phys. Chem.* **1991**, *95*, 10653−10658.
(76) 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.
(77) Bultinck, P.; Cooper, D. L.; Van Neck, D. Comparison of the Hirshfeld-I and iterated stockholder atoms in molecules schemes. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3424−3429.
(78) Baerends, E.; Ellis, D.; Ros, P. Self-consistent molecular Hartree−Fock−Slater calculations. I. The computational procedure. *Chem. Phys.* **1973**, *2*, 41−51.
(79) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On first-row diatomic molecules and local density models. *J. Chem. Phys.* **1979**, *71*, 4993.
(80) Fonseca Guerra, C.; Snijders, J. G.; Te Velde, G.; Baerends, E. J. Towards an order-N DFT method. *Theor. Chem. Acc.* **1998**, *99*, 391−403.
(81) Stewart, R. F. V. One-Electron Density Functions and Many-Centered Finite Multipole Expansions. *Isr. J. Chem.* **1977**, *16*, 124−131.
(82) Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J. G. Improved Density Dependent Correction for the Description of London Dispersion Forces. *J. Chem. Theory Comput.* **2013**, *9*, 4293−4299.
(83) Bučko, T.; Lebègue, S.; Ángyán, J. G.; Hafner, J. Extending the applicability of the Tkatchenko−Scheffler dispersion correction via iterative Hirshfeld partitioning. *J. Chem. Phys.* **2014**, *141*, 034114.
(84) Gould, T.; Bucko, T. $C_6$ Coefficients and Dipole Polarizabilities for All Atoms and Many Ions in Rows 1−6 of the Periodic Table. *J. Chem. Theory Comput.* **2016**, in press (DOI: 10.1021/acs.jctc.6b00361).
(85) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(86) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724−728.
(87) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark structures and binding energies of small water clusters with anharmonicity corrections. *J. Phys. Chem. A* **2011**, *115*, 12034−12046.
(88) Hayward, J. S.; Reimers, J. R. Unit cells for the simulation of hexagonal ice. *J. Chem. Phys.* **1997**, *106*, 1518.
(89) Lazzaro, A.; Gloeß, A.; Seitsonen, A. P.; Mundy, C.; Pousa, C.; Golze, D.; Mohamed, F.; Schiffmann, F.; Tabacchi, G.; Forbert, H.; Bani-Hashemian, H.; Bethune, I.; Kuo, I.-F. W.; Wilhelm, J.; VandeVondele, J.; Hutter, J.; Tong, L.; Walewski, L.; Schoenherr, M.; Guidon, M.; Mauri-Iannuzzi, M.; McGrath, M.; Krack, M.; Ben, M. D.; Ceriotti, M.; Schuett, O.; Seewald, P.; Andermatt, S.; Laino, T.; Chassaing, T.; Kuehne, T.; Borstnik, U.; Weber, V.; Rybkin, V. *CP2K 2.6.0*; 2014; Available via the Internet at: http://www.cp2k.org (accessed June 14, 2016).
(90) VandeVondele, J.; Hutter, J. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.* **2003**, *118*, 4365.
(91) Vandevondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103−128.
(92) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: Atomistic simulations of condensed matter systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 15−25.
(93) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(94) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(95) VandeVondele, J.; Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **2007**, *127*, 114105.
(96) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703−1710.
(97) Krack, M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor. Chem. Acc.* **2005**, *114*, 145−152.
(98) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(99) Horn, H.; Swope, W.; Pitera, J.; Madura, J.; Dick, T.; Hura, G.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(100) 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.
(101) Burnham, C. J.; Anick, D. J.; Mankoo, P. K.; Reiter, G. F. The vibrational proton potential in bulk liquid water and ice. *J. Chem. Phys.* **2008**, *128*, 154519.
(102) Yu, W.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Six-site polarizable model of water based on the classical Drude oscillator. *J. Chem. Phys.* **2013**, *138*, 034508.
(103) Mei, Y.; Simmonett, A. C.; Pickard, F. C.; DiStasio, R. A.; Brooks, B. R.; Shao, Y. Numerical Study on the Partitioning of the Molecular Polarizability into Fluctuating Charge and Induced Atomic Dipole Contributions. *J. Phys. Chem. A* **2015**, *119*, 5865−5882.
(104) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.
(105) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141−151.



DOI: 10.1021/acs.jctc.6b00456
*J. Chem. Theory Comput.* 2016, 12, 3894−3912

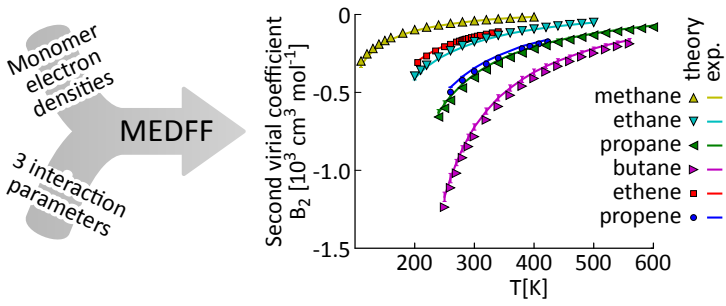
(106) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 4285−4292.
(107) Férey, G.; Latroche, M.; Serre, C.; Millange, F.; Loiseau, T.; Percheron-Guégan, A. Hydrogen adsorption in the nanoporous metal-benzenedicarboxylate $M(OH)(O_2C-C_6H_4-CO_2)$ ($M = Al^{3+}$, $Cr^{3+}$), MIL-53. *Chem. Commun.* **2003**, 2976−2977.
(108) Singh, U. C.; Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129−145.
(109) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.
(110) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.
(111) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(112) Löwdin, P.-O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365−375.
(113) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735−746.
(114) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(115) Manz, T. A.; Limas, N. G. *Chargemol program for performing DDEC analysis*, Version 09.15.2014; 2014; Available via the Internet at: http://ddec.sourceforge.net (accessed June 14, 2016).
(116) Keith, T. A. *AIMAll, Version 11.06.19*; TK Gristmill Software: Overland Park, KS, USA, 2014; Available via the Internet at: http://aim.tkgristmill.com (accessed June 14, 2016).
(117) Ibrahim, M. A. A. Molecular mechanical study of halogen bonding in drug discovery. *J. Comput. Chem.* **2011**, *32*, 2564−2574.
(118) Bader, R.; Carroll, M.; Cheeseman, J.; Chang, C. Properties of atoms in molecules: atomic volumes. *J. Am. Chem. Soc.* **1987**, *109*, 7968−7979.
(119) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497−4514.
(120) Devereux, M.; Popelier, P. L. A.; McLay, I. M. Toward an *ab initio* fragment database for bioisosterism: Dependence of QCT properties on level of theory, conformation, and chemical environment. *J. Comput. Chem.* **2009**, *30*, 1300−1318.
(121) Steinmann, S.; Corminboeuf, C. A System-Dependent Density-Based Dispersion Correction. *J. Chem. Theory Comput.* **2010**, *6*, 1990−2001.
(122) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(123) Watson, R. E. Analytic Hartree-Fock Solutions for $O^{=}$. *Phys. Rev.* **1958**, *111*, 1108−1110.



DOI: 10.1021/acs.jctc.6b00456
*J. Chem. Theory Comput.* 2016, 12, 3894−3912

Paper II

# The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Non-Covalent Interactions




S. Vandenbrande, M. Waroquier, V. Van Speybroeck and T. Verstraelen

*J. Chem. Theory Comput.*, **2017**, *13* (1), 161-179

S. Vandenbrande developed the methodology, performed the simulations, and wrote the manuscript.

JCTC Journal of Chemical Theory and Computation
Article
pubs.acs.org/JCTC

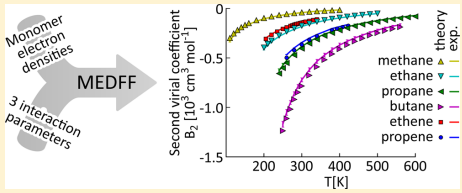
# The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions

Steven Vandenbrande, Michel Waroquier, Veronique Van Speybroeck, and Toon Verstraelen*

Center for Molecular Modeling (CMM), QCMM Ghent−Brussels Alliance, Ghent University, Technologiepark 903, B9000 Ghent, Belgium

Supporting Information

**ABSTRACT:** We propose a methodology to derive pairwise-additive noncovalent force fields from monomer electron densities without any empirical input. Energy expressions are based on the symmetry-adapted perturbation theory (SAPT) decomposition of interaction energies. This ensures a physically motivated force field featuring an electrostatic, exchange-repulsion, dispersion, and induction contribution, which contains two types of parameters. First, each contribution depends on several fixed atomic parameters, resulting from a partitioning of the monomer electron density. Second, each of the last three contributions (exchange-repulsion, dispersion, and induction) contains exactly one linear fitting parameter. These three so-called interaction parameters in the model are initially estimated separately using SAPT reference calculations for the S66x8 database of noncovalent dimers. In a second step, the three interaction parameters are further refined simultaneously to reproduce CCSD(T)/CBS interaction energies for the same database. The limited number of parameters that are fitted to dimer interaction energies (only three) avoids ill-conditioned fits that plague conventional parameter optimizations. For the exchange-repulsion and dispersion component, good results are obtained for all dimers in the S66x8 database using one single value for the associated interaction parameters. The values of those parameters can be considered *universal* and can also be used for dimers not present in the original database used for fitting. For the induction component such an approach is only viable for the dispersion-dominated dimers in the S66x8 database. For other dimers (such as hydrogen-bonded complexes), we show that our methodology remains applicable. However, the interaction parameter needs to be determined on a case-specific basis. As an external validation, the force field predicts interaction energies in good agreement with CCSD(T)/CBS values for dispersion-dominated dimers extracted from an HIV-II protease crystal structure with a bound ligand (indinavir). Furthermore, experimental second virial coefficients of small alkanes and alkenes are well reproduced.




## 1. INTRODUCTION

Classical force-field simulations are an attractive tool to study molecular systems because the phase space can be sampled at a very low computational cost (compared to ab initio methods). This computational efficiency allows one to study phenomena at the nanoscale such as the competitive adsorption of $H_2O$ and $CO_2$ in zeolite 13X,[1] the mechanical energy absorption ability of the MIL-53(Al) metal−organic framework,[2] or the composition and behavior of disordered protein states.[3] These are only a few examples of successful applications of force fields. The noncovalent interactions are most often based on experimental data.[4−8] This is of course a limitation, as it is potentially very interesting to study systems for which no experimental data is available. Additionally, in most cases it is impossible to assess a priori the accuracy of such force fields as the final results critically depend on the specific terms included in the force-field expression and on the parametrization procedure. Force fields derived from ab initio calculations offer in a way the best of two worlds. On the one hand the analytical energy expression can be evaluated efficiently, allowing simulations of large systems on a long time scale. On the other hand, these force fields succeed reasonably well in reproducing ab initio data and can be derived for any molecular system for which reliable ab initio data can be generated. In this work, we will therefore focus on describing noncovalent interactions with only first-principles calculations as input. Constructing such noncovalent force fields exclusively based on ab initio data is a long-standing problem. A *conventional* and straightforward approach is to generate relevant configurations of the system(s) of interest, determine the intermolecular interaction energy and/or forces for these configurations using an ab initio method, and then fit (many) parameters of the noncovalent force field by minimizing a cost function with ab initio results as reference data. This procedure has been applied successfully numerous times.[9−11] There are however also some drawbacks of this straightforward matching of ab initio and force-field results, such as the introduction of many parameters and the difficulty in giving a physical interpretation to each of the contributions. In this work we will try to circumvent some

Received: September 30, 2016
Published: November 17, 2016





DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

of these problems associated with deriving noncovalent force fields. Typically atom types are introduced leading to a large number of parameters that need to be fitted. Even though the number of unknowns in the cost function can be considerably reduced by using mixing rules, the fitting in a high-dimensional parameter space will often be ill-conditioned.[12] A second drawback is that the physical origin of interactions is not properly incorporated. For example, most noncovalent force fields omit the penetration effect, which has been shown to be substantial,[13] when describing electrostatics. This contribution is then erroneously included in the van der Waals energy. Indeed it has been shown that most force fields rely to a large extent on compensation of errors.[14] A third and final deficiency of the conventional approach is that scanning the PES with an ab initio method is computationally very costly. Therefore, one is usually limited to density functional theory (DFT) or second-order Møller−Plesset perturbation theory (MP2), which both have some known limitations (mostly relevant for intermolecular interactions): DFT cannot properly describe electron correlation, while the perturbation term in MP2 is rather large, leading to overestimated dispersion interactions in for example stacked aromatic−aromatic systems.[15,16] It would therefore be very interesting to base noncovalent force fields on coupled-cluster theory with single and double excitations and a perturbation correction for the triples contribution (CCSD(T)) in the complete basis set (CBS) limit, the current gold standard in quantum chemistry.[17]

The issues mentioned above have been partly addressed in earlier work, and these efforts are briefly summarized hereafter. The quantum mechanically derived force field[18] (QMDFF) is based on a limited number of universal parameters (28 in total for the noncovalent interactions) combined with atomic charges from a partitioned molecular electron density. QMDFF succeeds well in reproducing B3LYP-D3/def2-QZVP reference interaction energies for gas-phase dimers. However, effects such as charge penetration are not included and the parameters are obtained, at least partially, by manual inspection, thus rendering the physical interpretation of the different terms unclear. To correctly estimate parameters, it is desirable to match the force field term by term with an intermolecular perturbation theory that provides a decomposition of the interaction energy into physically recognizable components, such as symmetry-adapted perturbation theory[19] (SAPT). It has been shown that this approach leads to physically motivated noncovalent force fields that can predict second virial coefficients and gas adsorption in metal−organic frameworks in quantitative agreement with experiment.[20−23] Although the resulting parameters have been shown to be transferable to other chemical environments, the distribution of molecular properties over atomic contributions remains a tedious work. The study of SAPT exchange energies at small intermolecular distances showed that improved models (compared to traditional Lennard-Jones or Born−Mayer functional forms) lead to a more accurate, transferable and robust description of short-range interactions.[24] The SAPT decomposition has also been used for the construction of intermolecular potentials for small organic molecules using the CamCASP suite of programs, and this methodology is proven to be successful for the pyridine system.[25] A related approach yielded similar success for the water dimer.[26] Recent advances in the AMOEBA force field use SAPT to improve the description of electrostatics[27] and the sum of exchange-repulsion and dispersion,[28] aiming to improve the accuracy and transferability of the potential. Another method to ensure physical parameters is to use the results from a density-based energy decomposition analysis[29] (DEDA). This has been applied in a force-field context only to a few selected systems such as rare-gas dimers[30] or directional dependence of hydrogen bonds.[31] Also the SIBFA[32] (sum of interactions between fragments ab initio computed) force field should be mentioned, as it successfully applies a decomposition of the total interaction energy similar to the one used in this work.

We believe that the added value of the current work is mainly that several beneficial properties of the efforts discussed above are combined. In this way we take a step forward to tackle the three deficiencies mentioned earlier. At most three linear parameters are calibrated at a time, thus avoiding ill-conditioned fits. Furthermore, all terms are physically inspired (for instance including in a natural way the charge penetration effect using distributed charges) and a limited number of high-level ab initio calculations suffices to confidently determine all parameters. Our results show that the description of induction is the weakest point of the model and can be used with one universal value for the interaction parameter only for dispersion-dominated complexes. For hydrogen-bonded complexes, the interaction parameter needs to be determined on a case-specific basis. We thus gain insight into apparent limitations of the proposed methodology for the description of the SAPT induction for systems that feature hydrogen bonds. We note that many-body effects are very important in the induction energy for such systems and these need to be implicitly captured in our, by construction, pairwise-additive force field. Concerning the exchange and dispersion contributions, however, we show that good results are obtained for all dimers using a single universal value for the corresponding interaction parameters. Moreover our model succeeds in the explicit description of the electrostatic penetration effect using distributed charges. The combination of SAPT (to ensure physical values) and CCSD(T)/CBS (to ensure accurate values) reference data in the fitting process, through the introduction of prior knowledge about the parameters in the cost function, is a novel aspect of the presented work.

To conclude this introduction, we discuss the structure of the remainder of this paper. In section 2, we present the analytical expression for the noncovalent contributions to the energy of our force field. All terms are based on the partitioning of the ab initio monomer density, resulting in a force field containing only a very limited number of free parameters to be fitted, the so-called interaction parameters. In section 3, we will discuss how these interaction parameters are determined. In first instance this will be done by fitting to SAPT components of the interaction energy of a database of dimer configurations. In a second step the parameters are refined by fitting to CCSD(T)/CBS interaction energies for the same database of dimers. In section 4, we will perform an external validation of the constructed force field. First of all this will be done by comparing with ab initio interaction energies for another database of dimers, next by comparing with experimental second virial coefficients for small alkanes and alkenes. Finally we will also investigate whether the force field is suitable to predict properties of condensed-phase systems. Section 5 summarizes our main findings.

## 2. FORCE-FIELD COMPONENTS

The primary goal is to reproduce ab initio interaction energies of dimers with a force field. We are not concerned with



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

intramolecular interactions, as the force field proposed in this work only describes noncovalent interactions. We therefore start by presenting the analytical expression for the interaction energy *between* two molecules, for which only intermolecular interactions are present.

To construct a force field for a dimer, we require the ab initio molecular ground-state electron density of the separate, isolated molecules where the nuclear positions are taken from the monomer geometries in the dimer. In other words, the nuclear positions are not relaxed when going from the dimer configuration to the monomer configurations. As DFT provides reliable ground-state electron densities for most molecules, it will be used in this paper to compute the monomer properties that serve as input for the force field. More accurate wave function based methods, SAPT and CCSD(T), are only used in this work to determine the three interaction parameters related to intermolecular interactions. Using DFT for the monomer calculations makes it possible to apply the proposed method to relatively large molecular systems. The ab initio molecular electron densities of both isolated molecules will be represented by $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$ respectively. The minimal basis iterative stockholder (MBIS) scheme[33] is used to partition the ab initio molecular electron density of the isolated molecules into atomic contributions. This partitioning scheme allows one to approximate the molecular electron density as a sum of s-type Slater functions:

$$\rho_1(\mathbf{r}) \approx \sum_{A=1}^{N_{\text{atoms}}^1} \sum_{i=1}^{m_A} \frac{N_{Ai}}{8\pi\sigma_{Ai}^3} \exp\left(-\frac{|\mathbf{r}-\mathbf{R}_A|}{\sigma_{Ai}}\right) \tag{1}$$

where the sum over $A$ runs over all $N_{\text{atoms}}^1$ atoms in molecule 1 and $i$ runs over all $m_A$ shells of atom $A$. The position of atom $A$ is represented by $\mathbf{R}_A$, and the parameters $N_{Ai}$ and $\sigma_{Ai}$ give the population and *width* of the s-type Slater function of shell $i$ on atom $A$. The width of the s-type Slater function is the characteristic length over which the density decays appreciably. In other words the width $\sigma_{Ai}$ is inversely proportional to the rate of decay of the Slater function. The parameters $N_{Ai}$ and $\sigma_{Ai}$ are determined in such a way that the information loss (expressed as the Kullback−Leibler divergence) from the ab initio reference density to the prodensity is minimized, under constraints on the total charge and atomic populations. This is the most simple and concise model for an electron density that is physically reasonable, inspired by the piecewise exponential ansatz from statistical models for atomic densities.[34,35] The MBIS charges and widths are the only molecule-specific information used in the proposed force field and are independent of the other monomer present in a dimer. These *atomic* parameters are not fitted to dimer interaction energies, but obtained directly from the monomer electron density. This is in sharp contrast with the *interaction* parameters (introduced later in this section), which are the fitting parameters of the model.

For force-field applications, we further simplify the approximation of the molecular electron density by condensing the core Slater functions onto the nuclear charge. The force-field approximation of the valence molecular electron density $\tilde{\rho}_1(\mathbf{r})$ is then given by

$$\tilde{\rho}_1(\mathbf{r}) = \sum_{A=1}^{N_{\text{atoms}}^1} \frac{N_{A,\text{v}}}{8\pi\sigma_{A,\text{v}}^3} \exp\left(-\frac{|\mathbf{r}-\mathbf{R}_A|}{\sigma_{A,\text{v}}}\right) \tag{2}$$

where $N_{A,\text{v}}$ and $\sigma_{A,\text{v}}$ are the population and width of the valence Slater function of atom $A$. We therefore call $N_{A,\text{v}}$ and $\sigma_{A,\text{v}}$ the valence population and valence width, respectively. The effective core charge $q_{A,\text{c}}$ of atom $A$ is given by

$$q_{A,\text{c}} = Z_A - \sum_{i=1}^{m_A-1} N_{Ai} \tag{3}$$

where $Z_A$ is simply the atomic number or nuclear charge of atom $A$. To clarify this approach and to get a sense for the magnitude of populations and widths, we present the parameters of a water molecule in Table 1. The oxygen atom has two shells of electrons, and thus the MBIS partitioning leads to a model density with two Slater functions. The first Slater function contains 1.66 electrons ($N_{\text{O1}} = 1.66$) and decays exponentially with a characteristic length of $\sigma_{\text{O1}} = 0.03$ Å. The second shell, containing valence electrons, contains 7.20 electrons ($N_{\text{O2}} = 7.20$) and decays exponentially with a characteristic length of $\sigma_{\text{O2}} = 0.22$ Å. As expected, the first Slater function decays a lot faster than the second one. This means that the first Slater function can be approximated accurately by a point charge in force-field simulations. The charge of this first Slater function ($N_{\text{O1}} = 1.66$) is subtracted from the nuclear charge ($Z_{\text{O}} = 8$), resulting in an effective core charge $q_{\text{O,c}} = 6.34$. The oxygen atom is modeled in the force field as a point charge $q_{\text{O,c}} = 6.34$ and a Slater function with $N_{\text{O,v}} = 7.20$ and $\sigma_{\text{O,v}} = 0.22$ Å. For the hydrogen atoms, the situation is less complicated. Because a hydrogen atom only has one electron shell, the MBIS partitioning leads to a model density with only one Slater function. This same model is used for the force field. We represent a hydrogen atom by a point charge of $q_{\text{H,c}} = 1.00$ (equal to the nuclear charge) and a Slater function with $N_{\text{H,v}} = 0.57$ and $\sigma_{\text{H,v}} = 0.19$ Å.

**Table 1. MBIS Parameters and Resulting Force-Field Parameters for a Model of the Electron Density of a Water Molecule (DFT with B3LYP[36]/aug-cc-pVTZ[37])**

| element | $N_{A1}$ | $\sigma_{A1}$[a] | $N_{A2}$ | $\sigma_{A2}$[a] | $Z_A$ | $N_{A,\text{v}}$ | $\sigma_{A,\text{v}}$[a] | $q_{A,\text{c}}$ |
|---|---|---|---|---|---|---|---|---|
| O | 1.66 | 0.03 | 7.20 | 0.22 | 8 | 7.20 | 0.22 | 6.34 |
| H | 0.57 | 0.19 | | | 1 | 0.57 | 0.19 | 1.00 |
| H | 0.57 | 0.19 | | | 1 | 0.57 | 0.19 | 1.00 |

[a]Units: Å.

The total force-field interaction energy for the molecular dimer is written as a sum of 4 terms, each one with a clear physical interpretation: electrostatics, exchange-repulsion, dispersion, and induction. The intermolecular energy can thus be expressed as

$$E_{\text{inter}} = E_{\text{elst}} + E_{\text{exch-rep}} + E_{\text{disp}} + E_{\text{ind,ct}} \tag{4}$$

This corresponds to the decomposition of the total interaction energy used in SAPT. Every term will be discussed separately in the following subsections. The last three terms, the exchange-repulsion, dispersion, and induction component, all feature a separate linear parameter. These three parameters are called *interaction* parameters because they are all used to scale a specific interaction contribution to the total energy. The meaning of the exchange-repulsion interaction parameter will be discussed in section 2.2, the meaning of the dispersion interacton parameter in section 2.3, and the meaning of the induction interaction parameter in section 2.4. Note that the



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

Journal of Chemical Theory and Computation Article

electrostatic contribution, discussed in section 2.1, is constructed without any interaction parameter.

We thus obtain a force field with two different *types* of parameters. On the one hand there are atomic parameters (such as the valence population $N_v$ and valence width $\sigma_v$) obtained from the MBIS partitioning of the monomer density. These parameters are derived from monomer properties and not fitted to dimer interaction energies. On the other hand, there are three interaction parameters which are fitted to reference data for dimer interaction energies. These three interaction parameters are the only fitting parameters present in the model.

**2.1. Electrostatics.** The frozen-density electrostatic interaction energy between two molecules is given by the Coulomb integral

$$E_{\text{elst}} = \sum_{A=1}^{N^1_{\text{atoms}}}\sum_{B=1}^{N^2_{\text{atoms}}}\frac{q_{A,c}q_{B,c}}{|\mathbf{R}_A-\mathbf{R}_B|} - \sum_{A=1}^{N^1_{\text{atoms}}} q_{A,c}\int d\mathbf{r}\frac{\tilde{\rho}_2(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_A|} - \sum_{B=1}^{N^2_{\text{atoms}}} q_{B,c}\int d\mathbf{r}\frac{\tilde{\rho}_1(\mathbf{r})}{|\mathbf{r}-\mathbf{R}_B|} + \int d\mathbf{r}_1\int d\mathbf{r}_2\frac{\tilde{\rho}_1(\mathbf{r}_1)\tilde{\rho}_2(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|} \tag{5}$$

where $N^1_{\text{atoms}}$ is the number of atoms in molecule 1 and $N^2_{\text{atoms}}$ is the number of atoms in molecule 2. The first term describes repulsion between effective core charges, the second and third terms describe attraction between effective core charges and valences electrons, and finally the fourth term describes repulsion between valence electrons. The Coulomb integral can be evaluated analytically for our force-field model of the density from eq 2 to give

$$E_{\text{elst}} = \sum_{A=1}^{N^1_{\text{atoms}}}\sum_{B=1}^{N^2_{\text{atoms}}} E^{AB}_{\text{elst}} \tag{6}$$

$$E^{AB}_{\text{elst}} = \frac{q_{A,c}q_{B,c}}{R_{AB}} - \frac{q_{A,c}N_{B,v}}{R_{AB}}[1-g(\sigma_{B,v},R_{AB})] - \frac{N_{A,v}q_{B,c}}{R_{AB}}[1-g(\sigma_{A,v},R_{AB})] + \frac{N_{A,v}N_{B,v}}{R_{AB}}[1-f(\sigma_{A,v},\sigma_{B,v},R_{AB}) - f(\sigma_{B,v},\sigma_{A,v},R_{AB})] \tag{7}$$

$$g(\sigma,r) = \left(1+\frac{r}{2\sigma}\right)\exp\left(-\frac{r}{\sigma}\right) \tag{8}$$

$$f(\sigma_i,\sigma_j,r) = \frac{\sigma_i^4}{(\sigma_i^2-\sigma_j^2)^2}\left[1+\frac{r}{2\sigma_i}-\frac{2\sigma_j^2}{\sigma_i^2-\sigma_j^2}\right]\exp\left(-\frac{r}{\sigma_i}\right) \tag{9}$$

The expression for the Coulomb interaction of s-type Slater densities has been presented before.[31] For convenience we provide a derivation of the above formula in the Supporting Information. By using the net atomic charges of atoms $A$ and $B$, $q_A = q_{A,c} - N_{A,v}$ and $q_B = q_{B,c} - N_{B,v}$, we can rewrite the electrostatic interaction energy between those atoms as

$$E^{AB}_{\text{elst}} = \frac{q_Aq_B}{R_{AB}} + E^{AB}_{\text{penetration}} \tag{10}$$

The first term is the well-known electrostatic interaction between point charges. The second term decays exponentially as the atoms are separated and describes the penetration effect of the electrostatic interaction energy:

$$E^{AB}_{\text{penetration}} = \frac{q_{A,c}N_{B,v}}{R_{AB}}g(\sigma_{B,v},R_{AB}) + \frac{N_{A,v}q_{B,c}}{R_{AB}}g(\sigma_{A,v},R_{AB}) - \frac{N_{A,v}N_{B,v}}{R_{AB}}[f(\sigma_{A,v},\sigma_{B,v},R_{AB}) + f(\sigma_{B,v},\sigma_{A,v},R_{AB})] \tag{11}$$

We remark that existing implementations to evaluate long-range electrostatics in periodic systems (such as the Ewald summation,[38,39] PPPM,[40] or FMM[41]) do not need to be modified to calculate the first term in eq 10. The second term, containing the penetration correction to the electrostatic energy, clearly decays exponentially and can be evaluated similarly to other short-ranged noncovalent terms. By tabulating the values of this second term, our force field remains compatible with existing force-field codes.

The importance of the penetration effect has been discussed before.[42] Distributed charges have already been employed successfully in force fields. Most often Gaussian functions are used,[43] some force fields feature Slater densities,[30,31] and one force field even uses contracted Gaussian functions to approximate a Slater function.[44] Distributed charges offer a natural way to describe the penetration effect and can give accurate results on condition that the model density is a good approximation to the ab initio density. It has been shown that this is the case for model densities resulting from MBIS partitioning.[33]

**2.2. Exchange-Repulsion.** The wave function of a dimer has to be antisymmetric when two electrons are exchanged. This antisymmetrized wave function leads to a higher energy expectation value than if a simple product of the two monomer wave functions would be used. The resulting repulsive energy is only present when the electron densities of the two molecules in the dimer overlap. This repulsive energy dominates the so-called exchange-repulsion energy.[45] The energy due to the exchange-repulsion effect has no classical analogue and cannot be expressed simply as a functional of the isolated molecular densities. We therefore base our model on the observed proportionality between exchange-repulsion energy and overlap of electron densities:[25,46−48]

$$E_{\text{exch-rep}} \approx U_{\text{exch-rep}}S_{12} \tag{12}$$

$$= U_{\text{exch-rep}}\int d\mathbf{r}\,\tilde{\rho}_1(\mathbf{r})\,\tilde{\rho}_2(\mathbf{r}) \tag{13}$$

If we plug in our force-field model of the electron densities from eq 2, we can rewrite this as a sum of atomic pairwise interactions:

$$E_{\text{exch-rep}} = \sum_{A=1}^{N^1_{\text{atoms}}}\sum_{B=1}^{N^2_{\text{atoms}}} E^{AB}_{\text{exch-rep}} \tag{14}$$

$$E^{AB}_{\text{exch-rep}} = U_{\text{exch-rep}}S^{AB} \tag{15}$$

$$S^{AB} = \frac{N_{A,v}N_{B,v}}{8\pi R_{AB}}[h(\sigma_{A,v},\sigma_{B,v},R_{AB}) + h(\sigma_{B,v},\sigma_{A,v},R_{AB})] \tag{16}$$



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

$$h(\sigma_i, \sigma_j, r) = \left[\frac{4\sigma_i^2\sigma_j^2}{(\sigma_j^2 - \sigma_i^2)^3} + \frac{\sigma_i}{(\sigma_j^2 - \sigma_i^2)^2} r\right] \exp\left(-\frac{r}{\sigma_i}\right) \tag{17}$$

A derivation of this expression is provided in the Supporting Information.

$U_{\text{exch-rep}}$ is the first *interaction* parameter in our model. We use the term interaction parameter for this proportionality constant to avoid confusion with atomic or atom-pair parameters. Here, the *total* exchange-repulsion interaction energy is proportional to $U_{\text{exch-rep}}$, which justifies the name interaction parameter. The value of the interaction parameter $U_{\text{exch-rep}}$ will be determined in section 3 by a simple linear regression of the overlap of the force-field densities with exchange-repulsion energies from SAPT calculations for a large database of dimer configurations. It will be shown that the interaction parameter $U_{\text{exch-rep}}$ is, to some extent, system independent. Using a universal value for all dimers in the database leads to accurate results. Notice the important difference with some earlier works[24,49] where the prefactor of the overlap is atom-type specific (i.e., $E_{\text{exch-rep}}^{AB} = U_{\text{exch-rep}}^{AB} S^{AB}$). Such an approach of course involves more paramaters to be fitted to intermolecular reference data than the approach presented here.

**2.3. Dispersion.** The correlated movement of electrons in interacting molecules causes an attractive force. This effect is called dispersion. The dispersion energy between two molecules that are far apart can be written in second-order perturbation theory using a power expansion in terms of the intermolecular distance: $E_{\text{disp}} = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - \ldots$. From this expansion we only retain the two leading terms. For our force field, we write the energy as a sum of atomic pairwise contributions to the dispersion interaction. Finally we use a damping function to control the behavior for small distances. Our force-field model of the dispersion energy then has the following form:

$$E_{\text{disp}} = \sum_{A=1}^{N_{\text{atoms}}^1} \sum_{B=1}^{N_{\text{atoms}}^2} E_{\text{disp}}^{AB} \tag{18}$$

$$E_{\text{disp}}^{AB} = -f_6(x_{AB}) \frac{C_{AB}^6}{R_{AB}^6} - f_8(x_{AB}) \frac{U_{s8} C_{AB}^8}{R_{AB}^8} \tag{19}$$

$$f_n(x_{AB}) = 1 - \left(\sum_{k=0}^{n} \frac{x_{AB}^k}{k!}\right) \exp(-x_{AB}) \tag{20}$$

with $U_{s8}$ another interaction parameter that is used to linearly scale all interatomic $C_{AB}^8$ coefficients and thus the total $C^8$ dispersion interaction energy. The dispersion energy is damped at short interatomic distances using the Tang–Toennies damping function[50] with as argument the interatomic distance divided by the average of the Slater widths of the atoms involved: $x_{AB} = \frac{R_{AB}}{(\sigma_{A,v} + \sigma_{B,v})/2}$.

To determine the interatomic $C_{AB}^6$ coefficients, we use the Tkatchenko–Scheffler method.[51] In this method, the $C_{AB}^6$ coefficients of atoms in a molecule are computed using free-atom reference values that are modified based on the partitioning of the molecular electron density into atomic contributions. In the original paper, the Hirshfeld partitioning scheme[52] was used. To maintain consistency with the other force-field terms, we will use the MBIS partitioning scheme. It has been shown that the combination of the Tkatchenko–Scheffler method and MBIS partitioning leads to reliable $C_{AB}^6$ coefficients.[33] For more information about the computation of $C_{AB}^6$ coefficients we refer to the Supporting Information.

To compute the $C_{AB}^8$ coefficients, we use a recursion relation proposed by Starkschall:[53]

$$C_{AB}^8 = \frac{3}{2} C_{AB}^6 \sqrt{\frac{\langle r_A^4 \rangle}{\langle r_A^2 \rangle} + \frac{\langle r_B^4 \rangle}{\langle r_B^2 \rangle}} \tag{21}$$

The expectation values of powers of $r_A$ and $r_B$ are evaluated using the free-atom densities.[54] In the energy expression eq 19 the $C_{AB}^8$ coefficients are uniformly scaled with the interaction parameter $U_{s8}$, to account for the higher order contributions to the dispersion energy and possible deficiencies in the calculation of the $C_{AB}^8$ coefficients and damping function.

**2.4. Induction.** The electrostatic interaction given by eq 5 assumes a frozen-density approach, where the molecular densities are kept frozen when the two molecules approach each other. In reality, the molecular densities will deform as both molecules approach each other, an effect called polarization. The resulting stabilization is called the induction energy. The construction of polarizable force fields that succeed in quantitatively reproducing this induction energy is not trivial and is an issue that has received a lot of attention.[55,56] The most common approach is to model each atom as an inducible dipole, either using a Drude oscillator (charge-on-spring) model[15,57−62] or as a Thole- or Gausian-damped point dipole.[60,63−65] Each inducible dipole interacts with the atomic point charges and all other dipoles. For a given geometry, the energy of the inducible dipoles is minimized, in analogy to the Born–Oppenheimer approximation. In such models, at least two parameters per atom are needed: an atomic dipole polarizability and a damping parameter. We have tested models with inducible point dipoles, where realistic atomic polarizabilities are obtained with the Tkatchenko–Scheffler method.[51] However, the resulting induction energies were too small (in absolute value) when compared to SAPT reference values, especially at short intermolecular distances. The same conclusion is obtained for the induction term of the AMOEBA[66] force field. The results of those two induction models are presented in section 5 of the Supporting Information. This investigation reveals that another phenomenon plays a dominant role in the SAPT induction component. As an alternative, some of the authors recently proposed a polarizable force field where all parameters can be evaluated as expectation values from a quantum-mechanical wave function.[67,68] This method is however not readily applicable to classical molecular dynamics simulations, as it would require evaluating the quantum-mechanical wave function at every time step.

We therefore resort to an approximation with a simpler analytical expression: the induction energy is modeled as a pairwise-additive term with one interaction parameter, $U_{\text{ind,ct}}$.

$$E_{\text{ind,ct}} = \sum_{A=1}^{N_{\text{atoms}}^1} \sum_{B=1}^{N_{\text{atoms}}^2} E_{\text{ind,ct}}^{AB} \tag{22}$$

$$E_{\text{ind,ct}}^{AB} = -U_{\text{ind,ct}} S^{AB} \tag{23}$$

The expression for the overlap of 1s-type Slater densities $S^{AB}$ is given in eq 16. The proposed proportionality of the intermolecular induction energy is inspired by the exponentially



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

decaying nature of charge transfer.[23,32,69−73] More specifically, the proportionality between exchange-repulsion and the charge-transfer interaction has been demonstrated in the context of SAPT.[74] We expect that this assumption will work fairly well for dispersion-dominated complexes, where the induction energy is dominated by contributions from charge-transfer effects because classical polarization plays a minor role in such complexes. The induction energy is in general attractive, so we expect a positive value for $U_{ind,ct}$. The pairwise-additive expression for the induction energy is a serious approximation, and its applicability will be assessed in section 3. For complexes where the proposed method is not viable we will discuss possible causes of the observed deviations and extensions of the model.

In summary, our force field for noncovalent interactions contains three interaction parameters: $U_{exch\text{-}rep}$ gives the proportionality between the exchange-repulsion energy and monomer density overlap, $U_{s8}$ is a uniform scaling factor for the $C^8$ coefficients used to compute the dispersion energy, and finally $U_{ind,ct}$ gives the proportionality between the induction energy and monomer density overlap. Note that the total interaction energy is a linear function of these three interaction parameters.

## 3. PARAMETER CALIBRATION

In the previous section, analytical expressions have been proposed for the four terms composing the total intermolecular interaction energy: $E_{inter} = E_{elst} + E_{exch\text{-}rep} + E_{disp} + E_{ind,ct}$, representing electrostatic, exchange-repulsion, induction, and dispersion energy, containing three interaction parameters in total. These interaction parameters together with the MBIS partitioned density of the molecules at hand form a complete force field to describe the interaction between those two molecules. High-level ab initio interaction energies obtained for a large database of dimers will be used to determine the three interaction parameters. At the same time, the degree of reproduction of the high-level ab initio intermolecular interaction energies by the force field serves as a first validation that our force-field model is appropriate and that the interaction parameters for the exchange-repulsion and dispersion component are to a large extent system independent and their values can be considered universal.

We used the S66x8[75] database as a training set to determine the values of the interaction parameters. This database covers the most common types of noncovalent interactions in biomolecules, keeping a balanced representation of electrostatic-dominated, dispersion-dominated. and mixed-influence complexes. Based on a SAPT decomposition of the interaction energies, all 66 complexes in the database are classified either as hydrogen-bonded (23 complexes), dispersion-dominated (23 complexes), or "others" (20 complexes). In order to construct a total of 66 dimers, various combinations of 14 monomers are considered. The 14 selected monomers are acetic acid, acetamide, benzene, cyclopentane, ethene, ethyne, neopentane, *n*-pentane, methylamine, methanol, *N*-methylacetamide, pyridine, uracil, and water. All these monomers are composed exclusively of hydrogen, carbon, nitrogen, and oxygen atoms. All complexes are relatively small (6 to 34 atoms), such that CCSD(T)/CBS and SAPT calculations are feasible for the entire data set. The S66x8 database not only contains equilibrium geometries of the 66 dimers but also scans the dissociation curve with displacements relative to the equilibrium distance of 0.90, 0.95, 1.00, 1.05, 1.10, 1.25, 1.50, and 2.00, resulting in 528 = 66 × 8 data points. We note that a relative displacement of 1.00 corresponds approximately with the equilibrium separation distance. Finally we note that the monomer geometries remain fixed along the dissociation curve, which means that there are no changes in the covalent energy along the dissociation curve. This allows us to assess the accuracy of the noncovalent force field.

**3.1. SAPT Energies for the S66x8 Database.** One of the ultimate goals of the force field is the capability to accurately reproduce dissociation curves obtained using the CCSD(T)/CBS method. We proposed analytical expressions for four contributions composing the total intermolecular interaction energy: electrostatic, exchange, dispersion, and induction. A decomposition of the CCSD(T)/CBS energy into these four categories is not straightforward and not unambiguous. In the context of force-field development it is however instructive to also study methods that provide more chemical insight into the interactions, by decomposing the energy into meaningful contributions. It should be stressed that a decomposition of the total interaction energy is not unique and is somewhat arbitrary. Indeed, several decomposition schemes proposed in the literature, such as Kitaura−Morokuma[76] (based on the Hartree−Fock approximation), constrained space orbital variation,[77] or density-based energy decomposition analysis,[29] will all lead to different results. We will use SAPT[19] as a benchmark, as it directly computes electrostatics, exchange-repulsion, dispersion, and induction contributions separately. SAPT is an ab initio method and gives total interaction energies in reasonable agreement with CCSD(T)/CBS values.[78] Unfortunately, due to its double perturbative nature (the intermolecular interaction and the residual monomer correlation interactions), it is not always clear how to classify higher-order terms, as they show a mixture of exchange, dispersion, and induction. It is also clear that the perturbation expansion is not converged in third order, the highest order currently available in SAPT implementations. This is why we start by studying different SAPT approximations and compare them to CCSD(T)/CBS values. Our results are in line with conclusions from a similar analysis by Li et al.[78]

We used PSI4[79] for all SAPT and coupled-cluster calculations. For SAPT, the aug-cc-pVTZ basis set[37] was used with density fitting[80,81] to avoid evaluation of four-center integrals. Core orbitals were frozen and MP2 natural orbitals[82] were used to speed up the evaluation of second-order $T_2$ amplitudes and the triples contribution to the dispersion energy. All exchange-type terms are scaled with the factor $\frac{E_{exch}^{(10)}}{E_{exch}^{(10)}(S^2)}$, as recommended in the literature.[83] The CCSD(T)/CBS energies were computed as the sum of the Hartree−Fock energy in the aug-cc-pVQZ basis set,[37] an MP2 correction was extrapolated to the complete basis set limit using Helgaker's formula[84] applied to aug-cc-pVTZ and aug-cc-pVQZ basis sets,[37] and finally a CCSD(T) correction was computed with the aug-cc-pVDZ basis set.[37] This procedure corresponds to the one used originally for the S66 data set.[75]

In Table 2 we compare several SAPT levels of theory with CCSD(T)/CBS values. As reported before for the S66 data set[83] (only for the equilibrium separation distance), SAPT2+(3) performs better than SAPT2+3. This is an indication that the SAPT expansion is not converged, because an exact treatment of third-order perturbations (SAPT2+3) performs worse than an approximate treatment (SAPT2+(3)). Convergence properties of SAPT have been discussed



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

**Table 2. Rmsd of CCSD(T)/CBS and SAPT Interaction Energies in kJ mol$^{-1}$ for the S66x8 Data Set**

| relative displacement | SAPT0 | SAPT2+(3) | SAPT2+3 |
|---|---|---|---|
| 0.90 | 10.4 | 2.6 | 4.6 |
| 0.95 | 8.2 | 1.8 | 3.2 |
| 1.00 | 6.6 | 1.4 | 2.4 |
| 1.05 | 5.4 | 1.1 | 1.8 |
| 1.10 | 4.4 | 0.9 | 1.4 |
| 1.25 | 2.7 | 0.6 | 0.8 |
| 1.50 | 1.5 | 0.4 | 0.4 |
| 2.00 | 0.7 | 0.2 | 0.2 |

extensively, see refs 19, 85, and 86 and references therein. We note that although SAPT2+(3)$\delta$MP2/aug-cc-pVTZ[37] is considered as the optimal SAPT level of theory,[83] it cannot be used for the present application because it is no longer possible to clearly separate induction from dispersion contributions. We therefore use SAPT2+(3) in the remainder of this work, with the caveat that it may show some deficiencies, especially at short intermolecular distances.

**3.2. Parameter Calibration by Comparing with SAPT Energy Components.** We will now investigate each term of the force field separately and compare it with its SAPT counterpart, in order to determine the introduced interaction parameters. All the interaction parameters are fitted to the 23 dispersion-dominated complexes present in the S66x8 data set, except if noted otherwise. Although this selection is only really needed for the induction term, as will be explained below, other parameters were treated consistently. The electron densities of the isolated monomers in the S66x8 database were calculated using DFT with a B3LYP[36] functional and aug-cc-pVTZ basis set[37] using Gaussian.[87] The molecular electron density was then partitioned with the MBIS scheme using HORTON,[88] with the built-in ultrafine grid.

*3.2.1. Electrostatics.* No interaction parameters are included in the electrostatic term $E_{elst}$ (eq 5) of the force field. It simply consists of the Coulomb interaction between the total charge densities of the two monomers determined by the effective core point charges and the Slater-type charge distributions representing the valence electrons. Figure 1 shows a correlation diagram of the force-field electrostatic energies and the reference SAPT values for the 528 structures of the S66x8 data set. As expected, most of the electrostatic interaction energies are negative (510 out of 528 structures). For practical reasons, we only display those cases where the SAPT electrostatic energy is below −1 kJ mol$^{-1}$. It is clear that the force-field electrostatic model is able to capture the important penetration effect accurately. In Figure 2 we provide a similar

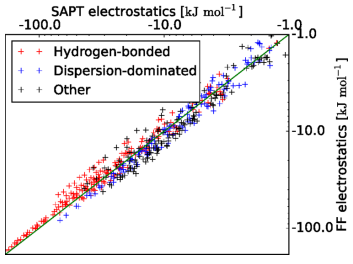



**Figure 1.** Correlation plot of SAPT and force-field electrostatic energies for the S66x8 database. Note the use of a logarithmic scale.

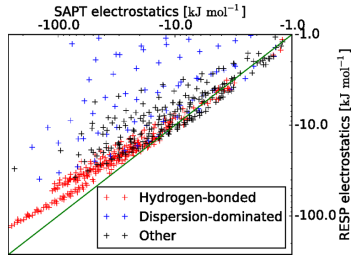



**Figure 2.** Correlation plot of SAPT and RESP fitted point charge (HF/6-31G*) electrostatic energies for the S66x8 database. Note the use of a logarithmic scale.

plot to investigate the performance of RESP[89] fitted point charges (obtained using antechamber[90]) from a HF/6-31G* density, a method often used in force fields. It is clear that the agreement is unsatisfactory. We note that point charges obtained with other methods (partitioning schemes such as Hirshfeld,[52] iterative-Hirshfeld,[91] or ESP fits with other levels of theory) result in similar deviations because of their intrinsic incapability to model the penetration effect.[33] Also the electrostatic part of the AMOEBA force field,[64,66,92,93] calculated using point multipoles up to quadrupoles ($l = 2$), is significantly improved when including an empirical penetration model.[27]

*3.2.2. Exchange-Repulsion.* The exchange-repulsion term of our force field contains one interaction parameter, $U_{exch\text{-}rep}$. We determine this parameter by minimizing a cost function $\chi^2$ that expresses the mean-square-deviation between force-field and SAPT exchange-repulsion energy:

$$\chi^2 = \frac{1}{N_d}\sum_{n=1}^{N_d}[E^{FF}_{exch\text{-}rep}(n) - E^{SAPT}_{exch\text{-}rep}(n)]^2 \quad (24)$$

The summation over $n$ runs over all $N_d$ data points that we consider. Because our force-field exchange-repulsion energy $E^{FF}_{exch\text{-}rep}(n) = U_{exch\text{-}rep}S(n)$ is a linear function of $U_{exch\text{-}rep}$ (with $S(n)$ the overlap of our molecular force-field densities), minimizing the cost function $\chi^2$ effectively amounts to a linear regression with solution

$$U_{exch\text{-}rep} = \frac{\sum_n S(n)E^{SAPT}_{exch\text{-}rep}(n)}{\sum_n S(n)\,S(n)} \quad (25)$$

We first determine the value of $U_{exch\text{-}rep}$ for every dimer separately, and Figure 3 justifies our approximation that the value of the interaction parameter $U_{exch\text{-}rep}$ is indeed universal, as it takes on about the same value for all dimers independent of their classification (hydrogen-bonded, dispersion-dominated, or other). A notable outlier is the neopentane dimer (index 36). By fitting to all dispersion-dominated complexes from the S66x8 database, we find $U_{exch} = 8.13$ in atomic units. In Figure 4 we show that our model is capable of reproducing SAPT exchange-repulsion energies, justifying the assumed functional expression that is proportional to the overlap of monomer electron densities. It turns out that for hydrogen-bonded complexes the model slightly underestimates the exchange-



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

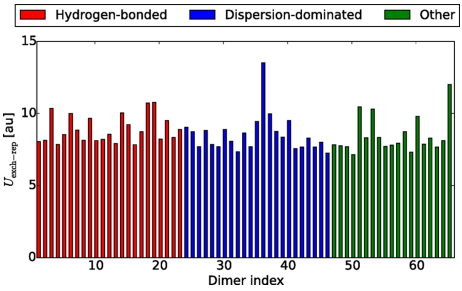



**Figure 3.** Values of $U_{exch-rep}$ specific for each dimer.

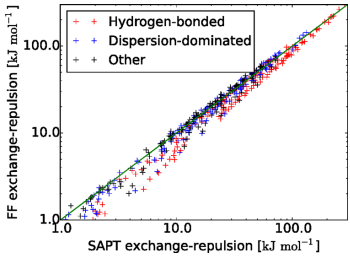



**Figure 4.** Correlation plot of SAPT and force-field exchange-repulsion energies for the S66x8 database. Note the use of a logarithmic scale.

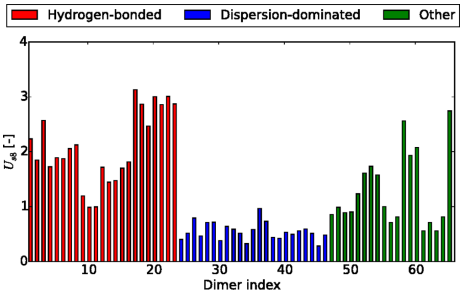



**Figure 5.** Values of $U_{s8}$ specific for each dimer.

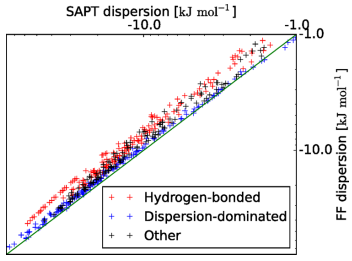



**Figure 6.** Correlation plot of SAPT and force-field dispersion energies for the S66x8 database. Note the use of a logarithmic scale.

repulsion energy. Fortunately, this error is (at least partly) compensated by an underestimation of the electrostatic attraction energy for the same complexes (see Figure 1). Our model seems to be better suited to reproduce the sum of exchange-repulsion and electrostatic energies, rather than both individual SAPT components. This can probably be related to the somewhat arbitrary classification of some terms, one of the limitations of the energy-decomposition methods mentioned earlier.

*3.2.3. Dispersion.* Our expression for the dispersion energy contains one interaction parameter $U_{s8}$ to scale the contribution of the term decaying as $r^{-8}$. To avoid interference from possible deficiencies of our Tang−Toennies damping function, we choose to fit $U_{s8}$ to relatively large intermolecular separations (relative displacements 1.50, 1.60, 1.70, 1.80, 1.90, 2.00, 2.25, and 2.50, some of which are not included in the original S66x8 database). The fitting is first performed for each dimer separately, and the resulting values of $U_{s8}$ are reported in Figure 5. We note a rather large spread on the values of the interaction parameter. Only within the class of dispersion-dominated complexes it is reasonable to attribute a universal value to $U_{s8}$ and we find $U_{s8}$ = 0.57. If we compare the SAPT dispersion energies with our model energies (for $U_{s8}$ = 0.57) we find a good agreement for the dispersion-dominated complexes in the S66x8 data set as shown in Figure 6. Surprisingly the model also performs acceptably well for the hydrogen-bonded and other complexes. This is due to a fortuitous error cancellation between the underestimated value for $U_{s8}$ and the overestimated damping for those complexes at short intermolecular separation. The value $U_{s8}$ = 0.57 can thus also be considered to be universally applicable.

*3.2.4. Induction.* Our current induction model has a simple pairwise-additive functional form, again involving only one interaction parameter $U_{ind,ct}$ that provides a linear relationship between the induction energy and monomer density properties. Figure 7 reveals that fitting $U_{ind,ct}$ for every dimer separately

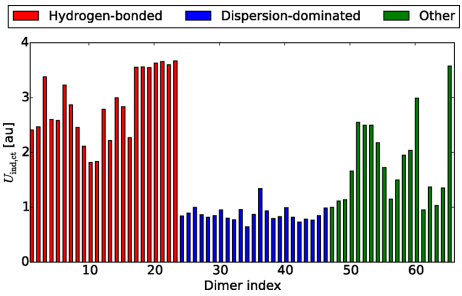



**Figure 7.** Values of $U_{ind,ct}$ specific for each dimer.

results in a large spread of values. It is notable, but perhaps not surprising, that $U_{ind,ct}$ is systematically higher for hydrogen-bonded than for dispersion-dominated complexes. It also becomes clear that only within the group of dispersion-dominated complexes does it make sense to define a universal value for $U_{ind,ct}$. Nevertheless we tested the approach with a universal value for $U_{ind,ct}$ for the entire data set on the one hand and the different subsets on the other hand, and report the



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

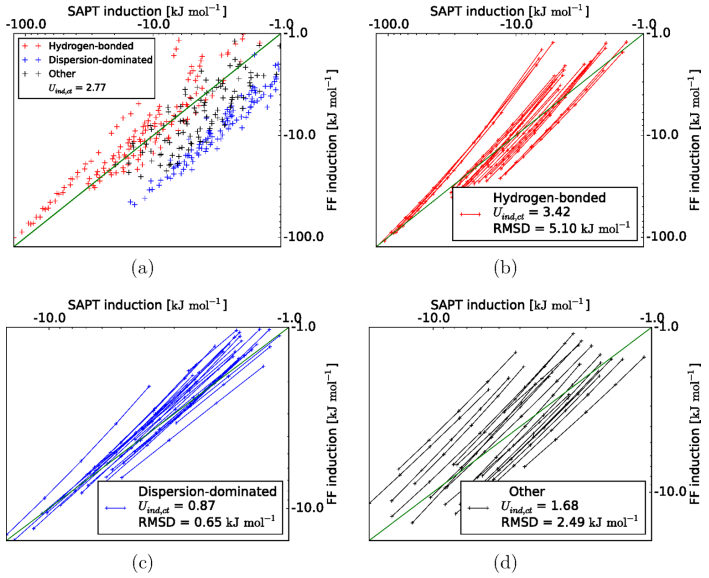



**Figure 8.** Correlation plots of SAPT and force-field induction energies. (a) Complete S66x8 database with $U_{ind,ct}$ = 2.77. (b) Only hydrogen-bonded complexes in cost function ($U_{ind,ct}$ = 3.42). (c) Only dispersion-dominated complexes in cost function ($U_{ind,ct}$ = 0.87). (d) Only "other" complexes in cost function ($U_{ind,ct}$ = 1.68). Note the use of a logarithmic scale. Data points on the dissociation curve of each dimer are connected with a line.

results in Figure 8. Figure 8c confirms that by setting $U_{ind,ct}$ = 0.87 we obtain a quantitatively accurate induction model for the dispersion-dominated complexes. From Figures 8b and 8d we learn that a similar approach is not viable for hydrogen-bonded and *other* complexes, although at the same time it seems that the correlation between SAPT induction and overlap is rather good for each separate complex (i.e., the curves connecting data points belonging to the same complex are relatively straight.)

We investigate this further on a complex-specific basis. In Figure 9a we compare our model induction with the SAPT counterpart for the water dimer (index 01 of the S66x8 data set, containing one hydrogen bond) where the parameter $U_{ind,ct}$ = 2.42 is obtained by fitting only to the 8 points on the dissociation curve of the water dimer. To make sure that the model still holds for other configurations on the water-dimer potential energy surface, we validate it for the geometries of the water dimer in the S66a8 data set.[94] This data set contains angular-displaced nonequilibrium geometries for the complexes present in the S66x8 data set. Similar plots are shown in Figures 9b, 9c, and 9d for the water−peptide dimer (index 04, containing one hydrogen bond), the acetamide dimer (index 21, containing two hydrogen bonds), and the benzene−methanol dimer (index 55, containing an OH−$\pi$ bond) respectively. In general the absolute value of the SAPT induction energy is underestimated at long intermolecular distances (small absolute values of SAPT induction energies) while the absolute value is overestimated at short intermolecular distances (large absolute values of SAPT induction energies). This is because at long range the SAPT induction energy decays as a power of $\frac{1}{r}$, while our model shows a much faster exponential decay. This could be solved by the inclusion of, for instance, inducible dipoles which are not present in our pairwise-additive force field. In general the MEDFF induction energies for the complexes shown in Figure 9 are in reasonable agreement with SAPT induction energies, and especially the transferability to angular-displaced geometries (very challenging for directional hydrogen bonds) is encouraging. We thus conclude that our induction model cannot describe all dimers in the S66x8 data set with just one universal value for $U_{ind,ct}$; only for the subset of dispersion-dominated dimers is this a viable approach. Other dimers such as hydrogen-bonded complexes can still be described relatively well using our induction model, provided one uses a value for $U_{ind,ct}$ specifically fitted to the dimer at hand. Plots similar to Figure 9 are included in the Supporting Information for all complexes of the S66x8 data set.

**3.3. Refining Parameters by Calibrating to CCSD(T)/CBS Energies.** By comparison of SAPT2+(3) interaction energies with CCSD(T)/CBS reference values, it became clear in section 3.1 that the perturbation expansion is not fully converged in SAPT2+(3), especially at shorter intermolecular distances. It is also important to note that the total force-field energy is composed of several terms that cancel each other to a large extent. The exchange-repulsion term is always positive, while induction, dispersion, and electrostatic are usually negative contributions. The typical behavior of all these terms is illustrated in Figure 10 for the dissociation of two dimers selected from the S66x8 database. An important consequence is that, in order to get reasonably accurate total energies, all the components need to be very accurate. Otherwise small relative errors on one component will result in large relative errors on



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

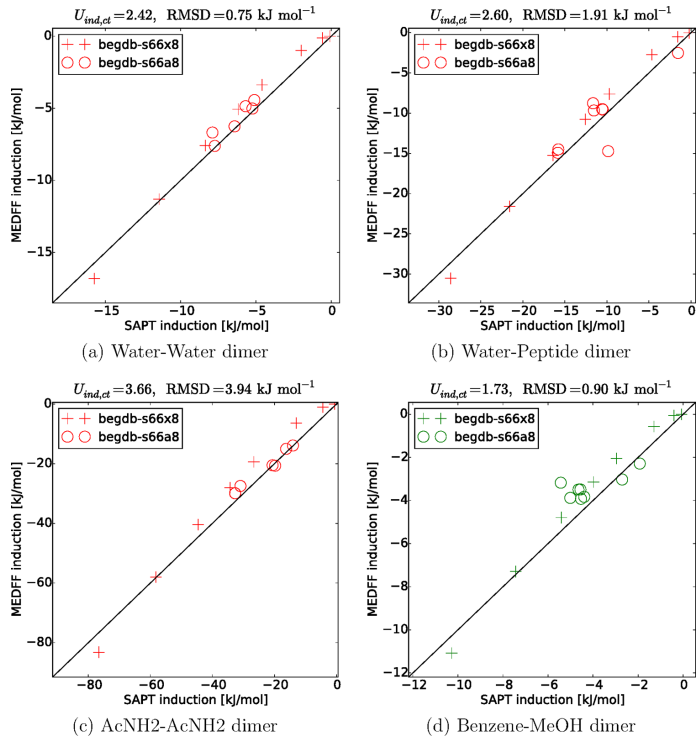



**Figure 9.** Correlation plots of SAPT and force-field induction energies. The parameter $U_{ind,ct}$ has a different value for each complex, obtained by fitting to the 8 geometries of that complex in the S66x8 data set. The rmsd is computed for all geometries of that complex in S66x8 and S66a8.

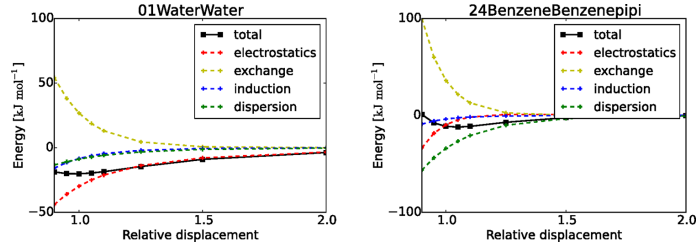



**Figure 10.** Dissociation curves showing components of the SAPT energy for the water dimer (left, hydrogen-bonded) and $\pi-\pi$ benzene dimer (right, dispersion-dominated).

the total energy. This is why we prefer to further refine the interaction parameters together by fitting to total interaction energies from CCSD(T)/CBS calculations. A conventional minimization of a quadratic cost function often leads to an ill-conditioned fitting procedure in a high-dimensional parameter space, with an imminent danger of overfitting. Even when the number of parameters is rather modest, so-called sloppy modes can lead to parameter sets that are unphysical.[95,96] Furthermore, all information deduced from the SAPT decomposition would be lost by using a conventional quadratic cost function. Although the total SAPT energy may not be very accurate, the interaction parameters of our force field fitted to individual SAPT components certainly are an excellent first guess for further refinement of the parameters to reproduce CCSD(T)/CBS energies. This is why we will use ridge regression in the next subsection to refine the interaction parameters.



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

Journal of Chemical Theory and Computation Article

We note that only the difference of the interaction parameters $U_{\text{exch-rep}} - U_{\text{ind,ct}}$ is important, as both $U_{\text{exch-rep}}$ and $-U_{\text{ind,ct}}$ are prefactors of the same functional form (the overlap of electron densities). We however choose to treat these parameters separately in the remainder, as knowing the contribution of exchange-repulsion and induction can be useful for the physical interpretation of the interactions in the systems at hand.

In the previous sections, we have shown that our force field is particularly suited for the dispersion-dominated complexes. We will refer to this force field as MEDFF (monomer electron density based force field) and will obtain universal values of the interaction parameters for MEDFF from fitting to the dispersion-dominated complexes from the S66x8 data set. It is clear that especially the induction term needs improvement to also describe hydrogen-bonded and other complexes. To demonstrate that the functional form of our force field is appropriate for those complexes, we show results for a few selected cases where the interaction parameters are determined specifically for each case.

*3.3.1. Ridge Regression or Tikhonov Regularization.* A refinement of the interaction parameters can be performed by applying Tikhonov regularization, sometimes also termed ridge regression, to the conventional quadratic cost function, similar to Bayesian statistics.[97−99] In this work we will use the following cost function:

$$\chi^2 = \frac{1}{2}\sum_{n=1}^{N_d}\frac{(E_{\text{FF}}^n[\mathbf{U}] - E_{\text{ref}}^n)^2}{N_d} + \frac{1}{2\sigma_{\text{pr}}^2}\sum_{\alpha=1}^{N_u}\left(\frac{U_\alpha - U_\alpha^{\text{SAPT}}}{U_\alpha^{\text{SAPT}}}\right)^2 \quad (26)$$

The cost function $\chi^2$ has dimension [energy$^2$] (note that the parameter $\sigma_{\text{pr}}$ has dimension [energy$^{-1}$] as all terms in the second summation are dimensionless). The sum over $n$ runs over all $N_d$ data points in the database, while the sum over $\alpha$ runs over all $N_u$ interaction parameters, thus $N_u = 3$ in the case of MEDFF. The vector $\mathbf{U}$ contains all interaction parameters, and the force-field energy $E_{\text{FF}}^n[\mathbf{U}]$ of dimer $n$ depends on these interaction parameters. The second term, which regularizes the cost function, is called the prior and its influence is controlled by the parameter $\sigma_{\text{pr}}$. Our prior term in the cost function penalizes large relative deviations of the interaction parameters from their values computed by fitting to individual SAPT energy components, denoted by $U_\alpha^{\text{SAPT}}$.

The parameter $\sigma_{\text{pr}}$ cannot be computed a priori, instead its optimal value has to be determined as part of the fitting procedure. This can be achieved by first computing the estimated prediction error (EPE) as a function of $\sigma_{\text{pr}}$. The EPE can be estimated from the training data set by cross validation (CV):

$$\text{EPE}^{\text{CV}} = \left(\sum_{n=1}^{N_d}\frac{(E_{\text{FF}}^n[\mathbf{U}(D_n)] - E_{\text{ref}}^n)^2}{N_d}\right)^{1/2} \quad (27)$$

Here each of the $N_d$ points in the database are in turn left out of the fitting procedure and the point that is left out contributes to $\text{EPE}^{\text{CV}}$. We denote the database without point $n$ as $D_n$. The notation $\mathbf{U}(D_n)$ is used to indicate that the parameters $\mathbf{U}$ are fitted to points in the database $D_n$, in this case the database that contains all data points except for point $n$. The optimal values of the parameters $\mathbf{U}$ depend on the value of $\sigma_{\text{pr}}$ in the cost function, and in this way also $\text{EPE}^{\text{CV}}$ depends on $\sigma_{\text{pr}}$. We computed $\text{EPE}^{\text{CV}}$ for several values of $\sigma_{\text{pr}}$. In Figure 11 we only consider dispersion-dominated complexes of the S66x8 database and use the force field for all energy contributions. We see that small values of $\sigma_{\text{pr}}$ result in a high EPE because all interaction parameters are constrained to their SAPT values, which results in large differences with CCSD(T)/CBS energies. As soon as $\sigma_{\text{pr}}$ is larger than 0.1 mol kJ$^{-1}$, the EPE does not decrease appreciably. Therefore, we consider the parameters that are found by minimizing the cost function with $\sigma_{\text{pr}} = 0.1$ mol kJ$^{-1}$ to offer an optimal combination of being physical (close to parameters from fitting to separate SAPT components) and at the same time leading to good correspondence between CCSD(T)/CBS and force-field energies. The fact that the EPE does not increase again indicates that we are not overfitting the parameters.

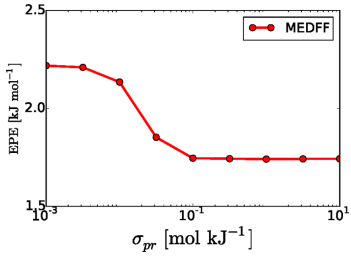



**Figure 11.** EPE initially decreases as $\sigma_{\text{pr}}$ increases (the prior contribution in the cost function diminishes). However, the EPE remains nearly constant as soon as $\sigma_{\text{pr}} \geq 0.1$ mol kJ$^{-1}$.

The values of the interaction parameters are given in Table 3 for different contributions of the prior to the cost function. If $\sigma_{\text{pr}} = 0.0$ mol kJ$^{-1}$, the prior dominates the cost function and we obtain the values from fitting to the SAPT components separately. On the other hand, if $\sigma_{\text{pr}} = \infty$, the prior does not contribute to the cost function and we obtain parameters from an unrestrained fit to CCSD(T)/CBS reference energies. If we choose $\sigma_{\text{pr}} = 0.1$ mol kJ$^{-1}$ for MEDFF, we obtain parameters that show small relative deviations from parameters fitted to SAPT components and at the same time result in energies close to CCSD(T)/CBS reference values. In the remainder of this paper we will work with the parameters obtained by minimizing the cost function with $\sigma_{\text{pr}} = 0.1$ mol kJ$^{-1}$.

**Table 3. Optimal Values of the MEDFF Interaction Parameters (All in Atomic Units) for Different Prior Contributions ($\sigma_{\text{pr}}$ in mol kJ$^{-1}$) to the Cost Function**

| $\sigma_{\text{pr}}$ [mol kJ$^{-1}$] | 0.0 | 0.1 | ∞ |
|---|---|---|---|
| $U_{\text{exch-rep}}$ | 8.13 | 8.43 | 8.27 |
| $U_{\text{ind,ct}}$ | 0.87 | 0.86 | 0.86 |
| $U_{s8}$ | 0.57 | 0.57 | 0.48 |

We note that the refinement of the parameters mainly influences the interaction energies at rather small intermolecular separations. The resulting differences are in general smaller than the remaining errors with respect to the reference ab initio interaction energies.

*3.3.2. Results.* With the refinement of the interaction parameters according to the cost function (eq 26), our force field MEDFF is completely constructed. The final values of the interaction parameters are tabulated in Table 3 in the column with $\sigma_{\text{pr}} = 0.1$ mol kJ$^{-1}$. In a first internal validation we check whether MEDFF succeeds in reproducing the reference data



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

that were used in the fitting procedure. We present the rmsd with respect to CCSD(T)/CBS interaction energies for our force field MEDFF in Table 4. We compare with the recently developed QMDFF,[18] where we note that QMDFF was fitted to B3LYP-D3/def2-QZVP reference data for the S22 data set, a subset of the S66 data set. MEDFF performs very well and can be considered chemically accurate (rmsd < 1 kcal mol$^{-1}$). We show the dissociation curves for a few selected dimers obtained with our force field MEDFF in Figure 12, and again compare with QMDFF. Dissociation curves for dispersion-dominated complexes are presented in the Supporting Information. These results reveal that the MEDFF force field performs remarkably well for all dispersion-dominated complexes. The neopentane dimer (index 36) is a notable exception, where also QMDFF fails to reproduce the dissociation curve correctly. A detailed analysis reveals that the main source of error is in the exchange-repulsion contribution, which is severely underestimated. This can also be seen in Figure 3, where it is clear that the value of the interaction parameter $U_{exch,rep}$ fitted only to the neopentane dimer (index 36) is significantly above average.

**Table 4. Rmsd of CCSD(T)/CBS and Force-Field Interaction Energies in kJ mol$^{-1}$ for the Dispersion-Dominated Complexes in the S66x8 Database**[a]

| relative displacement | MEDFF | QMDFF |
|---|---|---|
| 0.90 | 3.7 | 6.2 |
| 0.95 | 2.2 | 5.3 |
| 1.00 | 1.5 | 4.7 |
| 1.05 | 1.1 | 4.2 |
| 1.10 | 0.9 | 3.8 |
| 1.25 | 0.6 | 2.7 |
| 1.50 | 0.4 | 1.5 |
| 2.00 | 0.2 | 0.5 |

[a]The definition of relative displacements is discussed elsewhere.[75] CCSD(T)/CBS and force-field energies are always compared for the same geometry.

We now turn our attention to some dimers that are not dominated by dispersion. Just as in section 3.2 we will show the results for the water dimer (containing one hydrogen bond), the water−peptide dimer (containing one hydrogen bond), the acetamide dimer (containing two hydrogen bonds), and the benzene−methanol dimer (containing an OH−$\pi$ bond) where we determine the parameters specifically for those complexes using the regularization procedure described at the beginning of this section. Again we use the 8 geometries of each complex in the S66x8 data set as training data for the parameters for that specific complex. The geometries of the same complex in S66a8 are used afterward as a validation. All results for the 4 complexes mentioned are summarized in Figure 13. For the water dimer the results are not entirely satisfying. A detailed analysis reveals that this is due to errors in the electrostatic term, most likely because the MBIS model gives isotropic atomic densities while atomic multipoles are indispensable for this specific case. For the other cases shown here the agreement is acceptable and the transferability of the model to the angular-displaced geometries of the S66a8 data set is remarkable. The results for all complexes are provided in the Supporting Information.

In Table 5 we summarize the results for the force field where parameters are obtained specifically for each complex, in contrast to our universal approach presented earlier. The errors with respect to CCSD(T)/CBS are well below 1 kcal mol$^{-1}$ thereby demonstrating that our proposed energy expression can be used to reproduce high-level ab initio data. Of course one has to keep in mind that, to use the force field in this way for complexes not present in the S66x8 data set, a limited number of new ab initio calculations need to be performed.

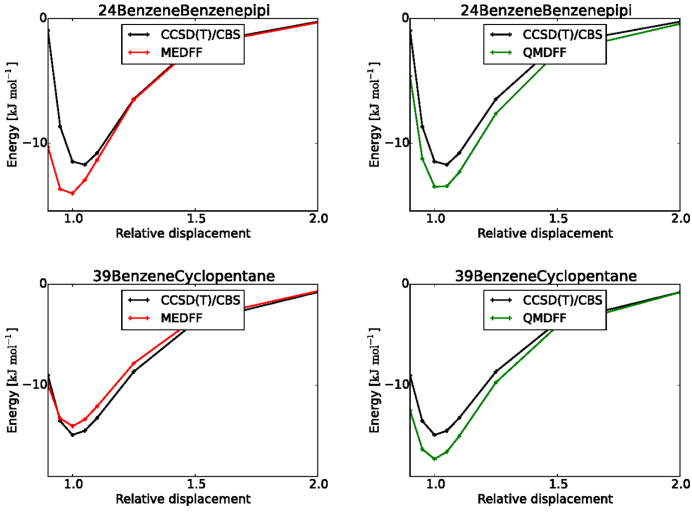



**Figure 12.** Dissociation curves of the $\pi$−$\pi$ pyridine dimer and the benzene−cyclopentane complex (both dispersion-dominated complexes).



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

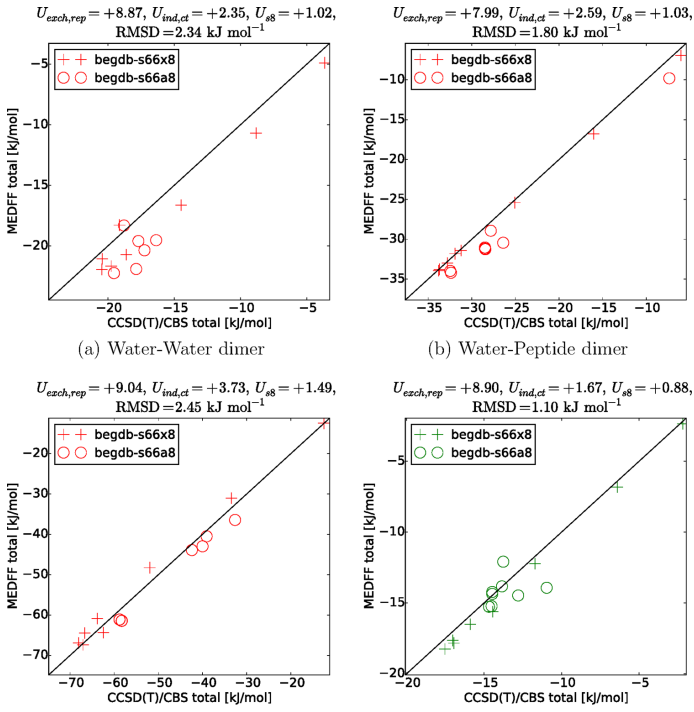



**Figure 13.** Correlation plots of CCSD(T) and force-field interactions energies. The parameters $U_{exch\text{-}rep}$, $U_{s8}$, and $U_{ind,ct}$ have a different value for each complex, obtained by fitting to the 8 geometries of that complex in the S66x8 data set. The rmsd is computed for all geometries of that complex in S66x8 and S66a8.

**Table 5. Rmsd of CCSD(T)/CBS and Force-Field Interaction Energies in kJ mol$^{-1}$ for All Complexes in the S66x8 Database**[a]

| relative displacement | hydrogen-bonded | dispersion-dominated | others | all |
|---|---|---|---|---|
| 0.90 | 1.89 | 0.72 | 1.34 | 1.40 |
| 0.95 | 0.46 | 0.39 | 0.88 | 0.60 |
| 1.00 | 1.50 | 0.58 | 1.13 | 1.13 |
| 1.05 | 2.39 | 0.69 | 1.32 | 1.64 |
| 1.10 | 2.94 | 0.73 | 1.38 | 1.94 |
| 1.25 | 3.31 | 0.68 | 1.20 | 2.10 |
| 1.50 | 2.26 | 0.51 | 0.72 | 1.42 |
| 2.00 | 0.70 | 0.24 | 0.26 | 0.46 |
| *All* | 2.15 | 0.59 | 1.09 | 1.45 |

[a]The force-field interaction energies are obtained using parameters determined specifically for each separate complex.

## 4. EXTERNAL VALIDATION

**4.1. External Validation on the hsg Data Set.** The next step in the validation of the new force field assesses to which extent MEDFF reproduces interaction energies of noncovalent complexes not included in the S66x8 data set. For this external validation we select the hsg data set.[100] The hsg data set consists of 21 interacting fragment pairs, extracted from an HIV-II protease crystal structure with a bound ligand (indinavir). The molecules included in this data set are, among others, 2-formamidoacetamide, *N*-methylacetamide, benzene, toluene, ethane, propane, butane, pyridine, methanol, and acetic acid. We only retain the 13 dimers that are dominated by dispersion in this study. The values of the interaction parameters of our force field are not reoptimized, but instead taken from the fitting for the S66x8 data set discussed earlier. This means that the computationally most expensive step to construct a force field for these complexes is a DFT calculation for the monomers. We note that a similar computational effort is required for force fields where charges are obtained from an ab initio density (for example from ESP fitting or partitioning).

We compare the performance of our force field with some other force fields (quantum mechanically derived force field[18] (QMDFF), the generalized amber force field[101] (GAFF), and the Merck molecular force field[102] (MMFF)) in Figure 14. It is encouraging to see that our force field, MEDFF, performs well for dispersion-dominated complexes, even though the molecules present in this data set were not used to fit the interaction parameters. Also the good performance of QMDFF for these



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

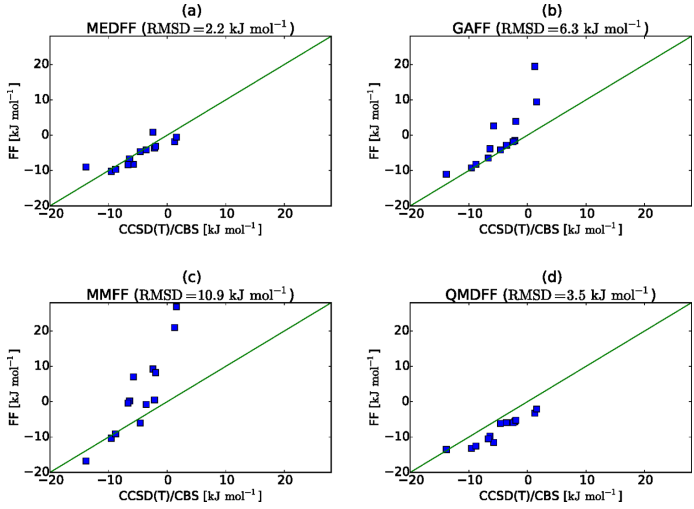



**Figure 14.** Performance of several force fields in reproducing CCSD(T)/CBS interaction energies for the dispersion-dominated complexes of the hsg data set.

complexes should be noted. Here we consider a complex to be dispersion-dominated if $\frac{E^{SAPT}_{dispersion}}{E^{SAPT}_{electrostatics}} \geq 1.7$, a definition corresponding to the classification used for the S66 data set.[75]

**4.2. Second Virial Coefficients.** To further validate our force field, we used it to compute second virial coefficients for some small alkanes and alkenes. Existing force fields can successfully predict this quantity, but such force fields are based on experimental data,[103] on first-principles calculations on alkane and alkene dimers,[23] or on both.[104] Here, most force-field parameters for these systems will be derived solely from monomer computations on alkanes and alkenes. The three interaction parameters retain the values that were determined earlier without fitting to calculations for the dimers considered here, making this a proper validation of our method. The second virial coefficient describes the nonideality of a gas in a first-order approximation. They form an interesting test case for our force field because experimental results are widely available and they only depend on the interaction between two molecules. The second virial coefficient $B_2$ can be computed as follows:[105]

$$B_2 = 2\pi \int (1 - \langle \exp[-\beta U_{inter}(r)] \rangle) r^2 \, dr \tag{28}$$

In this formula $r$ is the distance between the centers of mass of the two molecules and $\beta = \frac{1}{k_B T}$. The brackets $\langle ... \rangle$ indicate an average over all possible configurations and orientations of the molecules. This average is weighted with a Boltzmann factor of the *intramolecular* energies of the configurations. To sample the molecular configurations with these weights, we performed MD simulations of an isolated molecule in the NVT ensemble for 5 ns with a time step of 0.5 fs. Where necessary, the dihedral force constants are rescaled in order to correctly reproduce the energy barriers obtained from a dihedral scan using the same ab initio level of theory. For instance in butane this is crucial to correctly sample gauche and trans states. The resulting intramolecular parameters are provided in the Supporting Information. For noncovalent interactions, MEDFF is used. Note that we neglect noncovalent interactions between 1−2 and 1−3 neighbors. To compute the second virial coefficient, two random configurations were selected from the MD run and one of the monomers was randomly rotated. This procedure enables the computation of $B_2$ by numerical integration over $r$ with 1.5 Å $\leq r \leq$ 40 Å in steps of 0.02 Å. At each value of $r$, the orientational average was obtained by calculating the interaction energy for 100 different randomly generated orientations. The whole procedure was repeated 1000 times with different randomly selected configurations, in order to properly compute the average over all possible configurations.

As mentioned, our force field is expected to work well for molecules where polarization plays a minor role. Thus, for a series of alkanes and alkenes we expect good performance. We limit the simulations to methane, ethane, propane, butane, ethene, and propene, as for longer chains a proper intramolecular force field is strictly necessary to correctly sample molecular configurations. In general the agreement with experiment is rather good, as shown in Figure 15. The second virial coefficient is however slightly underestimated for the larger molecules propene, propane, and butane. We stress again that no dimer calculations for these molecules have to be performed to construct our force field. All necessary parameters come from monomer calculations with addition of a few interaction parameters obtained by fitting to reference data from the S66x8 database (accidentally ethene is present in the S66 database). This is in sharp contrast to most ab initio force fields where the parameters of a certain molecule are fitted to reproduce interaction energies for dimers containing that specific molecule.

**4.3. Liquid Properties.** As a final validation, we applied the force field to compute properties of liquids. Because our force



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161−179

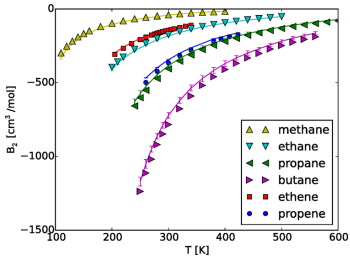



**Figure 15.** Comparison of second virial coefficients from our force field (symbols) with experiment (solid line). Experimental data for alkanes from ref 106, for propene from refs 107 and 108, and for ethene from ref 109.

field is derived entirely based on ab initio computations in the gas phase, condensed-phase simulations are a stringent test to validate the assumptions made in the construction of the force field.

*4.3.1. Computational Details.* We performed molecular dynamics simulations in the NPT-ensemble with our in-house code YAFF.[110] We used a Nosé–Hoover chain thermostat[111,112] with a time constant of 100 fs and chain length 3, in combination with a Martyna–Tobias–Klein barostat[112,113] with time constant 1000 fs. To treat the long-range electrostatic terms correctly, the Ewald summation[38,39] was used, while other nonbonded interactions were cut off at 12 Å using a third-order polynomial to ensure smoothness of the potential and forces. The penetration contribution to the electrostatic interaction (the second term in eq 10) is also cut off at 12 Å. This is justified by the exponential decay of this term, which is clear from eq 11. All simulations were carried out with 100 molecules in a cubic box and a time step of 0.5 fs. Every MD run lasted for 1 ns, of which the last 800 ps is used to collect data. The same force field as for the computation of virial coefficients, with an intramolecular force field obtained using QuickFF, was used.

The following quantities were computed based on the NPT simulation runs. The density $\rho$ follows trivially from $\rho = \frac{M}{V}$, with $M$ the mass of the system and $V$ the volume. The enthalpy of vaporization $\Delta H_{\mathrm{vap}}$ is approximated as

$$\Delta H_{\mathrm{vap}} \approx \langle E_{\mathrm{pot}}^{\mathrm{gas}} \rangle + k_{\mathrm{B}}T - \langle E_{\mathrm{pot}}^{\mathrm{liq}} \rangle \tag{29}$$

where $E_{\mathrm{pot}}^{\mathrm{gas}}$ is obtained from a 100 ps NVT simulation of a single molecule.

*4.3.2. Results.* In line with calculations of the second virial coefficients, we performed MD simulations of methane, ethane, propane, butane, ethene, and propene. We study liquid properties just below the boiling point of these molecules at a pressure of 1 atm. The experimental data for methane, ethane, propane, and butane are taken from ref 114, for ethene from ref 115, and for propene from ref 116. Results for the density and enthalpy of vaporization are shown in Table 6. In general the agreement with experiment is not very good, and we cannot claim that our force field predicts liquid properties that are quantitatively correct. The densities are systematically overestimated with errors ranging from +11% for ethane up to +20% for ethene. The enthalpies of vaporization are also

**Table 6. Comparison of Simulated and Experimental Liquid Density $\rho$ and Enthalpy of Vaporization $\Delta H_{\mathrm{vap}}$**

| | | $\rho$ [g l$^{-1}$] | | $\Delta H_{\mathrm{vap}}$ [kJ mol$^{-1}$] | |
|---|---|---|---|---|---|
| molecule | $T$ [K] | expt | simulation | expt | simulation |
| methane | 111.6 | 422.8 | 502.5 | 8.2 | 9.2 |
| ethane | 184.5 | 544.6 | 618.2 | 14.7 | 16.4 |
| propane | 231.1 | 581.2 | 684.1 | 18.8 | 22.8 |
| butane | 272.6 | 601.1 | 713.8 | 22.4 | 28.3 |
| ethene | 169.2 | 568.0 | 673.1 | 13.5 | 16.1 |
| propene | 225.4 | 609.1 | 674.0 | 18.5 | 18.8 |

overestimated for all molecules we studied, with errors ranging from +12% for ethane up to +23% for butane.

*4.3.3. Discussion.* In the condensed phase, many-body effects will have a much larger impact than in the gas phase. With our pairwise-additive force field, it is impossible to properly describe many-body effects such as induction or Axilrod–Teller–Muto dispersion. Recently, it was shown how pairwise-additive force fields can be extended with explicit three-body terms for dispersion and exchange.[117] In the same paper, it was shown that neglect of these terms in ab initio derived force fields can have a large impact on predicting quantities such as density and heat of vaporization. Similar conclusions have been presented for liquid argon,[118] where the two-body force field overestimates densities with an error of 12%, while the inclusion of three-body corrections leads to good agreement with experimental results. Indeed, the necessity of many-body terms in general has been discussed before.[119,120]

In force fields based on experimental data, all these effects are included *implicitly* in the parameters. We illustrate this for the case of butane with the optimized potential for liquid simulations all atom (OPLS-AA) force field.[121] The parameters of the intermolecular potential are determined to correctly reproduce experimental liquid densities, heats of vaporization, and radial distribution functions. For butane at the boiling point, this results in an error for the density of only −0.5% and only +1.4% for the heat of vaporization.[121] These errors are considerably smaller than errors obtained with MEDFF, +19.0% and +24.8% for liquid density and heat of vaporization, respectively. In Figure 16, however, we show that the OPLS-AA

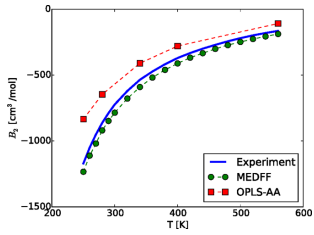



**Figure 16.** Second virial coefficient of butane. Experimental data from ref 106; OPLS-AA results from ref 122.

force field fails to correctly reproduce the second virial coefficient of butane.[122] As its name suggests, OPLS-AA is not well suited to compute gas-phase properties. The fact that MEDFF performs better at reproducing the second virial coefficient than the empirical OPLS-AA force field indicates indeed that the empirical OPLS-AA force field *implicitly*



DOI: 10.1021/acs.jctc.6b00969
J. Chem. Theory Comput. 2017, 13, 161–179

contains many-body effects (of critical importance in the condensed phase) in its parameters. Similar conclusions are obtained for the TraPPE force field for *n*-alkanes.[8]

We elaborate on this point by comparing our force field MEDFF with a two-body force field derived from SAPT.[117] For methane under similar conditions, the predicted relative errors on density and heat of vaporization there are +14% and +26% respectively. By including explicit three-body terms, the errors are reduced to +8% and +13% respectively. Omission of repulsive three-body terms seems to result in an overestimation of liquid densities and of enthalpies of vaporization. This observation provides further evidence that the errors in our current force field are to some extent due to its pairwise-additive character, and that these errors can be made significantly smaller by including explicit three-body terms in the force field. Another approach to include many-body dispersion effects is to map the valence-electron response onto a set of atom-centered quantum harmonic oscillators interacting through a dipole potential.[123−125] In this work, such effects are not taken into account yet, but this will be investigated in future extensions of the proposed method.

To obtain better results while maintaining the pairwise-additivity of our force field, we studied the following approach. Instead of deriving the MBIS parameters from the electron density in the gas phase, we derive them from the electron density from an ab initio calculation in the condensed phase. This offers an interesting way to *implicitly* include some many-body effects missing in the force field. We tested this approach for the case of methane. The resulting density and enthalpy of vaporization are indeed in closer agreement with experiment, compared to values obtained using MBIS parameters from gas-phase methane. The error is however still considerable, possibly because the interaction parameters have been derived using gas-phase interaction energies. In other words, not all many-body effects will be included using the suggested procedure. Another explanation is the limited accuracy of the two-body potential: after all MEDFF gives very good, but not perfect agreement with the second virial coefficient of methane. More details concerning this remark are given in section 6 of the Supporting Information.

## 5. SUMMARY AND CONCLUSIONS

We present a procedure to derive pairwise-additive noncovalent force fields based on monomer electron densities. Our approach differs from existing methods because we avoid ill-conditioned fitting of parameters, we employ physically motivated energy terms, and we are able to use CCSD(T)/CBS data as a reference, because a limited number of data points is sufficient to fit the three interaction parameters. Similar approaches based on universal parameters for non-covalent interactions have been presented before, but they usually involve more parameters. Furthermore, our model is based on a distributed model of monomer electron densities, which allows inclusion of the electrostatic penetration effect in an elegant and simple manner. By comparison with a SAPT decomposition of interaction energies for the S66x8 data set, we show that all terms of our force field correctly reproduce the reference data. For the exchange-repulsion and dispersion contribution, the associated interaction parameter values are shown to be universal, i.e., largely independent of which specific dimer is considered. For the induction term, the same holds only for the dispersion-dominated dimers. For complexes not dominated by dispersion (for example, hydrogen-bonded dimers) it is necessary to determine the induction parameter on a case-specific basis, so it takes on a different value for each complex studied. Next, the parameters for this pairwise-additive force field were refined by combining information from a SAPT energy decomposition with CCSD(T)/CBS interaction energies. The resulting force field is capable of accurately describing the dissociation curves that were used to fit the parameters.

As external validation, we first tested our force field on the hsg data set, consisting of dimers extracted from an HIV-II protease crystal structure with a bound ligand (indinavir). Again, our force field performs well for the dispersion-dominated complexes. Furthermore, experimental second virial coefficients for small alkanes and alkenes were successfully predicted. This is a remarkable success as the interaction parameters are not fitted specifically for these molecules and at no point is experimental input needed. Future extensions are needed to describe condensed-phase properties and hydrogen-bonded complexes. These extensions will cover many-body dispersion and exchange-repulsion terms, as well as a proper nonadditive polarization model to describe induction interactions.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00969.

Geometries, force-field parameters, and interaction energies (ZIP)

Analytical expression derivations, method review, scatter plots, and discussion (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: toon.verstraelen@ugent.be.

### ORCID

Toon Verstraelen: 0000-0001-9288-5608



### Funding

We acknowledge the Foundation of Scientific Research—Flanders (FWO), the Research Board of Ghent University (BOF), and BELSPO in the frame of IAP/7/05 for their financial support.


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure). We thank Stefan Grimme for providing us the results obtained with QMDFF.

## ■ REFERENCES


(1) Joos, L.; Swisher, J. A.; Smit, B. Molecular Simulation Study of the Competitive Adsorption of H2O and CO2 in Zeolite 13X. *Langmuir* **2013**, *29*, 15936−15942.

(2) Yot, P. G.; Boudene, Z.; Macia, J.; Granier, D.; Vanduyfhuys, L.; Verstraelen, T.; Van Speybroeck, V.; Devic, T.; Serre, C.; Férey, G.; Stock, N.; Maurin, G. Metal-organic frameworks as potential shock absorbers: the case of the highly flexible MIL-53(Al). *Chem. Commun.* **2014**, *50*, 9462−9464.

(3) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural




DOI: 10.1021/acs.jctc.6b00969
*J. Chem. Theory Comput.* 2017, 13, 161−179

Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(4) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.* **1984**, *106*, 6638−6646.
(5) Allinger, N.; Yuh, Y.; Lii, J. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.
(6) Rappé, A.; Casewit, C.; Colwell, K.; Goddard, W.; Skiff, W. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(7) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(8) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.
(9) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhys. Lett.* **1994**, *26*, 583.
(10) Brommer, P.; Gähler, F. Potfit: effective potentials from *ab initio* data. *Modell. Simul. Mater. Sci. Eng.* **2007**, *15*, 295.
(11) Li, A. H.-T.; Chao, S. D.; Chang, C.-C. Determination of a silane intermolecular force field potential model from an *ab initio* calculation. *Phys. Rev. A: At., Mol., Opt. Phys.* **2010**, *82*, 062520.
(12) Transtrum, M. K.; Machta, B. B.; Brown, K. S.; Daniels, B. C.; Myers, C. R.; Sethna, J. P. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *J. Chem. Phys.* **2015**, *143*, 010901.
(13) Spackman, M. A. The use of the promolecular charge density to approximate the penetration contribution to intermolecular electrostatic energies. *Chem. Phys. Lett.* **2006**, *418*, 158−162.
(14) 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.
(15) 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.
(16) Riley, K. E.; Platts, J. A.; Řezáč, J.; Hobza, P.; Hill, J. G. Assessment of the Performance of MP2 and MP2 Variants for the Treatment of Noncovalent Interactions. *J. Phys. Chem. A* **2012**, *116*, 4159−4169.
(17) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic Structure Theory*; John Wiley & Sons, LTD: Chichester, 2000.
(18) Grimme, S. A. General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 4497−4514.
(19) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chem. Rev.* **1994**, *94*, 1887−1930.
(20) Yu, K.; McDaniel, J. G.; Schmidt, J. R. Physically Motivated, Robust, ab Initio Force Fields for CO2 and N2. *J. Phys. Chem. B* **2011**, *115*, 10054−10063.
(21) McDaniel, J. G.; Yu, K.; Schmidt, J. R. Ab Initio, Physically Motivated Force Fields for CO2 Adsorption in Zeolitic Imidazolate Frameworks. *J. Phys. Chem. C* **2012**, *116*, 1892−1903.
(22) McDaniel, J. G.; Schmidt, J. R. Robust, Transferable, and Physically Motivated Force Fields for Gas Adsorption in Functionalized Zeolitic Imidazolate Frameworks. *J. Phys. Chem. C* **2012**, *116*, 14031−14039.
(23) McDaniel, J. G.; Schmidt, J. Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory. *J. Phys. Chem. A* **2013**, *117*, 2053−2066.
(24) 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.
(25) 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.
(26) Tafipolsky, M.; Ansorg, K. Toward a Physically Motivated Force Field: Hydrogen Bond Directionality from a Symmetry-Adapted Perturbation Theory Perspective. *J. Chem. Theory Comput.* **2016**, *12*, 1267−1279.
(27) Wang, Q.; Rackers, J. A.; He, C.; Qi, R.; Narth, C.; Lagardere, L.; Gresh, N.; Ponder, J. W.; Piquemal, J.-P.; Ren, P. General Model for Treating Short-Range Electrostatic Penetration in a Molecular Mechanics Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 2609−2618.
(28) Qi, R.; Wang, Q.; Ren, P. General van der Waals potential for common organic molecules. *Bioorg. Med. Chem.* **2016**, *24*, 4911−4919.
(29) Wu, Q.; Ayers, P. W.; Zhang, Y. Density-based energy decomposition analysis for intermolecular interactions with variationally determined intermediate state energies. *J. Chem. Phys.* **2009**, *131*, 164112.
(30) Zhou, N.; Lu, Z.; Wu, Q.; Zhang, Y. Improved parameterization of interatomic potentials for rare gas dimers with density-based energy decomposition analysis. *J. Chem. Phys.* **2014**, *140*, 214117.
(31) Lu, Z.; Zhou, N.; Wu, Q.; Zhang, Y. Directional Dependence of Hydrogen Bonds: A Density-Based Energy Decomposition Analysis and Its Implications on Force Field Development. *J. Chem. Theory Comput.* **2011**, *7*, 4038−4049.
(32) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.
(33) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *J. Chem. Theory Comput.* **2016**, *12*, 3894−3912.
(34) Wang, W.-P.; Parr, R. G. Statistical atomic models with piecewise exponentially decaying electron densities. *Phys. Rev. A: At., Mol., Opt. Phys.* **1977**, *16*, 891−902.
(35) Wang, W.-P. Fixed-shell statistical atomic models with piecewise exponentially decaying electron densities. *Phys. Rev. A: At., Mol., Opt. Phys.* **1982**, *25*, 2901−2912.
(36) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(37) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(38) Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys. (Berlin, Ger.)* **1921**, *369*, 253−287.
(39) de Leeuw, S. W.; Perram, J. W.; Smith, E. R. Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants. *Proc. R. Soc. London, Ser. A* **1980**, *373*, 27−56.
(40) Hockney, R.; Eastwood, J. *Computer Simulation Using Particles*; CRC Press: 1988.
(41) Greengard, L.; Rokhlin, V. A Fast Algorithm for Particle Simulations. *J. Comput. Phys.* **1997**, *135*, 280−292.
(42) Tafipolsky, M.; Engels, B. Accurate Intermolecular Potentials with Physically Grounded Electrostatics. *J. Chem. Theory Comput.* **2011**, *7*, 1791−1803.
(43) Piquemal, J.-P.; Cisneros, G. A.; Reinhardt, P.; Gresh, N.; Darden, T. A. Towards a force field based on density fitting. *J. Chem. Phys.* **2006**, *124*, 104101.
(44) 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.
(45) Stone, A. *The Theory of Intermolecular Forces*, 2nd ed.; Oxford University Press: 2013; Chapter 6.
(46) 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.



DOI: 10.1021/acs.jctc.6b00969
*J. Chem. Theory Comput.* 2017, 13, 161−179

(47) Kita, S.; Noda, K.; Inouye, H. Repulsive potentials for Cl−R and Br−R (R = He, Ne, and Ar) derived from beam experiments. *J. Chem. Phys.* **1976**, *64*, 3446−3449.
(48) Wheatley, R. J.; Price, S. L. An overlap model for estimating the anisotropy of repulsion. *Mol. Phys.* **1990**, *69*, 507−533.
(49) 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.
(50) 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.
(51) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.
(52) Hirshfeld, F. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129−138.
(53) Starkschall, G.; Gordon, R. G. Calculation of Coefficients in the Power Series Expansion of the Long-Range Dispersion Force between Atoms. *J. Chem. Phys.* **1972**, *56*, 2801−2806.
(54) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(55) Misquitta, A. J.; Stone, A. J. Accurate Induction Energies for Small Organic Molecules: 1. Theory. *J. Chem. Theory Comput.* **2008**, *4*, 7−18.
(56) 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.
(57) Dick, B. G.; Overhauser, A. W. Theory of the Dielectric Constants of Alkali Halide Crystals. *Phys. Rev.* **1958**, *112*, 90−103.
(58) Schröder, K.-P.; Sauer, J. Potential Functions for Silica and Zeolite Catalysts Based on ab Initio Calculations. 3. A Shell Model Ion Pair Potential for Silica and Aluminosilicates. *J. Phys. Chem.* **1996**, *100*, 11043−11049.
(59) Lewis, G. V.; Catlow, C. R. A. Potential models for ionic oxides. *J. Phys. C: Solid State Phys.* **1985**, *18*, 1149.
(60) Schmollngruber, M.; Lesch, V.; Schroder, C.; Heuer, A.; Steinhauser, O. Comparing induced point-dipoles and Drude oscillators. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14297−14306.
(61) Huang, L.; Roux, B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *J. Chem. Theory Comput.* **2013**, *9*, 3543−3556.
(62) Savelyev, A.; MacKerell, A. D. All-atom polarizable force field for DNA based on the classical drude oscillator model. *J. Comput. Chem.* **2014**, *35*, 1219−1239.
(63) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341−350.
(64) 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.
(65) Elking, D.; Darden, T.; Woods, R. J. Gaussian induced dipole polarization model. *J. Comput. Chem.* **2007**, *28*, 1261−1274.
(66) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143−3161.
(67) Verstraelen, T.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. ACKS2: Atom-condensed Kohn-Sham DFT approximated to second order. *J. Chem. Phys.* **2013**, *138*, 074108.
(68) Verstraelen, T.; Vandenbrande, S.; Ayers, P. W. Direct computation of parameters for accurate polarizable force fields. *J. Chem. Phys.* **2014**, *141*, 194114.
(69) Murrell, J. N.; Randic, M.; Williams, D. R. The Theory of Intermolecular Forces in the Region of Small Orbital Overlap. *Proc. R. Soc. London, Ser. A* **1965**, *284*, 566−581.
(70) Gresh, N.; Claverie, P.; Pullman, A. Computations of intermolecular interactions: Expansion of a charge-transfer energy contribution in the framework of an additive procedure. Applications to hydrogen-bonded systems. *Int. J. Quantum Chem.* **1982**, *22*, 199−215.
(71) Gresh, N.; Claverie, P.; Pullman, A. Intermolecular interactions: Elaboration on an additive procedure including an explicit charge-transfer contribution. *Int. J. Quantum Chem.* **1986**, *29*, 101−118.
(72) Hagberg, D.; Karlström, G.; Roos, B. O.; Gagliardi, L. The Coordination of Uranyl in Water: A Combined Quantum Chemical and Molecular Simulation Study. *J. Am. Chem. Soc.* **2005**, *127*, 14250−14256.
(73) Kumar, R.; Wang, F.-F.; Jenness, G. R.; Jordan, K. D. A second generation distributed point polarizable water model. *J. Chem. Phys.* **2010**, *132*, 014309.
(74) Stone, A. J.; Misquitta, A. J. Charge-transfer in Symmetry-Adapted Perturbation Theory. *Chem. Phys. Lett.* **2009**, *473*, 201−205.
(75) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.
(76) Morokuma, K.; Kitaura, K. *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Springer US: New York, 1981; pp 215−242.
(77) Bagus, P. S.; Hermann, K.; Bauschlicher, C. W. A new analysis of charge transfer and polarization for ligand-metal bonding: Model studies of Al4CO and Al4NH3. *J. Chem. Phys.* **1984**, *80*, 4378−4386.
(78) Li, A.; Muddana, H. S.; Gilson, M. K. Quantum Mechanical Calculation of Noncovalent Interactions: A Large-Scale Evaluation of PMx, DFT, and SAPT Approaches. *J. Chem. Theory Comput.* **2014**, *10*, 1563−1575.
(79) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. Psi4: an open-source ab initio electronic structure program. *WIREs Comput. Mol. Sci.* **2012**, *2*, 556−565.
(80) Hohenstein, E. G.; Sherrill, C. D. Density fitting of intramonomer correlation effects in symmetry-adapted perturbation theory. *J. Chem. Phys.* **2010**, *133*, 014101.
(81) Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *WIREs Comput. Mol. Sci.* **2012**, *2*, 304−326.
(82) Hohenstein, E. G.; Sherrill, C. D. Efficient evaluation of triple excitations in symmetry-adapted perturbation theory via second-order Møller-Plesset perturbation theory natural orbitals. *J. Chem. Phys.* **2010**, *133*, 104107.
(83) 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.
(84) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N2, and H2O. *Chem. Phys. Lett.* **1998**, *286*, 243−252.
(85) Patkowski, K.; Korona, T.; Jeziorski, B. Convergence behavior of the symmetry-adapted perturbation theory for states submerged in Pauli forbidden continuum. *J. Chem. Phys.* **2001**, *115*, 1137−1152.
(86) Szalewicz, K.; Patkowski, K.; Jeziorski, B. Intermolecular Interactions via Perturbation Theory: From Diatoms to Biomolecules. *Struct. Bonding* **2005**, *116*, 43−117.
(87) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.;



DOI: 10.1021/acs.jctc.6b00969
*J. Chem. Theory Comput.* 2017, 13, 161−179

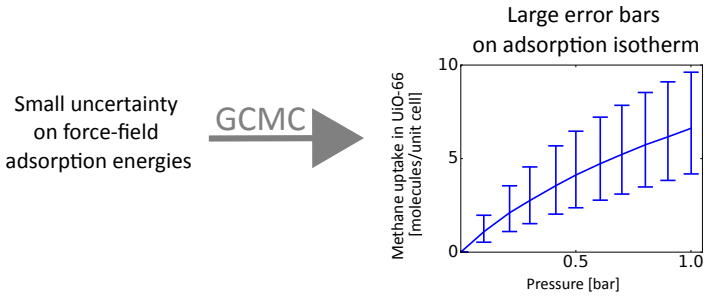
Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09 Revision D.01*; Gaussian Inc.: Wallingford, CT, 2013.
(88) Verstraelen, T.; Boguslawski, K.; Tecmer, P.; Heidar-Zadeh, F.; Chan, M.; Kim, T. D.; Zhao, Y.; Vandenbrande, S.; Yang, D.; González-Espinoza, C. E.; Limacher, P. A.; Berrocal, D.; Malek, A.; Ayers, P. W. *HORTON 2.0.0*; http://theochem.github.com/horton/, 2015.
(89) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(90) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.
(91) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007**, *126*, 144111.
(92) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.
(93) 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.
(94) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466−3470.
(95) Waterfall, J. J.; Casey, F. P.; Gutenkunst, R. N.; Brown, K. S.; Myers, C. R.; Brouwer, P. W.; Elser, V.; Sethna, J. P. Sloppy-Model Universality Class and the Vandermonde Matrix. *Phys. Rev. Lett.* **2006**, *97*, 150601.
(96) Gutenkunst, R. N.; Waterfall, J. J.; Casey, F. P.; Brown, K. S.; Myers, C. R.; Sethna, J. P. Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLoS Comput. Biol.* **2007**, *3*, 1−8.
(97) Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*; Springer-Verlag New York, Inc.: Secaucus, NJ, USA, 2006.
(98) Petzold, V.; Bligaard, T.; Jacobsen, K. Construction of New Electronic Density Functionals with Error Estimation Through Fitting. *Top. Catal.* **2012**, *55*, 402−417.
(99) Cailliez, F.; Pernot, P. Statistical approaches to forcefield calibration and prediction uncertainty in molecular simulation. *J. Chem. Phys.* **2011**, *134*, 054124.
(100) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, K. M. Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-Ligand Complexes. *J. Chem. Theory Comput.* **2011**, *7*, 790−797.
(101) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(102) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720−729.
(103) Rodríguez, A. L.; Vega, C.; Freire, J. J.; Lago, S. Potential parameters of methyl and methylene obtained from second virial coefficients of n-alkanes. *Mol. Phys.* **1991**, *73*, 691−701.
(104) Nagy, J., Jr.; Smith, V. H.; Weaver, D. F. Calculation of second virial coefficients of alkanes with the MM2 and MM3 force fields. *Mol. Phys.* **1994**, *81*, 1039−1047.
(105) McQuarrie, D. *Statistical Mechanics*; University Science Books: Sausalito, 2000.
(106) Dymond, J.; Cholinski, J.; Szafranski, A.; Wyrzykowska-Stankiewicz, D. Second virial coefficients for N-alkanes; recommendations and predictions. *Fluid Phase Equilib.* **1986**, *27*, 1−13.
(107) Glos, S.; Kleinrahm, R.; Wagner, W. Measurement of the ($p$, $\rho$, $T$) relation of propane, propylene, n-butane, and isobutane in the temperature range from (95 to 340) K at pressures up to 12 MPa using an accurate two-sinker densimeter. *J. Chem. Thermodyn.* **2004**, *36*, 1037−1059.
(108) Warowny, W.; Wielopolski, P.; Stecki, J. Compressibility factors and virial coefficients for propane, propene and their mixtures by the burnett method. *Phys. A* **1978**, *91*, 73−87.
(109) Nowak, P.; Kleinrahm, R.; Wagner, W. Measurement and correlation of the ($p$, $\rho$, $T$) relation of ethylene I. The homogeneous gaseous and liquid regions in the temperature range from 105K to 340K at pressures up to 12 MPa. *J. Chem. Thermodyn.* **1996**, *28*, 1423−1439.
(110) Vanduyfhuys, L.; Vandenbrande, S.; Verstraelen, T. *YAFF, Yet Another Force Field*; http://molmod.ugent.be/software/, accessed 09/30/2016.
(111) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **1992**, *97*, 2635−2643.
(112) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.* **1996**, *87*, 1117−1157.
(113) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994**, *101*, 4177−4189.
(114) Younglove, B. A.; Ely, J. F. Thermophysical Properties of Fluids. II. Methane, Ethane, Propane, Isobutane, and Normal Butane. *J. Phys. Chem. Ref. Data* **1987**, *16*, 577−798.
(115) Smukala, J.; Span, R.; Wagner, W. New Equation of State for Ethylene Covering the Fluid Region for Temperatures From the Melting Line to 450K at Pressures up to 300 MPa. *J. Phys. Chem. Ref. Data* **2000**, *29*, 1053−1121.
(116) Lemmon, E.; McLinden, M.; Friend, D. Thermophysical Properties of Fluid Systems. In *NIST Chemistry WebBook*, 1st (revised) ed.; National Institute of Standards and Technology: Gaithersburg, MD, 1996.
(117) 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.
(118) Desgranges, C.; Delhommelle, J. Many-Body Effects on the Thermodynamics of Fluids, Mixtures, and Nanoconfined Fluids. *J. Chem. Theory Comput.* **2015**, *11*, 5401−5414.
(119) DiStasio, R. A., Jr.; Gobre, V. V.; Tkatchenko, A. Many-body van der Waals interactions in molecules and condensed matter. *J. Phys.: Condens. Matter* **2014**, *26*, 213202.
(120) 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.
(121) Kaminski, G.; Duffy, E. M.; Matsui, T.; Jorgensen, W. L. Free Energies of Hydration and Pure Liquid Properties of Hydrocarbons from the OPLS All-Atom Model. *J. Phys. Chem.* **1994**, *98*, 13077−13082.
(122) Nagy, J.; Weaver, D. F.; Smith, V. H., Jr. A Comprehensive Study of Alkane Nonbonded Empirical Force Fields. Suggestions for Improved Parameter Sets. *J. Phys. Chem.* **1995**, *99*, 8058−8065.
(123) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.
(124) Tkatchenko, A.; Ambrosetti, A.; DiStasio, R. A., Jr. Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem. *J. Chem. Phys.* **2013**, *138*, 074106.
(125) Ambrosetti, A.; Reilly, A. M.; DiStasio, R. A., Jr.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.* **2014**, *140*, 18A508.



DOI: 10.1021/acs.jctc.6b00969
*J. Chem. Theory Comput.* 2017, 13, 161−179

Paper III

# Methane Adsorption in Zr-Based MOFs: Comparison and Critical Evaluation of Force Fields




S. Vandenbrande, T. Verstraelen, J.J. Gutiérrez-Sevillano, M. Waroquier, V. Van Speybroeck

*J. Phys. Chem. C*, **2017**, *121* (45), 25309-25322

S. Vandenbrande performed the simulations, analyzed the results, and wrote the manuscript.

THE JOURNAL OF PHYSICAL CHEMISTRY C

Article

Cite This: J. Phys. Chem. C 2017, 121, 25309-25322

pubs.acs.org/JPCC

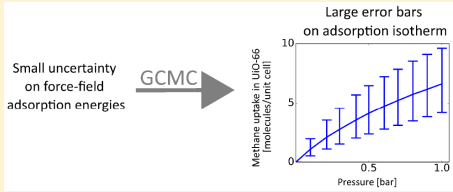
# Methane Adsorption in Zr-Based MOFs: Comparison and Critical Evaluation of Force Fields

Steven Vandenbrande, Toon Verstraelen, Juan José Gutiérrez-Sevillano, Michel Waroquier, and Veronique Van Speybroeck*

Center for Molecular Modeling (CMM), Ghent University (Member of the QCMM Ghent−Brussels Alliance), Technologiepark 903, B9000 Ghent, Belgium

Supporting Information

**ABSTRACT:** The search for nanoporous materials that are highly performing for gas storage and separation is one of the contemporary challenges in material design. The computational tools to aid these experimental efforts are widely available, and adsorption isotherms are routinely computed for huge sets of (hypothetical) frameworks. Clearly the computational results depend on the interactions between the adsorbed species and the adsorbent, which are commonly described using force fields. In this paper, an extensive comparison and in-depth investigation of several force fields from literature is reported for the case of methane adsorption in the Zr-based Metal−Organic Frameworks UiO-66, UiO-67, DUT-52, NU-1000, and MOF-808. Significant quantitative differences in the computed uptake are observed when comparing different force fields, but most qualitative features are common which suggests some predictive power of the simulations when it comes to these properties. More insight into the host−guest interactions is obtained by benchmarking the force fields with an extensive number of ab initio computed single molecule interaction energies. This analysis at the molecular level reveals that especially ab initio derived force fields perform well in reproducing the ab initio interaction energies. Finally, the high sensitivity of uptake predictions on the underlying potential energy surface is explored.




## 1. INTRODUCTION

For the past two decades, metal−organic frameworks (MOFs) have received considerable attention in scientific literature. This is in part due to the seemingly unlimited number of possible combinations of metal nodes coordinated by organic ligands, with each combination resulting in the unique charateristics of that MOF. The mixing and matching of these building blocks lead to a wide range of chemical environments, topologies, and pore sizes, with practical applications including heterogeneous catalysis[1,2] and gas storage or separation.[3] From the onset, MOFs were mostly used and studied for their gas adsorption properties.[4−6] Also from a computational side, MOFs attracted extensive attention, as evidenced by many high-throughput studies that aim to identify the most promising candidate materials for a given application.[7,8]

Many studies on the computational prediction of adsorption isotherms are available; however, the correct prediction of isotherms is particularly challenging. To compute the adsorption isotherm of a guest molecule in a rigid framework, one can use Grand-Canonical Monte Carlo (GCMC) simulations which are able to unveil the influence of the underlying potential energy surface (PES) on the uptake of the guest molecules. Typically, millions of energy evaluations are required in GCMC simulations, making them intractable with ab initio methods. Given the current limitations on computing power, only force-field simulations are feasible. There is an abundance of force fields in literature that can be used for adsorption simulations (note that we only consider the noncovalent part of the force fields in this study, as covalent terms are not relevant in rigid-framework simulations). An important distinction is the difference between so-called generic force fields on the one hand and ab initio derived force fields on the other hand. In brief, generic force fields are typically fitted to reproduce certain experimental results such as crystal structures, sublimation energies, or fluid properties and are then applied to other systems and/or properties of interest. Parameters of generic force fields are available for many chemical environments, which makes them (in principle) widely applicable.[9] In contrast, ab initio derived force fields are typically more tailored toward one specific application. Previous classification gives only a brief distinction between the two major classes of force fields (generic versus ab initio derived). In reality many force fields were developed which have much more subtle nuances. A more complete description of force fields used in this work is given in section 2.1. For an in-depth review, we refer to the work by Coudert et al.[10]

Hereafter, we particularly highlight some computational studies on gas adsorption which are relevant for this paper, in which methane adsorption in Zr-based MOFs is taken as a case

Received: September 8, 2017
Revised: October 23, 2017
Published: October 24, 2017

DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

The Journal of Physical Chemistry C Article

study. Vasanth Kumar et al. studied $H_2$ and $CH_4$ adsorption in UiO-66, UiO-67, and UiO-68, showing that linker−guest interactions are the main driving force for adsorption.[11] Their study was performed using GCMC simulations with a combination of different generic force fields (Dreiding, UFF, and TraPPE). In contrast to the current work, the obtained results were not validated using experimental adsorption data or ab initio computed adsorption energies. Snurr et al.[12] simulated the effect of missing linker defects on water and $CO_2$ adsorption in UiO-66, using also a combination of different generic force fields (Dreiding, UFF, TraPPE, and TIP4P) and with additional account of framework atomic charges which were derived from ab initio calculations using the REPEAT method. It was shown that defect sites render the material more hydrophilic and that the location of the defects has an appreciable impact on uptake. Düren et al.[13] presented an adsorption study of methane in CuBTC in which direct information from the true potential energy surface (calculated with the DFT/CC[14] ab initio method) was used. For this MOF with coordinatively unsaturated sites, the presented approach provided quantitative agreement with experiment over a wide range of temperatures and pressures. However, it was observed that using other ab initio methods as a reference may provide deviating results, which gives an indication of the sensitiviy of the adsorption isotherm on the underlying potential energy surface. Finally, Schmidt et al.[15] compared a generic force field with an ab initio derived force field in a screening study involving 424 MOFs. Significant differences in $CO_2$ and $CH_4$ gas adsorption isotherms were observed; however, both types of force fields predict a similar ranking (with respect to uptake at a certain pressure and temperature) of the frameworks. The authors also found that the generic force field may benefit from a compensation of errors. Consequently, the physical interpretation of results obtained with generic force fields should be done with caution. Obviously, computational studies of gas adsorption in MOFs have covered a much wider array of frameworks and guest molecules than the few examples mentioned before.

In this work, we study methane adsorption in Zr-based MOFs, which proves an ideal case study for a comparison and evaluation of force fields to predict adsorption isotherms. At first sight it might be regarded as a rather easy problem from a computational point of view for two reasons. First, for the study at hand, the methane guest molecule does not feature a dipole or quadrupole moment, rendering electrostatic interactions with the framework unimportant (as shown in section S3), and polarization effects can be assumed negligible. Second, the frameworks that will be studied here do not feature coordinatively unsaturated sites (note that such sites can be present in defective UiO-66 structures, but those cases are not considered in this work). Such open metal sites can have a dramatic impact on adsorption properties and often require additional force-field terms not included in generic force fields.[16,17] Generic force fields perform better in defect-free materials with fully saturated metal centers. This is convincingly demonstrated in ref 18, where adsorption isotherms have been computed with generic force fields for $CO_2$ and $CH_4$ in dehydroxylated UiO-66. An excellent agreement with experiment is even obtained if Dreiding, UFF, and TraPPE parameters are selectively used for the $CH_4$-MOF and $CO_2$-MOF interactions. In another recent study, it was concluded that generic force fields are suitable for the qualitative prediction of methane adsorption in MOFs in the low-loading regime.[19] In the same paper, it is demonstrated that these force fields are capable of providing detailed molecular-level information, although in some systems such conclusions should be approached with caution.

In this paper, we study methane adsorption in Zr-based MOFs using five different force fields. Three of them are so-called generic force fields while we also included two force fields based on an ab initio description of the potential energy surface. Adsorption isotherms are calculated using GCMC simulations. Insight at the atomic level is gained from single molecule adsorption energies, obtained from density functional theory (DFT) calculations on the periodic lattice. The main goal of this paper is to critically assess which factors contribute to the overall reproduction of the adsorption isotherms, both quantitatively and qualitatively. The paper is organized as follows. section 2.1 provides an elaborate description of the five used force fields. In section 2.2, one reads computational details. The main results of the GCMC simulations and the comparison with ab initio adsorption energies are presented in sections 3.1 and 3.2, respectively. The conclusions are formulated in section 4.

## 2. METHODS

**2.1. Description of the Force Fields.** Many different force fields are available in literature, which can be distinguished by their functional form, parametrization method, and design philosophy. This section discusses these features of the five force fields that will be used in this work.

The first two force fields are so-called generic force fields, based on generic, nonsystem-specific parameters. The first variant will be further designated UFF/TraPPE. The potential energy is a pairwise additive sum of Lennard-Jones potentials:

$$E = \sum_{(i,j)} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$$

where the sum runs over all pairs of atoms where atom $i$ is part of the framework and atom $j$ is part of the guest molecule. The parameters $\sigma_{ij}$ and $\epsilon_{ij}$ are determined using Lorentz−Berthelot mixing rules respectively:

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \quad (1)$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad (2)$$

The atomic parameters $\sigma_i$ and $\epsilon_i$ of the MOF frameworks are taken from the Universal Force Field (UFF)[20] and are in this case uniquely determined by the atomic number of atom $i$. The MOFs we will study (UiO-66, UiO-67, DUT-52, MOF-808, and NU-1000) consist of Zr, O, C, and H atoms, leading to 8 parameters. The UFF parameters have been developed by making use of a combination of ab initio calculations, literature values, and empirical rules. Methane is described using the transferable potential for phase equilibria (TraPPE),[21] where a united-atom model with only one site is used for methane. Because methane is a neutral molecule, this site has zero charge which means that no electrostatic interactions are taken into account. The $\sigma_i$ and $\epsilon_i$ of methane are parametrized to reproduce the experimental critical properties and coexistence densities. The second force field, designated as Dreiding-UFF/TraPPE, is very similar and shares the same functional form and number of parameters. The only difference with the UFF/TraPPE force field is that the framework parameters $\sigma_i$ and $\epsilon_i$ are obtained from the Dreiding force field.[22] Again, parameters are only based on the atomic number of the atom and take modified values that were based on



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

experimental crystal structures and sublimation enthalpies. Unfortunately, no Dreiding parameters are available for the Zr atom and therefore the ones available from UFF are taken. Both the Dreiding-UFF/TraPPE and the UFF/TraPPE force fields have been used before to model methane adsorption in UiO-66.[18]

The third force field, MM3-MBIS, uses generic parameters for the van der Waals contribution combined with system-specific charges to model electrostatic contributions:

$$E = \sum_{(i,j)} \left( \epsilon_{ij} \left[ 1.84 \times 10^5 e^{-12 r_{ij}/\sigma_{ij}} - 2.25 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}} \right) \tag{3}$$

The van der Waals contribution is modeled using the Buckingham potential. The parameters $\sigma_{ij}$ and $\epsilon_{ij}$ are again determined from atomic values using following mixing rules:

$$\sigma_{ij} = \sigma_i + \sigma_j \tag{4}$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{5}$$

Note that $\sigma_{ij}$ is defined as the sum of the atomic radii, rather than as the average of atomic radii (which is the case for the UFF and Dreiding force fields, see eq 1). The definition of $\sigma_{ij}$ as the sum of the atomic radii is in line with the convention used by the authors of the MM3 force field.[23] The atomic parameters in turn are parametrized to reproduce experimental results such as molecular geometry, crystal unit cell parameters, and heat of sublimation.[23−25] For some elements of the periodic table, parameters are obtained by interpolation. Note that for MM3, the values of $\sigma_i$ and $\epsilon_i$ are not uniquely determined by the atomic number of atom $i$, for instance a $sp^3$ hybridized C atom has different parameters than a $sp^2$ hybridized C atom. In the original MM3 work, electrostatic interactions are described using bond dipoles, but for simplicity, we will employ point charges instead. These atomic partial charges are obtained using the minimal basis iterative stockholder (MBIS) partitioning scheme.[26] It should be mentioned that for the nonpolar methane molecule electrostatic interactions have a modest impact, as shown in section S3.

The fourth force field is a purely ab initio based force field introduced by Schmidt et al. and is called SAPTFF.[27,28] The potential energy of this force field is decomposed into terms that correspond to the physically distinct intermolecular interaction energies from SAPT:

$$U_{tot} = E_{exch} + E_{elec} + E_{pol} + E_{disp} + E_{\delta hf} \tag{6}$$

where $E_{exch}$, $E_{elec}$, $E_{pol}$, $E_{disp}$, and $E_{\delta hf}$ correspond to the exchange-repulsion energy, electrostatic energy, polarization energy, dispersion energy, and higher-order contributions to polarization/exchange, respectively. For a methane molecule, the polarizability is taken to be zero and point-charge electrostatic interactions are not included.[15] Taking these simplifications into account, the above expression can be written as

$$U_{tot} = \sum_{(i,j)} \left( [A_{ij}^{elec} + A_{ij}^{exch} + A_{ij}^{ind} + A_{ij}^{\delta hf}] \exp(-B_{ij} r_{ij}) - \sum_{n=6,8,10,12} f_n(B_{ij} r_{ij}) \frac{C_n^{ij}}{r_{ij}^n} \right) \tag{7}$$

The dispersion interactions are damped using the Tang-Toennies damping function $f_n(x) = 1 - e^{-x} \sum_{m=0}^{n} \frac{x^m}{m!}$ and combination rules are as follows:

$$B_{ij} = (B_i + B_j) \frac{B_i B_j}{B_i^2 + B_j^2} \tag{8}$$

$$C_n^{ij} = \sqrt{C_n^i C_n^j} \tag{9}$$

$$|A_{ij}^{type}| = \sqrt{|A_i^{type}||A_j^{type}|} \tag{10}$$

where $A_{ij}^{type}$ is always positive for exchange, always negative for electrostatics and induction, and positive for $\delta$hf if and only if both $A_i^{\delta hf}$ and $A_j^{\delta hf}$ are positive. This functional expression is more complicated than the Lennard-Jones or Buckingham potentials used in the previously discussed generic force fields, but there is also an important difference with respect to the way parameters are determined. The SAPTFF parameters are determined based on monomer properties and dimer interactions, all computed using ab initio calculations. To generate the necessary reference data, cluster models of the frameworks under study are used. Different atom types are introduced of which 11 are necessary to describe the systems investigated in this paper. For each atom type, 9 parameters are introduced which leads to a total of 99 parameters as reported in section S6. SAPTFF has been used before in a screening study, which included methane adsorption in Zr-based MOFs.[15]

The fifth and final force field considered in this work is the monomer electron density based force field (MEDFF), which was proposed by the present authors.[29] It is also based on ab initio data and shares some common features with SAPTFF, although much less parameters need to be fitted to dimer interaction energies. The functional form is as follows:

$$E = \sum_{(i,j)} \left( E_{elst}^{ij} + (U_{exch-rep} - U_{ind}) S^{ij} - f_6(x_{ij}) \frac{C_6^{ij}}{r_{ij}^6} - U_{s8} f_8(x_{ij}) \frac{C_8^{ij}}{r_{ij}^8} \right) \tag{11}$$

The first term, $E_{elst}^{ij}$, represents electrostatic interactions including the penetration effect. The second term, $(U_{exch-rep} - U_{ind})S^{ij}$, describes exchange-repulsion and short-range induction interactions based on the overlap of the monomer electron densities. The final terms, $f_6(x_{ij}) \frac{C_6^{ij}}{r_{ij}^6}$ and $U_{s8} f_8(x_{ij}) \frac{C_8^{ij}}{r_{ij}^8}$, are dispersion terms that are damped at short-range using the Tang-Toennies damping function. Contrary to most other force fields, no atom types are defined in MEDFF. Instead the atomic parameters are obtained directly from an ab initio computed electron density through the MBIS partitioning scheme.[26] For instance, $S^{ij}$ is the overlap of the model electron densities, which can be accurately computed from the MBIS parameters of the atoms involved, as MBIS provides an analytically simple expression for electron density. Next to these atomic parameters, MEDFF also features three so-called interaction parameters: $U_{exch-rep}$, $U_{ind}$, and $U_{s8}$. The values for these parameters can not be obtained from the MBIS partitioning but are fitted to reproduce ab initio interaction energies (both SAPT and CCSD(T) reference data are used in the fitting procedure). In this work, we use the interaction parameters as determined in the original MEDFF paper[29] for a data set of dispersion-dominated complexes. This data set contains dimers of small molecules, such as benzene, ethene, and neopentane. As dispersion interactions



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

are the driving force for methane adsorption, we expect these parameter values to result in a proper description of the systems at hand.

**2.2. Computational Details.** *Monte Carlo Simulations.* Monte Carlo simulations are an important tool for the computational study of adsorption. In this work, adsorption isotherms are calculated using GCMC with the RASPA software package.[30,31] Each run consists of 50000 equilibration cycles and at least 100000 production cycles, where a cycle consists of max(20, $N$) move attempts (with $N$ the current number of adsorbed molecules). The error bars, computed using block averaging, indicate that increasing the number of cycles (to obtain better sampling of the visited parts of phase space) is unlikely to change the results by more than 1%. The applied chemical potential is computed from the fugacity, which is converted from the input pressure using the Peng−Robinson equation of state. Experimental values for the critical temperature, critical pressure, and acentric factor (which are input parameters of the Peng−Robinson equation) of methane are used.[32] Translation, rotation, reinsertion, deletion, and insertion moves are all attempted with a probability of 20%. For the all-atom force fields (MM3-MBIS, SAPTFF, and MEDFF) methane is considered as a rigid molecule with a geometry optimized at the B3LYP[33,34]/aug-cc-pVTZ[35] level-of-theory. For the frameworks UiO-66, UiO-67, and DUT-52, a 2 × 2 × 2 supercell of the conventional unit cell is employed. For NU-1000, a 1 × 1 × 2 supercell is used, while for MOF-808, the conventional unit cell is employed. During all RASPA simulations, the framework geometries are taken from experiment, except for the defective UiO-66 frameworks where the geometries are DFT optimized. In any case, the framework geometry is considered to be rigid during the simulations. The impact of this approximation is limited for the MOFs studied in this work, as we show in section S1. The vdW energy of each atom of the guest molecule and the electrostatic potential in the framework (and derivatives of these quantities) are tabulated on a cubic grid with a spacing of about 0.15 Å. During simulations, a tricubic interpolation scheme is used to obtain the host−guest energy using these grids.[36] Henry coefficients are also calculated with RASPA using 50000 Widom insertions.

The surface area (calculated with RASPA) is obtained geometrically based on the overlap of a helium probe with the framework atoms. Atoms are considered to overlap when they are closer than the minimum of the corresponding Lennard-Jones interaction, where the necessary radii are obtained from the Dreiding and UFF force fields. The available pore volume is calculated by RASPA as the amount of argon that is adsorbed at 87 K and a pressure of $0.5p_0$ using UFF parameters for the framework atoms and the parameters from Maitland et al. for argon.[37] The uptake (in moles) is converted to a volume by multiplying by the molar volume of bulk argon under the same conditions.

*Force-Field Settings.* All Monte Carlo simulations in this work are performed using force fields to evaluate the required energies, and therefore we now briefly discuss some settings related to the force-field energy evaluations. An important consideration for force fields in periodic systems is the truncation of the nonbonded interactions. Because of the slow decay of point-charge electrostatics (present only in MM3-MBIS and MEDFF), we employ the Ewald summation,[38,39] which allows one to exactly compute the electrostatic interaction of a periodic system under tinfoil boundary conditions. Similar techniques exist to exactly compute the contribution of dispersion terms decaying as $r^{-n}$ (where typically $n$ = 6,8,10, ...) but this is not implemented in RASPA. We therefore truncate the van der Waals interactions at a radius of 14 Å but employ tail-corrections to approximate the contributions beyond this cutoff. The influence of this setting is studied in more detail in the section S2. Note that this is the recommended treatment of long-range van der Waals interactions in the TraPPE model.

The MM3-MBIS and MEDFF force fields require MBIS parameters of both the frameworks and the methane guest molecule. For the MOF frameworks, the MBIS parameters are computed from the partitioning of the PBE[40] electron density computed with GPAW.[41,42] These calculations are performed using a real-space uniform grid with spacing of about 0.20 Å, with Γ-point-only sampling of the Brillouin zone, employing the supplied default PAW data set and standard convergence settings for the electronic self-consistent loop. The MBIS parameters for methane are obtained from the B3LYP[33,34]/aug-cc-pVTZ[35] electron density computed using Gaussian 09.[43]

*Ab Initio Adsorption Energies.* To gain more fundamental insight into the adsorption mechanisms that govern the GCMC simulations, we also compared force-field and ab initio computed single molecule adsorption energies. Adsorption energies of methane in UiO-66 were calculated with the PBE[40] functional with modified D3 dispersion corrections with Becke-Johnson damping (D3MBJ),[44] including the three-body contribution.[45] The three-body contribution consists of an Axilrod−Teller−Muto (ATM) term, and therefore, this level-of-theory will be referred to as PBE-D3MBJ-ATM. Periodic DFT calculations are performed using VASP.[46−48] The conventional unit cell of UiO-66 (a cubic cell with sides 20.75 Å containing 4 inorganic bricks) is used, which allows accurate sampling of the Brillouin-zone with the Γ point only. GW PAW potentials are used, the plane-wave basis set kinetic cutoff energy is set to 800 eV, and the electronic self-consistent loop is considered converged as soon as the change in energy drops below $10^{-6}$ eV. D3MBJ corrections together with the three-body term are calculated using the *dftd3* program and are added to the PBE energies.

The Henry coefficient $K_H$ of methane in UiO-66 is calculated using the same level-of-theory as follows. First, a cubic grid with spacing 0.262 Å (80 grid points in each direction, 512000 grid points in total) is constructed. For each grid point, the interaction energy is computed using Dreiding-UFF/TraPPE and if it is more repulsive than 1000 kJ mol$^{-1}$ then this point is not considered for the ab initio calculations as it will not contribute to the Henry coefficient. This allows one to discard 327780 grid points. Of the remaining 184220 grid points, only 2401 points are unique because of the symmetry of the framework. For each of these grid points, several random rotations are considered and the number of rotations for grid point $i$ is indicated as $N_{\mathbf{\Omega}}^i$. At least 5 random points are included in the final set, but additional samples are generated "on the fly" to ensure proper convergence. The Henry coefficient is finally simply calculated by numerical integration over the cubic grid:

$$K_H = \frac{\beta}{\rho}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N_{\mathbf{\Omega}}^i}\sum_{j=1}^{N_{\mathbf{\Omega}}^i}\exp[-\beta E_{\text{ads}}(\mathbf{r}_i, \mathbf{\Omega}_j)] \tag{12}$$

## 3. RESULTS AND DISCUSSION

**3.1. Adsorption Isotherms of Methane in Zr-Based MOFs.** *Comparison with Experiment for Three Isoreticular MOFs.* First of all, we study the adsorption of methane in three isoreticular, defect-free zirconium-based MOFs for which



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

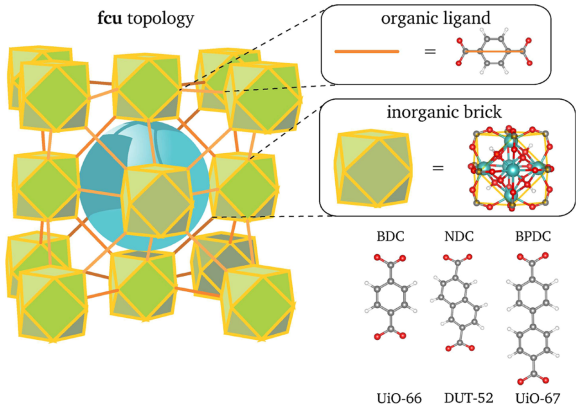



**Figure 1.** Depiction of the **fcu** topology which is shared by UiO-66, DUT-52, and UiO-67. Also the inorganic brick is common between these materials, which differ only in the organic linker. Zirconium atoms are shown in cyan, oxygen atoms in red, carbon atoms in gray, and hydrogen atoms in white. Reprinted from ref 51. Copyright 2016 American Chemical Society.

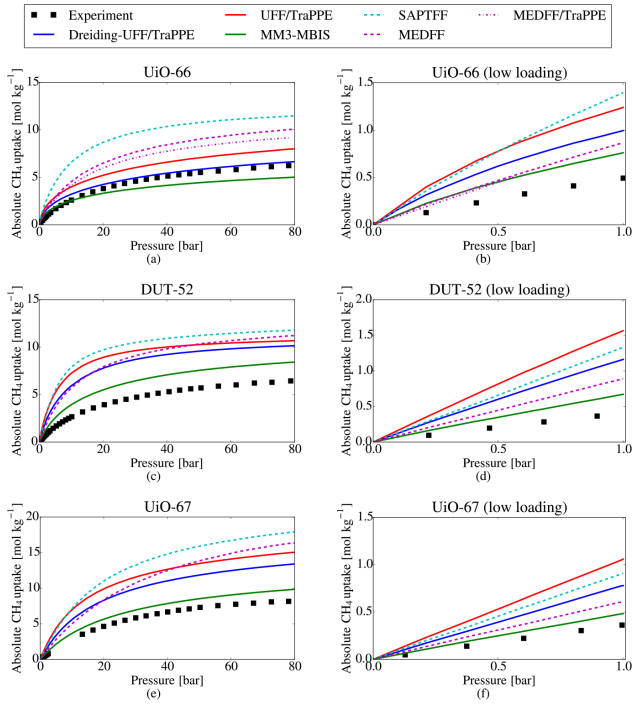



**Figure 2.** Comparison of experimental and simulated gravimetric adsorption isotherms of methane in a series of isoreticular MOFs. Experimental data for (a), (c), (d), (e), and (f) from Cavka et al.,[52] experimental data for (b) from Walton et al.[54]

experimental data is available: UiO-66,[49] UiO-67,[49] and DUT-52,[50] which are illustrated in Figure 1.

All these materials are composed of $Zr_6(\mu_3\text{-}O)_4(\mu_3\text{-}OH)_4$ bricks connected by ditopic organic ligands to form a network



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

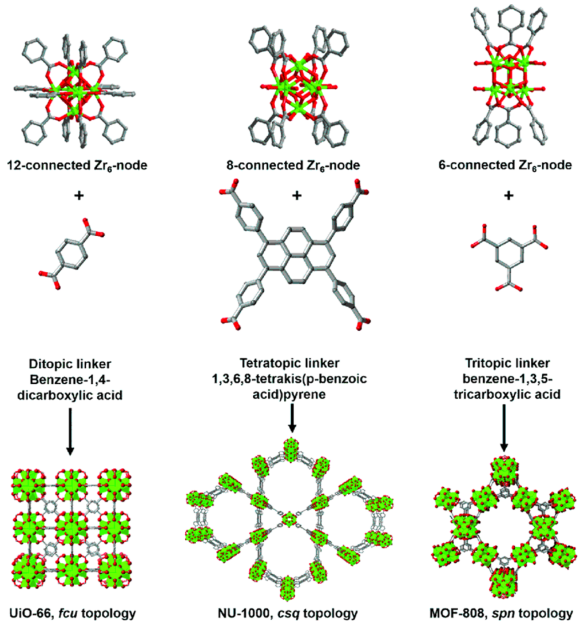



**Figure 3.** Node connectivity, linker and topology for UiO-66, NU-1000, and MOF-808. Reproduced with permission from ref 59. Copyright 2017 Royal Society of Chemistry.

with **fcu** topology. The three MOFs differ only in the organic linker, which is benzene-1,4-dicarboxylate (BDC) in UiO-66, 2,6-naphthalenedicarboxylate (NDC) in the case of DUT-52, and biphenyl-4,4′-dicarboxylate (BPDC) in UiO-67. The influence of the linker on methane adsorption was studied experimentally by Cavka et al.[52] It was found that the maximum $CH_4$ gravimetric loading (i.e., uptake per mass of the framework) is ordered as UiO-66 < DUT-52 < UiO-67, which is the same ordering as for the surface areas and pore volumes of these materials.

We simulated the methane uptake for pressures between 0.1 and 80 bar (the experimental range) using the force fields described in the previous section. All simulations are performed with a rigid framework, where the geometry is obtained from single-crystal X-ray diffraction by Lillerud et al.[53] for UiO-66 and UiO-67 and by Senkovska et al.[50] for DUT-52. The experimentally measured excess uptake is converted to an absolute amount of adsorbed gas as done in the original work reporting the experimental results. The resulting absolute adsorption isotherms are shown in Figure 2 together with the experimental curves. Hereafter, we discuss these isotherms and their agreement with the experiment both at low and higher pressures.

We start with a discussion of methane adsorption in UiO-66 at relatively high pressures (from 30 to 80 bar), which is reported in Figure 2a. As reported in earlier work, the Dreiding-UFF/TraPPE model offers a good agreement with experiment while the UFF/TraPPE model leads to an overestimation of adsorbed molecules.[18] The MM3-MBIS uptake is substantially lower, while both ab initio force fields (MEDFF) and (SAPTFF) predict uptakes substantially higher than the experiment. Perhaps surprisingly, the ab initio force fields are outperformed by a generic force field such as Dreiding-UFF/TraPPE, even though the latter is in no way fitted specifically for the system studied here. It has however been shown before that DFT-derived force fields are not necessarily superior to a generic force field for describing noble gas adsorption in MOFs.[55] A possible explanation for the apparent poor performance of MEDFF and SAPTFF lies in the description of the guest−guest interactions, which play an important role at higher pressures and thus higher loadings. Calculation of the vapor−liquid coexistence curve of bulk methane shows that both SAPTFF and MEDFF overestimate the liquid-phase density, whereas TraPPE offers a very good agreement with the experiment (see section S9). To investigate the influence of the guest−guest interactions on adsorption isotherms at higher pressures, we also performed simulations where the host−guest interactions are described using MEDFF, while the guest−guest interactions are described using TraPPE. The only difference between Dreiding-UFF/TraPPE and MEDFFF/TraPPE is thus in the description of the host−guest interactions. As seen in Figure 2a, the uptake predicted by MEDFF/TraPPE is indeed lower than the uptake predicted by MEDFF; however, changing the guest−guest interactions does not fully resolve the discrepancy with respect to the experiment.

The results presented at higher pressures already show the high sensitivity of the predicted isotherms of the used force field. We now focus on the low loading regime, where uptake is almost exclusively determined by host−guest interactions. In Figure 2b, the methane adsorption in UiO-66 is plotted for pressures lower



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

than 1 bar. Experimental data are taken from Walton et al.[54] because Cavka et al.[52] only report one data point below 1 bar. Note that both experimental sources agree reasonably well for common pressures. Dreiding-UFF/TraPPE reproduces fairly well isotherms at higher pressures but, as all the other force fields, predicts a too large uptake at low pressures. A possible explanation is that the Dreiding-UFF/TraPPE model does not offer a fundamentally proper description of the host−guest interactions, as evidenced by the overestimation at low pressure. At higher pressures, however, this is compensated by deficiencies in the description of guest−guest interactions. In other words, the Dreiding-UFF/TraPPE model might, in this case, benefit from a compensation of errors. Also UFF/TraPPE and SAPTFF largely overestimate the experimental values at low pressure. The MM3-MBIS and MEDFF isotherms are in slightly better agreement with the experiment for pressures below 1 bar but still overestimate the experimental values. The systematic overestimation of all force fields suggests it could be worthwhile to investigate error bars on the experimental data points. Indeed, an ab initio calculation of the Henry constant of methane in UiO-66 is higher than the experimental value (as discussed later in section 3.2 and Table 3). Several pitfalls are present when comparing experimental and computional adsorption,[10] but this discussion will not be pursued in this paper.

To investigate whether similar features hold for the other isoreticular Zr-based MOFs, a similar analysis is performed for DUT-52 and UiO-67 for which also experimental data are available. The results are shown in Figure 2 (panels c and e) for higher pressures. All force fields now predict too high uptakes in this pressure regime. Despite the similarity of these MOFs with UiO-66, the Dreiding-UFF/TraPPE model also considerably overestimates methane uptake at pressures above 30 bar compared to the experiment. A possible explanation might be the incomplete activation of the samples used to measure adsorption or the presence of defects in these samples. A procedure that has been suggested in literature to empirically remedy this problem is the rescaling of the entire adsorption isotherm. The rescaling factor can be determined as the ratio of the theoretical and experimental surface area[56] or as the ratio of the theoretical and experimental pore volume.[57] We have investigated both of these methods in section S11. The analysis reveals that the suggested rescaling procedure does not systematically improve the correspondence with simulated results. It has been noted before that a simple rescaling is rather artificial, as the rescaling factor can depend on the loading.[58]

Thus, far our assessment reveals that there is no systematic trend in the prediction of the adsorption trends among the three isoreticular MOFs. The only observed feature is that the absolute $CH_4$ uptake at higher pressures is predicted to be the highest for the SAPTFF, followed by MEDFF, UFF/TraPPE, Dreiding-UFF/TraPPE, and MM3-MBIS for all materials. This is not true at low pressures, where in general UFF/TraPPE is the highest, followed by SAPTFF, Dreiding-UFF/TraPPE, MEDFF, and MM3-MBIS. There is however an exception to this rule, as for UiO-66 MEDFF predicts a lower uptake than MM3-MBIS.

*Qualitative Comparison of Zr-Based MOFs.* An alternative way to assess the various force fields is to compare absolute $CH_4$ uptake for a series of materials on a qualitative basis. Therefore, we have extended our study with two other materials belonging to the Zr-based family (i.e., sharing the same inorganic brick) but with distinct other features, namely NU-1000 and MOF-808. For these latter two materials, no experimental values are available for $CH_4$ adsorption. The topology and linkers of NU-1000 (**csq** topology and tetratopic linker) and MOF-808 (**spn** topology and tritopic linker) are shown in Figure 3 and contrasted with the UiO-66 framework. In Table 1, we report geometric properties of these Zr-based MOFs: the density, surface area, pore volume, and pore diameters. The following order is obtained when the MOFs are ranked according to pore volume per framework mass:

**Table 1. Geometric Properties of the Five Zr-Based MOFs Considered in the Qualitative Comparison**

| | UiO-66 | DUT-52 | MOF-808 | UiO-67 | NU-1000 |
|---|---|---|---|---|---|
| density (g cm$^{-3}$) | 1.238 | 0.955 | 0.840 | 0.725 | 0.499 |
| pore volume (cm$^3$ g$^{-1}$) | 0.40 | 0.62 | 0.85 | 0.87 | 1.62 |
| surface area (m$^2$ g$^{-1}$) | 1113 | 2040 | 2042 | 2949 | 3217 |
| small pore diameter (Å) | 7.3 | 8.6 | 4.8 | 10.1 | 9.8 |
| large pore diameter (Å) | 8.8 | 9.3 | 18.4 | 13.0 | 29.1 |

$$\text{UiO-66} < \text{DUT-52} < \text{MOF-808} < \text{UiO-67} < \text{NU-1000}$$

This order is reversed when considering the framework density, which is to be expected as all frameworks are chemically rather similar. When the surface area is considered, the order is basically the same as the order for the pore volume:

$$\text{UiO-66} < \text{DUT-52} \approx \text{MOF-808} < \text{UiO-67} < \text{NU-1000}$$

We conclude that the surface area and pore volume show the same order for the five MOFs we consider. The adsorption isotherms of the five Zr-based MOFs are compared to each other and shown in Figure 4, for each force field separately. First we study the absolute methane uptake at a relatively high pressure of 30 bar. At these pressures, both host−guest and guest−guest interactions are important to compute the adsorption. We focus on qualitative features predicted by the force fields and therefore determine the ranking of the MOFs at the given pressure. If we order the MOFs according to their high-pressure uptake, Dreiding-UFF/TraPPE and MM3-MBIS predict:

$$\text{UiO-66} < \text{MOF-808} < \text{DUT-52} < \text{UiO-67} < \text{NU-1000}$$

while for UFF/TraPPE, SAPTFF, and MEDFF, a slightly different order is obtained:

$$\text{MOF-808} < \text{UiO-66} < \text{DUT-52} < \text{UiO-67} < \text{NU-1000}$$

If we exclude MOF-808 from the list, all force fields predict the same order, and this order also corresponds to the ranking with respect to the pore volume and surface area. Note that MOF-808 shows a considerable slope and is far from saturation even at 80 bar, which explains why it appears to be an outlier. At even higher pressures, the uptakes are strictly ordered according to the pore volume as shown in section S10. Such high pressures (above 100 bar) are however not relevant for many applications. The ranking at relatively high pressures is a qualitative feature on which all force fields agree, although it should be mentioned that even a very simple geometric calculation (surface area or pore volume) can suffice to predict this qualitative feature.

*Influence of Defects in UiO-66.* The importance of missing linker defects in UiO-66 has been studied both experimentally and computationally.[12,60,61] Recently, Lillerud et al. demonstrated that missing cluster defects are predominant and have a large impact on nitrogen adsorption.[62] Computations confirm that **reo** structures (with missing cluster defects) profoundly affect $CO_2$ adsorption.[63]

To assess the influence of defects on methane adsorption computationally, we considered four defect structures of UiO-66.



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

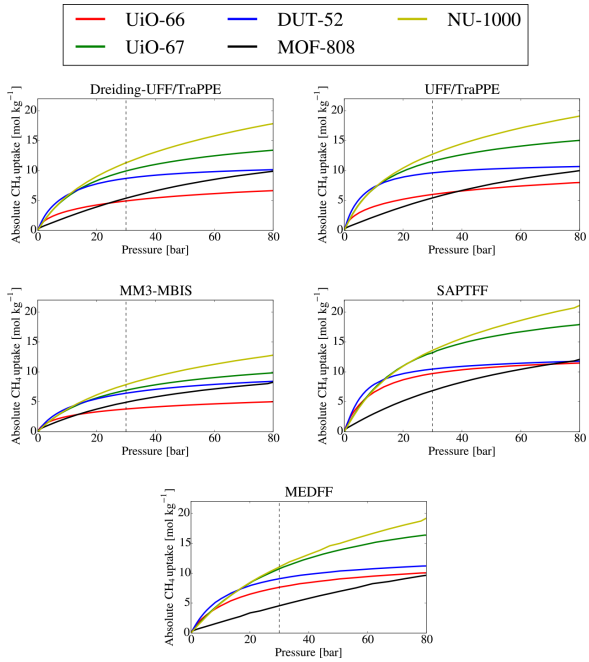



**Figure 4.** Comparison of methane adsorption in different Zr-based MOFs. The dashed vertical line indicates $P$ = 30 bar.

All adsorption simulations are performed using rigid frameworks, with geometries optimized using DFT.[64] Defect structures of UiO-66 can be created by removing BDC linkers from the pristine material, and this can be done in many ways depending on the number and position of removed BDC linkers.

The first defect structure ("1 missing linker") has one missing linker in the conventional 4-brick unit cell and this leaves two bricks 11-fold coordinated while the two other bricks remain 12-fold coordinated. This structure is classified as (11,11,12,12), according to the nomenclature recently introduced in literature.[64] The second structure ("2 missing linkers") has two missing linkers in the conventional 4-brick unit cell making all four bricks 11-fold coordinated and is classified as the $(11,11,11,11)_4$ structure. The third structure ("3 missing linkers") has three missing linkers in the conventional 4-brick unit cell, leaving two bricks 9-fold coordinated and two bricks 12-fold coordinated $(9_a,9_a,12,12)_{224}$. Next to these linker defect structures, we also consider a "1 missing cluster" defect. Here, one of the four Zr-bricks of the unit cell is removed together with all connected linkers, which is also referred to as the **reo** phase. All defects are terminated using a formate group, which ensures that no coordinatively unsaturated metal sites are introduced. A schematic representation of the defects structures considered here is shown in Figure S12.

In Figure 5, we plot the gravimetric methane adsorption isotherms of these defect structures. This allows one to show that all force fields share the same qualitative features. More specifically, in all cases, the introduction of linker defects has a small influence on methane adsorption, with deviations with respect to the pristine material being lower than 10%. In general, all force fields predict a slightly lower methane uptake at lower pressures compared to the pristine material. The introduction of a missing cluster defect has a bigger impact and leads to a considerably lower gravimetric uptake at low pressures and considerably higher gravimetric uptake at higher pressures, again relative to the pristine UiO-66 structure. Although there are important differences between the force fields (in absolute amount of methane adsorbed and in the pressure at which the "1 missing cluster" isotherm crosses the "pristine" isotherm), in this case all computations lead to qualitatively similar conclusions.

**3.2. Single Molecule Adsorption Energies in UiO-66 at the ab Initio and Force-Field Level.** In the previous section, we have shown that distinct force fields can predict very dissimilar isotherms. We now try to obtain more fundamental insight at the atomic level by studying the interaction of a single methane molecule with the framework making use of ab initio calculations. It is expected that such analysis yields insight into the adsorption features at low pressures, because in the low-pressure regime, the concentration of guest molecules is so low that host−guest interactions completely determine the uptake. In this section, we focus on methane in pristine UiO-66.

We computed the adsorption energy $E_{ads}$ of a methane molecule in UiO-66 using periodic DFT at the PBE level-of-theory with D3MBJ-ATM dispersion corrections. The adsorption energy $E_{ads}$ is defined as the energy of the host+guest system minus the energy of the separate host and guest systems:

$$E_{ads} = E_{host+guest} - E_{host} - E_{guest}$$



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

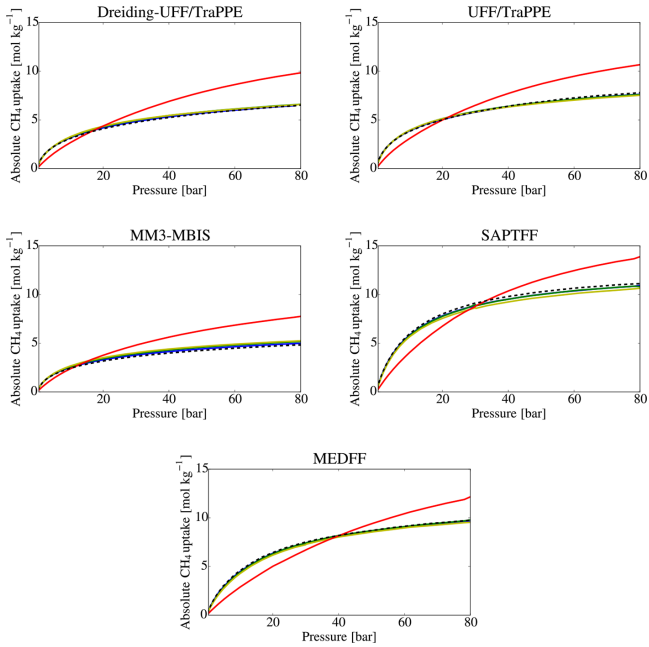



**Figure 5.** Adsorption isotherms of methane in defective UiO-66 structures.

To validate the used level-of-theory, we performed a series of CCSD(T)/CBS calculations for a set of methane-terephthalic acid dimers, which is a good model for the interaction between the guest molecule and the framework BDC linkers. The set of 80 dimer configurations is extracted from a GCMC simulation employing MEDFF. Counterpoise-corrected PBE-D3MBJ-ATM/aug-cc-pVTZ interaction energies show an RMSD of 0.39 kJ mol$^{-1}$ with respect to the CCSD(T)/CBS reference. This RMSD is much smaller than errors in the force-field interaction energies, which justifies the use of PBE-D3MBJ-ATM as the reference level-of-theory for the periodic DFT calculations. More details on the dimer calculations are provided in section S4.1.

An important point to discuss is the choice of configurations of the methane molecule in the periodic UiO-66 framework, as it is necessary to sample all energetically favorable sites. An efficient method to generate such configurations, is to extract snapshots from a GCMC simulation. It has been shown that generic force fields might not sample all relevant portions of the potential energy surface (PES).[65] We therefore extracted 100 configurations from snapshots of GCMC simulations (at 298 K and 1 bar) using all five force fields considered in this work, leading to a total set of 500 configurations. For the united-atom models, only the position of the carbon atom of methane can be extracted from the GCMC simulations: the orientation of the hydrogen atoms is in these cases determined randomly.

In Figure 6, we show scatter plots of the adsorption energies for the five force fields. It is important to stress that none of the considered force fields are fitted specifically to reproduce the ab initio data presented here, not even the ab initio derived force fields SAPTFF and MEDFF. In each case, the adsorption energies obtained with the target force field are compared with the PBE-D3MBJ-ATM data. Each plot shows the 100 configurations sampled from a Dreiding-UFF/TraPPE GCMC simulation as blue dots as well as the 100 configurations sampled from a SAPTFF GCMC simulation as red dots. The RMSD for all five force fields is also indicated in the plots and varies from 1.31 kJ mol$^{-1}$ for MEDFF to 3.24 kJ mol$^{-1}$ for UFF/TraPPE for the configurations sampled from the Dreiding-UFF/TraPPE GCMC simulation (blue dots). When considering the configurations sampled from the SAPTFF GCMC simulation (red dots), the deviations with respect to the ab initio data are notably larger, which indicates that SAPTFF indeed samples different regions of the PES. The RMSD values are the lowest for the two ab initio force fields. In the case of UFF/TraPPE, Dreiding-UFF/TraPPE, and SAPTFF, a large fraction of points is found below the diagonal, meaning that the configurations are too much stabilized compared to the ab initio adsorption data, and this will lead to an overestimation of the adsorption isotherm at low pressure. In Figure 2b, it is indeed indicated that these three force fields predict the highest methane uptake in UiO-66 at low pressures. Similar figures for all other force fields and a table summarizing all errors are provided in section S5.

We visualize the potential energy surface of methane in UiO-66 by plotting the isosurface at $E_{ads} = -8$ kJ mol$^{-1}$ for Dreiding-UFF/TraPPE (left) and SAPTFF (right) in Figure 7. Regions enclosed by this isosurface show adsorption energies that are more favorable than −8 kJ mol$^{-1}$. For SAPTFF, the adsorption energy at a point **r** is calculated as a rotational average:



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

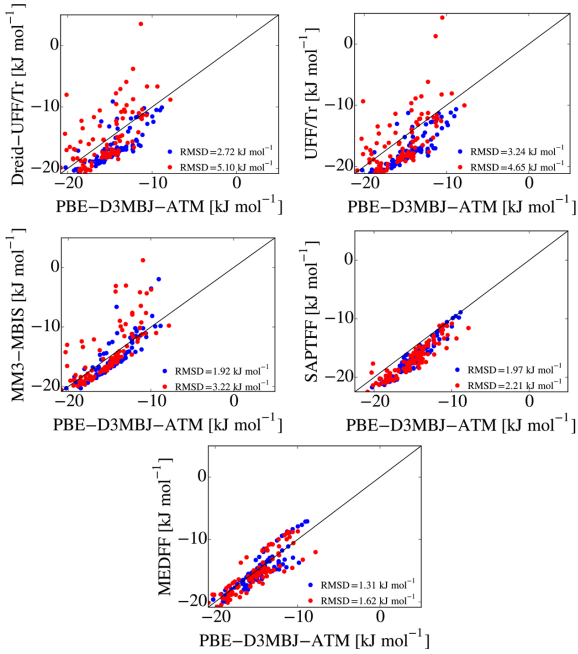



**Figure 6.** Scatter plots of adsorption energies for 200 configurations of methane in UiO-66 (blue dots, sampled from Dreiding-UFF/TraPPE GCMC; red dots, sampled from SAPTFF GCMC). Force-field energies are compared with the PBE-D3MBJ-ATM results.

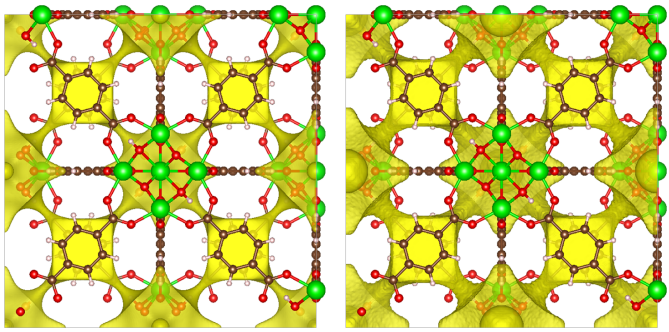


**Figure 7.** Isosurface ($E_{ads}$ = −8 kJ mol$^{-1}$) of methane in UiO-66 for Dreiding-UFF/TraPPE (left) and SAPTFF (right).

$$E_{ads}(\mathbf{r}) = -\frac{1}{\beta}\log\left[\frac{1}{N}\sum_{j=1}^{N} e^{-\beta E_{ads}(\mathbf{r},\Omega_j)}\right] \quad (13)$$

where $N$ = 100 random rotations are considered and $\beta = \frac{1}{k_B T}$ with $T$ = 298 K. Because TraPPE describes methane as a single site, this rotational averaging is not necessary for Dreiding-UFF/TraPPE. Clearly, SAPTFF predicts a larger portion of the tetrahedral pores to be favorable adsorption sites. This can be quantified by considering the volume fraction of adsorption sites that are more stabilized than $E_{ads}$ = −8 kJ mol$^{-1}$ as shown in Table 2. The difference in potential energy surface, as quantified in Table 2, explains for example why MEDFF/TraPPE predicts a higher uptake than Dreiding-UFF/TraPPE at pressures higher than 30 bar, as observed in Figure 2a. Indeed, because these two force fields share a common description of the guest−guest interactions, the difference in the number of attractive adsorption sites completely explains the different uptake at higher pressures.

*Henry Coefficients.* At low pressures, only host−guest interactions determine the uptake as the guest−guest inter-



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

**Table 2. Volume Fraction of Adsorption Sites More Stable than $E_{ads} = -8$ kJ mol$^{-1}$ in UiO-66**

| force field | volume fraction |
|---|---|
| Dreiding-UFF/TraPPE | 10.9% |
| UFF/TraPPE | 11.8% |
| MM3-MBIS | 11.5% |
| SAPTFF | 14.3% |
| MEDFF | 12.9% |

actions are unimportant at low loadings. This means there is a close correlation between single molecule adsorption energies and uptake at low pressure, which will be investigated hereafter by means of the Henry coefficient. At sufficiently low pressures, the uptake is proportional to the pressure, and the proportionality factor is called the Henry coefficient $K_H$. The value of $K_H$ for methane computed with the five force fields is given in Table 3 as well as the value estimated from low-pressure experimental data[54] and the ab initio computed value.

**Table 3. Henry Coefficient $K_H$ (mol g$^{-1}$ bar$^{-1}$) of Methane in UiO-66 at $T = 298$ K**

| | |
|---|---|
| Dreiding-UFF/TraPPE | 1.78 |
| UFF/TraPPE | 2.36 |
| MM3-MBIS | 1.21 |
| SAPTFF | 1.87 |
| MEDFF | 1.08 |
| experiment | 0.60 |
| PBE-D3MBJ-ATM | 0.91 |

The results can be linked to the comparison of ab initio and force-field single molecule adsorption energies, which are summarized in Figure 8 by showing the RMSD (left) for each force field for the data set of 500 configurations of methane in UiO-66. The RMSD values for the ab initio derived force fields MEDFF and SAPTFF are the smallest followed by MM3-MBIS, while the generic force fields Dreiding-UFF/TraPPE and UFF/TraPPE perform significantly worse. When considering adsorption, it is also important to study the mean deviation (MD): a systematic overbinding will lead to a much higher predicted uptake and Henry coefficient than a systematic underbinding, while both scenarios can give rise to the same RMSD value. The MD is shown on the right-hand side of Figure 8, and from this we conclude that MEDFF and MM3-MBIS offer the smallest MD for the periodic data set, which is in line with the observations about the Henry coefficients.

Although the ab initio computed Henry coefficient is closer to the experimental value than any of the force fields we studied, there is still a significant discrepancy. Next to possible deficiencies of the ab initio method, there are also experimental uncertainties on the reported values. These originate from imperfections or shortcomings in the experimental set up, such as slow adsorption kinetics, incomplete sample activation, external surfaces of the crystal sample, and the presence of defects.[10] A more extensive discussion on experimental uncertainties is not within the scope of the current work.

The large variations in the Henry coefficients predicted by the five force fields merit special attention, in view of the relatively small errors noticed in the force-field single molecule adsorption energies (Figure 8), which are below of 1 kcal mol$^{-1}$ (smaller than the threshold for chemical accuracy). We therefore propose to perform a sensitivity analysis of the Henry coefficient (or uptake at low pressures). We investigate the influence on the Henry coefficient by slightly varying the parameters of the Dreiding-UFF/TraPPE force field. The applied procedure for the sensitivity analysis is outlined in section S8. We illustrate the outcome of this analysis in Figure 9, where we correlate the change in the adsorption energies with the relative change on the Henry constant (in both cases with the original Dreiding-UFF/TraPPE model as a reference). By changing the Dreiding-UFF/TraPPE parameters in such a way that adsorption energies change with an RMSD of kJ mol$^{-1}$, the Henry coefficient (and thus the uptake predicted at low pressures) changes by as much as 40%. For slightly larger deviations on the RMSD (but still less than 2 kJ mol$^{-1}$), this number can rise above 80%. This extreme sensitivity is not completely unexpected, as the uptake at low pressures is proportional to the Boltzmann factor $e^{-\beta E_{ads}}$. In other words, small changes in the potential energy surface are exponentially amplified in the corresponding predicted uptake.

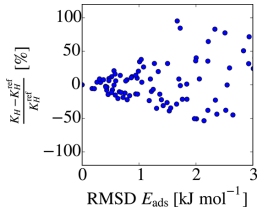



**Figure 9.** Correlation plot to investigate sensitivity of the Dreiding-UFF/TraPPE force-field parameters.

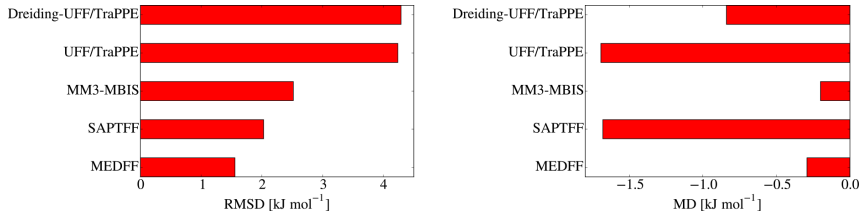



**Figure 8.** Errors of force-field single molecule adsorption energies with respect to ab initio reference data for a data set of 500 configurations of methane in the UiO-66 framework.



DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

## 4. CONCLUSIONS AND OUTLOOK

Within this paper, we studied the adsorption of methane in a series of Zr-based MOFs with the aim to critically assess the sensitivity of a diverse set of force fields to produce isotherms and single molecule adsorption energies. The selected materials include UiO-66, UiO-67, DUT-52, NU-1000, and MOF-808 and show distinctly different properties in pore volume and surface area, which is reflected in the uptake of methane. As generally known in literature, isotherms are very sensitive to the applied force field. However, to further unravel the physical origin of the observed correspondence between experimental and theoretical isotherms, we performed a systematic investigation of single methane adsorption energies using five different force fields and compared them with adsorption energies produced with periodic density functional theory data. To this end, 500 different configurations of methane adsorbed in the UiO-66 material were taken from the GCMC calculations. We find that some generic force fields such as UFF/TraPPE give an acceptable agreement with the experiment in the UiO-66 framework for pressure between 30 and 80 bar. However, these force fields fail to reproduce accurately single molecule adsorption energies (errors larger than 4 kJ mol$^{-1}$ are found), and the good correspondence between theory and experiment for the isotherm at these higher pressures should be ascribed to a fortuitous cancellation of errors. The two ab initio derived force fields, SAPTFF and MEDFF, yield a remarkable accuracy of the individual adsorption energies with deviations of less than 2 kJ mol$^{-1}$ on an overall adsorption energy (RMS value) of 15 kJ mol$^{-1}$ for methane in UiO-66. Still, such accuracy does not guarantee a quantitative reproduction of adsorption isotherms, since the uptake at low pressures is proportional to the Boltzmann factor $e^{-\beta E_{ads}}$, and thus the errors are exponentially amplified in the corresponding predicted uptake. The required accuracy for single molecule adsorption energies in order to quantitatively reproduce the isotherms in the low pressure limit is very hard to achieve with current available force fields and even ab initio methods. Apart from these quantitative differences between various methods, we find that all five force fields yield overall similar trends for the reproduction of isotherms of methane in the five different materials, which is rewarding since this validates the common usage of force field based GCMC calculations for the study of adsorption isotherms in the field of nanoporous materials. Yet when constructing force fields for computational simulation of adsorption data, it is advisible to start from ab initio derived force fields as they succeed in reproducing single molecule adsorption energies with high accuracy, which underlines the proper inclusion of host−guest interactions in these models. For higher pressures, a specifically designed force field such as TraPPE may be useful to describe guest−guest interactions, as it was specifically designed for the description of phase equilibria. In order to generalize the conclusions found here, it might be useful to extend the current study to polar adsorbates and frameworks with coordinatively unsaturated sites, which introduce specific host−guest interactions.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.7b08971.

Influence of the rigid framework approximation, the vdW cutoff radius and electrostatic interactions; results of dimer interaction energies and adsorption energies are provided; the parameters of SAPTFF are given, the defective UiO-66 structures are visualized, and the sensitivity analysis is explained in detail; and results for bulk methane simulations, isotherms up to very high pressures, and rescaling of isotherms are provided (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: veronique.vanspeybroeck@ugent.be.

### ORCID

Toon Verstraelen: 0000-0001-9288-5608

Veronique Van Speybroeck: 0000-0003-2206-178X


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 Programme/ERC Grant Agreement 64755 DYNPOR. We acknowledge the Foundation of Scientific Research-Flanders (FWO), the Research Board of Ghent University (BOF), and BELSPO in the frame of IAP/7/05 for their financial support. S.V. is a Ph.D. fellow funded by the Foundation of Scientific Research-Flanders (FWO). The computational resources and services used were provided by Ghent University (Stevin Supercomputer Infrastructure).

## ■ REFERENCES


(1) Liu, J.; Chen, L.; Cui, H.; Zhang, J.; Zhang, L.; Su, C.-Y. Applications of Metal-Organic Frameworks in Heterogeneous Supramolecular Catalysis. *Chem. Soc. Rev.* **2014**, *43*, 6011−6061.

(2) Rogge, S. M. J.; Bavykina, A.; Hajek, J.; García, H.; Olivos-Suarez, A. I.; Sepúlveda-Escribano, A.; Vimont, A.; Clet, G.; Bazin, P.; Kapteijn, F.; et al. Metal-Organic and Covalent Organic Frameworks as Single-Site Catalysts. *Chem. Soc. Rev.* **2017**, *46*, 3134−3184.

(3) Li, B.; Wen, H.-M.; Zhou, W.; Chen, B. Porous Metal-Organic Frameworks for Gas Storage and Separation: What, How, and Why? *J. Phys. Chem. Lett.* **2014**, *5*, 3468−3479.

(4) Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. Selective Gas Adsorption and Separation in Metal-Organic Frameworks. *Chem. Soc. Rev.* **2009**, *38*, 1477−1504.

(5) Düren, T.; Bae, Y.-S.; Snurr, R. Q. Using Molecular Simulation to Characterise Metal-Organic Frameworks for Adsorption Applications. *Chem. Soc. Rev.* **2009**, *38*, 1237−1247.

(6) Colón, Y. J.; Snurr, R. Q. High-Throughput Computational Screening of Metal-Organic Frameworks. *Chem. Soc. Rev.* **2014**, *43*, 5735−5749.

(7) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal-Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26*, 6185−6192.

(8) Bobbitt, N. S.; Chen, J.; Snurr, R. Q. High-Throughput Screening of Metal-Organic Frameworks for Hydrogen Storage at Cryogenic Temperature. *J. Phys. Chem. C* **2016**, *120*, 27328−27341.

(9) Fairen-Jimenez, D.; Lozano-Casal, P.; Düren, T. In *Characterisation of Porous Solids VIII*; Seaton, N., Reinoso, F. R., Llewellyn, P., Kaskel, S., Eds.; The Royal Society of Chemistry: Cambridge, UK, 2009; pp 80−87.

(10) Coudert, F.-X.; Fuchs, A. H. Computational Characterization and Prediction of Metal-Organic Framework Properties. *Coord. Chem. Rev.* **2016**, *307*, 211−236.

(11) Vasanth Kumar, K.; Charalambopoulou, G.; Kainourgiakis, M. E.; Stubos, A.; Steriotis, T. Insights on the Physical Adsorption of Hydrogen and Methane in UiO Series of MOFs Using Molecular Simulations. *Comput. Theor. Chem.* **2015**, *1061*, 36−45.




DOI: 10.1021/acs.jpcc.7b08971
J. Phys. Chem. C 2017, 121, 25309−25322

(12) Ghosh, P.; Colón, Y. J.; Snurr, R. Q. Water Adsorption in UiO-66: The Importance of Defects. *Chem. Commun.* **2014**, *50*, 11329−11331.
(13) Chen, L.; Grajciar, L.; Nachtigall, P.; Düren, T. Accurate Prediction of Methane Adsorption in a Metal-Organic Framework with Unsaturated Metal Sites by Direct Implementation of an Ab Initio Derived Potential Energy Surface in GCMC Simulation. *J. Phys. Chem. C* **2011**, *115*, 23074−23080.
(14) Bludský, O.; Rubeš, M.; Soldán, P.; Nachtigall, P. Investigation of the Benzene-Dimer Potential Energy Surface: DFT/CCSD(T) Correction Scheme. *J. Chem. Phys.* **2008**, *128*, 114102.
(15) McDaniel, J. G.; Li, S.; Tylianakis, E.; Snurr, R. Q.; Schmidt, J. R. Evaluation of Force Field Performance for High-Throughput Screening of Gas Uptake in Metal-Organic Frameworks. *J. Phys. Chem. C* **2015**, *119*, 3143−3152.
(16) Chen, L.; Morrison, C. A.; Düren, T. Improving Predictions of Gas Adsorption in Metal-Organic Frameworks with Coordinatively Unsaturated Metal Sites: Model Potentials, Ab Initio Parameterization, and GCMC Simulations. *J. Phys. Chem. C* **2012**, *116*, 18899−18909.
(17) Fischer, M.; Gomes, J. R.; Jorge, M. Computational Approaches to Study Adsorption in MOFs with Unsaturated Metal Sites. *Mol. Simul.* **2014**, *40*, 537−556.
(18) Yang, Q.; Wiersum, A. D.; Jobic, H.; Guillerm, V.; Serre, C.; Llewellyn, P. L.; Maurin, G. Understanding the Thermodynamic and Kinetic Behavior of the CO2/CH4 Gas Mixture Within the Porous Zirconium Terephthalate UiO-66(Zr): A Joint Experimental and Modeling Approach. *J. Phys. Chem. C* **2011**, *115*, 13768−13774.
(19) Lennox, M. J.; Bound, M.; Henley, A.; Besley, E. The Right Isotherms for the Right Reasons? Validation of Generic Force Fields for Prediction of Methane Adsorption in Metal-Organic Frameworks. *Mol. Simul.* **2017**, *43*, 828−837.
(20) Rappé, A. K.; Casewit, C.; Colwell, K.; Goddard, W. A., III; Skiff, W. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.
(21) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.
(22) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.
(23) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. Molecular Mechanics. the MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.
(24) Allinger, N. L.; Li, F.; Yan, L. Molecular Mechanics. The MM3 Force Field for Alkenes. *J. Comput. Chem.* **1990**, *11*, 848−867.
(25) Allinger, N. L.; Zhou, X.; Bergsma, J. Molecular Mechanics Parameters. *J. Mol. Struct.: THEOCHEM* **1994**, *312*, 69−83.
(26) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *J. Chem. Theory Comput.* **2016**, *12*, 3894−3912.
(27) McDaniel, J. G.; Schmidt, J. R. Robust, Transferable, and Physically Motivated Force Fields for Gas Adsorption in Functionalized Zeolitic Imidazolate Frameworks. *J. Phys. Chem. C* **2012**, *116*, 14031−14039.
(28) McDaniel, J. G.; Schmidt, J. R. Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory. *J. Phys. Chem. A* **2013**, *117*, 2053−2066.
(29) Vandenbrande, S.; Waroquier, M.; Van Speybroeck, V.; Verstraelen, T. The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *J. Chem. Theory Comput.* **2017**, *13*, 161−179.
(30) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials. *Mol. Simul.* **2016**, *42*, 81−101.
(31) Dubbeldam, D.; Torres-Knoop, A.; Walton, K. S. On the Inner Workings of Monte Carlo Codes. *Mol. Simul.* **2013**, *39*, 1253−1292.
(32) Poling, B.; Prausnitz, J.; Connell, J. *The Properties of Gases and Liquids SE*; McGraw Hill, 2000.
(33) Becke, A. D. A New Mixing of Hartree-Fock and Local density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(34) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(35) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(36) Lekien, F.; Marsden, J. Tricubic Interpolation in Three Dimensions. *Int. J. Numer. Methods Eng.* **2005**, *63*, 455−471.
(37) Maitland, G.; Rigby, M.; Smith, E. B.; Wakeham, W. A. *Intermolecular Forces: Their Origin and Determination*; Clarendon Press: Oxford, UK, 1981.
(38) Ewald, P. P. Die Berechnung Optischer Und Elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369*, 253−287.
(39) de Leeuw, S. W.; Perram, J. W.; Smith, E. R. Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants. *Proc. R. Soc. London, Ser. A* **1980**, *373*, 27−56.
(40) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(41) Mortensen, J. J.; Hansen, L. B.; Jacobsen, K. W. Real-Space Grid Implementation of the Projector Augmented Wave Method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2005**, *71*, 035109.
(42) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dułak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; et al. Electronic Structure Calculations with GPAW: A Real-Space Implementation of the Projector Augmented-Wave Method. *J. Phys.: Condens. Matter* **2010**, *22*, 253202.
(43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. *Gaussian Inc 09*, revision D.01; Gaussian Inc.: Wallingford, CT, 2013.
(44) Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *J. Phys. Chem. Lett.* **2016**, *7*, 2197−2203.
(45) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(46) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 558−561.
(47) Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15−50.
(48) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169−11186.
(49) Cavka, J. H.; Jakobsen, S.; Olsbye, U.; Guillou, N.; Lamberti, C.; Bordiga, S.; Lillerud, K. P. A New Zirconium Inorganic Building Brick Forming Metal Organic Frameworks with Exceptional Stability. *J. Am. Chem. Soc.* **2008**, *130*, 13850−13851.
(50) Bon, V.; Senkovska, I.; Weiss, M. S.; Kaskel, S. Tailoring of Network Dimensionality and Porosity Adjustment in Zr- and Hf-Based MOFs. *CrystEngComm* **2013**, *15*, 9572−9577.
(51) Rogge, S. M. J.; Wieme, J.; Vanduyfhuys, L.; Vandenbrande, S.; Maurin, G.; Verstraelen, T.; Waroquier, M.; Van Speybroeck, V. Thermodynamic Insight in the High-Pressure Behavior of UiO-66: Effect of Linker Defects and Linker Expansion. *Chem. Mater.* **2016**, *28*, 5721−5732.
(52) Cavka, J. H.; Grande, C. A.; Mondino, G.; Blom, R. High Pressure Adsorption of CO2 and CH4 on Zr-MOFs. *Ind. Eng. Chem. Res.* **2014**, *53*, 15500−15507.
(53) Øien, S.; Wragg, D.; Reinsch, H.; Svelle, S.; Bordiga, S.; Lamberti, C.; Lillerud, K. P. Detailed Structure Analysis of Atomic Positions and Defects in Zirconium Metal-Organic Frameworks. *Cryst. Growth Des.* **2014**, *14*, 5370−5372.



DOI: 10.1021/acs.jpcc.7b08971
*J. Phys. Chem. C* 2017, 121, 25309−25322

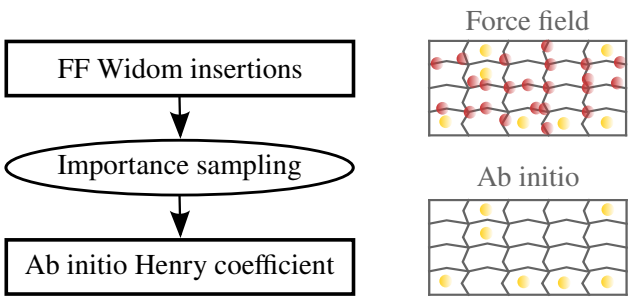
(54) Cmarik, G. E.; Kim, M.; Cohen, S. M.; Walton, K. S. Tuning the Adsorption Properties of UiO-66 Via Ligand Functionalization. *Langmuir* **2012**, *28*, 15606−15613.
(55) Demir, H.; Greathouse, J. A.; Staiger, C. L.; Perry, J. J., IV; Allendorf, M. D.; Sholl, D. S. DFT-Based Force Field Development for Noble Gas Adsorption in Metal Organic Frameworks. *J. Mater. Chem. A* **2015**, *3*, 23539−23548.
(56) Rosenbach, N., Jr; Ghoufi, A.; Deroche, I.; Llewellyn, P. L.; Devic, T.; Bourrelly, S.; Serre, C.; Ferey, G.; Maurin, G. Adsorption of Light Hydrocarbons in the Flexible MIL-53(Cr) and Rigid MIL-47(V) Metal-Organic Frameworks: A Combination of Molecular Simulations and Microcalorimetry/Gravimetry Measurements. *Phys. Chem. Chem. Phys.* **2010**, *12*, 6428−6437.
(57) Jorge, M.; Fischer, M.; Gomes, J. R. B.; Siquet, C.; Santos, J. C.; Rodrigues, A. E. Accurate Model for Predicting Adsorption of Olefins and Paraffins on MOFs with Open Metal Sites. *Ind. Eng. Chem. Res.* **2014**, *53*, 15475−15487.
(58) Chowdhury, P.; Bikkina, C.; Meister, D.; Dreisbach, F.; Gumma, S. Comparison of Adsorption Isotherms on Cu-BTC Metal Organic Frameworks Synthesized from Different Routes. *Microporous Mesoporous Mater.* **2009**, *117*, 406−413.
(59) Bobbitt, N. S.; Mendonca, M. L.; Howarth, A. J.; Islamoglu, T.; Hupp, J. T.; Farha, O. K.; Snurr, R. Q. Metal-Organic Frameworks for the Removal of Toxic Industrial Chemicals and Chemical Warfare Agents. *Chem. Soc. Rev.* **2017**, *46*, 3357−3385.
(60) Wu, H.; Chua, Y. S.; Krungleviciute, V.; Tyagi, M.; Chen, P.; Yildirim, T.; Zhou, W. Unusual and Highly Tunable Missing-Linker Defects in Zirconium Metal-Organic Framework UiO-66 and Their Important Effects on Gas Adsorption. *J. Am. Chem. Soc.* **2013**, *135*, 10525−10532.
(61) Liang, W.; Coghlan, C. J.; Ragon, F.; Rubio-Martinez, M.; D'Alessandro, D. M.; Babarao, R. Defect Engineering of UiO-66 for CO2 and H2O Uptake - a Combined Experimental and Simulation Study. *Dalton Trans.* **2016**, *45*, 4496−4500.
(62) Shearer, G. C.; Chavan, S.; Bordiga, S.; Svelle, S.; Olsbye, U.; Lillerud, K. P. Defect Engineering: Tuning the Porosity and Composition of the Metal-Organic Framework UiO-66 Via Modulated Synthesis. *Chem. Mater.* **2016**, *28*, 3749−3761.
(63) Thornton, A. W.; Babarao, R.; Jain, A.; Trousselet, F.; Coudert, F.-X. Defects in Metal-Organic Frameworks: A Compromise Between Adsorption and Stability? *Dalton Trans.* **2016**, *45*, 4352−4359.
(64) De Vos, A.; Hendrickx, K.; Van Der Voort, P.; Van Speybroeck, V.; Lejaeghere, K. Missing Linkers: An Alternative Pathway to UiO-66 Electronic Structure Engineering. *Chem. Mater.* **2017**, *29*, 3006−3019.
(65) Kulkarni, A. R.; Sholl, D. S. DFT-Derived Force Fields for Modeling Hydrocarbon Adsorption in MIL-47(V). *Langmuir* **2015**, *31*, 8453−8468.



DOI: 10.1021/acs.jpcc.7b08971
*J. Phys. Chem. C* 2017, 121, 25309−25322

Paper IV

# Ab initio Evaluation of Henry Coefficients Using Importance Sampling




S. Vandenbrande, M. Waroquier, T. Verstraelen, V. Van Speybroeck

*J. Chem. Theory Comput.*, **2018**, *14* (12), 6359-6369

S. Vandenbrande developed the methodology, performed the simulations, and wrote the manuscript.

JCTC Journal of Chemical Theory and Computation
Article
Cite This: J. Chem. Theory Comput. 2018, 14, 6359−6369
pubs.acs.org/JCTC

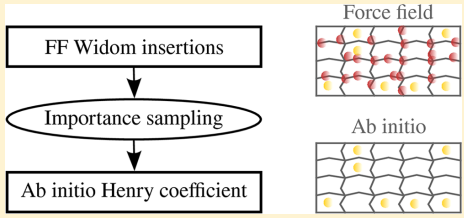
# Ab Initio Evaluation of Henry Coefficients Using Importance Sampling

Steven Vandenbrande, Michel Waroquier, Veronique Van Speybroeck, and Toon Verstraelen*

Center for Molecular Modeling (CMM), Ghent University, Technologiepark 903, 9052 Zwijnaarde, Belgium

Supporting Information

**ABSTRACT:** We present a new algorithm that allows for an efficient evaluation of the Henry coefficient of a guest molecule inside a porous material, which permits to use ab initio energy calculations. The Widom insertion method, which is currently used to compute these Henry coefficients, typically requires millions of energy evaluations. Our new methodology reduces this number by more than 1 order of magnitude, enabling the use of an ab initio potential energy surface. The methodology we propose is reminiscent of the well-known importance sampling technique which is frequently used in Monte Carlo integrations. First, a conventional Widom insertion simulation is performed using a force field. In the second step, the Widom results are used to select a limited number of configurations and only for these configurations the ab initio evaluation of the energy is required. Finally, by appropriately reweighting the latter energies, an accurate estimation of the ab initio Henry coefficient is possible at a moderate computational cost. We apply our methodology to the adsorption of $CO_2$ in Mg-MOF-74, a prototypical case where interactions of a polar guest molecule with unsaturated metal sites dominate the adsorption mechanism. In this case generic force fields such as UFF or Dreiding are inappropriate and the use of ab initio methods is indispensable. In a second case study, we compute Henry coefficients of methane in UiO-66 using different levels of theory. We pay particular attention to the influence of the dispersion corrections and the role of many-body effects. For the final example, we qualitatively investigate adsorption features for a series of functionalized UiO-66 frameworks. Overall the cases we present show that accurate computations of Henry coefficients is extremely challenging, as different levels of theory provide strongly varying results. At the same time ab initio calculations have added value compared to force fields, as they provide a physically more sound description of the adsorption mechanism and in some cases clearly improve correspondence with experiment.




## 1. INTRODUCTION

The separation and storage of gases are important industrial processes in modern-day society. Example applications include capture of the greenhouse gas $CO_2$, storage of methane or hydrogen gas for cleaner energy sources, and separation of toxic gases such as CO or ammonia to reduce environmental pollution. Porous materials (activated carbons, silicas, or zeolites) are often used in this respect, but recently much attention has been devoted to a new class of materials built from inorganic metal nodes and organic linkers: metal−organic frameworks (MOFs).[1] Because of their tunable structure and modifiable functionality, MOFs are very promising materials for the gas storage and separation applications mentioned earlier.[2] This paper deals with the computational analysis and prediction of the gas adsorption properties of MOFs, although the presented methodology is applicable to any porous medium.

There are many studies in which the adsorption of small guest molecules in MOFs has been simulated, often in a high-throughput fashion where calculations are performed for all structures in a database. Chung et al. constructed the CoRE MOF database and performed Grand-Canonical Monte Carlo (GCMC) simulations of methane adsorption for all of the more than 4700 structures.[3] Li et al. considered the same database and calculated the Henry coefficient (which completely characterizes adsorption at low pressures) for $CO_2$, $H_2O$, and $N_2$ in order to perform an initial screening for MOFs with a high selectivity.[4] In a study by Banerjee et al., the Henry coefficient of noble gases Xe and Kr in 125 000 MOF structures has been computed.[5] All of these high-throughput studies have in common that a generic force field (such as UFF[6] or Dreiding[7]) is used in order to keep the computational effort under control.

It has been demonstrated that generic force fields provide results in agreement with experimental adsorption isotherms for many MOFs.[8−10] At the same time there are some specific cases in which these classical force fields fail to properly describe the adsorption mechanism.[11] Grajciar et al. performed a systematic investigation of different guest molecules adsorbing on $Cu^{2+}$ and $Fe^{3+}$ clusters, which are models of coordinatively unsaturated metal sites (CUS).[12] The generic

Received: August 31, 2018
Published: October 30, 2018





DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

Journal of Chemical Theory and Computation Article

UFF potential performed surprisingly well in the description of interactions with organic linkers. The adsorption at CUS was however underestimated by more than 10 kJ $mol^{-1}$, and in such cases the use of ab initio methods is required. In the same study it was noted that the accuracy of often-used levels of theory, such as dispersion-corrected density functional theory (DFT), depends on the specific combination of adsorbate and adsorbent. This has been linked to the inaccuracy of DFT for the electronic structure of (open-shell) metal cations.[11]

In the context of CUS, the MOF-74 structure is an often studied example. For instance, Valenzano et al. investigated the adsorption of CO, $N_2$, and $CO_2$ in Mg-MOF-74 both computationally and experimentally.[13] Adsorption enthalpies are calculated by adding thermal corrections and zero-point energies to the electronic energy of the adsorbed complex extracted from periodic, dispersion-corrected B3LYP calculations. The experimental adsorption enthalpy of $CO_2$ is −47 kJ $mol^{-1}$, in reasonable agreement with the calculated value of −38 kJ $mol^{-1}$. Similarly, Kundu et al. studied CO and $N_2$ in the same MOF making use of ab initio calculations.[14] Even in this case study, in which only one structure is considered, performing ab initio GCMC simulations is computationally not feasible due to current hardware limitations. Those authors therefore propose a model in which the Gibbs free energy of a single adsorption site is computed using ab initio methods and combine this with the empirical Bragg-Williams/Langmuir model in order to compute an adsorption isotherm, which was in close agreement with experiment. Similarly Alonso et al. employed a combination of periodic DFT calculations with the dual-site Langmuir model in order to investigate the effect of $SO_2$ poisoning on the adsorption of $CO_2$ in Mg-MOF-74.[15]

Such an approach however is only appropriate on condition that there are a few clearly identifiable adsorption sites. Additionally, Lee et al. showed that, even if this is the case, the correlation between heats of adsorption and the Henry coefficient is not perfect.[16] This demonstrates the necessity for a complete ab initio evaluation of the Henry coefficient, and hereafter we give a short discussion on models aiming to do so. Peirs et al. reported ab initio computed Henry constants, separation constants and heats of adsorption for $N_2$ and $O_2$ in faujasite.[17] By using a cluster model for the cage and exploiting symmetry, the necessary integrals were computed on a cubic grid at the HF/6-31G* level of theory. Chen et al. also computed adsorption energies on a cubic grid, using methane in CuBTC as a case study.[18] To keep the computational cost under control, no rotational averaging was performed, or in other words, methane was approximated as a spherical molecule. By combining the grid (containing guest-framework interactions) with an empirical model for guest−guest interactions, an isotherm was constructed in better agreement with experiment than corresponding results obtained from classical force fields. Lee et al. recently proposed a biased Widom insertion method and applied it to compute the Henry coefficient for $CO_2$, $N_2$, $CH_4$, and $C_2H_2$ in Zn-MOF-74 and Mg-MOF-74.[16] This methodology relies on partitioning the framework into three regions, representing strongest, intermediate, and weakest adsorption domains based on the $CO_2$/MOF binding energies. Insertions are still random, as in the conventional Widom insertion scheme, but biased into the strongly adsorbing regions to quickly achieve good statistics on the required integrals.

Clearly, many methodologies have been developed and applied successfully in the context of gas adsorption in porous materials, specifically MOFs. Most often classical force fields are used because the large number of GCMC or Widom insertion simulations are computationally too expensive to rely on an ab initio potential energy surface (PES). Yet in many cases it would be very interesting, and sometimes indispensable, to study adsorption properties using an approach based on ab initio principles to enhance the accuracy of energy predictions.[19] Next to increased accuracy, an ab initio approach offers the additional advantage that it is more generally applicable and is able describe a wide range of chemical environments. Contrary to many force fields, there is no need for a time-consuming calibration specific to the specific combination of adsorbate and adsorbent. It should also be noted that empirical force fields can sometimes provide results in agreement with experiment, but without correctly capturing the underlying physical interactions.[20] This is another point where an appropriate ab initio level of theory provides added value.

In this work, we present a novel methodology to compute Henry coefficients, based on the concept of importance sampling that is generally applicable. We show that the number of required energy evaluations to compute the Henry coefficient is more than 1 order of magnitude smaller than using conventional Widom insertion. This drastic reduction makes the use of an ab initio PES realistic. First we present the theoretical development of our methodology. Its efficiency will then be demonstrated on some selected test cases.

## 2. METHODOLOGY

**2.1. Theoretical Development.** The adsorption of guest molecules in a porous material is described by an adsorption isotherm, which relates the number of adsorbed guest molecules to the external pressure (or the chemical potential, which is tightly coupled to the external pressure) at a certain temperature. Such an adsorption isotherm can be computed using for example GCMC simulations. At sufficiently low pressures, however, the number of adsorbed guest molecules is proportional to the external pressure $P$ and the adsorption isotherm is described by

$$\rho_a = K_H P \tag{1}$$

with $\rho_a$ the volumetric density of adsorbed guest molecules. This is essentially Henry's law, and therefore $K_H$ is called the Henry coefficient. The Henry coefficient $K_H$ at a certain temperature completely characterizes the adsorption in the low-pressure regime and is related to the excess chemical potential $\mu^{ex}$:[21]

$$K_H = \beta \exp\left(-\beta\mu^{ex}\right) \tag{2}$$

where $\beta = \frac{1}{k_B T}$. The excess chemical potential of the guest molecule, which is considered to be rigid, in the porous material can be computed using the Widom insertion method.[22]

$$\mu^{ex} = -\frac{1}{\beta}\ln\frac{\int \mathrm{d}\mathbf{s}\langle\exp(-\beta\Delta U)\rangle_{host}}{\int \mathrm{d}\mathbf{s}} \tag{3}$$

In this formula $\langle\cdots\rangle_{host}$ denotes a canonical ensemble average over configurations of the porous material, the integration $\int \mathrm{d}\mathbf{s}$ runs over all configurations of the guest molecule and $\Delta U = U_{host+guest} - U_{host} - U_{guest}$ is the adsorption energy. In this work we will employ the often-used rigid framework approximation,



DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

which means that the average over host configurations is reduced to a single configuration. The Henry coefficient can then be calculated using the Widom insertion method as follows:

$$K_{\mathrm{H}} = \beta \frac{\int \mathrm{d}\mathbf{s} \exp\left(-\beta \Delta U\right)}{\int \mathrm{d}\mathbf{s}} \tag{4}$$

In practice, the integral is approximated by generating a random unbiased set of $N$ configurations $\{\mathbf{s}_i\}$, calculating the corresponding set of adsorption energies $\{\Delta U_i\}$ and averaging the integrand:

$$\hat{K}_{\mathrm{H}} = \beta \frac{1}{N} \sum_{i=1}^{N} \exp\left(-\beta \Delta U_i\right) \tag{5}$$

The hat symbol on $\hat{K}_{\mathrm{H}}$ indicates that this represents a Monte Carlo estimation of the true value. The presented Widom insertion method requires a high number of configurations that need to be sampled in order to obtain an accurate estimation of $\hat{K}_{\mathrm{H}}$: typically $N$ should be of the order of millions. For pairwise-additive force fields this incurs an acceptable computational cost. For ab initio calculations on the other hand, this is prohibitive (except perhaps for some very small systems). Below we present a novel approach that renders the ab initio estimation of Henry coefficients computationally feasible.

In order to gain more insight into why a large number of configurations is required in the Widom insertion method, we study the sampling variance of the Monte Carlo estimator, which is related to the error bar on the computed integral:

$$\mathrm{VAR}[\hat{K}_H] = \frac{1}{N(N-1)} \sum_{i=1}^{N} \left(\beta \exp(-\beta \Delta U_i) - \hat{K}_{\mathrm{H}}\right)^2 \tag{6}$$

The term between brackets will fluctuate over many orders of magnitude because of the presence of the Boltzmann factor, which results in a large variance. Another way to look at this problem is as follows: most of the randomly generated configurations will not be favorable (i.e., $\Delta U_i$ is a positive number that is much larger than $k_B T$) and will make a negligible contribution to $\hat{K}_{\mathrm{H}}$. On the other hand, the most favorable adsorption sites (i.e., $\Delta U_i$ is a negative number) that contribute the most to the integral, will only be sampled ever so often. This is a problem that is often encountered in random sampling and a well-known solution is to use importance sampling. To this end we introduce a biasing potential energy surface $\tilde{U}$, of which the adsorption energies are represented by $\Delta\tilde{U}$. The expression for the Henry coefficient can now be rewritten as

$$K_{\mathrm{H}} = \beta \frac{\int \mathrm{d}\mathbf{s} \exp\left(-\beta \Delta U\right)}{\int \mathrm{d}\mathbf{s}} \tag{7}$$

$$= \beta \frac{\int \mathrm{d}\mathbf{s} \exp\left(-\beta[\Delta U - \Delta\tilde{U}]\right) \exp\left(-\beta \Delta\tilde{U}\right)}{\int \mathrm{d}\mathbf{s}} \tag{8}$$

$$= \frac{\int \mathrm{d}\mathbf{s} \exp\left(-\beta[\Delta U - \Delta\tilde{U}]\right) \exp\left(-\beta \Delta\tilde{U}\right)}{\int \mathrm{d}\mathbf{s} \exp\left(-\beta \Delta\tilde{U}\right)} \beta \frac{\int \mathrm{d}\mathbf{s} \exp\left(-\beta \Delta\tilde{U}\right)}{\int \mathrm{d}\mathbf{s}} \tag{9}$$

$$= \langle \exp(-\beta[\Delta U - \Delta\tilde{U}]) \rangle_{e^{-\beta \Delta \tilde{U}}} \tilde{K}_{\mathrm{H}} \tag{10}$$

where $\tilde{K}_{\mathrm{H}}$ is the Henry coefficient for the PES $\tilde{U}$. This approach is only useful if the biasing potential $\tilde{U}$ is computationally cheap (in practice a force field) and the original potential $U$ is computationally expensive (in practice an ab initio method). In this case, the Henry coefficient $\tilde{K}_{\mathrm{H}}$ can be accurately computed using the Widom insertion method by using a large value for the number of samples $N$. Ab initio calculations are only required to evaluate the expectation value $\lambda_{\tilde{U}}^{U} = \langle \exp\left(-\beta[\Delta U - \Delta\tilde{U}]\right) \rangle_{e^{-\beta \Delta \tilde{U}}}$. This expectation value is estimated by generating a set of $n$ configurations $\{\mathbf{s}_\alpha\}$ where the probability to include a configuration is given by $p_\alpha \approx \exp\left(-\beta \Delta\tilde{U}_\alpha\right)$:

$$\hat{\lambda}_{\tilde{U}}^{U} = \frac{1}{n} \sum_{\alpha=1}^{n} \exp\left(-\beta[\Delta U_\alpha - \Delta\tilde{U}_\alpha]\right) \tag{11}$$

A crucial difference with eq 5 is that, on condition that the biasing potential $\tilde{U}$ and the original potential $U$ are sufficiently similar, the integrand of $\lambda_{\tilde{U}}^{U}$ will always be close to unity and therefore show a much smaller variance. The advantage is that here ab initio computation are performed only for a limited number of configurations which are truly important. In other words, the number of evaluations of the ab initio PES $U$ is much smaller than in the Widom insertion method.

The method we present can be interpreted and summarized as follows. We start by computing the Henry coefficient $\tilde{K}_{\mathrm{H}}$ with a force-field potential $\tilde{U}$ using the Widom insertion method. As explained, this requires a large number of energy evaluations which is however feasible for the computationally cheap force field. Next, we select a fraction of configurations from this simulation where the probability to include configuration $\alpha$ is proportional to $\exp\left(-\beta \tilde{U}_\alpha\right)$. In other words, the biasing potential selects favorable adsorption sites (which contribute most to the Henry coefficient) and ensures that the computationally expensive ab initio calculations are only performed for this limited set of configurations that are most important. By computing the average of $\exp(-\beta[\Delta U_\alpha - \Delta\tilde{U}_\alpha])$ for these biased configurations, we can finally estimate the ab initio value of the Henry coefficient using eq 10.

To conclude, we note that the method outlined above can also be used to compute other quantities of interest. For example the average adsorption energy (related to the adsorption enthalpy as $\Delta H_{\mathrm{ads}} = E_{\mathrm{ads}} - \frac{1}{\beta}$) can be calculated as follows:

$$E_{\mathrm{ads}} = \frac{\int \mathrm{d}\mathbf{s} \Delta U e^{-\beta \Delta U}}{\int \mathrm{d}\mathbf{s} e^{-\beta \Delta U}} \tag{12}$$

$$= \frac{\int \mathrm{d}\mathbf{s} \Delta U e^{-\beta[\Delta U - \Delta\tilde{U}]} e^{-\beta \Delta\tilde{U}}}{\int \mathrm{d}\mathbf{s} e^{-\beta[\Delta U - \Delta\tilde{U}]} e^{-\beta \Delta\tilde{U}}} \tag{13}$$



DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

$$= \frac{\frac{\int \mathrm{d}\mathbf{s}\Delta U e^{-\beta[\Delta U-\Delta\tilde{U}]}e^{-\beta\Delta\tilde{U}}}{\int \mathrm{d}\mathbf{s}e^{(-\beta\Delta\tilde{U})}}}{\frac{\int \mathrm{d}\mathbf{s}e^{-\beta[\Delta U-\Delta\tilde{U}]}e^{-\beta\Delta\tilde{U}}}{\int \mathrm{d}\mathbf{s}e^{(-\beta\Delta\tilde{U})}}} \quad (14)$$

$$= \frac{\langle \Delta U \exp(-\beta[\Delta U - \Delta\tilde{U}])\rangle_{e^{-\beta\Delta\tilde{U}}}}{\langle \exp(-\beta[\Delta U - \Delta\tilde{U}])\rangle_{e^{-\beta\Delta\tilde{U}}}} \quad (15)$$

Again the expectation values will converge quickly with increasing number of samples on condition that $U$ and $\tilde{U}$ are sufficiently similar. Note that our procedure to calculate the ab initio value of the Henry coefficient also provides all necessary ingredients to compute the average adsorption energy as expressed in eq 15.

**2.2. Validation of the Methodology on $CH_4$ Adsorption in UiO-66.** We validated our approach by comparing with the numerical evaluation of the integral in eq 4 on a regular grid, considering $CH_4$ adsorption in UiO-66 as a test case. Details about this regular grid, that samples translations of the center-of-mass as well as rotations of the guest molecule, are provided in the Supporting Information. This numerical integration requires a lot of computational power and is therefore not a recommended approach: we use it here for one particular case in order to obtain a reference value for the Henry coefficient, which allows to validate the derivation and implementation of our importance sampling method.

Next to ab initio energies, force-field energies for all configurations considered in the regular grid were computed, thus constructing a database of ab initio and corresponding force-field adsorption energies to which we applied our methodology as outlined above. The ab initio level of theory is PBE+D3(BJ)[23,24] and the necessary periodic single-point DFT calculations are performed using VASP[25−28] version 5.4 using the projector augmented wave (PAW) method,[29,30] with the supplied PAW−PBE potentials. Different force fields are considered for the biasing potential $\tilde{U}$: MEDFF,[31,32] MM3,[33] Dreiding[7]/TraPPE,[34,35] and UFF[6]/TraPPE. Note that only host−guest force-field terms are required, as the frameworks and guest molecules are considered rigid and furthermore only one guest molecule is considered so guest−guest interactions are absent. All force-field calculations are performed using the in-house code YAFF.[36] The results are shown in Figure 1,

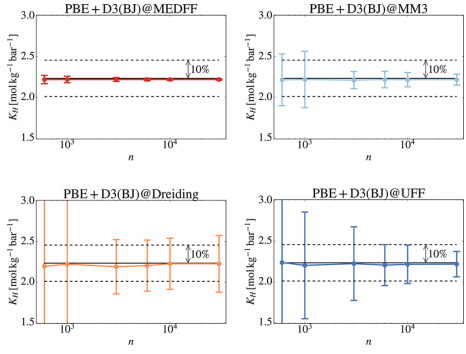



**Figure 1.** Comparison of our method with a grid-based approach for the Henry coefficient $K_H$ of $CH_4$ in UiO-66 at room temperature.

where the estimated Henry coefficient at room temperature as a function of the number of samples ($n$ in eq 11) is plotted in color. We use the notation PBE+D3(BJ)@MEDFF to indicate for instance that the Henry coefficient is calculated for the PBE+D3(BJ) level of theory, using our importance sampling approach with MEDFF as the biasing potential. For each sample size $n$, the simulation was repeated 1000 times: the dots indicate the average of these 1000 simulations and the error bar indicates the standard deviation. Irrespective of which force field is used as the biasing potential, the average values are always very close to the reference value from the grid-based calculation (shown as a full black line). Concerning the error bars, however, there are striking differences between the force fields. If MEDFF is used as the biasing potential, the error bar is only a few percent even if only about 1000 ab initio calculations are considered. For MM3, slightly larger sample sizes are required to reach similar accuracy, while for UFF and Dreiding the error is still about 10% even if 10 000 ab initio calculations are included. The reason for this has been mentioned in the theoretical derivation: it is required that the original potential $U$ and the biasing potential $\tilde{U}$ are sufficiently similar. Indeed, in our previous work we have shown that MEDFF adsorption energies for this system are in good agreement with ab initio values, while this is not the case for generic force fields such as Dreiding or UFF.[20]

The number of ab initio calculations that need to be considered to reach convergence, will obviously depend on the system of interest. It is not possible to estimate this number a priori, as it depends for instance strongly on the biasing force field. It could however be suspected that the symmetry of the adsorbent will play an important role. For the UiO-66 structure considered here, there are 24 symmetry operators. For a more complicated structure where this symmetry is completely absent, a significantly larger computational effort may be required.

The previous findings would suggest that pure force-field predictions of the Henry coefficient using MEDFF would also give good agreement with the PBE+D3(BJ) value, as both PES agree rather well. However, this is contradicted by the numerical values for the Henry coefficients taken up in Table 1, which show that on the contrary the UFF Henry

**Table 1. Henry Coefficient and Average Adsorption Energy of $CH_4$ in UiO-66 at Room Temperature: Comparison of Ab Initio and Force-Field Results**

| level of theory | $K_H$ [mol kg$^{-1}$] | $\Delta H_{ads}$ [kJ mol$^{-1}$] |
|---|---|---|
| PBE+D3(BJ) (grid) | 2.23 | −19.9 |
| PBE+D3(BJ) (importance sampling) | 2.22 | −19.9 |
| MEDFF (Widom) | 1.17 | −18.0 |
| MM3 (Widom) | 1.15 | −18.4 |
| Dreiding (Widom) | 1.72 | −19.4 |
| UFF (Widom) | 2.24 | −19.9 |

coefficient is very close to the ab initio PBE+D3(BJ) result. This apparent paradox is resolved by comparing ab initio and force-field adsorption energies. The RMSD between PBE+D3(BJ) and MEDFF is 2.7 kJ mol$^{-1}$, considering all configurations of the regular grid for which the PBE+D3(BJ) adsorption energy is negative. The RMSD between PBE+D3(BJ) and UFF on the other hand is 14.5 kJ mol$^{-1}$, indicating a much worse correlation for this force field with the ab initio reference. Scatter plots (shown in the Supporting



DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

Information) also reveal that UFF sometimes provides adsorption energies much more attractive than the reference value, while in many other cases it is severely more repulsive. The reason for the apparent good agreement (concerning the Henry coefficient) is due to a fortuitous cancellation of errors (sometimes overbinding, sometimes underbinding) when averaging is done as in eqs 4 or 12. As indicated in our earlier study on the prediction of adsorption isotherms, this is another case where apparent good results of generic force fields must be treated with caution.[20]

An interesting question is whether a limited number of ab initio calculations, which are needed anyway to apply the importance sampling method, could be used to reparametrize the force field that was used as the biasing potential. This reparametrized force field could then be used to calculate the Henry coefficient and should provide a result close to the ab initio value. We have investigated this approach and report the results in the Supporting Information. In short, we find that it can give acceptable approximations to the ab initio result (although less precise than the importance sampling method), provided the original force field is fairly accurate and features a lot of force-field parameters that are optimized. If this is not the case or one is interested in a very precise ab initio value, our importance sampling methodology is usually preferred because it can be fully automated and systematically converges to the correct value.

From the results presented in this section we conclude, by comparing with a grid-based approach, that our method provides correct results for the ab initio prediction of Henry coefficients. In principle any force field can be used as the biasing potential, but we showed that MEDFF leads to a faster convergence because it predicts adsorption energies that correlate well with ab initio values. Therefore, in the remainder of this work we will always employ MEDFF as the biasing potential. These conclusions were drawn for the PBE+D3(BJ) level of theory, but in the Supporting Information we show that they are valid for other ab initio methods as well.

## 3. RESULTS AND DISCUSSION

We now apply the developed procedure to two relevant test cases. First we discuss $CO_2$ adsorption in Mg-MOF-74, where we show that in general DFT dispersion corrected calculations improve the agreement with experiment compared to generic force fields. The second case is the adsorption of $CO_2$, $CH_4$, and $N_2$ in UiO-66 and its functionalized variants. There we mainly focus on the comparison of different DFT levels of theory, discussing among others the influence of many-body dispersion correction schemes.

**3.1. Adsorption of $CO_2$ on Mg-MOF-74.** The adsorption of $CO_2$ in MOF-74 frameworks is considered very challenging from a computational point of view. It is a prototypical case where the main adsorption mechanism is due to interactions between coordinatively unsaturated sites of the framework and a polar guest molecule. This typically results in generic force fields underestimating adsorption at low pressure by a few orders of magnitude, because such force fields cannot describe the interactions with open metal sites.[14]

Here we consider the MOF-74 variant with magnesium as metal element, which is referred to as Mg-MOF-74, Mg-CPO-27, or $Mg_2$(dobdc). The $Mg^{2+}$ metal nodes are linked by 2,5-dioxido-1,4-benzenedicarboxylate ($dobdc^{4-}$) ligands (as shown in Figure 2), with each magnesium atom 5-fold coordinated by oxygen atoms. Variants containing transition metals (Mn, Fe, Co, Ni, and Cu) exhibit multiple magnetic configurations.[37] Additionally, DFT provides a poor description of interactions concerning localized d electrons, which means corrections such as the Hubbard $U$ model are required.[38] To avoid these additional computational complications, we do not consider frameworks containing transition metals.

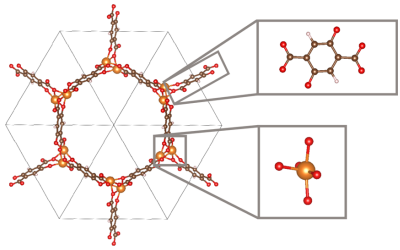


**Figure 2.** Graphical representation of the primitive unit cell of Mg-MOF-74, viewed along the ⟨1,1,1⟩ direction. Insets show the $dobdc^{4-}$ ligand and the unsaturated $MgO_5$ polyhedron. Color codes: Mg (orange), O (red), C (brown), and H (white).

We constructed a database of $50 \times 10^6$ randomly generated configurations of $CO_2$ in Mg-MOF-74 (geometry from a full optimization of the empty framework, see the Supporting Information) and computed the corresponding MEDFF adsorption energies. Next, in order to test the convergence of the conventional Widom method, 100 different subsets were chosen randomly, each subset containing $N$ samples. For each subset the Henry coefficient and adsorption enthalpy were estimated using eqs 5 and 12. By repeating this procedure for different values of $N$, we can plot the convergence of the Widom insertion method as a function of the number of considered samples as shown in Figure 3 on the left (in these violin plots the thickness indicates the probability to find the corresponding value, while bars indicate minimal, mean, and maximal value encountered). To obtain a Henry coefficient that is within a range of 10% of the converged result, more than $10^6$ Widom insertions are required. Concerning the adsorption enthalpy, the same number of Widom insertions is

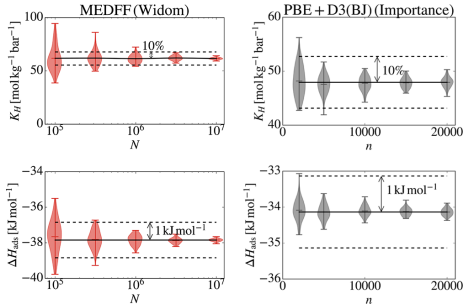



**Figure 3.** Convergence of the Henry coefficient and adsorption enthalpy of $CO_2$ in Mg-MOF-74 as a function of the number of samples. Left: MEDFF using conventional Widom insertion, right: PBE+D3(BJ) using our importance sampling method. Dotted lines indicate deviations with respect to the converged value.



DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

Journal of Chemical Theory and Computation Article

sufficient to obtain a result with less than 1 kJ mol$^{-1}$ uncertainty (with respect to the "correct" result for the given PES). We also applied our importance sampling methodology to this example with MEDFF as the biasing potential. As mentioned earlier, a generic force field is probably a poor option to choose as the biasing potential for this application. We assume that MEDFF will be a better choice, as it will provide a better correspondence with the ab initio PES because it has been shown to be very robust and particularly suitable for the prediction of interaction energies.[20,31] This assumption is confirmed by the obervation that MEDFF predicts Henry coefficients and adsorption enthalpies closer to ab initio results for $CO_2$ in Mg-MOF-74, which will be discussed further on in this section. Then, 20 000 configurations were selected, with probability to include configuration $\alpha$ proportional to $\exp(-\beta\tilde{U}_\alpha)$ where $\tilde{U}_\alpha$ is the MEDFF adsorption energy. For each of the selected 20 000 configurations the ab initio adsorption energy was calculated using various levels of theory. The electronic energy is calculated using VASP, while Grimme's dispersion corrections are calculated using the *dftd3*[39] program (see the Supporting Information for more details). Again, 100 different subsets of $n$ samples were chosen and for each subset the Henry coefficient and adsorption enthalpy were estimated, now using eqs 10 and 15. The results obtained using this importance sampling scheme are shown in Figure 3 on the right for the PBE+D3(BJ) ab initio method. Clearly the results converge much faster than for the conventional Widom simulation: about 10 000 samples are sufficient to obtain an accuracy of 10% on the Henry coefficient and 1 kJ mol$^{-1}$ on the adsorption enthalpy (again with respect to the fully converged value for the given PES). This demonstrates again that the importance sampling method reduces the required number of calculations by about a factor 100, making it computationally feasible for ab initio methods. Note that we compared the convergence behavior with two different methods (Widom insertions vs importance sampling) but also with two different PES (MEDFF vs PBE+D3(BJ)). Although the shape of the PES can also influence the convergence, we suspect that this has a minor effect considering that MEDFF and PBE+D3(BJ) adsorption energies correlate well. This means that the main difference in convergence is indeed due to the choice of the sampling methodology.

We now turn our attention to a comparison of the adsorption properties from force-field and ab initio simulations as well as from experiments for $CO_2$ in Mg-MOF-74. An overview of results is provided in Table 2, where experimental enthalpies of adsorption are obtained as $\Delta H = -Q_{st}$ with $Q_{st}$ the isosteric heat of adsorption. Ab initio results are obtained using the importance sampling method with MEDFF as biasing potential while force-field results are obtained using Widom insertion. Each simulation was performed on two different framework geometries: the DFT optimized structure of the empty framework and the DFT optimized structure of the framework loaded with 6 $CO_2$ molecules per unit cell, with each $CO_2$ located at a primary adsorption site. The presence of the $CO_2$ molecules induces subtle changes in the framework geometry as visualized in Figure 4, where a cutout of the open-metal site in the empty framework is shown, overlaid with the same atoms from the loaded framework shown in gray. A notable difference is the increase of the distance between the Mg atom and the apical oxygen (this is the oxygen opposite of the open side of the metal atom), which is 2.035 Å for the empty framework and 2.062 Å in the loaded framework. The cell parameters change ever so slightly, with the volume going from 1351.5 Å$^3$ in the empty framework to 1350.6 Å$^3$ for the loaded framework. In previous work it was found that the difference in binding energy changes by about 2.5 kJ mol$^{-1}$ when allowing the framework to relax in the presence of $CO_2$, which we deem significant.[40]

The computed Henry coefficients and enthalpies of adsorption for both geometries are compared in Table 2 (upper). Despite the seemingly small differences in atomic positions, the impact on adsorption properties is considerable: the Henry coefficients are a factor 3 to 4 larger in the geometry of the loaded framework and adsorption enthalpies are about 4 kJ mol$^{-1}$ lower, so the effect is slightly larger than in previous work.[40] The values obtained for the loaded framework tend to be in better agreement with experiment as discussed below.

**Table 2. Henry Coefficients in mol kg$^{-1}$ bar$^{-1}$ and Enthalpies of Adsorption in kJ mol$^{-1}$ of $CO_2$ in Mg-MOF-74 at 298 K**

| | geometry empty | | geometry loaded | |
|---|---|---|---|---|
| level of theory | $K_H$ | $\Delta H_{ads}$ | $K_H$ | $\Delta H_{ads}$ |
| PBE+D2 | 34 | −34.2 | 124 | −38.1 |
| PBE+D3(BJ) | 48 | −34.1 | 168 | −38.2 |
| PBE+D3M(BJ) | 112 | −37.5 | 434 | −41.4 |
| vdW-DF2 | 130 | −37.8 | 522 | −42.1 |
| MEDFF | 62 | −37.8 | 488 | −44.8 |
| MM3 | 28 | −31.9 | 71 | −35.1 |
| Dreiding | 27 | −30.6 | 45 | −32.5 |
| UFF | 31 | −31.2 | 47 | −32.8 |
| experimental source | | $K_H$ | | $\Delta H_{ads}$ |
| Britt et al.[50] | | | | −39 |
| Mason et al.[51] | | 384 | | −42 |
| Yu et al.[37] | | 414 | | −42 |
| Queen et al.[52] | | 407 | | −44 |

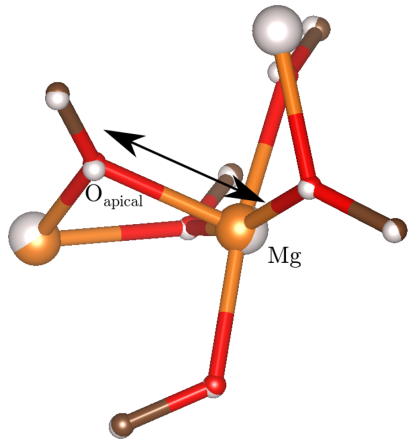



**Figure 4.** Cutout of the open-metal site in the empty framework (Mg, orange; O, red; C, brown). The same atoms in the framework loaded with $CO_2$ are shown in gray. The arrow indicates the distance between the metal atom and the oxygen atom opposite of the open side of the metal atom.



DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

The previous observations imply that the rigid framework approach is not valid for this material, despite its application in earlier studies.[16,19,41] An extension of the approach followed here using a hybrid MC/MD approach (where the MD part allows changes in framework geometry due to the presence of guest molecules) might be interesting for future computational studies but is out of the scope of this work.

Finally we discuss the comparison with experiment: several experimental results are compiled in Table 2 (lower). As noted before in the literature, generic force fields (UFF, Dreiding, MM3) underestimate the experimental Henry coefficient by more than 1 order of magnitude and the absolute value of the adsorption enthalpy by about 10 kJ $mol^{-1}$, irrespective of which framework geometry is considered. MEDFF fares slightly better, and for the loaded framework geometry it is very close to experimental values. All ab initio methods predict higher Henry coefficients than the generic force fields and are thus "closer" to experiment: again, for the loaded framework geometry the agreement with experiment is acceptable. The spread on the results from different levels of theory is also remarkable. For instance, PBE+D3M(BJ)[42] only differs from PBE+D3(BJ)[24] in the data set that was used to fit the damping parameters of the dispersion correction scheme. Yet, the former Henry coefficient is more than twice the value of the latter while the adsorption enthalpies differ by about 3 kJ $mol^{-1}$. On the other hand, the results obtained with another functional vdW-DF2[43−46] are relatively close to those from PBE+D3M(BJ). The PBE+D2[47] and PBE+D3(BJ) predictions of the adsorption enthalpy of $CO_2$ in Mg-MOF-74 are systematically underestimated with respect to vdW-DF2. This finding has been confirmed by a recent work[40] where the performance of several van der Waals corrected functionals was compared.

It has been suggested in the literature that computed isotherms (and Henry coefficients) should be rescaled to account for the limited availability of adsorption sites in experimental samples, compared to ideal crystal structures used in computations.[14,48] Based on experimental adsorption isotherms, a scaling factor of 0.765 is suggested for the case of Mg-MOF-74.[49] By rescaling the Henry coefficient predicted by vdW-DF2 and MEDFF for the loaded framework, the agreement with experiment improves. In all other cases howevers, the agreement is worse. Note that results in Table 2 are not rescaled.

We conclude that the Henry coefficient and enthalpy of adsorption are rather sensitive quantities, both to the PES and to the framework geometry. This makes a consistent quantitative agreement with experiment very challenging.

**3.2. Influence of (Many-Body) Dispersion Corrections on $CH_4$ Adsorption in UiO-66.** We used our methodology to study the adsorption of $CH_4$ in the pristine UiO-66[53] MOF using various ab initio methods, always with MEDFF as the biasing potential. The force-field value is determined using $N$ = 10 000 000 Widom insertions and the ab initio value using $n$ = 5000 importance sampled configurations. All results reported in this section are calculated using the PBE functional. The electronic part of the adsorption energy is calculated using VASP, Grimme's dispersion corrections using the *dftd3* and many-body dispersion corrections using the Pymbd[54] program. Using pure PBE energies, nearly no $CH_4$ adsorption is predicted because of the lack of (long-range) dispersion interactions in GGA functionals. Many dispersion correction schemes (which add an energy contribution on top of the PBE energy) are available in the literature, and we now present a comparison of such schemes paying special attention to the impact of many-body effects.

The most widely used dispersion schemes are variants of DFT-D as originally proposed by Grimme,[47] which add an $r^{-6}$ interaction term between all pairs of atoms, appropriately damped at short-range. The method has been refined by adjusting the damping (for instance Becke-Johnson damping D3(BJ)[24]) or by simply reparametrizing on a larger data set (D3M(BJ)[42]). Additionally, three-body interactions have been added using the expression derived by Axilrod−Teller−Muto (ATM).[55] The Henry coefficient and adsorption enthalpy of $CH_4$ in UiO-66 is reported in Table 3 for three variants of the D3 dispersion scheme. For the first column only two-body dispersion corrections are included, while for the second column the three-body ATM term is taken into account. The difference in adsorption enthalpy between PBE+D3 and PBE+D3M(BJ) is 2.5 kJ $mol^{-1}$, which can be considered significant considering that this means a deviation of more than 10%. The inclusion of the three-body term consistently leads to more repulsive adsorption energies, and the absolute value of the adsorption enthalpy decreases by about 2 kJ $mol^{-1}$. As a result, the Henry coefficient decreases by about 50% which can be rationalized by considering that $\exp(-\beta 2.0\ \text{kJ mol}^{-1}) \approx 0.45$ at room temperature. The experimental value for $K_H$ is 0.6 mol $kg^{-1}$ $bar^{-1}$,[56] so the inclusion of three-body dispersion interactions improves the agreement. We conclude that, even for a small guest molecule such as $CH_4$, three-body dispersion interactions have a non-negligible impact on adsorption.

**Table 3. Henry Coefficients [ mol $kg^{-1}$ $bar^{-1}$] and Enthalpies of Adsorption [ kJ $mol^{-1}$] of $CH_4$ in UiO-66 at 298 K Using Various Levels of Theory**

| | without ATM | | with ATM | |
|---|---|---|---|---|
| level of theory | $K_H$ | $\Delta H_{ads}$ | $K_H$ | $\Delta H_{ads}$ |
| PBE+D3 | 3.7 | −21.6 | 1.8 | −19.5 |
| PBE+D3(BJ) | 2.2 | −19.9 | 1.1 | −17.8 |
| PBE+D3M(BJ) | 1.7 | −19.1 | 0.8 | −17.1 |

The many-body dispersion (MBD) method was developed by Tkatchenko et al.[57] and also adds a van der Waals contribution to the underlying electronic structure calculation. Here, many-body effects up to infinite order are treated using the coupled fluctuating dipole model, which can be much better motivated from a physical point of view than the three-body corrections employed in DFT+$D^{ATM}$ schemes. The adsorption energies resulting from the MBD calculations were partitioned into contributions from second up to sixth order, which allowed to calculate the Henry coefficient and adsorption enthalpy including up to Nth order many-body effects for $N$ = 2, ..., 6. The resulting graph shown in Figure 5 reveals that, by adding three-body interactions, the Henry coefficient as well as the absolute value of the adsorption enthalpy decreases, indicating that three-body interactions are generally repulsive. By additionally including four-body interactions, the Henry coefficient increases again which means that this has a generally more attractive effect, in line with derivations from perturbation theory. Finally we note that omitting higher than two-body dispersion effects leads in this case to an overestimation of the Henry coefficient by a factor 2 with respect to the value obtained from the full inclusion of many-body effects up to infinite order. On the other hand, the



DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

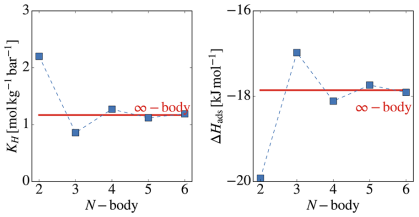



**Figure 5.** Henry coefficient (left) and adsorption enthalpy (right) of $CH_4$ in UiO-66 computed with PBE with MBD corrections, as a function of the order of included many-body effects (dotted lines are only present to guide the eye).

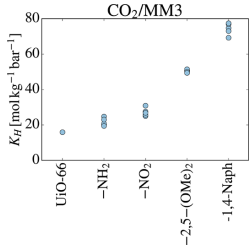



**Figure 7.** Henry coefficients for $CO_2$ at 298 K in different versions of the functionalized UiO-66 materials, computed with the MM3 force field.

value quickly converges as soon as three-body and certainly four-body interactions are treated. Again we note that the Henry coefficient computed with many-body dispersion interactions is closer to the experimental value of $K_H$ = 0.6 mol kg$^{-1}$ bar$^{-1}$ than the Henry coefficient computed with only two-body dispersion interactions.

**3.3. Linker Functionalizations in UiO-66.** In the previous section we demonstrated that adsorption properties are rather sensitive to the underlying PES that is used in the simulation. In many cases, especially high-throughput studies, one is more concerned with trends in properties rather than the absolute value of these properties. We therefore now turn our attention to a comparison of the Henry coefficients of $CO_2$, $CH_4$, and $N_2$ in a series of functionalized UiO-66 frameworks. The pristine UiO-66 MOF features benzene dicarboxylate (bdc) organic linkers. By functionalizing the benzene rings with amino ($-NH_2$), nitro ($-NO_2$), methoxy ($-2,5\text{-}OMe_2$), and naphthyl (−1,4-Naph) groups (as shown in Figure 6), we obtain four additional frameworks.

*3.3.1. Location of Functional Groups.* We considered functionalized versions of the UiO-66 by adding a functional group on each organic linker. Because these functional groups can be added on multiple locations of the linker, it is possible to construct many different functionalized frameworks with the same chemical formula. For each functional group, we constructed five different frameworks by placing the functionalization on a random position for each linker. This allows us to investigate whether the precise placement of functional groups has an important effect on the adsorption properties of the framework. We calculated the Henry coefficient of $CO_2$ in all frameworks with randomly placed functional groups using the MM3 force field as shown in Figure 7 (similar plots for other force fields and for $CH_4$ as the guest molecule are included in the Supporting Information). For each of the functional groups, there are five dots in this figure with each dot representing the Henry coefficient in one randomly constructed version of that specific functionalized framework. These results show that the precise placement of the functional groups certainly influences the adsorption properties, as there is a certain spread on the Henry coefficients for the different versions of each functionalized framework. At the same time, it should be realized that the spread is in general smaller than differences between different functional groups. In other words, if one is mostly interested in a comparison of different functionalized groups it suffices to study one specific placement for each group, as we have done here. Finally note that, when we compare results from various levels of theory, the same framework is considered, which enables a consistent comparison of different PES. A proper comparison with experiment is however more difficult, as it is often uncertain whether the simulated functionalized frameworks faithfully represent the ones used in experiment. Although Cmarik et al.[56] performed adsorption experiments on functionalized UiO-66 frameworks, these results will not be discussed here for this reason.

*3.3.2. Results.* As mentioned before, we focus on qualitative features in this section, and a comparison of force-field and ab initio results already reveals insights in this respect. We calculated the Henry coefficient of $CO_2$, $CH_4$, and $N_2$ at room temperature in all frameworks at the PBE+D3(BJ) and PBE+MBD level of theory using our importance sampling method with MEDFF as the reference potential ($N$ = 10 000 000 Widom insertions and $n$ = 5000 importance samples for $CH_4$ and $N_2$, $N$ = 50 000 000 Widom insertions and $n$ = 20 000 importance samples for $CO_2$). Additionally, the UFF values (a popular choice of PES for high-throughput studies) are calculated using Widom insertion and shown together with the ab initio results in Figure 8. The PBE+D3(BJ) and PBE+MBD results share a similar pattern, with the PBE+MBD Henry coefficients being generally lower. Following the discussion in the previous section, this can be attributed to the lack of (generally repulsive) many-body dispersion effects

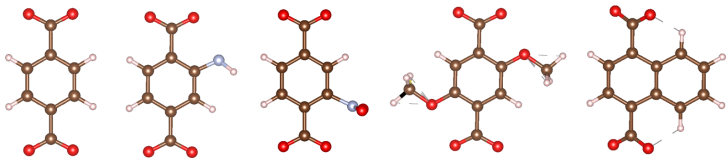


**Figure 6.** From left to right: bdc linker of pristine UiO-66, $-NH_2$, $-NO_2$, $-2,5\text{-}OMe_2$, and −1,4-Naph functionalized.



DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

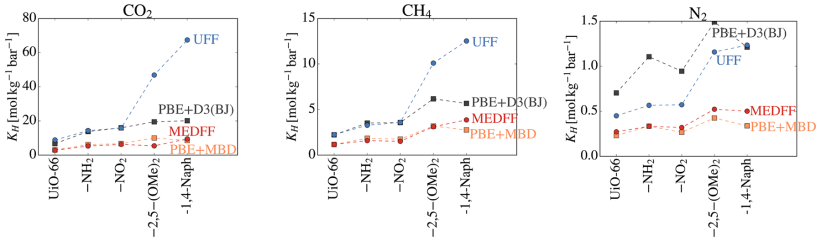



**Figure 8.** Henry coefficients at room temperature of $CO_2$, $CH_4$, and $N_2$ in functionalized versions of the UiO-66 MOF.

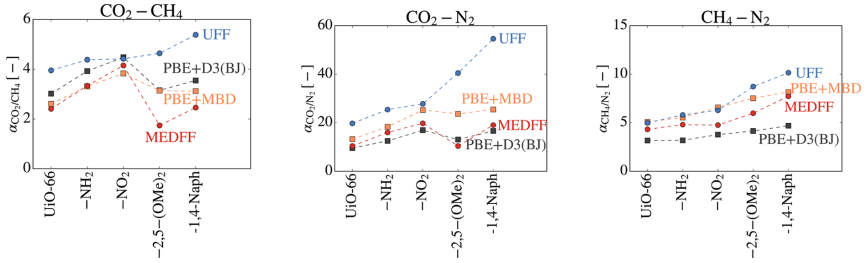



**Figure 9.** Selectivities of equimolar mixtures at room temperature in functionalized versions of the UiO-66 MOF.

in PBE+D3(BJ). For $CO_2$ and $CH_4$, the UFF results are in good agreement with the PBE+D3(BJ) results for UiO-66, UiO-66-$NH_2$, and UiO-66-$NO_2$. For UiO-66−2,5-$(OMe)_2$ and UiO-66−1,4-Naph, however, UFF predicts a steep increase in the adsorption which is not observed in the ab initio results. This indicates that some caution is warranted when using such a generic force field, even if only a qualitative comparison of some similar frameworks is made. MEDFF on the other hand is generally quite close to the ab initio values, especially if the PBE+MBD level of theory is considered as reference. Also trends predicted by MEDFF are in better agreement with the ab initio results.

We finally turn our attention to the selectivities of equimolar mixtures of $CO_2/CH_4$, $CO_2/N_2$, and $CH_4/N_2$, which in the low-pressure regime are simply provided as the ratios of the respective Henry coefficients. If a certain PES is systematically more attractive than another for all guest molecules, such deviations would compensate when looking at selectivities, which makes this an interesting property to look for qualitative differences between PES. Figure 9 reveals for instance that UiO-66-$NH_2$ and UiO-66-$NO_2$ show a higher $CO_2/CH_4$ selectivity than UiO-66−2,5-$(OMe)_2$ and UiO-66−1,4-Naph, according to both PBE+D3(BJ) and PBE+MBD. The generic UFF predicts exactly the opposite. Again this a clear example where one should be very careful when dealing with force-field results, even when simply extracting qualitative adsorption properties. For these cases, MEDFF provides patterns that are more similar to the ab initio curves, although some small inconsistencies are still present. This suggests that it might be worth considering ab initio based force fields (such as MEDFF) in favor of generic force fields (such as UFF) in high-throughput screening applications, despite the additional computational cost of the former PES.

## 4. CONCLUSIONS

In this work, a novel methodology to compute the Henry coefficient of small guest molecules in porous materials was developed, which allows to use ab initio evaluations of the PES. It comprises two stages: the first stage is a conventional Widom insertion simulation (using a computationally cheap force field PES), in the second stage ab initio adsorption energies are calculated only for configurations selected from the first stage using importance sampling. This approach enables an efficient evaluation of ab initio Henry coefficients, provided that force-field and ab initio adsorption energies correlate fairly well. It was demonstrated that MEDFF fulfills this condition in all investigated cases. MEDFF is a recently derived force field which is very robust because only a limited number of parameters are fitted to interaction energies from highly accurate CCSD(T)/CBS calculations. This robustness renders the current method applicable to a wide range of systems.

The adsorption of $CO_2$ in Mg-MOF-74 was investigated as a prototypical case where generic force fields fail to properly describe interactions between a polar guest molecule and open metal sites. Indeed, ab initio Henry coefficients were generally closer to experimental values than force-field results, but the detailed geometric features of the framework were shown to be important, thus hindering a fair comparison between experiment and simulation. Also the influence of the choice of level of theory was remarkably high, confirming the sensitivity of the Henry coefficient. More specifically it turned out that many-body dispersion effects have a significant impact, as shown for adsorption of $CH_4$ in the UiO-66 framework. Finally, qualitative features of adsorption in functionalizationed UiO-66 materials are discussed and here it is observed that generic force fields may predict dissimilar rankings compared to more advanced methods.



DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

In conclusion, we estimate that our methodology will help to investigate adsorption in porous materials whenever force fields are either not available or are not sufficiently accurate. Furthermore, it can also be a useful tool for benchmarking force fields to higher-level methods enabling a more judicious selection of a certain potential for the system at hand.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00892.

Details of computational settings, an explanation of the grid-based approach, a comparison of force-field and ab initio interaction energies, additional plots concerning the functionalized UiO-66 frameworks, a discussion on reparametrizing force fields, and numerical values for all Henry coefficients (PDF)

Geometries of all frameworks (ZIP)

Finally, scripts are included that demonstrate a Python implementation of the methodology (ZIP)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: toon.verstraelen@ugent.be.

### ORCID

Veronique Van Speybroeck: 0000-0003-2206-178X

Toon Verstraelen: 0000-0001-9288-5608


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 Programme/ERC Grant Agreement No. 64755 DYNPOR. We acknowledge the Foundation of Scientific Research-Flanders (FWO), the Research Board of Ghent University (BOF), and BELSPO in the frame of IAP/7/05 for their financial support. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation-Flanders (FWO) and the Flemish Government−department EWI.

## ■ REFERENCES


(1) Zhou, H.-C.; Long, J. R.; Yaghi, O. M. Introduction to Metal-Organic Frameworks. *Chem. Rev.* **2012**, *112*, 673−674.

(2) Li, H.; Wang, K.; Sun, Y.; Lollar, C. T.; Li, J.; Zhou, H.-C. Recent Advances in Gas Storage and Separation Using Metal-Organic Frameworks. *Mater. Today* **2018**, *21*, 108−121.

(3) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal-Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26*, 6185−6192.

(4) Li, S.; Chung, Y. G.; Snurr, R. Q. High-Throughput Screening of Metal-Organic Frameworks for $CO_2$ Capture in the Presence of Water. *Langmuir* **2016**, *32*, 10368−10376.

(5) Banerjee, D.; Simon, C. M.; Plonka, A. M.; Motkuri, R. K.; Liu, X.; Jian Chen; Smit, B.; Parise, J. B.; Haranczyk, M.; Thallapally, P. K. Metal-Organic Framework with Optimally Selective Xenon Adsorption and Separation. *Nat. Commun.* **2016**, *7*, ncomms11831.

(6) Rappé, A. K.; Casewit, C.; Colwell, K.; III Goddard, W. A.; Skiff, W. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(7) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.

(8) Düren, T.; Sarkisov, L.; Yaghi, O. M.; Snurr, R. Q. Design of New Materials for Methane Storage. *Langmuir* **2004**, *20*, 2683−2689.

(9) Yazaydın, A. Ö.; Snurr, R. Q.; Park, T.-H.; Koh, K.; Liu, J.; LeVan, M. D.; Benin, A. I.; Jakubczak, P.; Lanuza, M.; Galloway, D. B.; et al. Screening of Metal-Organic Frame-works for Carbon Dioxide Capture from Flue Gas Using a Combined Experimental and Modeling Approach. *J. Am. Chem. Soc.* **2009**, *131*, 18198−18199.

(10) Yang, Q.; Wiersum, A. D.; Llewellyn, P. L.; Guillerm, V.; Serre, C.; Maurin, G. Functionalizing Porous Zirconium Terephthalate UiO-66(Zr) for Natural Gas Upgrading: A Computational Exploration. *Chem. Commun.* **2011**, *47*, 9603−9605.

(11) Grajciar, L.; Rubeš, M.; Bludský, O.; Nachtigall, P. In *Metal-Organic Frameworks: Materials Modeling Towards Engineering Applications*; Jiang, J., Ed.; Pan Stanford, 2015; pp 175−206.

(12) Grajciar, L.; Nachtigall, P.; Bludský, O.; Rubeš, M. Accurate Ab Initio Description of Adsorption on Coordinatively Unsaturated $Cu^{2+}$ and Fe3+ Sites in MOFs. *J. Chem. Theory Comput.* **2015**, *11*, 230−238.

(13) Valenzano, L.; Civalleri, B.; Chavan, S.; Palomino, G. T.; Areán, C. O.; Bordiga, S. Computational and Experimental Studies on the Adsorption of CO, $N_2$, and $CO_2$ on Mg-MOF-74. *J. Phys. Chem. C* **2010**, *114*, 11185−11191.

(14) Kundu, A.; Piccini, G.; Sillar, K.; Sauer, J. Ab Initio Prediction of Adsorption Isotherms for Small Molecules in Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2016**, *138*, 14047−14056.

(15) Alonso, G.; Bahamon, D.; Keshavarz, F.; Giménez, X.; Gamallo, P.; Saýos, R. Density Functional Theory-Based Adsorption Isotherms for Pure and Flue Gas Mixtures on Mg-MOF-74. Application in $CO_2$ Capture Swing Adsorption Processes. *J. Phys. Chem. C* **2018**, *122*, 3945−3957.

(16) Lee, Y.; Poloni, R.; Kim, J. Probing Gas Adsorption in MOFs Using an Efficient Ab Initio Widom Insertion Monte Carlo Method. *J. Comput. Chem.* **2016**, *37*, 2808−2815.

(17) Peirs, J.; De Proft, F.; Baron, G.; Van Alsenoy, C.; Geerlings, P. Non-Empirical Quantum Chemical Calculation of Henry and Separation Constants and Heats of Adsorption for Diatomic Gases in Faujasite. *Chem. Commun.* **1997**, 531−532.

(18) Chen, L.; Grajciar, L.; Nachtigall, P.; Düren, T. Accurate Prediction of Methane Adsorption in a Metal-Organic Framework with Unsaturated Metal Sites by Direct Im-plementation of an Ab Initio Derived Potential Energy Surface in GCMC Simulation. *J. Phys. Chem. C* **2011**, *115*, 23074−23080.

(19) Mercado, R.; Vlaisavljevich, B.; Lin, L.-C.; Lee, K.; Lee, Y.; Mason, J. A.; Xiao, D. J.; Gonzalez, M. I.; Kapelewski, M. T.; Neaton, J. B.; et al. Force Field Development from Periodic Density Functional Theory Calculations for Gas Separation Applications Using Metal-Organic Frameworks. *J. Phys. Chem. C* **2016**, *120*, 12590−12604.

(20) Vandenbrande, S.; Verstraelen, T.; Gutiérrez-Sevillano, J. J.; Waroquier, M.; Van Speybroeck, V. Methane Adsorption in Zr-Based MOFs: Comparison and Critical Evaluation of Force Fields. *J. Phys. Chem. C* **2017**, *121*, 25309−25322.

(21) Frenkel, D.; Smit, B. In *Understanding Molecular Simulation*, 2nd ed.; Frenkel, D., Smit, B., Eds.; Academic Press: San Diego, 2002; pp 167−200 and 269−287.

(22) Widom, B. Some Topics in the Theory of Fluids. *J. Chem. Phys.* **1963**, *39*, 2808−2812.

(23) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(24) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(25) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B* **1993**, *47* (47), 558−561.




DOI: 10.1021/acs.jctc.8b00892
J. Chem. Theory Comput. 2018, 14, 6359−6369

(26) Kresse, G.; Hafner, J. Ab Initio Molecular-Dynamics Simulation of the Liquid-Metal-Amorphous-Semiconductor Transition in Germanium. *Phys. Rev. B* **1994**, *49*, 14251−14269.
(27) Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15−50.
(28) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54*, 11169−11186.
(29) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953−17979.
(30) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758−1775.
(31) Vandenbrande, S.; Waroquier, M.; Van Speybroeck, V.; Verstraelen, T. The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions. *J. Chem. Theory Comput.* **2017**, *13*, 161−179.
(32) Verstraelen, T.; Vandenbrande, S.; Heidar-Zadeh, F.; Vanduyfhuys, L.; Van Speybroeck, V.; Waroquier, M.; Ayers, P. W. Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development. *J. Chem. Theory Comput.* **2016**, *12*, 3894−3912.
(33) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. Molecular Mechanics. the MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.
(34) Martin, M. G.; Siepmann, J. I. Transferable Potentials for Phase Equilibria. 1. United- Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569−2577.
(35) Potoff, J. J.; Siepmann, J. I. Vapor-Liquid Equilibria of Mixtures Containing Alkanes, Carbon Dioxide, and Nitrogen. *AIChE J.* **2001**, *47*, 1676−1682.
(36) Vanduyfhuys, L.; Vandenbrande, S.; Rogge, S. M. J.; Wieme, J.; Verstraelen, T. YAFF 1.4.1, 2017; http://www.github.com/molmod/yaff/.
(37) Yu, D.; Yazaydin, A. O.; Lane, J. R.; Dietzel, P. D. C.; Snurr, R. Q. a Combined Experimental and Quantum Chemical Study of $CO_2$ Adsorption in the Metal-Organic Framework CPO-27 with Different Metals. *Chem. Sci.* **2013**, *4*, 3544−3556.
(38) Mann, G. W.; Lee, K.; Cococcioni, M.; Smit, B.; Neaton, J. B. First-Principles Hubbard U Approach for Small Molecule Binding in Metal-Organic Frameworks. *J. Chem. Phys.* **2016**, *144*, 174104.
(39) Grimme, S. Dftd3 3.2, 2016; https://www.chemie.uni-Bonn.de/pctc/mulliken-Center/software/dft-D3/get-The-Current-Version-Of-Dft-D3.
(40) Vlaisavljevich, B.; Huck, J.; Hulvey, Z.; Lee, K.; Mason, J. A.; Neaton, J. B.; Long, J. R.; Brown, C. M.; Alfè, D.; Michaelides, A.; et al. Performance of Van Der Waals Corrected Functionals for Guest Adsorption in the M2(dobdc) Metal-Organic Frameworks. *J. Phys. Chem. A* **2017**, *121*, 4139−4151.
(41) Becker, T. M.; Heinen, J.; Dubbeldam, D.; Lin, L.-C.; Vlugt, T. J. H. Polarizable Force Fields for $CO_2$ and $CH_4$ Adsorption in M-MOF-74. *J. Phys. Chem. C* **2017**, *121*, 4659−4673.
(42) Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *J. Phys. Chem. Lett.* **2016**, *7*, 2197−2203.
(43) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van Der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.
(44) Román-Pérez, G.; Soler, J. M. Efficient Implementation of a Van Der Waals Density Functional: Application to Double-Wall Carbon Nanotubes. *Phys. Rev. Lett.* **2009**, *103*, 096102.
(45) Lee, K.; Murray, É. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-Accuracy Van Der Waals Density Functional. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 081101.
(46) Klimeš, J.; Bowler, D. R.; Michaelides, A. Chemical Accuracy for the Van Der Waals Density Functional. *J. Phys.: Condens. Matter* **2010**, *22*, 022201.
(47) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(48) Sillar, K.; Kundu, A.; Sauer, J. Ab Initio Adsorption Isotherms for Molecules with Lateral Interactions: $CO_2$ in Metal-Organic Frameworks. *J. Phys. Chem. C* **2017**, *121*, 12789−12799.
(49) Dietzel, P. D. C.; Besikiotis, V.; Blom, R. Application of Metal-Organic Frameworks with Coordinatively Unsaturated Metal Sites in Storage and Separation of Methane and Carbon Dioxide. *J. Mater. Chem.* **2009**, *19*, 7362−7370.
(50) Britt, D.; Tranchemontagne, D.; Yaghi, O. M. Metal-Organic Frameworks with High Capacity and Selectivity for Harmful Gases. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11623−11627.
(51) Mason, J. A.; Sumida, K.; Herm, Z. R.; Krishna, R.; Long, J. R. Evaluating Metal-Organic Frameworks for Post-Combustion Carbon Dioxide Capture Via Temperature Swing Adsorption. *Energy Environ. Sci.* **2011**, *4*, 3030−3040.
(52) Queen, W. L.; Hudson, M. R.; Bloch, E. D.; Mason, J. A.; Gonzalez, M. I.; Lee, J. S.; Gygi, D.; Howe, J. D.; Lee, K.; Darwish, T. A.; et al. Comprehensive Study of Carbon Dioxide Adsorption in the Metal-Organic Frameworks $M_2$(dobdc) (M = Mg, Mn, Fe, Co, Ni, Cu, Zn). *Chem. Sci.* **2014**, *5*, 4569−4581.
(53) Cavka, J. H.; Jakobsen, S.; Olsbye, U.; Guillou, N.; Lamberti, C.; Bordiga, S.; Lillerud, K. P. A New Zirconium Inorganic Building Brick Forming Metal Organic Frameworks with Exceptional Stability. *J. Am. Chem. Soc.* **2008**, *130*, 13850−13851.
(54) Hermann, J. Pymbd, 2017; https://github.com/azag0/pymbd.
(55) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *ab initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(56) Cmarik, G. E.; Kim, M.; Cohen, S. M.; Walton, K. S. Tuning the Adsorption Properties of UiO-66 Via Ligand Functionalization. *Langmuir* **2012**, *28*, 15606−15613.
(57) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body Van Der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.



DOI: 10.1021/acs.jctc.8b00892
*J. Chem. Theory Comput.* 2018, 14, 6359−6369

# B

# Computational Details

This appendix provides the computational details concerning unpublished results included in this thesis.

## B.1 Path-integral Molecular Dynamics

PIMD simulations of liquid methane where performed using the i-PI[176, 177] framework for the path-integral sampling, interfaced to YAFF to evaluate the PES as explained in Appendix C. The Bussi-Zykova-Parrinello barostat (only allowing isotropic cell fluctuations) with a time constant of $1000\,\mathrm{fs}$ is used to control the pressure. The barostat particles are coupled to a Langevin thermostat with time constant $100\,\mathrm{fs}$ while the atoms are thermostatted using a local centroid PILE thermostat with time constant $100\,\mathrm{fs}$. A time step of $0.25\,\mathrm{fs}$ is used and each simulation lasts at least $2\,\mathrm{ns}$, considering the first $0.5\,\mathrm{ns}$ as equilibration and the remainder to collect results.

# C

# Software Development

In the context of this Ph.D. a substantial amount of time was devoted to software development. Below, an overview of the different simulation packages to which contributions were made is provided.

### Yaff

Yaff is a pythonic force-field code developed at the Center for Molecular Modeling. Its primary intention was to facilitate the implementation of newly developed methods, and therefore it is very flexible and thoroughly tested. Many features were implemented during this Ph.D. , for example the addition of the pair potentials used by MEDFF. Yaff is distributed as open source software under the conditions of the GPL license version 3 and is available at http://www.github.com/molmod/yaff/.

### LAMMPS

LAMMPS is a classical molecular dynamics code, and an acronym for Large–scale Atomic/Molecular Massively Parallel Simulator. LAMMPS is an open-source code, distributed freely under the terms of the GNU Public License (GPL) and is available at http://www.github.com/lammps/lammps/. Some bugfixes were contributed, which were discovered by comparison with Yaff, which could be considered as a reference implementation thanks to its extensive unit tests.

Another contribution relates to the coupling of LAMMPS and Yaff. As stated before, the main design principles of Yaff were extensibility and code-readability. This unfortunately results in code that executes rather slowly when compared to a highly-optimized (parallel) package such as LAMMPS. In most classical molecular dynamics simulations, the computational bottleneck is the evaluation of noncovalent interactions. Therefore the best properties of both Yaff and LAMMPS were combined: Yaff is used as the main driver (taking care of the sampling and optionally evaluation of covalent interactions), while the computationally most intensive parts are outsourced to LAMMPS. This allowed the simulation of larger systems on longer time scales, which would not have been possible using only Yaff.[60, 63, 226]

### i-PI

i-PI is a Python interface for path integral molecular dynamics simulations. i-PI itself only propagates (path integral) dynamics of the nuclei and relies on an external code for the evaluation of energies, forces and virial tensor. An interface that allows using Yaff as the external code was implemented, allowing path integral simulation employing in-house developed force fields.[227] i-PI is available at http://www.github.com/i-pi/i-pi/.

### QuickFF

QuickFF is a software package developed at the Center for Molecular Modeling, designed to derive accurate force fields for isolated and complex molecular systems in a quick and easy manner.[57, 58] An important contribution to this package was made by extending it to handle periodic systems. This was achieved by a coupling with Yaff to evaluate force field energies while constructing them. QuickFF is distributed as open source software under the conditions of the GPL at license version 3 at http://www.github.com/molmod/QuickFF/.

### HORTON

HORTON is a Helpful Open-source Research TOol for N-fermion systems. HORTON is not intended to be a replacement for existing quantum chemistry software, but a helpful supplement to it. The ambition is to provide a research tool that is computationally-efficient enough to be helpful, without compromising code-readability and user-friendliness. Contributions in the context of this Ph.D. include the implementation of the Poisson solver.[72] HORTON is distributed under the conditions of the GPL License version 3 (GPLv3) at http://theochem.github.io/horton/.

# D

# List of Publications

## Publications in international peer–reviewed journals

1. **Modeling Gas Adsorption in Flexible Metal–Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes**
   Sven M.J. Rogge†, Ruben Goeminne†, Ruben Demuynck, Juan José Gutiérrez-Sevillano, Steven Vandenbrande, Louis Vanduyfhuys, Michel Waroquier, Toon Verstraelen, Veronique Van Speybroeck
   *Adv. Theory Simul.*, **2019**, 2 (4), 1800177
   †authors contributed equally

2. **Ab Initio Evaluation of Henry Coefficients Using Importance Sampling**
   Steven Vandenbrande, Michel Waroquier, Toon Verstraelen, Veronique Van Speybroeck
   *J. Chem. Theory Comput.*, **2018**, *14* (12), 6359-6369

3. **i-PI 2.0: A Universal Force Engine for Advanced Molecular Simulations**
   V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R.H. Meißner, D. Wilkins, P. Juda, S.P.

Bienvenue, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. Kühne, D. Manolopoulos, T. Markland, J. Richardson, A. Tkatchenko, G.A. Tribello, V. Van Speybroeck, M. Ceriotti
*Comput. Phys. Commun.*, **2018**, *236*, 214-223

4. **Extension of the QuickFF Force Field Protocol for an Improved Accuracy of Structural, Vibrational, Mechanical and Thermal Properties of Metal–Organic Frameworks**
Louis Vanduyfhuys, Steven Vandenbrande, Jelle Wieme, Michel Waroquier, Toon Verstraelen, and Veronique Van Speybroeck
*J. Comput. Chem*, **2018**, *39* (16), 999–1011

5. **Influence of a Confined Methanol Solvent on the Reactivity of Active Sites in UiO–66**
Chiara Caratelli, Julianna Hajek, Sven M.J. Rogge, Steven Vandenbrande, Evert Jan Meijer, Michel Waroquier, and Veronique Van Speybroeck
*ChemPhysChem*, **2018**, *19* (4), 420–429

6. **Thermodynamic Insight into Stimuli-Responsive Behaviour of Soft Porous Crystals**
Louis Vanduyfhuys, Sven M.J. Rogge, Jelle Wieme, Steven Vandenbrande, Guillaume Maurin, Michel Waroquier, and Veronique Van Speybroeck
*Nat. Commun.*, **2018**, *9* (1), 204

7. **Methane Adsorption in Zr-Based MOFs: Comparison and Critical Evaluation of Force Fields**
Steven Vandenbrande, Toon Verstraelen, Juan José Gutiérrez-Sevillano, Michel Waroquier, and Veronique Van Speybroeck
*J. Phys. Chem. C*, **2017**, *121* (45), 25309–25322

8. **Alternating Copolymer of Double Four Ring Silicate and Dimethyl Silicone Monomer - PSS-1**
Sam Smet, Steven Vandenbrande, Pieter Verlooy, Stef Kerkhofs, Eric Breynaert, Christine E.A. Kirschhock, Charlotte Martineau, Francis Taulelle, Veronique Van Speybroeck, and Johan A. Martens

*Chem. Eur. J*, **2017**, *23* (47), 11286–11293

9. **The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Non-Covalent Interactions**
   Steven Vandenbrande, Michel Waroquier, Veronique Van Speybroeck, and Toon Verstraelen
   *J. Chem. Theory Comput.*, **2017**, *13* (1), 161–179

10. **Thermodynamic Insight in the High-Pressure Behavior of UiO-66: Effect of Linker Defects and Linker Expansion**
    Sven M.J. Rogge, Jelle Wieme, Louis Vanduyfhuys, Steven Vandenbrande, Guillaume Maurin, Toon Verstraelen, Michel Waroquier, and Veronique Van Speybroeck
    *Chem. Mater.*, **2016**, *28* (16), 5721–5732

11. **Minimal Basis Iterative Stockholder: Atoms-in-Molecules for Force-Field Development**
    Toon Verstraelen, Steven Vandenbrande, Farnaz Heidar-Zadeh, Louis Vanduyfhuys, Veronique Van Speybroeck, Michel Waroquier, and Paul W. Ayers
    *J. Chem. Theory Comput.*, **2016**, *12* (8), 3894–3912

12. **QuickFF: A program for a quick and easy derivation of force fields for Metal–Organic Frameworks from ab initio input**
    Louis Vanduyfhuys, Steven Vandenbrande, Toon Verstraelen, Rochus Schmid, Michel Waroquier, and Veronique Van Speybroeck
    *J. Comput. Chem*, **2015**, *36* (13), 1015–1027

13. **Direct Computation of Parameters for Accurate Polarizable Force Fields**
    Toon Verstraelen, Steven Vandenbrande, and Paul W. Ayers
    *J. Chem. Phys*, **2014**, *141*, 194144

# Conference contributions

## Oral presentations

1. **Towards modeling long-range disorder in MOFs: Development of a computational toolbox to extend the length scale in molecular simulations**
   Sven M.J. Rogge, Ruben Demuynck, Arhur De Vos, Kevin Hendrickx, Kurt Lejaeghere, Guillaume Maurin, Steven Vandenbrande, Sander Vandenhaute, Pascal Van der Voort, Louis Vanduyfhuys, Toon Verstraelen, Michel Waroquier, Jelle Wieme, Veronique Van Speybroeck
   *CECAM workshop: Multi-scale modelling of flexible and disordered porous materials*, Paris, France, June 11–13, 2018

2. **Towards a molecular level understanding of chemical and physical phenomena in metal–organic frameworks**
   Jelle Wieme, Chiara Caratelli, Ruben Demuynck, Arhur De Vos, Julianna Hajek, A.E.J. Hoffman, Aran Lamaire, Kurt Lejaeghere, Sven M.J. Rogge, Steven Vandenbrande, Louis Vanduyfhuys, Michel Waroquier, Veronique Van Speybroeck
   *Congrès français des MOFs*, Paris, France, May 16–17, 2018

3. **A Full Computational Characterization of the Impact of Defects on the Mechanical Stability and Gas Separation Ability of MOFs**
   Sven M.J. Rogge, Jelle Wieme, Ruben Demuynck, Louis Vanduyfhuys, Steven Vandenbrande, Toon Verstraelen, Michel Waroquier, Guillaume Maurin, and Veronique Van Speybroeck
   *docMOF Symposium on Porous Coordination Polymers and Metal–Organic Frameworks*, Munich, Germany, April 29 – May 2, 2018

4. **Thermodynamic Insight into Stimuli–Responsive Behavior of Soft Porous Crystals**
   Louis Vanduyfhuys, Sven M.J. Rogge, Jelle Wieme, Steven Vandenbrande, Guillaume Maurin, Michel Waroquier, and Veronique Van Speybroeck
   *EuroMOF 2017*, Delft, The Netherlands, October 29 – November 1, 2017

5. **A Thermodynamic Approach to Accurately Determine the Flexibility and Loss of Crystallinity in Metal–Organic Frameworks**
   Sven M.J. Rogge, Jelle Wieme, Louis Vanduyfhuys, Steven Vandenbrande, Guillaume Maurin, Toon Verstraelen, Michel Waroquier, and Veronique Van Speybroeck

*EUROMAT 2017*, Thessaloníki, Greece, September 17–22, 2017

6. **A Thermodynamic Characterization of Mechanical, Thermal and Adsorption Properties of Flexible Metal–Organic Frameworks**
   Louis Vanduyfhuys, Sven M.J. Rogge, Jelle Wieme, Steven Vandenbrande, Michel Waroquier, and Veronique Van Speybroeck
   *EUROMAT 2017*, Thessaloníki, Greece, September 17–22, 2017

7. **A Thermodynamic Approach to Accurately Determine Pressure Profiles for Metal–Organic Frameworks to Assess Their Flexibility and Loss of Crystallinity**
   Sven M.J. Rogge, Jelle Wieme, Louis Vanduyfhuys, Steven Vandenbrande, Guillaume Maurin, Toon Verstraelen, Michel Waroquier, and Veronique Van Speybroeck
   *Thermodynamics 2017*, Edinburgh, United Kingdom, September 5–8, 2017

8. **Insights in the Behaviour of MOFs Through Molecular Modeling: From Force Field Derivation to Thermodynamic Analysis**
   Louis Vanduyfhuys, Steven Vandenbrande, Ruben Demuynck, Sven M.J. Rogge, Jelle Wieme, An Ghysels, Toon Verstraelen, Guillaume Maurin, Rochus Schmid, Michel Waroquier, and Veronique Van Speybroeck
   *DAMP*, Montpellier, France, November 17, 2016

9. **Constructing complete non-covalent force fields based on ab initio monomer densities**
   Steven Vandenbrande, Louis Vanduyfhuys, Michel Waroquier, Veronique Van Speybroeck, and Toon Verstraelen
   *Many-Body Interactions: From Quantum Mechanics to Force Fields*, Telluride (CO), US, July 12–16, 2016

10. **Modeling Charge-Transfer Interactions with the ACKS2 Model**
    Toon Verstraelen, Steven Vandenbrande, Juan José Gutiérrez-Sevillano, and Paul W. Ayers
    *Density- and response density-based models for Intermolecular Interactions in Molecular Assemblies and in Solids*, Nancy, France, June 20–23, 2016

11. **Constructing complete noncovalent force fields based on ab initio monomer densities**
    Steven Vandenbrande, Louis Vanduyfhuys, Michel Waroquier,

Veronique Van Speybroeck, and Toon Verstraelen
*Solvay Workshop - Conceptual Quantum Chemistry: Present Aspects and Challenges for the Future*, Brussels, Belgium, April 5–8, 2016

12. **Force fields derived from electronic structure computations**
Toon Verstraelen, Steven Vandenbrande, Louis Vanduyfhuys, and Veronique Van Speybroeck
*IAP 7/05, PL3 Meeting, Multiscale modeling and model guided design*, Ghent, Belgium, March 1, 2016

13. **Minimal Basis Iterative Stockholder: a Self–Consistent Hirshfeld Method Tailored to Force–Field Development**
Toon Verstraelen, Steven Vandenbrande, Louis Vanduyfhuys, and Veronique Van Speybroeck
*Chemical Physics Symposium*, Waterloo (ON), Canada, November 6–8, 2015

14. **Unraveling the Breathing in Metal–Organic Frameworks: From Force Fields to Thermodynamic Insights**
Louis Vanduyfhuys, Steven Vandenbrande, Ruben Demuynck, Sven M.J. Rogge, An Ghysels, Toon Verstraelen, Michel Waroquier, and Veronique Van Speybroeck
*FEMS EUROMAT 2015*, Warsaw, Poland, September 20–24, 2015

15. **Recent Developments in Hirshfeld Density Partitioning Methods**
Toon Verstraelen, Steven Vandenbrande, Louis Vanduyfhuys, Veronique Van Speybroeck
*FEMS EUROMAT 2015*, Warsaw, Poland, September 20, 2015

16. **Understanding the Flexibility in Metal–Organic Frameworks: From Molecular to Thermodynamic Insights**
Louis Vanduyfhuys, Steven Vandenbrande, Ruben Demuynck, Sven M.J. Rogge, An Ghysels, Toon Verstraelen, Michel Waroquier, and Veronique Van Speybroeck
*CHITEL 2015*, Turin, Italy, July 26–31, 2015

17. **Unraveling the Breathing in Metal–Organic Frameworks: From Force Fields to Thermodynamic Insights**
Louis Vanduyfhuys, Toon Verstraelen, An Ghysels, Steven Vandenbrande, Sven M.J. Rogge, Ruben Demuynck, Rochus Schmid, Michel Waroquier, and Veronique Van Speybroeck
*IAP–PL3 Meeting*, Mons, Belgium, February 5, 2015

18. **How to Ensure the Accuracy of Polarizable Force Fields?**
    Toon Verstraelen, Steven Vandenbrande, and Paul W. Ayers
    *Visit ALGC, Free University of Brussels (VUB)*, Brussels, Belgium, November 21, 2014

19. **Modeling Electrostatic Penetration Effects with Atoms in Molecules**
    Toon Verstraelen, Steven Vandenbrande, Louis Vanduyfhuys, Veronique Van Speybroeck, and Michel Waroquier
    *Symposium on Molecular Electrostatic Potentials*, Brussels, Belgium, October 22, 2014

20. **An efficient protocol to derive reliable additive nonbonding force fields**
    Toon Verstraelen, Steven Vandenbrande, Louis Vanduyfhuys, and Veronique Van Speybroeck
    *WATOC 2014 Satellite Meeting on Large Condensed and Biological Systems*, Concepción, Chili, October 13–14, 2014

21. **QuickFF: Toward a Generally Applicable Methodology to Quickly Derive Accurate Force Fields for Metal–Organic Frameworks from Ab Initio Input**
    Louis Vanduyfhuys, Toon Verstraelen, Steven Vandenbrande, Michel Waroquier, and Veronique Van Speybroeck
    *Telluride Workshop on Many–Body Interactions: From Quantum Mechanics to Force Fields*, Telluride (CO), US, June 15–19, 2014

22. **Improved force–field electrostatics: decomposition of ab initio interaction energies**
    Steven Vandenbrande,
    *IAP - PL3 meeting (P7/05)*, Ghent, Belgium, February 7, 2014

23. **COK–18, A chain–like POSiSil**
    Sam Smet, Steven Vandenbrande, Pieter Verlooy, Lennart Joos, Louis Vanduyfhuys, Toon Verstraelen, Veronique Van Speybroeck, Johan A. Martens
    *Annual IAP - WP2 Meeting (P7/05)*, Gent, Belgium, February 5, 2014

## Poster presentations

1. **Influence of a Confined Methanol Solvent on the Reactivity of Active Sites in UiO–66**

Chiara Caratelli, Julianna Hajek, Sven M.J. Rogge, Steven Vandenbrande, Evert Jan Meijer, Michel Waroquier, and Veronique Van Speybroeck

*NCCC XIX*, Noordwijkerhout, The Netherlands, March 5–7, 2018

2. **Assessing the Flexibility and Loss of Crystallinity in Metal–Organic Frameworks From a Thermodynamic Point of View**

   Sven M.J. Rogge, Jelle Wieme, Louis Vanduyfhuys, Steven Vandenbrande, Guillaume Maurin, Toon Verstraelen, Michel Waroquier, and Veronique Van Speybroeck

   *EuroMOF 2017*, Delft, The Netherlands, October 29 – November 1, 2017

3. **Ab initio computed AIM densities as the basis for complete noncovalent force fields**

   Steven Vandenbrande, Michel Waroquier, Louis Vanduyfhuys, Veronique Van Speybroeck, and Toon Verstraelen

   *Density- and response density-based models for Intermolecular Interactions in Molecular Assemblies and in Solids*, Nancy, France, June 20–23, 2016

4. **Investigating the Heat Capacity of Metal–Organic Frameworks via Path Integral Molecular Dynamics Simulations**

   Jelle Wieme, Louis Vanduyfhuys, Sven M.J. Rogge, Steven Vandenbrande, and Veronique Van Speybroeck

   *CECAM/Psi–k School on Path Integral Quantum Mechanics: Theory, Simulation and Application*, Lausanne, Switzerland, June 13–17, 2016

5. **Extracting complete noncovalent force fields from ab initio computed electron densities**

   Steven Vandenbrande, Toon Verstraelen, Louis Vanduyfhuys, Michel Waroquier, and Veronique Van Speybroeck

   *Quantum Chemistry in Belgium 12*, Leuven, Belgium, February 16, 2016

6. **Minimal Basis Iterative Stockholder (MBIS) Atoms-in-Molecules for Force-Field Development**

   Toon Verstraelen, Steven Vandenbrande, Louis Vanduyfhuys, Michel Waroquier, and Veronique Van Speybroeck

   *Quantum Chemistry in Belgium 12*, Leuven, Belgium, February 16, 2016

7. **Extracting complete noncovalent force fields from ab initio computed electron densities**

   Steven Vandenbrande, Toon Verstraelen, Louis Vanduyfhuys, Michel Waroquier, and Veronique Van Speybroeck

*Euromat 2015*, Warschau, Poland, September 21–23, 2015

8. **Minimal Basis Iterative Stockholder: a Concise and Reliable Representation of the Electron Density**
   Toon Verstraelen, Louis Vanduyfhuys, Steven Vandenbrande, Michel Waroquier, and Veronique Van Speybroeck
   *Annual IAP Meeting*, Hasselt, Belgium, September 11, 2015

9. **Modeling Breathing of MIL–53: From Force Fields to Thermodynamic Insights**
   Louis Vanduyfhuys, Toon Verstraelen, An Ghysels, Steven Vandenbrande, Ruben Demuynck, Sven M.J. Rogge, Rochus Schmid, Guillaume Maurin, Michel Waroquier, and Veronique Van Speybroeck
   *Flexibility and Disorder in Metal–Organic Frameworks*, Paris, France, June 3–5, 2015

10. **Complete non-bonding force field derived from monomer electron densities**
    Steven Vandenbrande, Toon Verstraelen, Louis Vanduyfhuys, Michel Waroquier, and Veronique Van Speybroeck
    *Workshop on Force Fields: From Atoms to Materials*, Jülich, Germany, November 3–5, 2014

11. **QuickFF: Toward a generally applicable methodology to quickly derive accurate force fields for Metal–Organic Frameworks from ab initio input**
    Louis Vanduyfhuys, Steven Vandenbrande, Toon Verstraelen, Rochus Schmid, Michel Waroquier, and Veronique Van Speybroeck
    *Workshop on Force Fields: From Atoms to Materials*, Jülich, Germany, November 3–5, 2014

12. **Ab initio simulation of the IR spectrum of COK–18, a chain–like POSiSil**
    Steven Vandenbrande, Pieter Verlooy, Sam Smet, Lennart Joos, Louis Vanduyfhuys, Toon Verstraelen, Johan A. Martens, and Veronique Van Speybroeck
    *Meeting of the Dutch Zeolite Association: on Hybrids and zeolites*, Ghent, Belgium, October 7, 2014

13. **An efficient protocol to derive reliable additive nonbonding force fields**
    Toon Verstraelen, Steven Vandenbrande, Louis Vanduyfhuys, and Veronique Van Speybroeck

*WATOC 2014*, Santiago de Chile, Chile, October 5–10, 2014

14. **An efficient protocol to derive reliable additive nonbonding force fields**
    Toon Verstraelen, Steven Vandenbrande, Louis Vanduyfhuys, and Veronique Van Speybroeck
    *Annual IAP Meeting (P7/05)*, Louvain-La-Neuve, Belgium, September 19, 2014

15. **Ab initio simulation of the IR spectrum of COK-18, a chain-like POSiSil**
    Steven Vandenbrande, Pieter Verlooy, Sam Smet, Lennart Joos, Louis Vanduyfhuys, Toon Verstraelen, Johan A. Martens, and Veronique Van Speybroeck
    *Annual IAP Meeting (P7/05)*, Louvain-La-Neuve, Belgium, September 19, 2014

16. **QuickFF: a generally applicable methodology to quickly derive accurate force fields from ab initio input**
    Louis Vanduyfhuys, Toon Verstraelen, Steven Vandenbrande, Michel Waroquier, and Veronique Van Speybroeck
    *Annual IAP Meeting*, Ghent, Belgium, September 18, 2013

# Bibliography


[1] E. Schrödinger, "An Undulatory Theory of the Mechanics of Atoms and Molecules," *Phys. Rev.*, **28**, 1049 (1926).

[2] B. Bransden, C. Joachain, and T. Plivier, *Physics of Atoms and Molecules*, Pearson Education (Prentice Hall, 2003).

[3] M. Born and R. Oppenheimer, "Zur Quantentheorie Der Molekeln," *Ann. Phys.*, **389**, 457 (1927).

[4] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Computational science series (Elsevier Science, 2001).

[5] M. E. Tuckerman and A. Hughes, *Classical and Quantum Dynamics in Condensed Phase Simulations* (World Scientific, 1998).

[6] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory* (John Wiley & Sons, LTD, Chichester, 2000).

[7] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, Cambridge Molecular Science (Cambridge University Press, 2009).

[8] A. E. DePrince and C. D. Sherrill, "Accuracy and Efficiency of Coupled-Cluster Theory Using Density Fitting/Cholesky Decomposition, Frozen Natural Orbitals, and a t1-Transformed Hamiltonian," *J. Chem. Theory Comput.*, **9**, 2687 (2013).

[9] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Phys. Rev.*, **136**, B864 (1964).

[10] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecule*, International Series of Monographs on Chemistry (Oxford University Press, USA, 1994).

[11] J. P. Perdew and K. Schmidt, "Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy," *AIP Conference Proceedings*, **577**, 1 (2001).

[12] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, "Van Der Waals Density Functional for General Geometries," *Phys. Rev. Lett.*, **92**, 246401 (2004).

[13] S. Grimme, "Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction," *J. Comput. Chem.*, **27**, 1787 (2006).

[14] J. Wellendorff, K. T. Lundgaard, A. Møgelhøj, V. Petzold, D. D. Landis, J. K. Nørskov, T. Bligaard, and K. W. Jacobsen, "Density Functionals for surface science: Exchange-Correlation Model Development with Bayesian Error Estimation," *Phys. Rev. B*, **85**, 235149 (2012).

[15] H. Peng, Z.-H. Yang, J. P. Perdew, and J. Sun, "Versatile Van Der Waals Density Functional Based on a Meta-Generalized Gradient Approximation," *Phys. Rev. X*, **6**, 041005 (2016).

[16] A. Tkatchenko and M. Scheffler, "Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data," *Phys. Rev. Lett.*, **102**, 073005 (2009).

[17] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A Consistent and Accurate *ab initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu," *J. Chem. Phys.*, **132**, 154104 (2010).

[18] A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler, "Accurate and Efficient Method for Many-Body Van Der Waals Interactions," *Phys. Rev. Lett.*, **108**, 236402 (2012).

[19] J. VandeVondele, U. Borštnik, and J. Hutter, "Linear Scaling Self-Consistent Field Calculations with Millions of Atoms in the Condensed Phase," *J. Chem. Theory Comput.*, **8**, 3565 (2012).

[20] S. M. J. Rogge, A. Bavykina, J. Hajek, H. García, A. I. Olivos-Suarez, A. Sepúlveda-Escribano, A. Vimont, G. Clet, P. Bazin, F. Kapteijn, M. Daturi, E. V. Ramos-Fernández, F. X. Llabrés i Xamena, V. Van Speybroeck, and J. Gascon, "Metal–Organic and Covalent Organic Frameworks as Single-Site Catalysts," *Chem. Soc. Rev.*, **46**, 3134 (2017).

[21] P. Xu, E. B. Guidez, C. Bertoni, and M. S. Gordon, "Perspective: Ab Initio Force Field Methods Derived from Quantum Mechanics," *J. Chem. Phys.*, **148**, 090901 (2018).

[22] A. Stone, *The Theory of Intermolecular Forces*, 2nd ed. (Oxford University Press, 2013) Chap. 4.

[23] A. Stone, *The Theory of Intermolecular Forces*, 2nd ed. (Oxford University Press, 2013) Chap. 6.

[24] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, **21**, 1087 (1953).

[25] S. L. Mayo, B. D. Olafson, and W. A. Goddard, "DREIDING: A Generic Force Field for Molecular Simulations," *J. Phys. Chem.*, **94**, 8897 (1990).

[26] J. Gasteiger and M. Marsili, "Iterative Partial Equalization of Orbital Electronegativity-a Rapid Access to Atomic Charges," *Tetrahedron*, **36**, 3219 (1980).

[27] D. E. Williams and S. R. Cox, "Nonbonded Potentials for Azahydrocarbons: The Importance of the Coulombic Interaction," *Acta Crystallogr. Section B*, **40**, 404 (1984).

[28] A. K. Rappé, C. Casewit, K. Colwell, W. A. Goddard III, and W. Skiff, "UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations," *J. Am. Chem. Soc.*, **114**, 10024 (1992).

[29] A. K. Rappé and W. A. Goddard, "Charge Equilibration for Molecular Dynamics Simulations," *J. Phys. Chem.*, **95**, 3358 (1991).

[30] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, "A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins," *J. Am. Chem. Soc.*, **106**, 765 (1984).

[31] A. T. Hagler, E. Huler, and S. Lifson, "Energy Functions for Peptides and Proteins. I. Derivation of a Consistent Force Field Including the Hydrogen Bond from Amide Crystals," *J. Am. Chem. Soc.*, **96**, 5319 (1974).

[32] N. L. Allinger, Y. H. Yuh, and J.-H. Lii, "Molecular Mechanics. the MM3 Force Field for Hydrocarbons. 1," *J. Am. Chem. Soc.*, **111**, 8551 (1989).

[33] J. H. Lii and N. L. Allinger, "Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 3. the Van Der Waals' Potentials and Crystal Data for Aliphatic and Aromatic Hydrocarbons," *J. Am. Chem. Soc.*, **111**, 8576 (1989).

[34] N. L. Allinger, F. Li, and L. Yan, "Molecular Mechanics. The MM3 Force Field for Alkenes," *J. Comput. Chem.*, **11**, 848 (1990).

[35] W. Heitler and F. London, "Wechselwirkung Neutraler Atome Und Homöopolare Bindung Nach Der Quantenmechanik," *Z. Phys.*, **44**, 455 (1927).

[36] M. Born and J. E. Mayer, "Zur Gittertheorie Der Ionenkristalle," *Z. Phys.*, **75**, 1 (1932).

[37] R. A. Buckingham and J. E. Lennard-Jones, "The Classical Equation of State of Gaseous Helium, Neon and Argon," *Proc. R. Soc. London, Ser. A*, **168**, 264 (1938).

[38] N. L. Allinger, "Conformational Analysis. 130. MM2. a Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms," *J. Am. Chem. Soc.*, **99**, 8127 (1977).

[39] M. G. Martin and J. I. Siepmann, "Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes," *J. Phys. Chem. B*, **102**, 2569 (1998).

[40] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations," *J. Comput. Chem.*, **4**, 187 (1983).

[41] I. Y. Kanal, J. A. Keith, and G. R. Hutchison, "A Sobering Assessment of Small-Molecule Force Field Methods for Low Energy Conformer Predictions," *Int. J. Quantum Chem.*, **118**, e25512 (2018).

[42] J. Pérez-Pellitero, H. Amrouche, F. Siperstein, G. Pirngruber, C. Nieto-Draghi, G. Chaplais, A. Simon-Masseron, D. Bazer-Bachi, D. Peralta, and N. Bats, "Adsorption of $CO_2$, $CH_4$, and $N_2$ on Zeolitic Imidazolate Frameworks: Experiments and Simulations," *Chem. Eur. J.*, **16**, 1560 (2010).

[43] X. Wu, J. Huang, W. Cai, and M. Jaroniec, "Force Field for ZIF-8 Flexible Frameworks: Atomistic Simulation of Adsorption, Diffusion of Pure Gases as CH4, H2, CO2 and N2," *RSC Adv.*, **4**, 16503 (2014).

[44] E. Boulanger, L. Huang, C. Rupakheti, A. D. MacKerell, and B. Roux, "Optimized Lennard-Jones Parameters for Druglike Small Molecules," *J. Chem. Theory Comput.*, **14**, 3121 (2018).

[45] M. A. Addicoat, N. Vankova, I. F. Akter, and T. Heine, "Extension of the Universal Force Field to Metal-Organic Frameworks," *J. Chem. Theory Comput.*, **10**, 880 (2014).

[46] D. E. Coupry, M. A. Addicoat, and T. Heine, "Extension of the Universal Force Field for Metal-Organic Frameworks," *J. Chem. Theory Comput.*, **12**, 5215 (2016).

[47] P. Brommer and F. Gähler, "Effective Potentials for Quasicrystals from Ab-Initio Data," *Philos. Mag.*, **86**, 753 (2006).

[48] P. Brommer and F. Gähler, "Potfit: Effective Potentials from *ab Initio* Data," *Modell. Simul. Mater. Sci. Eng.*, **15**, 295 (2007).

[49] P. Brommer, A. Kiselev, D. Schopf, P. Beck, J. Roth, and H.-R. Trebin, "Classical Interaction Potentials for Diverse Materials from Ab Initio Data: A Review of Potfit," *Modell. Simul. Mater. Sci. Eng.*, **23**, 074002 (2015).

[50] F. Ercolessi and J. B. Adams, "Interatomic Potentials from First-Principles Calculations: The Force-Matching Method," *Europhys. Lett.*, **26**, 583 (1994).

[51] M. S. Daw, S. M. Foiles, and M. I. Baskes, "The Embedded-Atom Method: A Review of Theory and Applications," *Mater. Sci. Reports*, **9**, 251 (1993).

[52] B. Waldher, J. Kuta, S. Chen, N. Henson, and A. E. Clark, "ForceFit: A Code to Fit Classical Force Fields to Quantum Mechanical Potential Energy Surfaces," *J. Comput. Chem.*, **31**, 2307 (2010).

[53] L.-P. Wang, J. Chen, and T. Van Voorhis, "Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data," *J. Chem. Theory Comput.*, **9**, 452 (2013).

[54] I. Cacelli and G. Prampolini, "Parametrization and Validation of Intramolecular Force Fields Derived from DFT Calculations," *J. Chem. Theory Comput.*, **3**, 1803 (2007).

[55] V. Barone, I. Cacelli, N. De Mitri, D. Licari, S. Monti, and G. Prampolini, "Joyce and Ulysses: Integrated and User-Friendly Tools for the Parameterization of Intramolecular Force Fields from Quantum Mechanical Data," *Phys. Chem. Chem. Phys.*, **15**, 3736 (2013).

[56] I. Cacelli, A. Cimoli, P. R. Livotto, and G. Prampolini, "An Automated Approach for the Parameterization of Accurate Intermolecular Force-Fields: Pyridine as a Case Study," *J. Comput. Chem.*, **33**, 1055 (2012).

[57] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, and V. Van Speybroeck, "QuickFF: A Program for a Quick and Easy Derivation of Force Fields for Metal-Organic Frameworks from Ab Initio Input," *J. Comput. Chem.*, **36**, 1015 (2015).

[58] L. Vanduyfhuys, S. Vandenbrande, J. Wieme, M. Waroquier, T. Verstraelen, and V. Van Speybroeck, "Extension of the QuickFF Force Field Protocol for an Improved Accuracy of Structural, Vibrational, Mechanical and Thermal Properties of Metal-Organic Frameworks," *J. Comput. Chem.*, **39**, 999 (2018).

[59] S. M. J. Rogge, L. Vanduyfhuys, A. Ghysels, M. Waroquier, T. Verstraelen, G. Maurin, and V. Van Speybroeck, "A Comparison of Barostats for the Mechanical Characterization of Metal-Organic Frameworks," *J. Chem. Theory Comput.*, **11**, 5583 (2015).

[60] S. M. J. Rogge, J. Wieme, L. Vanduyfhuys, S. Vandenbrande, G. Maurin, T. Verstraelen, M. Waroquier, and V. Van Speybroeck, "Thermodynamic Insight in the High-Pressure Behavior of UiO-66: Effect of Linker Defects and Linker Expansion," *Chem. Mater.*, **28**, 5721 (2016).

[61] J. Wieme, L. Vanduyfhuys, S. M. J. Rogge, M. Waroquier, and V. Van Speybroeck, "Exploring the Flexibility of MIL-47(V)-Type Materials Using Force Field Molecular Dynamics Simulations," *J. Phys. Chem. C*, **120**, 14934 (2016).

[62] R. Demuynck, S. M. J. Rogge, L. Vanduyfhuys, J. Wieme, M. Waroquier, and V. Van Speybroeck, "Efficient Construction of Free Energy Profiles of Breathing Metal–Organic Frameworks Using Advanced Molecular Dynamics Simulations," *J. Chem. Theory Comput.*, **13**, 5861 (2017).

[63] L. Vanduyfhuys, S. M. J. Rogge, J. Wieme, S. Vandenbrande, G. Maurin, M. Waroquier, and V. Van Speybroeck, "Thermodynamic Insight into Stimuli-Responsive Behaviour of Soft Porous Crystals," *Nat. Commun.*, **9**, 5583 (2018).

[64] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, III, "ReaxFF: A Reactive Force Field for Hydrocarbons," *J. Phys. Chem. A*, **105**, 9396 (2001).

[65] P. Ren and J. W. Ponder, "Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanics Calculations," *J. Comput. Chem.*, **23**, 1497 (2002).

[66] P. Ren and J. W. Ponder, "Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation," *J. Phys. Chem. B*, **107**, 5933 (2003).

[67] A. Grossfield, P. Ren, and J. W. Ponder, "Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field," *J. Am. Chem. Soc.*, **125**, 15671 (2003).

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

[69] L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Y. Ren, J. W. Ponder, and J.-P. Piquemal, "Tinker-HP: A Massively Parallel Molecular Dynamics Package for Multiscale Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields," *Chem. Sci.*, **9**, 956 (2018).

[70] B. T. Thole, "Molecular Polarizabilities Calculated with a Modified Dipole Interaction," *Chem. Phys.*, **59**, 341 (1981).

[71] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier, "ACKS2: Atom-Condensed Kohn-Sham DFT Approximated to Second Order," *J. Chem. Phys.*, **138**, 074108 (2013).

[72] T. Verstraelen, S. Vandenbrande, and P. W. Ayers, "Direct Computation of Parameters for Accurate Polarizable Force Fields," *J. Chem. Phys.*, **141**, 194114 (2014).

[73] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K.-R. Müller, "Machine Learning of Accurate Energy-Conserving Molecular Force Fields," *Sci. Adv.*, **3** (2017).

[74] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, "Towards Exact Molecular Dynamics Simulations with Machine-Learned Force Fields," *Nat. Commun.*, **9** (2018).

[75] S. Vandenbrande, M. Waroquier, V. Van Speybroeck, and T. Verstraelen, "The Monomer Electron Density Force Field (MEDFF): A Physically Inspired Model for Noncovalent Interactions," *J. Chem. Theory Comput.*, **13**, 161 (2017).

[76] P. L. A. Popelier, "Molecular simulation by Knowledgeable Quantum Atoms," *Phys. Scr.*, **91**, 033007 (2016).

[77] T. L. Fletcher and P. L. A. Popelier, "FFLUX: Transferability of Polarizable Machine-Learned Electrostatics in Peptide Chains," *J. Comput. Chem.*, **38**, 1005 (2017).

[78] P. L. A. Popelier, "QCTFF: On the Construction of a Novel Protein Force Field," *Int. J. Quantum Chem.*, **115**, 1005 (2015).

[79] R. F. W. Bader, "A Quantum Theory of Molecular Structure and Its Applications," *Chem. Rev.*, **91**, 893 (1991).

[80] G. A. Cisneros, J.-P. Piquemal, and T. A. Darden, "Intermolecular Electrostatic Energies Using Density Fitting," *J. Chem. Phys.*, **123**, 044109 (2005).

[81] J.-P. Piquemal, G. A. Cisneros, P. Reinhardt, N. Gresh, and T. A. Darden, "Towards a Force Field Based on Density Fitting," *J. Chem. Phys.*, **124**, 104101 (2006).

[82] G. A. Cisneros, J.-P. Piquemal, and T. A. Darden, "Generalization of the Gaussian Electrostatic Model: Extension to Arbitrary Angular Momentum, Distributed Multipoles, and Speedup with Reciprocal Space Methods," *J. Chem. Phys.*, **125**, 184101 (2006).

[83] R. J. Wheatley and S. L. Price, "An Overlap Model for Estimating the Anisotropy of Repulsion," *Mol. Phys.*, **69**, 507 (1990).

[84] P. S. Bagus, K. Hermann, and C. W. Bauschlicher, "A New Analysis of Charge Transfer and Polarization for Ligand–metal Bonding: Model Studies of Al4CO and Al4NH3," *J. Chem. Phys.*, **80**, 4378 (1984).

[85] N. Gresh, G. A. Cisneros, T. A. Darden, and J.-P. Piquemal, "Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. a Bottom-up Strategy," *J. Chem. Theory Comput.*, **3**, 1960 (2007).

[86] B. Jeziorski, R. Moszynski, and K. Szalewicz, "Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of Van Der Waals Complexes," *Chem. Rev.*, **94**, 1887 (1994).

[87] J. G. McDaniel and J. R. Schmidt, "Physically-Motivated Force Fields from Symmetry-Adapted Perturbation Theory," *J. Phys. Chem. A*, **117**, 2053 (2013).

[88] A. Stone and M. Alderton, "Distributed Multipole Analysis," *Mol. Phys.*, **56**, 1047 (1985).

[89] J. G. McDaniel and J. R. Schmidt, "First-Principles Many-Body Force Fields from the Gas Phase to Liquid: A "Universal" Approach," *J. Phys. Chem. B*, **118**, 8042 (2014).

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

[91] T. C. Lillestolen and R. J. Wheatley, "Redefining the Atom: Atomic Charge Densities Produced by an Iterative Stockholder Approach," *Chem. Commun.*, 5909 (2008).

[92] A. J. Misquitta, A. J. Stone, and F. Fazeli, "Distributed Multipoles from a Robust Basis-Space Implementation of the Iterated Stockholder Atoms Procedure," *J. Chem. Theory Comput.*, **10**, 5405 (2014).

[93] M. J. Van Vleet, A. J. Misquitta, A. J. Stone, and J. R. Schmidt, "Beyond Born-Mayer: Improved Models for Short-Range Repulsion in Ab Initio Force Fields," *J. Chem. Theory Comput.*, **12**, 3851 (2016).

[94] M. J. Van Vleet, A. J. Misquitta, and J. R. Schmidt, "New Angles on Standard Force Fields: Toward a General Approach for Treating Atomic-Level Anisotropy," *J. Chem. Theory Comput.*, **14**, 739 (2018).

[95] A. J. Misquitta and A. J. Stone, "ISA-Pol: Distributed Polarizabilities and Dispersion Models from a Basis-Space Implementation of the Iterated Stockholder Atoms Procedure," *Theor. Chem. Acc.*, **137**, 153 (2018).

[96] L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier, and V. Van Speybroeck, "Ab Initio Parametrized Force Field for the Flexible Metal–Organic Framework MIL-53(Al)," *J. Chem. Theory Comput.*, **8**, 3217 (2012).

[97] L. Chen, C. A. Morrison, and T. Düren, "Improving Predictions of Gas Adsorption in Metal-Organic Frameworks with Coordinatively Unsaturated Metal Sites: Model Potentials, Ab Initio Parameterization, and GCMC Simulations," *J. Phys. Chem. C*, **116**, 18899 (2012).

[98] R. Mercado, B. Vlaisavljevich, L.-C. Lin, K. Lee, Y. Lee, J. A. Mason, D. J. Xiao, M. I. Gonzalez, M. T. Kapelewski, J. B. Neaton, and B. Smit, "Force Field Development from Periodic Density Functional Theory Calculations for Gas Separation Applications Using Metal-Organic Frameworks," *J. Phys. Chem. C*, **120**, 12590 (2016).

[99] T. M. Becker, J. Heinen, D. Dubbeldam, L.-C. Lin, and T. J. H. Vlugt, "Polarizable Force Fields for $CO_2$ and $CH_4$ Adsorption in M-MOF-74," *J. Phys. Chem. C*, **121**, 4659 (2017).

[100] T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck, M. Waroquier, and P. W. Ayers, "Minimal Basis Iterative Stockholder: Atoms in Molecules for Force-Field Development," *J. Chem. Theory Comput.*, **12**, 3894 (2016).

[101] U. C. Singh and P. A. Kollman, "An Approach to Computing Electrostatic Charges for Molecules," *J. Comput. Chem.*, **5**, 129 (1984).

[102] B. H. Besler, K. M. Merz, and P. A. Kollman, "Atomic Charges Derived from Semiempirical Methods," *J. Comput. Chem.*, **11**, 431 (1990).

[103] C. M. Breneman and K. B. Wiberg, "Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis," *J. Comput. Chem.*, **11**, 361 (1990).

[104] H. Hu, Z. Lu, and W. Yang, "Fitting Molecular Electrostatic Potentials from Quantum Mechanical Calculations," *J. Chem. Theory Comput.*, **3**, 1004 (2007).

[105] C. Campañá, B. Mussard, and T. K. Woo, "Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional," *J. Chem. Theory Comput.*, **5**, 2866 (2009).

[106] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model," *J. Phys. Chem.*, **97**, 10269 (1993).

[107] R. S. Mulliken, "Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I," *J. Chem. Phys.*, **23**, 1833 (1955).

[108] P.-O. Löwdin, "On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals," *J. Chem. Phys.*, **18**, 365 (1950).

[109] A. E. Reed, R. B. Weinstock, and F. Weinhold, "Natural Population Analysis," *J. Chem. Phys.*, **83**, 735 (1985).

[110] X. Qian, J. Li, L. Qi, C.-Z. Wang, T.-L. Chan, Y.-X. Yao, K.-M. Ho, and S. Yip, "Quasiatomic Orbitals for Ab Initio Tight-Binding Analysis," *Phys. Rev. B*, **78**, 245112 (2008).

[111] R. E. Duke, O. N. Starovoytov, J.-P. Piquemal, and G. A. Cisneros, "GEM*: A Molecular Electronic Density-Based Force Field for Molecular Dynamics Simulations," *J. Chem. Theory Comput.*, **10**, 1361 (2014).

[112] F. Hirshfeld, "Bonded-Atom Fragments for Describing Molecular Charge Densities," *Theor. Chim. Acta*, **44**, 129 (1977).

[113] P. Bultinck, C. Van Alsenoy, P. W. Ayers, and R. Carbó-Dorca, "Critical Analysis and Extension of the Hirshfeld Atoms in Molecules," *J. Chem. Phys.*, **126**, 144111 (2007).

[114] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, and M. Waroquier, "Hirshfeld-E Partitioning: AIM Charges with an Improved Trade-off Between Robustness and Accurate Electrostatics," *J. Chem. Theory Comput.*, **9**, 2221 (2013).

[115] T. A. Manz and D. S. Sholl, "Chemically Meaningful Atomic Charges That Reproduce the Electrostatic Potential in Periodic and Nonperiodic Materials," *J. Chem. Theory Comput.*, **6**, 2455 (2010).

[116] T. A. Manz and D. S. Sholl, "Improved Atoms-In-Molecule Charge Partitioning Functional for Simultaneously Reproducing the Electrostatic Potential and Chemical States in Periodic and Nonperiodic Materials," *J. Chem. Theory Comput.*, **8**, 2844 (2012).

[117] T. A. Manz and N. G. Limas, "Introducing DDEC6 Atomic Population Analysis: Part 1. Charge Partitioning Theory and Methodology," *RSC Adv.*, **6**, 47771 (2016).

[118] R. G. Parr, P. W. Ayers, and R. F. Nalewajski, "What Is an Atom in a Molecule?" *J. Phys. Chem. A*, **109**, 3957 (2005).

[119] A. D. Becke, "Density-Functional Thermochemistry. III. the Role of Exact Exchange," *J. Chem. Phys.*, **98**, 5648 (1993).

[120] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density," *Phys. Rev. B*, **37**, 785 (1988).

[121] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields," *J. Phys. Chem.*, **98**, 11623 (1994).

[122] R. Ditchfield, W. J. Hehre, and J. A. Pople, "Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules," *J. Chem. Phys.*, **54**, 724 (1971).

[123] J. Řezáč, K. E. Riley, and P. Hobza, "S66: A Well-Balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures," *J. Chem. Theory Comput.*, **7**, 2427 (2011).

[124] J. Řezáč and P. Hobza, "Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods," *J. Chem. Theory Comput.*, **8**, 141 (2012).

[125] J. Řezáč, K. E. Riley, and P. Hobza, "Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules," *J. Chem. Theory Comput.*, **8**, 4285 (2012).

[126] T. Verstraelen, E. Pauwels, F. De Proft, V. Van Speybroeck, P. Geerlings, and M. Waroquier, "Assessment of Atomic Charge Models for Gas-Phase Computations on Polypeptides," *J. Chem. Theory Comput.*, **8**, 661 (2012).

[127] M. Zgarbová, M. Otyepka, J. Sponer, P. Hobza, and P. Jurecka, "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.*, **12**, 10476 (2010).

[128] K. Kitaura and K. Morokuma, "A New Energy Decomposition Scheme for Molecular Interactions Within the Hartree-Fock Approximation," *Int. J. Quantum Chem.*, **10**, 325 (1976).

[129] K. Morokuma and K. Kitaura, in *Chemical Applications of Atomic and Molecular Electrostatic Potentials* (Springer US, 1981) pp. 215–242.

[130] J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, and T. D. Crawford, "Psi4: An Open-Source Ab Initio Electronic Structure Program," *WIREs Comput. Mol. Sci.*, **2**, 556 (2012).

[131] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, "Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability," *J. Chem. Theory Comput.*, **13**, 3185 (2017).

[132] Y. S. Kim, S. K. Kim, and W. D. Lee, "Dependence of the Closed-Shell Repulsive Interaction on the Overlap of the Electron Densities," *Chem. Phys. Lett.*, **80**, 574 (1981).

[133] S. Kita, K. Noda, and H. Inouye, "Repulsive Potentials for Cl–R and Br–R (R=He, Ne, and Ar) Derived from Beam Experiments," *J. Chem. Phys.*, **64**, 3446 (1976).

[134] K. T. Tang and J. P. Toennies, "An Improved Simple Model for the Van Der Waals Potential Based on Universal Damping Functions for the Dispersion Coefficients," *J. Chem. Phys.*, **80**, 3726 (1984).

[135] G. Starkschall and R. G. Gordon, "Calculation of Coefficients in the Power Series Expansion of the Long-Range Dispersion Force Between Atoms," *J. Chem. Phys.*, **56**, 2801 (1972).

[136] A. J. Misquitta, "Charge Transfer from Regularized Symmetry-Adapted Perturbation Theory," *J. Chem. Theory Comput.*, **9**, 5313 (2013).

[137] S. Deng, Q. Wang, and P. Ren, "Estimating and Modeling Charge Transfer from the SAPT Induction Energy," *J. Comput. Chem.*, **38**, 2222 (2017).

[138] B. G. Dick and A. W. Overhauser, "Theory of the Dielectric Constants of Alkali Halide Crystals," *Phys. Rev.*, **112**, 90 (1958).

[139] A. G. Donchev, V. D. Ozrin, M. V. Subbotin, O. V. Tarasov, and V. I. Tarasov, "A Quantum Mechanical Polarizable Force Field for Biomolecular Interactions," *PNAS of the United States of America*, **102**, 7829 (2005).

[140] K.-P. Schröder and J. Sauer, "Potential Functions for Silica and Zeolite Catalysts Based on Ab Initio Calculations. 3. a Shell Model Ion Pair Potential for Silica and Aluminosilicates," *J. Phys. Chem.*, **100**, 11043 (1996).

[141] G. V. Lewis and C. R. A. Catlow, "Potential Models for Ionic Oxides," *J. Phys. C: Solid State Phys.*, **18**, 1149 (1985).

[142] M. Schmollngruber, V. Lesch, C. Schroder, A. Heuer, and O. Steinhauser, "Comparing Induced Point-Dipoles and Drude Oscillators," *Phys. Chem. Chem. Phys.*, **17**, 14297 (2015).

[143] L. Huang and B. Roux, "Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data," *J. Chem. Theory Comput.*, **9**, 3543 (2013).

[144] A. Savelyev and A. D. MacKerell, "All-Atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model," *J. Comput. Chem.*, **35**, 1219 (2014).

[145] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, "Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins," *J. Chem. Theory Comput.*, **9**, 4046 (2013).

[146] D. Elking, T. Darden, and R. J. Woods, "Gaussian Induced Dipole Polarization Model," *J. Comput. Chem.*, **28**, 1261 (2007).

[147] P. Ren, C. Wu, and J. W. Ponder, "Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules," *J. Chem. Theory Comput.*, **7**, 3143 (2011).

[148] J. N. Murrell, M. Randic, and D. R. Williams, "The Theory of Intermolecular Forces in the Region of Small Orbital Overlap," *Proc. Roy. Soc. London, Series A*, **284**, 566 (1965).

[149] N. Gresh, P. Claverie, and A. Pullman, "Computations of Intermolecular Interactions: Expansion of a Charge-Transfer Energy Contribution in the Framework of an Additive Procedure. Applications to Hydrogen-Bonded Systems," *Int. J. Quantum Chem.*, **22**, 199 (1982).

[150] N. Gresh, P. Claverie, and A. Pullman, "Intermolecular Interactions: Elaboration on an Additive Procedure Including an Explicit Charge-Transfer Contribution," *Int. J. Quantum Chem.*, **29**, 101 (1986).

[151] D. Hagberg, G. Karlström, B. O. Roos, and L. Gagliardi, "The Coordination of Uranyl in Water: A Combined Quantum Chemical and Molecular Simulation Study," *J. Am. Chem. Soc.*, **127**, 14250 (2005).

[152] R. Kumar, F.-F. Wang, G. R. Jenness, and K. D. Jordan, "A Second Generation Distributed Point Polarizable Water Model," *J. Chem. Phys.*, **132**, 014309 (2010).

[153] A. J. Stone and A. J. Misquitta, "Charge-Transfer in Symmetry-Adapted Perturbation Theory," *Chem. Phys. Lett.*, **473**, 201 (2009).

[154] P. Söderhjelm, G. Karlström, and U. Ryde, "Comparison of Overlap-Based Models for Approximating the Exchange-Repulsion Energy," *J. Chem. Phys.*, **124**, 244101 (2006).

[155] S. Grimme, "A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations," *J. Chem. Theory Comput.*, **10**, 4497 (2014).

[156] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006).

[157] V. Petzold, T. Bligaard, and K. Jacobsen, "Construction of New Electronic Density Functionals with Error Estimation Through Fitting," *Top. Catal.*, **55**, 402 (2012).

[158] F. Cailliez and P. Pernot, "Statistical Approaches to Forcefield Calibration and Prediction Uncertainty in molecular simulation," *J. Chem. Phys.*, **134**, 054124 (2011).

[159] J. C. Faver, M. L. Benson, X. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill, and K. M. Merz, "Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein-Ligand Complexes," *J. Chem. Theory Comput.*, **7**, 790 (2011).

[160] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and Testing of a General Amber Force Field," *J. Comput. Chem.*, **25**, 1157 (2004).

[161] T. A. Halgren, "MMFF VI. MMFF94s Option for Energy Minimization Studies," *J. Comput. Chem.*, **20**, 720 (1999).

[162] W. L. Jorgensen, J. D. Madura, and C. J. Swenson, "Optimized Intermolecular Potential Functions for Liquid Hydrocarbons," *J. Am. Chem. Soc.*, **106**, 6638 (1984).

[163] D. McQuarrie, *Statistical Mechanics* (University Science Books, 2000).

[164] J. Dymond, J. Cholinski, A. Sząfranski, and D. Wyrzykowska-Stankiewicz, "Second Virial Coefficients for N-Alkanes; Recommendations and Predictions," *Fluid Phase Equilib.*, **27**, 1 (1986).

[165] S. Glos, R. Kleinrahm, and W. Wagner, "Measurement of the ($p$, $\rho$, $T$) Relation of Propane, Propylene, N-Butane, and Isobutane in the Temperature Range from ($95$ to $340$) K at Pressures up to $12$MPa Using an Accurate Two-Sinker Densimeter," *J. Chem. Thermodyn.*, **36**, 1037 (2004).

[166] W. Warowny, P. Wielopolski, and J. Stecki, "Compressibility Factors and Virial Coefficients for Propane, Propene and Their Mixtures by the Burnett Method," *Physica A*, **91**, 73 (1978).

[167] P. Nowak, R. Kleinrahm, and W. Wagner, "Measurement and Correlation of the ($p$, $\rho$, $T$) Relation of Ethylene I. the Homogeneous Gaseous and Liquid Regions in the Temperature Range from $105$K to $340$K at Pressures up to 12 MPa," *J. Chem. Thermodyn.*, **28**, 1423 (1996).

[168] B. A. Younglove and J. F. Ely, "Thermophysical Properties of Fluids. II. Methane, Ethane, Propane, Isobutane, and Normal Butane," *J. Phys. Chem. Ref. Data*, **16**, 577 (1987).

[169] J. Smukala, R. Span, and W. Wagner, "New Equation of State for Ethylene Covering the Fluid Region for Temperatures from the Melting Line to $450$K at Pressures up to $300$MPa," *J. Phys. Chem. Ref. Data*, **29**, 1053 (2000).

[170] E. Lemmon, M. McLinden, and D. Friend, *"Thermophysical Properties of Fluid Systems" in NIST Chemistry WebBook*, 1st ed. (National Institute of Standards and Technology, Gaithersburg MD, 20899, 1996).

[171] G. Kaminski, E. M. Duffy, T. Matsui, and W. L. Jorgensen, "Free Energies of Hydration and Pure Liquid Properties of Hydrocarbons from the OPLS All-Atom Model," *J. Phys. Chem.*, **98**, 13077 (1994).

[172] J. Nagy, D. F. Weaver, and V. H. Smith Jr., "A Comprehensive Study of Alkane Nonbonded Empirical Force Fields. Suggestions for Improved Parameter Sets," *J. Phys. Chem.*, **99**, 8058 (1995).

[173] F. Paesani, "Getting the Right Answers for the Right Reasons: Toward Predictive Molecular Simulations of Water with Many-Body Potential Energy Functions," *Acc. Chem. Res.*, **49**, 1844 (2016).

[174] O. Anatole von Lilienfeld and A. Tkatchenko, "Two- and Three-Body Interatomic Dispersion Energy Contributions to Binding in Molecules and Solids," *J. Chem. Phys.*, **132**, 234109 (2010).

[175] A. Ambrosetti, A. M. Reilly, R. A. DiStasio Jr., and A. Tkatchenko, "Long-Range Correlation Energy Calculated from Coupled Atomic Response Functions," *J. Chem. Phys.*, **140**, 18A508 (2014).

[176] M. Ceriotti, J. More, and D. E. Manolopoulos, "i-PI: A Python Interface for Ab Initio Path Integral Molecular Dynamics Simulations," *Comput. Phys. Commun.*, **185**, 1019 (2014).

[177] V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, B. A. Helfrecht, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. V. Speybroeck, and M. Ceriotti, "i-PI 2.0: A Universal Force Engine for advanced Molecular simulations," *Comput. Phys. Commun.* (2018).

[178] A. Lamaire, J. Wieme, S. M. J. Rogge, M. Waroquier, and V. Van Speybroeck, "On the Importance of Anharmonicities and Nuclear Quantum Effects in Modelling the Structural Properties and Thermal Expansion of MOF-5," *J. Chem. Phys.*, **150**, 094503 (2019).

[179] L. Pereyaslavets, I. Kurnikov, G. Kamath, O. Butin, A. Illarionov, I. Leontyev, M. Olevanov, M. Levitt, R. D. Kornberg, and B. Fain, "On the Importance of Accounting for Nuclear Quantum Effects in Ab Initio Calibrated Force Fields in Biological Simulations," *PNAS*, **115**, 8878 (2018).

[180] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics," *J. Comput. Phys.*, **117**, 1 (1995).

[181] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids," *J. Am. Chem. Soc.*, **118**, 11225 (1996).

[182] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe, and O. M. Yaghi, "Systematic Design of Pore Size and Functionality in Isoreticular MOFs and Their Application in Methane Storage," *Science*, **295**, 469 (2002).

[183] O. M. Yaghi, M. O'Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi, and J. Kim, "Reticular Synthesis and the Design of New Materials," *Nature*, **423**, 705 (2003).

[184] S. Horike, S. Shimomura, and S. Kitagawa, "Soft Porous Crystals," *Nat. Chem.*, **1**, 695 (2009).

[185] J.-R. Li, R. J. Kuppler, and H.-C. Zhou, "Selective Gas Adsorption and Separation in Metal-Organic Frameworks," *Chem. Soc. Rev.*, **38**, 1477 (2009).

[186] B. Li, H.-M. Wen, W. Zhou, and B. Chen, "Porous Metal-Organic Frameworks for Gas Storage and Separation: What, How, and Why?" *J. Phys. Chem. Lett.*, **5**, 3468 (2014).

[187] H. Li, K. Wang, Y. Sun, C. T. Lollar, J. Li, and H.-C. Zhou, "Recent Advances in Gas Storage and Separation Using Metal-Organic Frameworks," *Mater. Today*, **21**, 108 (2018).

[188] S. M. J. Rogge, M. Waroquier, and V. Van Speybroeck, "Reliably Modeling the Mechanical Stability of Rigid and Flexible Metal-Organic Frameworks," *Acc. Chem. Res.*, **51**, 138 (2018).

[189] S. Vandenbrande, T. Verstraelen, J. J. Gutiérrez-Sevillano, M. Waroquier, and V. Van Speybroeck, "Methane Adsorption in Zr-Based MOFs: Comparison and Critical Evaluation of Force Fields," *J. Phys. Chem. C*, **121**, 25309 (2017).

[190] J. H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga, and K. P. Lillerud, "A New Zirconium Inorganic Building Brick Forming Metal Organic Frameworks with Exceptional Stability," *J. Am. Chem. Soc.*, **130**, 13850 (2008).

[191] V. Bon, I. Senkovska, M. S. Weiss, and S. Kaskel, "Tailoring of Network Dimensionality and Porosity Adjustment in Zr- and Hf-Based MOFs," *CrystEngComm*, **15**, 9572 (2013).

[192] J. E. Mondloch, W. Bury, D. Fairen-Jimenez, S. Kwon, E. J. DeMarco, M. H. Weston, A. A. Sarjeant, S. T. Nguyen, P. C. Stair, R. Q. Snurr, O. K. Farha, and J. T. Hupp, "Vapor-Phase Metalation by Atomic Layer Deposition in a Metal-Organic Framework," *J. Am. Chem. Soc.*, **135**, 10294 (2013).

[193] H. Furukawa, F. Gándara, Y.-B. Zhang, J. Jiang, W. L. Queen, M. R. Hudson, and O. M. Yaghi, "Water Adsorption in Porous Metal–Organic Frameworks and Related Materials," *J. Am. Chem. Soc.*, **136**, 4369 (2014).

[194] N. S. Bobbitt, M. L. Mendonca, A. J. Howarth, T. Islamoglu, J. T. Hupp, O. K. Farha, and R. Q. Snurr, "Metal-Organic Frameworks for the Removal of Toxic Industrial Chemicals and Chemical Warfare Agents," *Chem. Soc. Rev.*, **46**, 3357 (2017).

[195] N. L. Allinger, X. Zhou, and J. Bergsma, "Molecular Mechanics Parameters," *J. Mol. Struct. THEOCHEM*, **312**, 69 (1994).

[196] J. G. McDaniel and J. R. Schmidt, "Robust, Transferable, and Physically Motivated Force Fields for Gas Adsorption in Functionalized Zeolitic Imidazolate Frameworks," *J. Phys. Chem. C*, **116**, 14031 (2012).

[197] D. Dubbeldam, S. Calero, D. E. Ellis, and R. Q. Snurr, "RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials," *Mol. Simul.*, **42**, 81 (2016).

[198] Q. Yang, A. D. Wiersum, H. Jobic, V. Guillerm, C. Serre, P. L. Llewellyn, and G. Maurin, "Understanding the Thermodynamic and Kinetic Behavior of the CO2/CH4 Gas Mixture Within the Porous Zirconium Terephthalate UiO-66(Zr): A Joint Experimental and Modeling Approach," *J. Phys. Chem. C*, **115**, 13768 (2011).

[199] J. H. Cavka, C. A. Grande, G. Mondino, and R. Blom, "High Pressure Adsorption of CO2 and CH4 on Zr-MOFs," *Ind. Eng. Chem. Res.*, **53**, 15500 (2014).

[200] G. E. Cmarik, M. Kim, S. M. Cohen, and K. S. Walton, "Tuning the Adsorption Properties of UiO-66 Via Ligand Functionalization," *Langmuir*, **28**, 15606 (2012).

[201] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, and R. Q. Snurr, "Computation-Ready, Experimental Metal-Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals," *Chem. Mater.*, **26**, 6185 (2014).

[202] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, **77**, 3865 (1996).

[203] D. G. A. Smith, L. A. Burns, K. Patkowski, and C. D. Sherrill, "Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory," *J. Phys. Chem. Lett.*, **7**, 2197 (2016).

[204] A. R. Kulkarni and D. S. Sholl, "DFT-Derived Force Fields for Modeling Hydrocarbon Adsorption in MIL-47(V)," *Langmuir*, **31**, 8453 (2015).

[205] M. J. Lennox, M. Bound, A. Henley, and E. Besley, "The Right Isotherms for the Right Reasons? Validation of Generic Force Fields for Prediction of Methane Adsorption in Metal-Organic Frameworks," *Mol. Simul.*, **43**, 828 (2017).

[206] S. Vandenbrande, M. Waroquier, V. Van Speybroeck, and T. Verstraelen, "Ab Initio Evaluation of Henry Coefficients Using Importance Sampling," *J. Chem. Theory Comput.*, **14**, 6359 (2018).

[207] B. Widom, "Some Topics in the Theory of Fluids," *J. Chem. Phys.*, **39**, 2808 (1963).

[208] S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the Damping Function in Dispersion Corrected Density Functional Theory," *J. Comput. Chem.*, **32**, 1456 (2011).

[209] A. Kundu, G. Piccini, K. Sillar, and J. Sauer, "Ab Initio Prediction of Adsorption Isotherms for Small Molecules in Metal-Organic Frameworks," *J. Am. Chem. Soc.*, **138**, 14047 (2016).

[210] D. Britt, D. Tranchemontagne, and O. M. Yaghi, "Metal-Organic Frameworks with High Capacity and Selectivity for Harmful Gases," *PNAS*, **105**, 11623 (2008).

[211] J. A. Mason, K. Sumida, Z. R. Herm, R. Krishna, and J. R. Long, "Evaluating Metal-Organic Frameworks for Post-Combustion Carbon Dioxide Capture Via Temperature Swing Adsorption," *Energy Environ. Sci.*, **4**, 3030 (2011).

[212] D. Yu, A. O. Yazaydin, J. R. Lane, P. D. C. Dietzel, and R. Q. Snurr, "a Combined Experimental and Quantum Chemical Study of $CO_2$ Adsorption in the Metal-Organic Framework CPO-27 with Different Metals," *Chem. Sci.*, **4**, 3544 (2013).

[213] W. L. Queen, M. R. Hudson, E. D. Bloch, J. A. Mason, M. I. Gonzalez, J. S. Lee, D. Gygi, J. D. Howe, K. Lee, T. A. Darwish, M. James, V. K. Peterson, S. J. Teat, B. Smit, J. B. Neaton, J. R. Long, and C. M. Brown, "Comprehensive Study of Carbon Dioxide Adsorption in the Metal-Organic Frameworks $M_2$(dobdc) (M = Mg, Mn, Fe, Co, Ni, Cu, Zn)," *Chem. Sci.*, **5**, 4569 (2014).

[214] G. Román-Pérez and J. M. Soler, "Efficient Implementation of a Van Der Waals Density Functional: Application to Double-Wall Carbon Nanotubes," *Phys. Rev. Lett.*, **103**, 096102 (2009).

[215] K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, "Higher-Accuracy Van Der Waals Density Functional," *Phys. Rev. B*, **82**, 081101 (2010).

[216] J. Klimeš, D. R. Bowler, and A. Michaelides, "Chemical Accuracy for the Van Der Waals Density Functional," *J. Phys.: Condens. Matter*, **22**, 022201 (2010).

[217] B. Vlaisavljevich, J. Huck, Z. Hulvey, K. Lee, J. A. Mason, J. B. Neaton, J. R. Long, C. M. Brown, D. Alfè, A. Michaelides, and B. Smit, "Performance of Van Der Waals Corrected Functionals for Guest Adsorption in the M2(dobdc) Metal-Organic Frameworks," *J. Phys. Chem. A*, **121**, 4139 (2017).

[218] I. T. Todorov, W. Smith, K. Trachenko, and M. T. Dove, "DL_POLY_3: New Dimensions in Molecular Dynamics Simulations Via Massive Parallelism," *J. Mater. Chem.*, **16**, 1911 (2006).

[219] T. Verstraelen, P. Tecmer, F. Heidar-Zadeh, C. E. González-Espinoza, M. Chan, T. D. Kim, K. Boguslawski, S. Fias, S. Vandenbrande, and P. W. Ayers, *HORTON 2.1.1, Http://theochem.github.com/horton/* (2017).

[220] T. Bereau, R. A. DiStasio, A. Tkatchenko, and O. A. von Lilienfeld, "Non-Covalent Interactions Across Organic and Biological Subsets of Chemical Space: Physics-Based Potentials Parametrized from Machine Learning," *J. Chem. Phys.*, **148**, 241706 (2018).

[221] J. Hermann, R. A. DiStasio, and A. Tkatchenko, "First-Principles Models for Van Der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications," *Chem. Rev.*, **117**, 4714 (2017).

[222] S. M. J. Rogge, R. Goeminne, R. Demuynck, J. J. Gutiérrez-Sevillano, S. Vandenbrande, L. Vanduyfhuys, M. Waroquier, T. Verstraelen, and V. Van Speybroeck, "Modeling Gas Adsorption in Flexible Metal–Organic Frameworks Via Hybrid Monte Carlo/Molecular Dynamics Schemes," *Advanced Theory and Simulations*, **2**, 1800177 (2019).

[223] P.-O. Löwdin, "Some Comments on the Present Situation of Quantum Chemistry in View of the Discussions a the Dubrovnik Workshop on the Electronic Structure of Molecules," *Pure Applied Chemistry*, **61**, 2185 (1989).

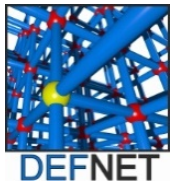
[224] J. Řezáč, L. Šimová, and P. Hobza, "CCSD[T] Describes Noncovalent Interactions Better Than the CCSD(T), CCSD(TQ), and CCSDT Methods," *J. Chem. Theory Comput.*, **9**, 364 (2013).

[225] P.-O. Löwdin, "Twenty-Five Years of Sanibel Symposia: A Brief Historic and Scientific Survey," *Int. J. Quantum Chem.*, **28**, 19 (1985).

[226] S. M. J. Rogge, S. Caroes, R. Demuynck, M. Waroquier, V. Van Speybroeck, and A. Ghysels, "The Importance of Cell Shape Sampling to Accurately Predict Flexibility in Metal–Organic Frameworks," *J. Chem. Theory Comput.*, **14**, 1186 (2018).

[227] V. Kapil, J. Wieme, S. Vandenbrande, A. Lamaire, V. Van Speybroeck, and M. Ceriotti, "Modeling the Structural and Thermal Properties of Loaded Metal-Organic Frameworks. an Interplay of Quantum and Anharmonic Fluctuations," *J. Chem. Theory Comput.* (2019).


This research was supported by the Research Foundation Flanders (FWO) through a personal mandate.



Funding was also received from the European Union's Horizon 2020 research and innovation programme under the European Research Council consolidator grant agreement no. 647755 (project acronym: DYNPOR), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska–Curie grant agreement no. 641887 (project acronym: DEFNET), and BELSPO in the frame of IAP/7/05.





The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO, and the Flemish Government – department EWI.