



UvA-DARE (Digital Academic Repository)

Design, Parameterization, and Implementation of Atomic Force Fields for Adsorption in Nanoporous Materials

Dubbeldam, D.; Walton, K.S.; Vlugt, T.J.H.; Calero, S.

DOI

[10.1002/adts.201900135](https://doi.org/10.1002/adts.201900135)

Publication date

2019

Document Version

Final published version

Published in

Advanced Theory and Simulations

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Dubbeldam, D., Walton, K. S., Vlugt, T. J. H., & Calero, S. (2019). Design, Parameterization, and Implementation of Atomic Force Fields for Adsorption in Nanoporous Materials. *Advanced Theory and Simulations*, 2(11), [1900135]. <https://doi.org/10.1002/adts.201900135>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Design, Parameterization, and Implementation of Atomic Force Fields for Adsorption in Nanoporous Materials

David Dubbeldam,* Krista S. Walton, Thijs J. H. Vlugt, and Sofia Calero

Molecular simulations are an excellent tool to study adsorption and diffusion in nanoporous materials. Examples of nanoporous materials are zeolites, carbon nanotubes, clays, metal-organic frameworks (MOFs), covalent organic frameworks (COFs) and zeolitic imidazolate frameworks (ZIFs). The molecular confinement these materials offer has been exploited in adsorption and catalysis for almost 50 years. Molecular simulations have provided understanding of the underlying shape selectivity, and adsorption and diffusion effects. Much of the reliability of the modeling predictions depends on the accuracy and transferability of the force field. However, flexibility and the chemical and structural diversity of MOFs add significant challenges for engineering force fields that are able to reproduce experimentally observed structural and dynamic properties. Recent developments in design, parameterization, and implementation of force fields for MOFs and zeolites are reviewed.

1. Introduction

Nanoporous materials are materials with pore shapes and sizes comparable to the typical size of small molecules. Selective adsorption is the process of exploiting this confinement using differences in molecular configurations. Zeolites are 3D, microporous, crystalline solids with well-defined structures that mainly

Dr. D. Dubbeldam
Van 't Hoff Institute for Molecular Sciences
University of Amsterdam
Science Park 904, 1098XH, Amsterdam, The Netherlands
E-mail: D.Dubbeldam@uva.nl

Prof. K. S. Walton
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology
311 Ferst Dr. NW, Atlanta, GA 30332-0100, USA

Prof. T. J. H. Vlugt
Delft University of Technology
Process & Energy Department
Leeghwaterstraat 39, 2628CB Delft, The Netherlands

Prof. S. Calero
Department of Physical
Chemical and Natural Systems
University Pablo de Olavide
Sevilla 41013, Spain

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adts.201900135>

© 2019 The Authors. Published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/adts.201900135

contain silicon and oxygen. They are readily available, very stable, and relatively cheap. The material should have the right combination of high adsorption selectivity, combined with adequate capacity for use in fixed-bed devices. Recently, new classes of nanoporous materials have been designed that have stability, high void volumes, and well defined tailorable cavities of uniform size. Example are metal-organic frameworks (MOFs),^[1–7] covalent organic frameworks (COFs),^[8,9] and zeolitic imidazolate frameworks (ZIFs).^[10] These novel materials possess almost unlimited structural variety because of the many combinations of building blocks that can be imagined. The building blocks self-assemble during synthesis into crystalline materials that, after evacuation of the structure,

can find applications in adsorption separations, air purification, gas storage, chemical sensing, and catalysis.^[4,11] The judicious selection of building blocks allows the pore volume and functionality to be tailored in a rational manner. Because of the design-principle underlying the synthesis, it is important to understand structure–properties relations, such as the mechanisms of gas separation as a function of shape and size of the pore system, in order to create “blueprints to MOF-design.” Computer simulation is cost-effective, fast, and allows for rapid identification of important structural parameters affecting adsorption.

Most of the published works on molecular simulation studying zeolites have used the Kiselev model.^[12] In this model, zeolite atoms are fixed and interactions of guest atoms with silicon atoms is accounted for by an effective interaction only with the surrounding oxygen atoms. Interactions between sorbate molecules and the host are modeled by placing Lennard–Jones sites and partial charges on all framework atoms and sorbate molecules. The Kiselev model is attractive because of its simplicity and computational efficiency. This model has shown significant success, both in using the molecular dynamics method to compute, for example, the diffusivities via the Einstein relation, and using the Monte Carlo method to, for example, calculate the sorption of hydrocarbons. A step forward for zeolite modeling was the TraPPE-zeo model.^[13] In this model, the Lennard–Jones interaction sites and partial charges are placed at both the oxygen and the silicon atoms of the zeolite lattice. This allows for a better balance of dispersive and first-order electrostatic interactions than is achievable with the Lennard–Jones potential used only for the oxygen atoms.

Early MOF work initially also adapted the Kiselev approach where the framework has been kept rigid. However, the

force field was replaced by a more generic solution such as DREIDING,^[14] or UFF,^[15] for example, to tackle the larger chemical diversity of MOFs. These approaches were very successful without the need for re-parameterization.^[16–19] Over the years several challenges were found. The first major force field challenge is to take flexibility of the framework into account.^[20] All MOFs exhibit some form of flexibility^[21–23] ranging from lattice vibrations around equilibrium positions to large-scale structural transformations upon external stimuli. An anomalous structural transformation due to temperature is negative thermal expansion.^[24,25] Guest adsorption can induce phase transitions such as breathing.^[26,27] Another newly discovered stimuli for breathing is the electric field.^[28] A related interesting phenomenon is the gate-opening behavior observed in ZIF-8. These flexible behaviors can have important implications for MOF application prospects^[29,30] by enabling, for instance, exceptional gas storage or separation performances that is not possible in traditional rigid materials. Soft vibrational modes (terahertz regime) determine large scale framework flexibility.^[31,32] The most difficult and important part of the force field to get right is the lower frequency modes that are responsible for large-scale flexibility and structural rearrangement as a function of external stimuli.^[20] The second major force field challenge are open metal-sites as found for example in Cu-BTC and MOF-74. Sorbent materials containing open-metal sites are promising materials for the separation olefin-paraffin mixtures.^[33] Unfortunately, conventional force fields do not capture the olefin-metal interactions. Ideal adsorbed solution theory (IAST)^[34] breaks down for olefin adsorption in open-metal site materials due to non-ideal donor–acceptor interactions.^[35]

Here, we review design, parameterization, and implementation of force fields that lie at the foundation of computing adsorption and diffusion in nanoporous materials. First, we briefly describe *ab initio* and classical simulations approaches. At the heart of classical simulation methodology lies the “force field,” an equation describing the dependence of the energy of a system on the coordinates of its particles using chemical concepts like bonds, bends, torsions, and non-bonded terms.^[36] The various intra-molecular and inter-molecular potentials that are used in the field of adsorption/diffusion in nanoporous materials are discussed and the implementation of the energy, gradients, and second-derivatives with respect to position and strain are detailed. Then we focus on the details of the long-range interactions, in particular charge–charge interaction using the Ewald summation and the inclusion of bond-dipoles, polarization using induced dipoles, MD/MC-implementation details, and how to deal with a net-charge in the system. The latter is also needed when decomposing the energy into framework, cations, and adsorbate contribution when the framework itself is net-charged but compensated by cations. Next, the theory of molecular mechanics is reviewed which takes the implemented force fields to accurately compute zero Kelvin properties like unit cell size and shape, crystal structure, elastic constants, and vibrational spectra. To be able to compute these properties fast and accurately, allows for a common force field parameterization workflow to optimize the force field parameters to be as close to *ab initio* results as possible. We also discuss other parameterization approaches like constructing empirical force fields based on reproducing experimental data, for example, fitting on adsorption isotherms and



David Dubbeldam received his B.Sc. and Ph.D. degree (with honors) from the University of Amsterdam, in Computer Science and Computational Chemistry, respectively. From 2006 until 2009, he carried out a post-doctoral stay at Northwestern University in the group of Professor Randall Q. Snurr, working on modeling of adsorption and diffusion in flexible metal-organic frame-

works. In 2010 he joined the Computational Chemistry Group at the University of Amsterdam. He is the main author of the Monte Carlo and Molecular Dynamics code “RASPA” aimed at simulations and visualization of adsorption, diffusion, and reactions of adsorbates in nanoporous materials like metal-organic frameworks, zeolites, and clays.



Prof. Krista S. Walton received a B.S.E. degree from the University of Alabama–Huntsville and Ph.D. from Vanderbilt University, both in chemical engineering. She joined the faculty of the Georgia Institute of Technology in the School of Chemical & Biomolecular Engineering in August 2009 and is now Professor and Robert “Bud” Moeller Faculty Fellow. Prof. Walton currently serves as an Associate

Editor for the ACS Journal Industrial & Engineering Chemistry Research, and is the Director and Lead PI of Georgia Tech’s DOE Energy Frontier Research Center, UNCAGE-ME.



Sofia Calero is Full Professor at the University Pablo de Olavide (UPO), in Seville and group leader of the RASPA group. She received her Ph.D. in Chemistry in 2000 from the University Complutense of Madrid, Spain. From 2001–2003 she was a postdoctoral researcher as a Marie Curie Fellow in the Chemical Engineering Department of the University of Amsterdam, in the Netherlands. In

2004 she moved to the UPO as Ramon y Cajal Fellow, Profesor Contratado Doctor (2006), Profesor Titular de Universidad (2009) and Catedrático de Universidad (2017). She serves on the board of directors of the International Adsorption Society (IAS).

vapor–liquid experimental data, and fitting on thermal expansion and elasticity properties. We then review specific force fields that have been developed for flexible zeolites and MOFs, and end with an outlook and discuss the aspects involved in implementing force field on graphic cards, and designing and constructing a force field in terms of energy versus entropy, bottom-up versus a top-down approach, and the advantage and pitfalls when using machine learning approaches to implement force fields.

2. Ab Initio vs Classical Simulations

2.1. Ab Initio

The term “ab initio” is Latin for “from the beginning.” The computational framework is derived from purely theoretical considerations with no inclusion of experimental data.^[37–42] The most common type of ab initio calculation is called a Hartree Fock (HF) calculation, an extension of molecular orbital theory, in which the correlated electron–electron repulsion is not specifically taken into account; only its average effect is included in the calculation. Most ab initio calculations make the Born–Oppenheimer approximation, which greatly simplifies the underlying Schrödinger equation by assuming that the nuclei remain fixed during the calculation. Present algorithms in computational chemistry can routinely calculate the properties of molecules that contain up to about 50 electrons with sufficient accuracy. Errors for energies can be less than a few kJ mol^{-1} . For geometries, bond lengths can be predicted within a few picometers and bond angles within 0.5 degrees. The treatment of larger molecules that contain hundreds of electrons is computationally tractable by approximate methods such as density functional theory (DFT).

In DFT, the total energy is expressed in terms of the total electron density, rather than the wave function.^[43–46] In this type of calculation, there is an approximate Hamiltonian and an approximate expression for the total electron density. DFT methods can be accurate for little computational cost. DFT methods that do not include HF exchange can scale better than HF. Some methods combine the density functional exchange functional with the HF exchange term and are known as hybrid functional methods. Despite recent improvements, there are still difficulties in using DFT to accurately describe inter molecular interactions, especially van der Waals forces (dispersion). They cannot be reproduced at the HF level because they are a pure correlation effect. Exact DFT would include all correlation effects, including the dispersion force. However, exact DFT contains a functional of the electron density which is unknown, and most probably unknowable (except by solving the full Schrödinger equation). DFT is currently unable to accurately describe adsorption in nanoporous materials. Phase equilibria are extremely sensitive to potential parameters of empirically determined *classical* force field. Importantly, this also applies to *first-principles* parameters of the system.^[47] The development of new DFT methods to overcome this problem is an active research topic. Current research direction include alteration to the functional,^[48] the inclusion of additive terms,^[49] van der Waals density functional (VDW-DF),^[50] and the use of maximally localized Wannier functions.^[51]

VDW-DF has been used to study CO_2 binding in ZIFs.^[52] For reaching chemical accuracy for the energies of adsorption, hybrid high-level (MP2) low-level (DFT + D) quantum chemical method is argued to be required.^[53,54] Gradient-corrected density functional with empirical dispersion corrections (PBE-D) as well as a nonlocal correlation functional (vdW-DF2) were found to reproduce available experimental adsorption enthalpies of alkanes in silicalite and HZSM-5 zeolite only in a limited fashion.^[55] Regular GGAs do not describe dispersion-governed crystal types well.^[56] Currently, dispersion-corrected DFT produces a stronger associated liquid phase when compared to experiment.^[57–59] Fang et al. systematically compared several dispersion-corrected DFT methods for CO_2 in zeolites.^[60] The semi-empirical method by Grimme^[48] was found to give the best agreement with high level quantum and experimental results. The other methods substantially overestimated the binding energies.

The non-local van der Waals density functional (vdW-DF) of Dion et al.^[50] is a very promising scheme for the efficient treatment of dispersion bonded systems.^[61,62] The Strongly Constrained and Appropriately Normed (SCAN) semilocal density functional^[63] is a major improvement over PBE (and much more so over LSDA), at nearly the same computational cost. rVV10^[64] is a simple revision of the VV10 nonlocal density functional by Vydrov and Van Voorhis^[65] for dispersion interactions. Unlike the original functional this modification allows nonlocal correlation energy and its derivatives to be efficiently evaluated in a plane wave framework. Peng et al. constructed a “best-of-both-worlds” van der Waals (vdW) density functional: SCAN + rVV10.^[66] The resultant vdW density functional yielded excellent interlayer binding energies and spacings, as well as intralayer lattice constants in 28 layered materials. Trans et al. tested a wide variety of nonlocal van der Waals functionals for solids. Only three functionals provided reasonably small errors for all properties (intralayer and interlayer lattice constants and interlayer binding energy) and for most solids: PBE + rVV10L, SCAN + rVV10, and rev-vdW-DF2.^[67]

In flexible MOFs, there is a delicate balance between dispersion and entropy that governs their transformation.^[68] Quantum simulations of adsorption (binding energies) of adsorbates in nanoporous materials are relatively scarce. The reason is that adsorption is due to dispersion interactions and therefore a very high level of theory is needed. This significantly reduces the amount of atoms that one can reliably include in the simulations. Therefore most such studies compute binding energies of an adsorbate with a small, representative cluster. However, this is a severe approximation for nanoporous materials where the molecule is usually (strongly) confined. It is more common to see quantum mechanics applied to the properties of the host material itself, for example, infra-red (IR) spectra, unit cell size, and elastic constants.^[69–71] Moreover, quantum mechanical tools are essential to obtain reliable atomic information on the positions of framework atoms when experimental crystallographic data are missing or only partially known. However, to study adsorption in nanoporous materials at, for example, physiological conditions we must resort to classical methodology. Three main categories of classical simulation methodologies are Molecular Mechanics (MM), Molecular Dynamics (MD), and Monte Carlo (MC).

2.2. Molecular Mechanics (MM)

The basic ideas behind MM date back to the 1930s and 1940s.^[72–74] MM assumes that matter consists of atoms and for every set of positions of the atoms the potential energy can be defined. In 1950–1970 the advent of computers caused MM methods to grow in popularity at a rapid rate, and currently MM is one of the standard methods of structural chemistry. The MM model is defined in terms of an energy equation. The classical molecular energy \mathcal{U} can be described as an Taylor expansion in bonds, bends, torsions, etc.^[75,76]

$$\begin{aligned} \mathcal{U} = & \sum_{\text{bonds}} \mathcal{U}_b(r) + \sum_{\text{bends}} \mathcal{U}_\theta(\theta) + \sum_{\text{torsions}} \mathcal{U}_\phi(\phi) + \sum_{\text{out-of-plane bends}} \mathcal{U}_\chi(\chi) \\ & + \sum_{\text{non-bonded}} \mathcal{U}_{nb}(r) + \sum_{\text{bond-bond}} \mathcal{U}_{bb'}(r, r') + \sum_{\text{bond-bend}} \mathcal{U}_{b\theta}(r, \theta) \\ & + \sum_{\text{bend-bend}} \mathcal{U}_{\theta\theta'}(\theta, \theta') + \sum_{\text{bond-torsion}} \mathcal{U}_{r\phi}(r, \phi, r') \\ & + \sum_{\text{bend-torsion}} \mathcal{U}_{\theta\phi}(\theta, \phi, \theta') + \dots \end{aligned} \quad (1)$$

This expansion captures all the chemical entities we are used to, such as atoms, bonds, bond-angles and torsions, and physical properties like equilibrium structures, vibrational spectra, etc. On top of the terms in Equation (1) one can add ad hoc terms, such as hydrogen bonding, that are not adequately accounted for otherwise. Equation (1) is historically referred to as an *force field*. The name arose from the lowest order approximation using only springs with *force constants*. The prototypical MM application is energy minimization, for example to obtain the best binding modes of two 3D molecules (“docking”). Some examples of its uses are:^[77]

- To obtain reasonable input geometries for more computational expensive types of calculations,
- To obtain good geometries (and perhaps energies) for small- to medium-sized molecules,
- To calculate the geometries and energies of very large molecules, usually polymeric biomolecules (proteins and nucleic acids),
- To generate the potential energy function under which molecules move, for MD or MC calculations,
- As a (usually quick) guide to the feasibility of, or likely outcome of, reactions in organic synthesis.

2.3. Molecular Dynamics (MD)

MD mimics nature by computing the positions $\mathbf{r}(t)$ and velocities $\mathbf{v}(t)$ of particles as a function of time t according to the equations of motion of Newton: $\mathbf{f} = m\mathbf{a}$, which relates the force \mathbf{f} to the mass m times the acceleration \mathbf{a} .^[78] MD neglects zero point motion and the quantisation of vibration. It is therefore not applicable to the very low temperature regime (at low temperature a path-integral formalism is required).^[79] The first step in a numerical scheme to solve this equation is to rewrite it as two coupled

first-order differential equations:

$$\frac{d\mathbf{v}(t)}{dt} = \mathbf{a}(t) \quad \mathbf{a}(t) \equiv \mathbf{a}(\mathbf{r}, \mathbf{v}, t) \quad (2)$$

$$\frac{d\mathbf{r}(t)}{dt} = \mathbf{v}(t) \quad (3)$$

We can discretize the time in small time steps Δt where the current time t at step n is $t_n = t_0 + n\Delta t$ with respect to an arbitrary time origin t_0 . To integrate the equations of motion we need to compute the values of \mathbf{r}_{n+1} and \mathbf{v}_{n+1} at time $t_{n+1} = t_n + \Delta t$. A commonly used integration scheme is the velocity Verlet scheme^[78]:

1. $\mathbf{v}(t + \frac{1}{2}\Delta t) = \mathbf{v}(t) + \frac{1}{2}\mathbf{a}(t)\Delta t$
2. $\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t + \frac{1}{2}\Delta t)\Delta t$
3. Compute $\mathbf{a}(t + \Delta t)$ ($= \mathbf{f}(t + \Delta t)/m$) from the interactions using positions $\mathbf{r}(t + \Delta t)$
4. $\mathbf{v}(t + \Delta t) = \mathbf{v}(t + \frac{1}{2}\Delta t) + \frac{1}{2}\mathbf{a}(t + \Delta t)\Delta t$.

The integration scheme is stable for sufficiently small time steps Δt . The permitted time step depends on the system. Stiff bonds require a smaller time step (like 0.5 fs), while for rigid molecules a larger time step can be used (like 2 fs). The main criteria to validate the chosen time step is the long-time conservation of the conserved quantity of the equations of motion. The main use of MD is to generate a thermodynamic ensemble. Newton’s equation of motion corresponds to the NVE-ensemble, constant number of particle N , constant volume V , and constant energy E . Other ensembles can be generated by extending the equation of motion with thermo- and barostats.^[80,81] MD can be applied to comparatively large systems for relatively long times compared to QM. In contrast to MC, MD can be used to compute dynamic properties (time correlation functions and transport properties). Practical aspects of MD are reviewed in refs. [36,82,83] and further details on the algorithms and computational aspects can be found in refs. [78,84–87].

2.4. Monte Carlo (MC)

In the Markov Chain Monte Carlo method (MCMC)^[88,89] the system evolves from state to state by applying “MC moves” that modifies the current state, denoted as “old” state, to generate a “new” state. The “acceptance rule” for the MC moves guarantees that all states of the Markov chain are visited with the appropriate probability. Hence, no weighting is necessary, and averages can be computed as averages over the chain. For an infinite Markov chain the expression is exact. The MC scheme makes use of the fact that only the relative probability of visiting points in configuration space is needed, *not* the absolute probability. Common moves are to translate and/or rotate a molecule. Such an attempt can be “accepted” or “rejected” by the acceptance rule, which means that the state has changed or that the new state is simply equal to the old one, respectively. Note that in MC, for *particle*-moves (like translation and rotation), the energy difference used in the MC moves has to be computed only for the atoms that move. Typical MC *system*-moves are volume changes and MC/MD hybrid moves, which are much more expensive

since the energy has to be recomputed for the entire system. Advantages of MC are that (for most moves) only energy evaluations are required and that different ensembles are easily handled. Examples are the ensembles used to compute adsorption, the Gibbs ensemble and the grand-canonical ensemble, where the chemical potential is imposed and the number of particles of the adsorbed-phase fluctuates. Downsides of MC are that MC is considerably harder to apply to chemically complex molecules than MD, and highly correlated movements are hard to simulate. Also only static properties can be computed in MC, because there is no time involved in an MC move. Time does not enter in the partition function (i.e., no velocities, only positions) and therefore the involved units are energy and distance (not time). However, it is possible to artificially enter time using schemes like “kinetic MC.” Practical aspects of MC are reviewed in ref. [90] and further details on the algorithms and computational aspects can be found in refs. [78,84,91–94].

3. Functional Form and Design

3.1. Functional Form

The energy equation Equation (1) is the central workhorse in MM, MD, and MC. The majority of the computational time will be spent here, in computing energies and energy derivatives (e.g., the forces). The term “force field” refers to the functional form and parameter sets of the energy, that is, a force field consists of two parts:

1. An analytical expression for the inter-atomic potential energy $U(\mathbf{r})$ as a function of the atomic coordinates \mathbf{r} .
2. Definitions and parameters for “atom types.” Chemical elements are classified based on bonding and environment (e.g., carbon: sp^3 , sp^2 , sp , aromatic, carbonyl). Parameters are assigned based on the atom types involved (the CC bond length and force constant are different for sp^3 - sp^3 vs sp^2 - sp^2).

The forces are given by the gradients of the potential energy with respect to the internal coordinates of the molecule.

Several types of force fields can be distinguished:

- Rule-based force fields
Force fields that have generally broad applicability across the periodic table. This type of force field utilizes generic rules to generate parameters for a range of interactions from a small set of atomic-based parameters (in contrast to using a set of explicit parameters determined for each type of interaction). The rule-based strategy provides a systematic and extensible way to develop a force field.
- First-generation, class I force fields
Harmonic functions for stretch and bend, no cross-terms, Lennard–Jones for dispersion interactions. Class I force fields are also called “harmonic” or “diagonal” and are common for large systems like proteins or DNA because of their computational efficiency. It is believed that class I force fields are sufficiently accurate for large bio-molecules near room-temperature.
- Second-generation, class II force fields

Class II force fields add cross-terms, cubic, or quartic expansions of stretch and bend, exponential-type potentials for dispersion interactions, all in an effort to increase the accuracy of the potential energies, the structure, and especially the vibrational frequencies of the system. Because of the additional complexity and computational expense, class II force fields are intended for materials science and small to medium molecules.

- Third-generation, class III force fields

Type III force fields include hyperconjugation and polarization.^[95–97] Polarization is included using either induced dipoles (AMOEBA, SIBFA, NEMO, and others), the drude model (polarizable CHARMM) or fluctuating charges.

Table 1 lists a selection of common force fields.

There are many advantages of the classical approach. Force field-based simulations can handle large systems, and are several orders of magnitude faster (and cheaper) than quantum-based calculations. But also the analysis of the energy contributions can be done at the level of individual classes of interactions, and the energy expression can be modified to bias the computations. However, classical approaches have difficulties with electronic transitions (photon absorption), electron transport phenomena, proton transfer (acid/base reactions), and chemical reactions. Therefore vibrational properties like intensities cannot be described by classical MD, and it is not possible to compute electronic (band structure, eDOS) or magnetic properties (NMR spectra). A recent approach to handle chemical reactions in classical simulations is the reactive force field (ReaxFF)^[161] which is bond order-based. The following characteristics of force fields are a trade-off:

- Accuracy (how accurate are properties of interest computed),
- Transferability (prediction of properties or systems outside of the fitting set),
- Computational efficiency.

3.2. Derivatives

Potential energy functions and their derivatives play an essential ingredient of molecule simulations. For MC methodology the potential energy function U suffice, but MD requires forces \mathbf{f} and hence first derivatives of the potential energy function, that is, $\mathbf{f} = -\partial U / \partial \mathbf{r}$. Computing normal modes requires the Hessian matrix, a matrix of second derivatives with respect to position. The Hessian matrix is particularly important, not only for geometry optimization, but also for the characterization of stationary points as minima, transition states, or hilltops, and for the calculation of IR spectra.^[77] A generalized Hessian matrix can be formed by including strain-derivatives such that also cell shape and volume changes are incorporated. Elastic constants at zero Kelvin can directly be computed at machine precision from the generalized Hessian (see Section 5.1), and the Hessian is also of great use in efficient minimization methodology^[162,163] (see Section 5.3). Expressions of first and second derivatives for stretching, linear, nonlinear, and out-of-plane bending and torsional motion have appeared in many articles.^[164–176] First and second derivatives for the Ewald summation have also appeared

Table 1. Examples of common classical atomic force fields.

Acronym	Full name	Class	Application
AMBER	Assisted Model Building with Energy Refinement	I	biochemistry, nucleic acids and proteins ^[98–105]
AMB/OPLS	AMBER/OPLS	I	proteins, ^[106] all-atom hydrocarbons, ^[107] organic liquids, ^[108] carbohydrates ^[109]
AMOEBA	Atomic Multipole Optimized Energetics for Biomolecular Applications	III	Nucleic acids ^[110]
CFF91	Consistent Force Field	II	hydrocarbons, proteins, protein-ligand interactions ^[111]
CFF93	Consistent Force Field	II	alkyl group and alkanes ^[112]
CFF95	Consistent Force Field	II	general-purpose ^[113,114]
CHARMM	Chemistry at HARvard Molecular Mechanics	I	biochemistry, proteins, nucleic acids, organic molecules, metal-ions ^[115–118]
COMPASS	Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies	II	small molecules and polymers ^[119,120]
CVFF	Consistent-Valence Force Field	I	biochemistry, peptides and proteins ^[121]
DREIDING	Generic force field	rule	general-purpose ^[14]
DRF90	Direct Reaction Field	III	small (solvent) molecules ^[122]
ECEPP/5	Empirical Conformational Energy Program for Peptides	I	peptides, proteins, and organic molecules ^[123]
ESFF	Extensible Systematic Force Field	rule	organic, inorganic, and organometallic systems ^[124]
GAFF	Generalized Amber Force Field	I	general-purpose ^[125]
GROMOS	GROnigen MOlecular Simulation	I	biochemistry, proteins, lipids, nucleic acids, polymers, water-protein, ^[126] oligosaccharides, ^[127] biomolecules in explicit water ^[128]
INTERFACE-PCFF	Thermodynamically CFF	II	inorganic, organic, and biological nanostructures ^[129]
OPLS	Optimized Potentials for Liquid Simulations	I	proteins, nucleic acids, liquid hydrocarbons, ^[130] amides and peptides, ^[131] liquid sulfur compounds, ^[132] liquid alcohols ^[133]
PCFF	Polymer Consistent Force-Field	II	organic polymers, (inorganic) metals, and zeolites ^[134]
PIPF	Polarizable Intermolecular Potential Function	III	liquid amides and alkanes ^[135]
MM2	Molecular Mechanics	II	hydrocarbons, ^[136] hydrogen-bonding, ^[137] peptides ^[138]
MM3	Molecular Mechanics	II	amides, polypeptides and proteins, ^[139] hydrogen-bonding ^[140]
MM4	Molecular Mechanics	II	saturated hydrocarbons, ^[141,142] conjugated hydrocarbons, ^[143] alkenes, ^[144,145] sulfides and mercaptans ^[146]
MMFF94	the Merck Molecular Force Field	II	general-purpose, all organic molecules for drug design ^[147–150]
NEMO	Non-Empirical Molecular Orbital	III	small molecules and systems ^[151]
SIBFA	Sum of Interactions Between Fragments Ab initio computed	III	small molecules and flexible proteins ^[152]
TraPPE	TRAnsferable Potentials for Phase Equilibria	I	small molecules ^[153–159]
UFF	Universal Force Field	rule	general-purpose ^[15]
X-Pol	Explicit POLarization	III	fluid systems ^[160]

in literature.^[177,178] Both the gradient and the Hessian can be evaluated in $\mathcal{O}(N^2)$ steps.^[175] Third-order derivatives are used in free energy minimization.^[162] By combining free-energy minimization with empirical fitting, it is possible to determine interatomic potentials with correct treatment of thermal effects and the zero-point energy.^[179] In free energy minimization the derivatives of the phonon frequencies are needed, which require the phased third derivatives of the energy with respect to either three Cartesian coordinates, or two Cartesian coordinates and one strain, for internal and external variables, respectively.^[162] Higher-order derivatives are also required for the analysis of thermal expansion at zero Kelvin. A harmonic approach does not suffice in this case because there will be no lattice expansion due to the symmetrical displacement of atoms about their equilibrium positions. Note

that the energy and derivative computation can be parallelized with nearly 100% efficiency for *internal* coordinates.^[180] The notational convention in this review is that $\alpha, \beta, \gamma, \nu = \{x, y, z\}$ denote the vector components, latin subscripts i, j, k, l denote atom indices, \mathbf{r}_i denotes the position of atom i , and $\varepsilon_{\alpha\beta}$ denotes the Lagrangian strain-tensor.

Derivatives of a function $f(\gamma)$, where γ is a function of x_1, x_2, \dots , are given by applications of the chain rule

$$\frac{\partial f(\gamma)}{\partial x_1} = \frac{\partial f(\gamma)}{\partial \gamma} \frac{\partial \gamma}{\partial x_1} \quad (4)$$

$$\frac{\partial^2 f(\gamma)}{\partial x_1 \partial x_2} = \frac{\partial f(\gamma)}{\partial \gamma} \frac{\partial^2 \gamma}{\partial x_1 \partial x_2} + \frac{\partial^2 f(\gamma)}{\partial \gamma^2} \frac{\partial \gamma}{\partial x_1} \frac{\partial \gamma}{\partial x_2} \quad (5)$$

$$\begin{aligned} & \frac{\partial^3 f(\gamma)}{\partial x_1 \partial x_2 \partial x_3} \\ &= \frac{\partial f(\gamma)}{\partial \gamma} \frac{\partial^3 \gamma}{\partial x_1 \partial x_2 \partial x_3} \\ &+ \frac{\partial^2 f(\gamma)}{\partial \gamma^2} \left(\frac{\partial \gamma}{\partial x_1} \frac{\partial^2 \gamma}{\partial x_2 \partial x_3} + \frac{\partial \gamma}{\partial x_2} \frac{\partial^2 \gamma}{\partial x_1 \partial x_3} + \frac{\partial \gamma}{\partial x_3} \frac{\partial^2 \gamma}{\partial x_1 \partial x_2} \right) \\ &+ \frac{\partial^3 f(\gamma)}{\partial \gamma^3} \frac{\partial \gamma}{\partial x_1} \frac{\partial \gamma}{\partial x_2} \frac{\partial \gamma}{\partial x_3} \end{aligned} \quad (6)$$

Here, f could be a potential energy function $f(r_{ij})$ that depends on the distance r_{ij} which in turn depends on the atomic positions $\mathbf{r}_i, \mathbf{r}_j$ of the atoms i and j involved, or a potential energy function $f(\theta_{ijk})$ that depends on the bend-angle θ_{ijk} which in turn depends on the atomic positions $\mathbf{r}_i, \mathbf{r}_j$ and \mathbf{r}_k for the three involved atoms i, j, k , etc. Hence, using the chain rule we can split the derivative in a derivative of the potential energy function form with respect to the r, θ and/or ϕ and the derivative of these with respect to the atomic positions.^[170]

Computing the gradient $\nabla_{\mathbf{r}_i} \mathcal{U}(r_{ij})$ of a pair potential $\mathcal{U}(r_{ij})$ requires the use of the chain rule

$$\nabla_{\mathbf{r}_i} \mathcal{U}(r_{ij}) = \left(\frac{d\mathcal{U}(r_{ij})}{dr_{ij}} \right) \nabla_{\mathbf{r}_i} r_{ij} \quad (7)$$

Hence we need derivatives of the distance vector r_{ij}

$$r_{ij} = \sqrt{(r_i^x - r_j^x)^2 + (r_i^y - r_j^y)^2 + (r_i^z - r_j^z)^2} \quad (8)$$

We will encounter $\partial r_{ij} / \partial \mathbf{r}_q$ and $\partial^2 r_{ij} / \partial \mathbf{r}_p \partial \mathbf{r}_q$ in single and double derivatives, respectively. We find^[176]

$$\frac{\partial r_{ij}}{\partial \mathbf{r}_q} = \delta_{iq} \frac{\mathbf{r}_{ij}}{r_{ij}} - \delta_{jq} \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (9)$$

$$\frac{\partial^2 r_{ij}}{\partial \mathbf{r}_p \partial \mathbf{r}_q} = \delta_{iq} \left(\frac{\partial}{\partial \mathbf{r}_p} \frac{\mathbf{r}_{ij}}{r_{ij}} \right) - \delta_{jq} \left(\frac{\partial}{\partial \mathbf{r}_p} \frac{\mathbf{r}_{ij}}{r_{ij}} \right) \quad (10)$$

$$= (\delta_{iq} - \delta_{jq}) (\delta_{ip} - \delta_{jp}) \left[\frac{r_{ij}^2 - \mathbf{r}_{ij}^2}{r_{ij}^3} \right] \quad (11)$$

where δ is the Kronecker delta. As an example, consider the harmonic bond potential $\mathcal{U}(r) = \frac{1}{2} k(r - r_0)^2$. The first derivative is given by

$$\frac{\partial \mathcal{U}(r)}{\partial \mathbf{r}_i} = \frac{\partial \mathcal{U}(r)}{\partial r} \frac{\partial r}{\partial \mathbf{r}_i} = k(r - r_0) \frac{\mathbf{r}_{ij}}{r} \quad (12)$$

For strain derivatives we first define the metric tensor by^[181,182]

$$M_{\alpha\beta} = h_{\alpha\mu} \tilde{h}_{\mu\beta}^{-1} \quad (13)$$

where the matrix $h_{\alpha\mu} = \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ is composed of the cell vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ of the simulation cell and $\tilde{\mathbf{h}}$ refers to \mathbf{h} in the reference

state (usually taken as the zero strain configuration). Using $M_{\alpha\beta}$, a homogeneous deformation (i.e., strain) can be applied by relating the position of an atom to its original position in the reference state of the body

$$\mathbf{r}^\alpha = M_{\alpha\beta} \tilde{\mathbf{r}}^\beta \quad (14)$$

The strain tensor $\varepsilon_{\alpha\beta}$ can be expressed in terms of the metric tensor as^[181,182]

$$\varepsilon_{\alpha\beta} = \frac{1}{2} \left(M_{\alpha\mu}^\ddagger M_{\mu\beta} - \delta_{\alpha\beta} \right) \quad (15)$$

where \ddagger denotes the transpose of the tensor. Derivatives of any general thermodynamic variable A that is an explicit function of $M_{\alpha\beta}$ can be derived as^[181,182]

$$\frac{\partial A}{\partial \varepsilon_{\alpha\beta}} = \frac{1}{2} \left(\frac{\partial A}{\partial M_{\alpha\gamma}^\ddagger} M_{\gamma\beta}^{\ddagger-1} + M_{\alpha\gamma}^{-1} \frac{\partial A}{\partial M_{\gamma\beta}} \right) \quad (16)$$

$$\frac{\partial^2 A}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} = \frac{1}{4} (D_{\alpha\beta\mu\nu} + D_{\beta\alpha\mu\nu} + D_{\beta\alpha\nu\mu} + D_{\alpha\beta\nu\mu}) A \quad (17)$$

where

$$D_{\alpha\beta\mu\nu} A = M_{\alpha\gamma}^{-1} \frac{\partial^2 A}{\partial M_{\gamma\beta}^\ddagger \partial M_{\mu\omega}^\ddagger} M_{\omega\nu}^{\ddagger-1} - M_{\alpha\gamma}^{-1} M_{\gamma\nu}^{\ddagger-1} M_{\beta\omega}^{-1} \frac{\partial A}{\partial M_{\omega\mu}} \quad (18)$$

An overview of popular function forms for atomic potentials can be found in ref. [162]. In the next sections, we highlight the first and second derivatives of the most common potentials used in zeolite and MOF work. These expressions are taken from literature but written down using a uniform notation (and hence changed from the original notation). An implementation for the analytical expressions as well as their numerical validation can be found as Supporting Information.

3.3. Bond-Stretching Potentials

Covalent forces are complicated in nature as the electronic structure changes dramatically upon bond formation. The strength-range of the covalent bonds is about 200–800 kJ mol⁻¹, and the force operates over very short distances of the order of 1–2 Å.^[183] MOFs consist of a coordinating metal atom (or cluster of atoms) with one or more electron rich ligands attached to it by so-called “coordination bonds.” These bonds are weaker (between 50 and 200 kJ mol⁻¹) and the force is directional acting on a relevant distance of 1.5–2.5 Å.^[184] In a course-grained approach, bonding can be modeled by bond-potentials that describe the change in energy as the bond stretches and compresses. At high temperatures the asymmetry is apparent as it is harder to compress a bond than it is to stretch it. The bond force constants are very high and motion is usually restricted to the harmonic well of the potential.

The bond vector \mathbf{r}_{ij} can be defined based on the position of the atoms $\mathbf{r}_i = \mathbf{r}_i - \mathbf{r}_j$ (see **Figure 1**) and points from \mathbf{r}_j toward \mathbf{r}_i . The distance r_{ij} between atom i and j is $|r_{ij}| = \sqrt{\mathbf{r}_{ij} \cdot \mathbf{r}_{ij}}$. The

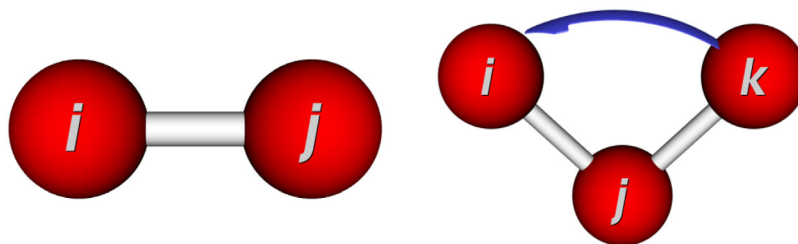


Figure 1. The bond distance r (left) defined by the distance formed by atoms i - j and (right) the bend angle θ defined by the angle between the vector i - j and k - j .

Table 2. Common bond potential functional forms.

TraPPE ^[153]	Fixed bond-lengths
OPLS ^[106] , UFF ^[15]	$U(r_{ij}) = \frac{1}{2}k(r_{ij} - r_0)^2$
AMBER ^[98] , GAFF ^[125] , CHARMM ^[116]	$U(r_{ij}) = k(r_{ij} - r_0)^2$
MM3 ^[139] , AMOEBA ^[110]	$U(r_{ij}) = \frac{1}{2}k(r_{ij} - r_0)^2[1 - 2.55(r_{ij} - r_0) + \frac{7}{12}2.55^2(r_{ij} - r_0)^2]$
CVFF ^[121] , UFF ^[15]	$U(r_{ij}) = k(1 - e^{-\alpha(r_{ij}-r_0)})^2$
CFF91 ^[111]	$U(r_{ij}) = k_2(r_{ij} - r_0)^2 + k_3(r_{ij} - r_0)^3 + k_4(r_{ij} - r_0)^4$

most common functional form for the bond potential U_b is the expansion

$$U_b(r) = \frac{k_1}{2}(r - r_0)^2 + \frac{k_2}{3}(r - r_0)^3 + \frac{k_3}{4}(r - r_0)^4 + \dots \quad (19)$$

where r is the current bond length, r_0 the reference bond length, and k_1, k_2, k_3 are the bond force constants. Often the expansion is truncated after the first term. The “Morse” potential is anharmonic and often provides a much better description of the potential

$$U_b(r) = D\left(1 - e^{-\alpha(r-r_0)}\right)^2 \quad (20)$$

Expanding around the equilibrium value leads to

$$U_b(r) = D\alpha^2(r - r_0)^2\left[1 - \alpha(r - r_0) + \frac{7}{12}\alpha^2(r - r_0)^2 \dots\right] \quad (21)$$

The first term is the harmonic potential (with $k = 2D\alpha^2$) and for organic structures where distortions from equilibrium are small, the difference between the potentials is small. For larger deviations the Morse potential provides a significantly better description. The Morse potential provides a restoring force which goes to zero at long distances. For minimization starting far from equilibrium this could result in non-convergence. Some force fields solved this problem by using modifications of Hook’s law. MM2 added a cubic term making the bond an-harmonic. However, this leads to large negative energies for poor initial geometries with large distortions. MM3 added the quartic term to solve this. Note the 7/12 terms in the MM2/3 functional forms originate from the Taylor expansion of the Morse potential, and the cubic and quartic terms are chosen to mimic the Morse potentials for moderate distortions. Some common bond potentials are listed in **Table 2**.

It is convenient to separate the bond potential functional form from the other derivatives so that a more general framework can be used. We therefore define the bond potential functional derivatives as

$$f_1 = \frac{dU(r)}{dr} \quad (22)$$

$$f_2 = \frac{d^2U(r)}{dr^2} \quad (23)$$

We define $\mathbf{u} = \mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $u = |\mathbf{u}| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$. The derivatives of $U(r)$ with respect to Cartesian coordinates are given by^[164]

$$\frac{\partial U(r)}{\partial r_i^\alpha} = f_1 \frac{u^\alpha}{u} \quad (24)$$

$$\frac{\partial U(r)}{\partial r_j^\alpha} = -f_1 \frac{u^\alpha}{u} \quad (25)$$

$$\frac{\partial^2 U(r)}{\partial r_i^\alpha \partial r_i^\beta} = \frac{\partial^2 U(r)}{\partial r_j^\alpha \partial r_j^\beta} = \left(f_2 - \frac{f_1}{u}\right) \frac{1}{u^2} u^\alpha u^\beta + \delta_{\alpha\beta} \frac{1}{u} f_1 \quad (26)$$

$$\frac{\partial^2 U(r)}{\partial r_i^\alpha \partial r_j^\beta} = \frac{\partial^2 U(r)}{\partial r_j^\alpha \partial r_i^\beta} = -\left(f_2 - \frac{f_1}{u}\right) \frac{1}{u^2} u^\alpha u^\beta - \delta_{\alpha\beta} \frac{1}{u} f_1 \quad (27)$$

The derivatives of $U(r)$ with respect to strain are derived by Ray^[185] and Lutsko^[181]

$$\frac{\partial U(r)}{\partial \epsilon_{\alpha\beta}} = f_1 \frac{1}{u} u^\alpha u^\beta \quad (28)$$

$$\frac{\partial U(r)}{\partial \epsilon_{\alpha\beta} \partial r_i^\gamma} = \left(f_2 - \frac{f_1}{u}\right) \frac{1}{u^2} u^\alpha u^\beta u^\gamma + \frac{1}{u} (\delta_{\alpha\gamma} f_1 u^\beta + \delta_{\beta\gamma} f_1 u^\alpha) \quad (29)$$

$$\frac{\partial U(r)}{\partial \epsilon_{\alpha\beta} \partial r_j^\gamma} = -\left(f_2 - \frac{f_1}{u}\right) \frac{1}{u^2} u^\alpha u^\beta u^\gamma - \frac{1}{u} (\delta_{\alpha\gamma} f_1 u^\beta + \delta_{\beta\gamma} f_1 u^\alpha) \quad (30)$$

$$\begin{aligned} \frac{\partial U(r)}{\partial \epsilon_{\alpha\beta} \partial \epsilon_{\mu\nu}} &= \left(f_2 - \frac{f_1}{u}\right) \frac{1}{u^2} u^\alpha u^\beta u^\mu u^\nu \\ &+ \frac{1}{2} \left[\delta_{\beta\nu} \frac{\partial U(r)}{\partial \epsilon_{\alpha\mu}} + \delta_{\alpha\nu} \frac{\partial U(r)}{\partial \epsilon_{\beta\mu}} + \delta_{\beta\mu} \frac{\partial U(r)}{\partial \epsilon_{\alpha\nu}} + \delta_{\alpha\mu} \frac{\partial U(r)}{\partial \epsilon_{\beta\nu}} \right] \end{aligned} \quad (31)$$

The explicit strain derivative terms appearing in the last term in Equation (31) arise from the action of the strain derivative on the explicit cell dependence of the microscopic stress.^[181]

Table 3. Common Urey–Bradley potential functional forms.

CHARMM ^[116] , AMOEBA ^[110]	$\mathcal{V}(\theta_{ijk}) = k(r_{ik} - r_0)^2$
---	---

3.4. Urey–Bradley

The Urey–Bradley term is a special case of cross-terms and consists of a harmonic potential as a function of the distance between the (non-bonded) atoms i and k of an i - j - k angle (i.e., as if there exists an extra bond stretching term between atoms i and k). This Urey–Bradley term is coupled with the i - j and j - k bond stretching terms and the i - j - k angle bending term through basic trigonometric relationships, and is a computationally elegant way of reproducing bond-bond coupling effects in vibrational spectra. However, compared to the more conventional cross-terms, it is poorly transferable to various combinations of three atoms with different reference values for the bonds or angle. Some common Urey–Bradley potentials are listed in **Table 3**. Derivatives are the same as for bond-potentials.

3.5. Bond Bending Potentials

We define

$$\mathbf{u} = \mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j, \quad u = |\mathbf{u}| = \sqrt{\mathbf{u} \cdot \mathbf{u}} \quad (32)$$

$$\mathbf{v} = \mathbf{r}_{kj} = \mathbf{r}_k - \mathbf{r}_j, \quad v = |\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} \quad (33)$$

The bend angle θ_{ijk} is formed by the angle between bond vector \mathbf{u} and \mathbf{v} (see Figure 1)

$$\cos \theta_{ijk} = \frac{\mathbf{u} \cdot \mathbf{v}}{uv} \quad (34)$$

To describe bond bending between three atoms one can use

$$\mathcal{V}(\theta) = \frac{k_1}{2}(\theta - \theta_0)^2 + \frac{k_2}{3}(\theta - \theta_0)^3 + \frac{k_3}{4}(\theta - \theta_0)^4 + \dots \quad (35)$$

Angles are much softer than bonds, especially in zeolites where a Si–O–Si angle ranges between 135 and 180 degrees.^[186] Also in three- or four-membered rings (cyclopropane, cyclobutane) large deviations from the usual bond angles are observed. A problem with all polynomial representations of angles is that angles of 180 degrees results in a singular point (unless the reference angle is 180 degrees). The case of 0 degree is not possible due to repulsion of the i and k atoms in the i - j - k bend. The singularity is due to a factor $1/\sin(\theta)$ in the force expression. A common solution is to use a trigonometric function^[187]

$$\mathcal{V}(\theta) = \frac{1}{2}k[\cos(\theta) - \cos(\theta_0)]^2 \quad (36)$$

Note that close to the maximum this potential has no restoring force, but for small distortions this is not a problem. The MM force fields use higher order terms. A six power term was needed to describe the highly bent bicyclo[1.1.1]pentane. Cubic terms and higher become desirable when the bending is more than

Table 4. Common bend angle potential functional forms.

TraPPE ^[153] , OPLS ^[106] , CVFF ^[121]	$\mathcal{V}(\theta_{ijk}) = \frac{1}{2}k(\theta_{ijk} - \theta_0)^2$
AMBER ^[98] , GAFF ^[125] , CHARMM ^[116]	$\mathcal{V}(\theta_{ijk}) = k(\theta_{ijk} - \theta_0)^2$
MM3 ^[139] , AMOEBA ^[110]	$\mathcal{V}(\theta_{ijk}) = \frac{1}{2}k(\theta_{ijk} - \theta_0)^2 \left(1 - 0.014(\theta_{ijk} - \theta_0) + 5.6 \times 10^{-5}(\theta_{ijk} - \theta_0)^2 - 7 \times 10^{-7}(\theta_{ijk} - \theta_0)^3 + 2.2 \times 10^{-8}(\theta_{ijk} - \theta_0)^4 \right)$
UFF ^[15]	$\mathcal{V}(\theta_{ijk}) = \frac{1}{2}k(\cos \theta_{ijk} - \cos \theta_0)^2 / \sin^2 \theta_0$
UFF9 ^[111]	$\mathcal{V}(\theta_{ijk}) = k_2(\theta_{ijk} - \theta_0)^2 + k_3(\theta_{ijk} - \theta_0)^3 + k_4(\theta_{ijk} - \theta_0)^4$
Cosine harmonic potential	$\mathcal{V}(\theta_{ijk}) = \frac{1}{2}k(\cos \theta_{ijk} - \cos \theta_0)^2$

10–15°. MM3 angle bending has been divided into in-plane and out-of-plane bending for planar trigonal centers. Otherwise the deviation from the ideal bond angle would force the atoms out of plane (e.g., C–C=O bond angle in cyclobutanone). The force constants for bending tend to be smaller than those for stretching by a factor of 5 to 8. Some common bend potentials are listed in **Table 4**.

It is convenient to separate the bend potential functional form from the other derivatives so that a more general framework can be used. We therefore define the bend potential functional derivatives as

$$f_1 = \frac{d\mathcal{V}(\theta)}{d \cos \theta} \quad (37)$$

$$f_2 = \frac{d^2\mathcal{V}(\theta)}{d \cos^2 \theta} \quad (38)$$

The derivative of the potential energy depends on the functional form of the potential energy function and can be obtained in general using^[164]

$$\frac{\partial \mathcal{V}(\theta)}{\partial r_i^\alpha} = f_1 \frac{\partial \cos \theta}{\partial r_i^\alpha} \quad (39)$$

$$\frac{\partial \mathcal{V}(\theta)}{\partial r_k^\alpha} = f_1 \frac{\partial \cos \theta}{\partial r_k^\alpha} \quad (40)$$

$$\frac{\partial \mathcal{V}(\theta)}{\partial r_j^\alpha} = -\left(\frac{\partial \mathcal{V}(\theta)}{\partial r_i^\alpha} + \frac{\partial \mathcal{V}(\theta)}{\partial r_k^\alpha} \right) \quad (41)$$

where

$$\frac{\partial \cos \theta}{\partial r_i^\alpha} = \frac{1}{u} \left(\frac{v^\alpha}{v} - \frac{u^\alpha}{u} \cos \theta \right) \quad (42)$$

$$\frac{\partial \cos \theta}{\partial r_k^\alpha} = \frac{1}{v} \left(\frac{u^\alpha}{u} - \frac{v^\alpha}{v} \cos \theta \right) \quad (43)$$

The second derivatives are given by

$$\frac{\partial \mathcal{V}(\theta)}{\partial r_i^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \theta}{\partial r_i^\alpha} \frac{\partial \cos \theta}{\partial r_i^\beta} + f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\alpha \partial r_i^\beta} \quad (44)$$

$$\begin{aligned} \frac{\partial \mathcal{U}(\theta)}{\partial r_j^\alpha \partial r_j^\beta} &= f_2 \frac{\partial \cos \theta}{\partial r_j^\alpha} \frac{\partial \cos \theta}{\partial r_j^\beta} + f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\alpha \partial r_i^\beta} + f_1 \frac{\partial^2 \cos \theta}{\partial r_k^\alpha \partial r_k^\beta} \\ &+ f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\alpha \partial r_k^\beta} + f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\beta \partial r_k^\alpha} \end{aligned} \quad (45)$$

$$\frac{\partial \mathcal{U}(\theta)}{\partial r_k^\alpha \partial r_k^\beta} = f_2 \frac{\partial \cos \theta}{\partial r_k^\alpha} \frac{\partial \cos \theta}{\partial r_k^\beta} + f_1 \frac{\partial^2 \cos \theta}{\partial r_k^\alpha \partial r_k^\beta} \quad (46)$$

$$\frac{\partial \mathcal{U}(\theta)}{\partial r_i^\alpha \partial r_j^\beta} = f_2 \frac{\partial \cos \theta}{\partial r_i^\alpha} \frac{\partial \cos \theta}{\partial r_j^\beta} - f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\alpha \partial r_i^\beta} - f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\beta \partial r_k^\alpha} \quad (47)$$

$$\frac{\partial \mathcal{U}(\theta)}{\partial r_j^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \theta}{\partial r_i^\beta} \frac{\partial \cos \theta}{\partial r_j^\alpha} - f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\beta \partial r_i^\alpha} - f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\alpha \partial r_k^\beta} \quad (48)$$

$$\frac{\partial \mathcal{U}(\theta)}{\partial r_i^\alpha \partial r_k^\beta} = f_2 \frac{\partial \cos \theta}{\partial r_i^\alpha} \frac{\partial \cos \theta}{\partial r_k^\beta} + f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\alpha \partial r_k^\beta} \quad (49)$$

$$\frac{\partial \mathcal{U}(\theta)}{\partial r_k^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \theta}{\partial r_i^\beta} \frac{\partial \cos \theta}{\partial r_k^\alpha} + f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\beta \partial r_k^\alpha} \quad (50)$$

$$\frac{\partial \mathcal{U}(\theta)}{\partial r_j^\alpha \partial r_k^\beta} = f_2 \frac{\partial \cos \theta}{\partial r_k^\beta} \frac{\partial \cos \theta}{\partial r_j^\alpha} - f_1 \frac{\partial^2 \cos \theta}{\partial r_k^\alpha \partial r_k^\beta} - f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\alpha \partial r_k^\beta} \quad (51)$$

$$\frac{\partial \mathcal{U}(\theta)}{\partial r_k^\alpha \partial r_j^\beta} = f_2 \frac{\partial \cos \theta}{\partial r_k^\alpha} \frac{\partial \cos \theta}{\partial r_j^\beta} - f_1 \frac{\partial^2 \cos \theta}{\partial r_k^\alpha \partial r_k^\beta} - f_1 \frac{\partial^2 \cos \theta}{\partial r_i^\beta \partial r_k^\alpha} \quad (52)$$

where

$$\frac{\partial^2 \cos \theta}{\partial r_i^\alpha \partial r_i^\beta} = \frac{\cos \theta}{u^2} \left(\frac{u^\alpha u^\beta}{u^2} - \delta_{\alpha\beta} \right) - \frac{1}{u^2} \frac{\partial \cos \theta}{\partial r_i^\alpha} u^\beta - \frac{1}{u^2} \frac{\partial \cos \theta}{\partial r_i^\beta} u^\alpha \quad (53)$$

$$\frac{\partial^2 \cos \theta}{\partial r_k^\alpha \partial r_k^\beta} = \frac{\cos \theta}{v^2} \left(\frac{v^\alpha v^\beta}{v^2} - \delta_{\alpha\beta} \right) - \frac{1}{v^2} \frac{\partial \cos \theta}{\partial r_k^\alpha} v^\beta - \frac{1}{v^2} \frac{\partial \cos \theta}{\partial r_k^\beta} v^\alpha \quad (54)$$

$$\frac{\partial^2 \cos \theta}{\partial r_i^\alpha \partial r_k^\beta} = \frac{1}{uv} \left(\frac{\cos \theta}{uv} u^\alpha v^\beta - \frac{1}{u^2} u^\alpha u^\beta - \frac{1}{v^2} v^\alpha v^\beta + \delta_{\alpha\beta} \right) \quad (55)$$

The strain derivatives are given by^[188]

$$\frac{\partial \mathcal{U}(\theta)}{\partial \varepsilon_{\alpha\beta}} = u^\alpha \frac{\partial \mathcal{U}(\theta)}{\partial r_i^\beta} + v^\alpha \frac{\partial \mathcal{U}(\theta)}{\partial r_k^\beta} \quad (56)$$

$$\begin{aligned} \frac{\partial \mathcal{U}(\theta)}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} &= f_2 \frac{\partial \cos \theta}{\partial \varepsilon_{\alpha\beta}} \frac{\partial \cos \theta}{\partial \varepsilon_{\mu\nu}} + f_1 \frac{\partial^2 \cos \theta}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} \\ &+ \frac{1}{2} \left[\delta_{\beta\nu} \frac{\partial \mathcal{U}(\theta)}{\partial \varepsilon_{\alpha\mu}} + \delta_{\alpha\nu} \frac{\partial \mathcal{U}(\theta)}{\partial \varepsilon_{\beta\mu}} \right. \\ &\left. + \delta_{\beta\mu} \frac{\partial \mathcal{U}(\theta)}{\partial \varepsilon_{\alpha\nu}} + \delta_{\alpha\mu} \frac{\partial \mathcal{U}(\theta)}{\partial \varepsilon_{\beta\nu}} \right] \end{aligned} \quad (57)$$

The explicit strain derivative terms appearing in the last term in Equation (57) arise from the action of the strain derivative on the

explicit cell dependence of the microscopic stress.^[181] The formulae for the strain derivatives are more conveniently written down in terms of the natural log of the cosine of the angles^[188]

$$\frac{\partial \cos \theta}{\partial \varepsilon_{\alpha\beta}} = \cos \theta \frac{\partial \ln \cos \theta}{\partial \varepsilon_{\alpha\beta}} \quad (58)$$

$$\frac{\partial^2 \cos \theta}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} = \cos \theta \left(\frac{\partial^2 \ln \cos \theta}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} + \frac{\partial \ln \cos \theta}{\partial \varepsilon_{\alpha\beta}} \frac{\partial \ln \cos \theta}{\partial \varepsilon_{\mu\nu}} \right) \quad (59)$$

where

$$\frac{\partial \ln \cos \theta}{\partial \varepsilon_{\alpha\beta}} = \frac{u^\alpha v^\beta + v^\alpha u^\beta}{\mathbf{u} \cdot \mathbf{v}} - \frac{u^\alpha u^\beta}{u^2} - \frac{v^\alpha v^\beta}{v^2} \quad (60)$$

$$\begin{aligned} \frac{\partial^2 \ln \cos \theta}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} &= - \frac{(v^\alpha u^\beta + u^\alpha v^\beta)(v^\mu u^\nu + u^\mu v^\nu)}{(\mathbf{u} \cdot \mathbf{v})^2} \\ &+ 2 \left(\frac{u^\alpha u^\beta u^\mu u^\nu}{u^4} + \frac{v^\alpha v^\beta v^\mu v^\nu}{v^4} \right) \end{aligned} \quad (61)$$

The cross-derivative of the strain derivative with respect to position can be expressed in terms of Hessian elements

$$\frac{\partial \mathcal{U}(\theta)}{\partial \varepsilon_{\alpha\beta} \partial r_i^\gamma} = u^\beta \frac{\partial \mathcal{U}(\theta)}{\partial r_i^\alpha \partial r_i^\gamma} + v^\beta \frac{\partial \mathcal{U}(\theta)}{\partial r_i^\alpha \partial r_k^\gamma} + \delta_{\gamma\beta} \frac{\partial \mathcal{U}(\theta)}{\partial r_i^\alpha} \quad (62)$$

$$\frac{\partial \mathcal{U}(\theta)}{\partial \varepsilon_{\alpha\beta} \partial r_j^\gamma} = u^\beta \frac{\partial \mathcal{U}(\theta)}{\partial r_i^\alpha \partial r_j^\gamma} + v^\beta \frac{\partial \mathcal{U}(\theta)}{\partial r_j^\alpha \partial r_k^\gamma} + \delta_{\gamma\beta} \frac{\partial \mathcal{U}(\theta)}{\partial r_j^\alpha} \quad (63)$$

$$\frac{\partial \mathcal{U}(\theta)}{\partial \varepsilon_{\alpha\beta} \partial r_k^\gamma} = u^\beta \frac{\partial \mathcal{U}(\theta)}{\partial r_i^\alpha \partial r_k^\gamma} + v^\beta \frac{\partial \mathcal{U}(\theta)}{\partial r_k^\alpha \partial r_k^\gamma} + \delta_{\gamma\beta} \frac{\partial \mathcal{U}(\theta)}{\partial r_k^\alpha} \quad (64)$$

3.6. Torsion Potentials

We define

$$\mathbf{u} = \mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j, \quad u = |\mathbf{u}| = \sqrt{\mathbf{u} \cdot \mathbf{u}} \quad (65)$$

$$\mathbf{v} = \mathbf{r}_{kj} = \mathbf{r}_k - \mathbf{r}_j, \quad v = |\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} \quad (66)$$

$$\mathbf{w} = \mathbf{r}_{lk} = \mathbf{r}_l - \mathbf{r}_k, \quad w = |\mathbf{w}| = \sqrt{\mathbf{w} \cdot \mathbf{w}} \quad (67)$$

with atom indices as defined in **Figure 2**. The structure of a molecule is largely determined by energy barriers to rotation about chemical bonds. The torsion energy varies with the dihedral angle ϕ . The dihedral angle is given by

$$\cos \phi_{ijkl} = \frac{\mathbf{m} \cdot \mathbf{n}}{m n} \quad (68)$$

where the following substitutions were made

$$\mathbf{m} = \mathbf{v} \times \mathbf{u} \quad (69)$$

$$\mathbf{n} = \mathbf{v} \times \mathbf{w} \quad (70)$$

The sign of ϕ is taken as the sign of $\mathbf{v} \cdot (\mathbf{m} \times \mathbf{n})$.^[170]

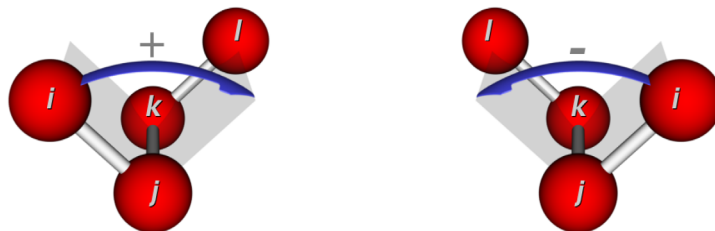


Figure 2. The dihedral angle ϕ defined by the angle between the planes formed by atoms i - j - k and j - k - l . A positive dihedral angle (left) and a negative dihedral angle (right) are shown.

Table 5. Common torsion potential functional forms.

CVFF ^[121]	$\mathcal{V}(\phi_{ijkl}) = k[1 + s \cos(n\phi_{ijkl})]$
AMBER ^[98] , GAFF ^[125]	$\mathcal{V}(\phi_{ijkl}) = \frac{1}{2}k[1 + \cos(n\phi_{ijkl} - \phi_0)]$
CHARMM ^[116]	$\mathcal{V}(\phi_{ijkl}) = k[1 + \cos(n\phi_{ijkl} + \phi_0)]$
UFF ^[15]	$\mathcal{V}(\phi_{ijkl}) = \sum_n \frac{1}{2}k[1 - \cos(n\phi_{ijkl}) \cos(n\phi_0)]$
CFF91 ^[111]	$\mathcal{V}(\phi_{ijkl}) = k_1[1 - \cos(\phi_{ijkl} - \phi_1)] + k_2[1 - \cos(2\phi_{ijkl} - \phi_2)] + k_3[1 - \cos(3\phi_{ijkl} - \phi_3)]$

A possible description with a physical interpretation is the three-term Fourier expansion^[86]

$$\mathcal{V}(\phi) = k_0 + k_1[1 + \cos(\phi)] + k_2[1 - \cos(2\phi)] + k_3[1 + \cos(3\phi)] \quad (71)$$

1. the onefold-term has been attributed to residual dipole-dipole interactions, to van der Waals interactions, or to any other direct interaction between atoms not accounted for otherwise,
2. the twofold arises from conjugation or hyper conjugation, being geometrically related to p orbitals,
3. and the threefold term has a steric (or bonding/anti-bonding) origin.

The values for fourfold or higher are small. It is unknown whether these are essential to include and perhaps the van der Waals and dipole interactions already take care of these effects. Torsions are even softer than bond angles. All possible ϕ -values can be found in structures. Therefore, the energy function must be valid over the entire range, the function must be periodic, and have stationary points at 0 and 180° (for reasons of symmetry). The periodicity is the number of minima for the potential, usually 3 for an sp^3 - sp^3 bond and 2 for a conjugate bond. The definition of a torsion includes two central and two terminal atoms. The term “torsional” means an internal rigid rotation and “dihedral” means a rotation of two vicinal bonds about a middle bond. In general, bonds and bend angles are high frequency modes, and generally not too important for thermodynamic quantities. Torsions, however, are *very* important. They determine the overall structure of a molecule, and improperly calibrated torsions can, for example, cause over stabilization of helices. Some common torsion potentials are listed in **Table 5**.

A common functional form, for example, for alkanes, is the Ryckaert–Bellemans potential^[189]

$$\mathcal{V}(\phi') = \sum_{n=0}^5 p_n \cos^n(\phi') \quad (72)$$

Note that $\phi' = \phi - \pi$ is defined according to the polymer convention ϕ' (trans) = 0. Functional forms for torsion potential can often be converted using trigonometric identities like $\cos 2x = -1 + 2 \cos^2 x$ and $\cos 3x = -3 \cos x + 4 \cos^3 x$.

It is convenient to separate the torsion potential functional form from the other derivatives so that a more general framework can be used. We therefore define the torsion potential functional derivatives as

$$f_1 = \frac{d\mathcal{V}(\phi)}{d \cos \phi} \quad (73)$$

$$f_2 = \frac{d^2\mathcal{V}(\phi)}{d \cos^2 \phi} \quad (74)$$

Using the chain rule we can split the derivative in a derivative of $\mathcal{V}(\phi_{ijkl})$ with respect to the dihedral angle ϕ and the derivative of the dihedral angle ϕ with respect to the atomic positions^[170]

$$\frac{\partial \mathcal{V}(\phi_{ijkl})}{\partial \mathbf{r}} = \frac{\partial \mathcal{V}(\phi_{ijkl})}{\partial \cos \phi_{ijkl}} \frac{\partial \cos \phi_{ijkl}}{\partial \mathbf{r}} \quad (75)$$

For first derivatives it is convenient to define vectors \mathbf{R} and \mathbf{S} ^[171]

$$\mathbf{R} = \mathbf{u} - d\mathbf{v} \quad R = |\mathbf{R}| = \sqrt{\mathbf{R} \cdot \mathbf{R}} \quad (76)$$

$$\mathbf{S} = \mathbf{w} - e\mathbf{v} \quad S = |\mathbf{S}| = \sqrt{\mathbf{S} \cdot \mathbf{S}} \quad (77)$$

with

$$d = \frac{1}{v^2} \mathbf{u} \cdot \mathbf{v} \quad (78)$$

$$e = \frac{1}{v^2} \mathbf{w} \cdot \mathbf{v} \quad (79)$$

such that only dot-products of vectors are involved.^[169,170]

$$\frac{\partial \cos \phi}{\partial r_i^\alpha} = \frac{1}{R} \left(\frac{S^\alpha}{S} - \frac{R^\alpha}{R} \cos \phi \right) \quad (80)$$

$$\frac{\partial \cos \phi}{\partial r_i^\alpha} = \frac{1}{S} \left(\frac{R^\alpha}{R} - \frac{S^\alpha}{S} \cos \phi \right) \quad (81)$$

$$\frac{\partial \cos \phi}{\partial r_j^\alpha} = (d - 1) \frac{\partial \cos \phi}{\partial r_i^\alpha} + e \frac{\partial \cos \phi}{\partial r_i^\alpha} \quad (82)$$

$$\frac{\partial \cos \phi}{\partial r_k^\alpha} = -(e + 1) \frac{\partial \cos \phi}{\partial r_i^\alpha} - d \frac{\partial \cos \phi}{\partial r_i^\alpha} \quad (83)$$

The second derivatives are given by^[87]

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_i^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} \quad (84)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_j^\alpha \partial r_j^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_j^\alpha} \frac{\partial \cos \phi}{\partial r_j^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_j^\alpha \partial r_j^\beta} \quad (85)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_k^\alpha \partial r_k^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_k^\alpha} \frac{\partial \cos \phi}{\partial r_k^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_k^\alpha \partial r_k^\beta} \quad (86)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_i^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} \quad (87)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_j^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_i^\alpha} \frac{\partial \cos \phi}{\partial r_j^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_j^\beta} \quad (88)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_j^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_i^\beta} \frac{\partial \cos \phi}{\partial r_j^\alpha} + f_1 \frac{\partial^2 \cos \phi}{\partial r_i^\beta \partial r_j^\alpha} \quad (89)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_k^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_i^\alpha} \frac{\partial \cos \phi}{\partial r_k^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_k^\beta} \quad (90)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_k^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_i^\beta} \frac{\partial \cos \phi}{\partial r_k^\alpha} + f_1 \frac{\partial^2 \cos \phi}{\partial r_i^\beta \partial r_k^\alpha} \quad (91)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_i^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} \quad (92)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_i^\beta} \frac{\partial \cos \phi}{\partial r_i^\alpha} + f_1 \frac{\partial^2 \cos \phi}{\partial r_i^\beta \partial r_i^\alpha} \quad (93)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_j^\alpha \partial r_k^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_j^\alpha} \frac{\partial \cos \phi}{\partial r_k^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_j^\alpha \partial r_k^\beta} \quad (94)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_k^\alpha \partial r_j^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_j^\beta} \frac{\partial \cos \phi}{\partial r_k^\alpha} + f_1 \frac{\partial^2 \cos \phi}{\partial r_j^\beta \partial r_k^\alpha} \quad (95)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_j^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_j^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_j^\alpha \partial r_i^\beta} \quad (96)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_j^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_j^\beta} \frac{\partial \cos \phi}{\partial r_i^\alpha} + f_1 \frac{\partial^2 \cos \phi}{\partial r_j^\beta \partial r_i^\alpha} \quad (97)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_k^\alpha \partial r_i^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_k^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} + f_1 \frac{\partial^2 \cos \phi}{\partial r_k^\alpha \partial r_i^\beta} \quad (98)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_k^\beta} = f_2 \frac{\partial \cos \phi}{\partial r_k^\beta} \frac{\partial \cos \phi}{\partial r_i^\alpha} + f_1 \frac{\partial^2 \cos \phi}{\partial r_k^\beta \partial r_i^\alpha} \quad (99)$$

where

$$\begin{aligned} \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} &= \frac{1}{R^2} \cos \phi \left(R^\alpha R^\beta + \frac{v^\alpha v^\beta}{v^2} - \delta_{\alpha\beta} \right) \\ &\quad - \frac{R^\alpha}{R} \frac{\partial \cos \phi}{\partial r_i^\beta} - \frac{R^\beta}{R} \frac{\partial \cos \phi}{\partial r_i^\alpha} \end{aligned} \quad (100)$$

$$\begin{aligned} \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} &= \frac{1}{S^2} \cos \phi \left(S^\alpha S^\beta + \frac{v^\alpha v^\beta}{v^2} - \delta_{\alpha\beta} \right) \\ &\quad - \frac{S^\alpha}{S} \frac{\partial \cos \phi}{\partial r_i^\beta} - \frac{S^\beta}{S} \frac{\partial \cos \phi}{\partial r_i^\alpha} \end{aligned} \quad (101)$$

$$\frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} = \frac{1}{R S} \left(\cos \phi R^\alpha S^\beta - R^\alpha R^\beta - S^\alpha S^\beta - \frac{1}{v^2} v^\alpha v^\beta + \delta_{\alpha\beta} \right) \quad (102)$$

$$\frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_j^\beta} = (d-1) \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} + e \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} + \frac{\partial d}{\partial r_i^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} \quad (103)$$

$$\frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_k^\beta} = -d \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} - (e+1) \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} - \frac{\partial d}{\partial r_i^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} \quad (104)$$

$$\frac{\partial^2 \cos \phi}{\partial r_j^\alpha \partial r_i^\beta} = (d-1) \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} + e \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} + \frac{\partial e}{\partial r_i^\beta} \frac{\partial \cos \phi}{\partial r_i^\alpha} \quad (105)$$

$$\frac{\partial^2 \cos \phi}{\partial r_k^\alpha \partial r_i^\beta} = -d \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} - (e+1) \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_i^\beta} - \frac{\partial e}{\partial r_i^\beta} \frac{\partial \cos \phi}{\partial r_i^\alpha} \quad (106)$$

$$\begin{aligned} \frac{\partial^2 \cos \phi}{\partial r_j^\alpha \partial r_j^\beta} &= (d-1) \frac{\partial^2 \cos \phi}{\partial r_i^\alpha \partial r_j^\beta} + e \frac{\partial^2 \cos \phi}{\partial r_j^\beta \partial r_i^\alpha} + \frac{\partial d}{\partial r_j^\beta} \frac{\partial \cos \phi}{\partial r_i^\alpha} \\ &\quad + \frac{\partial e}{\partial r_j^\beta} \frac{\partial \cos \phi}{\partial r_i^\alpha} \end{aligned} \quad (107)$$

$$\begin{aligned} \frac{\partial^2 \cos \phi}{\partial r_k^\alpha \partial r_k^\beta} &= -(e+1) \frac{\partial^2 \cos \phi}{\partial r_k^\alpha \partial r_i^\beta} - d \frac{\partial^2 \cos \phi}{\partial r_i^\beta \partial r_k^\alpha} - \frac{\partial d}{\partial r_k^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} \\ &\quad - \frac{\partial e}{\partial r_k^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} \end{aligned} \quad (108)$$

$$\begin{aligned} \frac{\partial^2 \cos \phi}{\partial r_j^\alpha \partial r_k^\beta} &= -d \frac{\partial^2 \cos \phi}{\partial r_i^\beta \partial r_j^\alpha} - (e+1) \frac{\partial^2 \cos \phi}{\partial r_j^\alpha \partial r_i^\beta} - \frac{\partial d}{\partial r_j^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} \\ &\quad - \frac{\partial e}{\partial r_j^\alpha} \frac{\partial \cos \phi}{\partial r_i^\beta} \end{aligned} \quad (109)$$

and where

$$\frac{\partial d}{\partial r_i^\alpha} = \frac{1}{v^2} v^\alpha \quad (110)$$

$$\frac{\partial d}{\partial r_j^\alpha} = \frac{1}{v} \left((2d-1) \frac{1}{v} v^\alpha - \frac{1}{v} u^\alpha \right) \quad (111)$$

$$\frac{\partial d}{\partial r_k^\alpha} = -\frac{1}{v} \left(2d \frac{1}{v} v^\alpha - \frac{1}{v} u^\alpha \right) \quad (112)$$

$$\frac{\partial d}{\partial r_i^\alpha} = 0 \quad (113)$$

$$\frac{\partial e}{\partial r_i^\alpha} = 0 \quad (114)$$

$$\frac{\partial e}{\partial r_j^\alpha} = \frac{1}{v} \left(2e \frac{1}{v} v^\alpha - \frac{1}{v} w^\alpha \right) \quad (115)$$

$$\frac{\partial e}{\partial r_k^\alpha} = -\frac{1}{v} \left((2e+1) \frac{1}{v} v^\alpha - \frac{1}{v} w^\alpha \right) \quad (116)$$

$$\frac{\partial e}{\partial r_l^\alpha} = \frac{1}{v^2} v^\alpha \quad (117)$$

The strain derivatives are given by^[188]

$$\frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\alpha\beta}} = u^\alpha \frac{\partial \mathcal{U}(\phi)}{\partial r_i^\beta} + v^\alpha \left(\frac{\partial \mathcal{U}(\phi)}{\partial r_k^\beta} + \frac{\partial \mathcal{U}(\phi)}{\partial r_l^\beta} \right) + w^\alpha \frac{\partial \mathcal{U}(\phi)}{\partial r_1^\beta} \quad (118)$$

$$\begin{aligned} \frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} &= f_2 \frac{\partial \cos \phi}{\partial \varepsilon_{\alpha\beta}} \frac{\partial \cos \phi}{\partial \varepsilon_{\mu\nu}} + f_1 \frac{\partial^2 \cos \phi}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} \\ &+ \frac{1}{2} \left[\delta_{\beta\nu} \frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\alpha\mu}} + \delta_{\alpha\nu} \frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\beta\mu}} \right. \\ &\left. + \delta_{\beta\mu} \frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\alpha\nu}} + \delta_{\alpha\mu} \frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\beta\nu}} \right] \quad (119) \end{aligned}$$

The explicit strain derivative terms appearing in the last term in Equation (119) arise from the action of the strain derivative on the explicit cell dependence of the microscopic stress.^[181] The formulae for the strain derivatives of the torsion are more conveniently written down in terms of the natural logarithm of the square of the cosine of the dihedral angle^[188]

$$\frac{\partial \cos \phi}{\partial \varepsilon_{\alpha\beta}} = \frac{1}{2} \cos \phi \frac{\partial \ln \cos^2 \phi}{\partial \varepsilon_{\alpha\beta}} \quad (120)$$

$$\frac{\partial^2 \cos \phi}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} = \frac{1}{2} \cos \phi \left(\frac{\partial^2 \ln \cos^2 \phi}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} + \frac{1}{2} \frac{\partial \ln \cos^2 \phi}{\partial \varepsilon_{\alpha\beta}} \frac{\partial \ln \cos^2 \phi}{\partial \varepsilon_{\mu\nu}} \right) \quad (121)$$

The first derivative with respect to strain is given by^[188]

$$\frac{\partial \ln \cos^2 \phi}{\partial \varepsilon_{\alpha\beta}} = \frac{2}{\mathbf{m} \cdot \mathbf{n}} \frac{\partial (\mathbf{m} \cdot \mathbf{n})}{\partial \varepsilon_{\alpha\beta}} - \frac{1}{m^2} \frac{\partial m^2}{\partial \varepsilon_{\alpha\beta}} - \frac{1}{n^2} \frac{\partial n^2}{\partial \varepsilon_{\alpha\beta}} \quad (122)$$

where

$$\begin{aligned} \frac{\partial (\mathbf{m} \cdot \mathbf{n})}{\partial \varepsilon_{\alpha\beta}} &= 2(\mathbf{w} \cdot \mathbf{u}) v^\alpha v^\beta + (\mathbf{v} \cdot \mathbf{v}) (w^\alpha u^\beta + u^\alpha w^\beta) \\ &- (\mathbf{w} \cdot \mathbf{v}) (v^\alpha u^\beta + u^\alpha v^\beta) - (\mathbf{v} \cdot \mathbf{u}) (w^\alpha v^\beta + v^\alpha w^\beta) \quad (123) \end{aligned}$$

$$\frac{\partial m^2}{\partial \varepsilon_{\alpha\beta}} = 2(v^2 u^\alpha u^\beta + u^2 v^\alpha v^\beta - (\mathbf{u} \cdot \mathbf{v}) (u^\alpha v^\beta + v^\alpha u^\beta)) \quad (124)$$

$$\frac{\partial n^2}{\partial \varepsilon_{\alpha\beta}} = 2(v^2 w^\alpha w^\beta + w^2 v^\alpha v^\beta - (\mathbf{w} \cdot \mathbf{v}) (w^\alpha v^\beta + v^\alpha w^\beta)) \quad (125)$$

The second derivative with respect to strain is given by^[188]

$$\begin{aligned} \frac{\partial \ln \cos^2 \phi}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} &= \frac{2}{(\mathbf{m} \cdot \mathbf{n})} \frac{\partial (\mathbf{m} \cdot \mathbf{n})}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} - \frac{2}{(\mathbf{m} \cdot \mathbf{n})} \frac{\partial (\mathbf{m} \cdot \mathbf{n})}{\partial \varepsilon_{\alpha\beta}} \frac{\partial (\mathbf{m} \cdot \mathbf{n})}{\partial \varepsilon_{\mu\nu}} \\ &+ \frac{1}{m^2} \frac{\partial m^2}{\partial \varepsilon_{\alpha\beta}} \frac{\partial m^2}{\partial \varepsilon_{\mu\nu}} + \frac{1}{n^2} \frac{\partial n^2}{\partial \varepsilon_{\alpha\beta}} \frac{\partial n^2}{\partial \varepsilon_{\mu\nu}} - \frac{1}{m^2} \frac{\partial^2 m^2}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} \\ &- \frac{1}{n^2} \frac{\partial^2 n^2}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} \quad (126) \end{aligned}$$

where

$$\begin{aligned} \frac{\partial (\mathbf{m} \cdot \mathbf{n})}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} &= 2(w^\alpha u^\beta + u^\alpha w^\beta) v^\mu v^\nu + 2v^\alpha v^\beta (w^\mu u^\nu + u^\mu w^\nu) \\ &- (w^\alpha v^\beta + v^\alpha w^\beta) (u^\mu v^\nu + v^\mu u^\nu) \\ &- (u^\alpha v^\beta + v^\alpha u^\beta) (w^\mu v^\nu + v^\mu w^\nu) \quad (127) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 m^2}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} &= 4(u^\alpha u^\beta v^\mu v^\nu + v^\alpha v^\beta u^\mu u^\nu) \\ &- 2(u^\alpha v^\beta + v^\alpha u^\beta) (u^\mu v^\nu + v^\mu u^\beta) \quad (128) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 n^2}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} &= 4(w^\alpha w^\beta v^\mu v^\nu + v^\alpha v^\beta w^\mu w^\nu) \\ &- 2(w^\alpha v^\beta + v^\alpha w^\beta) (w^\mu v^\nu + v^\mu w^\beta) \quad (129) \end{aligned}$$

The cross-derivative of the strain derivative with respect to position can again be expressed in terms of Hessian elements

$$\frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\alpha\beta} \partial r_i^\gamma} = u^\beta \frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_i^\gamma} + v^\beta \frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_k^\gamma} + (v^\beta + w^\beta) \frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_l^\gamma} + \delta_{\gamma\beta} \frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha} \quad (130)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\alpha\beta} \partial r_j^\gamma} = u^\beta \frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_j^\gamma} + v^\beta \frac{\partial \mathcal{U}(\phi)}{\partial r_j^\alpha \partial r_k^\gamma} + (v^\beta + w^\beta) \frac{\partial \mathcal{U}(\phi)}{\partial r_j^\alpha \partial r_l^\gamma} + \delta_{\gamma\beta} \frac{\partial \mathcal{U}(\phi)}{\partial r_j^\alpha} \quad (131)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\alpha\beta} \partial r_k^\gamma} = u^\beta \frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_k^\gamma} + v^\beta \frac{\partial \mathcal{U}(\phi)}{\partial r_k^\alpha \partial r_k^\gamma} + (v^\beta + w^\beta) \frac{\partial \mathcal{U}(\phi)}{\partial r_k^\alpha \partial r_l^\gamma} + \delta_{\gamma\beta} \frac{\partial \mathcal{U}(\phi)}{\partial r_k^\alpha} \quad (132)$$

$$\frac{\partial \mathcal{U}(\phi)}{\partial \varepsilon_{\alpha\beta} \partial r_l^\gamma} = u^\beta \frac{\partial \mathcal{U}(\phi)}{\partial r_i^\alpha \partial r_l^\gamma} + v^\beta \frac{\partial \mathcal{U}(\phi)}{\partial r_k^\alpha \partial r_l^\gamma} + (v^\beta + w^\beta) \frac{\partial \mathcal{U}(\phi)}{\partial r_l^\alpha \partial r_l^\gamma} + \delta_{\gamma\beta} \frac{\partial \mathcal{U}(\phi)}{\partial r_l^\alpha} \quad (133)$$

Using the chain rule we can split the derivative in a derivative of $\mathcal{U}(\phi_{ijkl})$ with respect to the dihedral angle ϕ and the derivative of the dihedral angle ϕ with respect to the atomic positions^[170]

$$f_i = -\frac{\partial \mathcal{U}(\phi)}{\partial \mathbf{r}_i} = -\frac{\partial \mathcal{U}(\phi)}{\partial \phi} \frac{\partial \phi}{\partial \cos \phi} \frac{\partial \cos \phi}{\partial \mathbf{r}_i} \quad (134)$$

$$f_i = -\frac{\partial \mathcal{U}(\phi)}{\partial \mathbf{r}_i} = -\frac{\partial \mathcal{U}(\phi)}{\partial \phi} \frac{\partial \phi}{\partial \mathbf{r}_i} \quad (135)$$

Table 6. Common out-of-plane potential functional forms.

CVFF ^[121] , CFF91 ^[111] , CHARMM ^[116]	$\mathcal{U}(\chi_{ijkl}) = k(\chi_{ijkl})^2$
--	---

The middle term in Equation (134) leads to a singularity term $1/\sin\phi$ when $\phi = 0$ or $\phi = \pi$. This is a problem for torsional expressions like for AMBER/GAFF and CHARMM that have phase angles ϕ_0 other than 0 or π . Therefore the equivalent chain-rule Equation 135 is more convenient.^[170]

3.7. Improper Torsions and Out-of-Plane Potentials

Common planar molecules that contain a double bond or sp^2 hybridization form planar groups with trigonal centers, for example, peptidic bonds, the carbon centers in benzene, and carbon and nitrogen centers in formamide. The mode of motion is different from bond stretching, bending, and internal rotation, and mandatory to control planarity. The associated harmonic potential is

$$\mathcal{U}(\chi) = \frac{1}{2}k(\chi)^2 \quad (136)$$

with χ the out-of-plane angle. Two common definitions in use are: i) the average angle between any bond that extends from the central atom and the plane defined by the other two bonds, and ii) the distance of the central atom from the plane defined by the other three atoms (pyramid height). The out-of-plane potential can also be used for non-planar structures, for example in chiral centers to avoid inversion. Another example of its use is coordination complexes where now the plane of the ligands need no longer be defined exactly. In square planar complexes it is necessary to define an average plane through the ligands (usually the least-square plane). Note that the definition include one central atom which is listed as the second in i - j - k - l : i , k , and l are bonded to the central atom j . The inversion angle potential is the average of the three possible inversion angle terms. The first and second derivatives for out-of-plane bending motion are special cases of linear bending motion of atoms.^[168]

Note that an alternative to the out-of-plane angle is the “improper torsion” potential

$$\mathcal{U}(\chi) = \frac{1}{2}k(1 - \cos 2\chi) \quad (137)$$

The improper torsion simply treats the four atoms in the plane as if they were bonded in the same way as in a true torsional angle. Note that the definition include one central atom which is listed as the second in i - j - k - l (i , k , and l are bonded to the central atom j). Improper torsions are often used to keep sp^2 atoms planar and sp^3 atoms in a tetrahedral geometry. The CHARMM convention is to list the central atom first, while there are no rules how to order the other three atoms. Hence, six possibilities exist for the definition of an improper torsion. The AMBER convention is that the out-of-plane atom is listed in the third position and the order of the other atoms is determined alphabetically by atom type, and by the atom number (i.e., the order in the molecule) when atom types are identical. Common out-of-plane potentials are listed in **Table 6**.

Table 7. Common intra-molecular cross potential functional forms.

	bond-bond coupling
CVFF ^[121] , CFF91 ^[111]	$\mathcal{U}(r_{ij}, r'_{ij}) = k(r_{ij} - r_0)(r'_{ij} - r'_0)$
	bond-bend coupling
CVFF ^[121] , CFF91 ^[111]	$\mathcal{U}(r_{ij}, \theta_{ijk}) = k(r_{ij} - r_0)(\theta_{ijk} - \theta_0)$
	bend-bend coupling
CVFF ^[121] , CFF91 ^[111]	$\mathcal{U}(\theta_{ijk}, \theta'_{ijk}) = k(\theta_{ijk} - \theta_0)(\theta'_{ijk} - \theta'_0)$
	bond-torsion coupling
CFF91 ^[111]	$\mathcal{U}(r_{ij}, \phi_{ikl}) = (r_{ij} - r_0)[k_1 \cos \phi + k_2 \cos(2\phi) + k_3 \cos(3\phi)]$
	$\mathcal{U}(r_{jk}, \phi_{jik}) = (r_{jk} - r_0)[k_1 \cos \phi + k_2 \cos(2\phi) + k_3 \cos(3\phi)]$
	bend-torsion coupling
CVFF ^[121]	$\mathcal{U}(\phi_{jikl}, \theta_{ijk}, \theta_{jkl}) = k \cos(\phi_{jikl})(\theta_{ijk} - \theta_0)(\theta_{jkl} - \theta'_0)$
CFF91 ^[111]	$\mathcal{U}(\theta_{ijk}, \phi_{jikl}) = (\theta - \theta_0)[k_1 \cos \phi_{jikl} + k_2 \cos(2\phi_{jikl}) + k_3 \cos(3\phi_{jikl})]$
	bend-torsion-bend coupling
CFF91 ^[111]	$\mathcal{U}(\theta_{ijk}, \phi_{jikl}, \theta'_{jkl}) = k \cos \phi(\theta_{ijk} - \theta_0)(\theta'_{jkl} - \theta'_0)$
Nicholas ^[190]	$\mathcal{U}(\theta_{ijk}, \phi_{jikl}) = S(\theta_{ijk})\{\frac{1}{2}k_1[1 + \cos \phi_{jikl}] + \frac{1}{2}k_2[1 - \cos(2\phi_{jikl})] + \frac{1}{2}k_3[1 + \cos(3\phi_{jikl})]\}S(\theta_{jkl})$
	out-of-plane-out-of-plane coupling
CVFF ^[121]	$\mathcal{U}(\chi_{ijkl}, \chi'_{ijkl}) = k\chi_{ijkl}\chi'_{ijkl}$

3.8. Cross Terms

The cross terms arise naturally from the energy Taylor expansion and are not ad hoc functions and allow to reproduce better the observed spectra of intra-molecular vibrations. Apart from anharmonic terms, class II and III force fields contain cross terms that reflect the coupling between adjacent bonds, angles and dihedrals. For example, bonds and bends interact, as the bend angle becomes smaller the bond lengths tend to increase. The inclusion of cross-terms leads to two advantages:

- they increase the accuracy of the force field (especially the vibrational frequencies),
- they increase the transferability of the diagonal terms $\mathcal{U}_b(r)$, $\mathcal{U}_\theta(\theta)$, $\mathcal{U}_\phi(\phi)$, $\mathcal{U}_\chi(\chi)$.

Some typical intra-molecular cross potentials are listed in **Table 7**.

For bond angles smaller than the equilibrium bond angle bond stretching is observed (e.g., cyclobutane), which can be described by a cross term that couples the bond-lengths and bend angle.

$$\mathcal{U}(r, \theta) = \frac{k_1}{2}(r - r_0)(\theta - \theta_0) \quad (138)$$

Similar to the stretch-bend term, the stretch can also couple to the torsion (stretch-torsion) and the bend couples to the torsion (bend-torsion cross-terms). These interactions are usually very small and several force fields neglect these potentials. An important use in zeolite models, however, is to slowly smooth the torsional energy to zero when the Si-O-Si bend approaches 180° (the torsion becomes ill-defined when three atoms are colinear).^[190,191]

3.9. Coulombic Potential and The Multi-Pole Expansion

So far we have discussed the intra-molecular bond, bend, torsion, and out-of-plane terms and these are sufficient to model small isolated molecules. The non-bonded term is the central work-horse of force fields and acts between atoms which are not linked by covalent bonds. It describes the dispersion and electrostatic interactions between molecules, as well as longer-range interactions within a molecule that are not already described by the bond, bend, torsion, and out-of-plane terms. It is clear that 1-2 and 1-3 pairs are excluded from Lennard-Jones and electrostatics (these are already taken into account by explicit bonding and bending terms), while 1-5 and more should be included. The 1-4 interaction is a somewhat gray area; some force fields neglect explicit torsions but include dispersion/electrostatics, others use only torsions and no dispersion/electrostatics, and some use both, often with scaling factors (generally 0.5; AMBER uses a scaling factor of 0.8, a scale factor of 0.75 is used in the MMFF94 force field,^[147] and a value of 0.5 is adopted in the OPLS-AA force field^[109]). The latter option (scaling) generally improves agreement and better reproduces the potential energy of rotation.

First, the electrostatic density of the molecules needs to be described. Consider two separate molecules *A* and *B* and their associated static charge densities ρ_A and ρ_B . The Coulombic interaction energy then reads^[192,193]

$$\mathcal{U}_{A,B} = \frac{1}{4\pi\epsilon_0} \iint \frac{\rho_A(\mathbf{r}')\rho_B(\mathbf{r}'')}{|\mathbf{r}' - \mathbf{r}''|} dV' dV'' \quad (139)$$

where ϵ_0 is the absolute dielectric permittivity of classical vacuum, and where the 3D integration is carried out over the volume V' of atom *A* as described by the vector \mathbf{r}' and over the volume V'' of atom *B* as described by the vector \mathbf{r}'' . If we assume that the extent of the charge distributions ρ_A and ρ_B is much smaller than their separation we may expand \mathcal{U}_{AB} in a multipole series with respect to the center of mass coordinates \mathbf{r}_A and \mathbf{r}_B . The first few multipole moments of the charge distribution ρ_A are determined as^[192,193]

$$q_A = \int \rho_A(\mathbf{r}') dV' \quad (140)$$

$$\boldsymbol{\mu}_A = \int \rho_A(\mathbf{r}')(\mathbf{r}' - \mathbf{r}_A) dV' \quad (141)$$

$$\overleftrightarrow{\mathbf{Q}}_A = \int \rho_A(\mathbf{r}') \left[3(\mathbf{r}' - \mathbf{r}_A)(\mathbf{r}' - \mathbf{r}_A) - \mathbf{I}|\mathbf{r}' - \mathbf{r}_A|^2 \right] dV' \quad (142)$$

and similar expressions for the charge distribution ρ_B . The monopole moment (q) is a scalar, the dipole moment ($\boldsymbol{\mu}$) is a vector, the quadrupole moment ($\overleftrightarrow{\mathbf{Q}}$) is a second-rank tensor, and so on. With these multipole moments we can express the interaction potential as^[194]

$$\mathcal{U}_{A,B}(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \left[\frac{q_A q_B}{r} + \frac{q_A \boldsymbol{\mu}_B \cdot \mathbf{r}}{r^3} - \frac{q_B \boldsymbol{\mu}_A \cdot \mathbf{r}}{r^3} + \frac{\boldsymbol{\mu}_A \cdot \boldsymbol{\mu}_B}{r^3} - \frac{3(\boldsymbol{\mu}_A \cdot \mathbf{r})(\boldsymbol{\mu}_B \cdot \mathbf{r})}{r^5} + \dots \right] \quad (143)$$

where $\mathbf{r} = \mathbf{r}_A - \mathbf{r}_B$. The first term in the expansion is the charge-charge interaction. It is only non-zero if both charge distributions carry a net charge. Charge-charge interactions span over long distances since the distance dependence is only $1/r$. The next two terms are charge-dipole interactions. They require that at least one particle carries a net charge. These interactions decay as $1/r^2$ and are therefore of shorter range than the charge-charge interaction. Finally, the fourth term is the dipole-dipole interaction. It is the most important interaction among neutral particles. The dipole-dipole interaction decays as $1/r^3$ and depends strongly on the dipole orientations. The next higher expansion terms are the quadrupole-charge, quadrupole-dipole, and quadrupole-quadrupole interactions, which decay approximately as $1/r^3$, $1/r^4$, and $1/r^5$, respectively.

By introducing the infinitesimal test charge δq at \mathbf{r} , the electrostatic potential $\phi(\mathbf{r})$ at \mathbf{r} is defined according to

$$\phi(\mathbf{r}) = \lim_{\delta q \rightarrow 0} \frac{\partial \mathcal{U}}{\partial \delta q} \quad (144)$$

where \mathcal{U} is the potential energy of the system including the test charge. The α -component of the electrostatic field $E_\alpha(\mathbf{r})$ at \mathbf{r} is defined as

$$E^\alpha(\mathbf{r}) \equiv -\frac{\partial \phi(\mathbf{r})}{\partial r^\alpha} \quad (145)$$

A non-uniform electrostatic field E distorts a charge distribution, thereby inducing multipole moments. The total moments may be expressed (to electric quadrupole order) as^[192,194]

$$\boldsymbol{\mu}_i = \boldsymbol{\mu}_i^{(0)} + \alpha_{ij} E_j + \frac{1}{2} a_{ijk} \nabla_k E_j + \frac{1}{6} b_{ijkl} \nabla_l \nabla_k E_j + \dots \quad (146)$$

$$\overleftrightarrow{\mathbf{Q}}_{ij} = \overleftrightarrow{\mathbf{Q}}_{ij}^{(0)} + a_{jki} E_k + \frac{1}{2} d_{ijkl} \nabla_l E_k + \dots \quad (147)$$

The tensors α_{ij} , a_{ijk} , b_{ijkl} , d_{ijkl} , ... are also properties of the undistorted distribution. They are referred to collectively as polarizabilities (α_{ij} is the dipole polarizability, a_{ijk} the quadrupole polarizability, b_{ijkl} the octopole polarizability, etc). The polarizabilities of all but spherically symmetric molecules are anisotropic, having different values along different molecular directions. This arises because the electronic polarizabilities of bonds, which are a measure of the response of the electrons to an external field, are anisotropic. If a charge distribution is unaffected by an applied field, it is said to be non-polarizable or rigid.

Molecules are often modeled using the lowest non-zero electric moment. Species such as Na^+ , Cl^- have the charge as their lowest non-zero moment. Molecules such as N_2 and CO_2 have the quadrupole as their lowest non-zero moment; the lowest non-zero moment for methane and tetrafluoromethane is the octopole. Each of these multiple moments can be represented by an appropriate amount of charges. A dipole can be represented using two charges placed an appropriate distance apart; a quadrupole can be represented using four charges and octopole by eight charges. A complete description of the charge distribution around a molecule requires all of the non-zero electric moments to be specified.

3.10. Van der Waals Potentials

At very small inter-atomic distances, the electron clouds of atoms overlap. A strong repulsive force arises that determines how close two atoms or molecules can ultimately approach each other. These repulsive forces are sometimes referred to as exchange repulsion, hard core repulsion, steric repulsion, or, for ions, the Born repulsion.^[183] They are characterized by having very short ranges and increasing very sharply as two molecules come together. They belong to the category of quantum mechanical forces, and unfortunately there is no general equation for describing their distance dependence. Instead, a number of empirical potential functions have been introduced over the years, all of which appear to be satisfactory as long as they have the property of a steeply rising repulsion at small separations. The three most common such potentials are the hard sphere potential, the inverse power-law potential, and the exponential potential.

Three distinct types of forces contribute to the total long-range interaction between polar molecules.^[183] They vary with the inverse sixth power of the distance and are collectively known as the van der Waals force.

1. For the dipole–dipole interaction, a Boltzmann averaging of the interaction energy over all orientations (which now involves averaging over two polar angles and one azimuthal angle) leads to an angle-averaged interaction free energy that varies with the inverse sixth power of the distance. The Boltzmann-averaged interaction between two permanent dipoles is usually referred to as the “orientation” or “Keesom interaction.”
2. For two molecules possessing permanent dipole moments and polarizabilities their net dipole-induced dipole energy is $1/r^6$ distance dependent. This is often referred to as the “Debye interaction” or the “induction interaction.”
3. Dispersion forces (often called “London dispersion forces”) make up the third and perhaps most important contribution to the total van der Waals force between atoms and molecules, and because they are always present (in contrast to the other types of forces that may or may not be present, depending on the properties of the molecules), they play a role in a host of important phenomena such as adhesion, surface tension, physical adsorption, wetting, the properties of gases, liquids, and thin films; the strengths of solids; the flocculation of particles in liquids; and the structures of condensed macro molecules such as proteins and polymers.^[183]

Today, the term “van der Waals” interactions refers often only to the London-dispersion forces. Dispersion forces are quantum mechanical in origin. Their origin may be understood intuitively as follows: For a nonpolar atom such as helium, the time average of its dipole moment is zero, but at any instant there exists a finite dipole moment given by the instantaneous positions of the electrons about the nuclear protons. This instantaneous dipole generates an electric field that polarizes any nearby neutral atom, inducing a dipole moment in it. The resulting interaction between the two dipoles gives rise to an instantaneous attractive force between the two atoms, and the time average of this force is finite. This can be summarized by calling the dispersion interaction an interaction between two mutually induced dipoles.^[195] The point

charge-induced dipole decays as $1/r^4$, induced dipole–dipole decays as $1/r^6$, the induced dipole–quadrupole as $1/r^8$, the induced dipole–octopole and the induced quadrupole–quadrupole as $1/r^{10}$. In general, the dispersion contribution is^[195]

$$U^{\text{disp}}(r) = -\frac{C_6^{AB}}{r^6} - \frac{C_7^{AB}}{r^7} - \frac{C_8^{AB}}{r^8} - \dots \quad (148)$$

The $1/r^6$ term dominates the dispersion and for computational simplicity the higher order terms are usually neglected. An interesting quirk of the multipole series is the $1/r^7$ term; this term is missing for atoms and molecules with an inversion center, or when one does rotational averaging, but in general it is present.^[195] This dispersion force results from instantaneous dipole/induced dipole (transitory dipole) interactions and thus is also observed acting between noble gas atoms and non polar molecules with closed shells of electrons. The strength of this force depends on the dipole moment of the atom inducing the second dipole and the polarizability of the second atom. The polarizability depends on the number of electrons in the outer shell of an atom. The same is true for the refraction of light, the so-called “dispersion” – hence the name dispersion forces. However, nothing is being dispersed by dispersion forces,^[195] and for better or worse we are stuck with the term. The dispersion coefficients can be written as^[196]

$$C_6^{AB} = \frac{3}{2} \left(\frac{E_1^A E_1^B}{E_1^A + E_1^B} \right) \alpha_1^A \alpha_2^B \quad (149)$$

$$C_8^{AB} = \frac{15}{4} \left[\frac{E_1^A E_2^B \alpha_1^A \alpha_2^B}{E_1^A + E_2^B} + \frac{E_2^A E_1^B \alpha_2^A \alpha_1^B}{E_2^A + E_1^B} \right] \quad (150)$$

$$C_{10}^{AB} = 7 \left[\frac{E_1^A E_3^B \alpha_1^A \alpha_3^B}{E_1^A + A_3^B} + \frac{E_3^A E_1^B \alpha_3^A \alpha_1^B}{E_3^A + A_3^B} + \frac{E_2^A E_2^B \alpha_2^A \alpha_2^B}{E_2^A + A_2^B} \right] \quad (151)$$

where E_A and E_B are the energies of the first electronic transition in both atoms, α_A and α_B the polarizability of atoms A and B, and where 1, 2, and 3 denote dipole, quadrupole and octopole, respectively. Nicholson et al.^[197] analyzed the magnitude, importance and origin of these terms for nonpolar molecules in silicalite and $\text{AlPO}_4\text{-5}$. The static polarizabilities for the charged lattice species were estimated from Auger data using a method given in previous work,^[198] and repulsive parameters were obtained by fitting to experimental Henry’s law constants. The oxygen atom two-body dispersion accounts for about 80% of the potential energy, while the total interaction of the Si species is of the order of 12%. There are significant contributions from the higher order (C_8 and C_{10}) terms. The induced interaction is extremely small in silicalite, being less than 0.3% of the total potential energy at the bottom of the well, even for polarizable xenon. This is due to cancellation of the O and Si contributions, which are individually quite large.

Unlike Coulomb forces, van der Waals forces are not generally pairwise additive: the force between any two molecules is affected by the presence of other molecules nearby, so one cannot simply add all the pair potentials of a molecule to obtain its net interaction energy with all the other molecules. This is because the field emanating from any one molecule reaches a second molecule both directly and by “reflection” from other molecules, since they,

too, are polarized by the field.^[183] This effect adds an additional contribution to the total van der Waals interaction energy. In general, we have the expansion

$$\mathcal{U}^{\text{disp}} = \sum_i \sum_{j>i} \mathcal{U}_2^{\text{disp}}(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j>i} \mathcal{U}_3^{\text{disp}}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots \quad (152)$$

The effect of the three-body interaction on the energy or force depends on the relative disposition of the molecules and can be positive or negative, so the net effect is usually small. In silicalite it turns out that the 3-body contribution to the total energy accounts for only 3% of the total potential energy for Ar, but is about 15% for Xe.^[196] The three body interaction can be quite significant for polar species. The overall effect of multiple- or multibody (including four-body, five-body, etc.) interactions usually results in an overall reduction in the strength of the summed pair interactions.

Dispersion interactions between two ground-state species are always attractive (for excited states it can be attractive or repulsive).^[195] Short-range interactions, also labeled exchange or overlap forces, are always unfavorable. They occur between electrons with the same spin but cannot occupy the same region in space (Pauli exclusion principle). The analytical form of the repulsion is unknown. The most common repulsion/dispersion functional form is the Lennard–Jones (LJ) potential often encountered in two flavors

$$\mathcal{U}^{\text{disp}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (153)$$

$$\mathcal{U}^{\text{disp}}(r) = \epsilon \left[\left(\frac{R_{\text{min}}}{r} \right)^{12} - 2 \left(\frac{R_{\text{min}}}{r} \right)^6 \right] \quad (154)$$

where ϵ is the well-depth and R_{min} is the distance of minimum interaction energy. Note that the potentials are related by $R_{\text{min}} = 2^{1/6}\sigma$. The LJ size-parameter σ is the distance at which the interparticle potential is zero. The potential should be viewed as an effective potential with many effects lumped in together: the van der Waals term must include not only the long-range induced-dipole–induced-dipole r^{-6} dispersion effect, but also the higher multipoles, the damping of the dispersion, the induction, and the steric repulsion when two atoms get close (sometimes called exchange repulsion), and it must also make up for errors in the highly approximate treatment of electrostatics.^[195] A common alternative to the LJ potential is the Hill (also called Buckingham) potential function,

$$\mathcal{U}^{\text{disp}}(r) = k_1 \exp(-k_2 r) - \left(\frac{\sigma}{r} \right)^6 \quad (155)$$

The Hill potential, having three adjustable parameters vs. two for Lennard–Jones might be slightly more accurate. A downside of the Hill potential is divergence to large negative energies at $r \rightarrow 0$. Therefore the potential needs to be “blocked” or changed to a polynomial repulsion (e.g., MM2) at short distances.

It is convenient to separate the VDW pair potential functional form from the other derivatives so that a more general framework can be used. We therefore define the pair potential

functional derivatives as

$$f_1 = \frac{d\mathcal{U}(r)}{dr} \quad (156)$$

$$f_2 = \frac{d^2\mathcal{U}(r)}{dr^2} \quad (157)$$

We define $r = |\mathbf{r}_{ij}| = \sqrt{\mathbf{r}_{ij} \cdot \mathbf{r}_{ij}}$. The derivatives of $\mathcal{U}(r)$ with respect to Cartesian coordinates are given by^[199]

$$\frac{\partial \mathcal{U}(r)}{\partial r_i^\alpha} = f_1 \frac{r_{ij}^\alpha}{r} \quad (158)$$

$$\frac{\partial \mathcal{U}(r)}{\partial r_j^\alpha} = -f_1 \frac{r_{ij}^\alpha}{r} \quad (159)$$

$$\frac{\partial^2 \mathcal{U}(r)}{\partial r_i^\alpha \partial r_i^\beta} = \frac{\partial^2 \mathcal{U}(r)}{\partial r_j^\alpha \partial r_j^\beta} = \left(f_2 - \frac{f_1}{r} \right) \frac{1}{r^2} r_{ij}^\alpha r_{ij}^\beta + \delta_{\alpha\beta} \frac{1}{r} f_1 \quad (160)$$

$$\frac{\partial^2 \mathcal{U}(r)}{\partial r_i^\alpha \partial r_j^\beta} = \frac{\partial^2 \mathcal{U}(r)}{\partial r_j^\alpha \partial r_i^\beta} = -\left(f_2 - \frac{f_1}{r} \right) \frac{1}{r^2} r_{ij}^\alpha r_{ij}^\beta - \delta_{\alpha\beta} \frac{1}{r} f_1 \quad (161)$$

The derivatives of $\mathcal{U}(r)$ with respect to strain are derived by Ray^[181,185,199]

$$\frac{\partial \mathcal{U}(r)}{\partial \epsilon_{\alpha\beta}} = f_1 \frac{1}{r} r_{ij}^\alpha r_{ij}^\beta \quad (162)$$

$$\frac{\partial \mathcal{U}(r)}{\partial \epsilon_{\alpha\beta} \partial r_i^\gamma} = \left(f_2 - \frac{f_1}{r} \right) \frac{1}{r^2} r_{ij}^\alpha r_{ij}^\beta r_{ij}^\gamma + \frac{1}{r} \left(\delta_{\alpha\gamma} f_1 r_{ij}^\beta + \delta_{\beta\gamma} f_1 r_{ij}^\alpha \right) \quad (163)$$

$$\frac{\partial \mathcal{U}(r)}{\partial \epsilon_{\alpha\beta} \partial r_j^\gamma} = -\left(f_2 - \frac{f_1}{r} \right) \frac{1}{r^2} r_{ij}^\alpha r_{ij}^\beta r_{ij}^\gamma - \frac{1}{r} \left(\delta_{\alpha\gamma} f_1 r_{ij}^\beta + \delta_{\beta\gamma} f_1 r_{ij}^\alpha \right) \quad (164)$$

$$\begin{aligned} \frac{\partial \mathcal{U}(r)}{\partial \epsilon_{\alpha\beta} \partial \epsilon_{\mu\nu}} &= \left(f_2 - \frac{f_1}{r} \right) \frac{1}{r^2} r_{ij}^\alpha r_{ij}^\beta r_{ij}^\mu r_{ij}^\nu \\ &+ \frac{1}{2} \left[\delta_{\beta\nu} \frac{\partial \mathcal{U}(r)}{\partial \epsilon_{\alpha\mu}} + \delta_{\alpha\nu} \frac{\partial \mathcal{U}(r)}{\partial \epsilon_{\beta\mu}} + \delta_{\beta\mu} \frac{\partial \mathcal{U}(r)}{\partial \epsilon_{\alpha\nu}} + \delta_{\alpha\mu} \frac{\partial \mathcal{U}(r)}{\partial \epsilon_{\beta\nu}} \right] \end{aligned} \quad (165)$$

The explicit strain derivative terms appearing in the last term in Equation (165) arise from the action of the strain derivative on the explicit cell dependence of the microscopic stress.^[181] Note that these expressions above are identical to the bond-stretching ones, as both are valid for pair potentials. Note that it is also common for pair potentials to define f_1^\dagger and f_2^\dagger to further simplify the equations

$$f_1^\dagger = \frac{1}{r} \frac{d\mathcal{U}(r)}{dr} = \frac{1}{r} f_1 \quad (166)$$

$$f_2^\dagger = \frac{1}{r} \frac{df_1^\dagger}{dr} = \left(f_2 - \frac{f_1}{r} \right) \frac{1}{r^2} \quad (167)$$

3.11. Rigid-Body Coordinates and Derivatives

Small molecules and phenyl-rings are often kept fixed in molecular simulations. A convenient formulation of rigid units is to use the center of mass (three degrees of freedom) and an orientation (zero, two, or three degrees of freedom, depending on whether it is a single atom, a linear molecule or a non-linear molecule, respectively). The six degrees of freedom for a non-linear rigid body can be defined in terms of the center of mass coordinates $\mathbf{r} = (r_1, r_2, r_3)$ and the three components of a vector $\mathbf{p} = (p_1, p_2, p_3) = \theta \hat{\mathbf{p}}$, which specifies both a normalised rotation axis $\hat{\mathbf{p}}$ through the center of mass, and a magnitude of rotation $\theta = \sqrt{p_1^2 + p_2^2 + p_3^2}$.^[200] For the rotation vector \mathbf{p} , the corresponding 3×3 rotation matrix \mathbf{R} can be expressed as (Rodrigues' rotation formula)

$$\mathbf{R} = \mathbf{I} + \tilde{\mathbf{p}}^2(1 - \cos \theta) + \tilde{\mathbf{p}} \sin \theta \quad (168)$$

where \mathbf{I} is a 3×3 identity matrix, and $\tilde{\mathbf{p}}$ is the skew-symmetric matrix obtained from $\hat{\mathbf{p}}$

$$\tilde{\mathbf{p}} = \frac{1}{\theta} \begin{pmatrix} 0 & -p_3 & p_2 \\ p_3 & 0 & -p_1 \\ -p_2 & p_1 & 0 \end{pmatrix} \quad (169)$$

The product of the skew-symmetric matrix $\tilde{\mathbf{p}}$ and any vector \mathbf{v} returns their cross product $\tilde{\mathbf{p}}\mathbf{v} = \mathbf{p} \times \mathbf{v}$. The matrix \mathbf{R} is then given by^[200]

$$\mathbf{R} = \begin{pmatrix} \cos \theta + \hat{p}_1^2(1 - \cos \theta) & \hat{p}_1\hat{p}_2(1 - \cos \theta) - \hat{p}_3 \sin \theta & \hat{p}_2 \sin \theta + \hat{p}_1\hat{p}_3(1 - \cos \theta) \\ \hat{p}_3 \sin \theta + \hat{p}_1\hat{p}_2(1 - \cos \theta) & \cos \theta + \hat{p}_2^2(1 - \cos \theta) & -\hat{p}_1 \sin \theta + \hat{p}_2\hat{p}_3(1 - \cos \theta) \\ -\hat{p}_2 \sin \theta + \hat{p}_1\hat{p}_3(1 - \cos \theta) & \hat{p}_1 \sin \theta + \hat{p}_2\hat{p}_3(1 - \cos \theta) & \cos \theta + \hat{p}_3^2(1 - \cos \theta) \end{pmatrix} \quad (170)$$

One can transform a quaternion $\mathbf{q} = [q_1, q_2, q_3, q_4]^T$ to the axis-angle representation

$$\hat{\mathbf{p}} = \frac{[q_1, q_2, q_3]^T}{\sqrt{q_1^2 + q_2^2 + q_3^2}} \quad (171)$$

$$\theta = 2 \arccos(q_4) \quad (172)$$

and the rotation matrix to the axis-angle representation

$$\theta = \arccos \left[\frac{1}{2} (R_{11} + R_{22} + R_{33} - 1) \right] \quad (173)$$

$$p_1 = \frac{R_{32} - R_{23}}{2 \sin \theta} \quad (174)$$

$$p_2 = \frac{R_{13} - R_{31}}{2 \sin \theta} \quad (175)$$

$$p_3 = \frac{R_{21} - R_{12}}{2 \sin \theta} \quad (176)$$

Denoting the derivative of the matrix \mathbf{A} , $\frac{\partial \mathbf{A}}{\partial p_k}$ by \mathbf{A}^k ($k = 1, 2, 3$) we have

$$\mathbf{R}^k = \frac{p_k \sin \theta}{\theta} \tilde{\mathbf{p}}^2 + (1 - \cos \theta) (\tilde{\mathbf{p}}_k \tilde{\mathbf{p}} + \tilde{\mathbf{p}} \tilde{\mathbf{p}}_k) + \frac{p_k \cos \theta}{\theta} \tilde{\mathbf{p}} + \sin \theta \tilde{\mathbf{p}}_k \quad (177)$$

Remembering $\theta = \sqrt{p_1^2 + p_2^2 + p_3^2}$ the derivatives of $\tilde{\mathbf{p}}$ are^[200]

$$\begin{aligned} \tilde{\mathbf{p}}_1 &= \frac{1}{\theta} \begin{pmatrix} 0 & \frac{p_1 p_3}{\theta^2} & -\frac{p_1 p_2}{\theta^2} \\ -\frac{p_1 p_3}{\theta^2} & 0 & -(1 - p_1^2/\theta^2) \\ \frac{p_1 p_2}{\theta^2} & 1 - p_1^2/\theta^2 & 0 \end{pmatrix} \\ \tilde{\mathbf{p}}_2 &= \frac{1}{\theta} \begin{pmatrix} 0 & \frac{p_2 p_3}{\theta^2} & 1 - p_2^2/\theta^2 \\ -\frac{p_2 p_3}{\theta^2} & 0 & \frac{p_1 p_2}{\theta^2} \\ -(1 - p_2^2/\theta^2) & -\frac{p_1 p_2}{\theta^2} & 0 \end{pmatrix} \\ \tilde{\mathbf{p}}_3 &= \frac{1}{\theta} \begin{pmatrix} 0 & -(1 - p_3^2/\theta^2) & -\frac{p_2 p_3}{\theta^2} \\ 1 - p_3^2/\theta^2 & 0 & \frac{p_1 p_3}{\theta^2} \\ \frac{p_2 p_3}{\theta^2} & -\frac{p_1 p_3}{\theta^2} & 0 \end{pmatrix} \end{aligned} \quad (178)$$

When $\theta = 0$ the singularity can be removed giving $\mathbf{R}^k = \tilde{\mathbf{p}}(k)$ where $\tilde{\mathbf{p}}(k)$ is obtained from Equation 169 using $\hat{\mathbf{p}}$ with the k th element set to unity.

The advantage of this scheme is that all the rigid-body coordinate information is stored in the space-fixed frame and the derivatives of the rotation matrix can all be programmed in general using a fixed subroutine, which can be called to deal with site-site isotropic potentials as well as single-site or site-site anisotropic potentials.

The current prescription can be extended to obtain the second derivatives by successive application of the chain rule. We then need to compute six additional 3×3 matrices of two types, namely $\frac{\partial^2 \mathbf{R}}{\partial p_k^2}$ and $\frac{\partial^2 \mathbf{R}}{\partial p_i \partial p_k}$ which are denoted by \mathbf{R}^{kk} and \mathbf{R}^{kl}

$$\begin{aligned} \mathbf{R}^{kk} &= \frac{2p_k \sin \theta}{\theta} (\tilde{\mathbf{p}}_k \tilde{\mathbf{p}} + \tilde{\mathbf{p}} \tilde{\mathbf{p}}_k) + \left(\frac{p_k^2 \cos \theta}{\theta^2} - \frac{p_k^2 \sin \theta}{\theta^3} + \frac{\sin \theta}{\theta} \right) \tilde{\mathbf{p}}^2 \\ &\quad + (1 - \cos \theta) (2\tilde{\mathbf{p}}_k^2 + \tilde{\mathbf{p}}_{kk} \tilde{\mathbf{p}} + \tilde{\mathbf{p}} \tilde{\mathbf{p}}_{kk}) \\ &\quad + \left(-\frac{p_k^2 \sin \theta}{\theta^2} - \frac{p_k^2 \cos \theta}{\theta^3} + \frac{\cos \theta}{\theta} \right) \tilde{\mathbf{p}} \\ &\quad + \frac{2p_k \cos \theta}{\theta} \tilde{\mathbf{p}}_k + \sin \theta \tilde{\mathbf{p}}_{kk} \end{aligned} \quad (179)$$

$$\begin{aligned} \mathbf{R}^{kl} &= \frac{p_k \sin \theta}{\theta} (\tilde{\mathbf{p}}_k \tilde{\mathbf{p}}_l + \tilde{\mathbf{p}}_l \tilde{\mathbf{p}}_k) + \left(\frac{p_k p_l \cos \theta}{\theta^2} - \frac{p_k p_l \sin \theta}{\theta^3} \right) \tilde{\mathbf{p}}^2 \\ &+ (1 - \cos \theta) (\tilde{\mathbf{p}}_{kl} \tilde{\mathbf{p}} + \tilde{\mathbf{p}}_k \tilde{\mathbf{p}}_l + \tilde{\mathbf{p}}_l \tilde{\mathbf{p}}_k + \tilde{\mathbf{p}}_{kl}) \\ &+ \frac{p_l \sin \theta}{\theta} (\tilde{\mathbf{p}}_k \tilde{\mathbf{p}} + \tilde{\mathbf{p}}_l \tilde{\mathbf{p}}_k) - \left(\frac{p_k p_l \sin \theta}{\theta^2} + \frac{p_k p_l \cos \theta}{\theta^3} \right) \tilde{\mathbf{p}} \\ &+ \frac{p_k \cos \theta}{\theta} \tilde{\mathbf{p}}_l + \frac{p_l \cos \theta}{\theta} \tilde{\mathbf{p}}_k + \sin \theta \tilde{\mathbf{p}}_{kl} \end{aligned} \quad (180)$$

The derivatives $\tilde{\mathbf{p}}_{kl}$ are^[200]

$$\tilde{\mathbf{p}}_{11} = \frac{1}{\theta^3} \begin{pmatrix} 0 & p_3(1-3p_1^2/\theta^2) & -p_2(1-3p_1^2/\theta^2) \\ -p_3(1-3p_1^2/\theta^2) & 0 & 3p_1(1-p_1^2/\theta^2) \\ p_2(1-3p_1^2/\theta^2) & -3p_1(1-p_1^2/\theta^2) & 0 \end{pmatrix} \quad (181)$$

$$\tilde{\mathbf{p}}_{12} = \frac{1}{\theta^3} \begin{pmatrix} 0 & -3p_1 p_2 p_3 / \theta^2 & -p_1(1-3p_2^2/\theta^2) \\ 3p_1 p_2 p_3 / \theta^2 & 0 & p_2(1-3p_1^2/\theta^2) \\ p_1(1-3p_2^2/\theta^2) & -p_2(1-3p_1^2/\theta^2) & 0 \end{pmatrix} \quad (182)$$

$$\tilde{\mathbf{p}}_{13} = \frac{1}{\theta^3} \begin{pmatrix} 0 & p_1(1-3p_3^2/\theta^2) & 3p_1 p_2 p_3 / \theta^2 \\ -p_1(1-3p_3^2/\theta^2) & 0 & p_3(1-3p_1^2/\theta^2) \\ -3p_1 p_2 p_3 / \theta^2 & -p_3(1-3p_1^2/\theta^2) & 0 \end{pmatrix} \quad (183)$$

$$\tilde{\mathbf{p}}_{21} = \frac{1}{\theta^3} \begin{pmatrix} 0 & -3p_1 p_2 p_3 / \theta^2 & -p_1(1-3p_2^2/\theta^2) \\ 3p_1 p_2 p_3 / \theta^2 & 0 & p_2(1-3p_1^2/\theta^2) \\ p_1(1-3p_2^2/\theta^2) & -p_2(1-3p_1^2/\theta^2) & 0 \end{pmatrix} \quad (184)$$

$$\tilde{\mathbf{p}}_{22} = \frac{1}{\theta^3} \begin{pmatrix} 0 & p_3(1-3p_2^2/\theta^2) & -3p_2(1-p_2^2/\theta^2) \\ -p_3(1-3p_2^2/\theta^2) & 0 & p_1(1-3p_2^2/\theta^2) \\ 3p_2(1-p_2^2/\theta^2) & -p_1(1-3p_2^2/\theta^2) & 0 \end{pmatrix} \quad (185)$$

$$\tilde{\mathbf{p}}_{23} = \frac{1}{\theta^3} \begin{pmatrix} 0 & p_2(1-3p_3^2/\theta^2) & -p_3(1-3p_2^2/\theta^2) \\ -p_2(1-3p_3^2/\theta^2) & 0 & -3p_1 p_2 p_3 / \theta^2 \\ p_3(1-3p_2^2/\theta^2) & 3p_1 p_2 p_3 / \theta^2 & 0 \end{pmatrix} \quad (186)$$

$$\tilde{\mathbf{p}}_{31} = \frac{1}{\theta^3} \begin{pmatrix} 0 & p_1(1-3p_3^2/\theta^2) & 3p_1 p_2 p_3 / \theta^2 \\ -p_1(1-3p_3^2/\theta^2) & 0 & p_3(1-3p_1^2/\theta^2) \\ -3p_1 p_2 p_3 / \theta^2 & -p_3(1-3p_1^2/\theta^2) & 0 \end{pmatrix} \quad (187)$$

$$\tilde{\mathbf{p}}_{32} = \frac{1}{\theta^3} \begin{pmatrix} 0 & p_2(1-3p_3^2/\theta^2) & -p_3(1-3p_2^2/\theta^2) \\ -p_2(1-3p_3^2/\theta^2) & 0 & -p_1(1-3p_2^2/\theta^2) \\ p_3(1-3p_2^2/\theta^2) & p_1(1-3p_3^2/\theta^2) & 0 \end{pmatrix} \quad (188)$$

$$\tilde{\mathbf{p}}_{33} = \frac{1}{\theta^3} \begin{pmatrix} 0 & 3p_3(1-p_3^2/\theta^2) & -p_2(1-3p_3^2/\theta^2) \\ -3p_3(1-p_3^2/\theta^2) & 0 & p_1(1-3p_3^2/\theta^2) \\ p_2(1-3p_3^2/\theta^2) & -p_1(1-3p_3^2/\theta^2) & 0 \end{pmatrix} \quad (189)$$

When $\theta = 0$ the singularity can be removed giving $\mathbf{R}^{kk} = \tilde{\mathbf{p}}(k)\tilde{\mathbf{p}}(k)$ and $\mathbf{R}^{kl} = \tilde{\mathbf{p}}(k)\tilde{\mathbf{p}}(l)$.

If the coordinates of two rigid bodies are denoted using the subscripts I and J , and the sites within each rigid body by subscripts i and j , then for site-site isotropic potentials the total energy is

$$\mathcal{U} = \sum_I \sum_{J < I} \sum_{i \in I} \sum_{j \in J} \mathcal{U}_{ij}^{\text{VDW}}(r_{ij}) \quad (190)$$

where $\mathcal{U}_{ij}^{\text{VDW}}$ is the pair potential between sites i and j , and

$$\mathbf{r}_{ij} = \mathbf{r}_I + \mathbf{R}_I \mathbf{r}_i^0 - \mathbf{r}_J - \mathbf{R}_J \mathbf{r}_j^0 \quad (191)$$

The first derivatives are given by^[200]

$$\frac{\partial \mathcal{U}_{ij}^{\text{VDW}}}{\partial \mathbf{r}_I^\alpha} = f_1 \frac{1}{r} r_{ij}^\alpha \quad (192)$$

$$\frac{\partial \mathcal{U}_{ij}^{\text{VDW}}}{\partial p_I^k} = f_1 \frac{1}{r} \mathbf{r}_{ij} \cdot (\mathbf{R}_I^k \mathbf{r}_i^0) \quad (193)$$

The second derivatives are^[200]

$$\frac{\partial^2 \mathcal{U}_{ij}^{\text{VDW}}}{\partial r_I^\alpha \partial r_J^\beta} = \begin{cases} (f_2 - \frac{1}{r} f_1) \frac{1}{r^2} r_{ij}^\alpha r_{ij}^\beta + f_1 \frac{1}{r} \delta_{\alpha\beta} & I = J \\ -(f_2 - \frac{1}{r} f_1) \frac{1}{r^2} r_{ij}^\alpha r_{ij}^\beta - f_1 \frac{1}{r} \delta_{\alpha\beta} & I \neq J \end{cases} \quad (194)$$

$$\frac{\partial^2 \mathcal{U}_{ij}^{\text{VDW}}}{\partial p_I^k \partial p_J^l} = \begin{cases} (f_2 - \frac{1}{r} f_1) \frac{1}{r^2} (\mathbf{r}_{ij} \cdot \mathbf{R}_I^k \mathbf{r}_i^0) (\mathbf{r}_{ij} \cdot \mathbf{R}_J^l \mathbf{r}_j^0) \\ \quad + f_1 \frac{1}{r} (\mathbf{R}_I^k \mathbf{r}_i^0) (\mathbf{R}_J^l \mathbf{r}_j^0) + f_1 \frac{1}{r} (\mathbf{r}_{ij} \cdot \mathbf{R}_I^k \mathbf{r}_i^0) \\ \quad (\mathbf{r}_{ij} \cdot \mathbf{R}_J^l \mathbf{r}_j^0) & I = J \\ -(f_2 - \frac{1}{r} f_1) \frac{1}{r^2} (\mathbf{r}_{ij} \cdot \mathbf{R}_I^k \mathbf{r}_i^0) (\mathbf{r}_{ij} \cdot \mathbf{R}_J^l \mathbf{r}_j^0) \\ \quad - f_1 \frac{1}{r} (\mathbf{R}_I^k \mathbf{r}_i^0) (\mathbf{R}_J^l \mathbf{r}_j^0) & I \neq J \end{cases} \quad (195)$$

$$\frac{\partial^2 \mathcal{U}_{ij}^{\text{VDW}}}{\partial r_I^\alpha \partial p_J^l} = \begin{cases} (f_2 - \frac{1}{r} f_1) \frac{1}{r^2} (\mathbf{r}_{ij} \cdot \mathbf{R}_I^l \mathbf{r}_i^0) r_{ij}^\alpha + f_1 \frac{1}{r} (\mathbf{R}_I^l \mathbf{r}_i^0) & I = J \\ -(f_2 - \frac{1}{r} f_1) \frac{1}{r^2} (\mathbf{r}_{ij} \cdot \mathbf{R}_J^l \mathbf{r}_j^0) r_{ij}^\alpha - f_1 \frac{1}{r} (\mathbf{R}_J^l \mathbf{r}_j^0) & I \neq J \end{cases} \quad (196)$$

The strain derivative for rigid units reads^[163]

$$\frac{\partial \mathcal{U}_{ij}^{\text{VDW}}}{\partial \epsilon_{\alpha\beta}} = \sum_{J \neq I} \sum_{i \in I} \sum_{j \in J} \left[f_1 \frac{1}{r} r_{ij}^\alpha r_{ij}^\beta - \frac{1}{2} f_1 \frac{1}{r} (r_{ij}^\alpha d_i^\beta + r_{ij}^\beta d_i^\alpha) \right] \quad (197)$$

where $\mathbf{d}_i = \mathbf{r}_{Ii} - \mathbf{r}_I$ is the distance vector between site i and the center of mass of rigid unit I . Note that the constraints result in a torque, and the instantaneous stress-tensor becomes asymmetric (remedied by explicitly symmetrizing the stress-tensor). The

presence of intra-molecular constraints results in a force term to be added to the strain derivative of an atomic system.

The derivatives of the energy with respect to strain and center of mass position and of the energy with respect to strain and orientation for pair potentials are given by^[163]

$$\begin{aligned} \frac{\partial \mathcal{U}^{\text{VDW}}}{\partial \varepsilon_{\alpha\beta} \partial r_I^\gamma} &= (f_2/r^2 - f_1/r^3) r_{ij}^\alpha r_{ij}^\beta r_{ij}^\gamma + f_1 \frac{1}{r} \left(\delta_{\alpha\gamma} r_{ij}^\beta + \delta_{\beta\gamma} r_{ij}^\alpha \right) \\ &+ \frac{1}{2} \left[d_{JI}^\beta \left((f_2/r^2 - f_1/r^3) r_{ij}^\gamma r_{ij}^\alpha + f_1 \frac{1}{r} \delta_{\alpha\gamma} \right) \right. \\ &\left. + d_{JI}^\alpha \left((f_2/r^2 - f_1/r^3) r_{ij}^\gamma r_{ij}^\beta + f_1 \frac{1}{r} \delta_{\beta\gamma} \right) \right] \end{aligned} \quad (198)$$

$$\begin{aligned} \frac{\partial \mathcal{U}^{\text{VDW}}}{\partial \varepsilon_{\alpha\beta} \partial p_I^\gamma} &= \frac{1}{2} \left[r_{ij}^\beta + d_{JI}^\beta \right] \left[(f_2/r^2 - f_1/r^3) (\mathbf{r}_{ij} \cdot \mathbf{R}_I^\gamma \mathbf{r}_i^0) r_{ij}^\alpha + f_1 \frac{1}{r} (\mathbf{R}_I^\gamma \mathbf{r}_i^0)^\alpha \right] \\ &+ \frac{1}{2} \left[r_{ij}^\alpha + d_{JI}^\alpha \right] \left[(f_2/r^2 - f_1/r^3) (\mathbf{r}_{ij} \cdot \mathbf{R}_I^\gamma \mathbf{r}_i^0) r_{ij}^\beta + f_1 \frac{1}{r} (\mathbf{R}_I^\gamma \mathbf{r}_i^0)^\beta \right] \end{aligned} \quad (199)$$

where $d_{JI}^\alpha = (d_{JI}^\alpha - d_{II}^\alpha)$.

For pair potentials we find the strain–strain derivative is given by^[163]

$$\begin{aligned} \frac{\partial \mathcal{U}^{\text{VDW}}}{\partial \varepsilon_{\alpha\beta} \partial \varepsilon_{\mu\nu}} &= \frac{1}{2} (f_2/r^2 - f_1/r^3) \left[(r_{ij}^\alpha + d_{JI}^\alpha) r_{ij}^\beta r_{ij}^\mu r_{ij}^\nu + r_{ij}^\alpha (r_{ij}^\beta + d_{JI}^\beta) r_{ij}^\mu r_{ij}^\nu \right] \\ &+ \frac{1}{2} f_1 \frac{1}{r} \left[\delta_{\beta\nu} (r_{ij}^\alpha + d_{JI}^\alpha) r_{ij}^\mu + \delta_{\alpha\nu} (r_{ij}^\beta + d_{JI}^\beta) r_{ij}^\mu \right] \\ &+ \delta_{\beta\mu} (r_{ij}^\alpha + d_{JI}^\alpha) r_{ij}^\nu + \delta_{\alpha\mu} (r_{ij}^\beta + d_{JI}^\beta) r_{ij}^\nu \\ &+ \frac{1}{4} (f_2/r^2 - f_1/r^3) \left[(r_{ij}^\alpha + d_{JI}^\alpha) r_{ij}^\beta \right. \\ &\left. + (r_{ij}^\beta + d_{JI}^\beta) r_{ij}^\alpha \right] \left(d_{JI}^\mu r_{ij}^\nu + d_{JI}^\nu r_{ij}^\mu \right) \\ &+ \frac{1}{4} f_1 \frac{1}{r} \left[\delta_{\beta\nu} (r_{ij}^\alpha + d_{JI}^\alpha) d_{JI}^\mu + \delta_{\alpha\nu} (r_{ij}^\beta + d_{JI}^\beta) d_{JI}^\mu \right. \\ &\left. + \delta_{\beta\mu} (r_{ij}^\alpha + d_{JI}^\alpha) d_{JI}^\nu + \delta_{\alpha\mu} (r_{ij}^\beta + d_{JI}^\beta) d_{JI}^\nu \right] \end{aligned} \quad (200)$$

which reduces to Equation (165) for atomic systems.

Miller et al. derived a Hamiltonian form of a minimal, nonsingular representation of rigid body rotations, the unit quaternion.^[201] The derived NVE integrator is the symplectic and time reversible integrator for molecules with an arbitrary level of rigidity. Thermo- and barostats can be combined with the Miller scheme to control the temperature and pressure.^[202–204]

3.12. QM-Corrected Potentials

Carefully calibrated Lennard–Jones models usually give quite satisfactory results. The model captures the first order effects, but because the parameters are adjustable parameters, also a large portion of the remaining physics is effectively incorporated. It performs well at moderate and high temperatures, where the potential energy is a relative small part of the total Hamiltonian.

Quantum effects are expected to become significant when the Broglie wavelength $\Lambda = h/\sqrt{2\pi m k_B T}$ is much larger than interparticle distance. For $T = 300\text{K}$ one has $\Lambda = 1\text{ \AA}$ for a hydrogen atom and $\Lambda = 0.19\text{ \AA}$ for a silicon atom. Except for the lightest ones, all atoms can be treated classically. However, with T becoming lower, quantum corrections may be needed. Using the path-integral formalism^[205] the Feynman–Hibbs (FH) potentials can be obtained that include quantum effects:

$$\begin{aligned} \mathcal{U}^{\text{FH}}(r) &= \mathcal{U}(r) + \frac{\beta \hbar^2}{24\mu} \left[\mathcal{U}'' + \frac{2}{r} \mathcal{U}' \right] \\ &+ \frac{\beta^2 \hbar^4}{1152\mu^2} \left[\frac{15}{r^3} \mathcal{U}' + \frac{4}{r} \mathcal{U}''' + \mathcal{U}'''' \right] + \dots \end{aligned} \quad (201)$$

where $m = 2\mu$ is the reduced mass of a pair of interacting particles, $\beta = 1/(k_B T)$ is the inverse temperature, \hbar is the reduced Planck constant, and where ' denotes differentiation with respect to r . The thermodynamics given by the FH-potential is modified.^[206] For example, the energy is given by

$$E = \frac{3Nk_B T}{2} + \left\langle \beta \sum_{i>j} \frac{\partial \mathcal{U}^{\text{FH}}(r)}{\partial \beta} \right\rangle + \left\langle \sum_{i<j} \mathcal{U}^{\text{FH}}(r) \right\rangle \quad (202)$$

where the second term on the right-hand side can be interpreted as a quantum correction to the classical kinetic energy. Liu et al.^[207] found that quantum effects are important for hydrogen adsorption in MOFs at cryogenic temperatures. For $\text{H}_2\text{-H}_2$ $\mu = m/2$ and $\beta \hbar^2/(24\mu) = 2.005/T$, while for H_2 -sorbent $\mu = m$ and $\beta \hbar^2/(24\mu) = 1.002/T$. The Feynman–Hibbs potential, truncated at second order in \hbar , was able to satisfactorily account for quantum effects.

3.13. Hydrogen Bonding

Ion–dipole interactions can be strong enough to lead to mutual alignment. Dipole–dipole interactions are usually too weak for that with the exception of the “hydrogen bond” (which is essentially just a particularly strong directional dipole–dipole interaction). Water is the prime example of hydrogen bonds but they are also found between molecules with O–H, N–H, F–H, and C–H bonds. The bond moment of $\text{O}^- \text{-H}^+$ is unusually large. The electronegative oxygen atom can get quite close to the highly polar $\text{O}^- \text{-H}^+$ groups and experience a very strong field. Hydrogen bond strength lies between 10 and 40 kJ mol^{-1} which makes them stronger than a typical van der Waals bond ($\approx 1\text{ kJ mol}^{-1}$) but still much weaker than covalent or ionic bonds ($\approx 500\text{ kJ mol}^{-1}$).^[183] Even though the hydrogen bond is believed to be purely electrostatics there is no simple equation for the interaction potential known. The electrostatics consist of charge-transfer, polarization, dispersion, and electron-exchange. One does find that the strengths of hydrogen bonds tend to follow a $1/r^2$ distance dependence. Hydrogen bonds can occur intermolecularly as well as intramolecularly and can exist in a nonpolar environment. While the van der Waals and electrostatic *point charge* interactions are more or less isotropic interactions, hydrogen bonds are directional short-range interactions. The energetically optimum arrangement occurs when the two dipole moments are colinear. Ab

Table 8. Common hydrogen bonding potential functional forms.

AMBER	$U(r_{ij}) = \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}$
CHARMM ^[116]	$U(r_{ij}) = \left[\frac{A}{r_{AD}^6} - \frac{B}{r_{A}^4} \right] \cos^m(\phi_{A-H-D}) \cos^n(\phi_{AD-A-H})$

initio quantum mechanical simulations have shown that this interaction can be accurately modeled by combining Coulomb's law with the Lennard–Jones potential if appropriate partial charges are used.^[208] Therefore several force fields do not model the hydrogen bond explicitly but account for this type of interaction by van der Waals and electrostatic interactions (MM2). Some force fields employ modified Lennard–Jones potentials (e.g., 12-10, 6-4, and 8-6 potential functions). Other force fields account for the directionality of the hydrogen bond (MM3) or even model lone-pairs explicitly. Some common functional forms for hydrogen bonding are listed in **Table 8**.

4. Coulomb Potential in Periodic Systems: Ewald Summation

The main challenge in calculating long-range potentials in periodic systems is their slow convergence with distance. The de facto method to deal with this problem is the Ewald summation. Suppose there are N point charges q_1, q_2, \dots, q_N at positions $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$ within the unit cell satisfying $q_1 + q_2 + \dots + q_N = 0$. The vectors \mathbf{a}_α , which need not be orthogonal, form the edges of the unit cell. The conjugate reciprocal vectors \mathbf{a}_α^* are defined by the relations

$$\mathbf{a}_\alpha^* \cdot \mathbf{a}_\beta = \delta_{\alpha\beta} \quad \alpha, \beta = \{x, y, z\} \quad (203)$$

Let \mathbf{a} be the 3×3 matrix having the lattice vectors \mathbf{a}_α as columns. Note that the volume V of the unit cell is given by the determinant of \mathbf{a} . Furthermore \mathbf{a}^{-1} is the 3×3 matrix having the reciprocal lattice vectors \mathbf{a}_α^* as rows. The point charge q_i at position \mathbf{r}_i has fractional coordinates s_{ai} defined by

$$s_{ai} = \mathbf{a}_\alpha^* \cdot \mathbf{r}_i \quad (204)$$

The charges interact according to Coulomb's law with periodic boundary conditions. Thus a point charge q_i at position \mathbf{r}_i interacts with all other charges q_j (with $j \neq i$) at positions \mathbf{r}_j as well as with all of their periodic images at positions $\mathbf{r}_j + n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ for all integers n_1, n_2, n_3 . It also interacts with its own periodic images at $\mathbf{r}_i + n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ for all integers n_1, n_2, n_3 not all zero. The electrostatic energy of the unit cell U can be written

$$U(\mathbf{r}_1, \dots, \mathbf{r}_N) = \frac{1}{2} \sum_{n'} \sum_{ij} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|} \quad (205)$$

where the outer sum is over the vectors $\mathbf{n} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, the prime indicating that terms with $i = j$ for $\mathbf{n} = \mathbf{0}$ are omitted. We define the reciprocal lattice vectors \mathbf{k} by

$$\mathbf{k} = 2\pi(k_1\mathbf{a}_1^* + k_2\mathbf{a}_2^* + k_3\mathbf{a}_3^*) \quad (206)$$

with k_1, k_2, k_3 integers not all zero and the structure factor $S(\mathbf{k})$ by

$$S(\mathbf{k}) = \sum_{j=1}^N q_j \exp(i\mathbf{k} \cdot \mathbf{r}_j) \quad (207)$$

The structure factor $S(\mathbf{k})$ can be viewed as a discrete Fourier transform of a set of charges placed irregularly within the unit cell.

The technique of Ewald summation is applicable to true periodic systems, that is, the simulation region or unit cell is effectively replicated in all spatial directions. For short-ranged potentials like the Lennard–Jones, the cutoff is usually chosen as smaller than half the shortest width of the unit cell, such that only summations over interactions in the main cell are required. For distances larger than half the shortest width of the unit cell, neighboring simulation volumes need to be taken into account. In the minimum image convention, the potential seen by a particle is summed over all other particles or their periodic images, whichever is closest. For long-range potentials this construction is inadequate because the contributions from more distant images at $2L, 3L$ etc. are no longer negligible. One might think that these contributions should approximately cancel and they nearly do. One has to take care in which order to do the sum as illustrated by an example. Consider a system of two oppositely charged ions, periodically extended to form an infinite 1D line of charges, each separated by a distance R . The potential energy of the reference ion with charge $-q$ is:^[38]

$$\begin{aligned} U &= -2q^2 \left(\frac{1}{R} - \frac{1}{2R} + \frac{1}{3R} - \frac{1}{4R} \dots \right) \\ &= -\frac{2q^2}{R} \left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \dots \right) \\ &= -\frac{2q^2}{R} \log 2 \end{aligned} \quad (208)$$

The factor $2 \log 2$ is the Madelung constant, which is of central importance in the theory of ionic crystals. The series is actually *conditionally convergent*, i.e. the results depends on the summation order. We can choose a different ordering, for example^[209]

$$1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \dots \quad (209)$$

that is, two positive terms followed by a negative term. Then it can be shown it now converges to $\frac{3}{2} \log 2$, which is 50% higher potential energy than before.^[209] The problem is in what order should we sum over periodic images. An intuitive and elegant way of doing this is to build up sets of images contained within successively larger spheres surrounding the simulation region. According to this scheme the energy is expressed as in Equation (205). However, using the Ewald transformation, the single *divergent* summation can be replaced with two *convergent* summations.

4.1. Ewald Transformation

For an infinite system, a meaningful energy can be defined by use of the Ewald transformation. The basic idea of the Ewald approach is as follows^[210,211]

$$\frac{1}{r} = \frac{\text{erf}(\eta r)}{r} + \frac{\text{erfc}(\eta r)}{r} \quad (210)$$

where the error function $\text{erf}(x)$ and its complement are defined as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (211)$$

$$\text{erfc}(x) = 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \quad (212)$$

In Equation (210), the first term goes to a constant ($\frac{2\eta}{\sqrt{\pi}}$) as $r \rightarrow 0$, but has a long tail as $r \rightarrow \infty$. The second term has a singular behavior as $r \rightarrow 0$, but vanishes exponentially as $r \rightarrow \infty$. Ewald's idea is to replace a single divergent summation with two convergent summations. The first summation has a convergent summation in the form of its Fourier transform and the second has a convergent direct summation. Thus the calculation of the electrostatic energy would be evaluated using

$$\begin{aligned} \mathcal{U} &= \frac{1}{8\pi\epsilon_0} \sum_{i \neq j} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (213) \\ &= \frac{1}{8\pi\epsilon_0} \left(\sum_{i \neq j} \frac{q_i q_j \text{erf}(\eta |\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i \neq j} \frac{q_i q_j \text{erfc}(\eta |\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|} \right) \end{aligned}$$

For an appropriate choice of the parameter η , the second summation converges quickly and can be evaluated directly. The first term in the summation must be transformed into Fourier space.

In order to describe these summations explicitly, we assume that we have a periodic lattice so that every ion can be located by $\mathbf{r}_i = \mathbf{r}'_i + \mathbf{n}$, that is a location \mathbf{r}'_i within a unit cell and a periodic translation vector \mathbf{n} . In this way, the summation becomes

$$\sum_{ij} = N \sum_{i,j,\mathbf{n}} \quad (214)$$

where N denotes the number of unit cells in the system. Since we have a periodic system, N is infinite, but the energy per unit cell \mathcal{U}/N is well defined. A sum over lattice translation \mathbf{n} may be transformed into an equivalent sum over reciprocal lattice translations \mathbf{k} according to the identity:

$$\sum_{\mathbf{n}} \delta^3(\mathbf{r} - \mathbf{n}) = \frac{1}{V} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (215)$$

where V denotes the unit cell volume. The first term thus becomes

$$\begin{aligned} &\sum_{i \neq j} \frac{q_i q_j \text{erf}(\eta |\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|} \\ &= \sum_{ij} \frac{q_i q_j \text{erf}(\eta |\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i=j} \frac{q_i q_j \text{erf}(\eta |\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (216) \end{aligned}$$

$$= N \left(\sum_{ij} q_i q_j \sum_{\mathbf{n}} \frac{\text{erf}(\eta |\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|)}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|} - \frac{2\eta}{\sqrt{\pi}} \sum_i q_i^2 \right) \quad (217)$$

where the last term comes from subtracting out the self-interaction ($i = j$) term from the complete lattice sum. Using $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, the lattice sum reads

$$\begin{aligned} &\frac{1}{V} \sum_{\mathbf{k}} \int e^{i\mathbf{k}\cdot\mathbf{r}} \frac{\text{erf}(\eta |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} d^3 r \\ &= \frac{4\pi}{V} \left(\sum_{\mathbf{k} \neq 0} \frac{e^{-i\mathbf{k}\cdot\mathbf{r}_{ij}} e^{-\frac{k^2}{4\eta^2}}}{k^2} + \frac{1}{2} \int_0^\eta \frac{1}{u^3} du \right) \quad (218) \end{aligned}$$

The last term in Equation (218) comes from the $\mathbf{k} = 0$ contribution. The omission of this term is called a "tin-foil" boundary condition and corresponds to a system embedded in a medium with infinite dielectric constant (a perfectly conducting medium).^[212] If the surrounding medium is a vacuum (dielectric constant of unity) then an additional surface energy term appears in the equations. However, tin-foil boundary conditions are standard and for ionic systems the use of this type of boundary condition is essential.^[84] The Ewald expression is found to be

$$\begin{aligned} \frac{\mathcal{U}}{V} &= \sum_{ij} \frac{q_i q_j}{8\pi\epsilon_0} \left(\sum_{\mathbf{n}'} \frac{\text{erfc}(\eta |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|} + \frac{4\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{e^{-i\mathbf{k}\cdot\mathbf{r}_{ij}} e^{-\frac{k^2}{4\eta^2}}}{k^2} - \frac{2\eta}{\sqrt{\pi}} \delta_{ij} \right) \quad (219) \\ &= \underbrace{\sum_{ij} \frac{q_i q_j}{8\pi\epsilon_0} \sum_{\mathbf{n}'} \frac{\text{erfc}(\eta |\mathbf{r}_{ij} + \mathbf{n}|)}{|\mathbf{r}_{ij} + \mathbf{n}|}}_{\text{Real part}} + \underbrace{\frac{1}{2V\epsilon_0} \sum_{\mathbf{k} \neq 0} \frac{e^{-\frac{k^2}{4\eta^2}}}{k^2} |S(\mathbf{k})|^2}_{\text{Reciprocal part}} \\ &\quad - \underbrace{\frac{1}{4\pi\epsilon_0} \sum_i \frac{\eta}{\sqrt{\pi}} q_i^2}_{\text{Self part}} \quad (220) \end{aligned}$$

where the ' in the summation over the lattice translations \mathbf{n} indicates that all self-interaction terms should be omitted, and the triple sum over \mathbf{k} , i , and j in the reciprocal term in Equation (219) has been replaced by a double sum over \mathbf{k} and i (in the structure factor $S(\mathbf{k})$) in Equation (220).

4.2. Units

The total electrostatic potential energy of interaction between point charges q_i at the positions \mathbf{r}_i is given by

$$\mathcal{U} = \frac{1}{4\pi\epsilon_0} \sum_{i < j} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (221)$$

A note on the prefactor of the Coulomb potential and conversion to internal simulation units. The dielectric constant of vacuum is

$\epsilon_0 = 8.8541878176 \times 10^{-12}$ F/m (where Farad $F = C^2/J$).

$$\frac{1}{4\pi\epsilon_0} = \frac{(\text{charge unit})^2}{(4\pi) \times (8.8541878176 \times 10^{-12}) \times (\text{length unit}) \times (\text{energy conversion})} \quad (222)$$

where the conversion of energy in internal units is (mass unit) \times (length unit)²/(time unit). For example, using Angstrom (10^{-10} m) as the length scale, picoseconds (10^{-12} s) as the time scale, Kelvin as the temperature, atomic mass ($1.6605402 \times 10^{-27}$ kg) as the unit of mass, and atomic charge ($1.60217733 \times 10^{-19}$ C per particle) as the unit of charge, then the units of energy will be 10 J mol^{-1} ($1.6605402 \times 10^{-23}$ J) and the Coulombic conversion will be 138935.48. This factor is needed to convert the electrostatic energy to the internal units at every evaluation. The factor corresponds to 167101.1 K, 1389.36 kJ mol⁻¹, and 332.067 kcal mol⁻¹. In the remainder, the prefactor will be omitted for notational convenience.

4.3. General Formalism Finite System

Nymand and Linse have provided a general formalism for atomic charges, dipoles, and polarizabilities and this section is taken from their article.^[213] Consider a finite system of interacting molecules possessing atomic charges, dipoles, and anisotropic polarizabilities. Throughout, the Einstein summation convention is used for Greek letter indices. The following interactions tensors are employed

$$\mathcal{T}_{ij} = \frac{1}{r_{ij}} \quad (223)$$

$$\mathcal{T}_{ij}^\alpha = \nabla_\alpha \mathcal{T}_{ij} = -\frac{r_{ij}^\alpha}{r_{ij}^3} \quad (224)$$

$$\mathcal{T}_{ij}^{\alpha\beta} = \nabla_\alpha \nabla_\beta \mathcal{T}_{ij} = \frac{3r_{ij}^\alpha r_{ij}^\beta}{r_{ij}^5} - \frac{\delta_{\alpha\beta}}{r_{ij}^3} \quad (225)$$

$$\mathcal{T}_{ij}^{\alpha\beta\gamma} = \nabla_\alpha \nabla_\beta \nabla_\gamma \mathcal{T}_{ij} = -\frac{15r_{ij}^\alpha r_{ij}^\beta r_{ij}^\gamma}{r_{ij}^7} + \frac{3(\delta_{\beta\gamma} r_{ij}^\alpha + \delta_{\alpha\gamma} r_{ij}^\beta + \delta_{\alpha\beta} r_{ij}^\gamma)}{r_{ij}^5} \quad (226)$$

Here, $r_{ij}^\alpha = r_i^\alpha - r_j^\alpha$, where r_i^α is the α -component of the position vector of atom i . After a generalization of the approach by Veseley,^[214] the total potential energy of a finite system containing a set of atoms possessing charges, dipoles, and anisotropic polarizabilities can be expressed as^[213]

$$\mathcal{U} = \mathcal{U}_{\text{int}} + \mathcal{U}_{\text{pol}} \quad (227)$$

$$\mathcal{U}_{\text{int}} = \underbrace{\frac{1}{2} \sum_{i,j \neq i} q_i \mathcal{T}_{ij} q_j}_{\text{charge-charge}} - \underbrace{\frac{1}{2} \sum_{i,j \neq i} q_i \mathcal{T}_{ij}^\alpha \mu_j^\alpha}_{\text{charge-total dipole}} + \underbrace{\frac{1}{2} \sum_{i,j \neq i} \mu_i^\alpha \mathcal{T}_{ij}^\alpha q_j}_{\text{total dipole-charge}} - \underbrace{\frac{1}{2} \sum_{i,j \neq i} \mu_i^\alpha \mathcal{T}_{ij}^{\alpha\beta} \mu_j^\beta}_{\text{total dipole-total dipole}} \quad (228)$$

$$\mathcal{U}_{\text{pol}} = \frac{1}{2} \sum_i \mu_{i,\alpha}^{\text{ind}} \alpha_{i,\alpha\beta}^{-1} \mu_{i,\beta}^{\text{ind}} \quad (229)$$

where q_i is the charge on atom i , μ_i^α the α -component of the total dipole moment of atom i , $\mu_{i,\alpha}^{\text{ind}}$ the α -component of the induced dipole moment of atom i , and $\alpha_{i,\alpha\beta}$ the $\alpha\beta$ -component of the polarizability tensor of atom i . In Equation (228) the four terms in \mathcal{U}_{int} represent the charge-charge, the charge-total dipole, the total dipole-charge, and the total dipole-total dipole interaction, respectively, whereas \mathcal{U}_{pol} represents the work of forming the induced dipoles. The total dipole moment of atom i is given by

$$\mu_i^\alpha = \mu_{i,\alpha}^{\text{stat}} + \mu_{i,\alpha}^{\text{ind}}, \quad (230)$$

where $\mu_{i,\alpha}^{\text{stat}}$ denotes the α -component of the permanent (static) dipole moment of atom i .

By introducing a infinitesimal test charge dq at \mathbf{r} , the electrostatic potential $\phi(\mathbf{r})$ at \mathbf{r} is defined according to

$$\phi(\mathbf{r}) \equiv \lim_{dq \rightarrow 0} \frac{\partial \mathcal{U}}{\partial dq} \quad (231)$$

where \mathcal{U} is the potential energy of the system including the test charge. The α -component of the electrostatic field $E_\alpha(\mathbf{r})$ is defined as

$$E^\alpha(\mathbf{r}) \equiv -\frac{\partial \phi(\mathbf{r})}{\partial r^\alpha}, \quad (232)$$

and the $\alpha\beta$ -component of the electric field gradient $E^{\alpha\beta}(\mathbf{r})$ at \mathbf{r} as

$$E^{\alpha\beta}(\mathbf{r}) \equiv \frac{\partial E^\alpha(\mathbf{r})}{\partial r^\beta}. \quad (233)$$

The inclusion of the test charge in the summation of Equation (227) and the application of Equations (231)–(233) on Equation (227) gives the electrostatic potential, field, and field gradients. The corresponding quantities on atom i become (after omission of any self-terms)^[213]:

$$\phi_i \equiv \phi(\mathbf{r}_i) = \sum_{j \neq i} (\mathcal{T}_{ij} q_j - \mathcal{T}_{ij}^\alpha \mu_{j,\alpha}), \quad (234)$$

$$E_i^\alpha \equiv E^\alpha(\mathbf{r}_i) = \sum_{j \neq i} (-\mathcal{T}_{ij}^\alpha q_j + \mathcal{T}_{ij}^{\alpha\beta} \mu_{j,\beta}), \quad (235)$$

$$E_i^{\alpha\beta} \equiv E^{\alpha\beta}(\mathbf{r}_i) = \sum_{j \neq i} (-\mathcal{T}_{ij}^{\alpha\beta} q_j + \mathcal{T}_{ij}^{\alpha\beta\gamma} \mu_{j,\gamma}). \quad (236)$$

The induced dipole moments are determined by a minimization of \mathcal{U} with respect to $\mu_{i,\alpha}^{\text{ind}}$

$$\mu_{i,\alpha}^{\text{ind}} = \alpha_{i,\alpha\beta} E_i^\beta, \quad (237)$$

where E_i^α is the α -component of the electrostatic field on atom i arising from other charges and total dipole moments as given by Equation (235).

The potential energy is often separated into two physically appealing contributions: an electrostatic and an induction term.

With the use of Equations (230) and (237), Equation (227) can be recast according to

$$\mathcal{U} = \mathcal{U}_{\text{elec}} + \mathcal{U}_{\text{ind}} \quad (238)$$

where the electrostatic energy is given by

$$\mathcal{U}_{\text{elec}} = \frac{1}{2} \sum_i \left(q_i \phi_i^{\text{stat}} - \mu_{i,\alpha}^{\text{stat}} E_{i,\alpha}^{\text{stat}} \right), \quad (239)$$

and the induction energy by

$$\mathcal{U}_{\text{ind}} = -\frac{1}{2} \sum_i \mu_{i,\alpha}^{\text{ind}} E_{i,\alpha}^{\text{stat}}, \quad (240)$$

with ϕ_i^{stat} denoting the electrostatic potential and $E_{i,\alpha}^{\text{stat}}$ the α -component of the electrostatic field at atom i arising from other charges and static dipoles. The quantities ϕ_i^{stat} and $E_{i,\alpha}^{\text{stat}}$ are obtained from Equation (234) and Equation (235) by replacing $\mu_{j,\alpha}^{\alpha}$ with $\mu_{j,\alpha}^{\text{stat}}$ according to

$$\phi_i^{\text{stat}} = \sum_{j \neq i} \left(\mathcal{T}_{ij} q_j - \mathcal{T}_{ij}^{\alpha} \mu_{j,\alpha}^{\text{stat}} \right), \quad (241)$$

and

$$E_{i,\alpha}^{\text{stat}} = \sum_{j \neq i} \left(-\mathcal{T}_{ij}^{\alpha} q_j + \mathcal{T}_{ij}^{\alpha\beta} \mu_{j,\beta}^{\text{stat}} \right). \quad (242)$$

By taking the derivative of the total potential energy \mathcal{U} given by Equation (227) with respect to the position of atom i and noting that \mathcal{U} is stationary with respect to variation of $\mu_{i,\alpha}^{\text{ind}}$, the α -component of the force on atom i arising from its interaction with atom j can be obtained according to

$$f_{ij}^{\alpha} = -q_i \mathcal{T}_{ij}^{\alpha} q_j + q_i \mathcal{T}_{ij}^{\alpha\beta} \mu_j^{\beta} - \mu_i^{\beta} \mathcal{T}_{ij}^{\beta\alpha} q_j + \mu_i^{\beta} \mathcal{T}_{ij}^{\beta\alpha\gamma} \mu_j^{\gamma}. \quad (243)$$

The α -component of the total force acting on atom i can conveniently be expressed as^[213]

$$f_i^{\alpha} = \sum_{j \neq i} f_{ij}^{\alpha} = q_i E_i^{\alpha} + \mu_i^{\beta} E_i^{\beta\alpha}, \quad (244)$$

where E_i^{α} is the α -component of the electrostatic field and $E_i^{\alpha\beta}$ the $\alpha\beta$ -component of the field gradient on atom i arising from other charges and total dipole moments as given by Equation (235) and Equation (236), respectively.

Finally, the virial Φ , which enters the expression for the pressure, $p = p_{\text{ideal}} - \Phi/3V$, is given by

$$\Phi = 3V \left\langle \frac{\partial \mathcal{U}}{\partial V} \right\rangle, \quad (245)$$

where V denotes the volume of the system. The virial for a molecular system composed of charges, static, and induced dipole moments formally is given by

$$\Phi = -\frac{1}{2} \sum_{i,j \neq i} \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} + \sum_i \mathbf{d}_i \cdot \mathbf{f}_i, \quad (246)$$

with \mathbf{f}_{ij} and \mathbf{f}_i given by Equation (243) and Equation (244), respectively.

4.4. General Formalism Periodic System

Nyman and Linse extended the formalism to periodic systems using the Ewald summation.^[213] Also this section is directly taken from their work. When applying truly periodic boundary conditions, the summations over $j \neq i$ are replaced by infinite lattice sums. In the Ewald formalism, the expression of a given property is conventionally divided into four different terms viz. i) a real term arising from the short-range interaction in the real space, ii) a reciprocal term arising from the long-range interaction in the reciprocal space iii) a self-term correcting an over-counting in the reciprocal space, and iv) a surface term appearing in systems with nonconducting surroundings. The modified interactions tensors are denoted as $\hat{\mathcal{T}}$. In practice, this surmounts to replacing $r^{-(2n+1)}$, $n \in \{0, 1, 2, 3\}$, in the interaction energy expressions by its screened counterpart $\widehat{r^{-(2n+1)}}$, where

$$\frac{\widehat{1}}{r} = \frac{\text{erfc}(\eta r)}{r} \quad (247)$$

and

$$\frac{\widehat{1}}{r^{2n+1}} = \frac{1}{r^2} \left[\frac{\widehat{1}}{r^{2n-1}} + \frac{(2\eta)^n}{\sqrt{\pi}\eta(2n-1)} e^{-\eta^2 r^2} \right], \quad (248)$$

for $n > 0$. Note that this can straightforwardly be extended to higher order moments through the use of the recursion formula Equation (248). We define $\mathcal{T}_{ij} = \mathcal{T}_{ij}^{\alpha} = T_{ij}^{\alpha\beta} = \mathcal{T}_{ij}^{\alpha\beta\gamma} \equiv 0$ when atoms i and j belong to the same molecule. The modified tensors read

$$\hat{\mathcal{T}}_{ij} = \frac{\text{erfc}(\eta r_{ij})}{r_{ij}} \quad (249)$$

$$\hat{\mathcal{T}}_{ij}^{\alpha} = \nabla_{\alpha} \hat{\mathcal{T}}_{ij} = \left(-\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} - \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) r_{ij}^{\alpha} \quad (250)$$

$$\begin{aligned} \hat{\mathcal{T}}_{ij}^{\alpha\beta} = \nabla_{\alpha} \nabla_{\beta} \hat{\mathcal{T}}_{ij} = & \left(\frac{6\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^4} + \frac{4\eta^3 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{3\text{erfc}(\eta r_{ij})}{r_{ij}^5} \right) r_{ij}^{\alpha} r_{ij}^{\beta} \\ & + \delta_{\alpha\beta} \left(-\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} - \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) \end{aligned} \quad (251)$$

$$\begin{aligned} \hat{\mathcal{T}}_{ij}^{\alpha\beta\gamma} = \nabla_{\alpha} \nabla_{\beta} \nabla_{\gamma} \hat{\mathcal{T}}_{ij} = & \left(-\frac{30\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^6} - \frac{20\eta^3 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^4} - \frac{8\eta^5 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} - \frac{15\text{erfc}(\eta r_{ij})}{r_{ij}^7} \right) r_{ij}^{\alpha} r_{ij}^{\beta} r_{ij}^{\gamma} \\ & + \left(\frac{6\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^4} + \frac{4\eta^3 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{3\text{erfc}(\eta r_{ij})}{r_{ij}^5} \right) \\ & \times \left[\delta_{\alpha\beta} r_{ij}^{\gamma} + \delta_{\alpha\gamma} r_{ij}^{\beta} + \delta_{\beta\gamma} r_{ij}^{\alpha} \right] \end{aligned} \quad (252)$$

Similar tensors denoted $\tilde{\mathcal{T}}_{ij} = \text{erf}(\eta d_{ij})/d_{ij}$ are defined

$$\tilde{\mathcal{T}}_{ij} = \frac{\text{erf}(\eta d_{ij})}{d_{ij}} \quad (253)$$

$$\tilde{\mathcal{T}}_{ij}^\alpha = \nabla_\alpha \tilde{\mathcal{T}}_{ij} = \left(\frac{2\eta e^{-\eta^2 d_{ij}^2}}{\sqrt{\pi} d_{ij}^2} - \frac{\text{erf}(\eta d_{ij})}{d_{ij}^3} \right) d_{ij}^\alpha \quad (254)$$

$$\begin{aligned} \tilde{\mathcal{T}}_{ij}^{\alpha\beta} = \nabla_\alpha \nabla_\beta \tilde{\mathcal{T}}_{ij} = & \left(-\frac{6\eta e^{-\eta^2 d_{ij}^2}}{\sqrt{\pi} d_{ij}^4} - \frac{4\eta^3 e^{-\eta^2 d_{ij}^2}}{\sqrt{\pi} d_{ij}^2} + \frac{3\text{erf}(\eta d_{ij})}{d_{ij}^5} \right) d_{ij}^\alpha d_{ij}^\beta \\ & + \delta_{\alpha\beta} \left(\frac{2\eta e^{-\eta^2 d_{ij}^2}}{\sqrt{\pi} d_{ij}^2} - \frac{\text{erf}(\eta d_{ij})}{d_{ij}^3} \right) \end{aligned} \quad (255)$$

$$\begin{aligned} \tilde{\mathcal{T}}_{ij}^{\alpha\beta\gamma} = \nabla_\alpha \nabla_\beta \nabla_\gamma \tilde{\mathcal{T}}_{ij} \\ = & \left(\frac{30\eta e^{-\eta^2 d_{ij}^2}}{\sqrt{\pi} d_{ij}^6} + \frac{20\eta^3 e^{-\eta^2 d_{ij}^2}}{\sqrt{\pi} d_{ij}^4} + \frac{8\eta^5 e^{-\eta^2 d_{ij}^2}}{\sqrt{\pi} d_{ij}^2} - \frac{15\text{erf}(\eta d_{ij})}{d_{ij}^7} \right) d_{ij}^\alpha d_{ij}^\beta d_{ij}^\gamma \\ & \times \left(-\frac{6\eta e^{-\eta^2 d_{ij}^2}}{\sqrt{\pi} d_{ij}^4} - \frac{4\eta^3 e^{-\eta^2 d_{ij}^2}}{\sqrt{\pi} d_{ij}^2} + \frac{3\text{erf}(\eta d_{ij})}{d_{ij}^5} \right) \\ & \times \left[\delta_{\alpha\beta} d_{ij}^\gamma + \delta_{\alpha\gamma} d_{ij}^\beta + \delta_{\beta\gamma} d_{ij}^\alpha \right] \end{aligned} \quad (256)$$

where d_{ij} is the intra-molecular distance between atom i and j . The tensors $\tilde{\mathcal{T}}_{ij}$ are used to remove *intra*-molecular electrostatics. In real space these interactions can simply be skipped. The Fourier part can then be corrected by subtracting real space value $\tilde{\mathcal{T}}_{ij}$ for excluded pairs, that is, by correcting in real space. An important note is that sometimes numerical divergences can occur. For example, in the core-shell models the shell can overlap with its core. This can be resolved by replacing them by analytical expressions for the limit $\mathbf{r} \rightarrow 0$ for small values of d , for example, $d < 10^{-4}$ Å. We have

$$\lim_{d \rightarrow 0} \tilde{\mathcal{T}}_{ij} = \frac{2\eta}{\sqrt{\pi}} \quad (257)$$

$$\lim_{d \rightarrow 0} \tilde{\mathcal{T}}_{ij}^\alpha = -\frac{4\eta^3}{3\pi} d_{ij}^\alpha \quad (258)$$

$$\lim_{d \rightarrow 0} \tilde{\mathcal{T}}_{ij}^{\alpha\beta} = \frac{8\eta^5}{4\pi} d_{ij}^\alpha d_{ij}^\beta - \delta_{\alpha\beta} \frac{4\eta^3}{3\pi} \quad (259)$$

$$\lim_{d \rightarrow 0} \tilde{\mathcal{T}}_{ij}^{\alpha\beta\gamma} = -\frac{16\eta^7}{7\pi} d_{ij}^\alpha d_{ij}^\beta d_{ij}^\gamma + \frac{8\eta^5}{4\pi} d_{ij}^\alpha \left[\delta_{\alpha\beta} d_{ij}^\gamma + \delta_{\alpha\gamma} d_{ij}^\beta + \delta_{\beta\gamma} d_{ij}^\alpha \right] \quad (260)$$

The real-space cancellation is imperfect and better cancellation can be achieved if the excluded interactions are subtracted out using a reciprocal-space expression.^[81,215]

Nymand and Linse provided expressions for \mathcal{U} , ϕ_i , E_i^α , $E_i^{\alpha\beta}$, f_i^α , and Φ for a truly infinite periodic system of charges, static dipoles, and anisotropic polarizabilities. The expressions correctly reduce to those for systems of only charges, or of only dipoles. The volume is defined by $V = L_x L_y L_z$, where L_x , L_y , and

L_z denote the lengths of the box edges, A_k is defined according to

$$A_k = \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}}, \quad (261)$$

and $Q^{q\mu}$ according to

$$Q^{q\mu} = Q^q + Q^\mu \quad (262)$$

$$Q^q = \sum_j q_j e^{-i\mathbf{k} \cdot \mathbf{r}_j} \quad (263)$$

$$Q^\mu = Q^{\mu(\text{stat})} + Q^{\mu(\text{ind})} \quad (264)$$

$$= \sum_j i(\boldsymbol{\mu}_j^{\text{stat}} \cdot \mathbf{k}) + \sum_j i(\boldsymbol{\mu}_j^{\text{ind}} \cdot \mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}_j}, \quad (265)$$

where $i^2 \equiv -1$. The bar-notation $\bar{Q}^{q\mu}$ is used for the complex conjugate of $Q^{q\mu}$.

The Potential Energy: For a system of charges, static dipoles, and anisotropic polarizabilities in an infinite lattice, the potential energy can be written as^[213]

$$\begin{aligned} \mathcal{U} = & \frac{1}{2} \underbrace{\sum_{i,j \neq i} \left(q_i \hat{\mathcal{T}}_{ij} q_j - q_i \hat{\mathcal{T}}_{ij}^\alpha \mu_j^\alpha + \mu_i^\alpha \hat{\mathcal{T}}_{ij}^\alpha q_j - \mu_i^\alpha \hat{\mathcal{T}}_{ij}^{\alpha\beta} \mu_j^\beta \right)}_{\text{Real part}} \\ & + \underbrace{\frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} A_k |Q^{q\mu}|^2}_{\text{Reciprocal part}} \\ & \underbrace{\left(-\sum_i \left(\frac{\eta}{\sqrt{\pi}} q_i^2 + \frac{2\eta^3}{3\sqrt{\pi}} \mu_i^2 \right) - \frac{1}{2} \sum_p \sum_{i,j \neq i; i,j \in p} \left(q_i \tilde{\mathcal{T}}_{ij} q_j - q_i \tilde{\mathcal{T}}_{ij}^\alpha \mu_j^\alpha + \mu_i^\alpha \tilde{\mathcal{T}}_{ij}^\alpha q_j - \mu_i^\alpha \tilde{\mathcal{T}}_{ij}^{\alpha\beta} \mu_j^\beta \right) \right)}_{\text{Self part}} \\ & + \underbrace{\frac{2\pi}{(2\epsilon_{\text{sur}} + 1)V} \left[\left(\sum_i q_i \mathbf{r}_i \right)^2 + 2 \left(\sum_i q_i \mathbf{r}_i \right) \cdot \left(\sum_j \boldsymbol{\mu}_j \right) + \left(\sum_i \boldsymbol{\mu}_i \right)^2 \right]}_{\text{Surface part}} \\ & + \mathcal{U}_{\text{pol}} \end{aligned} \quad (266)$$

where the four terms are the real, the reciprocal, the self, and the surface contribution, respectively, to the potential energy and \mathcal{U}_{pol} represents the work of forming the induced dipoles.

From Equation (266) and on, the Ewald convergence parameter η is assumed to be chosen such that the screening is sufficiently large to ensure that only particles in the primary box need to be considered when calculating the real space sum. In the self-term $\mathcal{U}_{\text{self}}$, the single sum over i is always present. In the

remaining part, the index p runs over all particles and the notation $i, j \in p$ means that atomic sites i and j both reside within particle p . Hence, the double sum over i and j contributes only if atom i and j belongs to the same particle.

Finally, the surface term \mathcal{U}_{sur} originates from the total dipole moment of the simulation box. The quantity ϵ_{sur} entering in \mathcal{U}_{sur} is the dielectric constant of the continuum surrounding the replicated sample, and is often chosen to be either ∞ (tin foil conditions) or 1 (vacuum conditions). For systems such as ionic and dipolar ones, where long-range interactions are important, Ewald summation with tin foil boundary conditions corresponds to the physically most desirable situation. The general energy expression reduces correctly for systems consisting of i) only charges by setting $\mu^{\text{stat}} = \mu^{\text{ind}} = 0$ and ii) only dipoles by setting $q = 0$ and $\mu^{\text{ind}} = 0$.

The Electrostatic Potential: The electrostatic potential on atom i arising from other charges and total dipole moments is obtained from Equation (266), and becomes^[213]

$$\begin{aligned} \phi_i = & \underbrace{\sum_{j \neq i} (\hat{\mathcal{T}}_{ij} q_j - \hat{\mathcal{T}}_{ij}^{\alpha} \mu_j^{\alpha})}_{\text{real part}} + \underbrace{\frac{4\pi}{V} \sum_{k \neq 0} A_k \text{Re} (e^{ik \cdot \mathbf{r}_i} \bar{Q}^{q\mu})}_{\text{reciprocal part}} \\ & - \underbrace{\frac{2\eta}{\sqrt{\pi}} q_i - \sum_{j \neq i; j \in p} (\tilde{\mathcal{T}}_{ij} q_j - \tilde{\mathcal{T}}_{ij}^{\alpha} \mu_j^{\alpha})}_{\text{self-term}} \\ & + \underbrace{\frac{4\pi}{(2\epsilon_{\text{sur}} + 1)V} \left[\mathbf{r}_i \cdot \sum_j (q_j \mathbf{r}_j + \boldsymbol{\mu}_j) \right]}_{\text{surface part}} \end{aligned} \quad (267)$$

For systems of only atoms the sum in the self-term vanishes. For a system of charge neutral molecules the surface term may be written as

$$\phi_i^{\text{sur}} = -\frac{2\pi}{(2\epsilon_{\text{sur}} + 1)V} \sum_j \left[q_j (\mathbf{r}_j - \mathbf{r}_i)^2 + 2\boldsymbol{\mu}_i \cdot (\mathbf{r}_j - \mathbf{r}_i) \right]. \quad (268)$$

The Electrostatic Field: Similarly, the expression for the corresponding α -component of the electrostatic field on atom i is^[213]

$$\begin{aligned} E_{i,\alpha} = & \underbrace{\sum_{j \neq i} \left(-\hat{\mathcal{T}}_{ij}^{\alpha} q_j + \hat{\mathcal{T}}_{ij}^{\alpha\beta} \mu_j^{\beta} \right)}_{\text{real part}} \\ & - \underbrace{\frac{4\pi}{V} \sum_{k \neq 0} A_k k_{\alpha} \left[\text{Re} (i e^{ik \cdot \mathbf{r}_i} \bar{Q}^{q\mu}) - \mu_i^{\beta} k_{\beta} \right]}_{\text{reciprocal part}} + \frac{4\pi}{3V} \mu_i^{\alpha} \\ & - \underbrace{\sum_{j \neq i; j \in p} \left(-\tilde{\mathcal{T}}_{ij}^{\alpha} q_j + \tilde{\mathcal{T}}_{ij}^{\alpha\beta} \mu_j^{\beta} \right)}_{\text{self part}} - \underbrace{\frac{4\pi}{(2\epsilon_{\text{sur}} + 1)V} \sum_j \left(q_j r_j^{\alpha} + \mu_j^{\alpha} \right)}_{\text{surface part}} \end{aligned} \quad (269)$$

In particular, for a system consisting of only charges the reciprocal term reduces to

$$E_{i,\alpha}^{\text{rec}} = -\frac{4\pi}{V} \sum_{k \neq 0} A_k k_{\alpha} \text{Re} (i e^{ik \cdot \mathbf{r}_i} \bar{Q}^q), \quad (270)$$

and for a system of only static dipoles it becomes

$$E_{i,\alpha}^{\text{rec}} = -\frac{4\pi}{V} \sum_{k \neq 0} A_k k_{\alpha} \left[\text{Re} (i e^{ik \cdot \mathbf{r}_i} \bar{Q}^{\mu(\text{stat})}) - \mu_{i,\beta}^{\text{stat}} k_{\beta} \right] + \frac{4\pi}{3V} \mu_{i,\alpha}^{\text{stat}} \quad (271)$$

where Q^q and $Q^{\mu(\text{stat})}$ were defined in Equations (263) and (265). For an atomic system, the self term is zero. The α -component of the electrostatic field on atom i arising from other induced dipoles only, $E_{i,\alpha}^{\text{ind}}$, is needed later, and it is obtained from Equation (269) by setting $Q_i = 0$ and by replacing μ_i^{α} by $\mu_{i,\alpha}^{\text{ind}}$ and $Q^{q\mu}$ by $Q^{\mu(\text{ind})}$.

For the periodic system, Nymand and Linse showed that i) the induced dipole moments are still given by Equation (237), but $E_{i,\alpha}$ is now given by Equation (269), and ii) the potential energy can again be separated into an electrostatic and induction term according to Equation (238). The electrostatic term \mathcal{U}_{ele} is still given by Equation (239) and the induction term by Equation (240), but ϕ_i^{stat} and $E_{i,\alpha}^{\text{stat}}$ are now given by Equation (267) and Equation (269), respectively, after replacing $\mu_{i,\alpha}$ with $\mu_{i,\alpha}^{\text{stat}}$ and $Q^{q\mu}$ by $Q^q + Q^{\mu(\text{stat})}$.

The Electrostatic Gradient: The $\alpha\beta$ -component of the field gradient i on atom i is given by^[213]

$$\begin{aligned} E_{i,\alpha\beta} = & \underbrace{\sum_{j \neq i} \left(-\hat{\mathcal{T}}_{ij}^{\alpha\beta} q_j + \hat{\mathcal{T}}_{ij}^{\alpha\beta\gamma} \mu_j^{\gamma} \right)}_{\text{real part}} \\ & + \underbrace{\frac{4\pi}{V} \sum_{k \neq 0} A_k k_{\alpha} k_{\beta} \left[\text{Re} (e^{ik \cdot \mathbf{r}_i} \bar{Q}^{q\mu}) - q_i \right] - \frac{4\pi}{3V} q_i \delta_{\alpha\beta}}_{\text{reciprocal part}} \\ & - \underbrace{\sum_{j \neq i; j \in p} \left(-\tilde{\mathcal{T}}_{ij}^{\alpha\beta} q_j + \tilde{\mathcal{T}}_{ij}^{\alpha\beta\gamma} \mu_j^{\gamma} \right)}_{\text{self term}} \end{aligned} \quad (272)$$

For an atomic system, the self-term is zero. Note that $E_{i,\alpha}^{\text{sur}}$ is zero.

The Force: The expressions for the α -component of the force on atom i is readily available from Equation (266) by taking its derivative with respect to $r_{i,\alpha}$ and reads^[213]

$$\begin{aligned} f_i^{\alpha} = & \underbrace{\sum_{j \neq i} \left(-q_i \hat{\mathcal{T}}_{ij}^{\alpha} q_j + q_i \hat{\mathcal{T}}_{ij}^{\alpha\beta} \mu_j^{\beta} - \mu_i^{\beta} \hat{\mathcal{T}}_{ij}^{\beta\alpha} q_j + \mu_i^{\beta} \hat{\mathcal{T}}_{ij}^{\beta\alpha\gamma} \mu_j^{\gamma} \right)}_{\text{real part}} \\ & - \underbrace{\frac{4\pi}{V} \sum_{k \neq 0} A_k k_{\alpha} \text{Re} [(i q_i - \mu_i^{\beta} k_{\beta}) e^{ik \cdot \mathbf{r}_i} \bar{Q}^{q\mu}]}_{\text{reciprocal part}} \end{aligned}$$

$$\underbrace{- \sum_{j \neq i; j, \in p} \left(-q_i \tilde{\mathcal{T}}_{ij}^\alpha q_j + q_i \tilde{\mathcal{T}}_{ij}^{\alpha\beta} \mu_j^\beta - \mu_i^\beta \tilde{\mathcal{T}}_{ij}^{\beta\alpha} q_j + \mu_i^\beta \tilde{\mathcal{T}}_{ij}^{\beta\alpha\gamma} \mu_j^\gamma \right)}_{\text{self part}}$$

$$- \underbrace{\frac{4\pi}{(2\epsilon_{\text{sur}} + 1)V} q_i \sum_j \left(q_j r_j^\alpha + \mu_j^\alpha \right)}_{\text{surface part}} \quad (273)$$

Nyman and Linse showed that the use of E_i^α (given by Equation (269)) and $E_i^{\alpha\beta}$ (given by Equation (272)) leads exactly to Equation (244), and that Equation (244) holds term-wise. For a system consisting of only charges the reciprocal term reduces to

$$f_{i,\alpha}^{\text{rec}} = \frac{4\pi}{V} \sum_{k \neq 0} A_k k_\alpha q_i \text{Im} \left[e^{i\mathbf{k} \cdot \mathbf{r}_i} \bar{Q}^q \right], \quad (274)$$

and for a system consisting of only static dipoles the reciprocal term reduces to

$$f_{i,\alpha}^{\text{rec}} = \frac{4\pi}{V} \sum_{k \neq 0} A_k k_\alpha k_\beta \mu_{i,\beta}^{\text{stat}} \text{Re} \left[e^{i\mathbf{k} \cdot \mathbf{r}_i} \bar{Q}^{\mu(\text{stat})} \right]. \quad (275)$$

Note that the self-term is not constant due to the presence of μ_i^α . For an atomic system, the self-term is zero.

The Virial: The virial can be expressed as^[213]

$$\Phi = \underbrace{-\frac{1}{2} \sum_{i,j \neq i} \mathbf{r}_{ij} \cdot \mathbf{f}_{ij}^{\text{real}}}_{\text{real part}}$$

$$- \underbrace{\frac{4\pi}{V} \sum_{k \neq 0} A_k \left[\text{Re} \left(Q^\mu \bar{Q}^{q\mu} \right) + \frac{1}{2} |Q^{q\mu}|^2 \left(1 - \frac{k^2}{2\eta^2} \right) \right]}_{\text{reciprocal part}}$$

$$- \underbrace{\frac{1}{2} \sum_p \sum_{i,j \in p} \mathbf{d}_{ij} \cdot \mathbf{f}_{ij}^{\text{self}}}_{\text{self part}} \underbrace{\left(\mathcal{U}^{\text{sur},qq} + 2\mathcal{U}^{\text{sur},q\mu} + 3\mathcal{U}^{\text{sur},\mu\mu} \right)}_{\text{surface part}}$$

$$+ \sum_i \mathbf{d}_i \cdot \left(\mathbf{f}_i^{\text{real}} + \mathbf{f}_i^{\text{rec}} + \mathbf{f}_i^{\text{sur}} \right) \quad (276)$$

where

$$\mathcal{U}^{\text{sur},qq} \equiv \frac{2\pi}{(2\epsilon_{\text{sur}} + 1)V} \left(\sum_i q_i \mathbf{r}_i \right)^2, \quad (277)$$

$$\mathcal{U}^{\text{sur},q\mu} \equiv \frac{4\pi}{(2\epsilon_{\text{sur}} + 1)V} \left(\sum_i q_i \mathbf{r}_i \right) \cdot \left(\sum_j \mu_j \right), \quad (278)$$

$$\mathcal{U}^{\text{sur},\mu\mu} \equiv \frac{2\pi}{(2\epsilon_{\text{sur}} + 1)V} \left(\sum_i \mu_i \right)^2 \quad (279)$$

4.5. Charge Only Systems

The expressions of the previous section reduce to the familiar ones for systems consisting of only charges. They are easily incorporated into the usual Ewald summation routines.

The Potential Energy: The potential energy can be written as

$$\mathcal{U} = \mathcal{U}^{\text{real}} + \mathcal{U}^{\text{rec}} + \mathcal{U}^{\text{self}} \quad (280)$$

where

$$\mathcal{U}^{\text{real}} = \frac{1}{2} \sum_{i,j \neq i} q_i q_j \frac{\text{erfc}(\eta r_{ij})}{r_{ij}} \quad (281)$$

$$\mathcal{U}^{\text{rec}} = \frac{2\pi}{V} \sum_{k \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} \left(\left| \sum_{i=1}^N q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 + \left| \sum_{i=1}^N q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 \right) \quad (282)$$

$$\mathcal{U}^{\text{self}} = - \sum_i \frac{\eta}{\sqrt{\pi}} q_i^2 \quad (283)$$

The Electrostatic Potential: The electrostatic potential in atom i arising from other charges is given by

$$\phi_i = \phi_i^{\text{real}} + \phi_i^{\text{rec}} + \phi_i^{\text{self}} \quad (284)$$

where

$$\phi_i^{\text{real}} = \frac{1}{2} \sum_{j \neq i} q_j \frac{\text{erfc}(\eta r_{ij})}{r_{ij}} \quad (285)$$

$$\phi_i^{\text{rec}} = \frac{4\pi}{V} \sum_{k \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} \left[\cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^N q_j \cos(\mathbf{k} \cdot \mathbf{r}_j) \right.$$

$$\left. + \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^N q_j \sin(\mathbf{k} \cdot \mathbf{r}_j) \right] \quad (286)$$

$$\phi_i^{\text{self}} = - \frac{2\eta}{\sqrt{\pi}} q_i \quad (287)$$

The Electrostatic Field: The α -component of the electrostatic field on atom i is given by

$$E_i^\alpha = E_{i,\alpha}^{\text{real}} + E_{i,\alpha}^{\text{rec}} \quad (288)$$

where

$$E_{i,\alpha}^{\text{real}} = \sum_{j \neq i} q_j \left(\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) r_{ij}^\alpha \quad (289)$$

$$E_{i,\alpha}^{\text{rec}} = \frac{4\pi}{V} \sum_{k \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} k_\alpha \left[\sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_j q_j \cos(\mathbf{k} \cdot \mathbf{r}_j) \right.$$

$$\left. - \cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_j q_j \sin(\mathbf{k} \cdot \mathbf{r}_j) \right] \quad (290)$$

The Electrostatic Field Gradient: The $\alpha\beta$ -component of the field gradient on atom i is given by

$$E_i^{\alpha\beta} = E_{i,\alpha\beta}^{\text{real}} + E_{i,\alpha\beta}^{\text{rec}} \quad (291)$$

where

$$E_{i,\alpha\beta}^{\text{real}} = - \sum_{j \neq i} q_j \left(\frac{6\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^4} + \frac{4\eta^3 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{3\text{erfc}(\eta r_{ij})}{r_{ij}^5} \right) r_{ij}^\alpha r_{ij}^\beta + \delta_{\alpha\beta} \sum_{j \neq i} q_j \left(\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) \quad (292)$$

$$E_{i,\alpha\beta}^{\text{rec}} = \frac{4\pi}{V} \sum_{k \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} k_\alpha k_\beta \left[\cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^N q_j \cos(\mathbf{k} \cdot \mathbf{r}_j) + \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^N q_j \sin(\mathbf{k} \cdot \mathbf{r}_j) - q_i \right] - \frac{4\pi}{3V} q_i \delta_{\alpha\beta} \quad (293)$$

The Force: The α -component of the force on atom i is given by

$$f_i^\alpha = f_{i,\alpha}^{\text{real}} + f_{i,\alpha}^{\text{rec}} \quad (294)$$

where

$$f_{i,\alpha}^{\text{real}} = q_i \sum_{j \neq i} q_j \left(\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) r_{ij}^\alpha \quad (295)$$

$$f_{i,\alpha}^{\text{rec}} = q_i \frac{4\pi}{V} \sum_{k \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} k_\alpha \left[\sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_j q_j \cos(\mathbf{k} \cdot \mathbf{r}_j) - \cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_j q_j \sin(\mathbf{k} \cdot \mathbf{r}_j) \right] \quad (296)$$

The Hessian: The atomic expression for the Hessian can be written out explicitly as^[163]

$$\frac{\partial^2 \mathcal{U}^{\text{real}}}{\partial r_i^\alpha \partial r_i^\beta} = \frac{\partial^2 \mathcal{U}^{\text{real}}}{\partial r_i^\alpha \partial r_i^\beta} = \sum_{i < j} q_i q_j \left(\frac{6\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^4} + \frac{4\eta^3 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{3\text{erfc}(\eta r_{ij})}{r_{ij}^5} \right) r_{ij}^\alpha r_{ij}^\beta - \delta_{\alpha\beta} \sum_{i < j} q_i q_j \left(\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) \quad (297)$$

$$\frac{\partial^2 \mathcal{U}^{\text{real}}}{\partial r_i^\alpha \partial r_j^\beta} = \frac{\partial^2 \mathcal{U}^{\text{real}}}{\partial r_j^\alpha \partial r_i^\beta} = - \sum_{i < j} q_i q_j \left(\frac{6\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^4} + \frac{4\eta^3 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{3\text{erfc}(\eta r_{ij})}{r_{ij}^5} \right) r_{ij}^\alpha r_{ij}^\beta + \delta_{\alpha\beta} \sum_{i < j} q_i q_j \left(\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) \quad (298)$$

$$\frac{\partial^2 \mathcal{U}^{\text{rec}}}{\partial r_i^\alpha \partial r_j^\beta} = \frac{4\pi}{V} \sum_{k \neq 0} A_k k_\alpha k_\beta \left\{ -\delta_{ij} \left[\left(\sum_{l=1}^N q_l \cos(\mathbf{k} \cdot \mathbf{r}_l) \right) q_j \cos(\mathbf{k} \cdot \mathbf{r}_j) + \left(\sum_{l=1}^N q_l \sin(\mathbf{k} \cdot \mathbf{r}_l) \right) q_j \sin(\mathbf{k} \cdot \mathbf{r}_j) \right] + [q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) q_j \cos(\mathbf{k} \cdot \mathbf{r}_j) + q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) q_j \sin(\mathbf{k} \cdot \mathbf{r}_j)] \right\} \quad (299)$$

and is similar to the expression given by Krishnan and Balasubramanian.^[177] Note that the Hessian requires a double summation over particles.

The Strain Derivatives: The strain derivative of the Ewald summation reads^[182]

$$\frac{\partial \mathcal{U}^{\text{real}}}{\partial \epsilon_{\alpha\beta}} = - \sum_{i < j} q_i q_j \left(\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) r_{ij}^\alpha r_{ij}^\beta \quad (300)$$

$$\frac{\partial \mathcal{U}^{\text{rec}}}{\partial \epsilon_{\alpha\beta}} = - \frac{2\pi}{V} \sum_{k \neq 0} A_k \left[\left(\sum_{i=1}^N q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right)^2 + \left(\sum_{i=1}^N q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \right)^2 \right] \Theta_{\alpha\beta} \quad (301)$$

where

$$\Theta_{\alpha\beta} = \delta_{\alpha\beta} - 2 \frac{k_\alpha k_\beta}{\lambda^2} \quad (302)$$

$$\frac{1}{\lambda^2} = \frac{1}{4\eta^2} + \frac{1}{k^2} \quad (303)$$

The Ewald expressions for positional cross derivatives read^[163]

$$\frac{\partial \mathcal{U}^{\text{real}}}{\partial \epsilon_{\alpha\beta} \partial r_i^\gamma} = \sum_{i < j} q_i q_j \left(\frac{6\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^4} + \frac{4\eta^3 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{3\text{erfc}(\eta r_{ij})}{r_{ij}^5} \right) r_{ij}^\alpha r_{ij}^\beta r_{ij}^\gamma - \sum_{i < j} q_i q_j \left(\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) (\delta_{\alpha\gamma} r_{ij}^\beta + \delta_{\beta\gamma} r_{ij}^\alpha) \quad (304)$$

$$\frac{\partial \mathcal{U}^{\text{real}}}{\partial \epsilon_{\alpha\beta} \partial r_j^\gamma} = - \sum_{i < j} q_i q_j \left(\frac{6\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^4} + \frac{4\eta^3 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{3\text{erfc}(\eta r_{ij})}{r_{ij}^5} \right) r_{ij}^{\alpha} r_{ij}^{\beta} r_{ij}^{\gamma} + \sum_{i < j} q_i q_j \left(\frac{2\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \right) (\delta_{\alpha\gamma} r_{ij}^{\beta} + \delta_{\beta\gamma} r_{ij}^{\alpha}) \quad (305)$$

$$\frac{\partial^2 \mathcal{U}^{\text{rec}}}{\partial \epsilon_{\alpha\beta} \partial r_i^\gamma} = \frac{4\pi}{V} \sum_{k \neq 0} A_k \left[\left(\sum_{j=1}^N q_j \cos(\mathbf{k} \cdot \mathbf{r}_j) \right) q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) - \left(\sum_{j=1}^N q_j \sin(\mathbf{k} \cdot \mathbf{r}_j) \right) q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right] \Theta_{\alpha\beta} k_\gamma \quad (306)$$

The strain-strain derivative for the Ewald summation is given by^[182]

$$\frac{\partial \mathcal{U}^{\text{real}}}{\partial \epsilon_{\alpha\beta} \partial \epsilon_{\mu\nu}} = \sum_{i < j} q_i q_j \left(\frac{6\eta e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^4} + \frac{4\eta^3 e^{-\eta^2 r_{ij}^2}}{\sqrt{\pi} r_{ij}^2} + \frac{3\text{erfc}(\eta r_{ij})}{r_{ij}^5} \right) r_{ij}^{\alpha} r_{ij}^{\beta} r_{ij}^{\mu} r_{ij}^{\nu} + \frac{1}{2} \left[\delta_{\beta\nu} \frac{\partial \mathcal{U}}{\partial \epsilon_{\alpha\mu}} + \delta_{\alpha\nu} \frac{\partial \mathcal{U}}{\partial \epsilon_{\beta\mu}} + \delta_{\beta\mu} \frac{\partial \mathcal{U}}{\partial \epsilon_{\alpha\nu}} + \delta_{\alpha\mu} \frac{\partial \mathcal{U}}{\partial \epsilon_{\beta\nu}} \right] \quad (307)$$

$$\frac{\partial \mathcal{U}^{\text{rec}}}{\partial \epsilon_{\alpha\beta} \partial \epsilon_{\mu\nu}} = \frac{2\pi}{V} \sum_{k \neq 0} A_k \left[\left(\sum_{i=1}^N q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right)^2 + \left(\sum_{i=1}^N q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \right)^2 \right] \Omega_{\alpha\beta\mu\nu} \quad (308)$$

where

$$\Omega_{\alpha\beta\mu\nu} = \Theta_{\alpha\beta} \Theta_{\mu\nu} + (\delta_{\alpha\mu} \delta_{\beta\nu} + \delta_{\alpha\nu} \delta_{\beta\mu}) + 4 \frac{k_\alpha k_\beta k_\mu k_\nu}{k^4} - 2 \frac{k_\alpha k_\nu \delta_{\beta\mu} + k_\beta k_\nu \delta_{\alpha\mu} + k_\beta k_\mu \delta_{\alpha\nu} + k_\alpha k_\mu \delta_{\beta\nu}}{\lambda^2} \quad (309)$$

4.6. Polarization via Induced Dipoles

There are basically three methods to generate induced dipoles^[95,216]

1. The Drude oscillator model (also called Core-Shell model).

In the shell model, the electronic polarizability is incorporated by representing an atom as a two-particle system: a “core” charge $z_i + q_i$ is attached by a harmonic spring (with spring constant k) to a “shell” to a shell charge $-q_i$.^[216] Atomic polarizability α_i , is related to force constant k of the harmonic spring connecting the core and shell and is determined by^[96,216]

$$\alpha_i = q_i^2 / k \quad (310)$$

The shells are massless. For minimization of core-shell models the Hessian matrix \mathcal{H} contains both the core and the shells because the shells need to be minimized with respect to the cores too.^[217] The set of vibrational frequencies can be obtained by diagonalization of the dynamical matrix D

$$D = M^{-\frac{1}{2}} \left[\mathcal{H}_{\text{core-core}} - \mathcal{H}_{\text{core-shell}} \mathcal{H}_{\text{shell-shell}}^{-1} \mathcal{H}_{\text{shell-core}} \right] M^{-\frac{1}{2}} \quad (311)$$

where M is a diagonal matrix that contains the masses of the atoms. For use in MD, Mitchell and Fincham assigned small masses to the shells (typically taken to be less than 10% of the total particle mass) and proposed a method to determine their motion using adiabatic dynamics.^[218]

2. The Fluctuation–Charge (FQ) model,

The FQ model is based on the principle of electronegativity equalization (EE): a charge flows between atoms until electronegativities of the atoms become equalized.^[96] The partial charges on each atom q_i of the molecule are found by minimizing the electrostatic energy of the system (equivalent to equalizing electronegativities) in a given configuration subject to the constraint that the total charge is conserved. Since charges depend on interactions with other charges located on the same molecule or other molecules, their values change with every time step or sampled configuration during a simulation. During the charge equalization process, charges may be constrained to movement within a molecule or movement between any atom pairs. Application of the latter option leads to charge-transfer effects. The method is further described in Subsection 4.13.

3. The point-polarizable dipole model,

In the classical point dipole approach, polarization energy is described as the interaction between static point charges and static dipoles and the dipole moments they induce.^[96] A straightforward way to choose the sites of the inducible dipoles is to put them on all atoms of the system.^[95] In a molecular simulation, numerical problems may occur when two inducible dipoles come spatially too close to each other, and the dipolar interaction between them will mutually enhance their induced dipoles to infinite size. The polarization catastrophe can be avoided in several ways,^[95] for example by choosing VDW parameters large enough such that interatomic distances are always larger than this critical distance. Choosing dipole sites that do not reside on atoms is possible but leads to more complicated expressions for the forces on these atoms.

The procedure for induced dipoles in the formulation of Nyman and Linse for calculating the electrostatic and induction contributions of the energy, forces, and the virial in a molecular system, where each molecule is described by atomic charges, dipole moments, and anisotropic polarizabilities, is as follows^[213]

1. Calculate the electrostatic potential ϕ_i^{stat} and the electrostatic field $E_{i,\alpha}^{\text{stat}}$ from charges and static dipole moments, according to Equation (241) and Equation (242), respectively.
2. Calculate the induced dipole moments $\mu_{i,\alpha}^{\text{ind}}$ either by prediction, or self-consistent iteration using Equation (237) and $E_i^\alpha = E_{i,\alpha}^{\text{stat}} + E_{i,\alpha}^{\text{ind}}$ with $E_{i,\alpha}^{\text{ind}} = \sum_{j \neq i} \mathcal{T}_{ij}^{\alpha\beta} \mu_{j,\beta}^{\text{ind}}$.

3. Calculate the electrostatic energy $\mathcal{U}_{\text{elec}}$ and the induction energy \mathcal{U}_{ind} from Equation (239) and Equation (240), respectively.
4. If the induced dipole moments were predicted above, calculate $E_{i,\alpha}^{\text{ind}}$.
5. Calculate the electrostatic field gradient $E_{i,\alpha\beta}$ from charges and total dipole moments according to Equation (236).
6. Calculate the forces f_i^α from Equation (244).
7. Calculate the virial Φ from Equation (246).

Due to the many-body nature of polarization, the interactions between all molecules change and have to be re-computed for every step. In MD simulations the full recomputation has to be done any way, but simple MC moves work via computing energy-differences between the new and old state for only the atoms that are changed. Having to recompute the *full* energy for the new and old state means an order-of-magnitude penalty in efficiency. Lachet et al. therefore proposed a procedure to mitigate this limitation by using only the first term of the multipole expansion.^[219] The induction energy is then given by

$$U_{\text{ind}} = -\frac{1}{2} \sum_{i \in \text{ads}} \alpha_i \cdot |\mathbf{E}_i|^2 \quad (312)$$

where α_i is the atomic polarizability of interaction site i of the adsorbate atoms, and \mathbf{E}_i is the electrostatic field at the position of atom i due to the partial charges of all the framework atoms. The approach accounts solely for polarization between the framework and neglects polarization caused by induced dipoles (called the “back-polarization”). In this way, the difference in induction energy to the previous configuration can simply be added as another energy term in the acceptance rule of the Monte Carlo algorithm, and the computational cost is similar to simulations without considering explicit polarization. Lachet et al. showed that for xylene NaY zeolite system this procedure captures around 94% of the total induction energy (the back-polarization effect accounts for only about 6% of the total induction energy).^[219]

When explicitly accounting for polarization, one has to ensure that the force field parameters describing the remaining interactions do not include an implicit polarization contribution which would have to be removed. Otherwise, the contribution of polarization would be double counted, once implicitly and once explicitly. The removal of implicit polarization is necessary if a standard force field is used as starting point for the development of a polarizable force field, because current force fields are likely to be calibrated to reproduce certain experimentally observed properties.

4.7. Electric-Field Dependent Potentials

Cicu et al. proposed a general method to include electric-field-dependent terms in empirical potential functions representing interatomic interactions,^[220] and this subsection is taken from their work. Let us consider a system of N charged particles, interacting with each other via a potential energy function depending on their positions and on the electric field acting upon each of them and generated by other particles (which also depend on the

positions of the particles):

$$\mathcal{U} = \mathcal{U}(\mathbf{r}_1, \dots, \mathbf{r}_N | \mathbf{E}_1, \dots, \mathbf{E}_N) = \mathcal{U}(\{\mathbf{r}_i\} | \mathbf{E}_i \{\mathbf{r}_j\}) \quad (i, j = 1, \dots, N) \quad (313)$$

where \mathbf{r}_i are the positions of the particles and \mathbf{r}_i are the electric fields. The force acting on a particle i is given by^[220]

$$\mathbf{f}_i = -\frac{\partial \mathcal{U}}{\partial \mathbf{r}_i} = -\left(\frac{\partial \mathcal{U}}{\partial \mathbf{r}_i}\right)_{\mathbf{E}} - \sum_j \frac{\partial \mathcal{U}}{\partial \mathbf{E}_j} \frac{\partial \mathbf{E}_j}{\partial \mathbf{r}_i} \quad (314)$$

where \mathbf{E} stands for $\{\mathbf{E}_i\}$. In particular, it is worth to note that the terms in Equation (314) including the $3N \times 3N$ matrix,

$$\frac{\partial \mathbf{E}_j}{\partial \mathbf{r}_i} \quad (315)$$

corresponds to many-body forces involving in principle *all* the particles of the system as it is better illustrated if the atom pair approximation is assumed. In this approximation, the potential energy is represented by the sum of pairwise interactions $\mathcal{U}_p(\mathbf{r}_{ij} | \mathbf{E})$ between i and j . Therefore Equation (314) and Equation (315) become^[220]

$$\mathcal{U} = \sum_i \sum_{j>i} \mathcal{U}_p(\mathbf{r}_{ij} | \mathbf{E}) \quad (316)$$

$$\mathbf{f}_i = -\left(\sum_{j>i} \frac{\partial \mathcal{U}_p(\mathbf{r}_{ij} | \mathbf{E})}{\partial \mathbf{r}_i}\right)_{\mathbf{E}} - \sum_{j>i} \frac{\partial \mathcal{U}_p(\mathbf{r}_{ij} | \mathbf{E})}{\partial \mathbf{E}_i} \frac{\partial \mathbf{E}_i}{\partial \mathbf{r}_i} - \sum_{k \neq i} \sum_{j>k} \frac{\partial \mathcal{U}_p(\mathbf{r}_{kj} | \mathbf{E})}{\partial \mathbf{E}_k} \frac{\partial \mathbf{E}_k}{\partial \mathbf{r}_i} \quad (317)$$

where, again, \mathbf{E} stands for $\{\mathbf{E}_i\}$ and the terms containing the electric field gradients have been grouped in two sums, the former including the gradient of the electric field acting on particle i for which the force is calculated, and the latter including the gradient of the electric field *acting on all the other particles multiplied by the derivatives, with respect to these electric fields, of the potential function terms involving particle j , even though these terms do not include explicitly any interaction with the particle i* . It is worthy to note that in general the last group of terms is not zero: indeed it ensured that the third law of the dynamics (as generalized for many-body forces) is obeyed.

4.8. Charged Frameworks

A minor complication arises when the framework has a non-zero net electric charge compensated by counter ions. Let us assume the framework is kept fixed, but the cations are allowed to move. Although the system a whole may be electrically neutral, the omission of framework-framework interactions from the calculation also means the Ewald Fourier-contributions should be splitted into separate sums which are each net-charged. Not only the interactions of a framework atoms with other framework atoms should be omitted, but also the interactions with all its images.

Splitting of the potential energy into separate contribution involves computing cross terms between component \mathcal{A} and \mathcal{B} . In real space this is trivially accomplished, but the reciprocal separation is more difficult. Cross term interaction energies in reciprocal space are given by

$$\mathcal{U}_{\mathcal{A},\mathcal{B}}^{\text{rec}} = \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} \left[\left(\sum_{i \in \mathcal{A}} q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right) \left(\sum_{i \in \mathcal{B}} q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right) + \left(\sum_{i \in \mathcal{A}} q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \right) \left(\sum_{i \in \mathcal{B}} q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \right) \right] \quad (318)$$

It is important to note that Equation (318) only applies when both the separate sums over species \mathcal{A} and \mathcal{B} are charge neutral. This is in general not the case. For example, in MC adsorption simulations one often encounters a negatively charged zeolite charge-compensated by either cations or protons. All-silica zeolites devoid of cations are neutral, but framework ions like aluminum induce a net-charge.

It is customary to treat charged periodic systems via a uniform neutralizing background plasma. However, this leads to serious artifacts in both system energy and pressure and leads to unrealistic behavior.^[221] Bogusz et al. corrected these artifacts by instituting a net-charge correction term that consists of subtracting off the Ewald sum for a single particle with charge equal to the net charge.^[221] This correction implicitly restores η -independence in net-charged systems. We define \mathcal{U}_ζ as the reciprocal energy of single ion place at the center of charge \mathbf{o}

$$\mathcal{U}_\zeta = \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} \left[(\cos(\mathbf{k} \cdot \mathbf{o}))^2 + (\sin(\mathbf{k} \cdot \mathbf{o}))^2 \right] - \frac{\eta}{\sqrt{\pi}}. \quad (319)$$

If only the energy is needed the particle can simply be placed at the origin. Note that the term then only depends on the shape and size of the simulation cell and has be computed only once if the cell does not change. The charge–charge interaction energy $\mathcal{U}_{\mathcal{A},\mathcal{B}}$ of component \mathcal{A} and \mathcal{B} is given by

$$\mathcal{U}_{\mathcal{A},\mathcal{B}} = \mathcal{U}_{\mathcal{A},\mathcal{B}}^{\text{real}} + \mathcal{U}_{\mathcal{A},\mathcal{B}}^{\text{rec}} + \mathcal{U}_{\mathcal{A},\mathcal{B}}^{\text{self}} + \mathcal{U}^{\text{corr}} \quad (320)$$

$$\mathcal{U}_{\mathcal{A},\mathcal{B}}^{\text{self}} = -(1 - \delta_{\mathcal{A}\mathcal{B}}) \frac{\eta}{\sqrt{\pi}} \sum_{i \in \mathcal{A},\mathcal{B}} q_i^2 \quad (321)$$

$$\mathcal{U}^{\text{corr}} = \left(\sum_{i \in \mathcal{A}} q_i \right) \left(\sum_{i \in \mathcal{B}} q_i \right) \mathcal{U}_\zeta \quad (322)$$

where $\delta_{ij} = 1$ if $i = j$, zero otherwise, denotes the Kronecker delta. These expressions are valid in any kind of net-charge arrangement, including a net-charge of the total system.

4.9. MD Implementation

In an MD simulation, all atoms move according to Newton's equations of motion and the forces computed on the atoms. Equation (296) shows that we basically need to compute two

sums:

$$\sum_{\mathbf{k} \neq 0} \left[\sum_{i=1}^N q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right] \quad \text{and} \quad \sum_{\mathbf{k} \neq 0} \left[\sum_{i=1}^N q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \right] \quad (323)$$

However, cosine and sine are very expensive to compute, on modern CPU's roughly 10 times more expensive than multiplication and 15 times more expensive than addition or subtraction. One trick is to compute the sin and cos only for the first \mathbf{k} -vector and use recursion to build up the higher terms. Chebyshev method is a recursive algorithm for finding the n th multiple angle formula knowing the $(n-1)$ th and $(n-2)$ th values

$$\cos(nx) = 2 \cos(x) \cos((n-1)x) - \cos((n-2)x) \quad (324)$$

$$\sin(nx) = 2 \cos(x) \sin((n-1)x) - \sin((n-2)x) \quad (325)$$

Alternative, we can start with $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ and substitute $nx = a$ and $x = b$ and obtain

$$\cos(nx+x) = \cos(nx)\cos(x) - \sin(nx)\sin(x) \quad (326)$$

This way we obtain for $\cos(nx)$ and $\sin(nx)$

$$\cos(nx) = \cos((n-1)x)\cos(x) - \sin((n-1)x)\sin(x) \quad (327)$$

$$\sin(nx) = \sin((n-1)x)\cos(x) + \cos((n-1)x)\sin(x) \quad (328)$$

This means we have to compute $\cos(x)$ and $\sin(x)$ once, and build up the multiple angles via the recursion formula using addition/subtraction and multiplication only.

A doubling in efficiency can be gained by making use of the symmetry of the reciprocal lattice. The loop over all \mathbf{k} -vectors $\mathbf{k} = 2\pi(\frac{n_x}{L_x}, \frac{n_y}{L_y}, \frac{n_z}{L_z})$ range over^[222]:

- n_x ranges over the value 0 to k_x^{max}
- n_y ranges over the values $-k_y^{\text{max}}$ to k_y^{max} except when $n_x = 0$ where it ranges over 0 to k_y^{max}
- n_z ranges over the values $-k_z^{\text{max}}$ to k_z^{max} except when $n_x = n_y = 0$ where it ranges over 1 to k_z^{max}

The Ewald method has an execution time which scales as $N^{3/2}$ for appropriate η and number of \mathbf{k} -vectors.^[223] When the total time is optimized it is roughly equally divided between real and reciprocal space parts of the calculation.^[223,224] The parameter η dictates the rate of convergence of the real and Fourier-space sums. The real space sum is truncated at r_{cut} and η must be chosen such that the contributions beyond r_{cut} are negligible. The errors for certain cutoffs can be estimated.^[224–226] The cutoff r_{cut} is usually set to $r_{\text{cut}} = L/2$ and in practice η is close to $6/L$ and typically 200–300 wavectors are used in the real-space sum.^[78] DL_POLY computes the η and \mathbf{k}_{max} -vectors given an input precision ϵ and cutoff r_{cut} ^[222]

$$\epsilon' = \min(|\epsilon|, 0.5) \quad (329)$$

$$\xi = \sqrt{|\log(\epsilon' r_{\text{cut}})|} \quad (330)$$

$$\eta = \sqrt{|\log(\epsilon' r_{\text{cut}} \xi)|} / r_{\text{cut}} \quad (331)$$

$$\chi = \sqrt{-\log(\epsilon' r_{\text{cut}} (2\xi\eta)^2)} \quad (332)$$

$$k_x^{\text{max}} = \text{rint}(0.25 + L_x \eta \chi / \pi) \quad (333)$$

$$k_y^{\text{max}} = \text{rint}(0.25 + L_y \eta \chi / \pi) \quad (334)$$

$$k_z^{\text{max}} = \text{rint}(0.25 + L_z \eta \chi / \pi) \quad (335)$$

where L_x , L_y , and L_z are the cell lengths, and the function "rint" rounds off to the nearest integer. These estimates for the Ewald parameters are largely based on relative error estimates for the real and reciprocal space terms

$$\epsilon \approx \text{erfc}(\eta r_{\text{cut}}) / r_{\text{cut}} \approx \exp[-(\eta r_{\text{cut}})^2] / r_{\text{cut}} \quad (336)$$

$$\epsilon \approx \exp[-k_{\text{max}}^2 / (4\eta^2)] / k_{\text{max}}^2 \quad (337)$$

In LAMMPS,^[227] the η parameter is chosen as

$$\xi = \epsilon \sqrt{N r_{\text{cut}} L_x L_y L_z} / \left(2 \sum_i q_i^2 \right) \quad (338)$$

$$\eta = \begin{cases} \sqrt{-\log \xi} & \xi < 1.0 \\ 1.35 - 0.15 \log(\epsilon) / r_{\text{cut}} & \xi \geq 1.0 \end{cases} \quad (339)$$

The k_{max} are obtained by iterating from $k_{\text{max}} = 1$ until the error estimate

$$\epsilon = \left(2 \sum_i q_i^2 \right) (\eta / L) \sqrt{\pi k_{\text{max}} N} \exp \left[- \left(\frac{\pi k_{\text{max}}}{\eta L} \right)^2 \right] \quad (340)$$

is smaller than the desired precision.

As mentioned, the structure factor $S(\mathbf{k})$ (Equation (207)) can be viewed as a discrete Fourier transform of a set of charges placed irregularly within the unit cell. The Smooth Particle Mesh Ewald (SPME) method is such an FFT-based method for the fast evaluation of electrostatic interactions under periodic boundary conditions.^[228] The reciprocal-space cost of the Ewald summation is reduced to $\mathcal{O}(N \log N)$ and hence SPME is much faster for large systems. However, the pre-factor is much larger, and it is estimated that the cross-over point happens around 10 000 ions.^[217] Cisneros et al. reviewed the classical Ewald summations, particle–particle particle–mesh, particle–mesh Ewald algorithms, multigrid, fast multipole, and local methods.^[229] As we shall see in the next section, conventional Ewald is more convenient for MC algorithms. An alternative for condensed matter systems is to omit the Fourier part because in these systems the inherent screening results in an *effective* short-ranged interaction. An example is the Wolf direction summation method.^[230] Rahbari et al. showed that this direct summation method gave similar results as the Ewald summation in aqueous methanol mixtures.^[231]

4.10. MC Implementation

Most MC moves are based on changing the coordinates of a single molecule and computing the energy difference of a new *trial*-molecule with respect to the old molecule. As the positions of all the other molecules are unchanged it is unnecessary to recompute the reciprocal energy of the non-moving molecules. An important example is a rigid zeolite. The precomputing and storing of the Ewald sums makes the Ewald-summation method the preferred method for Monte Carlo simulations.

Let's define

$$Q(\mathcal{A}, \mathbf{k}) = \sum_{i \in \mathcal{A}} q_i e^{i(\mathbf{k} \cdot \mathbf{r}_i)} \quad (341)$$

We assume the simulation starts with a full computation of Equation (322), and the sums per \mathbf{k} vector are stored in $Q(\mathcal{A}, \mathbf{k})$. For each of the n -species the sums $Q(\mathcal{A}, \mathbf{k})$ and the net-charges per species needs to be stored. After an accepted move the sums and net-charges are updated

$$\forall \mathbf{k} Q(\mathcal{A}, \mathbf{k}) \rightarrow Q(\mathcal{A}, \mathbf{k}) + Q(\text{New}, \mathbf{k}) - Q(\text{Old}, \mathbf{k}) \quad (342)$$

$$\mathcal{N}(\mathcal{A}) \rightarrow \mathcal{N}(\mathcal{A}) + \sum_{i \in \text{New}} q_i - \sum_{i \in \text{Old}} q_i \quad (343)$$

The difference of a selected molecule in the new state and the old molecule belonging to species \mathcal{A} can be calculated by

$$\begin{aligned} \Delta \mathcal{U}_{\mathcal{A}}^{\text{rec}} = & \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} \left[|Q(\mathcal{A}, \mathbf{k}) + Q(\text{New}, \mathbf{k}) - Q(\text{Old}, \mathbf{k})|^2 \right. \\ & \left. - |Q(\mathcal{A}, \mathbf{k})|^2 \right] + \frac{\eta}{\sqrt{\pi}} \left[\sum_{i \in \text{New}} q_i^2 - \sum_{i \in \text{Old}} q_i^2 \right] \\ & + \left[\left(\sum_{i \in \mathcal{A}} q_i + \sum_{i \in \text{New}} q_i - \sum_{i \in \text{Old}} q_i \right)^2 - \left(\sum_{i \in \mathcal{A}} q_i \right)^2 \right] \mathcal{U}_{\zeta} \quad (344) \end{aligned}$$

$$\begin{aligned} \mathcal{U}_{\mathcal{A}, \mathcal{B}}^{\text{rec}} = & \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} \text{Re} \left[(Q(\mathcal{A}, \mathbf{k}) + Q(\text{New}, \mathbf{k}) \right. \\ & \left. - Q(\text{Old}, \mathbf{k})) \bar{Q}(\mathcal{B}, \mathbf{k}) - Q(\mathcal{A}, \mathbf{k}) \bar{Q}(\mathcal{B}, \mathbf{k}) \right] \\ & + \left[\left(\sum_{i \in \mathcal{B}} q_i \right) \left(\sum_{i \in \text{New}} q_i \right) - \left(\sum_{i \in \mathcal{B}} q_i \right) \left(\sum_{i \in \text{Old}} q_i \right) \right] \mathcal{U}_{\zeta} \quad (345) \end{aligned}$$

4.11. Rigid Frameworks

The effect of flexibility on adsorption in zeolites is known to be small.^[232] Significant speedups can be obtained by keeping the framework fixed and pre-computing the potential energy surface induced by the framework.^[233,234] Instead of looping over all framework atoms in order to compute the host-adsorbate energy at each time step, one can construct a 3D grid before the simulation and then obtain the energy by interpolation during the simulation. The more points in the grid the higher the accuracy. RASPA^[235] implements the triclinic grid interpolation scheme in

three dimensions of Lekien and Marsden.^[236,237] The algorithm is based on a specific 64×64 matrix that provides the relationship between the derivatives at the corners of the elements and the coefficients of the tricubic interpolant for this element. The cubic interpolant and its first three derivatives are continuous and consistent. The same grids can therefore be used for both MC and MD with no additional energy drift besides the drift due to the integration scheme, that is, the energy gradients are the *exact* derivatives of the energy at each point in the element.

For LJ potentials a separate grid needs to be generated for each σ parameter. However, for the real part of the Ewald summation a single grid generated with a probe charge of unity will suffice. The real part of the Ewald summation can be obtained by simple multiplication of the grid-value with the actual charge of the atom. Note that the grid depends on η , and hence the real-space grid depends on the number of used unit cells.

4.12. Bond-Dipoles in Periodic Systems

The MM2 and MM3 force fields are able to calculate the electrostatic interaction in non-charged polar molecules by using bond dipoles. The electrostatic interaction is then given by

$$U_{\mu} = \frac{\mu_i \mu_j}{\epsilon r_{ij}^3} (\cos \chi - 3 \cos \alpha_i \cos \alpha_j) \quad (346)$$

where χ is the angle between the two dipoles and α_i and α_j are the angles between the dipoles and the vector connecting them. There is little difference if properly parameterized. However, almost all force fields prefer the charge model because it is easier to parameterize. Amirjalayer et al. extended the MM3 force field and used a building block approach for the parameterization of IRMOF-1.^[238] We will derive here the expressions to integrate the host bond-dipole model with adsorbate models that contain atomic charges using the Ewald summation framework.

Consider bond-dipoles between pairs of atoms a and b. We denote \mathbf{r}_i^{ab} as the separation vector or bond vector of the two atoms in the bond-dipole i , r_i^{ab} is the bond distance, and $\boldsymbol{\mu}_i$ is the bond-dipole with predefined (fixed) magnitude μ_i

$$\mathbf{r}_i^{ab} = \mathbf{r}_i^b - \mathbf{r}_i^a \quad (347)$$

$$r_i^{ab} = \sqrt{\mathbf{r}_i^{ab} \cdot \mathbf{r}_i^{ab}} \quad (348)$$

$$\boldsymbol{\mu}_i = \mu_i \hat{\mathbf{r}}_i^{ab} \equiv \mu_i \frac{\mathbf{r}_i^{ab}}{r_i^{ab}} \quad (349)$$

$$\mathbf{r}_i = (1/2)(\mathbf{r}_i^a + \mathbf{r}_i^b) \quad (350)$$

The dipole has a direction along the normalized bond separation vector and is located at the center \mathbf{r}_i of the bond. Using $B_0 = 1/r$, $B_1 = 1/r^3$, $B_2 = 3/r^5$, $B_3 = 15/r^7$ we can write the energies for the charge-charge, charge-bond-dipole, and bond-dipole-bond-dipole as

$$U_{\text{ch-ch}} = \sum_{i=1}^{N_{\text{ch}}-1} \sum_{j=i+1}^{N_{\text{ch}}} B_0 q_i q_j \quad (351)$$

$$U_{\text{ch-bdp}} = \sum_{i=1}^{N_{\text{bd}}} \sum_{j=1}^{N_{\text{ch}}} -B_1 q_j (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij}) \quad (352)$$

$$U_{\text{bdp-bdp}} = \sum_{i=1}^{N_{\text{bd}}-1} \sum_{j=i+1}^{N_{\text{bd}}} [B_1 (\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) - B_2 (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})] \quad (353)$$

where N_{ch} is the number of charges, N_{bd} the number of bond-dipoles, $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ the distance between charges and/or centers of bond-dipoles, and q_j denote the charge of ion j . Note that although the energy expressions are the same for bond-dipoles and point dipoles, the expressions for forces and stress/pressure are not. For bond-dipoles, the forces are exerted on the two atoms of the bond-dipole.

The force expression for charge i and charge j is given by

$$f_i^{\alpha} = B_1 q_i q_j r_{ij}^{\alpha}, \quad \mathbf{f}_j = -\mathbf{f}_i \quad (354)$$

We derive the force expressions for bond-dipole i and charge j as

$$f_{i,a}^{\alpha} = -q_j \left[\frac{1}{2} (B_2 (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij}) r_{ij}^{\alpha} - B_1 \mu_{i,\alpha}) + B_1 \frac{\mu_i}{r_i^{ab}} r_{ij}^{\alpha} - B_1 \frac{(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})}{r_i^{ab} \mu_i} \mu_{i,\alpha} \right] \quad (355)$$

$$f_{i,b}^{\alpha} = -q_j \left[\frac{1}{2} (B_2 (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij}) r_{ij}^{\alpha} - B_1 \mu_{i,\alpha}) - B_1 \frac{\mu_i}{r_i^{ab}} r_{ij}^{\alpha} + B_1 \frac{(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})}{r_i^{ab} \mu_i} \mu_{i,\alpha} \right] \quad (356)$$

$$f_j^{\alpha} = q_j [B_2 (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij}) r_{ij}^{\alpha} - B_1 \mu_{i,\alpha}] \quad (357)$$

and the expression for bond-dipole i and bond-dipole j

$$f_{i,a}^{\alpha} = \left\{ \frac{1}{2} [B_3 (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) r_{ij}^{\alpha} - B_2 (\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) + (\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) \mu_{i,\alpha} + (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij}) \mu_{j,\alpha}] + [B_1 (\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) \frac{\mu_{i,\alpha}}{r_i^{ab} \mu_i} - B_2 (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) \frac{\mu_{i,\alpha}}{r_i^{ab} \mu_i} - B_1 \frac{\mu_i}{r_i^{ab}} \mu_{j,\alpha} + B_2 \frac{\mu_i}{r_i^{ab}} (\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) r_{ij}^{\alpha}] \right\} \quad (358)$$

$$f_{i,b}^{\alpha} = \left\{ \frac{1}{2} [B_3 (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) r_{ij}^{\alpha} - B_2 (\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) + (\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) \mu_{i,\alpha} + (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij}) \mu_{j,\alpha}] - [B_1 (\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) \frac{\mu_{i,\alpha}}{r_i^{ab} \mu_i} - B_2 (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) \frac{\mu_{i,\alpha}}{r_i^{ab} \mu_i} - B_1 \frac{\mu_i}{r_i^{ab}} \mu_{j,\alpha} + B_2 \frac{\mu_i}{r_i^{ab}} (\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) r_{ij}^{\alpha}] \right\} \quad (359)$$

$$f_{j,a}^\alpha = \left\{ -\frac{1}{2} \left[B_3(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})r_{ij}^\alpha - B_2(\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) \right. \right. \\ \left. \left. + (\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})\mu_{i,\alpha} + (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})\mu_{j,\alpha} \right] \right. \\ \left. + \left[B_1(\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) \frac{\mu_{j,\alpha}}{r_{ab}^j \mu_j} - B_2(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) \frac{\mu_{j,\alpha}}{r_{ab}^j \mu_j} \right. \right. \\ \left. \left. - B_1 \frac{\mu_j}{r_{ab}^j} \mu_{i,\alpha} + B_2 \frac{\mu_j}{r_{ab}^j} (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij}) r_{ij}^\alpha \right] \right\} \quad (360)$$

$$f_{j,b}^\alpha = \left\{ -\frac{1}{2} \left[B_3(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})r_{ij}^\alpha - B_2(\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) \right. \right. \\ \left. \left. + (\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})\mu_{i,\alpha} + (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})\mu_{j,\alpha} \right] \right. \\ \left. - \left[B_1(\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) \frac{\mu_{j,\alpha}}{r_{ab}^j \mu_j} - B_2(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij}) \frac{\mu_{j,\alpha}}{r_{ab}^j \mu_j} \right. \right. \\ \left. \left. - B_1 \frac{\mu_j}{r_{ab}^j} \mu_{i,\alpha} + B_2 \frac{\mu_j}{r_{ab}^j} (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij}) r_{ij}^\alpha \right] \right\} \quad (361)$$

The stress σ and pressure (related to the trace of the stress-tensor) can be obtained using the results for point-dipoles^[213] but corrected. The contribution for each charge-charge, charge-bond-dipole, and bond-dipole-bond-dipole interaction is

$$\sigma_{\alpha\beta}^{\text{ch-ch}} = \frac{1}{V} f_{i,b}^\alpha f_{j,\beta} \quad (362)$$

$$\sigma_{\alpha\beta}^{\text{bdp-ch}} = -\frac{1}{V} q_j \left[\frac{1}{2} (B_2(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})r_{ij}^\alpha - B_1\mu_{i,\alpha}) \right] r_{ij,\beta} \\ - f_{i,a}^\alpha (\mathbf{r}_{i,a}^\beta - \mathbf{r}_i^\beta) - f_{i,b}^\alpha (\mathbf{r}_{i,b}^\beta - \mathbf{r}_i^\beta) \quad (363)$$

$$\sigma_{\alpha\beta}^{\text{bdp-bdp}} = -\frac{1}{V} \left[B_3(\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})(\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})r_{ij}^\alpha - B_2(\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j) \right. \\ \left. + (\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij})\mu_{i,\alpha} + (\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij})\mu_{j,\alpha} \right] r_{ij,\beta} \\ - f_{i,a}^\alpha (\mathbf{r}_{i,a}^\beta - \mathbf{r}_i^\beta) - f_{i,b}^\alpha (\mathbf{r}_{i,b}^\beta - \mathbf{r}_i^\beta) \\ - f_{j,a}^\alpha (\mathbf{r}_{j,a}^\beta - \mathbf{r}_j^\beta) - f_{j,b}^\alpha (\mathbf{r}_{j,b}^\beta - \mathbf{r}_j^\beta) \quad (364)$$

where V is the volume and the corresponding force expressions are given above. This procedure is related to the conversion between atomic and molecular stress for rigid molecules. Usually this conversion produces a torque and an asymmetric stress which needs to be symmetrized.

The expressions given above still apply for the real part of the Ewald summation if the following expressions are used

$$B_0 = \frac{\text{erfc}(\eta r_{ij})}{r_{ij}} \quad (365)$$

$$B_1 = \frac{2\eta e^{-(\eta r_{ij})^2}}{\sqrt{\pi} r_{ij}^2} + \frac{\text{erfc}(\eta r_{ij})}{r_{ij}^3} \quad (366)$$

$$B_2 = \frac{6\eta e^{-(\eta r_{ij})^2}}{\sqrt{\pi} r_{ij}^4} + \frac{4\eta^3 e^{-(\eta r_{ij})^2}}{\sqrt{r_{ij}} r_{ij}^2} + \frac{3\text{erfc}(\eta r_{ij})}{r_{ij}^5} \quad (367)$$

$$B_3 = \frac{30\eta e^{-(\eta r_{ij})^2}}{\sqrt{\pi} r_{ij}^6} + \frac{20\eta^3 e^{-(\eta r_{ij})^2}}{\sqrt{r_{ij}} r_{ij}^4} + \frac{8\eta^5 e^{-(\eta r_{ij})^2}}{\sqrt{r_{ij}} r_{ij}^2} + \frac{15\text{erfc}(\eta r_{ij})}{r_{ij}^7} \quad (368)$$

where erfc is the error-function complement.

The Fourier part of the charge-charge, charge-bond-dipole, and bond-dipole-bond-dipole energies is

$$U_{\text{ch-ch}} = \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{e^{-k^2/4\eta^2}}{k^2} \left(\left| \sum_i q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 \right. \\ \left. + \left| \sum_i q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 \right) - \frac{\eta}{\sqrt{\pi}} \sum_i q_i^2 \quad (369)$$

$$U_{\text{ch-bdp}} = \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{e^{-k^2/4\eta^2}}{k^2} \left(\sum_i q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_j (\boldsymbol{\mu}_j \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_j) \right. \\ \left. - \sum_i q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_j (\boldsymbol{\mu}_j \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_j) \right) \quad (370)$$

$$U_{\text{bdp-bdp}} = \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{e^{-k^2/4\eta^2}}{k^2} \left(\left| \sum_i (\boldsymbol{\mu}_i \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 \right. \\ \left. + \left| \sum_i (\boldsymbol{\mu}_i \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 \right) - \sum_i \frac{2\eta^3 \mu_i^2}{3\sqrt{\pi}} \quad (371)$$

The force in direction α on a charge exerted by other charges is given by

$$f_{i,\alpha}^\alpha = q_i \frac{4\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{e^{-k^2/(4\eta^2)}}{k^2} k_\alpha \left(\sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \cos(\mathbf{k} \cdot \mathbf{r}_j) \right. \\ \left. - \cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \sin(\mathbf{k} \cdot \mathbf{r}_j) \right) \quad (372)$$

The force in direction α on a charge exerted by the bond-dipoles is given by

$$f_{i,\alpha}^\alpha = -q_i \frac{4\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{e^{-k^2/(4\eta^2)}}{k^2} k_\alpha \left(\cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_j (\boldsymbol{\mu}_j \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_j) \right. \\ \left. + \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_j (\boldsymbol{\mu}_j \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_j) \right) \quad (373)$$

The forces in direction α on atom a and b , respectively, of bond dipole i exerted by charges are given by

$$f_{i,\alpha}^a = \frac{4\pi}{V} \frac{e^{-k^2/(4\eta^2)}}{k^2} \left\{ \frac{1}{2} (\boldsymbol{\mu}_i \cdot \mathbf{k}) k_\alpha \left[\cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \cos(\mathbf{k} \cdot \mathbf{r}_j) + \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \sin(\mathbf{k} \cdot \mathbf{r}_j) \right] + \left(k_\alpha \frac{\mu_i}{r_i^{ab}} - (\boldsymbol{\mu}_i \cdot \mathbf{k}) \frac{\mu_{i,\alpha}}{\mu_i r_i^{ab}} \right) \times \left(\cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \sin(\mathbf{k} \cdot \mathbf{r}_j) - \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \cos(\mathbf{k} \cdot \mathbf{r}_j) \right) \right\} - \left(k_\alpha \frac{\mu_i}{r_i^{ab}} - (\boldsymbol{\mu}_i \cdot \mathbf{k}) \frac{\mu_{i,\alpha}}{\mu_i r_i^{ab}} \right) \times \left((\boldsymbol{\mu}_i \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^{nbd} (\boldsymbol{\mu}_j \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_j) - (\boldsymbol{\mu}_i \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^{nbd} (\boldsymbol{\mu}_j \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_j) \right) \quad (374)$$

$$f_{i,\alpha}^b = \frac{4\pi}{V} \frac{e^{-k^2/(4\eta^2)}}{k^2} \left\{ \frac{1}{2} (\boldsymbol{\mu}_i \cdot \mathbf{k}) k_\alpha \left[\cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \cos(\mathbf{k} \cdot \mathbf{r}_j) + \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \sin(\mathbf{k} \cdot \mathbf{r}_j) \right] - \left(k_\alpha \frac{\mu_i}{r_i^{ab}} - (\boldsymbol{\mu}_i \cdot \mathbf{k}) \frac{\mu_{i,\alpha}}{\mu_i r_i^{ab}} \right) \times \left(\cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \sin(\mathbf{k} \cdot \mathbf{r}_j) - \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_j \cos(\mathbf{k} \cdot \mathbf{r}_j) \right) \right\} \quad (375)$$

The forces in direction α on atom a and b , respectively, of bond dipole i exerted by other bond-dipoles are given by

$$f_{i,\alpha}^a = \frac{4\pi}{V} \frac{e^{-k^2/(4\eta^2)}}{k^2} \left\{ \frac{1}{2} k_\alpha (\boldsymbol{\mu}_i \cdot \mathbf{k}) \times \left((\boldsymbol{\mu}_i \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^{nbd} (\boldsymbol{\mu}_j \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_j) - (\boldsymbol{\mu}_i \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^{nbd} (\boldsymbol{\mu}_j \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_j) \right) + \left(k_\alpha \frac{\mu_i}{r_i^{ab}} - (\boldsymbol{\mu}_i \cdot \mathbf{k}) \frac{\mu_{i,\alpha}}{\mu_i r_i^{ab}} \right) \times \left((\boldsymbol{\mu}_i \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^{nbd} (\boldsymbol{\mu}_j \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_j) - (\boldsymbol{\mu}_i \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^{nbd} (\boldsymbol{\mu}_j \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_j) \right) \right\} \quad (376)$$

$$f_{i,\alpha}^b = \frac{4\pi}{V} \frac{e^{-k^2/(4\eta^2)}}{k^2} \left\{ \frac{1}{2} k_\alpha (\boldsymbol{\mu}_i \cdot \mathbf{k}) \times \left((\boldsymbol{\mu}_i \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^{nbd} (\boldsymbol{\mu}_j \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_j) - (\boldsymbol{\mu}_i \cdot \mathbf{k}) \cos(\mathbf{k} \cdot \mathbf{r}_i) \sum_{j=1}^{nbd} (\boldsymbol{\mu}_j \cdot \mathbf{k}) \sin(\mathbf{k} \cdot \mathbf{r}_j) \right) \right\}$$

In **Table 9** we have compiled numerical data on various systems with charges and/or bond-dipoles. The Ewald results differ from the results of a finite systems and the difference is due the infinite periodic images. For this box-size the results are close though and serve as a good check for the derived expressions. Some bond-dipole systems are chosen such that the results are identical in energy to the *point*-dipole systems discussed by Nyman and Linse.^[213] Note that the forces differ, albeit that they are related (the bond-dipole expressions contain a term identical to the one for point-dipoles).

4.13. Charge Equilibration Methods

Related to electrostatics is the problem of computing atomic charges in periodic systems. In the charge equilibration models, effective atomic charges are obtained by minimizing a potential energy function in which the adjustable parameters are atomic parameters.^[239] Ongari et al. recently evaluated various charge equilibration methods to generate electrostatic fields in nanoporous materials.^[240] The following section is based on the papers of Wilmer and Snurr.^[241,242] They begin by assuming a Taylor series form for the energy of an isolated (gas-phase) atom as a function of its charge, $E^A(q)$, centered about some particular partial charge q^*

$$E^A(q) = E^A(q^*) + \left(\frac{\partial E}{\partial q} \right)_{q=q^*} (q - q^*) + \frac{1}{2} \left(\frac{\partial^2 E}{\partial q^2} \right)_{q=q^*} (q - q^*)^2 + \dots \quad (378)$$

The original Q_{eq} method^[243] expands around $q^* = 0$ but when atoms are expected to have large positive charges, such as the transition metals in MOFs, the atomic energy given by a second-order Taylor series centered at $q^* = 0$ is unlikely to yield accurate results because the energy is based on an extrapolation far from the measured value. Significantly improved results are obtained for mono or divalent ions in metal-organic frameworks by Taylor expanding around its oxidation state. By adding and subtracting $E^A(q^* + 1)$ and $E^A(q^* - 1)$ from $E^A(q^*)$ expressions for the first and second partial derivatives can be obtained

$$\left(\frac{\partial E}{\partial q} \right)_{q=q^*} = \frac{1}{2} (I_{q^*+1} + I_{q^*}) \equiv \chi_{q^*} \quad (379)$$

$$\left(\frac{\partial^2 E}{\partial q^2} \right)_{q=q^*} = I_{q^*+1} - I_{q^*} \equiv J_{q^*} \quad (380)$$

Table 9. Calculated energy and forces for charge and bond-dipoles as obtained from the finite and infinite system (Ewald summation). Box size $L_x = L_y = L_z = 10$, reduced units: the unit of charge is $1e$, the unit of length is 1 \AA , and $4\pi\epsilon = 1$. The spherical cutoff is $L/2$. The Ewald results are corrected for non-zero net charge of the system using the method of Bogusz et al.,^[221] e.g. the uncorrected energy of a single $+1$ charge of the system is $\mathcal{U} = -0.1402549$, while the corrected energy is $\mathcal{U} = 0$. The definition of the bond-dipole is that it points from a to b ; toward the partially positively charged atom if the magnitude μ is positive.

charge i , $q_i = +1$, $r_i = (0, 0, 0)$			
Finite	$\mathcal{U}=0$	(0,0,0)	(0,0,0)
Ewald	$\mathcal{U}=0$	(0,0,0)	(0,0,0)
charge i , $q_i = +1$, $r_i = (-\frac{1}{2}, 0, 0)$, and charge j , $q_j = -1$, $r_j = (\frac{1}{2}, 0, 0)$			
Finite	$\mathcal{U}=-1.0$	(1,0,0,0)	(-1,0,0,0)
Ewald	$\mathcal{U}=-1.0021255$	(0.9956865,0,0)	(-0.9956865,0,0)
bond-dipole i , $\mu_i = +1$, $r_{i,a} = (-\frac{1}{2}, 0, 0)$, $r_{i,b} = (\frac{1}{2}, 0, 0)$			
Finite	$\mathcal{U}=0$	(0,0,0)	(0,0,0)
Ewald	$\mathcal{U}=-0.0020944$	(0,0,0)	(0,0,0)
charge i , $q_i = +1$, $r_i = (-\frac{1}{2}, 0, 0)$, and bond-dipole j , $\mu_j = 1$, $r_{j,a} = (0, 0, 0)$, $r_{j,b} = (1, 0, 0)$			
Finite	$\mathcal{U}=-1.0$	(2,0,0)	(-1,0,0)
Ewald	$\mathcal{U}=-0.9977809$	(2.0045637,0,0)	(-1.0022818,0,0)
charge i , $q_i = +1$, $r_i = (-\frac{1}{2}, 0, 0)$, and bond-dipole j , $\mu_j = 1$, $r_{j,a} = (\frac{1}{2}, -\frac{1}{2}, 0)$, $r_{j,b} = (\frac{1}{2}, \frac{1}{2}, 0)$			
Finite	$\mathcal{U}=0.0$	(0,-1,0)	(-1,-\frac{1}{2},0)
Ewald	$\mathcal{U}=-0.0020944$	(0,-0.9959987,0)	(-0.9956865,-0.4979993,0)
bond-dipole i : $\mu_i = +1$, $r_{i,a} = (-1, 0, 0)$, $r_{i,b} = (0, 0, 0)$; $\mu_j = +1$, $r_{j,a} = (0, 0, 0)$, $r_{j,b} = (1, 0, 0)$			
Finite	$\mathcal{U}=-2$	(3,0,0)	(3,0,0)
Ewald	$\mathcal{U}=-2.0087525$	(2.9996230,0,0)	(-2.9996230,0,0)
bond-dipole i : $\mu_i = +1$, $r_{i,a} = (-\frac{1}{2}, -\frac{1}{2}, 0)$, $r_{i,b} = (-\frac{1}{2}, \frac{1}{2}, 0)$; $\mu_j = +1$, $r_{j,a} = (\frac{1}{2}, -\frac{1}{2}, 0)$, $r_{j,b} = (\frac{1}{2}, \frac{1}{2}, 0)$			
Finite	$\mathcal{U} = 1.0$	(-1.5,0,0)	(-1.5,0,0)
Ewald	$\mathcal{U} = 0.9918097$	(-1.4998115,0,0)	(-1.4998115,0,0)

where χ_{q^*} is the *electronegativity* (also known as *Mulliken electronegativity* and J_{q^*} is the *idempotential* (also known as the atomic *hardness*). The atom energy can now be described solely as a function of its charge and measured parameters that are available in the literature for most of the periodic table (currently available up to $Z = 86$). This leads to a quadratic energy function for every atom based entirely on measured quantities

$$E^A(q) = E^A(q^*) + \chi_{q^*}(q - q^*) + \frac{1}{2}J_{q^*}(q - q^*)^2 \quad (381)$$

The energy of the entire system E^{sys} can be expressed as a sum of individual atomic energies and pairwise Coulombic interactions^[242]

$$\begin{aligned} E^{\text{sys}}(q_1, q_2, \dots, q_N) \\ = \sum_{k=1}^N \left(E_k^A(q_k^*) + \chi_{q_k^*}(q_k - q_k^*) + \frac{1}{2}J_{q_k^*}(q_k - q_k^*)^2 + E_k^{\text{Coul}} + E_k^{\text{Orb}} \right) \end{aligned} \quad (382)$$

where E_k^{Orb} is the orbital energy term for the k th atom (a pairwise damping term to prevent infinite charge separation when two atoms are brought arbitrarily close together), and E_k^{Coul} is the energy of the charge on the k th atom interacting with the charges of all other atoms in the system. For a non-periodic system, E_k^{Coul} can be computed using a direct sum, but for a periodic system the Ewald methodology is needed.

The charges q_1, q_2, \dots, q_N are determined by finding the minimum energy in Equation (382) in a noniterative manner (regardless of whether the system is periodic), subject to the constraint that

$$\sum_{k=1}^N q_k = q^{\text{tot}} \quad (383)$$

where q^{tot} is the net charge of the system (which must be zero in the periodic case). The general approach to minimizing a function of many variables subject to constraints is via Lagrange multipliers,

$$\nabla E^{\text{sys}} - \lambda \nabla \left(\sum_{k=1}^N q_k - q^{\text{tot}} \right) = 0 \quad (384)$$

where the gradient vector ∇ spans the N dimensional space defined by the partial charges. Matching the vector components of Equation (384), we get the following $N - 1$ equations:

$$\frac{\partial E^{\text{sys}}}{\partial q_1} = \frac{\partial E^{\text{sys}}}{\partial q_2} = \dots = \frac{\partial E^{\text{sys}}}{\partial q_N} \quad (385)$$

Combining Equations (385) and (383) provides N equations with N unknowns (the charges) which can be solved simultaneously, without iteration.^[242]

5. Molecular Mechanics

5.1. Energy Expansion

The potential energy surface \mathcal{U} of a (periodic) system can be Taylor-expanded around a configuration \mathbf{x} of the system^[162]

$$\mathcal{U}(\mathbf{x} + \delta\mathbf{x}) = \mathcal{U}(\mathbf{x}) + \mathbf{h}^\ddagger \delta\mathbf{x} + \frac{1}{2} \delta\mathbf{x}^\ddagger \mathcal{H} \delta\mathbf{x} + \dots \quad (386)$$

where $\mathbf{h} = (\frac{\partial \mathcal{U}}{\partial x_i^\alpha})$ is the gradient and $\mathcal{H} = \frac{\partial^2 \mathcal{U}}{\partial x_i^\alpha \partial x_j^\beta}$ is usually referred to as the Hessian matrix. The superscript \ddagger denotes the transpose of a vector or matrix. The real potential energy surfaces of common force fields are rarely harmonic, but still the expansion is usually truncated at the second order (i.e., harmonic analysis). This assumes the energy surface is at least locally quadratic and can iteratively be used to find the nearest local minimum. The steps through the energy landscape can therefore not be taken too large. If the potential energy surface truly were quadratic, a method like Newton–Raphson would be able to find the minimum energy configuration in a single step. At the minimum energy configuration the first derivatives are zero, and all harmonic information on the system is therefore described by the Hessian matrix. The generalized Hessian, having dimensions $(3N + 6) \times (3N + 6)$ with N being the number of atoms, is given as

$$\mathcal{H} = \begin{pmatrix} \mathcal{H}_{ij} = \frac{\partial^2 \mathcal{U}}{\partial r_i \partial r_j} & \mathcal{H}_{ie} = \frac{\partial^2 \mathcal{U}}{\partial r_i \partial \epsilon} \\ \mathcal{H}_{ei} = \frac{\partial^2 \mathcal{U}}{\partial \epsilon \partial r_i} & \mathcal{H}_{ee} = \frac{\partial^2 \mathcal{U}}{\partial \epsilon \partial \epsilon} \end{pmatrix} \quad (387)$$

with the force constant matrix \mathcal{H}_{ij} ($3N \times 3N$) and the Born term \mathcal{H}_{ee} (6×6) being the second-order derivative of the internal energy with respect to position and strain, respectively. The strain (and also stress tensor) is symmetric and can be simplified to a 6D vector using Voigt notation: $\epsilon = (\epsilon_{xx}, \epsilon_{yy}, \epsilon_{zz}, \epsilon_{yz}, \epsilon_{xz}, \epsilon_{xy}) = (\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6)$. The Born term accounts for distortions of the lattice, \mathcal{H}_{ie} and \mathcal{H}_{ei} are cross-terms. The 0 K elastic tensor is defined as the derivative of the stress with respect to the strain^[244,245] at zero gradient $\mathbf{h} = 0$ and can be expressed in terms of the generalized Hessian^[181]

$$C_{\alpha\beta\mu\nu} = -\frac{1}{V} \frac{\partial \sigma_{\alpha\beta}}{\partial \epsilon_{\mu\nu}} \Big|_{\mathbf{h}=0} = \frac{1}{V} \frac{\partial^2 \mathcal{U}}{\partial \epsilon_{\alpha\beta} \partial \epsilon_{\mu\nu}} \Big|_{\mathbf{h}=0} \quad (388)$$

$$= \frac{1}{V} \left[\underbrace{\frac{\partial^2 \mathcal{U}}{\partial \epsilon_{\alpha\beta} \partial \epsilon_{\mu\nu}}}_{\text{Born term}} - \underbrace{\frac{\partial^2 \mathcal{U}}{\partial \epsilon_{\alpha\beta} \partial r_{i\lambda}} (\mathcal{H}^{-1})_{i\lambda j \xi} \frac{\partial^2 \mathcal{U}}{\partial \epsilon_{\mu\nu} \partial r_{j\xi}}}_{\text{Relaxation term}} \right] \quad (389)$$

$$= \frac{1}{V} \left[\mathcal{H}_{ee} - \mathcal{H}_{ei} (\mathcal{H}_{ij})^{-1} \mathcal{H}_{ie} \right] \quad (390)$$

The relaxation term arises when more than one particle is present in the unit cell.^[181] When the system is strained the atoms need to relax relative to one another, because before and after the strain applied the system must be in a state of zero net force. Note

that the generalized Hessian \mathcal{H} encompasses the conventional Hessian \mathcal{H} and hence all information on the vibrational spectrum. In addition to the vibrational information, elastic constants also contain information on the gradients and second derivatives on the cell, and the cross terms of position gradient and cell gradients.

5.2. Normal Mode Analysis

Normal Mode Analysis (NMA) has become one of the standard techniques in the study of the dynamics of molecules.^[246] It is the study of motions of small amplitudes in harmonic potential wells by analytical means. Here, small means small enough that the harmonic approximation holds. Usually this means that harmonic vibrational analysis can provide a very good description of the system at low temperature. A NMA contains all timescales. However, disadvantages include limited motion around a single stable conformation and lack of anharmonic features that are small but sometimes important. Despite these obvious limitations, NMA has found widespread use.

The equation of motion of a molecule in a harmonic well is given by

$$\mathbf{M} \frac{d^2(\delta\mathbf{x})}{dt^2} = -\mathcal{H} \delta\mathbf{x} \quad (391)$$

where \mathbf{M} is a diagonal matrix that contains the masses of the atoms.^[87,246] Substitution of the general solution $\delta\mathbf{x} = \mathbf{A} \cos(\omega t + \phi)$ into the equation results in

$$\mathcal{H} \mathbf{A} = \omega^2 \mathbf{M} \mathbf{A} \quad (392)$$

To remove the dependence on the mass matrix on the right side, one can rearrange the equation as

$$(\mathbf{M}^{-1/2} \mathcal{H} \mathbf{M}^{-1/2}) (\mathbf{M}^{1/2} \mathbf{A}) = \omega^2 (\mathbf{M}^{1/2} \mathbf{A}) \quad (393)$$

and denoting the new quantities with new symbols we have

$$\mathcal{H}' \mathbf{A}' = \alpha \mathbf{A}' \quad (394)$$

Here, $\mathcal{H}' = \mathbf{M}^{-1/2} \mathcal{H} \mathbf{M}^{-1/2}$ is the mass-weighted Hessian matrix, $\mathbf{A}' = \mathbf{M}^{1/2} \mathbf{A}$ is the eigenvector of the mass-weighted Hessian matrix and needs to be unmass weighted for normal modes $\mathbf{A} = \mathbf{M}^{-1/2} \mathbf{A}'$, and $\alpha = \omega^2$. The mass-weighted Hessian is usually projected to remove the translation, rotation, and constraints.

The harmonic frequencies ν are related to the eigenvalues of the projected and mass-weighted Hessian by

$$\nu_i = \frac{\sqrt{\alpha_i}}{2\pi} \quad (395)$$

The normal mode i is given by $\mathbf{A}_i = \frac{A'_i}{\sqrt{m_i}}$. The eigenvectors give the direction and relative amplitude of each atomic displacement.

The value A_i/α_i is an arbitrary amplitude for displacement along normal mode i . If atoms are undergoing thermal fluctuations along each mode, the standard deviation of each atom is given by setting $A_i = \sqrt{2k_B T/m_n}$, where m_n is the atomic mass of atom n .^[87,246] For the generation of the trajectory of normal mode i using N frames, the following expression applies:

$$\delta \mathbf{x}_i(t) = \sqrt{\frac{2k_B T/M}{\alpha_i}} \hat{\mathbf{A}}_i \sin(2\pi t/N) \quad t = 0, \dots, N \quad (396)$$

The magnitude of the motion depends on temperature and is inversely proportional to its frequency. The largest contribution to the atomic displacement comes from the lowest frequency normal modes, whereas for high-frequency eigenvectors, only a few atoms contribute. As mentioned, the low-frequency modes determine large scale framework flexibility^[31,32] and are responsible for structural rearrangement as a function of external stimuli.^[20]

5.3. Optimization

One can differentiate the energy Taylor expansion Equation (386) with respect to $\delta \mathbf{x}$, set the result to zero and solve to obtain what is known as the Newton–Raphson step^[87]

$$\delta \mathbf{x} = -\mathcal{H}^{-1} \mathbf{h} \quad (397)$$

The Newton–Raphson step can also be expressed as a sum over the eigenvectors \mathbf{e}_i (called local principal modes) and eigenvalues λ_i of the Hessian matrix

$$\delta \mathbf{x} = - \sum_i \frac{(\mathbf{e}_i^T \mathbf{h})}{\lambda_i} \mathbf{e}_i \quad (398)$$

where $\mathbf{e}_i^T \mathbf{h}$ is the component of the gradient along the eigenmode \mathbf{e}_i . For zero eigenvalues, the corresponding step component is set to zero. A zero eigenvalue means that for a displacement in the direction of the eigenvector the energy does not change, while a positive and negative value mean an increase and decrease in energy, respectively. Therefore, a true minimum has all positive eigenvalues. A first-order saddle point has exactly one negative eigenvalue. The Newton–Raphson steps minimize along the eigenvectors with positive eigenvalues and maximize along eigenvectors with negative eigenvalues. Therefore, when the starting configuration is of the correct curvature (the desired number of negative eigenvalues) the Newton–Raphson step is a good one to take. In general, however, the step must be modified to obtain a structure of the desired curvature.^[248] A simple but very powerful modification is to use a shift parameter γ which shifts the value of the eigenvalues^[249]

$$\delta \mathbf{x} = - \sum_i \frac{\mathbf{e}_i^T \mathbf{h}}{\lambda_i - \gamma} \mathbf{e}_i \quad (399)$$

Simons et al.^[250] derived an equation to find the shift parameter

$$\gamma = - \sum_i \frac{(\mathbf{e}_i^T \mathbf{h})^2}{\gamma - \lambda_i} \quad (400)$$

which can be solved by iteration. Note that, besides minimization, the method can also be used to find saddle points.^[248] Taking a big step leads to an increased likelihood of moving out of the region where the Hessian was valid. Therefore it makes sense to set a maximum tolerance on the size of any calculated $\delta \mathbf{x}$ and scale down $\delta \mathbf{x}$ accordingly if this maximum is exceeded.^[250] The mode-following technique is also applicable to minimizations of large quantum mechanical systems. Here, the analytical evaluation of the Hessian might be computational and memory demanding, but efficient parallel distributed data algorithms have been developed.^[251]

An illustration, the minimization process using the generalized Hessian is shown in **Figure 3** for the MIL-53 with 0, 1, or 2 water molecules per unit cell. The initial configuration is the experimental crystal structure^[252] with the water at random positions. To find the lowest energy configuration the minimization was repeated 1000 times for the systems with water. Initially, there are many negative eigenvalues present and the number increases with the amount of water in the system. The number of minimization steps therefore depends on the amount and initial positions of the adsorbates. Slowly but steadily the number of negative eigenvalues is reduced to zero, although it is possible that the number increases before decreasing again. This usually corresponds to a substantial structural change (angle and shape change). As can be seen, once zero eigenvalues have been reached the convergence is very fast. The stopping criteria has been that each of the gradients on the particles and on the cell have to be smaller than 10^{-10} kJ mol⁻¹ Å². Initially the gradients are high and the volume increases. This behavior increases with water content, because the water is placed randomly inside the structure. During the minimization the structure relaxes and for no water or one water molecule per unit cell the volume stays close to the open-pore starting structure. However, when two water molecules are optimized, the structure collapsed to the narrow-pore structure.

The method can be of great utility in force field development and to assess the influence of adsorbates on the framework structure. It is able to provide the correct solution, that is, zero gradients and the desired number of negative eigenvalues, by construction, where other methods can give wrong solutions. The method can be applied to unit cell minimization of periodic systems such as metal-organic frameworks and zeolites, for vibrational analysis (IR spectra and mode analysis), for force field development and for the computation of elastic constants at zero Kelvin. In Subsection 6.9 we will discuss an application of this methodology: the fitting of MOF force fields on elastic constants.

6. Parameterization

6.1. Force Field Optimization of Biological Systems

As an example to see how a biological force field is optimized, let us discuss how the CHARMM force field is optimized as

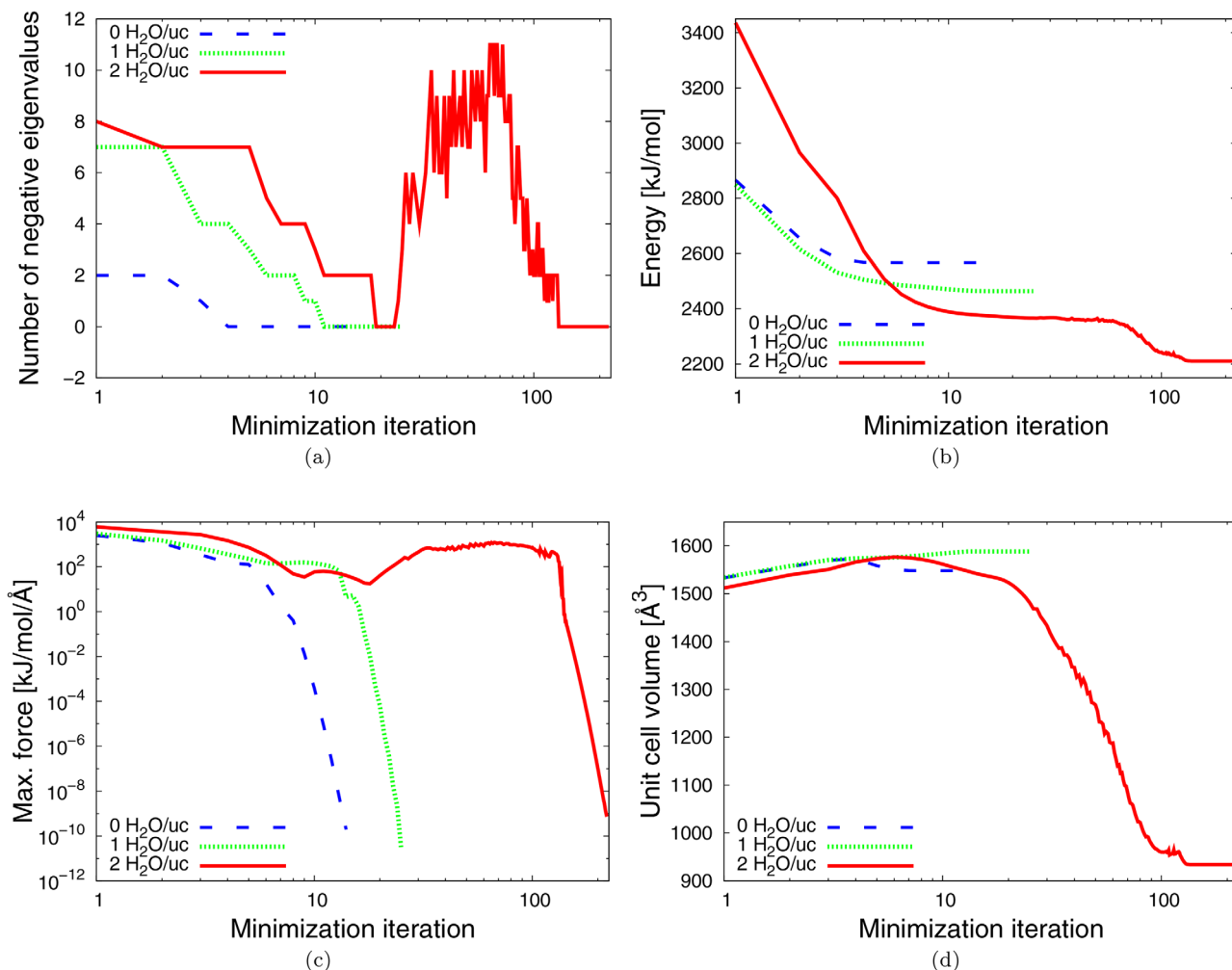


Figure 3. Mode-following minimization of MIL-53 using the model of Salles et al.^[247] with zero, one and two water molecules per unit cell: a) the number of negative eigenvalues, b) the convergence of the energy, c) the convergence of the maximum of the gradients on the particles (translation and rotational) and on the cell, and d) the volume of the unit cell. Reproduced with permission.^[163] Copyright 2009, American Chemical Society.

described in ref. [253]. The CHARMM potential energy function is^[118]

$$\begin{aligned}
 \mathcal{U} = & \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{UB}} K_{\text{UB}} (S - S_0)^2 + \sum_{\text{angle}} K_\theta (\theta - \theta_0)^2 \\
 & + \sum_{\text{dihedrals}} K_\phi (1 + \cos(n\phi - \delta)) + \sum_{\text{impropers}} K_{\text{imp}} (\chi - \chi_0)^2 \\
 & + \sum_{\text{nonbond}} \epsilon_{ij} \left[\left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_0 r_{ij}} \quad (401)
 \end{aligned}$$

Figure 4 is a flow diagram of the CHARMM parameter optimization procedure.^[253] Loops I, II, and III were included in the optimization of the CHARMM22 force field for nucleic acids, with loop IV representing an extension of that approach included in the CHARMM27 optimization.

In the CHARMM22 nucleic acid parameter optimization a variety of model compounds were selected with target data collected on those compounds. This target data included both experimental and ab initio data and solely acted as the basis for the parameter optimization. Empirical force field calculations were performed on the model compounds with the computed properties compared with the target data. The parameters were then manually adjusted to better reproduce the target data. Part of this process involved iterative procedures where, upon changing one class of parameters, a set of previously optimized parameters were re-adjusted if necessary (loops I, II, and III in Figure 4). For example, a set of partial atomic charges would be assigned to a model compound following which dihedral parameters would be adjusted to reproduce a target potential energy surface for that model compound. The partial atomic charges would then be re-investigated due to possible changes in geometry associated with optimization of the dihedral parameters that could effect the reproduction of the target data for the charge optimization. This approach yields a parameter set that accurately reproduces a variety

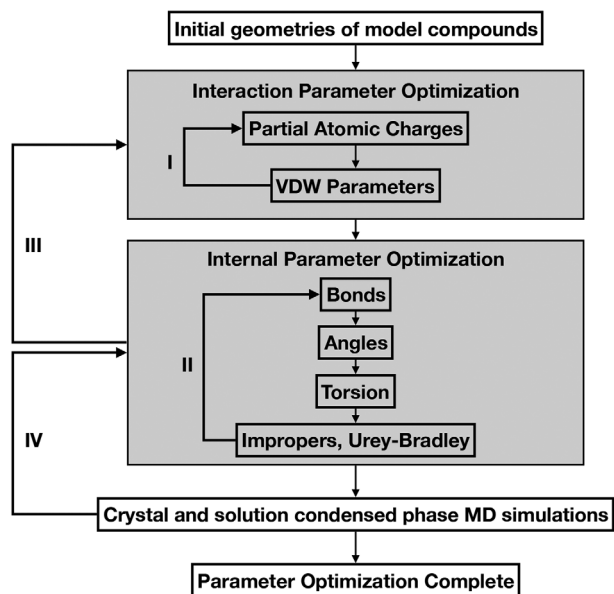


Figure 4. Flow diagram of the CHARMM parameter optimization. Iterative loops included in the parameterization are indicated by roman numerals. Adapted with permission.^[253] Copyright 2000, John Wiley & Sons, Inc.

of internal (e.g., geometries, vibrational spectra, conformational energetics) and interaction (e.g., interactions with water, heats, of sublimation) target data for the selected model compounds. Once the optimization procedure at the model compound level was complete, the resultant parameters were then used to perform simulations of B and Z DNA in their crystal environments, both of which yielded satisfactory agreement with experiment. At this point, the CHARMM22 parameterization was considered complete.

Re-optimization of the protein backbone parameters to systematically deviate from the QM energetic data led to improved properties for the protein backbone. This additional procedure is represented by loop IV in Figure 4. The need for this additional loop may reflect limitations associated with the level of theory of the QM data as well as the simplified form of the potential energy function in Equation (401), and emphasizes the importance of including macro-molecular properties as part of the target data for the parameter optimization procedure. The fitting procedure highlights that all the terms in Equation (401) influences each other. In iterative schemes, the long-range interactions are usually chosen first before optimizing the internal parameters. Recent developments and future trends in force fields for biomolecular simulations have been reviewed by Nerenberg en Head-Gordon.^[254]

6.2. Force Field Optimization for Adsorption and Diffusion in MOFs and Zeolites

In contrast to biological systems, MOFs and zeolites are fundamentally different and are usually considered a *host-guest* system. The host is the framework, the guests are small adsorbate molecules that are adsorbed inside the nanopores. The adsor-

bates experience confinement induced by the structure. Adsorption is driven by dispersion interactions and it is crucial to accurately describe the long-range interactions in the system. There is a natural separation of the system into a model for the adsorbates, a model for the adsorbate-framework interactions, and if the framework is taken as flexible, a model for the framework itself.

The long-range interactions are: 1) the Coulombic interactions, and 2) the van der Waals interactions. A common workflow is first to develop models for the adsorbates. The TraPPE models aim to have a high degree of accuracy in the prediction of thermophysical properties when applied to a range of different compounds, different state points, different compositions, and different properties. This makes TraPPE one of the few force fields generally suitable for materials and industrial applications. For nonbonded interactions, parameters for representative simple molecules are found by empirical fitting procedures that reproduce experimental phase equilibrium data, typically the vapor-liquid coexistence curve (VLCC) but also binary mixtures and vapor-solid equilibria. Fitting to phase equilibrium data provides accurate benchmarks against multiple phases, temperatures, and densities.

The advantage of adsorbate models that reproduce phase equilibrium data is that the saturation value of the adsorption isotherm is well-reproduced *by construction*. This is important, because it allows an examination of the state of the pores of the framework, that is, is there pore-blocking? are there remaining solvent or template molecules in the structure? The TraPPE model and many others use the approximation of fixed point charges. For small molecules, the charges follow from the dipole or quadrupole moment. However, this approximation immediately shows why water is difficult to model, it has a different dipole-moment in the gas-phase compared to the liquid-phase. Also, the bend-angle is altered in the adsorbed state. One approach to handle this is to make the potentials electric-field dependent.^[220] The framework-atoms are also usually modelled with the LJ+Coulombic approach. The next subsection focuses on how to obtain point charges. Once the charge is determined, the next step is to obtain the van der Waals parameters. We discuss two ways: the VLE fitting for the adsorbate-adsorbate interaction, and 2) fitting to inflection point in isotherms to fit the adsorbate-framework interactions. The next subsection deals with QM-derived potential for open-metal sites. Then, we discuss how to obtain models for the framework itself, and we close with a discussion on force fields for modeling mechanical properties.

6.3. Point Charges

Charge is not a quantum mechanical observable. Electrons are smeared out, and their charge is shared among nearby atoms. We are faced with the difficult problem of how to map the electron density onto classical atomic point charges.^[255] Many algorithms were defined over the past 60 years to divide molecules into atomic parts,^[256] for example Mulliken Population Analysis (MPA),^[257] Natural Population Analysis (NPA),^[258] Bader's AIM scheme,^[259] Lowdin Population Analysis (LPA),^[260] Hirshfeld partitioning,^[261] Hirshfeld-I partitioning.^[262] One approach is the concept of Atoms in Molecules (AIM)^[263] and then try to

introduce AIM subspaces within the molecule. The Hirshfeld-I has been extended to periodic bulk materials^[264] and Verstealen et al. proposed a simple Extended Hirshfeld (Hirshfeld-E or HE) scheme to solve an ambiguity in the Hirshfeld-I scheme when applied to inorganic oxide clusters.^[265] Manz and Sholl developed an AIM-based method called Density Derived Electrostatic and Chemical (DDEC) charges^[266–268] method. Here, the electron density is partitioned and assigned to each atom employing spherically symmetrical weighting functions^[269] with the partial charges optimized to simultaneously reproduce realistic chemical states, together with the QM electrostatic potential. This method gives accurate results for a variety of periodic and non-periodic materials including molecular systems, solid surfaces, porous solids, and nonporous solids.^[268]

A second class of methods is to fit the reproduced QM Electrostatic Potential (ESP):

$$\phi(\mathbf{r}) = \sum_{i=1}^N \frac{q_i}{|\mathbf{r} - \mathbf{R}_i|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' \quad (402)$$

This quantity can be used to fit point charges.^[270] Subsequently the method was refined by introducing least-squares fitting of point charges on a grid,^[271] ESP on spherical shells surrounding atom (CHELP), CHELP on a grid instead of spheres (ChelpG) which removed the problem with CHELP of not satisfying rotational invariance, and the RESP method (Restrained electrostatic potential fit) that applies restraints to partial charges on some atoms to remove the problem of erratic charges for buried atoms. In the REPEAT method,^[272] Campana et al. introduced a new error functional which acts on the relative differences of the potential and not on its absolute values, as it is currently done with molecular ESP charge derivation methods. Conventional ESP fitting procedures developed for molecular systems, in general, will not work for crystalline systems because the electrostatic potential in periodic systems is ill-defined up to a constant offset at each spatial position. The method produces sensible partial atomic charges for *periodic* systems. There is no need to extract chemically sensible and neutral moiety from the framework. Problems like capping/terminating the fragment to satisfy unfilled valences can be avoided. For many systems, minimal user intervention is required to generate the REPEAT charges, which is necessary for the automation of the charge derivation procedure. Such automation is very useful in high throughput computational studies of nanoporous materials. However, Watanabe et al. found for the ZIF-8 system that the REPEAT charges (especially for Zn) significantly depend on the details of the fitting-volume.^[267] The grid used in fitting the REPEAT charges should cover a volume similar to that which is available to an adsorbate molecule. It is therefore critical to exclude points near an atomic nucleus in the fitting of REPEAT charges.^[267,272] Wannabe et al.^[267] proposed to directly use the electrostatic potential energy calculated from plane wave DFT. This method removes an ambiguity that has been inherent in all previous models of these materials in which point charges are assigned to atoms in the adsorbent. By definition, the DDEC, Hirshfeld, and REPEAT charges sum over the unit cell to the correct total charge.

ESP methods have problems with buried atoms. Some alternative methods, which are not affected by the presence of

buried atoms, have been proposed. The dipole preserving and polarization consistent (DPPC) charges method,^[273] based on a Lagrangian optimization to reproduce atomic dipoles and multipoles,^[274] has been developed with the introduction of peculiar features, including charge redistribution based on relative atomic electronegativity and the treatment of molecular dipole polarization. Resulting charges are able to accurately reproduce the molecular dipole moments, thus ensuring a good description of the system local properties (which may be spoiled by basic ESP fitting). Two major improvements to the state-of-the-art REPEAT method, for generating accurate partial charges for molecular simulations of periodic structures, were developed by Gabrieli et al.^[275] The first, D-REPEAT, consists in the simultaneous fit of the ESP, together with the total dipole fluctuations (TDF) of the framework. The second, M-REPEAT, allows the fit of multiple ESP configurations at once. When both techniques are fused into one, DM-REPEAT method, the resulting charges become remarkably stable over a large set of fitting regions, giving a robust and physically sound solution to the buried atoms problem. The method capabilities were extensively studied in ZIF-8 framework, and subsequently applied to IRMOF-1 and ITQ-29 crystal structures.^[275] *InfiniCharges*^[276] is a tool for generating reliable partial charges for molecular simulations in periodic systems using the DM-REPEAT method. The stability of the resulting charges, over a large set of fitting regions, is obtained through the simultaneous fit of multiple electrostatic potential (ESP) configurations together with the total dipole fluctuations (TDF). The program can also perform standard REPEAT fit and its multi-frame extension (M-REPEAT), with the possibility to restrain the charges to an arbitrary value.

A third approach for the computation of effective charge is through charge equilibration models,^[277] as discussed in Subsection 4.13. Charge equilibration tries to find the set of partial atomic charges that minimize the energy of the system. The method of Rappe and Goddard required iteratively minimizing the system energy in a self-consistent manner.^[243] In contrast, the recently proposed method of Wilmer et al. requires no iteration,^[241,242] which makes it better suited for high throughput screening of materials and for calculating fluctuating charges on the fly throughout a molecular simulation. The method of Wilmer is applicable to MOFs: the Taylor series is expanded around the charged metal atom, and the ionization potentials and electron affinities corresponding to the cation are used. The charge equilibration is able to compute charges at only a tiny fraction of the computational cost of DFT methods. Charge equilibration methods have been evaluated for the generation of electrostatic fields in nanoporous materials.^[278]

Many charge sets have been published for frameworks like MOFs,^[279–281] COFs,^[282] ZIFs,^[283,284] and siliceous zeolites.^[285] The different methods available to calculate atomic partial charges in MOFs have been reviewed by Hamad et al.^[256] They demonstrated the influence of even small structural variations on atomic charges, and the influence of the choice of charges on computed properties. The decision about what method is the best is not a simple one, and the choice will depend on factors such as the knowledge and experience of the researcher, the codes that he or she has access to, the type of systems that will be studied, etc.^[256] A relatively new approach to charge assignment is Machine Learning.^[286–288] Also note that the force-match procedure,

Table 10. Selection of TraPPE models for molecules calibrated on experimental vapor-liquid equilibrium data. Cross terms are computed using Lorentz-Berthelot mixing rules.

Molecule type	All-atom	Type	ϵ/k_B [K]	σ [Å]	q [e]	Cutoff	Shifted
Alkanes ^[153]	no	CH ₄	148	3.73	–	14	no
	no	CH ₃	98	3.75	–	14	no
	no	CH ₂	46	3.95	–	14	no
Branched alkanes ^[154]	no	CH	10	4.68	–	14	no
	no	C	0.5	6.4	–	14	no
CO ₂ , bond-distance 1.16 Å ^[157]	yes	O	79.0	3.05	–0.35	10	no
	yes	C	27.0	2.80	–0.70	10	no
N ₂ , bond-distance 1.1 Å ^[157]	yes	N	36.0	3.31	–0.482	10	no
	–	Dummy	0.0	0.0	+0.964	10	no
CH ₄ ^[289]	no	CH ₄	158.5	3.72	0.0	12	yes

to be discussed in Subsection 6.7, naturally includes the fit of partial charges.

6.4. Vapor–Liquid Equilibrium Data

A computational very efficient model is the Transferable potentials for Phase Equilibria TraPPE force field by Martin and Siepmann.^[153,154] The force field describes linear, mono-branched and di-branched alkanes,^[153,154] benzene, pyridine, pyrimidine, pyrazine, pyridazine, thiophene, furan, pyrrole, thiazole, oxazole, isoxazole, imidazole, and pyrazole,^[298] primary, secondary, and tertiary amines, nitroalkanes and nitrobenzene, nitriles, amides, pyridine, and pyrimidine,^[299] ethers, glycols, ketones, and aldehydes,^[158] thiols, sulfides, disulfides, and thiophene,^[159] as well as some smaller molecule like CO₂ and N₂^[157] and ethane and ethylene.^[300] Despite the fact that the model lumps CH₃, CH₂, and CH into single interaction centers, it very accurately reproduces the experimental phase diagram and

critical points. This united atom approach allows for much longer simulation times and larger systems because each of the CH_x-groups is charge-neutral and charge-charge interaction can be omitted. Some TraPPE models for small molecules are listed in **Table 10**.

The TraPPE model is calibrated incrementally using Vapor–Liquid Equilibrium (VLE) data, for example, CH₄ using methane data, CH₃ using ethane data, CH₂ using propane data, CH using isobutane, and C using neopentane. This then constitutes the “calibration-set” (**Figure 5a**). With the final parameters in hand the remaining experimental data, for example, for heptane or isopentane, can then be used to validate the predictions of the model. If for this “validation set” unsatisfactory results are obtained and the experimental calibration data are not in doubt, then the model needs to be refined. A more sophisticated model could include explicit hydrogens^[155] or anisotropic sites.^[187,301] However, the united-atom TraPPE model for alkanes is able to accurately describe alkanes properties over a wide range of chain-lengths, densities, pressures, and temperatures (**Figure 5b**).

The TraPPE alkane fitting strategy results in a well depth that increases, and a diameter that decreases, with increasing number of hydrogens. At first it might be surprising that the LJ diameter of a methyl pseudo atom is smaller than that of a methylene group. However, it has been pointed out that the incremental volume of a methyl group remains much larger than the incremental volume of a methylene group as proposed by Bondi^[302] because of the large overlap between a methylene pseudoatom and its two neighbors.^[153,154] Furthermore, it is plausible that the diameter of a methylene pseudoatom containing two long C-C bonds and two (relatively) short C-H bonds might be larger than that of a pseudoatom containing a single long C-C bond and three short C-H bonds.^[153,154] The TraPPE-UA force field parameters for a given pseudo-atom do not depend upon its neighboring pseudo-atoms. While it could be argued that the parameters of pseudo-atoms should depend upon their nearest neighbors, such a force field for the alkanes would have 69 different types of pseudo-atoms.^[154]

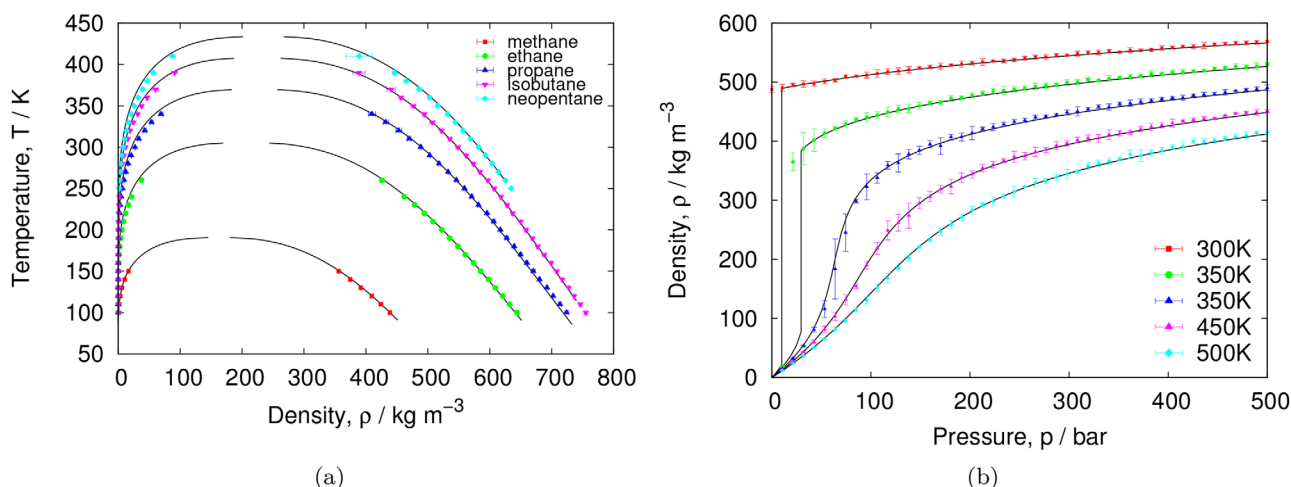


Figure 5. Alkane model based on VLE fitting (left) calibration of CH₄ using methane data, CH₃ using ethane data, CH₂ using propane data, CH using isobutane, and C using neopentane (right) density vs. pressure for a united-atom TraPPE-UA propane model. Experimental data from the NIST database. a) Reproduced with permission.^[90] Copyright 2013, The Authors, published by Taylor & Francis. b) Reproduced with permission.^[235] Copyright 2015, The Authors, published by Taylor & Francis.

Table 11. Selection of models by the Calero-group using shifted LJ potentials (cutoff 12 Å) for molecules calibrated on experimental vapor–liquid equilibrium data. These potentials can be used by both MC and MD. Cross terms are computed using Lorentz–Berthelot mixing rules.

Molecule type	All-atom	Type	ϵ/k_B [K]	σ [Å]	q [e]	Cutoff	Shifted
N ₂ , bond-distance 1.1 Å ^[290]	yes	N (N ₂)	38.298	3.306	−0.405	12	yes
(0.55 Å from N atom)	—	Dummy			0.810	12	yes
O ₂ , bond-distance 1.2 Å ^[290]	yes	O (O ₂)	53.023	3.045	−0.112	12	yes
(0.6 Å from O atom)	—	Dummy			0.224	12	yes
Ar ^[290]	yes	Ar	124.070	3.380	0.0	12	yes
CCl ₄ ^[290]	no	CCl ₄	519.730	5.140	0.0	12	yes
CO, bond-distance 1.128 Å ^[291]	yes	C	16.141	3.636	−0.2424	12	yes
(0.6443 Å from C atom)	yes	O	98.014	2.979	−0.2744	12	yes
	—	Dummy			0.5168	12	yes
CO ₂ , bond-distance 1.149 Å ^[292]	yes	O	85.671	3.017	−0.3256	12	yes
	yes	C	29.933	2.745	−0.6512	12	yes
SO ₂ , bond-distance 1.431 Å ^[293]	yes	S	189.353	3.41	0.402	12	yes
(bond angle of 119°)	yes	O	58.725	3.198	−0.201	12	yes
SF ₆ , bond-distance 1.565 Å ^[294]	yes	F	73.130	2.843	—	12	yes
	yes	S	—	—	—	12	yes
H ₂ S, bond-distance 1.34 Å ^[295]	yes	S	275	3.7	−0.32	12	yes
(bond angle of 92°)	yes	H	—	—	0.16	12	yes
Alkenes ^[296]	no	—	—	—	—	12	yes
propylene ^[297]	no	CH ₃	93.0	3.685	0.87	12	yes
	no	CH	51.0	4.0	0.87	12	yes
(Dummy-CH ₂ bond length 0.704 Å)	—	Dummy			−1.74	12	yes
	—	CH ₃	108.0	3.76	—	12	yes

Adjusting these force field parameters to VLE data is currently a cumbersome and computationally expensive task. For alkanes an incremental optimization (e.g. first CH₃ from ethane, then CH₂ from propane, etc) is possible due to the simplistic nature of alkanes, but this is an exception rather than the rule. As mentioned before, a force field optimization via the VLE of, for example, CO₂ involves optimizing the van der Waals interactions of the carbon and oxygen atom type simultaneously, therefore requiring many time-consuming iterations. An objective function can be defined as the squared deviation between the computed and the experimental data. Current fitting strategies involve lowering this function below a defined threshold using Simplex- or gradient-based methods. Each iteration point requires a full molecular simulation run. Van Westen et al. developed a method to overcome this limitation.^[303] The actual optimization is performed in a faster, different framework than molecular simulation, that is, using the PC-SAFT equation of state.^[304] The PC-SAFT is able to sufficiently capture the size and shape of molecules and the dispersion interactions of fluids to reduce the amount of iterations to only two or three.

The TraPPE model was mainly developed for MC purposes. For pure alkanes, the Radial Distribution Function (RDF) tends to converge to unity beyond 14 Å. This indicates that longer range order is absent. The Van der Waals potential is then truncated at this distance and the energy-correction due to this truncation is approximated (the “tail-correction”). The truncation distance and whether or not to use tail-correction should be considered

as part of the force field. For molecular dynamics the truncation in the energy leads to a divergence in the forces. Common approaches include the use of switching function where the energy is forced to smoothly go to zero, and to shift the Van der Waals potential to zero at the cutoff. An alternative is to use the Ewald summation to compute the Van der Waals term,^[305,306] but only a few codes have a that capability (an example is GULP^[307]). The Calero-group (re-)calibrated many small molecules model for a shifted Van der Waals potential and a cutoff of 12 Å (see **Table 11**) for use in both MC and MD. Compared to similar molecules in Table 10 we notice that the size parameter σ is slightly smaller and the strength parameter ϵ has become larger to make up for the lost attraction due to the shifting of the potential.

Subtle differences in energy and structure have been exploited to develop a force field for molecular simulation of martensitic phase transitions.^[308] Phase equilibria are extremely sensitive to potentials parameters of empirically determined classical force field. The reason the VDW parameters are so sensitive to the VLCC is that the molecule experiences two different environments, the low density and the high density, and the parameters need to fit both simultaneously. Likewise, inflections in isotherms are caused by the filling up of two (or more) adsorption sites that differ in energy. While VLE can fix the parameters for adsorbate–adsorbate interactions, the fitting on inflections in isotherms allows the accurate determination of adsorbate-framework parameters.

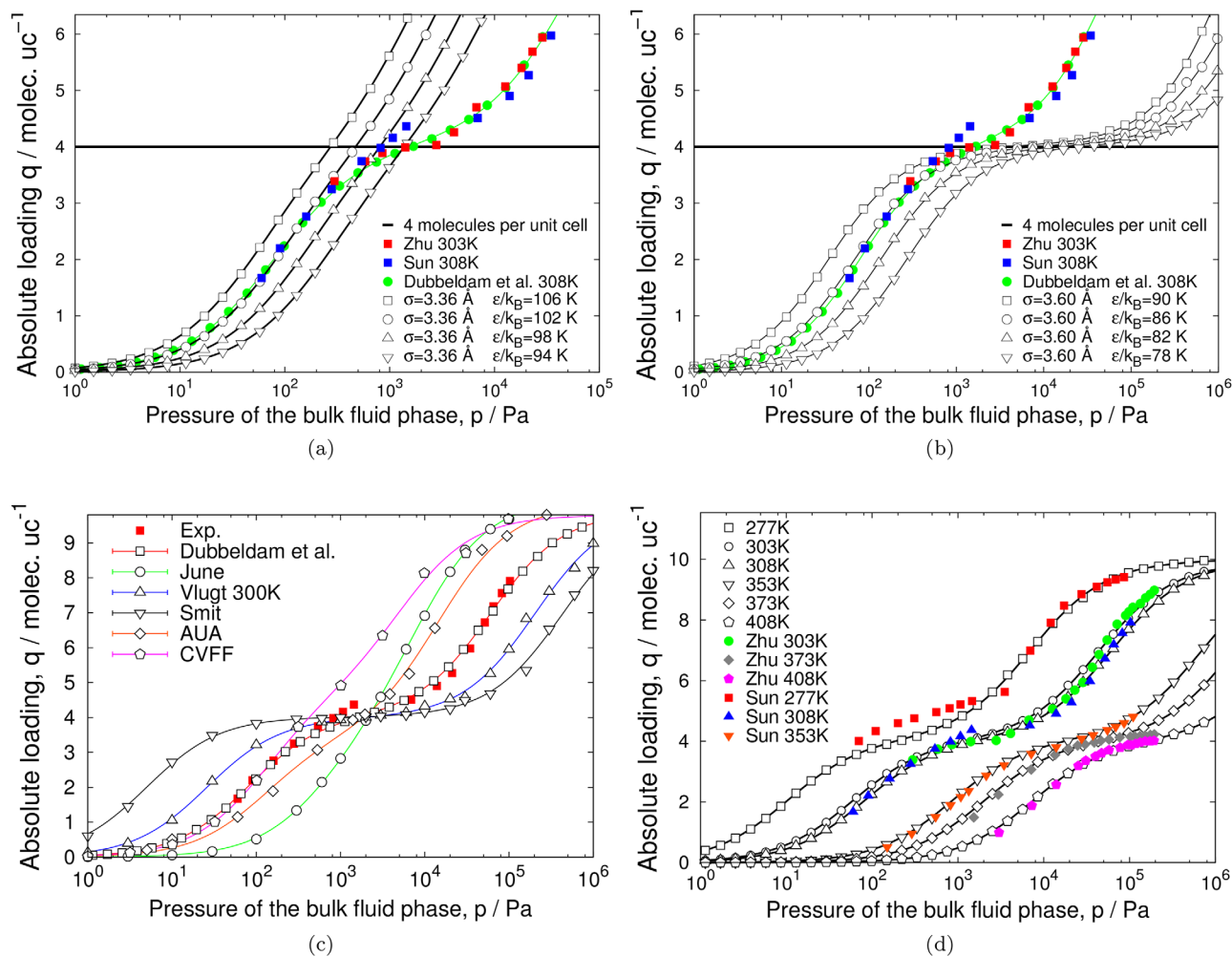


Figure 6. The calibration of isobutane in MFI zeolite. a) For a fixed σ that is too small the inflection is not reproduced, b) for a fixed σ that is too large the inflection is too large (for several different ϵ). c) Only a single ϵ , σ -pair is able to reproduce the experimental isotherm. Other models are not able to model the inflection of isobutane in MFI accurately. d) The resulting model predicts the experimental isobutane isotherms for two sources at various temperatures accurately. Reproduced with permission.^[309] Copyright 2004, American Chemical Society.

6.5. Fitting on Inflections in Adsorption Isotherms

Early work on alkanes in zeolites used the Kiselev-model where adsorbate-framework interactions were optimized to reproduce experimental Henry coefficient and heats of adsorption.^[310] The Kiselev approach models the zeolite as a rigid network of oxygen atoms.^[12] This is a very common approximation because the large oxygen atoms account for most of the potential energy and essentially shield the much smaller silicon atoms. The oxygen-adsorbate potential is an effective potential which implicitly includes the Si-contribution. The advantage of such a simplicity is that when combined with a united atom model, one can *uniquely* optimize an alkane-zeolite model (e.g., the obtained parameters are the best ones possible). Fitting on Henry coefficients and heats of adsorption does not uniquely fix the LJ-parameters. The methodology proposed by Dubbeldam et al. is based on fitting isotherms with inflection points.^[309,311] These inflection points are due to adsorption sites which differ significantly in energy. One site is predominantly filled up first, before the other sites fill

up (as a function of pressure). The inflection in the isotherm reflects this subtle difference in adsorption energetics of the sites. Inflections are extremely sensitive to the van der Waals parameters which is the dominating mechanism for physisorption.

It is instructive to discuss the role of the size-parameter $\sigma_{\text{O-CH}_x}$. In **Figure 6** we show the influence of the σ parameters on the inflection of 2-methylpropane in MFI. The O-CH parameters remain fixed at $\sigma = 3.92 \text{ \AA}$ and $\epsilon/k_B = 40 \text{ K}$, while $\epsilon_{\text{O-CH}_3}$ is examined over a range of reasonable values for two values of $\sigma_{\text{O-CH}_3}$: one significantly too small and one significantly too large. A crucial observation is that only a single strength/size parameter pair is able to describe the inflection and the entire isotherm properly. This is in contrast with the common belief that for each value of σ there is a corresponding ϵ that can describe the isotherm correctly.^[312] The shape of the isotherm and the inflection points are the most sensitive to the size-parameter of the interactions, whereas the loading at a given pressure is most sensitive to the strength-parameter of the interaction. A higher strength parameter ϵ induces an increased loading, and a lower strength

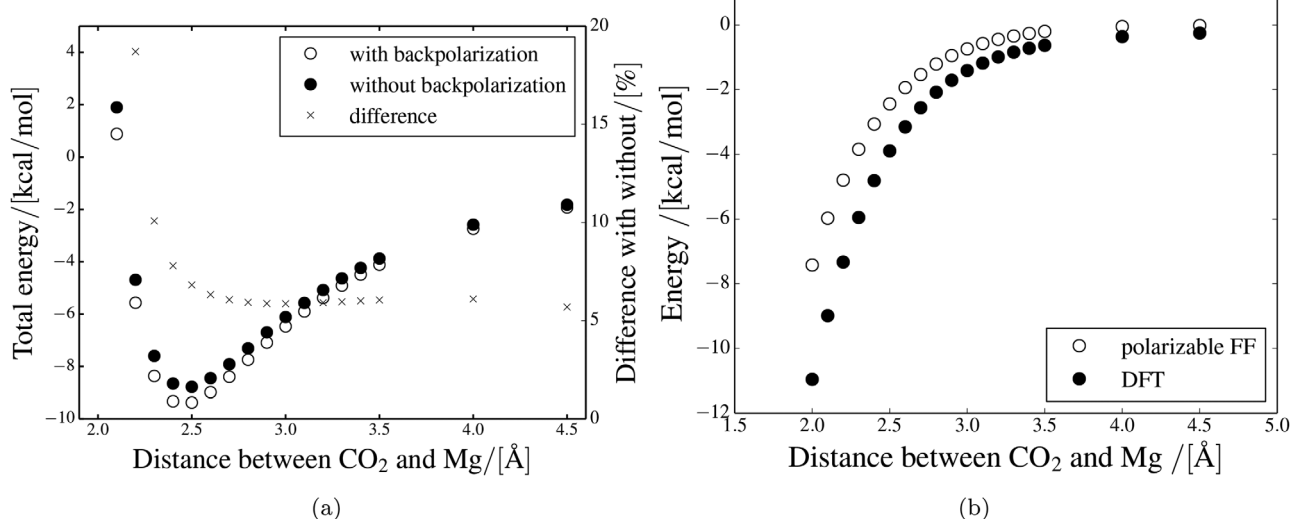


Figure 7. a) Total energy of a single CO₂ molecule in Mg-MOF-74 calculated using the developed polarizable force field as a function of the distance to the open-metal site. Comparison between interactions with and without back-polarization (units left axis), and the difference between with and without back-polarization (units right axis) b) Comparison of the polarization energy computed with the developed polarizable force field without considering back-polarization and the orbital interaction energy from DFT calculations as a function of the distance between a CO₂ molecule and the Mg ion.^[313] Reproduced with permission.^[313] Copyright 2017, American Chemical Society.

parameter results in a decrease in loading (for a fixed pressure). The amount of inflection is controlled by the size parameter σ . These properties can be exploited to obtain unique parameters.

In practice, one proceeds as follows. A reasonable starting size parameter is chosen. For this parameter one iteratively searches for the corresponding strength parameter that matches the experimental data at a pressure significantly below the inflection. The entire isotherm is then followed for increasing pressure until a deviation from the experimental data is observed. The updated size parameter is then found by choosing a higher value for a deviation to the left of the experimental data, and by choosing a lower value for the size parameter for a deviation to the right of the experimental data. This scheme proceeds iteratively until the entire experimental isotherm is accounted for.

The fitting to well-established inflection points in the isotherms has many advantages^[309,311]: i) a unique set of parameters is obtained directly related to a well defined physical property, ii) the parameters are determined accurately, iii) inflections are found at moderate pressures and here the experimental data is most reliable, iv) the inflection is directly related to the structure (e.g., for *n*-heptane and 2-methylpropane in MFI the inflection occurs exactly at four molecules per unit cell) and v) by explicitly fitting to full adsorption isotherms the proper reproduction of properties such as Henry coefficients, heats of adsorption, adsorption entropies, and maximum loadings is guaranteed.

6.6. Polarizable and QM-Derived Force Fields for Open-Metal Sites

When simple force fields fail to accurately describe experimental data, physically based next-generation force fields that efficiently

describe effects like polarization needs to be adapted.^[314] Mg-MOF-74 has been shown to be a promising candidate for carbon capture due to its high CO₂ uptake capacity at low partial pressures. Despite the significant progress, it is still a major challenge to accurately capture the change of interaction strength with varying metal ions in M-MOF-74 in molecular simulations. One approach is to use ab initio derived force fields,^[315,316] another to include explicitly polarization effects. Guest molecules are polarized in the vicinity of the open-metal sites in M-MOF-74 and this interaction contributes to the enhanced CO₂ affinity. Since this is *local* effect it is not possible to simply re-parametrized the guest-host interactions. Considering polarization explicitly can help to create force fields that overcome the shortcomings of current generic force fields.^[313,317–320]

Lachet et al. explored polarization in MC simulations using solely polarization between the framework and adsorbate molecules and neglecting polarization caused by induced dipoles, so-called back-polarization.^[219] Thereby, an iterative scheme is avoided and the computational costs of the method are drastically reduced. In fact, the computational costs can be similar to simulations without considering explicit polarization. **Figure 7a** shows the most favorable position for CO₂ in Mg-MOF-74 at approximately 2.4 Å, the difference in total energy is approximately 7%. This deviation seems to be acceptable in comparison with the considerable speedup of the simulations.

Note that one can not just add polarization on top of an existing force field. Most likely, the VDW parameters are fitted to experimental data, and hence implicitly contain some sort of average polarization. Becker et al. therefore removed the contribution of implicitly considered polarization to the interaction potential, by applying a global scaling parameter to all Lennard–Jones energy parameters developed without explicit polarization.^[313] In **Figure 7b**, we compare the polarization energy of a CO₂ molecule

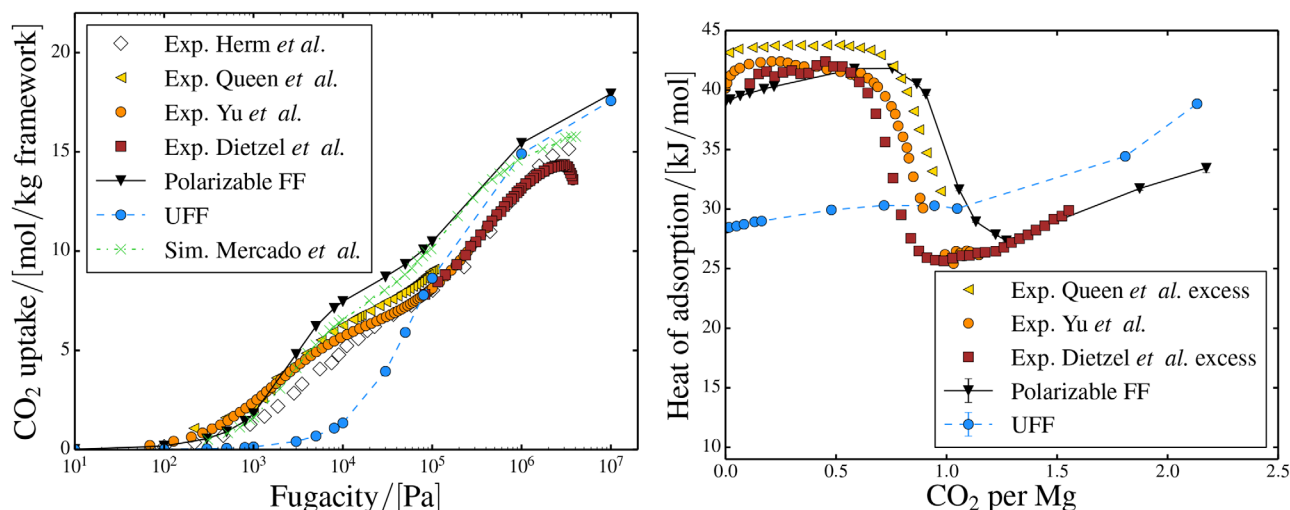


Figure 8. Comparison between the experimental results of Herm *et al.*^[321] (open), Queen *et al.*^[322] (yellow), Yu *et al.*^[323] (orange), and Dietzel *et al.*^[324] (brown) and simulation results using the polarizable force field of Becker *et al.*^[313] (black), the UFF force field (blue) and the DFT-derived non-polarizable force field of Mercado *et al.*^[325] (green) for CO₂ in Mg-MOF-74. a) adsorption isotherm at 298 K (Herm *et al.*^[321] 313K) (b) heat of adsorption as a function of uptake. Reproduced with permission.^[313] Copyright 2017, American Chemical Society.

approaching the Mg ion of Mg-MOF-74 estimated with the developed polarizable force field to the orbital interaction energy calculated from ADF. For relevant distances, both methods show a comparable trend for the energy contributions. The most relevant distance between the CO₂ molecule and the Mg ion is where the total energy is the lowest (i.e., 2.3–2.5 Å, as shown in Figure 7a). At this distance the polarizable force field predicts that the polarization energy of a single CO₂ molecule in Mg-MOF-74 has a significant contribution of around 30% to the total energy. For larger distances the contribution of polarization decreases rapidly.

The resulting adsorption isotherms for CO₂ in Mg-MOF-74 in comparison to experimental measurements, the UFF force field, and the DFT-derived non-polarizable force field of Mercado *et al.*^[325] are shown in Figure 8a. The simulation results with the polarizable force field clearly display the inflection of the experimental adsorption isotherm. The predicted behavior is significantly better than with the UFF force field. This is expected, because the scaling factors are adjusted to reproduce the experimental data. The overall agreement with the experimental measurements is comparable with the DFT-derived non-polarizable force field of Mercado *et al.*^[325] Both force fields can predict the low fugacity region which is particularly important for carbon capture. For higher fugacities, simulations with all compared force fields predict higher CO₂ uptakes in comparison to the experiments. This can be attributed to the fact that a certain degree of inaccessibility due to diffusion limitation or defects in the crystal structure is inherent with experimental structures. It should be noted that Mercado *et al.*^[325] scaled the calculated CO₂ uptakes with 0.85 to account for inaccessibility of open-metal sites. This scaling procedure mainly improves the agreement between experiments and computations for the high fugacity region. Figure 8b shows the heat of adsorption as a function of CO₂ molecules per metal ion. The distinct inflection of the adsorption isotherm caused by the strong affinity of the CO₂ molecule toward the metal ions is reflected by the change of

the heat of adsorption with increasing gas uptake. The calculated heat of adsorption has an inflection around one CO₂ per metal ion. Before and after the rapid decrease at one CO₂ molecule per metal ion, the heat of adsorption increases slightly. This increase can be related to a rise in the total number of adsorbed CO₂ and therefore a larger contribution of the CO₂–CO₂ interactions to the total energy.

6.7. Fitting of Force Field Parameters against Ab Initio Generated PES Data

Fitting algorithms are generally based on optimizing a “merit function” that corresponds to minimizing a weighted sum of square deviations between the classical data and reference data points from the QM PES.^[76,113] The key quantity is the “sum of squares” F which measures how good your fit is

$$F = \sum_{\text{all observables}} w(f_{\text{calc}} - f_{\text{obs}})^2 \quad (403)$$

where f_{calc} and f_{obs} are the calculated and observed quantities and w is a weighting factor. The aim of a fit is to minimize the sum of squares by varying the potential parameters. Also “Force Matching” can be done^[326] and even a more generalized match on energy E , forces \mathbf{F} , and stresses \mathbf{S}

$$\xi_1 = \sqrt{(E_{\text{dft}} - E_{\text{md}})^2} \quad (404)$$

$$\xi_2 = \sqrt{(\mathbf{F}_{\text{dft}} - \mathbf{F}_{\text{md}})^2} \quad (405)$$

$$\xi_3 = \sqrt{(\mathbf{S}_{\text{dft}} - \mathbf{S}_{\text{md}})^2} \quad (406)$$

$$F = \min (w_1 \xi_1 + w_2 \xi_2 + w_3 \xi_3) \quad \sum_i w_i = 1 \quad (407)$$

The weight factors are usually in the order: forces \gg stresses $>$ energies.

There are several codes able to do the parameterization for several classical force fields. GULP^[162] allows fitting on many observables like elastic constants, stresses, heat capacity, entropy, bond lengths and angles, relative energies, and on energy surfaces. To fit an energy surface it is basically necessary to input all the structures and the energies that correspond to them. *ForceFit* is a code to fit classical force fields to quantum mechanical potential energy surfaces.^[327] ForceFit interfaces with three molecular mechanics codes, DL_POLY, Amber, and LAMMPS. *Potfit* (<https://github.com/potfit>) is a free implementation of the force-matching algorithm to generate effective potentials from ab initio reference data.^[328] *DFTFIT* (<https://github.com/costrouc/dffit>) is a python code that uses ab initio data from DFT calculations such as VASP, Quantum Espresso, and Siesta to develop molecular dynamic potentials. DFTFIT uses the least square method to fit the stresses, total energy, and forces of a given set of configurations. The user has the flexibility of changing the weights for fitting and uses relative error to judge the fitness of parameters. *Atomicrex* is a versatile tool for the construction of advanced atomistic models.^[329] In a general sense, it allows one to develop models that describe a given scalar or vectorial property (e.g., total energies and forces) as a function of an atomic (or atom-like) configuration. *GAFit* is a software package based on a genetic algorithm that fits an analytic function to a given set of data points.^[330] The code was interfaced with the CHARMM and MOPAC programs in order to facilitate force field parameterizations. *GARFfield* (Genetic Algorithm based Reactive Force Field optimizer) is a multi-platform, multi-objective parallel hybrid genetic algorithm (GA)/conjugate-gradient (CG) based force field optimization framework.^[331] GARFfield currently supports a range of force field engines, via the LAMMPS Parallel Molecular Dynamics Simulator (including the adiabatic ReaxFF). *MEAM-fit* is an interatomic potential fitting code capable of fitting embedded atom method (EAM) and reference-free modified embedded atom method (RF-MEAM) potentials to energies and/or atomic forces produced by the VASP density functional theory (DFT) package.^[332] As a last example, we would like to mention machine learning interpolation of atomic potential energy surfaces to enable the nearly automatic construction of highly accurate atomic interaction potentials.^[333]

6.8. Genetic Algorithms

A common fitting approach is to use Genetic Algorithms (GA).^[334] A genetic algorithm has been used to generate plausible crystal structures from the knowledge of only the unit cell dimensions and constituent elements. Woodley et al. successfully generated 38 known binary oxides and various known ternary oxides with the Perovskite, Pyrochlore, and Spinel structures, from starting configurations which include no knowledge of the atomic arrangement in the unit cell.^[335]

There have been a number of approaches to parameterize a force field directly from the quantum chemical calculations (see

ref. [336] and references therein). Since the true potential energy surface (PES) can be approximated using quantum mechanical methods, a force field can be directly fit to a calculated QM PES by numerically matching the gradients or energy. Sufficiently accurate description of the PES enables fitting a variety of terms in the force field including harmonic bonding constants, as well as anharmonic and coupling terms, using, for example, a freely available code like “ForceFit.”^[327] Note that it is very hard, if not impossible, to obtain potentials from, for example, rotational energy curves. These profiles are the result of a complicated interaction of many potentials and to separate these special methods were developed. Although it is theoretically possible to include non-bonded interactions in the fitting, it is more common to obtain charges and van der Waals parameters separately and use these as input. Tafipolsky et al. parameterized a force field from first principles reference data by optimizing a novel objective function with a GA.^[334] The objective function was based on a representation of both structure and Hessian matrix in a redundant internal coordinate set, which can in principle be defined automatically by the connectivity of the molecular mechanics force field. The van der Waals parameters and electrostatics were given as input, as well as the functional forms of the bonded terms and a quantum-optimized representative cluster. The GA algorithm then is able to efficiently parameterize the bonded terms, which reproduce as close as possible the DFT results (structure and vibrational modes). It can be used in combination with any quantum-mechanical package that is able to provide the optimized structure and the Hessian matrix in Cartesian coordinates. The evaluation of the fitness of an individual parameter set comprises several basic steps:

1. MM geometry optimization of the reference structure based on the selected functional form (Tafipolsky et al. use MM3) and a trial set of parameters
2. Calculation of the Hessian matrix in Cartesian coordinates
3. Transformation of both the Cartesian coordinates and the Hessian matrix into the principal axes system
4. Transformation of the Hessian matrix from Cartesian to a pre-defined set of redundant internal coordinates
5. Calculation of the sum of mean square deviations between a trial solution and the reference data

The stopping criteria for the GA can be set up by separate convergence thresholds for the metric parameters (bonds, angles) and Hessian terms or by a fixed number of generations to be evolved.

6.9. Elastic Constants

Gale and coworkers used the elastic constant tensor for the optimization of core-shell zeolite models.^[162] However, only a hand-full of parameters need to be optimized and there is little ambiguity in the functional forms of the model. MOFs are chemically much more diverse, but fortunately there are many generic force fields available for the linker-molecules. The difficult part is to connect the linkers together via the hinges while making sure the model reproduces the correct mechanical behavior. Heinen et al.^[337] argue that reproducing the flexible

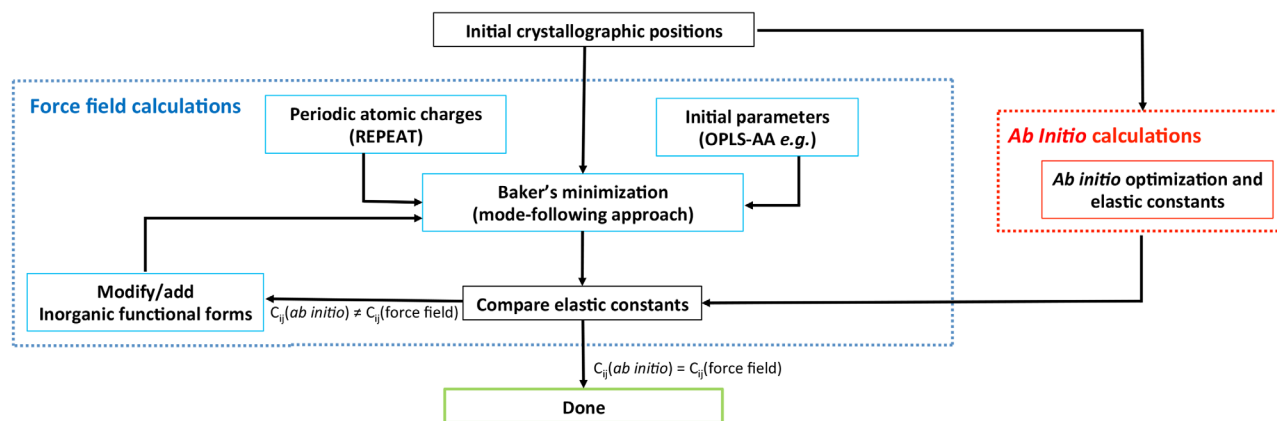


Figure 9. Elastic tensor-based force field parameterization approach: functional forms and corresponding parameters are changed and/or added until the classical force field elastic tensor agrees with the ab initio derived elastic tensor. Reproduced with permission.^[337] Copyright 2017, American Chemical Society.

behavior of the hinges in MOFs can be done by parameterizing the flexible force field on the elastic tensor. The elastic constant can be obtained from experiment when available, but otherwise can be computed by ab initio calculations. The ab initio computation of the elastic tensor for various MOFs have been performed by Tan et al.^[32,338,339] and Coudert et al.^[340,341] Understanding mechanical properties is crucial as it has been argued that the elastic tensor can predict flexible behavior.^[338]

The large-scale flexibility and low frequency modes in MOFs revolves around the behaviour of metal-clusters that are thought of as “hinges” acting between rigid “struts.”^[2] Heinen et al. propose to fit the parameters of the hinges of MOFs on DFT derived elastic constants.^[20,337] It is the hinges that determine the mechanical behavior and modeling these correctly is *vital*. The fourth-rank elastic tensor C is related to the second-rank stress σ and strain ϵ tensors via Hooke’s law^[245]

$$\sigma_{\alpha\beta} = \sum_{\mu\nu} C_{\alpha\beta\mu\nu} \epsilon_{\mu\nu} \quad (408)$$

with Greek indices denoting Cartesian components. For an orthorhombic unit cell, there are nine independent elastic constants when considering Voigt symmetry as shown below.

$$C_{ij} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{22} & C_{23} & 0 & 0 & 0 \\ C_{13} & C_{23} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix} \quad (409)$$

The first three diagonal components of the elastic tensor in Voigt notation C_{11} , C_{22} , and C_{33} represent the stiffness along the principal crystal axes. The other diagonal components C_{44} , C_{55} , and C_{66} are the shear coefficients that determine the resistance against angular deformation due to shear strain. The off-diagonal

components C_{12} , C_{13} , and C_{23} are the tensile-coupling interactions between two principal axes.

The proposed fitting procedure of the hinges is shown in **Figure 9**. First, target elastic constants are computed from DFT. This is non-trivial by itself as the energy-landscape around the hinges can be shallow. Very high precision computations are needed, and since second derivatives are usually computed numerically in DFT using finite differences from the forces, the dependence on the step-length of the differentiation needs to be investigated. Next, the parameters of hinges can be initialized with reasonable values based on previous work, other force fields, or intuition. The structure is minimized and the elastic constants are examined. By examining the values to the target values a new set of parameters is chosen. By applying an iterative process, the final set is obtained.

6.10. Machine Learning (ML)

A few examples of the use of machine learning in the field of classical simulations for materials science are the deep neural network learning of complex binary sorption equilibria from molecular simulation data,^[342] solving vapor–liquid flash problems using artificial neural networks,^[343] predicting thermodynamic properties of alkanes,^[344] charge assignment,^[286–288] prediction of partition functions,^[345] simulation of infrared spectra,^[346] predicting the mechanical properties of zeolite frameworks,^[347] CO₂ capture using MOFs,^[348,349] prediction of methane adsorption performance of MOFs,^[350,351] chemically intuited, large-scale screening of MOFs,^[352] screening for precombustion carbon capture using MOFs,^[353] screening of MOF Membranes for the separation of gas mixtures,^[354] and screening of MOFs for use as electronic devices.^[355] Using ML to combine the accuracy and flexibility of electronic structure calculations with the speed of classical potentials is a very active research field.^[356–358] The predictive accuracy of Machine Learning (ML) models of molecular properties depends on the choice of the molecular representation.^[359] Machine learning for atomistic force fields can be used in two different ways

- to find parameters for a force field with specified functional forms (for example, see ref. [360]),
- to replace the classical or QM force field altogether.

In the second approach the vectorial force on an atom is computed directly from its environment. By directly mapping quantum mechanical-derived force components to the numerical representation of an atoms local environment, accurate, and computationally inexpensive force fields can be developed.^[361–363] A typical workflow is to systematically construct reference data, representing the atomic environments with a numerical fingerprint, sampling nonredundant data, and learning the forces. The ML databases can be pre-computed or updated on the fly when required. An example of the on-the-fly approach is the paper of Li et al.^[364] Forces on atoms are either predicted or, if necessary, can be computed by on-the-fly QM calculations and added to a growing ML database, whose completeness is, thus, never required. As a result, the scheme is accurate and general, while progressively fewer QM calls are needed when a new chemical process is encountered for the second and subsequent times.

Atomistic Machine-learning Package (Amp) is an open-source package designed to easily bring machine learning to atomistic calculations.^[365] This allows one to predict (or really, interpolate) calculations on the potential energy surface, by first building up a regression representation from a “training set” of atomic images. The Amp calculator works by first learning from any other calculator (usually quantum mechanical calculations) that can provide energy and forces as a function of atomic coordinates. Depending upon the model choice, the predictions from Amp can take place with arbitrary accuracy, approaching that of the original calculator. Amp is designed to integrate closely with the Atomic Simulation Environment (ASE).

DeePMD-kit, a package written in Python/C++ that has been designed to minimize the effort required to build deep learning-based representation of potential energy and force field and to perform molecular dynamics.^[366] DeePMD-kit is interfaced with TensorFlow (<https://www.tensorflow.org>), one of the most popular deep learning frameworks, making the training process highly automatic and efficient. On the other end, DeePMD-kit is interfaced with high-performance classical molecular dynamics and quantum (path-integral) molecular dynamics packages, that is, LAMMPS and the i-PI, respectively. Thus, upon training, the potential energy and force field models can be used to perform efficient molecular simulations for different purposes. Based on DeePMD-kit, Zhang et al. used active learning to obtain uniformly accurate interatomic potentials for materials simulation.^[367]

Jpred (<http://www.compbio.dundee.ac.uk/jpred4>)^[368] predicts secondary structural elements for an amino acid sequence using machine learning approaches, based on known 3D structural information as available in the Protein Data Bank. Machine learning approaches can predict protein–ligand binding accurately.^[369]

Tangadpalliwar et al. developed “ChemSuite,”^[370] a stand-alone application for chemoinformatics calculations and machine-learning model development. Availability of multi-functional features makes it widely acceptable in various fields. Force field such as UFF is incorporated in tool for optimization of molecules. “SchNet” is another software package that can accurately predict a range of properties across chemical space for molecules and

Table 12. Functional form of the core-shell model for zeolites, metal-oxides and minerals. Note that the interaction between a core and its shell is skipped in the VDW and Ewald real-part, and a correction-term is added to the Fourier sum.

$$\begin{aligned}
 \mathcal{U}^{\text{total}} &= \mathcal{U}^{\text{core-shell}} + \mathcal{U}^{\text{3-body}} + \mathcal{U}^{\text{nonbonded}} \\
 \mathcal{U}^{\text{core-shell}} & \\
 \text{core-shell energy} & \quad \mathcal{U}^{\text{core-shell}}(r_{ij}) = \sum_{\text{core-shells}} \frac{1}{2} k(r_{ij})^2 \\
 & \quad \mathcal{U}^{\text{3-body}} \\
 \text{three body energy} & \quad \mathcal{U}^{\text{3-body}}(\theta_{ijk}) = \sum_{\text{bends}} \frac{1}{2} k(\theta_{ijk} - \theta_0)^2 \\
 \mathcal{U}^{\text{nonbonded}} &= \mathcal{U}^{\text{VDW}} + \mathcal{U}^{\text{Ewald}} \\
 \text{Van der Waals energy} & \quad \mathcal{U}^{\text{VDW}} = \sum_{i<j} A \exp[-r_{ij}/\rho_{ij}] - (C_{ij} r_{ij}^{-6}) \\
 \text{Ewald charge energy} & \quad \mathcal{U}^{\text{Ewald}} = \mathcal{U}^{\text{real}} + \mathcal{U}^{\text{rec}} + \mathcal{U}^{\text{self}} + \mathcal{U}^{\text{exclusion}} \quad \text{where} \\
 & \quad \mathcal{U}^{\text{real}} = \sum_{i<j} q_i q_j \frac{\text{erfc}(\eta r_{ij})}{r_{ij}} \\
 & \quad \mathcal{U}^{\text{rec}} = \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \frac{1}{k^2} e^{-\frac{k^2}{4\eta^2}} \left(\left| \sum_{i=1}^N q_i \cos(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 \right. \\
 & \quad \left. + \left| \sum_{i=1}^N q_i \sin(\mathbf{k} \cdot \mathbf{r}_i) \right|^2 \right) \\
 & \quad \mathcal{U}^{\text{self}} = -\frac{\eta}{\sqrt{\pi}} \sum_i q_i^2 \\
 & \quad \mathcal{U}^{\text{exclusion}} = -\sum_{i<j} q_i q_j \frac{\text{erf}(\eta r_{ij})}{r_{ij}}
 \end{aligned}$$

materials, where their model learns chemically plausible embeddings of atom types across the periodic table.^[371]

7. Zeolite and MOF Force Fields

7.1. Zeolite Force Fields

Metal-oxides are very successfully modeled using the Core-Shell method.^[69,307,372–374] To model polarization on the oxygen, the charge of the oxygen is divided over the “core” and the “shell” which are separated by a spring. Typically, a Buckingham potential is used for the metal (with no attraction term), while for the oxygen both the exponential repulsion and the $1/r^6$ attraction term are used. The attraction of the oxygen is very strong (see ref. [335] for a compilation of metal-oxide parameters). A common functional form of the model is shown in Table 12. In the shell model, the short-range repulsion and van der Waals interactions are taken to act between the shell particles. This finding has the effect of coupling electrostatic and steric interactions in the system: in a solid-state system where the nuclei are fixed at lattice positions, polarization can occur not only from the electric field generated by the neighboring atoms, but also from the short-range interactions with close neighbors. This ability to model both electrical and mechanical polarization is one reason for the success of shell models in solid-state ionic materials.^[216]

The core-shell modes are also very successfully applied to zeolites.^[69,307,374–378] The predicted crystal structures agreed very well with experimental crystallographic data. The core-shells approach allowed to study structural phase transitions, zeolite hydration processes, phonon-dispersion curves, and negative thermal expansion in zeolitic materials. However, it unfortunately proved difficult to combine these models with adsorbate models to study adsorption. The core-shell model requires formal charges on the atoms while charges based on ab initio approaches are significantly lower. In addition, this type of model increases the amount of particles in the system. This reduces the efficiency

of the calculation. To overcome these inherent drawbacks several other force fields for zeolite were proposed.

The Demontis model consists of an O-Si bond-potential and an Urey-Bradley Si-T-Si potential (bond potential between 1-3).^[379,380] The models are restricted to the simulation of adsorbates without the use of charge. This includes alkanes as simulated in the united-atom approach. The advantage of the model is not its accuracy, but its efficiency. The model was extended to hydrated aluminosilicates including cations and Coulombic interactions were added.^[381] The BKS force field was developed by van Beest et al. to describe the geometries, and mechanical properties of silicas and aluminophosphates.^[382] The parameterization is based on both ab initio and experimental data. The force field contains four atom types: Si, O, Al, and P. One important observation by the authors is that when the inter-atomic interactions goes beyond nearest neighbors it is inevitable to also use macroscopic information (here elastic constants) in the force field parameterization. Kramer et al. further outlined their route to develop effective inter-atomic force fields for zeolites from ab initio calculations on small clusters.^[383] The Nicholas model used bond-, bend-, torsion-potential, as well as Lennard-Jones and Coulombic long-ranged interactions.^[190] A Urey-Bradley term was included on the Si-Si nonbonded distance for each Si-O-Si angle to couple the bond-stretching to the bend-angle (in silicates the Si-O bond is known to lengthen as the Si-O-Si bond angle becomes smaller). When the Si-O-Si angle is close to linear, it is possible for it to invert, causing a discontinuity in the torsional potential that contains the Si-O-Si angle. Therefore, the torsion potential was coupled to the related Si-O-Si angle, so that the torsional energy goes smoothly to zero as the Si-O-Si bond becomes linear. The force field has been demonstrated to reproduce the structure and dynamics of silica sodalite with use of energy minimization, normal mode analysis, and molecular dynamics techniques. Schrimpf et al.^[384] developed a mode for zeolite NaY, accounting for the flexibility of the lattice. It is mainly based on effective pair potentials between neighboring atoms. The comparison of calculated and experimental data shows that the main structural and dynamical properties of the zeolite are reproduced.^[384] Smirnov and coworkers parameterized the simplified general valence force field (SGVFF) to study window fluctuations and Raman and infrared spectra of zeolites and aluminosilicates.^[385-388] In a comparison to experimental data of spectra generated by different force fields, Bueno-Pérez et al. found that the Nicholas and Hill force fields generate the most similar IR spectra.^[389] However, none of these force fields allowed the identification of frameworks based on their experimental spectra. Hill and Sauer derived a CFF force field for the simulation of aluminum-free zeolite structures^[390] and aluminosilicates^[391] from results of ab initio calculations on molecular models which represent typical structural elements of zeolites. In addition to the Nicholas forms, they also included bond-bond, bond-bend, bend-bend, and angle-angle-torsion terms. Flexibility is essential for studying properties of the zeolites themselves, and for example transition between phases. Yamahara et al. performed MD simulations on MFI and reproduced the temperature-induced phase transition between orthorhombic and monoclinic phases.^[392] Bordat et al.^[393] modified the BKS model with reduced partial atomic charges and re-optimized covalent bond potential wells and reproduced

both the structure of the framework and its monoclinic to orthorhombic transition. Combined with the SPC/E model of water, the model also reproduces the important result that pure unhydroxylated silica is hydrophobic.

Zeolites with germanium substitutions have also been modeled with force fields with Buckingham potentials.^[376] The study of these zeolites has been interesting for the synthesis of pure silica zeolites. These force fields allow the study of the position and distribution of germanium atoms in the structure.^[394,395] The predicted germanium distribution agreed very well with experimental NMR data.

To study diffusion of benzene in Na-Y and develop a transport theory for cationic zeolites, Auerbach modeled the zeolite fully flexible.^[396] The Si and Al atoms were replaced with an average tetrahedral-site atom (T-atom) because of the difficulty in experimentally determining Al distributions in disordered zeolites. The work was the first long length scale simulation of diffusion in cation containing zeolites. Jaramillo and Auerbach developed and validated a new force field for cations in zeolites, which explicitly distinguishes Si and Al atoms, as well as different types of oxygens in the framework.^[397] Their molecular dynamics simulations show that most cations are immobile in Na-X and Na-Y on the MD time scale, even at 1000 K. The MD simulations also show that cation movement is highly correlated. Most of the studies up to this time used force fields with the same partial charge on silicon and aluminum. However, because of the close proximity between cations and framework atoms this approach is unsuited for modeling cations in zeolites and it cannot account for the allocation of cations to specific positions. The model of Jaramillo and Auerbach was extended by Calero et al. to model adsorption of hydrocarbons in Na-Y and their results showed that these cations not only have a large influence on diffusion but also on adsorption.^[398]

Jeffroy et al. provide a general and transferable force field able to predict the structure of zeolites for various type of extra-framework cations.^[399] Based on simple functional forms and interactions, it can be easily implemented in most common molecular simulation codes. The optimized force field is validated on structural properties (lattice parameters and Si-O-Al angles) for a large variety of zeolites, including faujasites of different Si/Al ratio and different extra-framework cation types (Li^+ , Na^+ , K^+ , Mg^{2+} , Ca^{2+} , and Co^{2+}). The transferability of the force field was successfully tested on zeolites of different topologies such as FAU, LTA, MFI, FER, and TON. Shi et al. studied the temperature dependence of zeolite pore sizes using ab initio and classical simulations^[400] for all-silica zeolites SOD, FER, and MFI over the temperature range 300–900 K. Such simulations are important for understanding and predicting zeolite/guest fit, especially for relatively bulky guest species. Several force fields were tested and a modified BKS force field was found to give the best agreement with the simulated properties listed above, especially for MFI.

Some more recent work focuses on moving toward more generic force fields. Pedone et al. developed a new empirical pairwise potential model for ionic and semi-ionic oxides.^[401] Its transferability and reliability have been demonstrated by testing the potentials toward the prediction of structural and mechanical properties of a wide range of silicates. Gabrieli et al. created a new force field for fast MD simulations in flexible aluminosilicates by adapting to the CHARMM functional form.^[402] This

new force field allows the execution of large-scale simulations in a parallel environment via the most common packages available and correctly reproduce the crystallographic structures and the vibrational properties for silicalite and zeolites Na A, Ca A, Na Y, and Na X. The CHARMM functional form was also used by Praprotnik et al.^[403] to construct a new all-atom force field for molecular dynamics simulation of an AlPO₄-34 molecular sieve.

Zimmerman et al. proposed a modified CHARMM force-field (ZHB potential) with low point charges for silica.^[404] Lennard–Jones zeolite parameters were chosen to be compatible with existing CHARMM parameters. Sahoo and Nair modified the ZHB potential (MZHB) by reformulating its bonding potential, while retaining the nonbonding potential as in the ZHB force-field.^[405] They showed that several structural and dynamic properties of silica, like the IR spectrum, distribution functions, mechanical properties, and negative thermal expansion computed using the MZHB potential agree well with experimental data. Gu and Hammond developed a potential energy model for siliceous zeolites based in part on fits to infrared (IR) spectra by extending the MZHB potential.^[406] This potential is tested in terms of its ability to model seven silica polymorphs (α -quartz, α -cristobalite, and five siliceous zeolites: zeolite Y, sodalite, silicalite-1, zeolite A, and chabazite). A sensitivity analysis shows significant sensitivity of the lattice parameter to the atomic partial charges, indicating a possible target for the parameterization of the atomic partial charges for crystalline materials. The Hill–Sauer force field for silica has recently been improved and the transferability of the force field was tested on 13 experimentally known silica eight-ring zeolites.^[407]

7.2. MOF Force Fields

The first work on flexible force fields was on the prototypical IRMOF-1 structure and started around 2006. Greathouse and Allendorf studied water decomposition of IRMOF-1 using a model adapted from the generic CVFF force field.^[408] To incorporate the possible breakdown of the structure, the Zn–O interactions were modeled using a nonbonding approach. Using this model, Greathouse and Allendorf demonstrated a structural collapse occurring around 4% water content. An understanding of the water stability and adsorption concepts has been reviewed by Burtch, Jajusa, and Walton.^[409] The Greathouse and Allendorf model was later extended to other isorecticular zinc carboxylate coordination polymers.^[410] Three other flexible force fields for IRMOF-1 were introduced in 2007 by Amirjalayer et al.,^[238] Dubbeldam et al.,^[411] and Han and Goddard.^[412] Amirjalayer et al. extended the MM3 force field and used a building block approach for the parameterization of IRMOF-1. It correctly predicts the structure of IRMOF-1 and the MM3 force field was chosen for a more accurate reproduction of the vibrational modes. The parameters for the Zn₄O moiety were refined based on first principles DFT of nonperiodic model systems.^[413] Dubbeldam et al. re-parameterized the Greathouse and Allendorf force field by reproducing the experimental lattice parameters, and CO₂ and methane adsorption isotherms.^[411] These authors showed that IRMOFs have exceptionally large negative thermal expansion coefficients and that this phenomenon is likely a generic feature

of MOFs associated with the struts-and-spring, large-pore nature of MOFs. Negative thermal expansion in MOFs is related to relatively rigid linkers connected to rigid metal clusters via flexible groups such as carboxylate groups and the porosity of the structure that allows adequate volume for motion. The linker molecule undergoes transverse motions due to thermal vibrations such that at lower temperatures the linker becomes more rigid and stretches out. The simulations were in excellent agreement with later experiments using neutron powder diffraction and first-principles calculations using minimization of the free energy.^[414] Han and Goddard used the DREIDING force field to further elucidate the mechanism of negative thermal expansion in MOFs.^[412] The magnitude of the negative thermal expansion effect was found to be sensitive to the specific types of organic linker used. This change in thermal expansion can also change from negative to positive with the action of the pressure from adsorbed molecules.^[415] Huang et al. developed a force field for IRMOF-1 and calculated the phonon thermal conductivity and vibrational power spectra.^[416] Using the approach based on the MM-type force field Schmid and Tafipolsky,^[417] and Amirjalayer et al.^[418] developed force fields for COF-102 and boron based covalent organic frameworks (COFs) in general, respectively.

The Dubbeldam model for IRMOF-1 force field reproduced the experimental thermal expansion properties well. A later refinement was made to reproduce the ab initio elastic tensor.^[419] The classical model of IRMOF-1 of Dubbeldam et al. gave $C_{11} = 29.3$ GPa, $C_{12} = 11.9$ GPa, and $C_{44} = 0.985$ GPa. DFT calculations gave $C_{11} = 29.2$ GPa, $C_{12} = 13.1$ GPa, and $C_{44} = 1.4$ GPa for LDA quantum using VASP,^[70] and the experimental and DFT combined work of Zhou and Yildirim found $C_{11} = 29.42$ GPa, $C_{12} = 12.56$ GPa, and $C_{44} = 1.16$ GPa.^[71] In order to refine the model for reproduction of elastic constants the generalized Hessian matrix was used to optimize the IRMOF-1 structure to a true minimum with all positive eigenvalues. The second use of the generalized Hessian was to analytically compute the elastic constants with machine precision. This mode-following optimization mechanism was previously applied to core-shell models of minerals and zeolites and available in the GULP software.^[307] Using this optimization procedure on MOFs showed that, for example, the Greathouse and Allendorf minimum structure was a structure with 32 negative eigenvalues. A further optimization using generalized Hessian matrix showed that their model actually leads to an asymmetric and Zn₄O distorted structure (in disagreement with the experimental structure). Mode-following optimization is therefore an incredibly powerful tool to identify problems in force fields.

The IRMOF-1 structure does not show large-scale structural changes. Rather the atoms vibrate largely around their equilibrium positions. Ford et al. only found minor differences between the diffusion results from the rigid and flexible framework models for several short n-alkanes and benzene in IRMOF-1.^[420] Extreme examples of framework flexibility are the "breathing MOFs," with the MIL-53 structure as a prime example.^[26] Pore breathing structures can expand or shrink to admit guest molecules. MIL-53 is a three dimensional MOF containing unidirectional diamond-shaped channels. The open form of MIL-53 (MIL-53lp) is obtained upon calcination of the as-synthesized compound, while the adsorption of water at room temperature leads to the closed, narrow-pore form (MIL-53np). Upon

heating, the water adsorption is reversed, leading to a recovery of the open form.^[26,421] Salles et al. used MD simulations to study the unusually complex case of guest-induced structural transformations of MIL-53.^[247] The model successfully captures the two-step structural switching, as induced by CO₂ adsorption in the Cr-containing framework at finite temperatures. Coombes et al. develop a hybrid force field combining ionic-model potentials, used for modeling inorganic solids, with molecular mechanics terms for the organic part. This gives an energy difference of close to 30 kJ mol⁻¹ between the large- and narrow-pore forms.^[422] Water molecules are being more strongly stabilized in the narrow-pore form which then favor pore closure at loadings of more than one molecule per unit cell. The flexibility of MIL-53 is controlled by weak dispersion forces.^[423] Coudert et al. developed the osmotic thermodynamic model, a general thermodynamic framework for the understanding of guest-induced structural transitions in hybrid organic-inorganic materials.^[424] Another semi-analytical model for rationalizing the breathing of MIL-53 type materials was developed by Vanduyfhuys et al.^[425] The model can predict breathing behavior of other flexible materials, using as an input the free energy of the empty host, the interaction energy between a guest molecule and the host matrix and the pore volume accessible to the guest molecules. The MIL-53 models were further improved by Vanduyfhuys et al. by being calibrated against density functional theory calculations on nonperiodic clusters.^[426] The force field predicts geometries and cell parameters that compare well with the experimental values both for the large and narrow pore phases. The relative stability of the two phases critically depends on the VDW parameters, and the MM3 dispersion interaction has the tendency to overstabilize the np phase. Una-Triguero et al.^[427] studied a similar breathing effect of ZJU-198 MOF. They predicted a structural phase transition from a np to a lp phase upon the adsorption of acetylene, ethene, and carbon dioxide. However, methane and nitrogen are not able to produce structural changes in ZJU-198. The np phase corresponds with the as-synthesized structure, while the lp phase was obtained by a geometry optimization of the framework loaded with adsorbates. These calculations were performed by energy minimization simulation using the DREIDING inter-atomic potential. Monte Carlo simulations using the optimized structures were performed to obtain the adsorption isotherms of small gases in ZJU-198, showing a good agreement with the experimental data.

MIL-47 has V^{IV}O₆ octahedra linked by rigid terephthalate anions and, in contrast to MIL-53, is quite rigid at ambient conditions. Despite the same topology, the use of Al or Cr instead of V leads to large-scale motion. Because of that tight relation between MIL-47 and MIL-53, the MIL-47 is also heavily studied. Hamon et al. created a MIL-47(V) framework model by combining the generic Universal Force Field for the inorganic part of the crystalline lattice (the metal center with its oxygen environment) and the DREIDING inter-atomic potential for the organic part.^[428] MIL-47(V^{IV}) does not collapse even upon high mechanical pressure, but there is a structural transition for the MIL-47(V^{IV}) solid upon mechanical pressure.^[429] Yot et al. developed a bonded force field for treating the flexibility of the MOF framework that was inspired on the isostructural MIL-53(Cr^{III}). Wieme et al. developed a first-principles-based force field for MIL-47. Specific terms were added for the materials at hand to describe the asymmetry of the 1D vanadium-oxide chain and to account for the flexibility

of the organic linkers.^[430] The large-scale flexibility in MOFs revolves around the behavior of metal clusters that can be thought of as “hinges” acting between rigid “struts.”^[2,431] Based on this observation, Heinen and Dubbeldam used fitting on elastic constants from DFT to refine the parameters of the hinges of the Yot-model for MIL-47. To assess the effect of framework flexibility, it was shown for MIL-47(V) that framework flexibility has negligible influence on the adsorption and diffusion properties for small guest molecules.^[337]

DMOF-1 is a 3D, jungle-gym-type MOF that breathes upon adsorption of benzene or isopropyl alcohol, but remains rigid with CO₂ and H₂.^[432] Using a flexible model for DMOF-1, derived from ab initio data, Grosch and Paesani obtained fundamental insights into the molecular mechanisms that determine the breathing behavior upon adsorption of benzene and isopropyl alcohol.^[433] Changing the identity of the linear, ditopic ligand in either the 2D layer or the pillaring third dimension allows targeted modulation of the chemical functionality, porosity, and interpenetration of the framework.^[434] By varying the functional groups on the 1,4-benzenedicarboxylic acid (BDC) linker of DMOF, tune the kinetic water stability of this isostructural, pillared family of MOFs can be systematically tuned.^[435] Also cobalt-, nickel-, copper-, and zinc-based pillared MOFs have been made.^[436] Burtch et al. adapted the model of Grosch and Paesani to study the rotation of the 1,4-diazabicyclo [2.2.2] octane (DABCO) ligand in the pillared Zn-DMOF and Zn-DMOF-TM (TM=tetramethyl) structures.^[437] In another study, Burtch et al. studied properties that are inherent to the DMOF frameworks themselves such as thermal expansion and ligand flexibility.^[438]

Besides IRMOF-1, the most famous MOF is probably Cu-BTC, also known as HKUST-1.^[439] At the open metal site, competitive adsorption can take place that enhances adsorption of adsorbates like CO₂ when water molecules are present.^[440] Ryder et al. demonstrated that the co-existence of soft modes and intrinsic shear distortions connected to the THz lattice dynamics in Cu-BTC dictate a range of anomalous elastic phenomena, for example: negative Poisson's ratio (auxeticity), negative thermal expansion (NTE), and exceedingly low shear moduli properties.^[32] A flexible force field for Cu-BTC was developed by Tafipolsky et al.^[441] based on the established MM3 force field, extended by additional cross terms and specific bond-stretching and angle-bending terms for the square-planar CuO₄ coordination environment. Zhao et al. constructed a force field for Cu-BTC using density functional theory calculations and other parts from other force fields, which reproduce well the experimental crystal structure, negative thermal expansion, vibrational properties as well as adsorption behavior in Cu-BTC.^[442] STAM-1^[443] is a MOF constructed with the same organic linker (benzene-1,3,5-tricarboxylic acid) and metal center as Cu-BTC. A monomethyl esterification of the organic linker during the synthesis gives rise to a new MOF with unique properties. STAM-1 shows a gate-opening mechanism during the adsorption of light alcohols and water. However, this effect was not observed for non-polar molecules such as alkanes, carbon dioxide, oxygen, and nitrogen. This flexibility of the structure combined with the presence of hydrophilic/hydrophobic 1D channel system and open metal sites results in interesting adsorption properties of this material. Slawek et al.^[444] investigated the origin of the gate opening mechanism upon the adsorption of polar adsorbates

in STAM-1 by means of experimental and molecular simulation techniques. Slawek et al. used the flexible force field developed by Zhao et al.^[442] for Cu-BTC, due to the similarity between the two MOFs. MC simulations in the osmotic ensemble and MD simulations using this flexible force field provided an accurate description of the framework flexibility upon the adsorption of methanol in STAM-1. The prediction of accurate multi-component adsorption isotherms for mixtures containing olefin and paraffin molecules in open-metal site materials like Cu-BTC is difficult. Ideal Adsorption Solution Theory (IAST) fails, but Heinen et al. have shown that accurate results can be obtained by explicit binary-mixture simulations using a force field that captures the relevant orbital interactions using an additional charge-induced-dipole-like term.^[35] Recently, Durholt et al. parameterized a coarse grained force field for the copper paddle-wheel based HKUST-1,^[445] allowing to extend both length and timescale in the simulation of MOFs.

The zirconium-based MOF UiO-66(Zr)^[446] is comprised of zirconium oxide clusters connected by benzene dicarboxylate (BDC) linkers. The framework demonstrates excellent chemical, hydrothermal, and mechanical stability,^[446] possesses reverse shape selectivity in the adsorption of hexane and xylene isomers.^[447,448] Yang et al. studied CO₂ and CH₄ mixture adsorption and diffusion in UiO-66 using a fully flexible framework.^[449] The force constants for the organic ligand in UiO-66(Zr) were directly taken from the CVFF force field, while the equilibrium bond distances and angles were assigned by averaging the values obtained from the DFT optimized crystal structure. The interactions between the inorganic and organic parts, that is, the torsion and bend term, were re-parameterized. The LJ potential parameters were taken from the DREIDING force field. A more transferable model, the BTW-FF model, included UiO-66 and UiO-67 in the initial parameterization set.^[450]

ZIF materials combine the topology and stability of zeolites with the chemical diversity of MOFs. Typical flexibility effect in ZIFs are for example the gate-opening behavior observed in ZIF-8,^[451] and flexibility of ZIF-7 and its structural impact in alcohol adsorption.^[452] Zheng et al. presented a full set of force field parameters for ZIF-8, based on the AMBER database and on previously computed partial charges, well reproducing the ZIF-8 structural properties over a wide range of temperatures and pressures.^[453] The AMBER force field was also the basis of the flexible force field of Hu et al.,^[454] with several parameters taken from quantum chemical calculations. A structural transition from crystalline to amorphous as found in experiment was observed. Only with the structural flexibility included were the predicted diffusivities of CH₄ and CO₂ close to reported data in the literature. Also, generic force fields like DREIDING were used to model ZIF-8.^[455] Recently, more general and transferable intra-molecular ZIF force fields were developed by Verploegh et al.^[456], ZIF-FF by Weng et al.^[457] and MOF-FF for ZIFs by Durholt et al.^[458]

Several other general approaches and force field for MOFs have been developed. MOF-FF, a third generation force field of Schmid and coworkers is based on the building block methodology. The first generation has been discussed in the preceding paragraphs.^[238] In the second generation, a generic algorithm was used to derive force field potentials for the zinc benzoate and dilithium terephthalate.^[336] In the third generation, a new en-

ergy expression was proposed as well as a new parameterization scheme for Cu and Zn paddlewheel-based structures and Zn4O and Zr6(OH)4-based MOFs.^[459] A popular approach is to use a building block methodology utilizing ab initio calculations on nonperiodic clusters to fit parameters. QuickFF is a program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input.^[460] The QuickFF force field protocol was extended for an improved accuracy of structural, vibrational, mechanical, and thermal properties of metal-organic frameworks like IRMOF-1, MIL-53(Al), CAU-13, and NOTT-300.^[461] UFF4MOF^[462,463] is an extension of the original Universal force field UFF (UFF) that incorporates additional atom types found in the Computation-Ready Experimental (CoRE) Database.^[464] The BTW-FF model is a transferable inter-atomic potential that can be applied to MOFs regardless of metal or ligand identity. The initial parameterization set included IRMOF-1, IRMOF-10, IRMOF-14, UiO-66, UiO-67, and HKUST-1.^[450] By extending the BTW-FF model, Bristow et al. developed the vibrational metal-organic framework (VMOF) force field.^[465] Here, force field parameters are explicitly fitted on DFT-PBESol calculated phonon spectra of periodic binary oxides such as ZnO, ZrO₂, TiO₂, and Al₂O₃. A modified MM3 Buckingham potential for the metal-linker interaction was needed to reproduce the ab initio and experimental structural and mechanical properties of the binary oxides more accurately.

8. Outlook

Bottom-Up vs Top-Down: Long-range interactions necessitate the use of macroscopic information in the force field parameterization.^[382] That means that system properties like unit cell size/shape and elastic constants are used to fit local parameters in a top-down approach. In contrast, local interactions, like bond, bend, and torsions can be efficiently obtained from small clusters using ab initio data as a reference. Small clusters are unsuitable for obtaining long-range interactions, and more and more approaches are based on using the full crystallographic unit cell. Another reason for using full unit cell or even larger cells is the modeling of defects.^[466,467]

Ab Initio Approach vs Empirical Fitting: Parameters from one force field are in principle not transferable to another. Force field design and parameterization is highly non-trivial as evidenced by the large number of different force fields. For the different classes of force fields the explanation is rather obvious: they have very different function forms, but the reasons per class are different. For class I force fields, the lack of cross-terms makes the diagonal parameters less transferable. Class II force fields applied to systems where polarization and charge transfer is negligible are potentially more accurate, but require more parameters to optimize. However, most of the additional parameters are additional intra-molecular interactions that can be obtained from genetic algorithm using ab initio structural and vibrational information as target data. The main reason why force fields parameterization is difficult is the treatment of long-range interactions.

VDW interactions are notoriously difficult to compute accurately at the ab initio level. Yet they are responsible for most of interaction energies involved in adsorption and for the

large-scale structural changes of flexible MOFs. But even if you *could* compute them with unlimited precision, there is no one-to-one mapping of a classical force field on the QM PES. If one traces the energy for a CO₂ molecule in different orientations as a function of distance to a metal-site, then all of these energy-distance curves have different errors, but all of them have an intrinsic error. This error will be magnified at high temperature where entropy comes into play (and sometimes becomes dominant). Therefore, fitting the ab initio data for VDW potentials is often replaced by fitting to experimental data. Fitting VDW potentials to experimental data (such as VLE curves) has additional advantages. It “lumps” other errors, like the assumption of a rigid molecule, into the mix which then are optimized as a whole. Even when flexibility and polarization are not explicitly included in the model, they implicitly enter via the fitted parameters.

Machine Learning: Molecular simulation is a powerful tool to conduct “in-silico” experiments on atomic systems. DFT is currently applicable to hundreds of atoms, but using a classical formulation trillions of atoms can be used.^[468] However, while the increase in computational power and algorithms enables simulation of larger systems that can be run for longer times, one simulation itself is just one typical reproduction of the physics. To really get to the bottom of the physical-, chemical-, and biological phenomena that occur, it is vital to elucidate the what, why, and hows. Therefore, even if you could atomically simulate the system on large space and timescales, in many cases it is preferable to simplify the system significantly to focus on one particular aspect. For example, using transition state for rare events, the reaction coordinate provides valuable information what exactly happens during the transition and why it occurs.

The main use of molecular simulation is to provide understanding. For this purpose, qualitative agreement with experimental data is sufficient. Reproducing experiments is not the end point, but rather the start point for obtaining understanding. A decomposition of the total energy in terms with chemical meaning is essential to be able to investigate microscopic origins of physical behavior. For example, one could switch off terms like electrostatics to check the effect and influence, or increase point charges, or fix certain torsion angles, etc. Molecular simulations can therefore be viewed as “doing experiments on the computer.”

Machine learning can be used as a machinery to better parameterize an existing force field, but also as a new force field approach by itself. Replacing a force field by a machine learning approach has up- and downsides. A downside is that the machine learning force field might contain correlations that are artificial or potentially produce physics that is not real. However, if one sees this approach as an additional tool amongst many others, then one could take advantage of it in situations where it really could do great things. One such example is screening of a large set of structures for specific applications.^[347,352,353,469–471] Here, the problem of false positives would not matter too much because it is used to reduce the millions of structures to a small set of promising structures. This reduced dataset, usually a few dozen, are then re-evaluated and ranked using conventional simulation approaches. Zeolite databases on hypothetical structures have been used to analyze structure–property relations.^[472] The Computation-Ready Experimental (CoRE) Database is one

of the first attempts to construct a large set of MOF structures for analysis.^[464,473,474] For the purpose of screening, new methodologies are actively developed to generate large sets of hypothetical structures,^[475] as well as methods to detect and avoid duplicates.^[476,477] Potentially the results of screening could depend on the used force field, but for small molecules not much difference was found in the ranking of materials for the tested DREIDING and UFF force fields.^[478]

GPU Computing: Graphics processing units (GPUs) were initially designed to deliver computer graphics. It was soon realized that the massive parallelism could also be exploited in many other computational tasks, for example, molecular simulation. This leads to GPGPU (General Purpose GPU) programming. Modern GPUs contain hundreds of processing units and potentially offer large speedups compared to CPU-computing. The growth of the market used to be driven by evolving graphic games, but currently augmented reality (AR), virtual reality (VR), and AI on the GPU are gaining a lot of traction.

The last decade GPU-accelerated molecular modeling has really matured.^[479] Although capable of very high rates of floating-point arithmetic, GPUs are also intrinsically highly parallel processors and their effective exploitation typically requires extensive software refactoring and development.^[480] CPU parallelism can be achieved with SIMD, OPENMP, and MPI which is more task-parallel, but for the GPU specialist programming languages such as CUDA and OpenCL were developed. Many force field implementations have now GPU-support, for example, Amber,^[481–483] AMOEBA,^[484] and MFF and UFF.^[485] Examples of codes that are designed to run on the GPU are the OpenMM^[486] for MD, and GOMC^[487] for MC. The programming paradigms for GPU computing are vastly different in CUDA/OpenCL. In these languages, a computation must be mapped to work-groups of work-items that can be executed in parallel on the compute units. That means that new algorithms had to be developed that map optimally on GPU hardware. Examples of these modifications are: nonbonded-interactions^[488] and efficient cell list implementations,^[489] electrostatics,^[490–492] many-body potentials,^[493,494] ReaxFF,^[495] rigid-body constraints,^[496] Monte Carlo,^[487,497–501] and MD.^[502–504] Other use cases are to speed up post-processing,^[505] screening,^[506] accelerating analysis of void space in porous materials on multicore and GPU platforms,^[507] and boosting theoretical zeolitic framework generation for the determination of new materials structures.^[508] The iRASPA GPU-accelerated visualization software for materials allows interactive examination of energy surfaces and computation of surface areas and void volume^[509]

While on the CPU there is little difference in performance between double- and single-precision (except when using SIMD units), the GPU is optimized for single-precision (which is sufficient for the dominant use case: graphics). Besides that double-precision is *much* slower on the GPU, there are also significantly less computing units of it. Therefore, most GPU codes use only single-precision.^[510] However, for challenging systems the use of single-floating point precision may result in quantitatively and even physically wrong results.^[511] The goal of many algorithms is to obtain the same accuracy but better performance with more low-precision computations. The relation between computational precision and final accuracy is complicated. The challenge in using single-precision operations is the reduction of

cancellation or round-off errors in the calculation results. One solution is to use mixed precision models,^[512] another one is to use quasi double-precision methods to accelerate scientific applications.^[510] Many simulation methods require a large number of independent random variables generated at each step. Some solutions include generating random numbers on the CPU and uploading these sets to the GPU before use. However, also solutions have been proposed that generate random numbers on the GPU.^[513,514]

The future looks bright for GPGPU. Increasing higher performance will be achieved by increasing the number of computing units. Also *tensor*-cores have been recently introduced for acceleration of ML on the GPU. An alternative is that also on the CPU side more and more cores are and will be introduced. However, GPUs currently significantly outperform CPUs on both bandwidth and floating point operations, and this gap is growing.

Open Force Field Initiative: Current force fields are very successful at what they were constructed for, but at the same time limited in accuracy and transferability when applied to other systems. There is no “one force field to rule them all,” as evidenced by the large amount of different force fields in common use (Table 1). The core of present day force fields has remained almost the same since the 1980s and 1990s. Relatively few new efforts have been undertaken since then. Developing force field is hard and time-consuming.

Scientific software is called a “community code” if it is freely available, written by a team of developers who welcome user input, and has attracted users beyond the developers.^[515] Recent contributions to the LAMMPS code by other authors are, for example, an implementation of the neural network potential within the molecular dynamics package LAMMPS,^[516] OCTP, a tool for on-the-fly calculation of transport properties of fluids with the order-*n* algorithm,^[517] and PyLAT, an open-source Python-based post-processing routines to compute viscosities, self-diffusivities, ionic conductivities, molecule or ion pair lifetimes, dielectric constants, and radial distribution functions using best-practice methods.^[518]

The Open Force Field Consortium (<https://openforcefield.org>) is one such effort to bundle resources via community codes. The Open Force Field Consortium aims to solve these and other significant problems by developing a new generation of open force fields, along with the open software infrastructure and open datasets to advance the field and accelerate progress. The consortium seeks to improve force field science and accelerate progress. It aims to make force field fitting/extension straightforward by automating derivation of: i) numerical parameters ii) classification of atom types, bond types, etc., and iii) confidence metrics.

The initiative generate/curates open datasets necessary for producing high-accuracy (bio-) molecular force fields, but also develops a machinery to automatically parameterize molecular mechanics force fields. One important aspect is automating chemical perception, that is, the way force field parameters are assigned to a molecule based on chemical environment. The SMIRKS Native Open Force Field (SMIRNOFF) avoids the complexities of atom typing and the format allows for changes to easily be made in both the chemical perception and quantitative parameter space.^[519]

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was sponsored by NWO Exacte Wetenschappen (Physical Sciences) for the use of supercomputer facilities, with financial support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organization for Scientific Research, NWO). T.J.H. Vlucht would like to thank NWO-CW (Chemical Sciences) for a VICI grant and S. Calero would like to thank the Spanish Ministerio de Economía y Competitividad (CTQ2016-80206-P).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

force fields, metal-organic frameworks, molecular simulation, zeolites

Received: July 12, 2019

Revised: August 23, 2019

Published online: September 23, 2019

- [1] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O’Keeffe, O. M. Yaghi, *Science* **2002**, 295, 469.
- [2] O. M. Yaghi, M. O’Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi, J. Kim, *Nature* **2003**, 423, 705.
- [3] S. Kitagawa, R. Kitaura, S. Noro, *Angew. Chem. Int. Ed.* **2004**, 43, 2334.
- [4] R. Q. Snurr, J. T. Hupp, S. T. Nguyen, *AIChE J.* **2004**, 50, 1090.
- [5] U. Mueller, M. Schubert, F. Teich, H. Puetter, K. Schierle-Arndt, J. Pastré, *J. Mater. Chem.* **2006**, 16, 626.
- [6] G. Férey, *Chem. Soc. Rev.* **2008**, 37, 191.
- [7] J. R. Long, O. M. Yaghi, *Chem. Soc. Rev.* **2009**, 38, 1213.
- [8] A. P. Côté, A. I. Benin, N. W. Ockwig, M. O’Keeffe, A. J. Matzger, O. M. Yaghi, *Science* **2005**, 310, 1166.
- [9] H. M. El-Kaderi, J. R. Hunt, J. L. Mendoza-Cortés, A. P. Côté, R. E. Taylor, M. O’Keeffe, O. M. Yaghi, *Science* **2007**, 316, 268.
- [10] K. S. Park, Z. Ni, A. P. Côté, J. Y. Choi, R. Huang, F. J. Uribe-Romo, H. K. Chae, M. O’Keeffe, O. M. Yaghi, *P. Natl. Acad. Sci. U. S. A.* **2006**, 103, 10186.
- [11] J. Jiang, *Curr. Opin. Chem. Eng.* **2012**, 1, 138.
- [12] A. G. Bezus, A. V. Kiselev, A. A. Lopatkin, P. Q. Du, *J. Chem. Soc. Faraday Trans. II* **1978**, 74, 367.
- [13] P. Bai, M. Tsapatsis, J. I. Siepmann, *J. Phys. Chem. C* **2013**, 117, 24375.
- [14] S. L. Mayo, B. D. Olafson, W. A. Goddard, *J. Phys. Chem.* **1990**, 94, 8897.
- [15] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard, W. M. Skiff, *J. Am. Chem. Soc.* **1992**, 114, 10024.
- [16] T. Duren, L. Sarkisov, O. M. Yaghi, R. Q. Snurr, *Langmuir* **2004**, 20, 2683.
- [17] L. Sarkisov, T. Duren, R. Q. Snurr, *Mol. Phys.* **2004**, 102, 211.
- [18] T. Duren, R. Q. Snurr, *J. Phys. Chem. B* **2004**, 108, 15703.
- [19] L. Sarkisov, T. Duren, R. Q. Snurr, *Mol. Phys.* **2004**, 102, 211.

- [20] J. Heinen, D. Dubbeldam, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, 1363.
- [21] A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel, R. Fischer, *Chem. Soc. Rev.* **2014**, *43*, 6062.
- [22] Z. Chang, D.-H. Yang, J. Xu, T.-L. Hu, X.-H. Bu, *Adv. Mater.* **2015**, *27*, 5432.
- [23] F.-X. Coudert, *Chem. Mater.* **2015**, *27*, 1905.
- [24] G. D. Barrera, J. A. O. Bruno, T. H. K. Barron, N. L. Allan, *J. Phys.: Condens. Matter* **2005**, *17*, 217.
- [25] W. Miller, C. W. Smith, D. S. Mackenzie, K. E. Evans, *J. Mater. Sci.* **2009**, *44*, 5441.
- [26] C. Serre, F. Millange, C. Thouvenot, M. Nogues, G. Marsolier, D. Louer, G. Férey, *J. Am. Chem. Soc.* **2002**, *124*, 13519.
- [27] G. Y. Gor, P. Huber, N. Bernstein, *Appl. Phys. Rev.* **2017**, *4*, 011303.
- [28] R. Schmid, *ACS Central Sci.* **2017**, *3*, 369.
- [29] A. J. Fletcher, K. M. Thomas, M. J. Rosseinsky, *J. Solid State Chem.* **2005**, *178*, 2491.
- [30] N. C. Burtch, J. Heinen, T. D. Bennett, D. Dubbeldam, M. D. Allendorf, *Adv. Mater.* **2018**, *30*, 1704124.
- [31] M. R. Ryder, B. Civalieri, T. D. Bennett, S. Henke, R. Svemir, G. Cinque, F. Fernandez-Alonso, J.-C. J.-C. Tan, *Phys. Rev. Lett.* **2014**, *113*, 215502.
- [32] M. R. Ryder, B. Civalieri, G. Cinque, J.-C. Tan, *CrystEngComm* **2016**, *18*, 4303.
- [33] L. Grajciar, M. Rubes, O. Bludsky, P. Nachtigall in, *Metal-Organic Frameworks Materials Modeling towards Potential Engineering Applications* (Ed.: J. Jiang), Pan Stanford Publishing, **2015**, Ch. 4, p. 175.
- [34] A. L. Myers, J. M. Prausnitz, *AIChE J.* **1965**, *11*, 121.
- [35] J. Heinen, N. C. Burtch, K. S. Walton, C. F. Guerra, D. Dubbeldam, *Chem.: Eur. J.* **2016**, *22*, 18045.
- [36] M. A. González, *Collection SFN* **2011**, *12*, 169.
- [37] N. W. Ashcroft, D. Mermin, *Solid State Physics*, Brooks/Cole, Belmont, U. S. A. **1976**.
- [38] C. Kittel, *Introduction to Solid State Physics*, 8th ed., John Wiley & Sons, New York **2005**.
- [39] F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, West Sussex, England **1999**.
- [40] J. Kohanoff, *Electronic Structure Calculations for Solids and Molecules*, Cambridge University Press, Cambridge, UK **1999**.
- [41] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge University Press, Cambridge, UK **2004**.
- [42] D. Marx, J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*, Cambridge University Press, New York **2009**.
- [43] F. M. Bickelhaupt, E. J. Baerends, in *Rev. Comput. Chem.*, Vol. 15 (Eds.: K. B. Lipkowitz, D. B. Boyd), Wiley Online Library, New Jersey, U. S. A. **2000**, p. 1.
- [44] W. Koch, M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, 2nd ed., Wiley, Weinheim, Germany **2001**.
- [45] D. S. Sholl, J. A. Steckel, *Density Functional Theory*, Wiley Online Library, Hoboken, New Jersey **2009**.
- [46] E. Engel, R. M. Dreizler, *Density Functional Theory: An Advanced Course*, Springer-Verlag, Berlin, Germany **2011**.
- [47] M. J. McGrath, J. I. Siepmann, I. F. W. Kuo, C. J. Mundy, *Mol. Phys.* **2006**, *104*, 3619.
- [48] S. Grimme, *J. Chem. Phys.* **2006**, *124*, 034108.
- [49] S. Grimme, *J. Comput. Chem.* **2004**, *25*, 1463.
- [50] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* **2004**, *92*, 246401.
- [51] P. G. Silvestrelli, *Phys. Rev. Lett.* **2008**, *100*, 053002.
- [52] K. G. Ray, D. Olmsted, N. He, Y. Houndonougbo, B. B. Laird, M. Asta, *Phys. Rev. B* **2012**, *85*, 085410.
- [53] G.-M. Piccini, M. Alessio, J. Sauer, Y. Zhi, Y. Liu, R. Kolvenbach, A. Jentys, J. A. Lercher, *J. Phys. Chem. C* **2015**, *119*, 6128.
- [54] A. Kundu, G. M. Piccini, K. Sillar, J. Sauer, *J. Am. Chem. Soc.* **2016**, *138*, 14047.
- [55] C.-C. Chiu, G. N. Vayssilov, A. Genest, A. Borgna, N. Roesch, *J. Comput. Chem.* **2014**, *35*, 809.
- [56] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, S. Cottenier, *Crit. Rev. Solid State* **2014**, *39*, 1.
- [57] M. D. Baer, M. J. McGrath, I.-F. W. Kuo, J. I. Siepmann, D. J. Tobias, *J. Chem. Phys.* **2011**, *135*, 124712.
- [58] M. J. McGrath, I.-F. W. Kuo, J. N. Ghogomu, C. J. Mundy, J. I. Siepmann, *J. Phys. Chem. B* **2011**, *115*, 11688.
- [59] M. J. McGrath, I.-F. W. Kuo, J. I. Siepmann, *Phys. Chem. Chem. Phys.* **2011**, *13*, 19943.
- [60] H. Fang, P. Kamakoti, J. Zang, S. Cundy, C. Paur, P. I. Ravikovitch, D. S. Sholl, *J. Phys. Chem. C* **2012**, *116*, 10692.
- [61] J. Klimes, D. R. Bowler, A. Michaelides, *J. Phys. Condens. Matter* **2009**, *22*, 022201.
- [62] J. Klimes, D. R. Bowler, A. Michaelides, *Phys. Rev. B* **2011**, *83*, 195131.
- [63] J. Sun, A. Ruzsinszky, J. P. Perdew, *Phys. Rev. Lett.* **2015**, *115*, 036402.
- [64] R. Sabatini, T. Gorni, S. De Gironcoli, *Phys. Rev. B* **2013**, *87*, 041108.
- [65] O. A. Vydrov, T. Van Voorhis, *J. Chem. Phys.* **2010**, *133*, 244103.
- [66] H. Peng, Z.-H. Yang, J. P. Perdew, J. Sun, *Phys. Rev. X* **2016**, *6*, 041005.
- [67] F. Tran, L. Kalantari, B. Traoré, X. Rocquefelte, P. Blaha, *Phys. Rev. Mater.* **2019**, *3*, 063602.
- [68] J. Wieme, K. Lejaeghere, G. Kresse, V. Van Speybroeck, *Nat. Commun.* **2018**, *9*, 4899.
- [69] K.-P. Schröder, J. Sauer, *J. Phys. Chem.* **1996**, *110*, 11043.
- [70] A. Samanta, T. Furuta, J. Li, *J. Chem. Phys.* **2006**, *125*, 084714.
- [71] W. Zhou, T. Yildirim, *Phys. Rev. B* **2006**, *74*, 180301.
- [72] T. L. J. Hill, *J. Chem. Phys.* **1946**, *14*, 465.
- [73] F. H. Westheimer, J. E. Mayer, *J. Chem. Phys.* **1946**, *14*, 733.
- [74] F. H. Westheimer, *Steric Effects in Organic Chemistry*, Wiley, New York **1956**.
- [75] U. Dinur, A. T. Hagler, in *Rev. Comput. Chem.*, Vol. 2 (Eds.: K. B. Lipkowitz, D. B. Boyd), Wiley Online Library, New Jersey, USA **1991**, p. 99.
- [76] J. R. Hill, C. M. Freeman, L. Subramanian in, *Rev. Comput. Chem.*, Vol. 16 (Eds.: K. B. Lipkowitz, D. B. Boyd), Wiley Online Library, New Jersey, USA **2000**, p. 141.
- [77] E. G. Lewars, *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, 3rd ed., Springer International Publishing, Switzerland **2016**.
- [78] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, 2nd ed., Clarendon Press, Oxford **2017**.
- [79] C. Chakravarty, *Int. Rev. Phys. Chem.* **1997**, *16*, 421–444.
- [80] M. E. Tuckerman, G. J. Martyna, *J. Phys. Chem. B* **2000**, *104*, 159.
- [81] M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulations*, Oxford University Press, New York **2010**.
- [82] E. R. Hernandez, *AIP Conference Proceedings* **2008**, *1077*, 95.
- [83] E. Braun, J. Gilmer, H. B. Mayes, D. L. Mobley, J. I. Monroe, S. Prasad, D. M. Zuckerman, *Living J. Comp. Sci.* **2019**, *1*, 5957.
- [84] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, 2nd ed., Academic Press, London, UK **2002**.
- [85] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*, 2nd ed., Cambridge University Press, Cambridge **2004**.
- [86] A. R. Leach, *Molecular Modelling: Principles and Applications*, 2nd ed., Prentice Hall, Longman, Essex, England **2001**.
- [87] M. J. Field, *A practical introduction to the simulation of molecular systems*, Cambridge University Press, New York **2007**.
- [88] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.* **1953**, *21*, 1087.

- [89] J. P. Valleau, S. G. Whittington in, *Statistical Mechanics A. Modern Theoretical Chemistry, Vol. 5* (Ed.: B. Berne), Plenum Press, **1977**, p. 137.
- [90] D. Dubbeldam, A. Torres-Knoop, K. S. Walton, *Mol. Simulat.* **2013**, *39*, 1253.
- [91] T. J. H. Vlugt, J. P. J. M. van der Eerden, M. Dijkstra, B. Smit, D. Frenkel, Introduction to Molecular Simulation and Statistical Thermodynamics, **2008**, <http://homepage.tudelft.nl/v9k6y/imsst/index.html>.
- [92] K. Binder, *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, Oxford University Press, Oxford **1995**.
- [93] K. Binder, D. W. Heermann, *Monte Carlo Simulation in Statistical Physics: An Introduction*, 5th ed., Springer, **2010**.
- [94] D. P. Landau, K. Binder, *A Guide to Monte-Carlo Simulations in Statistical Physics*, 3rd ed., Cambridge University Press, Cambridge **2009**.
- [95] H. Yu, W. F. van Gunsteren, *Comput. Phys. Commun.* **2005**, *172*, 69.
- [96] P. Cieplak, F.-Y. Dupradeau, Y. Duan, J. Wang, *J. Phys.: Condens. Matter* **2009**, *21*, 333102.
- [97] Y. Shi, P. Ren, M. Schnieders, J.-P. Piquemal, in *Rev. Comput. Chem.*, Vol. 28 (Eds.: A. L. Parrill, K. B. Lipkowitz), Wiley Online Library, New Jersey, U. S. A. **2015**, p. 51.
- [98] P. K. Weiner, P. A. Kollman, *J. Comp. Chem.* **1981**, *2*, 287.
- [99] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta Jr, P. K. Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765.
- [100] S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *J. Comp. Chem.* **1986**, *7*, 230.
- [101] D. A. Pearlman, D. A. Case, J. C. Caldwell, G. L. Seibel, U. C. Singh, P. Weiner, P. A. Kollman, *AMBER 4.0*, University of California Press, San Francisco **1991**.
- [102] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- [103] R. Salomon-Ferrer, D. A. Case, R. C. Walker, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198.
- [104] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696.
- [105] L. P. Wang, K. A. McKiernan, J. Gomes, K. A. Beauchamp, T. Head-Gordon, J. E. Rice, W. C. Swope, T. J. Martinez, V. S. Pande, *J. Phys. Chem. B* **2017**, *121*, 4023.
- [106] W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657.
- [107] G. Kaminski, E. M. Duffy, T. Matsui, W. L. Jorgensen, *J. Phys. Chem.* **1994**, *98*, 13077.
- [108] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- [109] W. Damm, A. Frontera, J. Tirado-Rives, W. L. Jorgensen, *J. Comp. Chem.* **1997**, *18*, 1955.
- [110] C. Zhang, C. Lu, Z. Jing, C. Wu, J.-P. Piquemal, J. W. Ponder, P. Ren, *J. Chem. Theory Comput.* **2018**, *14*, 2084.
- [111] J. R. Maple, U. Dinur, A. T. Hagler, *Proc. Nat. Acad. Sci.* **1988**, *85*, 5350.
- [112] M.-J. Hwang, T. P. Stockfisch, A. T. Hagler, *J. Am. Chem. Soc.* **1994**, *116*, 195.
- [113] A. T. Hagler, C. S. Ewig, *Comp. Phys. Comm.* **1994**, *84*, 131.
- [114] J. R. Maple, M.-J. Hwang, T. P. Stockfisch, A. T. Hagler, *Israel J. Chem.* **1994**, *34*, 195.
- [115] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187.
- [116] F. A. Momany, R. Rone, *J. Comp. Chem.* **1992**, *13*, 888.
- [117] J. J. Pavelites, J. Gao, P. A. Bash, A. D. Mackerell Jr, *J. Comp. Chem.* **1997**, *18*, 221.
- [118] A. D. Mackerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. M. S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
- [119] H. Sun, P. Ren, J. R. Fried, *Comput. Theor. Polym. Sci.* **1998**, *8*, 229.
- [120] H. Sun, *J. Phys. Chem. B* **1998**, *102*, 7338.
- [121] P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, A. T. Hagler, *Proteins* **1988**, *4*, 31.
- [122] M. Swart, P. T. van Duijnen, *Mol. Simulat.* **2007**, *32*, 471.
- [123] Y. A. Arnautova, A. Jagielska, H. A. Scheraga, *J. Phys. Chem.* **2006**, *110*, 5025.
- [124] S. Shi, L. Yan, Y. Yang, J. Fisher-Shaulsky, T. Thacher, *J. Comput. Chem.* **2003**, *24*, 1059.
- [125] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157.
- [126] J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, J. P. M. Postma, *Biopolymers* **1984**, *23*, 1.
- [127] K.-H. Ott, B. Meyer, *J. Comp. Chem.* **1996**, *17*, 1068.
- [128] C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **2004**, *25*, 1656.
- [129] H. Heinz, T.-J. Lin, R. K. Mishra, F. S. Emami, *Langmuir* **2013**, *29*, 1754.
- [130] W. L. Jorgensen, J. D. Madura, C. J. Swenson, *J. Am. Chem. Soc.* **1984**, *106*, 6638.
- [131] W. L. Jorgensen, C. J. Swenson, *J. Am. Chem. Soc.* **1985**, *107*, 569.
- [132] W. L. Jorgensen, *J. Phys. Chem.* **1986**, *90*, 6379.
- [133] W. L. Jorgensen, *J. Phys. Chem.* **1988**, *90*, 1276.
- [134] H. Sun, S. J. Mumby, J. R. Maple, A. T. Hagler, *J. Am. Chem. Soc.* **1994**, *116*, 2978.
- [135] W. Xie, J. Pu, A. D. Mackerell, Jr, J. Gao, *J. Chem. Theory Comput.* **2007**, *3*, 1878.
- [136] N. L. Allinger, *J. Am. Chem. Soc.* **1977**, *99*, 8127.
- [137] N. L. Allinger, R. Kok, M. R. Imam, *J. Comp. Chem.* **1988**, *9*, 591.
- [138] J.-H. Lii, S. Gallion, C. Bender, H. Wikstrom, N. L. Allinger, K. M. Flurchick, M. M. Teeter, *J. Comp. Chem.* **1989**, *10*, 503.
- [139] J.-H. Lii, N. L. Allinger, *J. Comp. Chem.* **1991**, *12*, 186.
- [140] J.-H. Lii, N. L. Allinger, *J. Comp. Chem.* **1998**, *19*, 1001.
- [141] N. L. Allinger, K. Chen, J.-H. Lii, *J. Comp. Chem.* **1996**, *17*, 642.
- [142] N. L. Allinger, K. Chen, J. A. Katzenellenbogen, S. R. Wilson, G. M. Anstead, *J. Comp. Chem.* **1996**, *17*, 747.
- [143] N. Nevens, J.-H. Lii, N. L. Allinger, *J. Comp. Chem.* **1996**, *17*, 695.
- [144] N. Nevens, K. Chen, N. L. Allinger, *J. Comp. Chem.* **1996**, *17*, 669.
- [145] N. Nevens, N. L. Allinger, *J. Comp. Chem.* **1996**, *17*, 730.
- [146] N. L. Allinger, Y. Fan, *J. Comp. Chem.* **1997**, *18*, 1827.
- [147] T. A. Halgren, *J. Comp. Chem.* **1996**, *17*, 490.
- [148] T. A. Halgren, *J. Comp. Chem.* **1996**, *17*, 520.
- [149] T. A. Halgren, *J. Comp. Chem.* **1996**, *17*, 553.
- [150] T. A. Halgren, R. B. Nachbar, *J. Comp. Chem.* **1996**, *17*, 616.
- [151] J. M. Hermida-Ramón, S. Brdarski, G. Karlström, U. Berg, *J. Comput. Chem.* **2003**, *24*, 161.
- [152] N. Gresh, G. A. Cisneros, T. A. Darden, J.-P. Piquemal, *J. Chem. Theory Comput.* **2007**, *4*, 1960.
- [153] M. G. Martin, J. I. Siepmann, *J. Phys. Chem. B* **1998**, *102*, 2569.
- [154] M. G. Martin, J. I. Siepmann, *J. Phys. Chem. B* **1999**, *103*, 4508.
- [155] B. Chen, J. I. Siepmann, *J. Phys. Chem. B* **1999**, *103*, 5370.
- [156] C. D. Wick, M. G. Martin, J. I. Siepmann, *J. Phys. Chem. B* **2000**, *104*, 8008.
- [157] J. J. Potoff, J. I. Siepmann, *AIChE J.* **2001**, *47*, 1676.
- [158] J. M. Stubbs, J. J. Potoff, J. I. Siepmann, *J. Phys. Chem. B* **2004**, *108*, 17596.
- [159] N. Lubna, G. Kamath, J. J. Potoff, N. Rai, J. I. Siepmann, *J. Phys. Chem. B* **2005**, *109*, 24100.
- [160] J. Gao, *J. Phys. Chem.* **1997**, *101*, 657.
- [161] A. C. T. Van Duin, S. Dasgupta, F. Lorant, W. A. Goddard, *J. Phys. Chem. A* **2001**, *105*, 9396.

- [162] J. D. Gale, A. L. Rohl, *Mol. Simulat.* **2003**, 29, 291.
- [163] D. Dubbeldam, R. Krishna, R. Q. Snurr, *J. Phys. Chem. C* **2009**, 113, 19317.
- [164] E. B. Wilson, Jr., J. D. Decious, P. C. Cross, *Molecular Vibrations*, McGraw-Hill, New York **1955**.
- [165] H. B. Thompson, *J. Chem. Phys.* **1967**, 47, 3407.
- [166] I. H. Williams, *J. Mol. Spectrosc.* **1977**, 66, 288.
- [167] D. F. McIntosh, K. H. Michaelian, M. R. Peterson, *Can. J. Chem.* **1978**, 56, 1289.
- [168] K. J. Miller, R. J. Hinde, J. Anderson, *J. Comput. Chem.* **1989**, 10, 63.
- [169] T. Schlick, *J. Comput. Chem.* **1989**, 10, 951.
- [170] W. C. Swope, D. M. Ferguson, *J. Comput. Chem.* **1992**, 13, 585.
- [171] H. Bekker, H. J. C. Berendsen, W. F. Van Gunsteren, *J. Comput. Chem.* **1995**, 16, 527.
- [172] A. Blondel, M. Karplus, *J. Comput. Chem.* **1996**, 17, 1132.
- [173] S. H. Lee, K. Palmo, S. Krimm, *J. Comput. Chem.* **1999**, 20, 1067.
- [174] R. E. Tuzun, D. W. Noid, B. G. Sumpter, *J. Comput. Chem.* **2000**, 21, 553.
- [175] C. Lavor, *Physica D* **2007**, 227, 135.
- [176] C. Wade, MSc thesis, *Gdansk University of Technology*, Gdansk, Poland **2015**.
- [177] M. Krishnan, S. Balasubramanian, *Phys. Rev. B* **2003**, 68, 064304.
- [178] B. A. Wells, A. L. Chaffee, *J. Chem. Theory Comput.* **2015**, 11, 3684.
- [179] J. D. Gale, *Philos. Mag. B-Phys. Condens. Matter Stat. Mech. Electron. Opt. Magn. Prop.* **1996**, 73, 3.
- [180] M. Totrov, R. Abagyan, *J. Comput. Chem.* **1994**, 15, 1105.
- [181] J. F. Lutsko, *J. Appl. Phys.* **1989**, 65, 2991.
- [182] K. van Workum, K. Yoshimoto, J. J. de Pablo, J. F. Douglas, *Phys. Rev. E* **2005**, 71, 061102.
- [183] J. A. Israelachvili, *Intermolecular and Surface Forces*, 3rd ed., Elsevier, Amsterdam, The Netherlands **2011**.
- [184] J. V. Barth, *Annu. Rev. Phys. Chem.* **2007**, 58, 375.
- [185] J. R. Ray, *Comput. Phys. Rep.* **1988**, 8, 111.
- [186] W. H. Baur, R. X. Fischer, *Chem. Mat.* **2019**, 31, 2401.
- [187] S. Toxvaerd, *J. Chem. Phys.* **1990**, 93, 4290.
- [188] K. van Workum, G. Gao, J. D. Schall, J. A. Harrison, *J. Chem. Phys.* **2006**, 125, 144506.
- [189] J. P. Ryckaert, A. Bellemans, *Faraday Discuss.* **1978**, 66, 95.
- [190] J. B. Nicholas, A. J. Hopfinger, F. R. Trouw, L. E. Iton, *J. Am. Chem. Soc.* **1991**, 113, 4792.
- [191] M. P. Howard, A. Statt, A. Z. Panagiotopoulos, *J. Chem. Phys.* **2017**, 146, 2.
- [192] R. E. Raab, O. L. de Lange, *Multipole Theory in Electromagnetism: Classical, Quantum, and Symmetry Aspects, with Applications*, Clarendon Press, Oxford **2005**.
- [193] L. Novotny, B. Hecht, *Principles of Nano-Optics*, 2nd ed., Cambridge University Press, New York **2012**.
- [194] A. J. Stone, *Theory of Intermolecular Forces*, Oxford University Press, New York **1997**.
- [195] D. G. Truhlar, *J. Chem. Educ.* **2019**, 96, 1671.
- [196] R. J. M. Pellenq, D. Nicholson, *Mol. Phys.* **1998**, 95, 549.
- [197] D. Nicholson, A. Boutin, R. J. M. Pellenq, *Mol. Simulat.* **1996**, 17, 217.
- [198] R. J. M. Pellenq, D. Nicholson, *J. Chem. Soc. Faraday T.* **1993**, 89, 2499.
- [199] C. R. A. Catlow, W. C. Mackrodt, *Lecture Notes in Physics*, Berlin **1982**.
- [200] D. Chakrabarti, D. J. Wales, *Phys. Chem. Chem. Phys.* **2009**, 11, 1970.
- [201] T. F. Miller, M. Eleftheriou, P. Pattnaik, A. Ndirango, D. Newns, G. J. Martyna, *J. Chem. Phys.* **2002**, 116, 8649.
- [202] G. J. Martyna, M. Tuckerman, D. J. Tobias, M. L. Klein, *Mol. Phys.* **1996**, 87, 1117.
- [203] M. E. Tuckerman, J. Alejandre, R. Lopez-Rendon, A. L. Jochim, G. J. Martyna, *J. Phys. A* **2006**, 39, 5629.
- [204] T.-Q. Yu, J. Alejandre, R. Lopez-Rendon, G. J. Martyna, M. E. Tuckerman, *Chem. Phys.* **2010**, 370, 294.
- [205] R. P. Feynman, A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York **1965**.
- [206] R. M. Stratt, *J. Chem. Phys.* **1979**, 70, 3630.
- [207] J. C. Liu, J. T. Culp, S. Natesakhawat, B. C. Bockrath, B. Zande, S. G. Sankar, G. Garberoglio, J. K. Johnson, *J. Phys. Chem. C* **2007**, 111, 9305.
- [208] W. E. Reiher III, PhD thesis, Harvard University, Cambridge, U. S. A., **1985**.
- [209] P. Gibbon, G. Sutmann, *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, Lecture Notes, Jülich **2002**, p. 467.
- [210] P. P. Ewald, *Ann. Phys.* **1921**, 369, 253.
- [211] D. J. Adams, *J. Chem. Phys.* **1983**, 78, 2585.
- [212] S. W. de Leeuw, J. W. Perram, E. R. Smith, *P. Roy. Soc. Lond. A Mat.* **1980**, 373, 27.
- [213] T. W. Nymand, P. Linse, *J. Chem. Phys.* **2000**, 112, 6152.
- [214] F. J. Vesely, *J. Comput. Phys.* **1977**, 24, 361.
- [215] P. Procacci, M. Marchi, G. J. Martyna, *J. Chem. Phys.* **1998**, 108, 8799.
- [216] S. W. Rick, S. J. Stuart, in *Rev. Comput. Chem.*, Vol. 18 (Eds.: K. B. Lipkowitz, D. B. Boyd), Wiley Online Library, New Jersey, U. S. A. **2002**, p. 89.
- [217] J. D. Gale, *Rev. Mineral. Geochem.* **2001**, 42, 37.
- [218] P. J. Mitchell, D. Fincham, *J. Phys.: Condens. Matter* **1993**, 5, 1031.
- [219] V. Lachet, A. Boutin, B. Tavittian, A. H. Fuchs, *J. Phys. Chem. B* **1998**, 102, 9224.
- [220] P. Cicu, P. Demontis, S. Spanu, G. B. Suffritti, A. Tilocca, *J. Chem. Phys.* **2000**, 112, 8267.
- [221] S. Bogusz, T. E. Cheatham, B. R. Brooks, *J. Chem. Phys.* **1998**, 108, 7070.
- [222] W. Smith, T. R. Forester, *J. Mol. Graphics* **1996**, 14, 136.
- [223] J. W. Perram, H. G. Petersen, S. W. Deleeuw, *Mol. Phys.* **1988**, 65, 875.
- [224] D. Fincham, *Mol. Simulat.* **1994**, 13, 1.
- [225] J. Kolafa, J. W. Perram, *Mol. Simulat.* **1992**, 9, 351.
- [226] B. Linse, P. Linse, *J. Chem. Phys.* **2014**, 141, 184114.
- [227] S. Plimpton, *J. Comput. Phys.* **1995**, 117, 1.
- [228] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, 103, 8577.
- [229] G. A. Cisneros, V. Babin, C. Sagui in *Electrostatics Interactions in Classical Simulations*, Springer, New York **2013**, p. 243.
- [230] D. Wolf, P. Keblinski, S. R. Phillpot, J. Eggebrecht, *J. Chem. Phys.* **1999**, 110, 8254.
- [231] A. Rahbari, R. Hens, S. H. Jamali, M. Ramdin, D. Dubbeldam, T. J. H. Vlugt, *Mol. Simulat.* **2018**, 45, 336.
- [232] T. J. H. Vlugt, M. Schenk, *J. Phys. Chem. B* **2002**, 106, 12757.
- [233] R. L. June, A. T. Bell, D. N. Theodorou, *J. Phys. Chem.* **1990**, 94, 8232.
- [234] R. Q. Snurr, R. L. June, A. T. Bell, D. N. Theodorou, *Mol. Simulat.* **1991**, 8, 73.
- [235] D. Dubbeldam, S. Calero, D. E. Ellis, R. Q. Snurr, *Mol. Simulat.* **2016**, 42, 81.
- [236] F. Lekien, J. Marsden, *Int. J. Numer. Meth. Engng* **2005**, 63, 455.
- [237] A. Torres-Knoop, S. P. Balaji, T. Vlugt, D. Dubbeldam, *J. Chem. Theor. Comp.* **2014**, 10, 942.
- [238] S. Amirjalayer, M. Tafipolsky, R. Schmid, *Angew. Chem. Int. Ed.* **2007**, 46, 463.
- [239] W. J. Mortier, S. K. Ghosh, S. Shankar, *J. Am. Chem. Soc.* **1986**, 108, 4315.
- [240] D. Ongari, P. G. Boyd, O. Kadioglu, A. K. Mace, S. Keskin, B. Smit, *J. Chem. Theory Comput.* **2019**, 15, 382.
- [241] C. E. Wilmer, R. Q. Snurr, *Chem. Eng. J.* **2011**, 171, 775.
- [242] C. E. Wilmer, K. C. Kim, R. Q. Snurr, *J. Phys. Chem. Lett.* **2012**, 3, 2506.

- [243] A. K. Rappe, W. A. Goddard, *J. Phys. Chem.* **1991**, *95*, 3358.
- [244] D. C. Wallace, *Thermodynamics of Crystals*, John Wiley & Sons, New York **1972**.
- [245] J. F. Nye, *Physical Properties of Crystals*, Oxford University Press, Oxford **1957**.
- [246] B. R. Brooks, D. Janezic, M. Karplus, *J. Comp. Chem.* **1995**, *16*, 1522.
- [247] F. Salles, A. Ghoufi, G. Maurin, R. G. Bell, C. Mellot-Draznieks, G. Férey, *Angew. Chem. Int. Ed.* **2008**, *47*, 8487.
- [248] J. Baker, *J. Comp. Chem.* **1986**, *7*, 385.
- [249] C. J. Cerjan, W. H. Miller, *J. Chem. Phys.* **1981**, *75*, 2800.
- [250] J. Simons, P. Jorgensen, H. Taylor, J. Ozment, *J. Phys. Chem.* **1983**, *87*, 2745.
- [251] Y. Alexeev, M. W. Schmidt, T. L. Windus, M. S. Gordon, *J. Comput. Chem.* **2007**, *28*, 1685.
- [252] Y. Liu, J.-H. Her, A. Dailly, A. J. Ramirez-Cuesta, D. A. Neumann, C. M. Brown, *J. Am. Chem. Soc.* **2008**, *130*, 11813.
- [253] N. Foloppe, A. D. MacKerell, *J. Comput. Chem.* **2000**, *21*, 86.
- [254] P. S. Nerenberg, T. Head-Gordon, *Curr. Opin. Struct. Biol.* **2018**, *49*, 129.
- [255] M. M. Francl, L. E. Chirlian, in *Rev. Comput. Chem.*, Vol. 14 (Eds.: K. B. Lipkowitz, D. B. Boyd), Wiley Online Library, New Jersey, U. S. A. **2000**, p. 1.
- [256] S. Hamad, S. R. G. Balestra, R. Bueno-Perez, S. Calero, A. R. Ruiz-Salvador, *J. Solid State Chem.* **2015**, *223*, 144.
- [257] R. S. Mulliken, *J. Chem. Phys.* **1955**, *23*, 1833.
- [258] A. E. Reed, R. B. Weinstock, F. Weinhold, *J. Chem. Phys.* **1985**, *83*, 735.
- [259] R. F. W. Bader, *Phys. Rev. B* **1994**, *49*, 13348.
- [260] P. J. Lowdin, *J. Chem. Phys.* **1950**, *18*, 365.
- [261] F. L. Hirshfeld, *Theoret. Chim. Acta* **1977**, *44*, 129.
- [262] P. Bultinck, C. Van Alsenoy, P. W. Ayers, R. Carbó-Dorca, *J. Chem. Phys.* **2007**, *126*, 144111.
- [263] R. F. W. Bader, *Acc. Chem. Res.* **1985**, *18*, 9.
- [264] D. E. P. Vanpoucke, P. Bultinck, I. Van Driessche, *J. Comput. Chem.* **2013**, *34*, 405.
- [265] T. Verstraelen, P. W. Ayers, V. Van Speybroeck, M. Waroquier, *J. Chem. Theory Comput.* **2013**, *9*, 2221.
- [266] T. A. Manz, D. S. Sholl, *J. Chem. Theory Comput.* **2010**, *6*, 2455.
- [267] T. Watanabe, T. A. Manz, D. S. Sholl, *J. Phys. Chem. C* **2011**, *115*, 4824.
- [268] T. A. Manz, D. S. Sholl, *J. Chem. Theory Comput.* **2012**, *8*, 2844.
- [269] T. C. Lillestolen, R. J. Wheatley, *Chem. Comm.* **2008**, *45*, 5909.
- [270] F. A. Momany, *J. Phys. Chem.* **1978**, *82*, 592.
- [271] S. R. Cox, D. E. Williams, *J. Comput. Chem.* **1981**, *2*, 304.
- [272] C. Campana, B. Mussard, T. K. Woo, *J. Chem. Theory Comput.* **2009**, *5*, 2866.
- [273] P. Zhang, P. Bao, J. Gao, *J. Comput. Chem.* **2011**, *32*, 2127.
- [274] B. T. Thole, P. T. van Duijnen, *Theoret. Chim. Acta* **1983**, *63*, 209.
- [275] A. Gabrieli, M. Sant, P. Demontis, G. B. Suffritti, *J. Chem. Theory Comput.* **2015**, *11*, 3829.
- [276] M. Sant, A. Gabrieli, P. Demontis, G. B. Suffritti, *Comput. Phys. Commun.* **2016**, *200*, 190.
- [277] W. J. Mortier, K. V. Genechten, J. Gasteiger, *J. Am. Chem. Soc.* **1985**, *107*, 829.
- [278] D. Ongari, P. G. Boyd, O. Kadioglu, A. K. Mace, S. Keskin, B. Smit, *J. Chem. Theory Comput.* **2018**, *15*, 382.
- [279] Q. Xu, C. L. Zhong, *J. Phys. Chem. C* **2010**, *114*, 5035.
- [280] M. V. Parkes, C. L. Staiger, J. J. Perry IV, M. D. Allendorf, J. A. Greathouse, *Phys. Chem. Chem. Phys.* **2013**, *15*, 9093.
- [281] D. Nazarian, J. S. Camp, D. S. Sholl, *Chem. Mater.* **2016**, *28*, 785.
- [282] C. Zheng, C. Zhong, *J. Phys. Chem. C* **2010**, *114*, 9945.
- [283] M. K. Rana, F. G. Pazzona, G. B. Suffritti, P. Demontis, M. Masia, *J. Chem. Theory Comput.* **2011**, *7*, 1575.
- [284] J. J. Gutierrez-Sevillano, S. Calero, C. O. Ania, J. B. Parra, F. Kapteijn, J. Gascon, S. Hamad, *J. Phys. Chem. C* **2013**, *117*, 466.
- [285] J. J. Wolffs, D. E. P. Vanpoucke, A. Sharma, K. V. Lawler, P. M. Forster, *Micropor. Mesopor. Mat.* **2019**, *277*, 184.
- [286] P. Bleiziffer, K. Schaller, S. Riniker, *J. Chem. Inf. Model.* **2018**, *58*, 579.
- [287] A. E. Sifain, N. Lubbers, B. T. Nebgen, J. S. Smith, A. Y. Likhov, O. Isayev, A. E. Roitberg, K. Barros, S. Tretiak, *J. Phys. Chem. Lett.* **2018**, *9*, 4495.
- [288] B. Nebgen, N. Lubbers, J. S. Smith, A. E. Sifain, A. Likhov, O. Isayev, A. E. Roitberg, K. Barros, S. Tretiak, *J. Chem. Theory Comput.* **2018**, *14*, 4687.
- [289] M. G. Martin, A. P. Thompson, T. M. Nenoff, *J. Chem. Phys.* **2001**, *114*, 7174.
- [290] A. Martin-Calvo, E. Garcia-Perez, A. Garcia-Sanchez, R. Bueno-Perez, S. Hamad, S. Calero, *Phys. Chem. Chem. Phys.* **2011**, *13*, 11165.
- [291] A. Martin-Calvo, F. D. Lahoz-Martin, S. Calero, *J. Phys. Chem. C* **2012**, *116*, 6655.
- [292] A. Garcia-Sanchez, C. O. Ania, J. B. Parra, D. Dubbeldam, T. J. H. Vlugt, R. Krishna, S. Calero, *J. Phys. Chem. C* **2009**, *113*, 8814.
- [293] I. Matito-Martos, A. Martin-Calvo, J. J. Gutierrez-Sevillano, M. Haranczyk, M. Doblare, J. B. Parra, C. O. Ania, S. Calero, *Phys. Chem. Chem. Phys.* **2014**, *16*, 19884.
- [294] I. Matito-Martos, J. Alvarez-Ossorio, J. J. Gutierrez-Sevillano, M. Doblare, A. Martin-Calvo, S. Calero, *Phys. Chem. Chem. Phys.* **2015**, *17*, 18121.
- [295] J. J. Gutierrez-Sevillano, A. Martin-Calvo, D. Dubbeldam, S. Calero, S. Hamad, *Rsc Advances* **2013**, *3*, 14737.
- [296] B. Liu, B. Smit, F. Rey, S. Valencia, S. Calero, *J. Phys. Chem. C* **2008**, *112*, 2492.
- [297] J. J. Gutierrez-Sevillano, D. Dubbeldam, F. Rey, S. Valencia, M. Palomino, A. Martin-Calvo, S. Calero, *J. Phys. Chem. C* **2010**, *114*, 14907.
- [298] N. Rai, J. I. Siepmann, *J. Phys. Chem. B* **2007**, *111*, 10790.
- [299] C. D. Wick, J. M. Stubbs, N. Rai, J. I. Siepmann, *J. Phys. Chem. B* **2005**, *109*, 18974.
- [300] M. S. Shah, J. I. Siepmann, M. Tsapatsis, *Aiche J.* **2017**, *63*, 5098.
- [301] P. Ungerer, C. Beauvais, J. Delhommelle, A. Boutin, B. Rousseau, A. H. Fuchs, *J. Chem. Phys.* **2000**, *112*, 5499.
- [302] A. Bondi, *J. Phys. Chem.* **1964**, *68*, 441.
- [303] T. van Westen, T. J. H. Vlugt, J. Gross, *J. Phys. Chem. B* **2011**, *115*, 7872.
- [304] J. Gross, G. Sadowski, *Ind. Eng. Chem. Res.* **2001**, *40*, 1244.
- [305] N. Karasawa, W. A. Goddard, *J. Phys. Chem.* **1989**, *93*, 7320.
- [306] Z. M. Chen, T. Cagin, W. A. Goddard, *J. Comput. Chem.* **1997**, *18*, 1365.
- [307] J. D. Gale, *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 629.
- [308] S. C. Tumble, J. Anwar, J. D. Gale, *J. Am. Chem. Soc.* **2004**, *126*, 396.
- [309] D. Dubbeldam, S. Calero, T. J. H. Vlugt, R. Krishna, T. L. M. Maesen, B. Smit, *J. Phys. Chem. B* **2004**, *108*, 12301.
- [310] T. J. H. Vlugt, R. Krishna, B. Smit, *J. Phys. Chem. B* **1999**, *103*, 1102.
- [311] D. Dubbeldam, S. Calero, T. J. H. Vlugt, R. Krishna, T. L. M. Maesen, E. Beerdsen, B. Smit, *Phys. Rev. Lett.* **2004**, *93*, 088302.
- [312] B. Smit, *J. Phys. Chem.* **1995**, *99*, 5597.
- [313] T. M. Becker, J. Heinen, D. Dubbeldam, L. C. Lin, T. J. H. Vlugt, *J. Phys. Chem. C* **2017**, *121*, 4659.
- [314] H. Fang, H. Demir, P. Kamakoti, D. S. Sholl, *J. Mater. Chem. A* **2014**, *2*, 274.
- [315] A. L. Dzubak, L.-C. Lin, J. Kim, J. A. Swisher, R. Poloni, S. N. Maximoff, B. Smit, L. Gagliardi, *Nat. Chem.* **2012**, *4*, 810.
- [316] E. Haldoupis, J. Borycz, H. L. Shi, K. D. Vogiatzis, P. Bai, W. L. Queen, L. Gagliardi, J. I. Siepmann, *J. Phys. Chem. C* **2015**, *119*, 16058.
- [317] T. Pham, K. A. Forrest, K. McLaughlin, J. Eckert, B. Space, *J. Phys. Chem. C* **2014**, *118*, 22683.

- [318] T. Pham, K. A. Forrest, R. Banerjee, G. Orcajo, J. Eckert, B. Space, *J. Phys. Chem. C* **2015**, *119*, 1078.
- [319] T. M. Becker, D. Dubbeldam, L. C. Lin, T. J. H. Vlugt, *J. Comput. Sci.* **2016**, *15*, 86.
- [320] T. M. Becker, A. Luna-Triguero, J. M. Vicent-Luna, L. C. Lin, D. Dubbeldam, S. Calero, T. J. H. Vlugt, *Phys. Chem. Chem. Phys.* **2018**, *20*, 28848.
- [321] Z. R. Herm, J. A. Swisher, B. Smit, R. Krishna, J. R. Long, *J. Am. Chem. Soc.* **2011**, *133*, 5664.
- [322] 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, *Chem. Sci.* **2014**, *5*, 4569.
- [323] D. Yu, A. O. Yazaydin, J. R. Lane, P. D. C. Dietzel, R. Q. Snurr, *Chem. Sci.* **2013**, *4*, 3544.
- [324] P. D. C. Dietzel, V. Besikiotis, R. Blom, *J. Mater. Chem.* **2009**, *19*, 7362.
- [325] R. Mercado, B. Vlasisavljevich, L.-C. Lin, K. Lee, Y. Lee, J. A. Mason, D. J. Xiao, M. I. Gonzalez, M. T. Kapelewski, J. B. Neaton, B. Smit, *J. Phys. Chem. C* **2016**, *120*, 12590.
- [326] F. Ercolessi, J. B. Adams, *Europhys. Lett.* **1994**, *26*, 583.
- [327] B. Waldher, J. Kuta, S. Chen, N. Henson, A. E. Clark, *J. Comput. Chem.* **2010**, *31*, 2307.
- [328] P. Brommer, A. Kiselev, D. Schopf, P. Beck, J. Roth, H. R. Trebin, *Model. Simul. Mater. Sci. Eng.* **2015**, *23*, 074002.
- [329] A. Stukowski, E. Fransson, M. Mock, P. Erhart, *Model. Simul. Mater. Sci. Eng.* **2017**, *25*, 055003.
- [330] R. Rodriguez-Fernandez, F. B. Pereira, J. M. C. Marques, E. Martinez-Nunez, S. A. Vazquez, *Comput. Phys. Commun.* **2017**, *217*, 89.
- [331] A. Jaramillo-Botero, S. Naserifar, W. A. Goddard, *J. Chem. Theory Comput.* **2014**, *10*, 1426.
- [332] A. I. Duff, M. W. Finnis, P. Maugis, B. J. Thijsse, M. H. F. Sluiter, *Comput. Phys. Commun.* **2015**, *196*, 439.
- [333] N. Artrith, A. Urban, *Comput. Mater. Sci.* **2016**, *114*, 135.
- [334] R. Judson, in *Rev. Comput. Chem.*, Vol. 10 (Eds.: K. B. Lipkowitz, D. B. Boyd), Wiley Online Library, New Jersey, USA **1997**, p. 1.
- [335] S. M. Woodley, P. D. Battle, J. D. Gale, C. R. A. Catlow, *Phys. Chem. Chem. Phys.* **1999**, *1*, 2535.
- [336] M. Tafipolsky, R. Schmid, *J. Phys. Chem. B* **2009**, *113*, 1341.
- [337] J. Heinen, N. C. Burtch, K. S. Walton, D. Dubbeldam, *J. Chem. Theory Comput.* **2017**, *13*, 3722.
- [338] J.-C. Tan, B. Civalleri, C.-C. Lin, L. Valenzano, R. Galvelis, P.-F. Chen, T. D. Bennett, C. Mellot-Draznieks, C. M. Zicovich-Wilson, A. K. Cheetham, *Phys. Rev. Lett.* **2012**, *108*, 095502.
- [339] M. R. Ryder, B. Civalleri, J.-C. Tan, *Phys. Chem. Chem. Phys.* **2016**, *18*, 9079.
- [340] A. U. Ortiz, A. Boutin, A. H. Fuchs, F.-X. Coudert, *Phys. Rev. Lett.* **2012**, *109*, 195502.
- [341] A. U. Ortiz, A. Boutin, A. H. Fuchs, F.-X. Coudert, *J. Chem. Phys.* **2013**, *138*, 174703.
- [342] Y. Z. S. Sun, R. F. DeJaco, J. I. Siepmann, *Chem. Sci.* **2019**, *10*, 4377.
- [343] J. P. Poort, M. Ramdin, J. van Kranendonk, T. J. H. Vlugt, *Fluid Phase Equilib.* **2019**, *490*, 39.
- [344] Z. Gong, Y. Z. Wu, L. Wu, H. Sun, *J. Chem. Inf. Model.* **2018**, *58*, 2502.
- [345] C. Desgranges, J. Delhommelle, *J. Chem. Phys.* **2018**, *149*, 10.
- [346] M. Gastegger, J. Behler, P. Marquetand, *Chem. Sci.* **2017**, *8*, 6924.
- [347] J. D. Evans, F.-X. Coudert, *Chem. Mater.* **2017**, *29*, 7833.
- [348] M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji, T. K. Woo, *J. Phys. Chem. Lett.* **2014**, *5*, 3056.
- [349] R. Anderson, J. Rodgers, E. Argueta, A. Biong, D. A. Gomez-Gualdron, *Chem. Mat.* **2018**, *30*, 6325.
- [350] H. Ohno, Y. Mukae, *J. Phys. Chem. C* **2016**, *120*, 23963.
- [351] M. Pardakhti, E. Moharrer, D. Wanik, S. L. Suib, R. Srivastava, *ACS Comb. Sci.* **2017**, *19*, 640.
- [352] G. Borboudakis, T. Stergiannakos, M. Frysali, E. Klontzas, I. Tsamardinos, G. E. Froudakis, *Npj Comput. Mater.* **2017**, *3*, 40.
- [353] H. Dureckova, M. Krykunov, M. Z. Aghaji, T. Woo, *J. Phys. Chem. C* **2019**, *123*, 4133.
- [354] W. Y. Yang, H. Liang, F. Peng, Z. L. Liu, J. Liu, Z. W. Qiao, *Nanomaterials* **2019**, *9*, 13.
- [355] Y. P. He, E. D. Cubuk, M. D. Allendorf, E. J. Reed, *J. Phys. Chem. Lett.* **2018**, *9*, 4562.
- [356] J. Behler, *J. Chem. Phys.* **2016**, *145*, 170901.
- [357] H. Chan, B. Narayanan, M. J. Cherukara, F. G. Sen, K. Sasikumar, S. K. Gray, M. K. Y. Chan, S. K. R. S. Sankaranarayanan, *J. Phys. Chem. C* **2019**, *123*, 6941.
- [358] Y. F. Huang, J. Kang, W. A. Goddard, L. W. Wang, *Phys. Rev. B* **2019**, *99*, 064103.
- [359] B. Huang, O. A. von Lilienfeld, *J. Chem. Phys.* **2016**, *145*, 161102.
- [360] P. Narwelkar, H. S. Lee, S. H. Xie, W. Im, *Biophys. J.* **2019**, *116*, 289.
- [361] V. Botu, R. Ramprasad, *Int. J. Quantum Chem.* **2015**, *115*, 1074.
- [362] V. Botu, R. Ramprasad, *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92*, 094306.
- [363] V. Botu, R. Batra, J. Chapman, R. Ramprasad, *J. Phys. Chem. C* **2017**, *121*, 511.
- [364] Z. W. Li, J. R. Kermode, A. De Vita, *Phys. Rev. Lett.* **2015**, *114*, 096405.
- [365] A. Khorshidi, A. A. Peterson, *Comp. Phys. Commun.* **2016**, *207*, 310.
- [366] H. Wang, L. Zhang, J. Han, W. E., *Comp. Phys. Commun.* **2018**, *228*, 178.
- [367] L. F. Zhang, D. Y. Lin, H. Wang, R. Car, W. N. E., *Phys. Rev. Mater.* **2019**, *3*, 023804.
- [368] A. Drozdetskiy, C. Cole, J. Procter, G. J. Barton, *Nucleic Acids Res.* **2015**, *43*, 389.
- [369] L. J. Colwell, *Curr. Opin. Struct. Biol.* **2018**, *49*, 123.
- [370] S. R. Tangadpalliwar, S. Vishwakarma, R. Nimbalkar, P. Garg, *Chem. Biol. Drug Des.* **2019**, *93*, 960.
- [371] K. T. Schutt, H. E. Saucedo, P. J. Kindermans, A. Tkatchenko, K. R. Muller, *J. Chem. Phys.* **2018**, *148*, 241722.
- [372] B. G. Dick, A. W. Overhauser, *Phys. Rev.* **1958**, *112*, 90.
- [373] C. R. A. Catlow, A. N. Cormack, F. Theobald, *Acta Cryst. B* **1984**, *40*, 195.
- [374] M. Sierka, J. Sauer, *Faraday Discuss.* **1997**, *106*, 41.
- [375] C. R. A. Catlow, J. M. Thomas, C. M. Freeman, P. A. Wright, R. G. Bell, *P. R. Soc. A* **1993**, *442*, 85.
- [376] G. Sastre, J. D. Gale, *Chem. Mater.* **2003**, *15*, 1788.
- [377] G. Sastre, J. D. Gale, *Chem. Mater.* **2005**, *17*, 730.
- [378] S. R. G. Balestra, S. Hamad, A. R. Ruiz-Salvador, V. D. Garcia, P. J. Merklings, D. Dubbeldam, S. Calero, *Chem. Mater.* **2015**, *27*, 5657.
- [379] P. Demontis, G. B. Suffritti, S. Quartieri, E. S. Fois, A. Gamba, *Zeolites* **1987**, *7*, 522.
- [380] P. Demontis, G. B. Suffritti, S. Quartieri, E. S. Fois, A. Gamba, *J. Phys. Chem. Chem.* **1988**, *92*, 867.
- [381] P. Demontis, G. B. Suffritti, E. S. Fois, A. Gamba, G. Morosi, *Mater. Chem. Phys.* **1991**, *29*, 357.
- [382] B. W. H. Van Beest, G. J. Kramer, R. A. Van Santen, *Phys. Rev. Lett.* **1990**, *64*, 1955.
- [383] G. J. Kramer, N. P. Farragher, B. W. H. Van Beest, R. A. Van Santen, *Phys. Rev. B* **1991**, *43*, 5068.
- [384] G. Schrimpf, M. Schlenkrich, J. Brickmann, P. Bopp, *J. Phys. Chem.* **1992**, *96*, 7404.
- [385] K. S. Smirnov, D. Bougeard, *J. Raman Spectrosc.* **1993**, *24*, 255.
- [386] K. S. Smirnov, D. Bougeard, *J. Phys. Chem.* **1993**, *97*, 9434.
- [387] K. S. Smirnov, D. Bougeard, *Zeolites* **1994**, *14*, 203.
- [388] K. S. Smirnov, M. Le Maire, C. Brémard, D. Bougeard, *Chem. Phys.* **1994**, *179*, 445.
- [389] R. Bueno-Pérez, S. Calero, D. Dubbeldam, A. Conchi, J. Parra, P. Zaderenko, P. Merklings, *J. Phys. Chem. C* **2012**, *116*, 25797.
- [390] J. R. Hill, J. Sauer, *J. Phys. Chem.* **1994**, *98*, 1238.

- [391] J. R. Hill, J. Sauer, *J. Phys. Chem.* **1995**, *99*, 9536.
- [392] K. Yamahara, K. Okazaki, K. Kawamura, *Catal. today* **1995**, *23*, 397.
- [393] P. Bordat, P.-A. Cazade, I. Baraille, R. Brown, *J. Chem. Phys.* **2010**, *132*, 094501.
- [394] G. Sastre, A. Pulido, A. Corma, *Micropor. Mesopor. Mat.* **2005**, *82*, 159.
- [395] R. T. Rigo, S. R. G. Balestra, S. Hamad, R. Bueno-Perez, A. R. Ruiz-Salvador, S. Calero, M. A. Cambor, *J. Mater. Chem. A* **2018**, *6*, 15110.
- [396] S. M. Auerbach, N. J. Henson, A. K. Cheetham, H. I. Metiu, *J. Phys. Chem.* **1995**, *99*, 10600.
- [397] E. Jaramillo, S. M. Auerbach, *J. Phys. Chem. B* **1999**, *103*, 9589.
- [398] S. Calero, D. Dubbeldam, R. Krishna, B. Smit, T. J. H. Vlugt, J. F. M. Denayer, J. A. Martens, T. L. M. Maesen, *J. Am. Chem. Soc.* **2004**, *126*, 11377.
- [399] M. Jeffroy, C. Nieto-Draghi, A. Boutin, *Mol. Simulat.* **2014**, *40*, 6.
- [400] H. Shi, A. N. Miguez, S. M. Auerbach, *Green Chem.* **2014**, *16*, 875.
- [401] A. Pedone, G. Malavasi, M. C. Menziani, A. N. Cormack, U. Segre, *J. Phys. Chem. B* **2006**, *110*, 11780.
- [402] A. Gabrieli, M. Sant, P. Demontis, G. B. Suffritti, *J. Phys. Chem. C* **2013**, *117*, 503.
- [403] M. Praprotnik, S. Hocevar, M. Hodoscek, M. Penca, D. Janezic, *J. Comput. Chem.* **2008**, *29*, 122.
- [404] P. M. Zimmerman, M. Head-Gordon, A. T. Bell, *J. Chem. Theory Comput.* **2011**, *7*, 1695.
- [405] S. K. Sahoo, N. N. Nair, *J. Comput. Chem.* **2015**, *36*, 1562.
- [406] J. Guo, K. D. Hammond, *J. Phys. Chem. C* **2018**, *122*, 11345.
- [407] S. E. Boulfelfel, P. I. Ravikovitch, L. Koziol, D. S. Sholl, *J. Phys. Chem. C* **2016**, *120*, 14140.
- [408] J. A. Greathouse, M. D. Allendorf, *J. Am. Chem. Soc.* **2006**, *128*, 10678.
- [409] N. C. Burtch, H. Jasuja, K. S. Walton, *Chem. Rev.* **2014**, *114*, 10575.
- [410] J. A. Greathouse, M. D. Allendorf, *J. Phys. Chem. C* **2008**, *112*, 5795.
- [411] D. Dubbeldam, K. S. Walton, D. E. Ellis, R. Q. Snurr, *Angew. Chem. Int. Ed.* **2007**, *46*, 4496.
- [412] S. S. Han, W. A. Goddard, *J. Phys. Chem. C* **2007**, *111*, 15185.
- [413] M. Tafipolsky, S. Amirjalayer, R. Schmid, *J. Comput. Chem.* **2007**, *28*, 1169.
- [414] W. Zhou, H. Wu, T. Yildirim, J. R. Simpson, A. R. H. Walker, *Phys. Rev. B* **2008**, *78*, 054114.
- [415] S. R. G. Balestra, R. Bueno-Perez, S. Hamad, D. Dubbeldam, A. R. Ruiz-Salvador, S. Calero, *Chem. Mater.* **2016**, *28*, 8296.
- [416] B. L. Huang, A. J. H. McGaughey, M. Kaviani, *Int. J. Heat Mass Transf.* **2007**, *50*, 393.
- [417] R. Schmid, M. Tafipolsky, *J. Am. Chem. Soc.* **2008**, *130*, 12600.
- [418] S. Amirjalayer, R. Q. Snurr, R. Schmid, *J. Phys. Chem.* **2012**, *116*, 4921.
- [419] D. Dubbeldam, R. Krishna, R. Q. Snurr, *J. Phys. Chem. C* **2009**, *113*, 19317.
- [420] D. C. Ford, D. Dubbeldam, R. Q. Snurr, *Diffusion Fundamentals III* **2009**, *11*, 1.
- [421] S. Devautour-Vinot, G. Maurin, F. Henn, C. Serre, T. Devic, G. Férey, *Chem. Comm.* **2009**, *19*, 2733.
- [422] D. S. Coombes, F. Cora, C. Mellot-Draznieks, R. G. Bell, *J. Phys. Chem. C* **2009**, *113*, 544.
- [423] A. M. Walker, B. Civalieri, B. Slater, C. Mellot-Draznieks, F. Cora, C. M. Zicovich-Wilson, G. Roman-Perez, J. M. Soler, J. D. Gale, *Angew. Chem.-Int. Edit.* **2010**, *49*, 7501.
- [424] F.-X. Coudert, M. Jeffroy, A. H. Fuchs, A. Boutin, C. Mellot-Draznieks, *J. Am. Chem. Soc.* **2008**, *130*, 14294.
- [425] L. Vanduyfhuys, A. Ghysels, S. M. J. Rogge, R. Demuyne, V. Van Speybroeck, *Mol. Simulat.* **2015**, *41*, 1311.
- [426] L. Vanduyfhuys, T. Verstraelen, M. Vandichel, M. Waroquier, V. van Speybroeck, *J. Chem. Theory Comput.* **2012**, *8*, 3217.
- [427] A. Luna-Triguero, J. M. Vicent-Luna, S. Calero, *Chem.: Eur. J.* **2018**, *24*, 8530.
- [428] L. Hamon, H. Leclerc, A. Ghoufi, L. Oliviero, A. Travert, J. C. Lavalley, T. Devic, C. Serre, G. Férey, G. de Weireld, A. Vimont, G. Maurin, *J. Phys. Chem. C* **2011**, *115*, 2047.
- [429] P. G. Yot, Q. Ma, J. Haines, Q. Yang, A. Ghoufi, T. Devic, C. Serre, V. Dmitriev, G. Férey, C. Zhong, *Chem. Sci.* **2012**, *3*, 1100.
- [430] J. Wieme, L. Vanduyfhuys, S. M. J. Rogge, M. Waroquier, V. Van Speybroeck, *J. Phys. Chem. C* **2016**, *120*, 14934.
- [431] L. Sarkisov, R. L. Martin, M. Haranczyk, B. Smit, *J. Am. Chem. Soc.* **2014**, *136*, 2228.
- [432] D. Dybtsev, H. Chun, K. Kim, *Angew. Chem. Int. Ed.* **2004**, *43*, 5033.
- [433] J. S. Grosch, F. Paesani, *J. Am. Chem. Soc.* **2012**, *134*, 4207.
- [434] N. C. Burtch, K. S. Walton, *Acc. Chem. Res.* **2015**, *48*, 2850.
- [435] H. Jasuja, N. C. Burtch, Y.-G. Huang, Y. Cai, K. S. Walton, *Langmuir* **2013**, *29*, 633.
- [436] H. Jasuja, Y. Jiao, N. C. Burtch, Y.-G. Huang, K. S. Walton, *Langmuir* **2014**, *30*, 14300.
- [437] N. C. Burtch, A. Torres-Knoop, G. S. Foo, J. Leisen, C. Sievers, B. Ensing, D. Dubbeldam, K. S. Walton, *J. Phys. Chem. Lett.* **2015**, *6*, 812.
- [438] N. C. Burtch, D. Dubbeldam, K. S. Walton, *Mol. Simulat.* **2015**, *41*, 1379.
- [439] S. S.-Y. Chui, S. M.-F. Lo, J. P. H. Charmant, A. G. Orpen, I. D. Williams, *Science* **1999**, *283*, 1148.
- [440] A. O. Yazaydin, A. I. Benin, S. A. Faheem, P. Jakubczak, J. J. Low, R. R. Willis, R. Q. Snurr, *Chem. Mater.* **2009**, *21*, 1425.
- [441] M. Tafipolsky, S. Amirjalayer, R. Schmid, *J. Phys. Chem. C* **2010**, *114*, 14402.
- [442] L. Zhao, Q. Yang, Q. Ma, C. Zhong, J. Mi, D. Liu, *J. Mol. Model.* **2011**, *17*, 227.
- [443] M. I. H. Mohideen, B. Xiao, P. S. Wheatley, A. C. McKinlay, Y. Li, A. M. Z. Slawin, D. W. Aldous, N. F. Cessford, T. Düren, X. Zhao, *Nat. Chem.* **2011**, *3*, 304.
- [444] A. Slawek, J. M. Vicent-Luna, B. Marszalek, B. Gil, R. E. Morris, W. Makowski, S. Calero, *Chem. Mater.* **2018**, *30*, 5116.
- [445] J. P. Durholt, R. Galvelis, R. Schmid, *Dalton Trans.* **2016**, *45*, 4370.
- [446] J. H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga, K. P. Lillerud, *J. Am. Chem. Soc.* **2008**, *130*, 13850.
- [447] P. S. Bárcia, D. Guimarães, P. A. P. Mendes, J. A. C. Silva, V. Guillerme, H. Chevreau, C. Serre, A. E. Rodrigues, *Micropor. Mesopor. Mat.* **2011**, *139*, 67.
- [448] M. A. Moreira, J. C. Santos, A. F. P. Ferreira, J. Loureiro, F. Ragon, P. Horcajada, K.-E. Shim, Y.-K. Hwang, U.-H. Lee, J.-S. Chang, C. Serre, *Langmuir* **2012**, *28*, 5715.
- [449] Q. Yang, A. D. Wiersum, H. Jobic, V. Guillerme, C. Serre, P. Llewellyn, G. Maurin, *J. Phys. Chem. C* **2011**, *115*, 13768.
- [450] J. K. Bristow, D. Tiana, A. Walsh, *J. Chem. Theory Comput.* **2014**, *10*, 4644.
- [451] T. Chokbunpiam, S. Fritzsche, C. Chmelik, J. Caro, W. Janke, S. Hannongbua, *Chem. Phys. Lett.* **2016**, *648*, 178.
- [452] Y. Du, K. Mao, B. Wooler, A. K. Sharma, D. Colmyer, M. Nines, S. C. Weston, *J. Phys. Chem. C* **2017**, *121*, 28090.
- [453] B. Zheng, M. Sant, P. Demontis, G. B. Suffritti, *J. Phys. Chem. C* **2012**, *116*, 933.
- [454] Z. Q. Hu, L. L. Zhang, J. W. Jiang, *J. Chem. Phys.* **2012**, *136*, 244703.
- [455] E. Pantatosaki, G. Megariotis, A. K. Pusch, C. Chmelik, F. Stallmach, G. K. Papadopoulos, *J. Phys. Chem. C* **2012**, *116*, 201.
- [456] R. J. Verploegh, A. Kulkarni, S. E. Boulfelfel, J. C. Haydak, D. Tang, D. S. Sholl, *J. Phys. Chem. C* **2019**, *123*, 9153.
- [457] T. Weng, J. R. Schmidt, *J. Phys. Chem. A* **2019**, *123*, 3000.
- [458] J. P. Durholt, G. Fraux, F.-X. Coudert, R. Schmid, *J. Chem. Theory Comput.* **2019**, *15*, 2420.

- [459] S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy, R. Schmid, *Physica Status Solidi (b)* **2013**, 250, 1128.
- [460] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, V. Van Speybroeck, *J. Comput. Chem.* **2015**, 36, 1015.
- [461] L. Vanduyfhuys, S. Vandenbrande, J. Wieme, M. Waroquier, T. Verstraelen, V. Van Speybroeck, *J. Comput. Chem.* **2018**, 39, 999.
- [462] M. A. Addicoat, N. Vankova, I. F. Akter, T. Heine, *J. Chem. Theory Comput.* **2014**, 10, 880.
- [463] D. E. Coupry, M. A. Addicoat, T. Heine, *J. Chem. Theory Comput.* **2016**, 12, 5215.
- [464] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungeleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, R. Q. Snurr, *Chem. Mater.* **2014**, 26, 6185.
- [465] J. K. Bristow, J. M. Skelton, K. L. Svane, A. Walsh, J. D. Gale, *Phys. Chem. Chem. Phys.* **2016**, 18, 29316.
- [466] D. S. Sholl, R. P. Lively, *J. Phys. Chem. Lett.* **2015**, 6, 3437.
- [467] L. Sarkisov, *Dalton Trans.* **2016**, 45, 4203.
- [468] T. C. Germann, K. Kadau, *J. Mod. Phys. C* **2008**, 19, 1315.
- [469] G. Fraux, S. Chibani, F.-X. Coudert, *Phil. Trans. R. Soc. A* **2019**, 377, 20180220.
- [470] X. J. Wu, S. C. Xiang, J. Q. Su, W. Q. Cai, *J. Phys. Chem. C* **2019**, 123, 8550.
- [471] Z. Gulsoy, K. B. Sezginel, A. Uzun, S. Keskin, R. Yildirim, *ACS Comb. Sci.* **2019**, 21, 257.
- [472] E. Haldoupis, S. Nair, D. S. Sholl, *Phys. Chem. Chem. Phys.* **2011**, 13, 5053.
- [473] Y. J. Colon, R. Q. Snurr, *Chem. Soc. Rev.* **2014**, 43, 5735.
- [474] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, R. Q. Snurr, *Nat. Chem.* **2012**, 4, 83.
- [475] J. Keupp, R. Schmid, *Faraday Discuss.* **2018**, 211, 79.
- [476] D. C. Lonie, E. Zurek, *Comput. Phys. Commun.* **2012**, 183, 690.
- [477] S. Barthel, E. V. Alexandrov, D. M. Proserpio, B. Smit, *Cryst. Growth Des.* **2018**, 18, 1738.
- [478] D. Dokur, S. Keskin, *Ind. Eng. Chem. Res.* **2018**, 57, 2298.
- [479] J. E. Stone, D. J. Hardy, I. S. Ufimtsev, K. Schulten, *J. Mol. Graph.* **2010**, 29, 116.
- [480] M. J. Harvey, G. D. Fabritiis, *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2012**, 2, 734.
- [481] A. W. Gotz, M. J. Williamson, D. Xu, D. Poole, S. L. Grand, R. C. Walker, *J. Chem. Theory Comput.* **2012**, 8, 1542.
- [482] R. Salomon-Ferrer, A. W. Gotz, D. Poole, S. L. Grand, R. C. Walker, *J. Chem. Theory Comput.* **2013**, 9, 3878.
- [483] T. S. Lee, D. S. Cerutti, D. Mermelstein, C. Lin, S. LeGrand, T. J. Giese, A. Roitberg, D. A. Case, R. C. Walker, D. M. York, *J. Chem Inf. Model.* **2018**, 58, 2043.
- [484] S. Lindert, D. Bucher, P. Eastman, V. Pande, J. A. McCammon, *J. Chem. Theory Comput.* **2013**, 9, 4684.
- [485] A. Jasz, A. Rak, I. Ladjanski, G. Cserey, *J. Mol. Struct.* **2019**, 1188, 227.
- [486] P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L. P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts, V. S. Pande, *J. Chem. Theory Comput.* **2013**, 9, 461.
- [487] Y. Nejahi, M. S. Barhaghi, J. Mick, B. Jackman, K. Rushaidat, Y. Z. Li, L. Schwiebert, J. Potoff, *SoftwareX* **2019**, 9, 20.
- [488] P. Eastman, V. S. Pande, *J. Comput. Chem.* **2010**, 31, 1268.
- [489] K. Rushaidat, L. Schwiebert, B. Jackman, J. Mick, J. Potoff, *Proceedings of the Conference on Summer Computer Simulation*, San Diego, CA, U. S. A. **2015**, p. 1.
- [490] M. J. Harvey, G. D. Fabritiis, *J. Chem. Theory Comput.* **2009**, 5, 2371.
- [491] D. J. Hardy, J. E. Stone, K. Schulten, *Parallel Comput.* **2009**, 35, 164.
- [492] R. Anandakrishnan, T. R. W. Scogland, A. T. Fenley, J. C. Gordon, W. C. Feng, A. V. Onufriev, *J. Mol. Graph.* **2010**, 28, 904.
- [493] A. S. Minkin, A. A. Knizhnik, B. V. Potapkin, *Adv. Eng. Softw.* **2017**, 111, 43.
- [494] Z. Y. Fan, W. Chen, V. Vierimaa, A. Harju, *Comput. Phys. Commun.* **2017**, 218, 10.
- [495] M. Zheng, X. X. Li, L. Guo, *J. Mol. Graph.* **2013**, 41, 1.
- [496] T. D. Nguyen, C. L. Phillips, J. A. Anderson, S. C. Glotzer, *Comput. Phys. Commun.* **2011**, 182, 2307.
- [497] J. Kim, J. M. Rodgers, M. Athenes, B. Smit, *J. Chem. Theory Comput.* **2011**, 7, 3208.
- [498] J. Mick, E. Hailat, V. Russo, K. Rushaidat, L. Schwiebert, J. Potoff, *Comput. Phys. Commun.* **2013**, 184, 2662.
- [499] C. Hall, W. X. Ji, E. Blaisten-Barojas, *J. Comput. Phys.* **2014**, 258, 871.
- [500] E. Hailat, V. Russo, K. Rushaidat, J. Mick, L. Schwiebert, J. Potoff, *Int. J. Parallel Emergent Distrib. Syst.* **2014**, 29, 379.
- [501] Y. H. Liang, X. J. Xing, Y. H. Li, *J. Comput. Phys.* **2017**, 338, 252.
- [502] J. A. Anderson, C. D. Lorenz, A. Travesset, *J. Comput. Phys.* **2008**, 227, 5342.
- [503] M. S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A. L. Beberg, D. L. Ensign, C. M. Bruns, V. S. Pande, *J. Comput. Chem.* **2009**, 30, 864.
- [504] J. Glaser, T. D. Nguyen, J. A. Anderson, P. Lui, F. Spiga, J. A. Millan, D. C. Morse, S. C. Glotzer, *Comput. Phys. Commun.* **2015**, 192, 97.
- [505] B. G. Levine, J. E. Stone, A. Kohlmeier, *J. Comput. Phys.* **2011**, 230, 3556.
- [506] J. Kim, R. L. Martin, O. Rübél, M. Haranczyk, B. Smit, *J. Chem. Theory Comput.* **2012**, 8, 1684.
- [507] R. L. Martin, Prabhat, D. D. Donofrio, J. A. Sethian, M. Haranczyk, *Int. J. High Perform. Comput. Appl.* **2012**, 26, 347.
- [508] L. A. Baumes, F. Kruger, S. Jimenez, P. Collet, A. Corma, *Phys. Chem. Chem. Phys.* **2011**, 13, 4674.
- [509] D. Dubbeldam, S. Calero, T. J. H. Vlugt, *Mol. Simulat.* **2018**, 44, 653.
- [510] T. Narumi, T. Hamada, K. Nitadori, R. Sakamaki, K. Yasuoka, *Int. J. Comput. Methods* **2011**, 8, 561.
- [511] P. H. Colberg, F. Hofling, *Comput. Phys. Commun.* **2011**, 182, 1120.
- [512] S. Le Grand, A. W. Gotz, R. C. Walker, *Comput. Phys. Commun.* **2013**, 184, 374.
- [513] A. Zhmurov, K. Rybnikov, Y. Kholodov, V. Barsegov, *J. Phys. Chem. B* **2011**, 115, 5278.
- [514] H. Kargar, A. Minucmehr, A. Zolfaghari, *AIP Adv.* **2016**, 6, 7.
- [515] S. J. Plimpton, J. D. Gale, *Curr. Opin. Solid State Mat. Sci.* **2013**, 17, 271.
- [516] A. Singraber, J. Behler, C. Dellago, *J. Chem. Theory Comput.* **2019**, 15, 1827.
- [517] S. H. Jamali, L. Wolff, T. M. Becker, M. de Groen, M. Ramdin, R. Hartkamp, A. Bardow, T. J. H. Vlugt, O. A. Moutos, *J. Chem Inf. Model.* **2019**, 59, 1290.
- [518] M. T. Humbert, Y. Zhang, E. J. Maginn, *J. Chem Inf. Model.* **2019**, 59, 1301.
- [519] D. L. Mobley, C. C. Bannan, A. Rizzi, C. I. Bayly, J. D. Chodera, V. T. Lim, N. M. Lim, K. A. Beauchamp, D. R. Slochow, M. R. Shirts, *J. Chem. Theory Comput.* **2018**, 14, 6076.