From adsorptation to dissipation: insights from computer simulations of solid H2O and CO

Ferrari, B.C.

Chapter 2

Methods

This section covers a brief explanation of generic methodology used in different chapters of this thesis, including force fields, vibrational frequencies and molecular dynamics.

2.1 Force Fields

A force field is a vector field of non-contact forces acting on a particle at a certain position in space. Within theorectical chemistry, the term is used to refer to a potential that describes the forces and energies acting on a system of atomic particles. Various force fields have been used in the research carried out in this thesis. Although they are referred to as force fields, within theorectical chemistry the description focuses on terms contributing to the potential energy, which are traditionally physically motivated. Forces are naturally calculated by taking the negative gradient of the potential,

$$\vec{F} = -\vec{\nabla}V \tag{2.1}$$

where $\vec{F}$ is the force and V is the potential energy. The force fields used for the research within this thesis are explained within this section, along with a few general considerations.

2.1.1 General Considerations

The evaluation of a force field can be done with a distance-based cutoff, since at large distances interactions are negligible and ignoring them saves computational time. Additionally, to replicate bulk systems, force fields can be evaluated under periodic boundary conditions. In practice both of these techniques require careful implementation to avoid introducing artifacts in the potential energy of the system.

Smoothing Functions

In order to take the gradient of the potential, the potential must be differentiable (*ie.* smooth and continuous). A distance based cutoff for interactions would result in a non-smooth or continuous potential. As such, a smoothing function (also called switching function) needs to be introduced.

A smoothing function smoothly truncates the interaction energy at a given cutoff value. There are many smoothing functions available, an example is,

$$S(r) = \begin{cases} 1 & r \leq r_s \\ 2(\frac{r-r_s}{r_c-r_s})^3 - 3(\frac{r-r_s}{r_c-r_s})^2 + 1 & r_s < r \leq r_c \\ 0 & r > r_c \end{cases} \tag{2.2}$$

where r is the interaction distance, r_s is the distance at which smoothing starts, and r_c is the interaction cutoff. This smoothing function is then multiplied with the interaction energy to give the smoothed interaction potential. An important caveat from smoothing the interaction energy is that upon calculating the gradient, this will also result in additional terms for the forces. In the simplest case of an interaction potential depending only on distances, this results in:

$$-\frac{\partial}{\partial r} S(r)V(r) = -\frac{\partial S(r)}{\partial r} V(r) - S(r)\frac{\partial V(r)}{\partial r} \tag{2.3}$$

Without properly including the addtionaly forces terms the energy and forces will be divergent, which prevents geometry optimizations from succeeding and causes the total energy to not be conserved in NVE molecular dynamics simulations (*vide infra*).

Smoothing functions can be applied at the atomic level, or at the molecular level. Although both methods are possible it is important to note that frequently force fields are pair potentials, meaning the interaction energy is fitted for a pair interaction. In these cases applying a smoothing function at the atomic level will introduce artifacts, notably due to imbalances of the pair potential interactions. For these cases the smoothing function must be applied to all pair interactions based on the same pair distance (either the center of mass or a particular atom).

Periodic Boundary Conditions

To approximate the structure of bulk systems it is common to apply periodic boundary conditions (PBCs) to all particles within a unit cell. In Figure 2.1 an example of a

small unit cell under PBCs is shown.

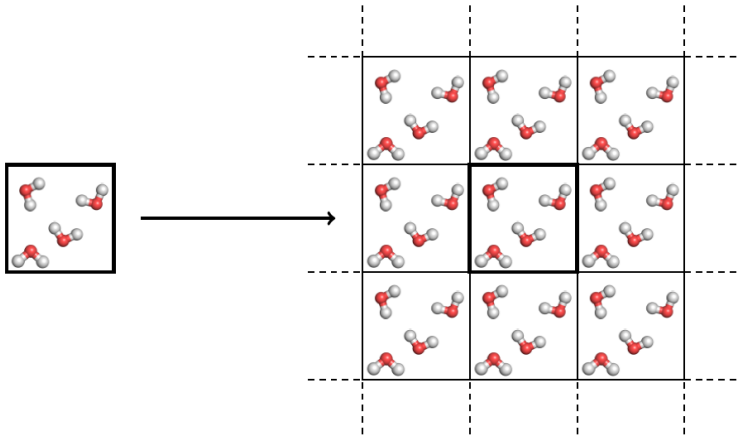

Figure 2.1: Illustration of how PBCs is used to approximate bulk structures. The prominent black box around 4 water molecules is the unit cell. Each cell surrounding the original unit cell contains images of the molecules within the unit cell.

An important consideration to avoid artifacts is that the total energy of the system should be consistent with the number of particles (or molecules) within the unit cell. As such, if all interactions between particles and images are counted, then interactions with images and particles can only add half their energy to the total energy of the system. Similarly, self-interactions (*ie.* a molecule interacting with its image) are forbidden, which arises from the extra total energy the system would have if these interactions are counted. A popular method for applying PBCs without violating these above stipulations is the minimum image convention (MIC).

Under the MIC only interactions between real particles and the closest representation (an image or real particle) of other particles is counted. In order to ensure that self-interactions are ignored under this convention, it is typically required that an interaction cutoff smaller than half the shortest cell length is used. To be clear, this means that two particles on opposite ends of a unit cell will only interact via the images that are closest to the other real particle.

2.1.2 Force Fields for H_2O

There exits a wide variety of force fields describing water, ranging from simple point charge models (SPC[1], SPC/E[2], SPC-F[3], and TIP4P[4]) to highly sophisticated many-body polarizable models (TTMx-F[5–10], POLI2VS[11], AMOEBA[12–17], SCME/f[18–21], and MB-pol[22–25]). The work in this thesis utilizes two of the highly sophisticated many-body polarizable models, MB-pol & SCME/f, which are briefly explained here.

MB–pol

MB-pol[22–25] is derived from the many-body expansion (MBE) of the interaction energy between water molecules. The MBE for a system of N molecules is given by,

$$V = \sum_{i=1}^{N} V_{1B}(i) + \sum_{i<j}^{N} V_{2B}(i,j) + \ldots + V_{NB}(1,\ldots,N) \tag{2.4}$$

where V_{1B} is the one-body potential, V_{2B} is the two-body potential, and V_{NB} is the N-body potential. The i and j indexes run over all atoms in each molecule for all pairs of molecules in the system. The MB-pol potential only extends to the three-body term in the MBE, and includes addtional dispersion and electrostatic terms. The total MB-pol energy is given by,

$$V = V_{1B} + V_{2B} + V_{3B} + V_{\text{Dispersion}} + V_{\text{Electrostatics}} \tag{2.5}$$

where $V_{\text{Dispersion}}$ is the dispersion interaction energy, and $V_{\text{Electrostatics}}$ is the electrostatic interaction energy.

The one-body term in MB-pol is taken from Partridge and Schwenke[26], which is the gold standard of one-body water potentials. The two-body potential is given by,

$$V_{2B} = \sum_{i<j}^{N} S(r_{O_i,O_j}) V_{2B-PIP}(X_i, X_j) \tag{2.6}$$

where N is the number of molecules, r_{O_i,O_j} is the distance between the ith and jth oxygen atom positions, X_i and X_j are the ith and jth molecule positions (all atoms plus two lone-pair sites), S is the switching function, and V_{2B-PIP} is the two-body PIP potential. Figure 2.2 shows the interaction sites of the MB-pol pair potential.

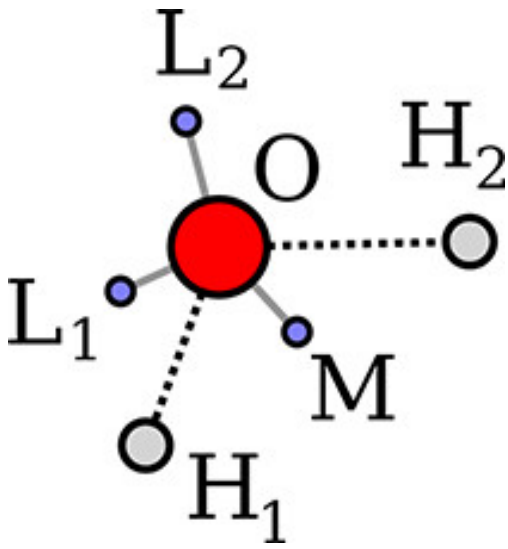

Figure 2.2: Interaction sites of MB-pol pair potential. Figure taken from Babin *et al.* [22].

The dispersion interaction is also a two-body interaction within MB-pol given by,

$$V_{\text{Dispersion}} = \sum_{i<j}^{N} f_6(r_{ij}) \frac{C_6^{(ij)}}{r_{ij}^6} + f_8(r_{ij}) \frac{C_8^{(ij)}}{r_{ij}^8} \tag{2.7}$$

where f_n are the Tang-Toennies damping functions [27], C_n are constants, and r_{ij} is the distance between the ith and jth atoms.

The electrostatic potential, within the MB-pol(2023) model, is composed of four terms; the charge–charge interactions (V_{qq}), charge–dipole interactions ($V_{q\mu}$), dipole–dipole interactions ($V_{\mu\mu}$), and the many-body polarization (V_{pol}).

$$V_{qq} = \sum_{b>a}^{n} q_a \hat{T}_{ab} q_b \tag{2.8}$$

$$V_{q\mu} = \sum_{b>a}^{n} (\mu_a^{\alpha} \hat{T}_{ab}^{\alpha} q_b - q_a \hat{T}_{ab}^{\alpha} \mu_b^{\alpha}) \tag{2.9}$$

$$V_{\mu\mu} = -\sum_{b>a}^{n} \mu_a^{\alpha} \hat{T}_{ab}^{\alpha\beta} \mu_b^{\beta} \tag{2.10}$$

$$V_{\text{pol}} = \frac{1}{2} \sum_{a=1}^{n} \mu_a \hat{\alpha}_a^{-1} \mu_a \tag{2.11}$$

In the above equations for the electrostatic terms n is the total number of electrostatic sites, q_a is the charge of site a, μ_a is the dipole of site a, $\hat{\alpha}_a$ is the polarizability of site a, Greek superscripts define Cartesian coordinates where Einstein notation is used for repeated superscripts, and $\hat{T}$ terms are the electrostatic tensors. The tensors are given by,

$$\hat{T}_{ab} = S_0(R_{ab}) \frac{1}{R_{ab}} \tag{2.12}$$

$$\hat{T}_{ab}^{\alpha} = \nabla_{\alpha} \hat{T}_{ab} = -S_1(R_{ab}) \frac{R_{ab}^{\alpha}}{R_{ab}^3} \tag{2.13}$$

$$\hat{T}_{ab}^{\alpha\beta} = \nabla_{\alpha} \hat{T}_{ab}^{\beta} = S_2(R_{ab}) \frac{3R_{ab}^{\alpha} R_{ab}^{\beta}}{R_{ab}^5} - S_1(R_{ab}) \frac{\delta_{ab}}{R_{ab}^3} \tag{2.14}$$

where R_{ab} is the distance between sites a and b, S_n are the screening functions, and δ is the Kronecker delta. The screening functions are given by,

$$S_0(R_{ab}) = 1 - \exp\left(-\omega(\frac{R_{ab}}{A})^4\right) + \frac{\omega^{\frac{1}{4}} R_{ab}}{A} \Gamma\left(\frac{3}{4}, \omega(\frac{R_{ab}}{A})^4\right) \tag{2.15}$$

$$S_1(R_{ab}) = 1 - \exp\left(-\omega(\frac{R_{ab}}{A})^4\right) \tag{2.16}$$

$$S_2(R_{ab}) = S_1(R_{ab}) - \frac{4\omega}{3} (\frac{R_{ab}}{A})^4 \exp\left(-\omega(\frac{R_{ab}}{A})^4\right) \tag{2.17}$$

where ω is the Thole damping factor, $A = (\alpha_a \alpha_b)^{\frac{1}{6}}$, and α is the polarizability.

SCME/f

The flexible single-center multipole expansion (SCME/f)[18–21] potential is another many-body water potential. The total interaction energy is given by,

$$V = V_{1B} + V_{2B} + V_{\text{Dispersion}} + V_{\text{Electrostatics}} \tag{2.18}$$

where V_{1B} is the one-body potential, V_{2B} is the two-body potential, $V_{\text{Dispersion}}$ is the dispersion interaction energy, and $V_{\text{Electrostatics}}$ is the electrostatic interaction energy. The one-body potential is identical to the one used in MB-pol (described above).

The two-body potential employs a Born-Mayer potential,

$$V_{2B} = \sum_{i<j} A r_{ij}^{-k} e^{-h r_{ij}} \tag{2.19}$$

where r_{ij} is the distance between the ith and jth oxygen atoms, and A, k, & h are constants which can be found in Jónsson *et al.*[21].

The dispersion energy is similar to that of MB-pol but with an additional term.

$$V_{\text{Dispersion}} = \sum_{i<j}^{N} f_6(r_{ij}) \frac{C_6^{(ij)}}{r_{ij}^6} + f_8(r_{ij}) \frac{C_8^{(ij)}}{r_{ij}^8} + f_{10}(r_{ij}) \frac{C_{10}^{(ij)}}{r_{ij}^{10}} \tag{2.20}$$

Again, here f_n are the Tang-Toennies damping functions[27], C_n are constants, and r_{ij} is the distance between the ith and jth atoms. The i and j indicies run over all atoms in each molecule for all pairs of molecules in the system. The electrostatic energy is given by,

$$V_{\text{electrostatic}} = V_{\text{in+pol}} + V_{\text{self}} \tag{2.21}$$

where $V_{\text{in+pol}}$ is the combination of the electrostatic interaction between intrinsic molecular moments and the field-induced polarization energy, and V_{self} is the self-energy. The self-energy is the energy required to polarize a molecule and is given by,

$$V_{\text{self}} = -\frac{1}{2} \sum_{i}^{N} (\Delta\mu_\alpha^i V_\alpha^i + \frac{1}{3} \Delta\theta_{\alpha\beta}^i) V_{\alpha\beta} \tag{2.22}$$

where Einstein notation is used for the Greek indicies, $\Delta\mu$ is the change in the dipole moment due to the field-induced polarization, and $\Delta\theta$ is the change in the quadrupole

moment due to the field-induced polarization. The other term is given by,

$$V_{\text{in+pol}} = \frac{1}{2}\sum_{i}^{N}[(\mu_\alpha^i(r^{ia}) + \Delta\mu_\alpha^i)V_\alpha^i + \frac{1}{3}(\theta_{\alpha\beta}^i(r^{ia}) + \Delta\theta_{\alpha\beta}^i)V_{\alpha\beta} + \frac{1}{15}\Omega_{\alpha\beta\gamma}^i V_{\alpha\beta\gamma}^i + \frac{1}{105}\Phi_{\alpha\beta\gamma\delta}^i V_{\alpha\beta\gamma\delta}^i] \quad (2.23)$$

where Ω is the octupole, and Φ is the hexadecapole.

2.1.3 Force Fields for CO

Compared to water there are significantly fewer force fields developed for studying carbon monoxide. The generic COMPASS force field was parameterized for 14 inorganic molecules, one of them being CO.[28] Beyond generic potentials being fitted to CO, there are two potentials focusing specifically on CO. The first is a site-site pair potential[29], and the other a neural network potential based on permutation-invariant polynomials (PIP–NN)[30].

Site-Site Pair Potential

This potential was originally developed by van Hemert *et al.*[29], where it models CO interactions as a site-site pair potential that was parameterized using CCSD(T) calculations for the CO dimer with the aug-cc-pVQZ basis set in combination with the Boys-Bernardi counterpoise correction. The values for all parameters in the force field are summarized in Table 2.1.

	$V_{\mathbf{Morse}}$		
	r_e (Å)	D_e (eV)	γ (Å^{-1})
	1.1282	11.23	2.3281
	$V_{\mathbf{exch}} + V_{\mathbf{disp}}$		
i,j	A_{ij} (eV)	B_{ij} (Å^{-1})	C_{ij} (eVÅ^6)
C,C	361.36	2.836	33.37
O,O	6370.10	4.253	10.52
C,O	1516.74	3.544	15.16
	$V_{\mathbf{el}}$		
i	Q_i^0 (e)	σ_i (Å^{-1})	
C	-0.47	3.845	
O	-0.615	2.132	

Table 2.1: CO-CO Potential Parameters

The interaction potential is composed of an intramolecular Morse potential (V_{Morse}), and intermolecular contributions that are intended to capture exchange (V_{exch}), dispersion (V_{disp}), and electrostatic (V_{el}) interactions between pairs of CO molecules:

$$V = V_{\text{Morse}} + V_{\text{exch}} + V_{\text{disp}} + V_{\text{el}} \tag{2.24}$$

The Morse potential is given by

$$V_{\text{Morse}} = \sum_{i=1}^{N} D_{\text{e}}(1 - e^{-\gamma(r_i - r_{\text{e}})})^2, \tag{2.25}$$

where D_{e} is the dissociation energy, r_i is the CO bond length of the ith molecule, r_{e} is the equilibrium CO bond length, N is the number of molecules in the system, and γ a constant given in Table 2.1. All of the Morse potential parameters were obtained from fitting to experimental data. A Buckingham potential is used for the exchange and dispersion contributions.

$$V_{\text{exch}} + V_{\text{disp}} = \sum_{ij \mid i<j} A_{ij} e^{-B_{ij} r_{ij}} - \frac{C_{ij}}{r_{ij}^6} \tag{2.26}$$

Here, r_{ij} is the distance between the ith and jth atoms, and all other terms (A_{ij}, B_{ij}, C_{ij}) are constants, given in Table 2.1. Note, i and j are atoms of different molecules. Finally, the electrostatic contribution is given by

$$V_{\text{el}} = \sum_{ij \mid i<j} \frac{1}{4\pi\epsilon_0} \frac{Q_i Q_j}{r_{ij}}, \tag{2.27}$$

where ϵ_0 is the vaccum dielectric constant, and the point charges Q_i and Q_j are each located on a different molecule at a distance r_{ij} with respect to each other. The C and the O atom form negative charge centers. A compensating total positive charge, $Q = -(Q_C + Q_O)$, is placed on the center of mass of each molecule, resulting in 9 Coulomb interactions between each pair of CO molecules. The charges mimic the *ab-initio* derived dipole and quadrupole moments. The moments were originally calculated using MCSCF/CCI calculations with the aug-cc-pVQZ basis set [29] and the coordinate-dependence of the charges is given by,

$$Q_i = Q_i^0 e^{\sigma_i(r_i - r_e)} \tag{2.28}$$

where r_i is the intramolecular CO distance, Q_i^0 is the charge distribution at the equilibrium bond length and σ_i is a constant, given in Table 2.1.

PIP–NN

This potential was originally developed by Chen *et al.*[30], and unlike potentials described by physically based interaction equations, PIP–NN potentials do not have parameters that can be directly linked to physical or chemical properties.

The total energy V is represented by the sum of the intramolecular V_{intra} and intermolecular V_{inter} potential energies.

$$V = V_{\text{intra}} + V_{\text{inter}} \tag{2.29}$$

The intramolecular energy is given by,

$$V_{\text{intra}} = b_2 + \vec{w}_2 \cdot \tanh(\vec{b}_1 + \vec{w}_1 x) \tag{2.30}$$

where b_2 & $\vec{b}_1$ are biases, $\vec{w}_1$ & $\vec{w}_2$ are weights and x is the bond length.[1] The intermolecular potential is a 2-layer PIP–NN given by,

$$V_{\text{inter}} = b_5 + \vec{w}_3 \cdot \tanh(\vec{b}_4 + W_2 \tanh(\vec{b}_3 + W_1 \vec{G})) \tag{2.31}$$

where $\vec{b}_3$, $\vec{b}_4$ & b_5 are biases, $\vec{w}_3$ is a weight, W_1 & W_2 are also weights but capitalized to denote matrices, and $\vec{G}$ is the symmetry functions. The symmetry functions are permutation invariant polynomials related to the distances between pairs of CO molecules.

$$\begin{aligned}
G_1 &= p_{O1,C1} + p_{O2,C2} \\
G_2 &= p_{O1,C2} + p_{O2,C1} \\
G_3^2 &= p_{O1,C1}^2 + p_{O2,C2}^2 \\
G_4^2 &= p_{O1,C2}^2 + p_{O2,C1}^2 \\
G_5^2 &= p_{O1,C1} p_{O1,C2} + p_{O2,C2} p_{O2,C1} \\
G_6 &= p_{O1,O2} \\
G_7 &= p_{C1,C2}
\end{aligned}$$

The above 7 functions are all the symmetry functions for this potential, with p_{ij} given

[1] tanh applied on a vector is done by applying tanh element-wise.

by,

$$p_{ij} = \exp(-\lambda r_{ij}) \tag{2.32}$$

where $\lambda = 0.3$[30] is an empirical parameter, and r_{ij} is the distance between atoms i and j.

An implementation of this potential in Julia can be found in YASS, which was developed during this PhD.

2.2 Geometry Optimization

Computing the optimized geometry, i.e., the atomic structure that minimizes the energy, of a system is often the first step in molecular simulations. For vibrational analysis this step ensures that the "imaginary modes", i.e., modes that represent motion toward a minimum on the systems potential energy surface, are not calculated. For molecular dynamics simulations, this step provides a stable starting point for simulations, which avoids unwanted simulation results.

Geometry optimization is done by using mathematical optimization techniques on the energy of the system as function of the $3N$ atomic coordinates of the system ($E(r_1, \ldots, r_{3N})$). The solution is one where the derivative of the energy with respect to all coordinates is zero ($\frac{\partial E}{\partial r_i} = 0$), and the Hessian matrix ($\frac{\partial^2 E}{\partial r_i \partial r_j}$) is positive semi-definite, i.e. all of its eigenvalues are non-negative. In computer simulations finding the solution is done iteratively, where several algorithms exist but the general concepts are consistent. Here I cover the gradient descent optimizer, also known as steepest descent, which serves as a simple model for explaining the concepts of geometry optimization.

Gradient descent assumes the objective function, the function being minimized, is defined and differentiable which is always the case for the energy of a system. It then follows that the object function decreases the fastest when traversing in the negative gradient of the function. In simulations this is done iteratively by traveling in steps along the negative gradient, typically via

$$R_{n+1} = R_n - \eta \nabla E(R) \tag{2.33}$$

where η is the step size, R is the $3N$ atomic coordinates and $E(R)$ the energy (or objective function). This is repeated until either reaching the maximum number of iterations or a convergence threshold on either the energy, gradient of the energy or the change in coordinates. Other more sophisticated optimization algorithms are typically used, however, the core concept of iteratively minimizing the energy until convergence

is consistent.

An important consideration here is that the optimization algorithm finds the nearest local minima, not the global minima. This means the optimized geometry found is dependent on the initial geometry supplied to the optimization routine. Algorithms that seek to find the global minia also exist but are more computationally demanding.

2.3 Vibrational Analysis

Vibrational frequency calculations are of great interest since vibrational properties of solids are accessible through different experimental techniques. Here I briefly describe two commonly used computational methods to calculate vibrational density of states (VDOS) that can be used with the force fields presented in the previous section.

2.3.1 Harmonic Approximation

To better understand the harmonic approximation it is beneficial to first briefly cover the physics of a much simpler system. Consider a frictionless system where a mass is attached to a massless spring attached to a wall (see Figure 2.3).

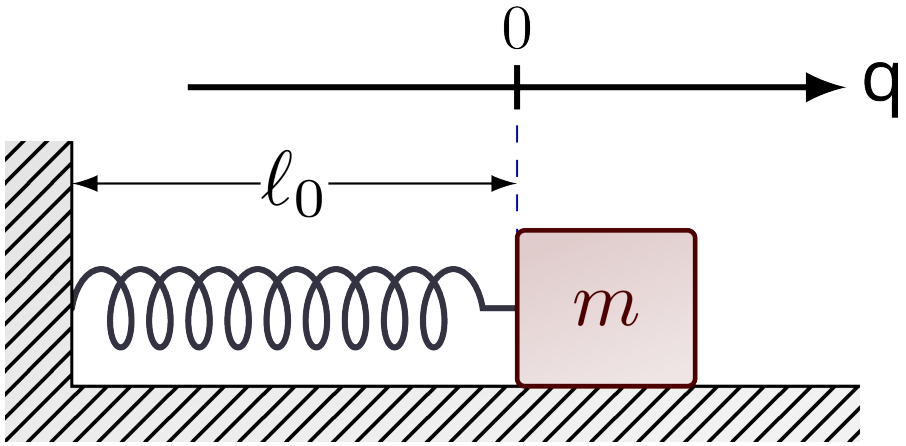

Figure 2.3: An illustration of a mass attached to a massless spring attached to a wall.

In order to simplify the problem we take the coordinate system where q is defined as the displacement of the mass (m) from its equilibrium position. The kinetic energy is then given by,

$$T = \frac{1}{2} m \dot{q}^2 \tag{2.34}$$

where T is kinetic energy, m is the mass, and $\dot{q}$ is the time derivative of the coordinate q. Assuming a spring constant of k, then the potential energy is as follows.

$$V = \frac{1}{2}kq^2 \tag{2.35}$$

The equations of motion for this system are then given by,

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}} + \frac{\partial V}{\partial q} = 0 \tag{2.36}$$

$$m\ddot{q} = -kq \tag{2.37}$$

where $\ddot{q}$ is the second time derivative of q. Solving this differential equation gives a description of the motion,

$$q(t) = A\sin(2\pi\sqrt{\frac{k}{m}}t + \phi) \tag{2.38}$$

where A is a constant describing the amplitude of the oscillations, ϕ is a constant descibing the phase of the oscillations, and the frequency of the oscillations is given by $\sqrt{\frac{k}{m}}$. This simple system is commonly referred to as a harmonic oscillator, and the solution to the differential equation is the normal mode of the system.

This same process can be applied to more complex systems to calculate the normal modes of the system. Consider a 3-dimensional system with N atoms at equilibrium. The problem can be simplified by changing the coordinates to mass weighted displacements of the atoms from equilibrium, ie., defining the coordinates $q_1 = \sqrt{m_1}\Delta x_1$, $q_2 = \sqrt{m_1}\Delta y_1$, ..., $q_{3N} = \sqrt{m_N}\Delta z_N$. The kinetic energy of the system is given by the following expression.

$$T = \sum_{i}^{3N} \dot{q}_i^2 \tag{2.39}$$

A generalized potential energy for the system, which is valid for any interaction potentials, is less straightforward. In order to achieve it, a Taylor series expansion of the potential is taken about the equilibrium position.

$$V = V_0 + \sum_{i}^{3N} \frac{\partial V}{\partial q_i}\bigg|_0 q_i + \sum_{i,j}^{3N} \frac{\partial^2 V}{\partial q_i \partial q_j}\bigg|_0 q_i q_j + \dots \tag{2.40}$$

The approximation in the term "harmonic approximation" comes in at this stage, where the vibrations are assumed to be small and so the potential energy is truncated

at the third term in the Taylor series expansion. The remaining expression is further simplified by defining that the potential energy at equilibrium is zero ($V_0 = 0$). Similarly, the gradient of the potential at equilibrium is zero ($\frac{\partial V}{\partial q_i}\Big|_0 = 0$), as defined by the optimization of the system (*ie.* the equilibrium point of a system is where the gradient of potential energy is zero). Resulting in the following potential energy for the system.

$$V = \sum_{i,j}^{3N} \frac{\partial V}{\partial q_i \partial q_j}\bigg|_0 q_i q_j = \sum_{i,j}^{3N} f_{ij} q_i q_j \tag{2.41}$$

Newton's equations of motion (one equation for each $j = 1, ..., 3N$ coordinate) for this system are given by the following expressions.

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_j} + \frac{\partial V}{\partial q_j} = 0 \tag{2.42}$$

$$\ddot{q}_j = -\sum_{i}^{3N} f_{ij} q_i \tag{2.43}$$

This gives $3N$ coupled differential equations, each of which bears a remarkable similarity to Equation (2.37). Analogous to the one-dimensional harmonic oscillator above, we make the following ansatz:

$$q_j(t) = A_j \sin(2\pi\sqrt{\lambda}t + \epsilon) \tag{2.44}$$

Here, A_j is the amplitude of the jth coordinate, λ is the frequency, and ϵ is the phase. To solve for the frequencies the ansatz is substituted into Equation (2.43).

$$-A_j\lambda \sin(\sqrt{\lambda}t + \epsilon) = -\sum_{i}^{3N} f_{ij} A_i \sin(\sqrt{\lambda}t + \epsilon) \tag{2.45}$$

$$-A_j\lambda = -\sum_{i}^{3N} f_{ij} A_i \tag{2.46}$$

$$\sum_{i}^{3N} f_{ij} A_i - A_j\lambda = 0 \tag{2.47}$$

$$\sum_{i}^{3N} (f_{ij} - \delta_{ij}\lambda) A_i = 0 \tag{2.48}$$

In Equation (2.48) δ_{ij} is the Kronecker delta. Only for certain values of λ does

Equation (2.48) have a non-trivial (*ie.* $A_i \neq 0$) solution. These particular values can be determined by solving the characteristic equation.

$$\begin{bmatrix} f_{11}-\lambda & f_{12} & \dots & f_{1,3N} \\ f_{21} & f_{22}-\lambda & \dots & f_{2,3N} \\ \dots & \dots & \dots & \dots \\ f_{3N,1} & \dots & \dots & f_{3N,3N}-\lambda \end{bmatrix} = 0 \tag{2.49}$$

In total there should be $3N$ values of λ that satisfy the non-trivial solution criteria. Some of these values may be identical to another value, these are called degenerate values. When V describes an internal interaction potential, $3N-6$ λ values correspond to the vibrational frequencies under the harmonic approximation, which are commonly referred to as normal mode frequencies. The exact solutions for the other 6 λ values is 0, and these solutions correspond to centre-of-mass translation and rotations of the entire system.

Figure 2.4 shows an example of the normal modes frequencies calculated for a bulk ice Ih cell with 768 molecules using the TIP4P/2005f potential. The histogram shows the OH stretching mode (3000 – 3400 cm^{-1}), the water bending mode ($\approx$ 1800 cm^{-1}), the water librational mode (500 – 1000 cm^{-1}), and the lattice modes ($<$ 500 cm^{-1}) of the hexagonal phase of ice (ice Ih).

Normal mode analysis results in $3N-6$ number of modes, where N is the number of atoms. When a large system is used and the modes are plotted as a histogram, the density of states becomes clear. For small systems the number of modes will not be large enough to produce a statistically converged sampling of the density of states of real systems. In computational practice, this method requires highly optimized systems, to avoid imaginary modes, which results in the frequencies being devoid of temperature information. In order to artificially smear the frequencies to model temperature effects the frequencies can be broadened with a Gaussian or Lorentzian lineshape. However, these artificial line broadening techniques do not capture the reallity of the thermal occupancy of the system, so care must be taken when using this technique.

2.3.2 Velocity Autocorrelation Function

Autocorrelation is the measure of correlation between a signal and a time-delayed copy of itself. This is used to identify periodic trends, even those obscured by noise, within timeseries. In atomistic simulations the VACF is a method for identifying

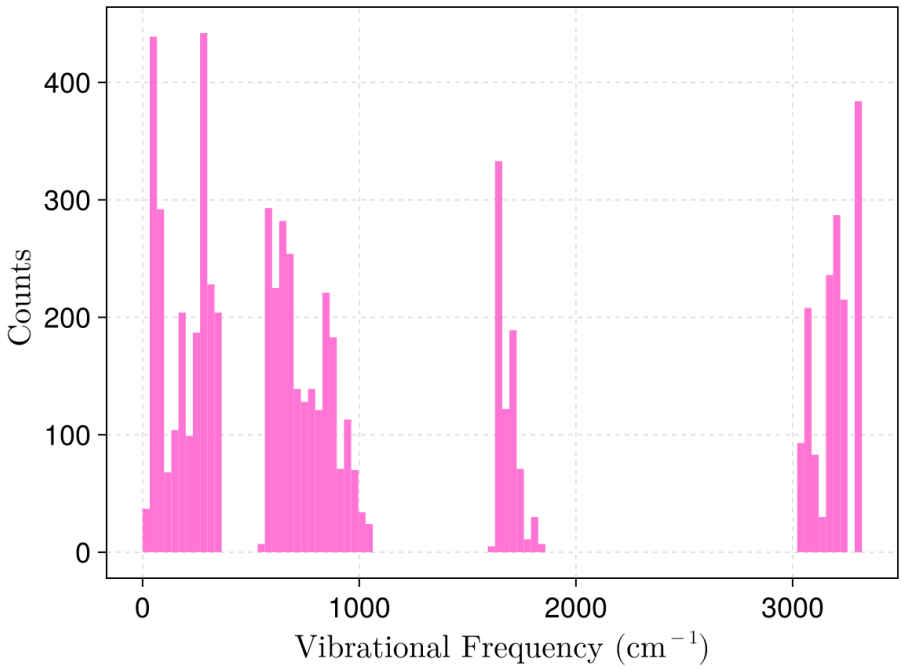

Figure 2.4: A histogram of the frequencies calculated from a normal mode analysis under the harmonic approximation. These frequencies are calculated for a bulk ice Ih cell using the TIP4P/2005f potential.

periodic motion of atoms (*ie.* vibrational modes). Taking the Fourier transform of the VACF gives the vibrational density of states (VDOS) of the system. For the sake of computational efficiency, the fast Fourier transform (FFT) is typically used. The VACF is given by,

$$< v(t) \cdot v(t - \Delta t) >= \frac{1}{M} \sum_{i}^{M} \frac{1}{N} \sum_{j}^{N} v_j(t_i) \cdot v_j(t_i - \Delta t) \tag{2.50}$$

where v is the velocity of an atom, M is the total number of timesteps, N is the total number of atoms, t is the time, and Δt is the timestep. In practice this expression is not used to calculate the VACF since it is slow, instead an FFT of the velocity timeseries is taken then multiplied by its conjugate and run through an inverse FFT. This is a significantly faster way of calculating the VACF, as long as the FFT algorithm is sufficiently fast. The majority of codes that calculate the VACF employ this technique.

The result of this is typically normalized by the total number of timesteps, resulting in a VDOS where the scale is proportional to the strength of the oscillation at a particular frequency.

Additional consideration also needs to be taken to avoid spectral leakage, a phenomenon in signal processing where the energy of a single frequency is spread across several frequency bins. Spectral leakage is caused by signals that are non-periodic within the time window. This will always be true when performing VACF analysis, but can be reduced by using a window function that smooths the signal edges. Some commonly used window functions are the Hann and Welch window functions.

The frequency resolution is determined via,

$$\Delta\omega = \frac{f_s}{cn} \tag{2.51}$$

where $\Delta\omega$ is the frequency resolution in wavenumber (cm^{-1}), f_s is the sampling frequency, c is the speed of light, and n is the number of timesteps in the trajectory. The maximum frequency is then given by,

$$\omega_{\max} = \Delta\omega\frac{n}{2} = \frac{f_s}{2c} \tag{2.52}$$

where the half factor comes from slicing the frequencies to be only those that are positive. From these equations it is clear that the frequency resolution and the largest frequency are determined by the MD timestep size and the length of the MD trajectory.

The velocitity data used in the VACF is taken from MD simulations, where the total number of timesteps determines the largest observable frequency and the size of the timestep the resolution of the VDOS. Additionally, the signal-to-noise of the VDOS can be improved by taking a long MD simulation (*ie.* 200 ps) slicing it into several timeseries of smaller duration (keeping in mind the length determines the highest frequency in the VDOS), then computing the VDOS of each timeseries and averaging them all.

Besides decreasing the size of the timestep in the MD simulation there are a few other tricks that can artificially increase the resolution of the calculated VDOS. One can mirror their VACF signal, pad the ends of the signal with zeros, or do both (note that zero padding must be done before mirroring to maintain symmetry). These tricks artificially increase the resolution of the VDOS, which means that small features might be blurred out despite the resolution improving.

Figure 2.5 shows the VDOS of a bulk Ih cell with 768 molecules calculated through

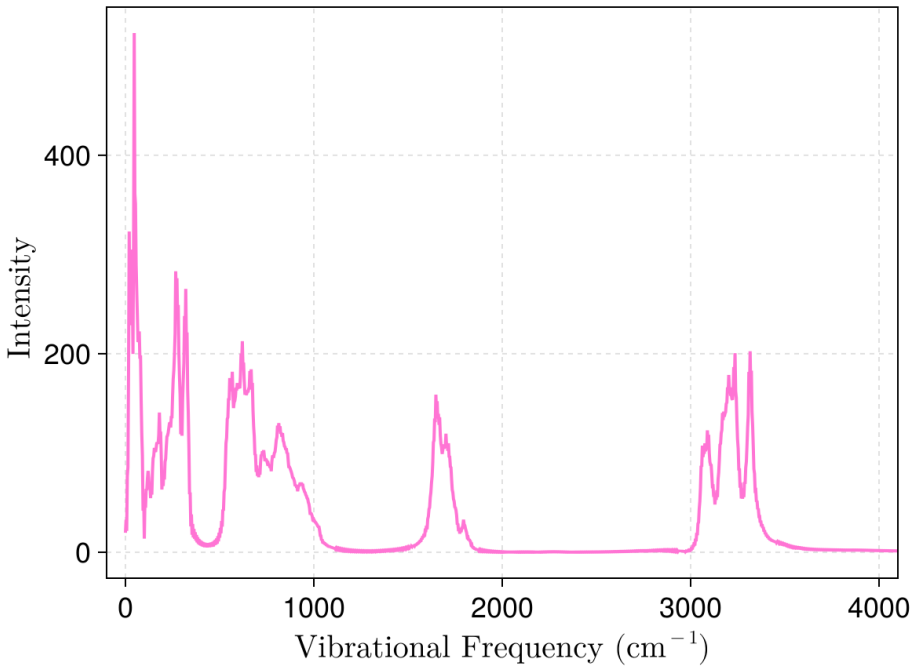

Figure 2.5: The VDOS of a bulk Ih cell with 768 molecules calculated through the VACF method using TIP4P/2005f. A 15 ps long MD simulation at 120 K (*vide infra*) was sliced into 3 5 ps long trajectories, where the VACF was calculated for each slice. A Welch window and mirroring was used on the autocorrelation signal, however, no zero-padding was used. The VDOS from the slices were averaged to produce this plot.

the VACF method. All of the same modes seen in Figure 2.4 can be seen here as well, however, here anharmonicity and temperature effects are included.

2.4 Molecular Dynamics

Molecular dynamics (MD) is a computational method for simulating the dynamics of atomic (or molecular) systems in a particular statistical ensemble. The statistical ensemble of the system, Figure 2.6 illustrates the possible ensembles, represents which thermodynamic properties of the system are held constant and which are free to vary with time. Within this thesis only the microcanonical and canonical ensembles are used, as such, only those two will be explained here. The dynamical motion of the system is calculated by solving Newton's equations of motion for a given system,

where the forces on the atoms are used to calculate the accelerations of the atoms. MD simulations provide insights into a wide range of dynamic properties, and are used across various fields of research such as physics, chemistry, materials science and biology.

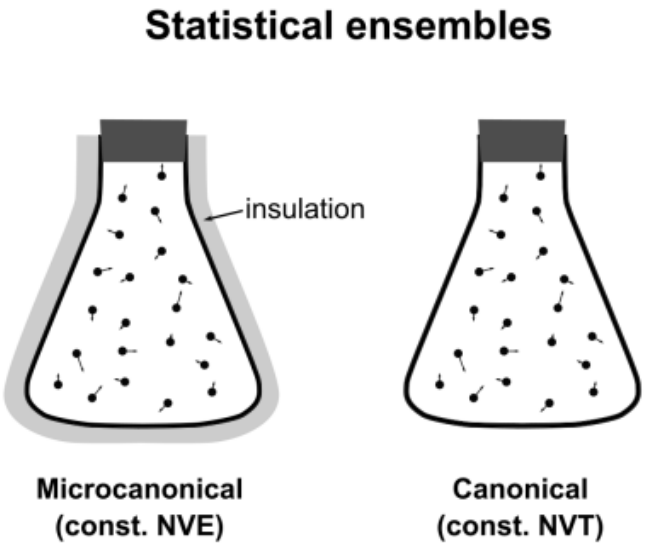

Figure 2.6: Illustration of the different statistical ensembles for thermodynamic systems. This illustration was taken from Wikimedia Commons and is free to use and share under the Creative Commons license.

A simple MD simulation only requires an integrator and a force field. As the name implies, the former integrates the equations of motion to update positions and velocities for the particles in the system. The most commonly employed MD integrator is the Velocity-Verlet integration scheme, which calculates both the velocity and position of a particle at a moment in time.

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \tag{2.53}$$

$$v(t + \Delta t) = v(t) + \frac{a(t) + a(t + \Delta t)}{2}\Delta t \tag{2.54}$$

Naturally, the accuracy of this numerical integration method depends on the choice of timestep (Δt), where smaller timesteps improve accuracy. The basic sequence of steps an MD simulation are shown in Figure 2.7, where the loop exits at the desired final simulation time.

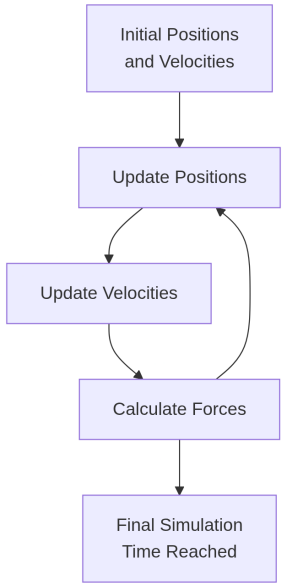

Figure 2.7: Typical sequence for MD simulations.

Prior to the extraction of observables from an MD trajectory, the system is typically equilibrated to make the kinetic and potential energy reach desired target values. However, certain phenomena require analysis of simulations that are not equilibrated, this is referred to as non-equilibrium molecular dynamics (NEMD). Frequently the topic of interest is how the system relaxes from the non-equilibrium state to an equilibrated state (*ie.* in VER studies), which is also done here in this thesis (see Chapter 6).

2.4.1 Microcanonical Ensemble

The microcanonical ensemble (NVE) describes a system that has constant number of particles (N), volume (V), and total energy (E). This ensemble is often used to study intrinsic properties of a system without external interactions. For instance, Figure 2.5 was produced from the analysis of an NVE simulation.

Within the NVE ensemble the total energy is conserved, the largest deviations in the total energy are correlated to the timestep of the simulation. Smaller timesteps allow for better energy conservation, because the numerical MD integrator produces more accurate trajectories. This relation is often also used to test newly developed or implemented potentials, by checking that the energy drift is reduced when the timestep is reduced. If the energy and forces in the potential are divergent, the energy drift will

not abide by this relation. In those cases the implementation of the potential needs to be debugged.

2.4.2 Canonical Ensemble

The canonical ensemble (NVT) describes a system that has constant number of particles (N), volume (V), and temperature (T). This ensemble is often used to study thermodynamic properties of a system at a specific temperature.

In order to maintain a constant temperature in an NVT simulations a "thermostat" is required. Thermostats are algorithms that regulate the temperature of the system, typically through scaling the velocities of the particles in the system. Careful consideration needs to be taken with the thermostat selection to avoid the so-called "flying ice cube" problem.[31,32] This issue is an atrifact of the velocity rescaling where high-frequency motions are drained into low-frequency, typically zero-frequency, motions like the translations and rotations of the whole system.

The Berendsen thermostat is one of the simplest thermostats, however, it is plagued by the flying ice cube problem. Regardless, the simplicity of the thermostat makes it perfect for explaining how thermostat algorithms work. The thermostat attempts to couple the temperature of the system to a heat bath,

$$\frac{dT}{dt} = \frac{1}{\tau}(T_0 - T(t)) \tag{2.55}$$

where T_0 is the bath temperature (the desired simulation temperature), $T(t)$ is the simulation temperature at time t, $\frac{dT}{dt}$ is the rate of change of the simulation temperature, and τ is a coupling parameter that describes the equilibration time according to which the Berendsen thermostat modifies the velocities. The temperature of the system is calculated by setting the total instantaneous kinetic energy of the system equal to the thermodynamic energy of the system and solving for temperature. This gives,

$$T = \frac{\sum_i^N m_i v_i^2}{3Nk_B} \tag{2.56}$$

where m_i is the mass of the ith particle, v_i the velocity of the ith particle, N the number of particles, and k_B the Boltzmann constant. The thermostat then rescales the velocities by the factor,

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}(\frac{T_0}{T(t)} - 1)} \tag{2.57}$$

where Δt is the timestep. Although the Berendsen thermostat should not be used for studies, the concepts carry over to other more exact thermostats. Within the works in this thesis, the canonical velocity rescaling thermostat by Bussi *et al.*[33] is used instead.

2.A Bibliography

[1] Herman JC Berendsen, James PM Postma, Wilfred F van Gunsteren, and Jan Hermans. Interaction models for water in relation to protein hydration. In *Intermolecular forces: proceedings of the fourteenth Jerusalem symposium on quantum chemistry and biochemistry held in jerusalem, Israel, april 13–16, 1981*, pages 331–342. Springer, 1981.

[2] Herman JC Berendsen, J-Raúl Grigera, and Tjerk P Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[3] Kahled Toukan and Aneesur Rahman. Molecular-dynamics study of atomic motions in water. *Physical Review B*, 31(5):2643, 1985.

[4] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[5] Christian J Burnham and Sotiris S Xantheas. Development of transferable interaction models for water. III. reparametrization of an all-atom polarizable rigid model (TTM2–R) from first principles. *The Journal of Chemical Physics*, 116(4): 1500–1510, 2002.

[6] Sotiris S Xantheas, Christian J Burnham, and Robert J Harrison. Development of transferable interaction models for water. II. accurate energetics of the first few water clusters from first principles. *The Journal of Chemical Physics*, 116(4): 1493–1499, 2002.

[7] Christian J Burnham and Sotiris S Xantheas. Development of transferable interaction models for water. IV. a flexible, all-atom polarizable potential (TTM2-F) based on geometry dependent charges derived from an ab initio monomer dipole moment surface. *The Journal of Chemical Physics*, 116(12):5115–5124, 2002.

[8] George S Fanourgakis and Sotiris S Xantheas. The flexible, polarizable, Thole-type interaction potential for water (TTM2-F) revisited. *The Journal of Physical Chemistry A*, 110(11):4100–4106, 2006.

[9] George S Fanourgakis and Sotiris S Xantheas. Development of transferable interaction potentials for water. V. extension of the flexible, polarizable, Thole-type model potential (TTM3-F, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *The Journal of Chemical Physics*, 128(7), 2008.

[10] CJ Burnham, DJ Anick, PK Mankoo, and GF Reiter. The vibrational proton potential in bulk liquid water and ice. *The Journal of Chemical Physics*, 128(15), 2008.

[11] Taisuke Hasegawa and Yoshitaka Tanimura. A polarizable water model for intramolecular and intermolecular vibrational spectroscopies. *The Journal of Physical Chemistry B*, 115(18):5545–5553, 2011.

[12] Pengyu Ren and Jay W Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B*, 107(24): 5933–5947, 2003.

[13] Pengyu Ren and Jay W Ponder. Temperature and pressure dependence of the AMOEBA water model. *The Journal of Physical Chemistry B*, 108(35):13427–13437, 2004.

[14] Jay W Ponder, Chuanjie Wu, Pengyu Ren, Vijay S Pande, John D Chodera, Michael J Schnieders, Imran Haque, David L Mobley, Daniel S Lambrecht, Robert A DiStasio Jr, *et al.* Current status of the AMOEBA polarizable force field. *The Journal of Physical Chemistry B*, 114(8):2549–2564, 2010.

[15] Lee-Ping Wang, Teresa Head-Gordon, Jay W Ponder, Pengyu Ren, John D Chodera, Peter K Eastman, Todd J Martinez, and Vijay S Pande. Systematic improvement of a classical molecular model of water. *The Journal of Physical Chemistry B*, 117(34):9956–9972, 2013.

[16] Chengwen Liu, Jean-Philip Piquemal, and Pengyu Ren. AMOEBA+ classical potential for modeling molecular interactions. *Journal of Chemical Theory and Computation*, 15(7):4122–4139, 2019.

[17] Chengwen Liu, Jean-Philip Piquemal, and Pengyu Ren. Implementation of geometry-dependent charge flux into the polarizable AMOEBA+ potential. *The Journal of Physical Chemistry Letters*, 11(2):419–426, 2019.

[18] Enrique Ricardo Batista. *Development of a new water-water interaction potential and application to molecular processes in ice*. University of Washington, 1999.

[19] K. T. Wikfeldt, E. R. Batista, F. D. Vila, and H. Jónsson. A transferable H_2O interaction potential based on a single center multipole expansion: SCME. *Phys. Chem. Chem. Phys.*, 15:16542–16556, 2013.

[20] Elvar Orn Jónsson, Asmus Ougaard Dohn, and Hannes Jonsson. Polarizable embedding with a transferable H_2O potential function I: formulation and tests on dimer. *Journal of Chemical Theory and Computation*, 15(12):6562–6577, 2019.

[21] Elvar Orn Jónsson, Soroush Rasti, Marta Galynska, Jorg Meyer, and Hannes Jónsson. Transferable potential function for flexible H_2O molecules based on the single-center multipole expansion. *Journal of Chemical Theory and Computation*, 18(12):7528–7543, 2022.

[22] Volodymyr Babin, Claude Leforestier, and Francesco Paesani. Development of a "first principles" water potential with flexible monomers: dimer potential energy surface, VRT spectrum, and second virial coefficient. *Journal of Chemical Theory and Computation*, 9(12):5395–5403, 2013.

[23] Volodymyr Babin, Gregory R Medders, and Francesco Paesani. Development of a "first principles" water potential with flexible monomers. II: trimer potential energy surface, third virial coefficient, and small clusters. *Journal of Chemical Theory and Computation*, 10(4):1599–1607, 2014.

[24] Gregory R Medders, Volodymyr Babin, and Francesco Paesani. Development of a "first-principles" water potential with flexible monomers. III. liquid phase properties. *Journal of Chemical Theory and Computation*, 10(8):2906–2910, 2014.

[25] Francesco Paesani. Getting the right answers for the right reasons: toward predictive molecular simulations of water with many-body potential energy functions. *Accounts of Chemical Research*, 49(9):1844–1851, 2016.

[26] Harry Partridge and David W Schwenke. The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio cal-

culations and experimental data. *The Journal of Chemical Physics*, 106(11): 4618–4639, 1997.

[27] KT Tang and J Peter Toennies. An improved simple model for the van der waals potential based on universal damping functions for the dispersion coefficients. *The Journal of Chemical Physics*, 80(8):3726–3741, 1984.

[28] Jie Yang, Yi Ren, An-min Tian, and Huai Sun. COMPASS force field for 14 inorganic molecules, He, Ne, Ar, Kr, Xe, H_2, O_2, N_2, NO, CO, CO_2, NO_2, CS_2, and SO_2, in liquid phases. *The Journal of Physical Chemistry B*, 104(20):4951–4957, 2000.

[29] Marc C van Hemert, Junko Takahashi, and Ewine F van Dishoeck. Molecular dynamics study of the photodesorption of CO ice. *The Journal of Physical Chemistry A*, 119(24):6354–6369, 2015.

[30] Jun Chen, Jun Li, Joel M Bowman, and Hua Guo. Energy transfer between vibrationally excited carbon monoxide based on a highly accurate six-dimensional potential energy surface. *The Journal of Chemical Physics*, 153(5):054310, 2020.

[31] Stephen C Harvey, Robert K-Z Tan, and Thomas E Cheatham III. The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition. *Journal of computational chemistry*, 19(7):726–740, 1998.

[32] Efrem Braun, Seyed Mohamad Moosavi, and Berend Smit. Anomalous effects of velocity rescaling algorithms: the flying ice cube effect revisited. *Journal of chemical theory and computation*, 14(10):5262–5272, 2018.

[33] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.