Nuclear quantum effects in solid water: new insights from computational modeling

Rasti, S.

CHAPTER 5

Transferable Potential Function for Flexible H_2O

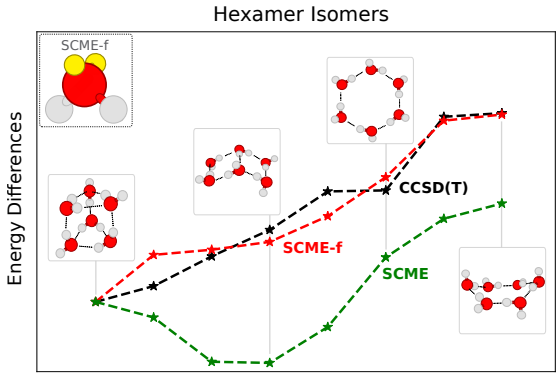
This chapter is based on:

E. Ö. Jónsson, S. Rasti, M. Galynska, J. Meyer and H. Jónsson, Transferable Potential Function for Flexible H_2O Molecules Based on the Single Center Multipole Expansion, *submitted to the Journal of Chemical Theory and Computation (JCTC)*

The submitted version of is available at arXiv:2007.06090v2 [physics.comp-ph].

Abstract

A potential function is presented for describing a system of flexible H_2O molecules based on the single center multipole expansion (SCME) of the electrostatic interaction. The model, referred to as SCME/f, includes the variation of the molecular quadrupole moment as well as the dipole moment with changes in bond length and angle so as to reproduce results of high level electronic structure calculations. The multipole expansion also includes fixed octupole and hexadecapole moments, as well as anisotropic dipole-dipole, dipole-quadrupole and quadrupole-quadrupole polarizability tensors. The model contains five adjustable parameters related to the repulsive interaction and damping functions in the electrostatic and dispersion interactions. Their values are adjusted to reproduce the lowest energy isomers of small clusters, $(H_2O)_n$ with $n = 2 - 6$, as well as measured properties of the ice Ih crystal. Subsequent calculations of the energy difference between the various isomer configurations of the clusters show that SCME/f gives good agreement with results of electronic structure calculations and represents a significant improvement over the previously presented rigid SCME potential function. Analysis of the vibrational frequencies of the clusters and structural properties of ice Ih crystal show the importance of accurately describing the variation of the quadrupole moment with molecular structure.



5.1 Introduction

The most commonly used potential energy functions for describing water molecules and their interaction are based on simple pairwise additive functions with fixed point charges [1–4], such as the well known TIPnP and SPC force fields. Extensions of these potential functions to describe flexible molecules have been developed, such as aSPC/Fw [5] and q-TIP4P/F [6], and they offer, for example, the possibility to include the effect of zero point energy. The point charge potential functions are typically parameterized in such a way as to reproduce a few thermally averaged properties of liquid water. The properties of water molecules are, however, strongly environment dependent as illustrated by the molecular dipole moment, which is 1.8 D in the gas phase and 3.1 D in ice Ih [7]. This large environment dependence needs to be modeled accurately in order to develop a transferable potential function applicable, for example, to small clusters and crystal structures as well as liquid water.

Such environment dependence is best described using well established physical laws, since empirical fitting to some limited set of data is likely not going to work well when the potential function is applied to configurations that are significantly different from the ones used in the fitting process. A systematic multipole expansion up to and including the hexadecapole, with dipole and quadrupole polarizability, has been shown to reproduce well the electrostatics in water clusters and ice [8]. A potential function based on this approach has been presented for rigid molecules and is referred to as the single center multipole expansion (SCME) potential function [9, 10]. In the present work, this approach is extended to flexible molecules.

By expanding the electrostatics around a single center on each molecule, the introduction of point charges is avoided and the correct long range distance dependence of the Coulomb potential built in naturally. The leading term, the dipole potential, decays as $1/R^3$, and combined with the polarization response of the molecules this makes it possible to use a long range cut-off for the electrostatic interaction between molecules in typical condensed matter simulations. [7]

Hybrid simulation schemes, where part of the system is simulated using a potential function while another part is described using electronic structure calculations, the so-called quantum mechanics / molecular mechanics (QM/MM) simulations, have been used in important simulation studies in various fields such has biochemistry [11–15], medicine [16], photochemistry [17] and solvation dynamics [18–21], nanostructures [22], and materials science [23]. In most cases, such simulations make use of fixed point charge models [24–27], thereby neglecting the mutual polarization of the charges in the MM subsystem by the QM subsystem – an effect that was, however, included in the inceptive work initiating the QM/MM approach [28]. The use of fixed point charge models to represent water molecules in the MM region results in errors that limit the applicability of the QM/MM method.

Several H_2O potential functions that include some level of polarizability exist [29–31]. These include the Thole-type multipole models such as the TTMn series [32–35], and HBB2-pol [36, 37]. The MB-pol [38–40] potential function has arguably reached the highest precision as it includes an explicit treatment of two-body and three-body interactions through an intricate permutationally invariant polynomial fit to data bases constructed with high level quantum chemistry calculations. However, inclusion of such explicit many body terms makes the interfacing with a QM region more challenging. Instead, simpler polarizable MM potential functions based on pair-wise potentials to describe the short-range interactions are used in so-called polarizable embedding QM/MM (PE-QM/MM) approach [26, 41–71]. The PE-QM/MM approach can be used to study the effects of solvation and solvent response to excitations and charge transfer in solvated species. However, such simulations have typically included only the molecular dipole-dipole response and make use of atomic point charges.

Here, an extension of the single-center multipole expansion [9, 10] (SCME) potential function is described, which has recently been integrated in a PE-QM/MM scheme [72, 73]. The extended potential function, SCME/f, includes flexibility of the internal geometry of the water molecules while still maintaining the single center description of the electrostatic interaction in terms of molecular moment tensors. The SCME/f model in-

cludes variable dipole and quadrupole moment tensors that depend on the geometry of the H_2O molecule. The dipole is described by the well established Partridge-Schwenke model, [74] but a new, geometrical model based on four sites is presented here for the quadrupole moment. It reproduces results of high-level multireference electronic structure calculations of the quadrupole moment to within 1.6% RMS over a broad range in its magnitude. This model for the quadrupole moment is found to provide better description than the so-called M-site models that have been used previously. [2, 32–40, 75–78]

There are five adjustable parameters in the description of the intermolecular interaction. They include parameters relating to the pair-wise repulsive interaction as well as damping parameter in the dispersion interaction and a screening parameter for the electrostatic interaction tensors. These parameters are optimized in such a way that the SCME/f reproduces the binding energy and intermolecular distance of the dimer, the interaction energy of the lowest energy conformation of water clusters $(H_2O)_n$ with n ranging from 3 to 6, calculated at the level of RI-MP2 with CCSD(T) corrections [79] and full CCSD(T) at the complete basis set limit [80] as well as measured properties of crystalline ice Ih taking into account the zero-point energy. The resulting parametrization of the model reproduces nicely trends in the relative energy of the conformers of the hexamer obtained from high level quantum chemistry calculations. Some discrepancies, however, exist in the series of pentamer isomers. An analysis of the frequency of vibrational modes of the various clusters and the structure of ice Ih crystal highlights the importance of an accurate model for the molecular quadrupole moment.

The article is organized as follows: The SCME/f potential function is described in Section 5.2. The dipole and quadrupole surfaces are presented in Section 5.3 and the calculation of atomic forces is described in Section 5.4. The fitting of the five adjustable parameters is described in Section 5.5 and comparison with *ab initio* data on the cluster conformer energy and vibrational frequencies of small clusters is described in Section 5.6. Discussion and conclusions are in Section 5.7.

5.2 Flexible SCME Model

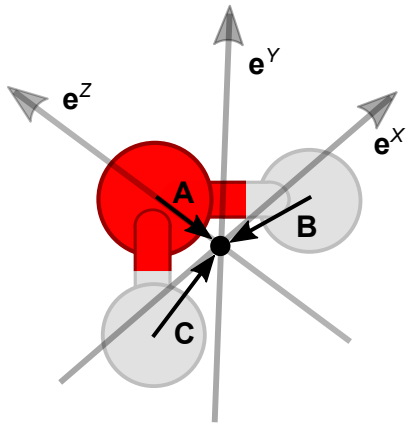


Figure 5.1: The definition of the principal vectors and local reference frame for the water molecule used in the SCME/f model. The black circle denotes the expansion center, chosen here to be at the center of mass. Black arrows show the three principal vectors **A**, **B** and **C** pointing *from* the oxygen and the hydrogen atoms *to* the expansion center. The gray opaque arrows show the local reference frame basis vectors $\{\mathbf{e}^X, \mathbf{e}^Y, \mathbf{e}^Z\}$. The principal vectors **B** and **C** define a local-to-global reference frame rotation matrix. Due to symmetry specific indexing of the atoms is omitted, and positions and scales are exaggerated for clarity.

Fig. 5.1 shows the principal vectors which define both the position of the expansion center and the local-to-global reference frame rotation matrix for the flexible water molecule. The local frame origin is placed at the center of mass (COM). In SCME/f each water molecule is ascribed a molecular dipole and quadrupole moments in terms of variable partial charges based on the internal geometry, $\mu^i_\alpha(\{\mathbf{r}^{ia}\})$ and $\theta^i_{\alpha\beta}(\{\mathbf{r}^{ia}\})$, respectively, where $\{\mathbf{r}^{ia}\} = \{\mathbf{r}^{i\mathrm{O}}, \mathbf{r}^{i\mathrm{H}_1}, \mathbf{r}^{i\mathrm{H}_2}\}$, and is the set of position vectors for atoms a in molecule i in the global reference frame. The details of the dipole moment and quadrupole moment surfaces are described in Section 5.3. The index i is used to denote both the specific water molecule, as well as the corresponding COM site. Furthermore, each water molecule is ascribed, in the local reference frame, a fixed octupole, $\Omega^{i'}_{\alpha\beta\gamma}$, and

hexadecapole, $\Phi^{i'}_{\alpha\beta\gamma\delta}$, moment tensors, as well as polarizability tensors including dipole-dipole, $\alpha^{i'}_{\alpha\beta}$, dipole-quadrupole, $A^{i'}_{\alpha\beta\gamma}$, and quadrupole-quadrupole, $C^{i'}_{\alpha\beta\gamma\delta}$, induction terms.

Lipparini *et al.* [81] describe commonly used local reference frames and associated rotation matrices. The derivation here follows closely their work, with some obvious sign changes. The expansion center is placed at the COM

$$\mathbf{r}^i = \sum_a^{n_i} \mathbf{r}^{ia} \frac{M^a}{M^i} \tag{5.1}$$

where n_i denotes the atomic sites $\{\mathrm{O},\mathrm{H}_1,\mathrm{H}_2\}$ of molecule i, and M^a and M^i is the mass of the atom and molecule, respectively. The principal vectors used to define the rotation are

$$\mathbf{B}^i = \mathbf{r}^i - \mathbf{r}^{i\mathrm{H}_1}, \quad \mathbf{C}^i = \mathbf{r}^i - \mathbf{r}^{i\mathrm{H}_2} \quad , \tag{5.2}$$

where in general, i.e. for a flexible H_2O molecule, $B^i \neq C^i$. Unit basis vectors are in terms of the principal vectors given by

$$\begin{aligned}
\mathbf{e}^{iZ} =& \frac{B^i\mathbf{C}^i + C^i\mathbf{B}^i}{|B^i\mathbf{C}^i + C^i\mathbf{B}^i|} \\
\mathbf{e}^{iX} =& \frac{\mathbf{B}^i - (\mathbf{B}^i \cdot \mathbf{e}^{iZ})\mathbf{e}^{iZ}}{|\mathbf{B}^i - (\mathbf{B}^i \cdot \mathbf{e}^{iZ})\mathbf{e}^{iZ}|} \\
\mathbf{e}^{iY} =& \mathbf{e}^{iZ} \times \mathbf{e}^{iX}
\end{aligned} \tag{5.3}$$

where $\mathbf{e}^{iZ}$ is, as defined above, the bisector between the two oxygen-hydrogen bonds. In terms of the unit basis vectors a unitary local-to-global reference frame rotation matrix is

$$\mathbf{R}^i = \begin{bmatrix} e^{iX}_x & e^{iX}_y & e^{iX}_z \\ e^{iY}_x & e^{iY}_y & e^{iY}_z \\ e^{iZ}_x & e^{iZ}_y & e^{iZ}_z \end{bmatrix} \tag{5.4}$$

Given the rotation matrix for each molecule the fixed moment and polarizability matrices are rotated into the global reference frame for each COM site i[1]

[1]Throughout this work Einstein notation was used, i.e. Cartesian vector spaces are indexed with Greek letters, $\alpha = \beta = \cdots = \nu \in \{x, y, z\}$, and repeated Greek indices are to be summed over.

$$M^i_{\alpha\ldots\delta} = R^i_{\eta\alpha}\ldots R^i_{\sigma\delta}M^{i'}_{\eta\ldots\sigma} \tag{5.5}$$

where $M^i_{\alpha\ldots\delta}$ is a generalized tensor of order t, requiring t rotation operations (e.g. $\alpha^i_{\alpha\beta} = R^i_{\eta\alpha}R^i_{\tau\beta}\alpha^{i'}_{\eta\tau}$). With the definitions above atomic forces are derived (see the Supplementary Information) from the contribution of the fixed moments and polarizabilities to the electrostatic interactions involving the single expansion center on each molecule.

General formulation, and notation, of the perturbative expansion of the electrostatic intermolecular interaction – resulting in the multipole moment model – can be found elsewhere [82]. Here only the main expressions are presented which are used to arrive at a self-consistent solution to polarized molecular moments at sites i in response to the external field due to all other neighboring molecules $j(\neq i)$.

Given the external field, V^i_α (negative of the electric field), and the field gradient, $V^i_{\alpha\beta}$, at the COM of i, the molecules are polarized resulting in induced dipole and quadrupole moments

$$\Delta\mu^i_\alpha = -\alpha^i_{\alpha\beta}V^i_\beta - \frac{1}{3}A^i_{\alpha\beta\gamma}V^i_{\beta\gamma} \tag{5.6}$$

$$\Delta\theta^i_{\alpha\beta} = -A^i_{\gamma\alpha\beta}V^i_\gamma - C^i_{\gamma\delta\alpha\beta}V^i_{\gamma\delta} \tag{5.7}$$

where the external field is given by

$$V^i_\alpha = \sum_{j\neq i}^{n} V^{ij}_\alpha \tag{5.8}$$

and the contribution to the external field at site i due to site j is given by

$$\begin{aligned} V^{ij}_\alpha =& -T^{ij}_{\alpha\beta}(\mu^j_\beta(\{\mathbf{r}^{jb}\}) + \Delta\mu^j_\beta) + \frac{1}{3}T^{ij}_{\alpha\beta\gamma}(\theta^j_{\beta\gamma}(\{\mathbf{r}^{jb}\}) + \Delta\theta^j_{\beta\gamma}) \\ &- \frac{1}{15}T^{ij}_{\alpha\beta\gamma\delta}\Omega^j_{\beta\gamma\delta} + \frac{1}{105}T^{ij}_{\alpha\beta\gamma\delta\epsilon}\Phi^j_{\beta\gamma\delta\epsilon} \end{aligned} \tag{5.9}$$

The field gradient – and higher order gradients – are given by the subsequent use of the gradient operator, $\nabla_\beta V^i_\alpha = V^i_{\alpha\beta}$, $\nabla_\gamma V^i_{\alpha\beta} = V^i_{\alpha\beta\gamma}$.

At the start the external field and field gradient due to the intrinsic moments is evaluated at each site. This results in an induced dipole and quadrupole moment, which

in turn results in a change in the external field and field gradient. A self-consistent solution to the non-linear relation between Eq. (5.6)–Eq. (5.9) is achieved with an iterative procedure and a suitable convergence threshold of the induced moments to achieve energy-force consistency (see the Supplementary Information).

As the point moments come close the multipole moment expansion breaks down – resulting in the so-called polarization catastrophe. [83] In order to avoid this screened interaction tensors are introduced [83–88] which effectively smear out the point moments. To zeroth order the Coulomb interaction tensors in Eq. (5.9) are defined as

$$T^{ij} = \frac{1}{|\mathbf{r}^j - \mathbf{r}^i|}\lambda_0(r) = \frac{1}{r}\lambda_0(r) \tag{5.10}$$

where $\lambda_0(r)$ is a short-range electrostatic interaction screening function. The gradient operators act to increase the order of the screened interaction tensors, for example

$$\nabla_\alpha T^{ij} = T^{ij}_\alpha \equiv -\frac{r_\alpha}{r^3}\lambda_1(r) \tag{5.11}$$

$$\nabla_\beta T^{ij}_\alpha = T^{ij}_{\alpha\beta} \equiv 3\frac{r_\alpha r_\beta}{r^5}\lambda_2(r) - \frac{\delta_{\alpha\beta}}{r^3}\lambda_1(r) \tag{5.12}$$

where $r_\alpha = (\mathbf{r}^j - \mathbf{r}^i)_\alpha$.

Most commonly used interaction tensor screening functions in the context of polarizable force fields are based on exponential decay of the point charges resulting in the Thole-type damped tensors. [83] Here we make use of screening functions derived from considering the overlap and resulting Coulomb electrostatic screening of Gaussian charge densities and multipoles. [87] In the equations above they are

$$\lambda_1(r) = \mathrm{erf}(S) - \frac{2}{\sqrt{\pi}} S e^{-S^2} \tag{5.13}$$

$$\lambda_2(r) = \mathrm{erf}(S) - \frac{2}{\sqrt{\pi}}\left(S + \frac{2}{3}S^3\right) e^{-S^2} \tag{5.14}$$

where S is the screened distance, $S = r/g$, and g is the screening length – describing the spatial extent of the Gaussian functions.

In the SCME/f model the total energy is a functional of the external field, V^i_α, at each

molecular COM site i and is given by

$$E_{\rm tot}[\{V_\alpha^i\}] = E_{\rm elst}[\{V_\alpha^i\}] + E_{\rm non-elst} + E_{\rm mon} \tag{5.15}$$

where the terms on the right hand side are, $E_{\rm elst}[\{V_\alpha^i\}]$, the total electrostatic energy functional, the non-electrostatic terms, $E_{\rm non-elst}$, which includes a pair-wise repulsive and a dispersion potential, and $E_{\rm mon}$, which is a sum of the internal energies described by the Partridge-Schwenke potential energy surface (PS–PES) of the water monomer. [74]

More explicitly the first term on the right hand side of Eq. (5.15) can be further separated into three terms describing the inter- and intramolecular contributions to the total electrostatic energy of the system, namely

$$E_{\rm elst}[\{V_\alpha^i\}] = E_{\rm in}[\{V_\alpha^i\}] + E_{\rm pol}[\{V_\alpha^i\}] + E_{\rm self}[\{V_\alpha^i\}] \tag{5.16}$$

where $E_{\rm in}[\{V_\alpha^i\}]$ is the electrostatic interaction between the intrinsic molecular moments and $E_{\rm pol}[\{V_\alpha^i\}]$ is the field-induced polarization energy. At self-consistency these terms combine to give

$$\begin{aligned} E_{\rm in+pol}[\{V_\alpha^i\}] =& \frac{1}{2}\sum_i^n \Big((\mu_\alpha^i(\{\mathbf{r}^{ia}\}) + \Delta\mu_\alpha^i)V_\alpha^i + \frac{1}{3}(\theta_{\alpha\beta}^i(\{\mathbf{r}^{ia}\}) + \Delta\theta_{\alpha\beta}^i)V_{\alpha\beta}^i \\ &+ \frac{1}{15}\Omega_{\alpha\beta\gamma}^i V_{\alpha\beta\gamma}^i + \frac{1}{105}\Phi_{\alpha\beta\gamma\delta}^i V_{\alpha\beta\gamma\delta}^i \Big) \end{aligned} \tag{5.17}$$

$E_{\rm self}$ is the on-site self-energy, given by

$$E_{\rm self}[\{V_\alpha^i\}] = -\frac{1}{2}\sum_i^n \left(\Delta\mu_\alpha^i V_\alpha^i + \frac{1}{3}\Delta\theta_{\alpha\beta}^i V_{\alpha\beta}^i \right) \tag{5.18}$$

and accounts for the change in internal energy required to polarize the molecules.

The non-electrostatic term is composed of two intermolecular pair-wise potentials centered on the oxygen atom

$$E_{\rm non-elst} = E_{\rm rep} + E_{\rm disp} \tag{5.19}$$

describing repulsion, $E_{\rm rep}$, and dispersion $E_{\rm disp}$. In the following expressions for the potentials the distance r refers to the oxygen-oxygen distance between pair i and j, or $r = |\mathbf{r}^{j\rm O} - \mathbf{r}^{i\rm O}|$.

Making use of the same dispersion coefficients as in the original SCME model [89]. The dispersion energy is

$$E_{\text{disp}} = -\sum_{i}^{n}\sum_{j<i}^{n}\left(\frac{C_6}{r^6}t_6(r) + \frac{C_8}{r^8}t_8(r) + \frac{C_{10}}{r^{10}}t_{10}(r)\right) \tag{5.20}$$

with isotropic coefficients up to tenth order from Wormer and Hettema [90]. At short range the interaction is smoothly switched off with a Tang-Toennies damping function [91]

$$t_m(r) = 1 - e^{-\tau_d r}\sum_{k=0}^{m}\frac{(\tau_d r)^k}{k!} \tag{5.21}$$

where the parameter τ_d represents the inverse decay length of the charge density.

In the rigid SCME [89] model a modified Born-Mayer potential is used, which includes a term which scales the magnitude of the repulsion depending on the local environment around the repulsion center – a molecular density dependent term. With the introduction of the Gaussian type interaction tensor screening function we find the molecular density dependence unnecessary and revert back to the basic Born-Mayer type potential. The pair-wise repulsion is

$$E_{\text{rep}} = \sum_{i}^{n}\sum_{j<i}^{n} A_{\text{rep}} r^{-k} e^{-hr} \tag{5.22}$$

The parameters of the non-electrostatic terms, τ_d, A_{rep}, k and h, are optimized to work with the new SCME/f model. The optimization also includes the screening length parameter g of Eq. (5.14). The fitting is described in Section 5.5.

5.3 The Dipole and Quadrupole Moment Surfaces

The internal energy as described by the PS-PES includes analytical atomic force components, [74] as well as an accurate mapping of the dipole moment surface (DMS) for an isolated water molecular as a function of the internal geometry. The DMS is given by

$$\mu_\alpha^i(\mathbf{r}^{i\text{O}}, \mathbf{r}^{i\text{H}_1}, \mathbf{r}^{i\text{H}_2}) = q^{i\text{H}_1} r_\alpha^{i\text{H}_1} + q^{i\text{H}_2} r_\alpha^{i\text{H}_2} + q^{i\text{O}} r_\alpha^{i\text{O}} \tag{5.23}$$

where $q^{i\mathrm{O}} = -(q^{i\mathrm{H}_1} + q^{i\mathrm{H}_2})$ and the partial charges of the two hydrogens are in turn a function of the internal geometry, fitted to recreate the calculated DMS. For example $q^{i\mathrm{H}_1} = q^{i\mathrm{H}_1}(r^{\mathrm{OH}_1}, r^{\mathrm{OH}_2}, \cos(\theta_{\mathrm{HOH}}))$, where r^{OH_1} and r^{OH_2} are the internal bond lengths between the oxygen and the two hydrogens, and θ_{HOH} the HOH angle. This mapping is used, and it was left unchanged.

The DMS partial charges are not suitable to describe a quadrupole moment surface (QMS) without modification. Instead the charge site associated with the oxygen is split up into two components and placed within a plane perpendicular to the symmetry plane of the hydrogens and oxygens. The sites are denoted L_1 and L_2, where the site positions are directly related to the length of the hydrogen bond lengths indexed H_1 and H_2, and the HOH angle. See Fig. 5.2. The QMS is written as

$$\theta^i_{\alpha\beta}(\mathbf{r}^{i\mathrm{O}}, \mathbf{r}^{i\mathrm{H}_1}, \mathbf{r}^{i\mathrm{H}_2}) = \sum_a^{n_i'} \frac{3}{2}\left\{ q^{ia}\left((\mathbf{r}^{ia} - \mathbf{r}^i)_\alpha (\mathbf{r}^{ia} - \mathbf{r}^i)_\beta - \frac{\delta_{\alpha\beta}}{3}||\mathbf{r}^{ia} - \mathbf{r}^i|| \right) \right\} \quad (5.24)$$

where n_i' denotes the sites $\{\mathrm{H}_1', \mathrm{H}_2', \mathrm{L}_1, \mathrm{L}_2\}$ associated with molecule i. The apostrophe on the hydrogen is used to discern their role in the QMS from their role in the DMS since the charges $q^{i\mathrm{H}_l{}'}$ are different from the DMS charges, and are

$$q^{i\mathrm{H}_l{}'} = \mathrm{A} q^{i\mathrm{H}_l} + \mathrm{B} q^{\mathrm{H}}_{\mathrm{eq}} \quad (5.25)$$

and for the L-sites they are

$$q^{i\mathrm{L}_l} = \mathrm{C} q^{i\mathrm{H}_l} + \mathrm{D} q^{\mathrm{H}}_{\mathrm{eq}} \quad (5.26)$$

where $q^{\mathrm{H}}_{\mathrm{eq}}$ is the DMS charge of the hydrogen in the equilibrium monomer configuration.

The position of the L_1 and L_2 charge sites is related to the atomic positions of each water molecule through a rotation operator times a scaling factor which controls the length of the rotated vector. A translation operator translates the vector to the COM position of molecular site i for completeness. Explicitly this operation is

$$r^{i\mathrm{L}_l}_\alpha = R^{i\mathrm{L}_l}_{\eta\alpha} e^{iZ}_\eta f(\mathbf{r}^{\mathrm{H}_l}) + r^i_\alpha \quad (5.27)$$

We make use of the unit basis vectors previously used to define the local-to-global rotation matrices in Eq. (5.1)–Eq. (5.3). The rotation matrices for the L_1 and L_2 sites

are

$$\mathbf{R}^{i\mathrm{L}_1} = \left(\cos(f(\theta))\mathbf{I} - \sin(f(\theta))\left[\mathbf{e}^{iX}\right]_\times\right) \tag{5.28}$$

$$\mathbf{R}^{i\mathrm{L}_2} = \left(\cos(f(\theta))\mathbf{I} + \sin(f(\theta))\left[\mathbf{e}^{iX}\right]_\times\right) \tag{5.29}$$

and is a simplification of the general Rodrigues' rotation operator [92] in terms of the local orthonormal basis vectors (shown in Fig. 5.1).

In order to allow for flexibility of the L-sites and correlate their positions to the change in the positions of the hydrogens, both the angle factor and length scale factor are defined in terms of the OH bond lengths and HOH angle through

$$f(\mathbf{r}^{\mathrm{H}_l}) = \mathrm{a} + \mathrm{b}(|\mathbf{r}^{i\mathrm{O}} - \mathbf{r}^{i\mathrm{H}_l}| - r_{\mathrm{eq}}) + \mathrm{c}(|\mathbf{r}^{i\mathrm{O}} - \mathbf{r}^{i\mathrm{H}_l}| - r_{\mathrm{eq}})^2 \tag{5.30}$$

$$f(\theta) = \mathrm{d} + \mathrm{e}(\theta - \theta_{\mathrm{eq}}) \tag{5.31}$$

where r_{eq} and θ_{eq} are the equilibrium hydrogen to oxygen bond length and HOH angle of the isolated PS–PES water molecule, respectively, see Fig. 5.2. It is found that a second order polynomial in terms of the change in bond length, and a linear term for the change in bond angles is adequate to capture the QMS with good accuracy. The charge scaling parameters A, B, C and D, and the geometric parameters a, b, c, d, e are fitted parameters, described below.

5.3.1 Ab initio QMS Calculations and Fit

The dipole and quadrupole moment is mapped using the ab initio quantum chemistry software ORCA [93, 94]. An iterative-configuration expansion configuration interaction (ICE-CI) method is used, with the aug-cc-pvqz basis set and the energy convergence threshold is set to 10^{-8} E_h. Eight correlated electrons are included and the active orbitals were chosen by including MP2 orbitals of natural orbital occupation numbers ranging between 1.99999 and 0.00001. The ICE-CI method is related to the CIPSI technique [95]. Note that this level of theory is necessary to accurately determine the dipole and quadrupole moment using their well defined charge density based operators, instead of resorting to energy based schemes to estimate these quantities. For example, it is found that coupled-cluster at the CCSD(T)/aug-cc-pvqz level of theory and

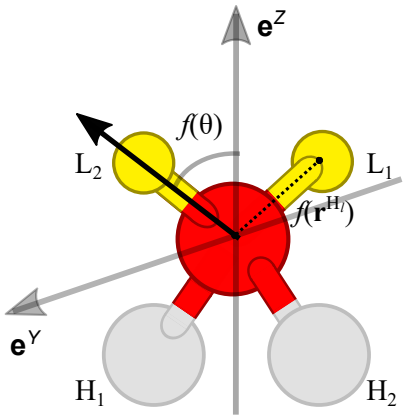


Figure 5.2: L-site placement (yellow) in the water monomer structure. The relationship of the angle to the unit basis vectors which describe the local reference frame is shown, Eq. (5.31) and eqs Eq. (5.28)–Eq. (5.29). For example, operating with the rotation vector corresponding to hydrogen indexed 1 on e_α^{iZ} results in $\left(\cos(f(\theta))e_\alpha^{iZ} - \sin(f(\theta))e_\alpha^{iY}\right)$. Due to symmetry specific indexing of the atoms is completely interchangeable, and either pair of H and L in the Figure above can serve as pair 1 or 2. The distance from the oxygen to a L-site, controlled with $f(\mathbf{r}^{\mathrm{H}_l})$ is a second order polynomial function depending on the position of one of the hydrogens (while the position of the other L-site depends on the other hydrogen), Eq. (5.30). Positions and scales are exaggerated for clarity.

orbital optimized coupled-cluster theory OOCCSD(T)/aug-cc-pvdz, did not provide a satisfactory agreement with the DMS of the PS-PES, when using the dipole moment operator $\mu_\alpha = \int \rho(\mathbf{r}) r_\alpha d\mathbf{r}$. See the Supporting Information for more details.

Starting from the ground state geometry in the local-frame as shown in Fig. 5.1 the internal bond lengths and HOH angle are systematically changed and range from 0.7 to 1.3 Å, and 60 to 175°, respectively. These intervals broadly represent the variation in the bond lengths and the angle of the water molecule in the liquid phase at ambient conditions. Fig. 5.3 shows a comparison between the internal energy change of each configuration as calculated by the ICE-CI method compared to the PS-PES. The

agreement is excellent, and justifies the use of the ab initio data to fit the QMS while retaining the original PS-PES energy mapping to describe the internal energy change and resulting atomic forces in our model. Fig. 5.4, left, presents a comparison between the ICE-CI DMS and the PS-PES DMS, again in an excellent agreement.

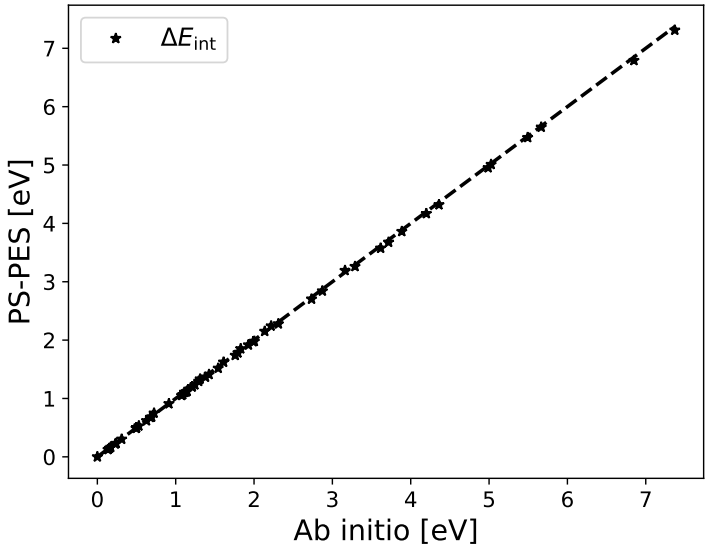


Figure 5.3: The relative internal energy difference between the different monomer configurations used in the QMS fit, compared between the ab initio results and the PS–PES. The good agreement between the two methods implies that the use of the ICE-CI data to fit the QMS justifies the use of the original PS–PES to represent internal energy changes and resulting atomic forces, as both potential energy surfaces are close with RMSD of 0.022 eV, within chemical accuracy (∼0.51 kcal/mol).

The QMS model parameters associated with the charges in eqs Eq. (5.25)–Eq. (5.26), A, B, C and D, as well as the geometric parameters of eqs Eq. (5.30)–Eq. (5.31), a, b, c, d and e, are fitted to best reproduce the principal quadrupole moment component. Considering the water molecule in the ground state configuration the symmetric

quadrupole moment tensor can be written as

$$\theta = \begin{bmatrix} \theta_T - \Delta & 0 & 0 \\ 0 & -\theta_T - \Delta & 0 \\ 0 & 0 & 2\Delta \end{bmatrix} \tag{5.32}$$

where $\theta_T = (\theta_{xx} - \theta_{yy})/2$.

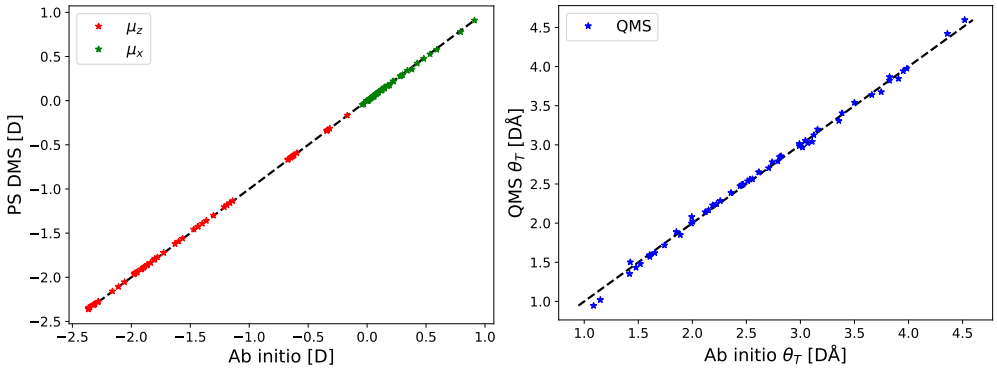


Figure 5.4: Left: comparison of the dipole z- and x-components, μ_z and μ_x respectively, as predicted by the DMS, Eq. (5.23) and compared to the ICE-CI μ_z and μ_x. Note that due to a choice of local reference frame the μ_y component is always numerically zero. The DMS of the PS-PES and ICE-CI are in an excellent agreement, with a RMSD of 0.004 D and within 0.5% on average. Right: comparison of the θ_T component mapped by the QMS, Eq. (5.24), with the ab initio ICE-CI data. The geometric QMS model of this work, which is fitted to best reproduce the ab initio results, captures the results to a good degree with low scatter, a mean absolute error of 0.04 DÅ, and an average RMS difference of around 1.6% (see Supplementary Information for the RMSD analysis).

The values of the QMS parameters are determined by carrying out a least-squares optimization, using a module freely available in the scientific computing package SciPy. [96] Table 5.1 presents the numerical values and units of the resulting best fit parameters, and Fig. 5.4, right, shows the resulting fit of the θ_T components, compared between the QMS fit and ab initio ICE-CI values. The overall fit is in good agreement with the ab initio values over a broad range of θ_T values, with very low scatter. The largest

deviation is found where θ_T is lowest, i.e. where the quadrupole moment interaction strength is the weakest.

Table 5.1: Numerical values and units of the quadrupole moment surface function, Eq. (5.24).

Geometry		Charges	
a [Å]	0.5149	A	0.9763
b	−1.1271	B	0.6418
c [$Å^{-1}$]	0.5146	C	0.7251
d [rad]	3.5908	D	−1.0603
e	−0.1081	$q_{\mathrm{eq}}^{\mathrm{H}}$	0.3310
r_{eq} [Å]	0.9578		
θ_{eq} [rad]	1.8240		

5.4 Forces

With the various expressions given in the preceding section analytical atomic force components can be obtained and are derived from the negative gradient of the total energy expression, Eq. (5.15), with respect to the position of atom a in molecule i, or

$$\begin{aligned} F_{\alpha}^{ia} &= -\frac{dE_{\mathrm{tot}}}{dr_{\alpha}^{ia}} \\ &= -\frac{\partial E_{\mathrm{elst}}}{\partial r_{\alpha}^{ia}} - \frac{\partial E_{\mathrm{non-elst}}}{\partial r_{\alpha}^{ia}} - \frac{\partial E_{\mathrm{mon}}}{\partial r_{\alpha}^{ia}} \end{aligned} \tag{5.33}$$

The first term on the right hand side result in several contributing factors to the atomic forces due to the definition of the principal axes, choice of expansion center and the DMS and QMS. The atomic forces resulting from the simple pair-wise potentials describing the non-electrostatic terms are omitted for the sake of brevity, and the atomic forces due to the monomer energy expression – the PS-PES – are accounted for in their original work. [74]

The first term on the right hand side of Eq. (5.33), the total intermolecular electrostatic interaction, can be further divided into four contributions

$$-\frac{\partial E_{\text{elst}}}{\partial r_\alpha^{ia}} = -\frac{\partial E_{\text{elst}}}{\partial \mu_\beta^j(\{\mathbf{r}^{jb}\})}\frac{\partial \mu_\beta^j(\{\mathbf{r}^{jb}\})}{\partial r_\alpha^{ia}} - \frac{\partial E_{\text{elst}}}{\partial \theta_{\beta\gamma}^j(\{\mathbf{r}^{jb}\})}\frac{\partial \theta_{\beta\gamma}^j(\{\mathbf{r}^{jb}\})}{\partial r_\alpha^{ia}} - \frac{\partial E_{\text{elst}}}{\partial V_{\beta\gamma\delta\epsilon\ldots\eta}^j}\frac{\partial V_{\beta\gamma\delta\epsilon\ldots\eta}^j}{\partial r_\alpha^{ia}} - \frac{\partial E_{\text{elst}}}{\partial R_{\eta\beta}^j}\frac{\partial R_{\eta\beta}^j}{\partial r_\alpha^{ia}} \tag{5.34}$$

which are, in order, the partial derivative of the DMS and QMS, partial derivative of the external field and gradients thereof, and partial derivatives of the local-to-global rotation matrices as defined in Eq. (5.1)–Eq. (5.4).

At self-consistency of the iterative process which minimizes the energy in terms of the polarized moments the following conditions apply

$$\frac{\partial E_{\text{elst}}}{\partial \Delta\mu_\alpha^i} = \frac{\partial E_{\text{elst}}}{\partial \Delta\theta_{\alpha\beta}^i} = \frac{\partial E_{\text{self}}}{\partial \Delta\mu_\alpha^i} = \frac{\partial E_{\text{self}}}{\partial \Delta\theta_{\alpha\beta}^i} = 0$$

There are no explicit force contributions from the self-energy terms due to the on-site external field as the self-energy can be written solely in terms of the on-site induced moments (see the Supplementary Information). This results in a non-trivial additional condition

$$\frac{\partial E_{\text{self}}}{\partial V_{\beta\gamma\delta\epsilon\ldots\eta}^j} = 0 \tag{5.35}$$

Due to these conditions of the self-energy a single force contribution arises and is due to the local-to-global transformation of the fixed polarizability tensors

$$-\frac{\partial E_{\text{self}}}{\partial r_\alpha^{ia}} = -\frac{\partial E_{\text{self}}}{\partial R_{\eta\beta}^j}\frac{\partial R_{\eta\beta}^j}{\partial r_\alpha^{ia}} \tag{5.36}$$

The total force contribution due to the intermolecular electrostatic interaction and intramolecular self-energy is then

$$-\left(\frac{\partial E_{\text{in+pol}}}{\partial r_\alpha^{ia}} + \frac{\partial E_{\text{self}}}{\partial r_\alpha^{ia}}\right) = -\frac{\partial E_{\text{in+pol}}}{\partial \mu_\beta^j(\{\mathbf{r}^{jb}\})}\frac{\partial \mu_\beta^j(\{\mathbf{r}^{jb}\})}{\partial r_\alpha^{ia}} - \frac{\partial E_{\text{in+pol}}}{\partial \theta_{\beta\gamma}^j(\{\mathbf{r}^{jb}\})}\frac{\partial \theta_{\beta\gamma}^j(\{\mathbf{r}^{jb}\})}{\partial r_\alpha^{ia}} - \frac{\partial E_{\text{in+pol}}}{\partial V_{\beta\gamma\delta\epsilon\ldots\eta}^j}\frac{\partial V_{\beta\gamma\delta\epsilon\ldots\eta}^j}{\partial r_\alpha^{ia}} - \left(\frac{\partial E_{\text{in+pol}}}{\partial R_{\eta\beta}^i} + \frac{\partial E_{\text{self}}}{\partial R_{\eta\beta}^i}\right)\frac{\partial R_{\eta\beta}^i}{\partial r_\alpha^{ia}} \tag{5.37}$$

The terms in the expression above are given explicitly in the Supporting Information. It is noted that in order to evaluate the first term on the right hand side, explicit partial charge derivatives with respect to atomic positions of the DMS are required, which were not included in the original work on the PS–PES. [74] These are provided by Burnham and Xantheas, first used in the development of a flexible Thole-type multipole moment expansion potential. [32]

5.5 Flexible Model Fit

With the introduction of the DMS and QMS, the Gaussian type interaction tensor screening functions, as well as the changes to the pair-wise repulsion function, all of the five model parameters which affect the intermolecular interactions g, τ_d, A_{ref}, k and h are re-fitted. The fitting is performed with the same least-squares optimization module used for the QMS fit. The same numerical values are used for the fixed octupole and hexadecapole, as well as the dipole-dipole, dipole-quadrupole and quadrupole-quadrupole polarizability as in the original SCME model. [89] The Fortran based SCME/f code is freely available online [97], and includes an interface to the Python based Atomic Simulation Environment [98, 99] library. The data set used for the fitting

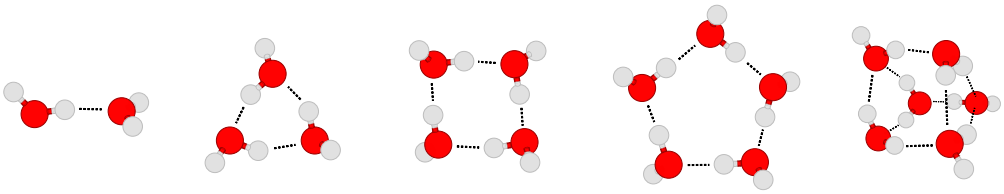


Figure 5.5: The lowest lying water cluster $(H_2O)_n$ isomers for n=2-6 used in the fitting procedure for the model parameters. From left to right; dimer (Cs), trimer (UUD), quadromer (S4), pentamer (cyclic, CYC) and hexamer (prism, PRI).

includes several points around the minimum of the dimer binding curve with the energy minimum and oxygen-oxygen distance corresponding to CCSD(T) calculations. [80] A single interaction energy for the lowest lying trimer, tetramer, pentamer and hexamer

is included. Fig. 5.5 shows the geometry of the lowest-lying water clusters $(H_2O)_n$ in the range $n = 2 - 6$. The reference calculations which are used here include the complete basis set limit CCSD(T) energies of the low-lying water hexamer structures by Bates and Tschumper. [79] For the other cluster sizes – trimers, tetramers and pentamers – complete basis set limit RI-MP2 calculations, with CCSD(T) corrections, are used. [80]

Table 5.2: Properties of crystal ice Ih evaluated with SCME [9, 10] and SCME/f, compared to experimental values. $\langle r_{OO} \rangle$ is the average oxygen-oxygen distance, a, b, c the lattice parameters for a dipole-free orthorhombic cell (containing eight molecules). V_0^{ZPE} (V_0) is the optimized cell volume, E_{lat}^{ZPE} (E_{lat}) and B_0^{ZPE} (B_0) are the lattice energy and bulk modulus with (and without) zero-point energy correction, all expressed per molecule.

Property	SCME	SCME/f	Exp.[1]
$\langle r_{OO} \rangle$ [Å]	2.742	2.751	2.751
a [Å]	4.470	4.478	4.497
b [Å]	7.747	7.777	7.789
c [Å]	7.287	7.331	7.321
V_0 [Å^3]	31.55	30.38	
V_0^{ZPE} [Å^3]		31.98	32.05
E_{lat} [eV]	−0.611	−0.645	−0.611
E_{lat}^{ZPE} [eV]		−0.489	−0.491
B_0 [GPa]	11.4	15.0	
B_0^{ZPE} [GPa]		12.2	10.9

1 Experimental values: average oxygen-oxygen distance is from Bjerrum [100], lattice parameters from Röttger *et al.* [101] (and resulting V_0^{ZPE}), enthalpy of vaporization (E_{lat}^{ZPE}) and lattice energy (E_{lat}) from Whalley [102], and bulk modulus from Hobbs [103].

In addition to the clusters, properties of hexagonal ice (ice Ih) have been considered, which is the most common ice phase. There are no high-level first-principles calculations with sufficient accuracy to serve as reference values. Instead, experimental data

for lattice constants, unit cell volume, bulk modulus and lattice energies need to be used, which generally include zero-point energy (ZPE) effects – and these effects are quite sizeable. [102, 104–106] Consequently, I have performed phonon calculations with the SCME/f model for proton disordered units cells of ice Ih containing 96 water molecules using the Parlinski-Li-Kawazoe finite-displacement method [107] as implemented in the PHONOPY package [108] using $3 \times 3 \times 3$ supercells and a displacement of 0.01 Å. For a unit cell with fixed cell vectors a geometry relaxation has been performed, which employs the analytical SCME/f forces with a force threshold of 10^{-3} eV/Å. Then, using a $10 \times 10 \times 10$ q-point sampling for the Brillouin zone integration, A numerically converged phonon density of states $g(\omega)$ has been obtained, the first moment of which provides the ZPE

$$E_{\mathrm{ZPE}} = \tfrac{\hbar}{2} \int_0^\infty \omega \, g(\omega) \, d\omega \quad . \tag{5.38}$$

Considering the dependence of the phonon frequencies on the unit cell volume $\omega = \omega(V)$ within the so-called quasi-harmonic approximation yields a ZPE-corrected energy-volume curve

$$E_{\mathrm{tot+ZPE}}(V) = E_{\mathrm{tot}}(V) + E_{\mathrm{ZPE}}(V) \quad , \tag{5.39}$$

where the energy zero is such that it describes infinitely separated (non-bound) individual water molecules. By fitting the Rose-Vinet equation of state [109] the minimum $E_{\mathrm{lat}}^{\mathrm{ZPE}} = E_{\mathrm{tot+ZPE}}(V_0^{\mathrm{ZPE}})$ of that curve together with the ZPE-corrected bulk modulus B_0^{ZPE} has been obtained (see supporting information for more details), which can be compared against accurate experimental data. [101–103] In order to include this data in the fitting process, an initial $E_{\mathrm{tot+ZPE}}(V)$ was calculated based on the SCME/f parameters first determined by fitting the data set derived from the water clusters. Then, $E_{\mathrm{tot}}(V)$ was improved by further parameter adjustments such that the expected ZPE correction would bring it close to the experimental values.

This trial and error scheme was found necessary since the phonon calculations are significantly more expensive than the calculation of the cluster properties. The end results based on a new set of phonon calculations is presented in Table 5.2 and shows good agreement with the experimental target properties. (The concomitant energy-volume

Table 5.3: Intermolecular interaction model parameters, numerical values and units.

Damping		Repulsion	
τ_d [Å^{-1}]	7.5548	A_{rep} [eV]	8149.63
g [Å]	1.1045	k	0.5515
		h [Å^{-1}]	3.4695

curves are shown in the supporting information.) Table 5.3 compiles the concomitant final optimized parameters of the SCME/f model.

Table 5.4 shows the resulting interaction energy and relative interaction energy versus the reference CCSD(T) calculations of the lowest lying isomers used in the fit. This includes a structural analysis comparing the relaxed SCME/f structure to the CCSD(T) reference structures, where the RMS deviation of nearest neighbor oxygen-oxygen distances, $\langle d\mathbf{r}_{\text{OO}}\rangle$, intramolecular oxygen-hydrogen bond lengths of the donor hydrogens, $\langle d\mathbf{r}_{\text{OH}}\rangle$, hydrogen bonding (H-bond) bond lengths, $\langle d\mathbf{r}_{\text{O}\cdots\text{H}}\rangle$, and angles between oxygen-hydrogen-oxygen in H-bonds, $\langle d\angle\text{OHO}\rangle$, are presented. The overall RMSD of the atomic positions is also presented, $\langle d\mathbf{r}^a\rangle$, and is evaluated with the Kabsch algorithm [110]. The interaction energies for the different cluster sizes are reproduced to a reasonable degree, with sub kcal/mol difference compared to the CCSD(T) results, except for the prism isomer of the hexamer where the interaction energy is overestimated by 1.18 kcal mol^{-1}. The resulting relaxed structures are in an overall very good agreement with the reference structures, with small variations in the second decimal in terms of atomic distances. Similarly, the angles between OHO in H-bonds are in a good agreement with the reference. The largest deviation is found in the angle between the donor-acceptor in the dimer.

5.6 Model Validation

For further validation of the new model the interaction energies and relative energy differences of all higher lying isomers of the pentamers and hexamers are calculated, which are not included in the fitting data set, and compared to the relative energies from

Table 5.4: Interaction energy (kcal/mol) and distances (Å) between atoms in the most stable configuration of clusters $(H_2O)_n$ with n=2,$\cdots$,6. $E_{\rm int}$ is the SCME/f calculated interaction energy of the clusters and $\Delta E_{\rm int}$ (kcal/mol) the difference with respect to the CCSD(T) values. $\langle d\mathbf{r}_{\rm OO}\rangle$, $\langle d\mathbf{r}_{\rm OH}\rangle$ and $\langle d\mathbf{r}_{\rm O\cdots H}\rangle$ are the RMSD of the oxygen-oxygen neighbour distances, intramolecular oxygen-hydrogen bond lengths of the donor-hydrogen and bonding oxygen$\cdots$hydrogen bond length distances, respectively, compared to the CCSD(T) obtained structures. [79, 80] $\langle d\mathbf{r}^a\rangle$ is the overall RMSD of the relaxed SCME/f structure evaluated using the Kabsch algorithm [110]. $\langle d\angle{\rm OHO}\rangle$ is the RMSD of the angle (in degrees) between the oxygen-hydrogen-oxygen in hydrogen bonds.

$(H_2O)_n$	$E_{\rm int}$	$\Delta E_{\rm int}$	$\langle d\mathbf{r}_{\rm OO}\rangle$	$\langle d\mathbf{r}_{\rm OH}\rangle$	$\langle d\mathbf{r}_{\rm O\cdots H}\rangle$	$\langle d\mathbf{r}^a\rangle$	$\langle d\angle{\rm OHO}\rangle$
2-Cs	−4.85	+0.18	0.011	0.000	0.017	0.017	5.923
3-UUD	−15.16	+0.54	0.035	0.010	0.037	0.037	2.489
4-S4	−27.51	−0.11	0.005	0.014	0.006	0.045	1.382
5-CYC	−36.72	−0.71	0.014	0.015	0.003	0.046	0.369
6-PRI	−47.10	−1.18	0.017	0.012	0.035	0.033	4.564

the quantum chemistry references. [79, 80] The trends are shown in Figs. 5.6 and 5.7, and the trend predicted with the rigid SCME is shown for comparison. All structures are relaxed with a force tolerance of $10^{-4}\,{\rm eV\,\AA}^{-1}$, and results collected in Table 5.5, which also presents the RMS difference between the relaxed SCME/f structures and the quantum chemistry reference structures.

For the pentamers, Fig. 5.6, most of the relative energy difference trend is captured with the exception of isomer FRA, whose relative stability is underestimated. Another key difference between SCME/f and the reference calculations is the series of CAA-CAB isomers, which have a cagelike structure. In particular the cage structure of isomers CAA and CAB are not stable and rearrange to isomers which are more akin to the fused ring structures of the FRA-FRC isomers. The resulting SCME/f structures of CAA and CAB are near identical, with an interaction energy difference of only 0.01

Table 5.5: Energies and relative energy and structural properties of the pentamer and hexamer isomers. See the caption of Table 5.4 for the definition of the table entries.

$(H_2O)_n$	$E_{\rm int}$	$\Delta E_{\rm int}$	$\langle d\mathbf{r}_{\rm OO}\rangle$	$\langle d\mathbf{r}_{\rm OH}\rangle$	$\langle d\mathbf{r}_{\rm O\cdots H}\rangle$	$\langle d\mathbf{r}^a\rangle$	$\langle d\angle{\rm OHO}\rangle$
5-FRB	−35.60	−0.72	0.026	0.012	0.041	0.036	4.008
5-CAC	−35.50	−0.81	0.053	0.012	0.089	0.136	8.548
5-CAA	−35.07	−0.53	0.060	0.012	0.053	0.254	9.281
5-CAB	−35.06	−1.23	0.065	0.011	0.080	0.237	6.107
5-FRC	−33.56	−1.12	0.025	0.013	0.026	0.043	1.859
5-FRA	−32.91	0.22	0.025	0.013	0.032	0.059	1.731
6-CAG	−46.44	−0.74	0.013	0.017	0.019	0.054	1.607
6-BK1	−46.37	−1.09	0.014	0.015	0.009	0.033	2.346
6-BK2	−46.26	−1.35	0.014	0.016	0.008	0.038	3.433
6-BAG	−45.90	−1.52	0.015	0.017	0.012	0.065	3.826
6-CYR	−45.36	−1.00	0.012	0.015	0.006	0.018	3.913
6-CB1	−44.57	−1.23	0.013	0.015	0.003	0.031	2.924
6-CB2	−44.49	−1.20	0.013	0.015	0.005	0.025	2.284

1 Pentamers; fused-ring-B (FRB), cage-C (CAC), cage-A (CAA), cage-B (CAB), fused-ring-C (FAC) and fused-ring-A (FRA); and the hexamers; cage (CAG), book-1 (BK1), book-2 (BK2), bag (BAG), cyclic-ring (CYR), cyclic-boat-1 (CB1) and cyclic-boat-2 (CB2).

kcal/mol. Only the CAC isomer keeps its cagelike structure, but one of the H-bonds is not stable (between a donor acceptor oxygen with distance greater than 3.0 Å), resulting in a rotation of one of the water monomers. Compared to the rigid SCME predecessor this represents an improvement, in particular for the FRB, CAC and CAA isomers, whose stability is greatly underestimated relative to the CYC isomer.

For the hexamers, Fig. 5.7, the overall trend in the relative interaction energies is captured to a good degree compared to the CCSD(T) reference, and is a substantial improvement over the rigid SCME model, which greatly underestimates the stability of the prism isomer relative to all other isomers. The bond lengths and angles of

the hexamer isomers are all in very good agreement with the reference structures, with small differences in the second or third decimal in terms of the bond lengths, and the H-bonded OHO angles deviate by only 2-4°. Tables 5.6 and 5.7 presents vibrational frequency analysis of the lowest lying isomers, including the cyclic ring isomer of the hexamer. The RMS deviation from near-CBS CCSD(T) calculations [111, 112] are presented for the different classes of modes. These are intermolecular low-energy vibrarional modes (10-1000 cm^{-1}), intramonomer bending modes (1600-1800 cm^{-1}) and H-bonded and non-bonded OH stretching modes (ca. 3200-3900 cm^{-1}). For comparison, the same analysis is performed for the SCME/f model, but with the quadrupole moment fixed and corresponding to the numerical value of the quadrupole moment for the ground state monomer configuration.

With the inclusion of the QMS (left column Tables 5.6 and 5.7) the low-energy vibrational modes and, in particular, the bending modes are in a good agreement with the reference calculations. The RMS deviation ranges from 18 to 23 cm^{-1} and 7 to 14 cm^{-1} for the two classes of modes, respectively. The maximum difference in the bending modes does not exceed 20 cm^{-1} for any of the clusters analyzed. The red shift of the H-bonded OH stretches is, however, not captured by our model, resulting in an overestimation of these modes, which becomes systematically larger with cluster size. This is due to the underlying monomer potential energy surface, whose limit in terms of hydrogen dissociation is $OH^{\cdot} + H^{\cdot}$ whereas should be in the condensed phase $OH^{-} + H^{+}$. The model potential does not describe this important change, and the resulting weakening of oxygen-hydrogen bonds in H-bonding OH. The high-frequency modes for the dimer are though in a reasonable agreement with the reference calculations.

A comparison to the same vibrational frequency analysis is performed with the quadrupole moment fixed (right column, Tables 5.6 and 5.7). Fixing the quadrupole moment results in a drastic change in the difference between all of the types of modes and for all cluster sizes, with for example a RMS deviation of up to three times greater for the bending modes. The overall agreement with the reference calculations of all modes is consistently worse, in particular for the larger cluster, n=4-6. Only the low-frequency

Table 5.6: Relative vibrational properties of three lowest lying water clusters $(H_2O)_n$ with n=2,3,4 (Cs, UUD, S4). The entries for each system correspond from top to bottom, the low-frequency intermolecular vibrational modes (l, 10 to 1000 cm^{-1}), intramonomer bending (b, 1600 to 1800 cm^{-1}) and high frequency stretching of H-bond OH and non-bonded OH bonds (h, 3200 to 3900 cm^{-1}). $\langle \Delta cm^{-1} \rangle$ is the RMSD between the frequencies in the low, medium and high range, as predicted with SCME/f compared to near-CBS CCSD(T) reference calculations. [111, 112] The last entry is RMSD for the total frequency range (**t**), where the value in the parenthesis excludes the overestimated H-bond OH stretches. $\max|\Delta cm^{-1}|$ is the maximum absolute difference for each entry. The two columns on the right are for the SCME/f model potential with the quadrupole moment set to a fixed value corresponding to the ground state water monomer configuration.

		quadrupole moment surface			fixed quadrupole moment		
$(H_2O)_n$		$\langle \Delta cm^{-1} \rangle$		$\max\|\Delta cm^{-1}\|$	$\langle \Delta cm^{-1} \rangle$		$\max\|\Delta cm^{-1}\|$
2-Cs	l	17.52		41.50	15.33		23.10
	b	6.71		9.10	9.18		12.70
	h	28.55		36.10	50.41		70.30
	t	**17.23**			**23.56**		
3-UUD	l	23.03		49.20	34.43		71.70
	b	7.75		12.20	18.41		23.90
	h	68.09		89.30	153.65		190.40
	t	**31.20**	(24.98)		**64.02**	(29.19)	
4-S4	l	20.87		44.40	29.43		58.10
	b	10.74		12.80	25.69		29.40
	h	155.58		200.20	266.11		334.10
	t	**59.19**	(21.06)		**100.25**	(26.49)	

Table 5.7: Same as Table 5.6, but for different water clusters $(H_2O)_n$ with n=5,6 (CYC, CYR, PRI).

		quadrupole moment surface			fixed quadrupole moment		
$(H_2O)_n$		$\langle\Delta\text{cm}^{-1}\rangle$		$\max\lvert\Delta\text{cm}^{-1}\rvert$	$\langle\Delta\text{cm}^{-1}\rangle$		$\max\lvert\Delta\text{cm}^{-1}\rvert$
5-CYC	l	18.47		35.20	35.08		65.30
	b	14.22		19.60	26.44		34.40
	h	179.34		229.60	282.41		358.10
	t	**66.02**	(20.13)		**105.22**	(31.38)	
6-CYR	l	21.10		44.30	35.16		75.40
	b	11.55		13.10	32.53		40.00
	h	185.38		239.30	280.06		356.30
	t	**67.75**	(23.24)		**103.48**	(32.56)	
6-PRI	l	21.70		87.60	27.74		56.20
	b	9.86		18.30	35.16		45.80
	h	208.58		408.60	313.99		571.80
	t	**75.79**	(24.60)		**113.85**	(36.40)	

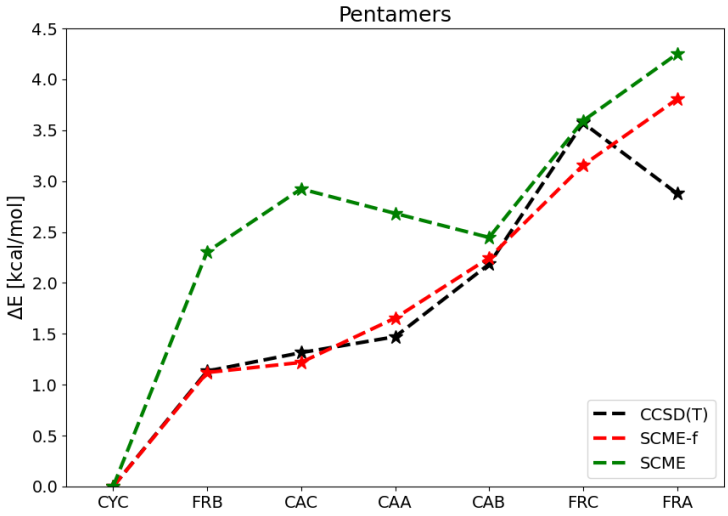


Figure 5.6: Relative energy difference for the lowest lying pentamers water cluster isomers. The results for the rigid version of SCME [9, 89] and SCME/f are compared. Relative energy differences from high level quantum chemistry calculations are also shown. RI-MP2 energies at the complete basis set limit with CCSD(T) corrections [80] (MP2/CBS+ΔCCSD(T)). The acronyms from left to right are the different isomers: Cyclic (CYC), fused-ring-B (FRB), cage-C (CAC), cage-A (CAA), cage-B (CAB), fused-ring-C (FAC) and fused-ring-A (FRA).

modes of the Cs dimer seem improved by fixing the quadrupole moment. While the parametrization of the intermolecular interaction parameters is with the QMS included, the structural properties and interaction energy of the small clusters are not drastically changed with the quadrupole moment fixed (see Supplementary Information).

It is also of interest to analyze the structure of the monomers in crystal ice Ih with or without the QMS included. Table 5.8 presents the average internal HOH angle of each water monomer in the crystal lattice, extracted at volume V_0^{ZPE}, and compares to the experimental value of the angle for the isolated monomer and in crystal ice Ih. The

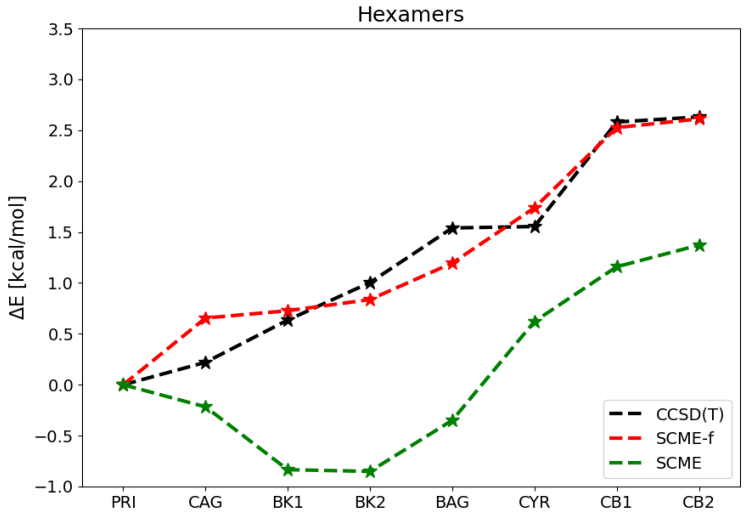


Figure 5.7: The same as Fig. 5.6 but for hexamer water cluster isomers. CCSD(T) energies at the complete basis set limit (CCSD(T)/CBS) [79]. The acronyms from left to right are the different isomers: Prism (PRI), cage (CAG), book-1 (BK1), book-2 (BK2), bag (BAG), cyclic-ring (CYR), cyclic-boat-1 (CB1) and cyclic-boat-2 (CB2).

experiments show a clear widening of the monomer HOH angle by about 3.5 degrees (104.5°–108.1°) going from the gas to crystal phase. Without the QMS the trend is opposite, with the angle favoring lower values by about 4.5 degrees (104.5°–99.95°), where the dipole moment is high. The correct trend is captured again with the inclusion of the QMS, with the angle widening by about 2 degrees (104.5°–106.51°). The QMS correctly balances the magnitude of the dipole moment and principal quadrupole moment in the lattice, and in such a way that a widening of the angle is favoured.

5.7 Discussion and Conclusions

An extension of the SCME potential function has been presented for water molecules to allow for distortion of the molecular structure. In addition to the dipole moment

Table 5.8: Average intramolecular HOH angles (in degrees) for the SCME/f model with and without the QMS. Experimental angles for the isolate water molecule (gas) and in crystal ice Ih (Ih) are presented for comparison.

	Exp (gas)	Exp (Ih)	SCME/f	SCME/f no QMS
$\langle\angle$HOH$\rangle$	104.5	108.1	106.51	99.95

surface, this flexible potential function, SCME/f, includes a mapping of the quadrupole moment surface which has not been previously included at this level of detail to our knowledge. A simpler model for the quadrupole moment that has been used in both rigid and flexible point charge based potential functions [2, 75–78], as well as more sophisticated polarizable models [32–40], make use of the so-called M-site. We now digress in a brief comparison between the QMS model described in this work and the M-site model.

In the M-site model the partial charge associated with the oxygen is moved off the atomic center to a position behind the oxygen and on to the bisector defined by the two OH bond vectors. The position of the M-site in the global coordinate frame is written as [113–115]

$$\mathbf{r}_{\mathrm{M}} = (1-\gamma)\mathbf{r}_{\mathrm{O}} + \frac{\gamma}{2}(\mathbf{r}_{\mathrm{H}_1} + \mathbf{r}_{\mathrm{H}_2}) \tag{5.40}$$

where $0 < \gamma \leq 1$. For any finite value of γ the partial charges are re-scaled according to

$$q^{\mathrm{H}_l^\gamma} = \frac{q^{\mathrm{H}_l}}{1-\gamma}, \qquad q^{\mathrm{M}} = -q^{\mathrm{H}_1^\gamma} - q^{\mathrm{H}_2^\gamma} \tag{5.41}$$

such that the dipole moment remains unchanged in the M-site frame, and a single set of three partial charges describes both the dipole and quadrupole moment.

More importantly, a value of γ can be derived such that the Δ component in Eq. (5.32) vanishes, resulting in the compactly written moment tensor

$$\theta = \begin{bmatrix} \theta_T & 0 & 0 \\ 0 & -\theta_T & 0 \\ 0 & 0 & 0. \end{bmatrix} \tag{5.42}$$

This illustrates that the principal quadrupole moment component θ_T is origin independent, and is the rational for placing the partial charge on the M-site and not on the oxygen center. The strength of the quadrupole moment interaction is determined by θ_T. For the ground state PE-PES water monomer configuration used in this work a $\gamma = 0.4071$ results in a compact tensor of the form in Eq. (5.42) (see the Supplementary Information). Similar values for γ are reported in potential functions based on the M-site. While such a three site partial charge model can capture both the dipole and principal quadrupole moment for a fixed ground state monomer configuration, the question is how the model holds up in the case of a flexible water monomer.

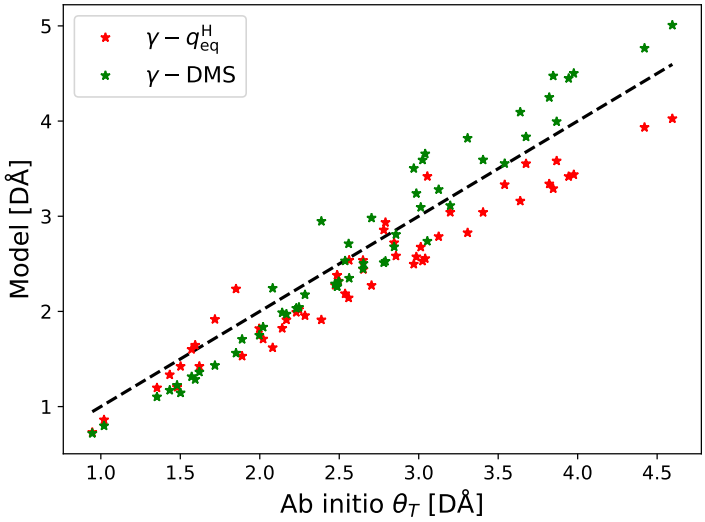


Figure 5.8: M-site model making use of a fixed charge ($\gamma - q_{eq}^{H}$, red) in the ground state monomer configuration or the DMS charges (γ – DMS, green). The fixed point charge model (red) tends to underestimate the strength of the quadrupole moment over the whole range, whereas when based on the DMS charges (green) the quadrupole moment is underestimated in the lower region and overestimated in the higher region (> 3.0 DÅ).

Using the ab inito ICE-CI quadrupole moment data four M-site models are considered and compared, and are representative of M-site models encountered in the literature.

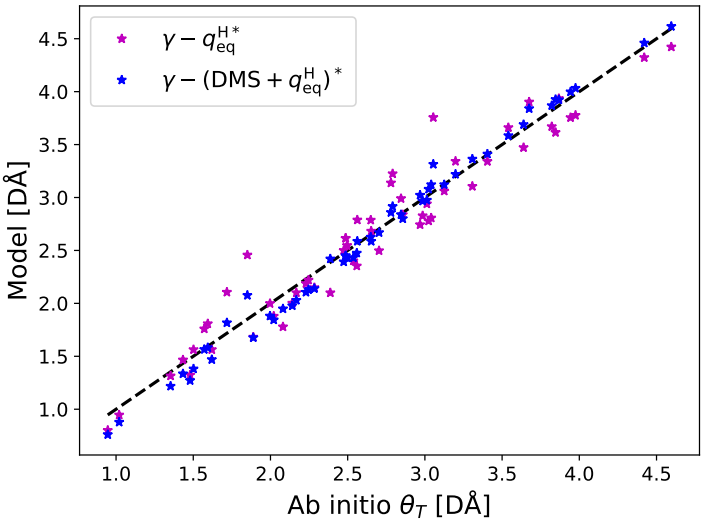


Figure 5.9: M-site model making use of scaled charge ($\gamma - q_{eq}^{H*}$, magenta) and a combination of scaled DMS and fixed charges as in Eq. (5.25) ($\gamma - (DMS + q_{eq}^{H})^*$, blue). The overall trend is better captured in Figs. 5.8 and 5.9, but the scatter is still substantial for the scaled fixed charge model (magenta). The agreement is improved substantially with the mixture of fixed charges and variable DMS charges (blue) compared to the other models, and the scatter is more concentrated in the region of low θ_T. See the Supplementary Information for details on the M-site models.

The details of the models and parameters are presented in the Supplementary Information. The first two models, Fig. 5.10 left, make use of $\gamma = 0.4071$ and a set of fixed partial charges ($\gamma - q_{eq}^{H}$) – corresponding to the partial charges of the ground state monomer configuration – or scaled ground state charges ($\gamma - q_{eq}^{H,*}$). The scaling parameter is fit such that the model best captures θ_T over the whole range. The fixed point charge model tends to underestimate the strength of the quadrupole moment over the whole range, whereas the scaling of the charge results in a change in the slope and overall better agreement. However, in both cases the scatter is substantial and the RMS difference between the trace components of the quadrupole moment versus the ab inito values is $> 10\%$ on average (see the Supplementary Information).

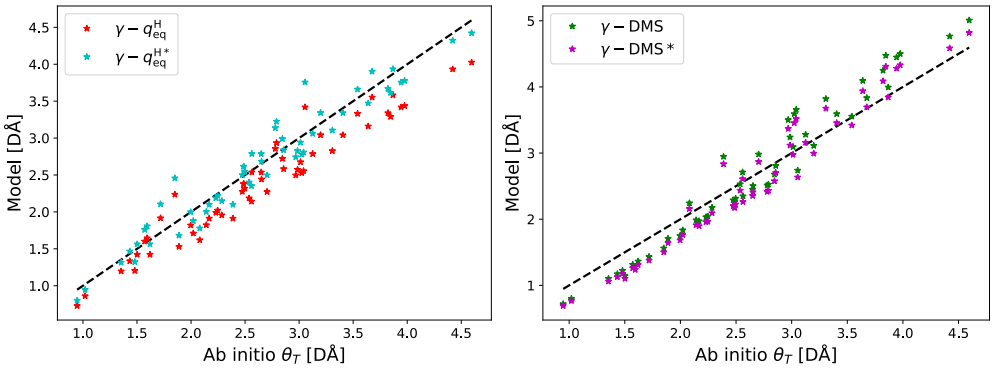


Figure 5.10: M-site models. See the Supplementary Information for details on the individual models. Left: M-site models making use of a fixed charge ($\gamma - q_{\mathrm{eq}}^{\mathrm{H}}$, red) corresponding to the ground state monomer configuration or scaled charge ($\gamma - q_{\mathrm{eq}}^{\mathrm{H},*}$, cyan). The fixed point charge model (red) tends to underestimate the strength of the quadrupole moment over the whole range, whereas the scaling results in a change in the slope and an overall better agreement. Right: M-site models making use of variable DMS charges ($\gamma - \mathrm{DMS}$, green) and scaled DMS charges ($\gamma - \mathrm{DMS}^*$), magenta). Both model tend to underestimate the strength of the quadrupole moment in region of low strength, whereas overestimate in the region of large strength. The agreement is only slightly improved with the scaled DMS model, corresponding to a shift of the M-site to $\gamma = 0.3838$.

In the third and fourth model, Fig. 5.10 right, the charge are described with the DMS charge. In the third model the optimal γ value is used ($\gamma - \mathrm{DMS}$) and in the fourth model the DMS charges are scaled ($\gamma - \mathrm{DMS}^*$) to best capture θ_T over the whole range. The qualitative trend is the same in both cases, with the strength of the quadrupole moment underestimated in the region of low strength, and overestimated in the region of large strength, and the overall agreement is only slightly improved with a change in the slope. Similar to the fixed charge models the scatter is substantial, and the RMS difference is found to be $\approx 10\%$, on average.

While the simple M-site models capture the overall qualitative trend in the change of the principal quadrupole moment over a broad range of configurations, an analysis of the RMS difference of the quadrupole moment components shows that they deviate significantly for monomer configurations different than the ground state configuration. Neither the fixed charge or DMS charge M-site models (scaled or not) seem to better capture the principal quadrupole component over the other, and in all cases the RMS difference is around 10% or greater. This illustrates that a three site model based on the M-site principle is not able to capture the variation of the quadrupole moment in a flexible water potential model to a good degree. The four site QMS model developed in this work, which captures the principal quadrupole moment with a mean absolute error of 0.04 DÅ, similarly has low scatter throughout the range with an average RMS difference of 1.6%, with greatest discrepancy in the region where the quadrupole moment interaction is the weakest.

Furthermore, the intermolecular interactions of the SCME/f model only depend on five parameters. The parameters have been fitted to reproduce high level quantum chemistry calculations for the water dimer energy surface near the equilibrium geometry and interaction energy of the lowest-lying water clusters up to and including the hexamer, as well as the properties of the Ih ice crystal – and in such a way that experimental values are reproduced to a good degree after including zero point energy corrections.

The simple parameterization of the flexible model and the use of a single center for

the electrostatic interactions allows for the seamless integration into our recently implemented PE-QM/MM interface [72, 73].

The calculated energy of the higher lying energy isomers of hexamer water cluster are found to be in a reasonable agreement with the results of CCSD(T) calculations in the complete basis set limit. [79] The relative trend in the energy differences between the isomers, as well as the overall structures are captured to a good degree. This represents a significant improvement over the rigid SCME potential function and is on par with the trend predicted with the HBB2-pol [36, 37] potential function, which explicitly models the N-body expansion up to the three-body terms in the interaction energy and is the predecessor of the MB-pol potential function. [38–40] However, discrepancies are present in the series of pentamer isomers, in particular the cage-like isomers. H-bonds in bonds where the distance is greater than 3 Å are found to be unstable, leading to a rearrangement of some of the SCME/f structures compared to the reference structures.

Analysis of the vibrational modes of the small water clusters reveal a substantial improvement with the QMS mapping included (as opposed to a fixed value). In particular are the intramolecular bending modes in the range 1600 to 1800 cm^{-1}, with maximum absolute deviation consistently less than 20 cm^{-1} with the QMS included, compared to near-CBS CCSD(T) calculations. [111, 112] Importantly, including the DMS only results in the opposite trend of the intramolecular angle widening in crystal ice Ih compared to the gas phase. The inclusion of the QMS recovers the correct trend due to the balance between the magnitude of the dipole and principal quadrupole moment which are functions of the internal geometry and strongly dependent on this angle.

While the results presented here represent an important step forward in the development of a single center multipole expansion model for water, there is room for improvement, and this will be addressed in future work. A natural next step to the mapping of the dipole and the quadrupole is to incorporate a mapping of the polarizability tensors. Work is ongoing to incorporate the intramolecular geometry dependent mapping of the dipole-dipole, dipole-quadrupole and quadrupole-quadrupole polarizability tensors by

Loboda *et al.* [116]. It has been suggested that a critical part of the H-bond OH softening lies in the correct mapping of the polarizability surface of the individual monomers. [35]

In particular, and in order to further address the overestimated H-bonded OH stretches, an improvement of the underlying water monomer potential energy surface – whose limit in terms of hydrogen dissociation is $OH^{\cdot} + H^{\cdot}$ – must be made when there are neighboring water molecules such that it approaches to some degree the dissociation limit in a condensed phase which is $OH^{-} + H^{+}$. In order to capture this one could modify the DMS and QMS charges to better represent this limit, and in a way which depends on the environment. Modifying the charge of the DMS has, for example, previously been considered in water potentials in order to capture the charge delocalization and resulting softening of the H-bond, such as in the TTM3-F model. [34] Further improvements to this flexible SCME model that are being pursued include a more elaborate repulsive part including deviations from spherical symmetry.

5.A Appendix

The supporting information includes a detailed derivation of the atomic forces corresponding to contributions presented in Eq. (5.37), as well as a comparison between the numerical and analytical forces as the convergence criteria of the induced moments is varied. The parameters used for the model M-site description of the principal quadrupole moment are presented, followed by an analysis of the RMSD between ab initio versus the QMS quadrupole as well as model M-site quadrupoles with respect to geometrical variation of the monomer. Binding energies and relative structural properties of the lowest-lying water clusters are given for the case where the quadrupole moment is set to a fixed value corresponding to the ground state monomer configuration. Finally, the evaluation of the bulk properties from fitting the energy-volume relation – with and without zero-point energy corrections – is described.

5.A.1 Analytical Forces

We further apply the chain rule considering the main electrostatic plus induction force expression in the main text

$$-\left(\frac{\partial E_{\text{ele+ind}}}{\partial r_\alpha^{ia}}+\frac{\partial E_{\text{self}}}{\partial r_\alpha^{ia}}\right)=-\frac{\partial E_{\text{ele+ind}}}{\partial \mu_\beta^j(\{\mathbf{r}^{jb}\})}\frac{\partial \mu_\beta^j(\{\mathbf{r}^{jb}\})}{\partial r_\alpha^{ia}}-\frac{\partial E_{\text{ele+ind}}}{\partial \theta_{\beta\gamma}^j(\{\mathbf{r}^{jb}\})}\frac{\partial \theta_{\beta\gamma}^j(\{\mathbf{r}^{jb}\})}{\partial r_\alpha^{ia}}$$
$$-\frac{\partial E_{\text{ele+ind}}}{\partial V_{\beta\gamma\delta\epsilon\ldots\eta}^j}\frac{\partial V_{\beta\gamma\delta\epsilon\ldots\eta}^j}{\partial r_\alpha^{ia}}-\left(\frac{\partial E_{\text{ele+ind}}}{\partial R_{\nu o}^j}+\frac{\partial E_{\text{self}}}{\partial R_{\nu o}^j}\right)\frac{\partial R_{\nu o}^j}{\partial r_\alpha^{ia}} \tag{5.43}$$

The last term on the right hand side describes the force contribution due to the definition of the local-to-global reference frame transformation, and is the only term which includes an explicit contribution to the atomic forces due to the self-energies. To see this we first write the MM induced dipoles and quadrupoles as

$$\Delta\mu_\alpha^i=-\alpha_{\alpha\beta}V_\beta^i-\frac{1}{3}A_{\alpha,\beta\gamma}V_{\beta\gamma}^i=\Delta\mu_\alpha^i(\alpha)+\Delta\mu_\alpha^i(A) \tag{5.44}$$

$$\Delta\theta_{\alpha\beta}^i=-A_{\gamma,\alpha\beta}V_\gamma^i-C_{\gamma\delta,\alpha\beta}V_{\gamma\delta}^i=\Delta\theta_{\alpha\beta}^i(A)+\Delta\theta_{\alpha\beta}^i(C) \tag{5.45}$$

where on the right hand side of the second equality the contribution from the external field and field gradient due to the on-site potential is split up. With these definitions it is easy to relate the external field and field gradient at site i to the self-consistent moments of molecule i

$$V^i_\beta = -\frac{\Delta\mu^i_\alpha(\alpha)}{\alpha_{\alpha\beta}} \tag{5.46}$$

$$V^i_\gamma = -\frac{\Delta\theta^i_{\alpha\beta}(A)}{A_{\gamma,\alpha\beta}} \tag{5.47}$$

and

$$V^i_{\gamma\delta} = -\frac{\Delta\theta^i_{\alpha\beta}(C)}{C_{\gamma\delta,\alpha\beta}} \tag{5.48}$$

$$V^i_{\beta\gamma} = -\frac{\Delta\mu^i_\alpha(A)}{A_{\alpha,\beta\gamma}}. \tag{5.49}$$

The self-energy of an induced dipole in linear response theory is

$$E^\mu_{\text{self}} = -\int_0^{\Delta\mu^i} V^i_\beta d\Delta\mu^i. \tag{5.50}$$

It gives the energy cost of inducing a first order moment in the potential field at site i. By inserting the relation in Eq. (5.46) into the equation above, and by considering only (for the moment) the induced dipole in in response to an external field gives

$$E^\mu_{\text{self}} = \int_0^{\Delta\mu^i(\alpha)} \frac{\Delta\mu^i_\alpha(\alpha)}{\alpha_{\alpha\beta}} d\Delta\mu^i = \frac{1}{2}\frac{\Delta\mu^i_\alpha(\alpha)\Delta\mu^i_\beta(\alpha)}{\alpha_{\alpha\beta}}. \tag{5.51}$$

For isotropic atomic polarization this becomes

$$E^{\text{iso}}_{\text{self}} = \frac{1}{2}\frac{(\Delta\mu^i)^2}{\alpha}. \tag{5.52}$$

This form is most frequently encountered in MM work based on isotropic atomic polarization and induced dipole in response to an external field. Similarly for the induced quadrupole

$$E^\theta_{\text{self}} = -\frac{1}{3}\int_0^{\Delta\theta^i} V^i_{\beta\gamma} d\Delta\theta^i \tag{5.53}$$

which expresses the energy cost of inducing a second order moment in the field gradient at site i. The factor of $1/3$ follows from the definition of the traceless Cartesian moments

[82] used in SCME. The total self-energy for a single site i in SCME is then

$$\begin{aligned}
E_{\text{self}} =& E^{\mu}_{\text{self}} + E^{\theta}_{\text{self}} \\
=& \int_0^{\Delta\mu^i} \frac{\Delta\mu^i_\alpha(\alpha)}{\alpha_{\alpha\beta}} d\Delta\mu^i + \frac{1}{3}\int_0^{\Delta\theta^i} \frac{\Delta\theta^i_{\alpha\beta}(C)}{C_{\gamma\delta,\alpha\beta}} d\Delta\theta^i \\
=& \frac{1}{2}\frac{\Delta\mu^i_\alpha(\alpha)}{\alpha_{\alpha\beta}}\left(\Delta\mu^i_\beta(\alpha) + \frac{1}{3}\Delta\mu^i_\beta(A)\right) + \frac{1}{6}\frac{\Delta\theta^i_{\alpha\beta}(C)}{C_{\gamma\delta,\alpha\beta}}\left(\Delta\theta^i_{\gamma\delta}(C) + \Delta\theta^i_{\gamma\delta}(A)\right) \\
=& \frac{1}{2}\frac{\Delta\mu^i_\alpha(\alpha)\Delta\mu^i_\beta(\alpha)}{\alpha_{\alpha\beta}} + \frac{1}{3}\frac{\Delta\mu^i_\alpha(\alpha)\Delta\theta^i_{\beta\gamma}(C)}{k_{\alpha,\beta\gamma}} + \frac{1}{6}\frac{\Delta\theta^i_{\alpha\beta}(C)\Delta\theta^i_{\gamma\delta}(C)}{C_{\gamma\delta,\alpha\beta}}
\end{aligned} \tag{5.54}$$

where the relations in Eqs. (5.46) and (5.49) are used. The matrix k is given by

$$k = \frac{\alpha C}{A} \tag{5.55}$$

This expression for the self-energies is very useful at self-consistency (SCF). First and foremost it shows that there are no force contributions arising from partial derivatives of the on-site potential field and field gradients when considering the self-energy terms, since at SCF we have

$$\frac{\partial E^{\text{sys}}_{\text{tot}}}{\partial\Delta\mu^i_\alpha} = \frac{\partial E^{\text{sys}}_{\text{tot}}}{\partial\Delta\theta^i_\alpha} = 0, \tag{5.56}$$

which implies

$$\frac{\partial E_{\text{self}}}{\partial V^{jb}_{\beta\gamma\delta\epsilon\ldots\eta}} = 0 \tag{5.57}$$

Contributions arise from the static octupole and static hexadecapole, as well as the dipole-dipole, dipole-quadrupole quadrupole-quadrupole polarizability matrices. The general contributions are of the following form

$$\frac{\partial E}{\partial R^j_{\nu o}}\frac{\partial R^j_{\nu o}}{\partial r^{ia}_\alpha} \to \delta_{ij}\delta_{\nu o,\beta\gamma} \tag{5.58}$$

resulting in for the static moments

$$\begin{aligned}
\frac{\partial E_{\text{ele+ind}}}{\partial R^j_{\nu o}}\frac{\partial R^j_{\nu o}}{\partial r^{ia}_\alpha} =& \frac{1}{15}\left(\frac{\partial R^i_{\eta\beta}}{\partial r^i_\alpha}R^i_{\tau\gamma}R^i_{\kappa\delta} + R^i_{\eta\beta}\frac{\partial R^i_{\tau\gamma}}{\partial r^i_\alpha}R^i_{\kappa\delta} + R^i_{\eta\beta}R^i_{\tau\gamma}\frac{\partial R^i_{\kappa\delta}}{\partial r^i_\alpha}\right)\Omega^{i'}_{\eta\tau\kappa}V^i_{\beta\gamma\delta} \\
&+ \frac{1}{105}\left(\frac{\partial R^i_{\eta\beta}}{\partial r^{ia}_\alpha}R^i_{\tau\gamma}R^i_{\kappa\delta}R^i_{\sigma\eta} + R^i_{\eta\beta}\frac{\partial R^i_{\tau\gamma}}{\partial r^{ia}_\alpha}R^i_{\kappa\delta}R^i_{\sigma\eta}\right. \\
&\left. + R^i_{\eta\beta}R^i_{\tau\gamma}\frac{\partial R^i_{\kappa\delta}}{\partial r^{ia}_\alpha}R^i_{\sigma\eta} + R^i_{\eta\beta}R^i_{\tau\gamma}R^i_{\kappa\delta}\frac{\partial R^i_{\sigma\eta}}{\partial r^{ia}_\alpha}\right)\Phi^{i'}_{\eta\tau\kappa\sigma}V^i_{\beta\gamma\delta\epsilon}
\end{aligned} \tag{5.59}$$

and for the polarizability matrices the contributions are

$$\left(\frac{\partial E_{\text{ele+ind}}}{\partial R^j_{\nu o}} + \frac{\partial E_{\text{self}}}{\partial R^j_{\nu o}}\right)\frac{\partial R^j_{\nu o}}{\partial r^{ia}_\alpha} = -\frac{1}{2}\left(\frac{\partial R^i_{\eta\beta}}{\partial r^i_\alpha}R^i_{\tau\gamma} + R^i_{\eta\beta}\frac{\partial R^i_{\tau\gamma}}{\partial r^i_\alpha}\right)\alpha^{i'}_{\eta\tau}V^i_\beta V^i_\gamma$$
$$-\frac{1}{3}\left(\frac{\partial R^i_{\eta\beta}}{\partial r^i_\alpha}R^i_{\tau\gamma}R^i_{\kappa\delta} + R^i_{\eta\beta}\frac{\partial R^i_{\tau\gamma}}{\partial r^i_\alpha}R^i_{\kappa\delta} + R^i_{\eta\beta}R^i_{\tau\gamma}\frac{\partial R^i_{\kappa\delta}}{\partial r^i_\alpha}\right)A^{i'}_{\eta\tau\kappa}V^i_\beta V^i_{\gamma\delta}$$
$$-\frac{1}{6}\left(\frac{\partial R^i_{\eta\beta}}{\partial r^i_\alpha}R^i_{\tau\gamma}R^i_{\kappa\delta}R^i_{\sigma\eta} + R^i_{\eta\beta}\frac{\partial R^i_{\tau\gamma}}{\partial r^i_\alpha}R^i_{\kappa\delta}R^i_{\sigma\eta}\right.$$
$$\left. + R^i_{\eta\beta}R^i_{\tau\gamma}\frac{\partial R^i_{\kappa\delta}}{\partial r^i_\alpha}R^i_{\sigma\eta} + R^i_{\eta\beta}R^i_{\tau\gamma}R^i_{\kappa\delta}\frac{\partial R^i_{\sigma\eta}}{\partial r^i_\alpha}\right)C^{i'}_{\eta\tau\kappa\sigma}V^i_{\beta\gamma}V^i_{\delta\eta} \tag{5.60}$$

where the factors 1/2, 1/3 and 1/6 are due to the self-energy terms – i.e. the contribution from the electrostatic plus induction interaction is reduced exactly by one-half due to net cancellation by the self-energy terms.

Different choices of local frames and principal vectors, as well as atomic force contributions, are detailed in the work of Lipparini *et al.* [81] The specific choices in this work which define the rotation matrix $R^i_{\Lambda\lambda}$ result in obvious sign changes compared to their work, so the atomic contributions are detailed below in compact form. We require the derivatives of the rotation matrix with respect to the atomic coordinates which can be written as

$$\frac{\partial R^i_{\Lambda\lambda}}{\partial r^{ia}_\alpha} = \begin{bmatrix} \frac{\partial e^{iX}_x}{\partial r^{ia}_\alpha} & \frac{\partial e^{iX}_y}{\partial r^{ia}_\alpha} & \frac{\partial e^{iX}_z}{\partial r^{ia}_\alpha} \\ \frac{\partial e^{iY}_x}{\partial r^{ia}_\alpha} & \frac{\partial e^{iY}_y}{\partial r^{ia}_\alpha} & \frac{\partial e^{iY}_z}{\partial r^{ia}_\alpha} \\ \frac{\partial e^{iZ}_x}{\partial r^{ia}_\alpha} & \frac{\partial e^{iZ}_y}{\partial r^{ia}_\alpha} & \frac{\partial e^{iZ}_z}{\partial r^{ia}_\alpha} \end{bmatrix} \tag{5.61}$$

The COM and principal vectors used to define the rotation are

$$\mathbf{r}^i = \sum_{a\in i}\mathbf{r}^{ia}\frac{M^a}{M^i}, \qquad \mathbf{B}^i = \mathbf{r}^i - \mathbf{r}^{i\mathrm{H}_1}, \qquad \mathbf{C}^i = \mathbf{r}^i - \mathbf{r}^{i\mathrm{H}_2} \tag{5.62}$$

Defining

$$\mathbf{D}^i = B^i\mathbf{C}^\mathbf{i} + C^i\mathbf{B}^\mathbf{i} \tag{5.63}$$

such that

$$\mathbf{e}^{iZ} = \frac{\mathbf{D}^i}{D^i} \tag{5.64}$$

where B^i, C^i and D^i are the euclidean norms of the vectors, the terms in the derivative of the rotation matrix, Eq. (5.61), are then

$$\frac{\partial e_\lambda^{iZ}}{\partial r_\alpha^{ia}} = \frac{\partial e_\lambda^{iZ}}{\partial D_\beta^i}\left(\frac{\partial D_\beta^i}{\partial r_\gamma^i}\frac{\partial r_\gamma^i}{\partial r_\alpha^{ia}} + \frac{\partial D_\beta^i}{\partial r_\alpha^{ia}}\right) \tag{5.65}$$

$$\frac{\partial e_\lambda^{iX}}{\partial r_\alpha^{ia}} = \frac{\partial e_\lambda^{iX}}{\partial B_\beta^i}\left(\frac{\partial B_\beta^i}{\partial r_\gamma^i}\frac{\partial r_\gamma^i}{\partial r_\alpha^{ia}} + \frac{\partial B_\beta^i}{\partial r_\alpha^{ia}}\right) + \frac{\partial e_\lambda^{iX}}{\partial e_\beta^{iZ}}\frac{\partial e_\beta^{iZ}}{\partial r_\alpha^{ia}} \tag{5.66}$$

$$\frac{\partial e_\lambda^{iY}}{\partial r_\alpha^{ia}} = \frac{\partial e_\lambda^{iY}}{\partial e_\beta^{iX}}\frac{\partial e_\beta^{iX}}{\partial r_\alpha^{ia}} + \frac{\partial e_\lambda^{iY}}{\partial e_\beta^{iZ}}\frac{\partial e_\beta^{iZ}}{\partial r_\alpha^{ia}} \tag{5.67}$$

where the leading terms are as follows

$$\frac{\partial e_\lambda^{iZ}}{\partial D_\beta^i} = \left(\frac{\mathbf{I}}{D^i} - \frac{\mathbf{D}^i \otimes \mathbf{D}^i}{(D^i)^3}\right)_{\lambda\beta} \tag{5.68}$$

$$\frac{\partial e_\lambda^{iX}}{\partial B_\beta^i} = \left(\frac{\mathbf{I} - \mathbf{e}^{iZ}\otimes\mathbf{e}^{iZ} - \mathbf{e}^{iX}\otimes\mathbf{e}^{iX}}{\mid \mathbf{B}^i - (\mathbf{B}^i\cdot\mathbf{e}^{iZ})\mathbf{B}^i \mid}\right)_{\lambda\beta} \tag{5.69}$$

$$\frac{\partial e_\lambda^{iX}}{\partial e_\beta^{iZ}} = \left(\frac{(\mathbf{B}^i\cdot\mathbf{e}^{iZ})\mathbf{e}^{iX}\otimes\mathbf{B}^i}{\mid \mathbf{B}^i - (\mathbf{B}^i\cdot\mathbf{e}^{iZ})\mathbf{B}^i \mid^2} - \frac{(\mathbf{B}^i\cdot\mathbf{e}^{iZ})\mathbf{I} + \mathbf{e}^{iZ}\otimes\mathbf{B}^i}{\mid \mathbf{B}^i - (\mathbf{B}^i\cdot\mathbf{e}^{iZ})\mathbf{B}^i \mid}\right)_{\lambda\beta} \tag{5.70}$$

$$\frac{\partial e_\lambda^{iY}}{\partial e_\beta^{iZ}} = \epsilon_{\lambda\sigma\tau}\delta_{\beta\sigma}e_\tau^{iX} \tag{5.71}$$

$$\frac{\partial e_\lambda^{iY}}{\partial e_\beta^{iX}} = \epsilon_{\lambda\sigma\tau}e_\sigma^{iZ}\delta_{\beta\tau} \tag{5.72}$$

where $\mathbf{I}$ is the 3×3 identity matrix and $\epsilon_{\alpha\beta\gamma}$ the Levi-Civita symbols. The latter terms are

$$\frac{\partial D_\beta^i}{\partial r_\gamma^i}\frac{\partial r_\gamma^i}{\partial r_\alpha^{ia}} = \left((B^i + C^i)\mathbf{I} + \frac{\mathbf{B}^i\otimes\mathbf{C}^i}{C^i} + \frac{\mathbf{C}^i\otimes\mathbf{B}^i}{B^i}\right)_{\beta\gamma}\delta_{\gamma\alpha}\frac{M^a}{M^i} \tag{5.73}$$

$$\frac{\partial D_\beta^i}{\partial r_\alpha^{i\mathrm{H}_1}} = -\left(B^i\mathbf{I} + \frac{\mathbf{B}^i\otimes\mathbf{C}^i}{C^i}\right)_{\beta\alpha} \tag{5.74}$$

$$\frac{\partial D_\beta^i}{\partial r_\alpha^{i\mathrm{H}_2}} = -\left(C^i\mathbf{I} + \frac{\mathbf{C}^i\otimes\mathbf{B}^i}{B^i}\right)_{\beta\alpha} \tag{5.75}$$

$$\frac{\partial B_\beta^i}{\partial r_\gamma^i}\frac{\partial r_\gamma^i}{\partial r^{ia}} = \delta_{\beta\gamma}\delta_{\gamma\alpha}\frac{M^a}{M^i} \tag{5.76}$$

$$\frac{\partial B_\beta^i}{\partial r_\alpha^{i\mathrm{H}_1}} = -\delta_{\beta\alpha} \tag{5.77}$$

For the DMS (see the main text) the first term on the right hand side is

$$\frac{\partial E_{\text{ele+ind}}}{\partial \mu^j_\beta(\{\mathbf{r}^{jb}\})} = \left(\frac{\partial E_{\text{ele+ind}}}{\partial \mu^j_\beta(\{\mathbf{r}^{jb}\})} + \frac{\partial E_{\text{ele+ind}}}{\partial V^k_{\gamma\delta\ldots\eta}} \frac{\partial V^k_{\gamma\delta\ldots\eta}}{\partial \mu^j_\beta(\{\mathbf{r}^{jb}\})} \right) = \frac{1}{2} V^j_\beta + \frac{1}{2} \sum_k^n \delta_{jk} V^k_\beta \tag{5.78}$$

$$\frac{\partial \mu^j_\beta(\{\mathbf{r}^{jb}\})}{\partial r^{ia}_\alpha} = \left(\frac{\partial q^{jb}}{\partial r^{ia}_\alpha} + \frac{\partial r^{jb}}{\partial r^{ia}_\alpha} \right) = \delta_{ji} \left(\sum_b^{n_i} \frac{\partial q^{jb}}{\partial r^{ia}_\alpha} r^{jb}_\beta + \delta_{ba} q^{jb} \delta_{\beta\alpha} \right) \tag{5.79}$$

$$\frac{\partial E_{\text{ele+ind}}}{\partial \mu^j_\beta(\{\mathbf{r}^{jb}\})} \frac{\partial \mu^j_\beta(\{\mathbf{r}^{jb}\})}{\partial r^{ia}_\alpha} = q^{ia} V^i_\alpha + \sum_b^{n_i} \frac{\partial q^{ib}}{\partial r^{ia}_\alpha} r^{ib}_\beta V^i_\beta \tag{5.80}$$

The derivatives of the DMS, $\frac{\partial q^{ib}}{\partial r^{ia}_\beta}$, with respect to the atomic positions are derived by Burnham and Xantheas [32] and are available in open source repositories.

For the force contribution due to the QMS we first rewrite the following expression

$$\theta^i_{\alpha\beta}(\mathbf{r}^{i\mathrm{O}}, \mathbf{r}^{i\mathrm{H}_1}, \mathbf{r}^{i\mathrm{H}_2}) = \sum_a^{\mathrm{H}'_1,\mathrm{H}'_2,\mathrm{L}_1,\mathrm{L}_2} \frac{3}{2} \left\{ q^{ia} \left((\mathbf{r}^{ia} - \mathbf{r}^i)_\alpha (\mathbf{r}^{ia} - \mathbf{r}^i)_\beta - \frac{\delta_{\alpha\beta}}{3} ||\mathbf{r}^{ia} - \mathbf{r}^i|| \right) \right\} \tag{5.81}$$

noting that the position of the L-sites in the global frame are

$$r^{i\mathrm{L}_l}_\alpha = R^{i\mathrm{H}_l}_{\eta\alpha} e^{iZ}_\eta f(\mathbf{r}^{\mathrm{H}_l}) + r^i_\alpha \tag{5.82}$$

we remove redundant terms and the expression for the QMS becomes

$$\begin{aligned} \theta^i_{\alpha\beta}(\mathbf{r}^{i\mathrm{O}}, \mathbf{r}^{i\mathrm{H}_1}, \mathbf{r}^{i\mathrm{H}_2}) = \frac{3}{2} \Bigg\{ & \sum_a^{\mathrm{H}'_1,\mathrm{H}'_2} q^{ia} \left((\mathbf{r}^{ia} - \mathbf{r}^i)_\alpha (\mathbf{r}^{ia} - \mathbf{r}^i)_\beta - \frac{\delta_{\alpha\beta}}{3} ||\mathbf{r}^{ia} - \mathbf{r}^i|| \right) \\ & + \sum_l^{1,2} q^{i\mathrm{L}_l} \left(dr^{i\mathrm{L}_l}_\alpha dr^{i\mathrm{L}_l}_\beta - \frac{\delta_{\alpha\beta}}{3} ||dr^{i\mathrm{L}_l}|| \right) \Bigg\} \end{aligned} \tag{5.83}$$

where

$$dr^{i\mathrm{L}_l}_\alpha = R^{i\mathrm{H}_l}_{\eta\alpha} e^{iZ}_\eta f(\mathbf{r}^{\mathrm{H}_l}) \tag{5.84}$$

Similar to the DMS we have

$$\frac{\partial E_{\text{ele+ind}}}{\partial\theta^j_{\beta\gamma}(\{\mathbf{r}^{jb}\})} = \left(\frac{\partial E_{\text{ele+ind}}}{\partial\theta^j_{\beta\gamma}(\{\mathbf{r}^{jb}\})} + \frac{\partial E_{\text{ele+ind}}}{\partial V^k_{\delta\kappa\ldots\eta}}\frac{\partial V^k_{\delta\kappa\ldots\eta}}{\partial\theta^j_{\beta\gamma}(\{\mathbf{r}^{jb}\})}\right) = \frac{1}{6}V^j_{\beta\gamma} + \frac{1}{6}\sum_k^n \delta_{jk}V^k_{\beta\gamma} \tag{5.85}$$

$$\begin{aligned}
\frac{\partial\theta^j_{\beta\gamma}(\{\mathbf{r}^{jb}\})}{\partial r^{ia}_\alpha} =& \frac{3}{2}\delta_{ji}\Bigg\{ \sum_b^{\mathrm{H}'_1,\mathrm{H}'_2} \frac{\partial q^{jb}}{\partial r^{ia}_\alpha}\left((\mathbf{r}^{jb}-\mathbf{r}^j)_\beta(\mathbf{r}^{jb}-\mathbf{r}^j)_\gamma - \frac{\delta_{\beta\gamma}}{3}||\mathbf{r}^{jb}-\mathbf{r}^j||\right) \\
&+ \sum_l^{1,2}\frac{\partial q^{j\mathrm{L}_l}}{\partial r^{ia}_\alpha}\left(dr^{j\mathrm{L}_l}_\beta dr^{j\mathrm{L}_l}_\gamma - \frac{\delta_{\beta\gamma}}{3}||dr^{j\mathrm{L}_l}||\right) \\
&+ \sum_b^{\mathrm{H}'_1,\mathrm{H}'_2} q^{jb}\Bigg(\delta_{\beta\alpha}\left(\delta_{ba}-\sum_c^{n_j}\delta_{ca}\frac{M^c}{M^j}\right)(\mathbf{r}^{jb}-\mathbf{r}^j)_\gamma \\
&\qquad + (\mathbf{r}^{jb}-\mathbf{r}^j)_\beta\delta_{\gamma\alpha}\left(\delta_{ba}-\sum_c^{n_j}\delta_{ca}\frac{M^c}{M^j}\right) \\
&\qquad + \delta_{\beta\gamma}\frac{2}{3}\left(\delta_{ba}-\sum_c^{n_j}\delta_{ca}\frac{M^c}{M^j}\right)\delta_{\alpha\delta}(\mathbf{r}^{jb}-\mathbf{r}^j)_\delta\Bigg) \\
&+ \sum_l^{1,2} q^{j\mathrm{L}_l}\Bigg(\delta_{\beta\alpha}\frac{\partial dr^{j\mathrm{L}_l}_\beta}{\partial r^{ia}_\alpha}dr^{j\mathrm{L}_l}_\gamma + dr^{j\mathrm{L}_l}_\beta\delta_{\gamma\alpha}\frac{\partial dr^{j\mathrm{L}_l}_\gamma}{\partial r^{ia}_\alpha} \\
&\qquad + \delta_{\beta\gamma}\frac{2}{3}\frac{\partial r^{j\mathrm{L}_l}_\delta}{\partial r^{ia}_\alpha}dr^{j\mathrm{L}_l}_\delta\Bigg)\Bigg\}
\end{aligned} \tag{5.86}$$

The terms involving the partial derivative of the charges for each site are readily available since through the definition of the QMS charges we have

$$q^{i\mathrm{H}'_l} = \mathrm{A}q^{i\mathrm{H}_l} + \mathrm{B}q^{i\mathrm{H}_l}_{\text{eq}} \tag{5.87}$$

$$q^{i\mathrm{L}_l} = \mathrm{C}q^{i\mathrm{H}_l} + \mathrm{D}q^{i\mathrm{H}_l}_{\text{eq}} \tag{5.88}$$

and hence the expression reduces to

$$\begin{aligned}\frac{\partial E_{\text{ele+ind}}}{\partial\theta^j_{\beta\gamma}(\{\mathbf{r}^{jb}\})}\frac{\partial\theta^j_{\beta\gamma}(\{\mathbf{r}^{jb}\})}{\partial r^{ia}_\alpha} =&\frac{1}{2}\sum_l^{1,2}\Bigg\{\frac{\partial q^{i\mathrm{H}_l}}{\partial r^{ia}_\alpha}\Bigg(\mathrm{A}\Bigg(dr^{i\mathrm{H}_l}_\beta dr^{i\mathrm{H}_1}_\gamma-\frac{\delta_{\beta\gamma}}{3}||dr^{i\mathrm{H}_l}||\Bigg)\\&+\mathrm{C}\Bigg(dr^{i\mathrm{L}_l}_\beta dr^{i\mathrm{L}_l}_\gamma-\frac{\delta_{\beta\gamma}}{3}||dr^{i\mathrm{L}_l}||\Bigg)\Bigg)\\&+q^{i\mathrm{H}'_l}\Bigg(2\left(\delta_{ba}-\sum_c^{n_i}\delta_{ca}\frac{M^c}{M^i}\right)dr^{i\mathrm{H}_1}_\gamma\delta_{\beta\alpha}\\&\quad-\frac{2}{3}\left(\delta_{ba}-\sum_c^{n_i}\delta_{ca}\frac{M^c}{M^i}\right)dr^{i\mathrm{H}_l}_\alpha\delta_{\beta\gamma}\Bigg)\\&+q^{i\mathrm{L}_l}\Bigg(2\frac{\partial dr^{i\mathrm{L}_l}_\beta}{\partial r^{ia}_\alpha}dr^{i\mathrm{L}_l}_\gamma-\frac{2}{3}\frac{\partial dr^{i\mathrm{L}}_\delta}{\partial r^{ia}_\alpha}dr^{i\mathrm{L}_l}_\delta\delta_{\beta\gamma}\Bigg)\Bigg\}V^i_{\beta\gamma}\end{aligned}\tag{5.89}$$

where

$$dr^{i\mathrm{H}_l}_\alpha=(\mathbf{r}^{i\mathrm{H}_l}-\mathbf{r}^i)_\alpha\tag{5.90}$$

The only unknowns are the derivatives of the position of the L-sites in the local frame reference. Applying the rotation operators

$$\mathbf{R}^{i\mathrm{L}_1}=\Big(\cos(\theta')\mathbf{I}-\sin(\theta')\left[\mathbf{e}^{iX}\right]_\times\Big)\tag{5.91}$$

$$\mathbf{R}^{i\mathrm{L}_2}=\Big(\cos(\theta')\mathbf{I}+\sin(\theta')\left[\mathbf{e}^{iX}\right]_\times\Big)\tag{5.92}$$

on $\mathbf{e}^{iZ}$ in the case of L_1 and L_2 the expression for the local frame vectors becomes

$$dr^{i\mathrm{L}_1}_\alpha=\left(\cos(\theta')e^{iZ}_\alpha-\sin(\theta')e^{iY}_\alpha\right)f(\mathbf{r}^{i\mathrm{H}_1})\tag{5.93}$$

$$dr^{i\mathrm{L}_2}_\alpha=\left(\cos(\theta')e^{iZ}_\alpha+\sin(\theta')e^{iY}_\alpha\right)f(\mathbf{r}^{i\mathrm{H}_2})\tag{5.94}$$

Derivatives of the expressions above are of the form

$$\begin{aligned}\frac{\partial dr^{i\mathrm{L}_l}_\beta}{\partial r^{ia}_\alpha}=&\frac{\partial dr^{i\mathrm{L}_l}_\beta}{\partial e^{iZ}_\gamma}\frac{\partial e^{iZ}_\gamma}{\partial r^{ia}_\alpha}+\frac{\partial dr^{i\mathrm{L}_l}_\beta}{\partial e^{iY}_\gamma}\frac{\partial e^{iY}_\gamma}{\partial r^{ia}_\alpha}+\frac{\partial dr^{i\mathrm{L}_l}_\beta}{\partial f(\mathbf{r}^{i\mathrm{H}_l})}\frac{\partial f(\mathbf{r}^{i\mathrm{H}_l})}{\partial r^{ia}_\alpha}\\&+\left(\frac{\partial dr^{i\mathrm{L}_l}_\beta}{\partial\cos(\theta')}\frac{\partial\cos(\theta')}{\partial\theta'}+\frac{\partial dr^{i\mathrm{L}_l}_\beta}{\partial\sin(\theta')}\frac{\partial\sin(\theta')}{\partial\theta'}\right)\frac{\partial\theta'}{\partial r^{ia}_\alpha}\end{aligned}\tag{5.95}$$

For $l = 1$ as an example the leading terms are

$$\frac{\partial dr_\beta^{i\mathrm{L}_l}}{\partial e_\gamma^{iZ}} = \cos(\theta')f(\mathbf{r}^{i\mathrm{H}_1})\delta_{\beta\gamma} \tag{5.96}$$

$$\frac{\partial dr_\beta^{i\mathrm{L}_l}}{\partial e_\gamma^{iY}} = -\sin(\theta')f(\mathbf{r}^{i\mathrm{H}_1})\delta_{\beta\gamma} \tag{5.97}$$

$$\frac{\partial dr_\beta^{i\mathrm{L}_1}}{\partial f(\mathbf{r}^{\mathrm{H}_1})} = \left(\cos(\theta')e_\beta^{iZ} - \sin(\theta')e_\beta^{iY}\right) \tag{5.98}$$

$$\frac{\partial dr_\beta^{i\mathrm{L}_l}}{\partial \cos(\theta')}\frac{\partial \cos(\theta')}{\partial \theta'} = e_\beta^{iZ}\left(\sin(\theta')\right)f(\mathbf{r}^{i\mathrm{H}_1}) \tag{5.99}$$

$$\frac{\partial dr_\beta^{i\mathrm{L}_l}}{\partial \sin(\theta')}\frac{\partial \sin(\theta')}{\partial \theta'} = e_\beta^{iY}\left(-\cos(\theta')\right)f(\mathbf{r}^{i\mathrm{H}_1}) \tag{5.100}$$

and the two remaining latter terms are

$$\frac{\partial f(\mathbf{r}^{\mathrm{H}_1})}{\partial r_\alpha^{i\mathrm{H}_1}} = -b\frac{(\mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_1})_\alpha}{\mid \mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_1} \mid} - 2cf(\mathbf{r}^{\mathrm{H}_1})\frac{(\mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_1})_\alpha}{\mid \mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_1} \mid} \tag{5.101}$$

$$\frac{\partial \theta'}{\partial r_\alpha^{i\mathrm{H}_1}} = \frac{1}{\sqrt{1-x^2}}\left(-\frac{(\mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_2})_\alpha}{\mid \mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_1} \mid\mid \mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_2} \mid} + x\frac{(\mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_1})_\alpha}{\mid \mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_1} \mid^2}\right) \tag{5.102}$$

where

$$x = \frac{(\mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_1}) \cdot (\mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_2})}{\mid \mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_1} \mid\mid \mathbf{r}^{\mathrm{O}} - \mathbf{r}^{\mathrm{H}_2} \mid}$$

and

$$\theta' = \arccos(x) \tag{5.103}$$

Finally, the third term on the right hand side in the force expression of equation 5.43 is the partial derivatives of the external field at each site with respect to the atomic positions. It is of the general form

$$\frac{\partial E_{\mathrm{ele+ind}}}{\partial V^j_{\beta\gamma\delta\epsilon\ldots\eta}}\frac{\partial V^j_{\beta\gamma\delta\epsilon\ldots\eta}}{\partial r^k_\delta}\frac{\partial r^k_\delta}{\partial r^{ia}_\alpha} \rightarrow O^i_{\beta\gamma\delta\epsilon\ldots\eta}V^i_{\alpha\beta\gamma\delta\epsilon\ldots\eta}\frac{M^a}{M^i} \tag{5.104}$$

where $O^i_{\beta\gamma\delta\epsilon\ldots\eta}$ is the $(n-1)$th ranked moment tensor (static plus induced) and $V^i_{\alpha\beta\gamma\delta\epsilon\ldots\eta}$ the corresponding nth ranked external potential gradient. Considering the electrostatic plus induction interaction expression (see the main text), this leads to

$$\frac{\partial E_{\mathrm{ele+ind}}}{\partial V^j_{\beta\gamma\delta\epsilon\ldots\eta}}\frac{\partial V^j_{\beta\gamma\delta\epsilon\ldots\eta}}{\partial r^k_\delta}\frac{\partial r^k_\delta}{\partial r^{ia}_\alpha} = \Bigg((\mu^i_\beta(\{\mathbf{r}^{ia}\}) + \Delta\mu^i_\beta)V^i_{\alpha\beta} + \frac{1}{3}(\theta^i_{\beta\gamma}(\{\mathbf{r}^{ia}\}) + \Delta\theta^i_{\beta\gamma})V^i_{\alpha\beta\gamma} + \frac{1}{15}\Omega^i_{\beta\gamma\delta}V^i_{\alpha\beta\gamma\delta} + \frac{1}{105}\Phi^i_{\beta\gamma\delta\epsilon}V^i_{\alpha\beta\gamma\delta\epsilon}\Bigg)\frac{M^a}{M^i} \tag{5.105}$$

The energy-force consistency of the formulation is checked against numerical forces, and at different convergence criteria of the induced moments, see figure 5.11.

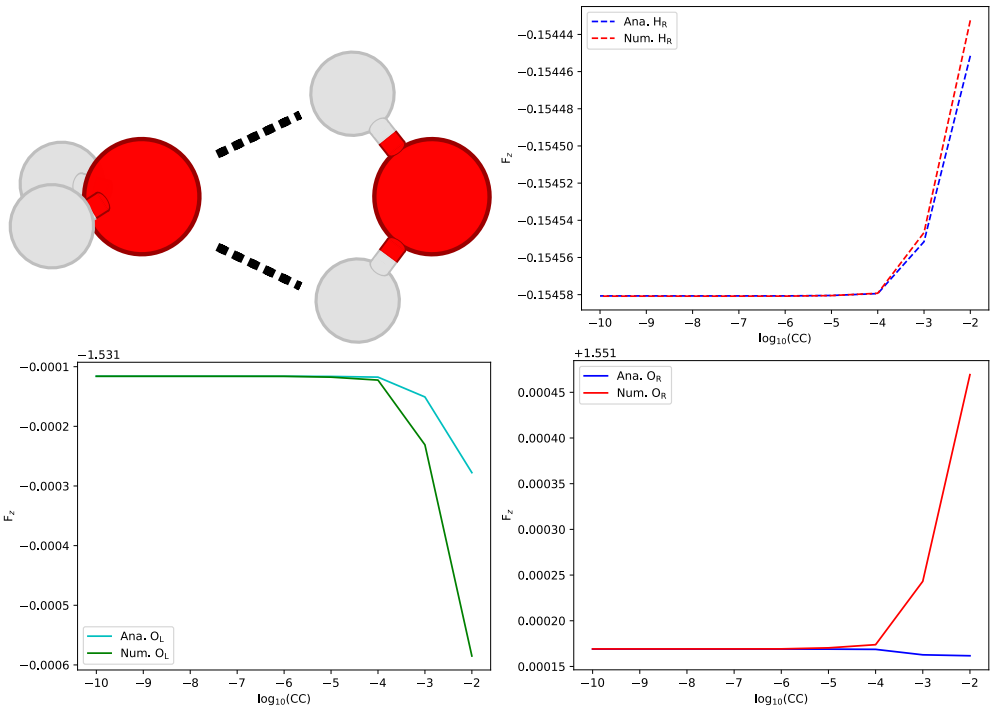


Figure 5.11: C2v isomer configuration for the water dimer (top left), used for the numerical versus analytical forces at different convergence criteria ranges for one of the right hydrogens, H_R (top right), as well as the left and right oxygen, O_L (bottom left) and O_R (bottom right). The convergence of the force components is shown versus the magnitude of the convergence criteria, CC, which is defined as $\sum_i \mid \Delta\mu^i_{n+1} - \Delta\mu^i_n \mid \leq$ CC, and similarly for the induced quadrupole moment. Good energy-force consistency is reached reliably at a criteria of 1e-6 D for the dipole (or DÅ for the quadrupole).

5.A.2 Deviation of M-site Models

The general expression for the %RMS difference is given by

$$\%\mathrm{RMS} = \sqrt{\frac{\sum_i^{x,y,z}(\theta_{ii} - \theta_{ii}^{\text{ab initio}})^2}{\sum_i^{x,y,z}(\theta_{ii}^{\text{ab initio}})^2}} \times 100\% \tag{5.106}$$

and reveals if there is a large deviation from the individual components of the quadrupole moment along the trace, $\{\theta_{xx}, \theta_{yy}, \theta_{zz}\}$, relative to the trace norm. While a quadrupole model may capture the principal quadrupole component θ_T, such as the M-site based $\gamma-(\mathrm{DMS}-q_{\mathrm{eq}}^{\mathrm{H}})^*$ model, to a reasonable degree this can simply be due to a net cancellation of errors in the individual components. The QMS model captures both the θ_T and the Δ component on average at around 1.6% evaluated over the whole range. There is no systematic correlation in the deviation of the quadrupole components of the QMS models and the RMSD of the monomer geometry from the equilibrium geometry. The RMSD of the geometry is evaluated with the Kabsch algorithm [110].

To solve for the γ factor and hence the position of the M-site such that Δ vanishes (see main text), one considers the water monomer in the equilibrium configuration with the oxygen placed at the origin. In this configuration and frame of reference the quadrupole moment tensor only has components along the trace $\{\theta_{xx}, \theta_{yy}, \theta_{zz}\}$, where $\theta_{zz} = 2\Delta$. This component, in terms of the charges and positions of the hydrogens and the M-site, is given by [117]

$$\theta_{zz} = q^{\mathrm{H}}(-(r_x^{\mathrm{H}})^2 + 2(r_z^{\mathrm{H}})^2 - 2(r_z^{\mathrm{M}})^2) \tag{5.107}$$

which is trivial to solve in order for Δ to vanish, by using the generalization of this expression in the global coordinate frame [113–115]

$$\mathbf{r}_{\mathrm{M}} = (1-\gamma)\mathbf{r}_{\mathrm{O}} + \frac{\gamma}{2}(\mathbf{r}_{\mathrm{H}_1} + \mathbf{r}_{\mathrm{H}_2}) \tag{5.108}$$

5.A.3 Energy-Volume Curves for Ice Ih

Fig. 5.13 shows energy-volume curves without ($E_{\mathrm{tot}}(V)$, blue) and with zero-point-energy correction according to the quasi-harmonic approximation ($E_{\mathrm{tot+ZPE}}(V)$, red).

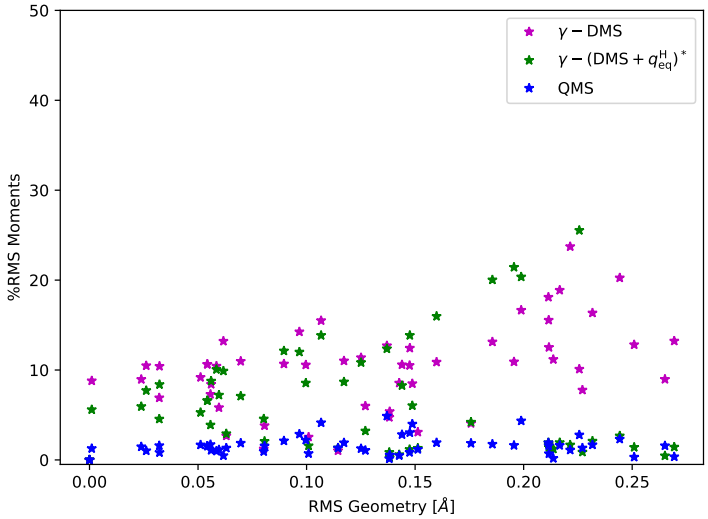


Figure 5.12: Root-mean-square percentage difference between the three diagonal components of the quadrupole moment tensors of two M-site models and the QMS model versus the ab initio (ICE-CI) quadrupole moment tensor respectively. The ab initio and the average RMS percentage QMS difference is around 1.6%. The largest deviation corresponds to the numerically lowest θ_T. There is no correlation between the magnitude of the relative geometrical change relative to the ground state monomer geometry and the deviation of the quadrupole moment.

The indicated 6 data points are the direct results from the calculations with the optimized (final) parameters for SCME/f described in the main article around the equilibrium volume. The lines are the results of least-square fits to the Rose-Vinet equation of state

$$E(\eta) = E_0 \; + \; \frac{2\,B_0\,V_0}{(B'-1)^2} \cdot \left[2 - (5 + 3B'\,(\eta - 1) - 3\eta) \cdot \exp\left(-\frac{3}{2}(B'-1)\,(\eta-1)\right)\right] \tag{5.109}$$

as implemented in the PHONOPY package [108], where $\eta = \frac{V}{V_0}$ and $E_0 = -E_{\mathrm{lat}}$. The corresponding fit parameters compiled in Table 5.9. We have verified that adding more points does not yield significant changes for the fit parameters, in particular those that are directly compared against experimental data (E_0, V_0 and B_0).

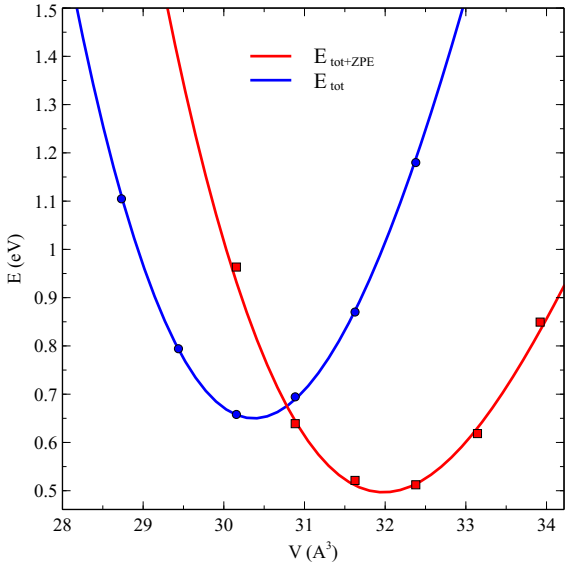


Figure 5.13: Energy-volume curves without ($E_{\mathrm{tot}}(V)$, blue) and with zero-point-energy correction ($E_{\mathrm{tot+ZPE}}(V)$, red) obtained with the optimized (final) parameters for SCME/f. Energies and volumes are per water molecule.

	E_0 (eV)	V_0 (Å^3)	B_0 (eV/Å^3)	B_0 (GPa)	B'
E_{tot}	0.645	30.38	0.094	15.0	5.39
$E_{\mathrm{tot+ZPE}}$	0.489	31.98	0.076	12.2	5.68

Table 5.9: Fit parameters for the Rose-Vinet equation (Eq. (5.109)) for the fits to $E_{\mathrm{tot}}(V)$ and $E_{\mathrm{tot+ZPE}}(V)$ shown in Fig. 5.13. All quantities are given per water molecule.

5.8 References

1. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **79**, 926–935 (1983).
2. W. L. Jorgensen, *J. Am. Chem. Soc.* **103**, 335–340 (1981).
3. H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, *et al.*, *J. Chem. Phys.* **120**, 9665 (2004).
4. J. Zielkiewicz, *J. Chem. Phys.* **123**, 104501 (2005).
5. K. Park, W. Lin, F. Paesani, *J. Phys. Chem. B* **116**, 343–352 (2012).
6. S. Habershon, T. E. Markland, D. E. Manolopoulos, *J. Chem. Phys.* **131**, 024501 (2009).
7. E. R. Batista, S. S. Xantheas, H. Jónsson, *J. Chem. Phys.* **109**, 4546 (1998).
8. E. R. Batista, S. S. Xantheas, H. Jónsson, *J. Chem. Phys.* **112**, 3285 (2000).
9. E. Batista, *Development of a New Water-Water Interaction Potential and Applications to Molecular Processes in Ice* (University of Washington, 1999).
10. K. T. Wikfeldt, E. R. Batista, F. D. Vila, H. Jónsson, *Phys. Chem. Chem. Phys.* **15**, 16542 (2013).
11. I. V. Smirnov, A. V. Golovin, S. D. Chatziefthimiou, A. V. Stepanova, Y. Peng, *et al.*, *Sci. Adv.* **2**, e1501695 (2016).
12. T. R. M. Barends, L. Foucar, A. Ardevol, K. Nass, A. Aquila, *et al.*, *Science* **350**, 445–450 (2015).
13. H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed. (English)* **48**, 1198–1229 (2009).
14. K. Senthilkumar, J. I. Mujika, K. E. Ranaghan, F. R. Manby, A. J. Mulholland, *et al.*, *J. Royal Soc. Interface* **5**, 207–216 (2008).
15. A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, *et al.*, *Chem. Rev.* **106**, 3210–3235 (2006).
16. F. Zheng, L. Xue, S. Hou, J. Liu, M. Zhan, *et al.*, *Nat. Commun.* **5**, 3457 (2014).
17. J. Knorr, P. Sokkar, S. Schott, P. Costa, W. Thiel, *et al.*, *Nat. Commun.* **7**, 12968 (2016).
18. V.-T. Pham, T. J. Penfold, R. M. van der Veen, F. Lima, A. E. Nahhas, *et al.*, *J. Am. Chem. Soc.* **133**, 12740–12748 (2011).
19. A. O. Dohn, E. O. Jónsson, K. S. Kjær, T. B. van Driel, M. M. Nielsen, *et al.*, *J. Phys. Chem. Lett.* **5**, 2414–2418 (2014).
20. A. O. Dohn, K. S. Kjær, T. B. Harlang, S. E. Canton, M. M. Nielsen, *et al.*, *Inorg. Chem.* **55**, 10637–10644 (2016).
21. G. Levi, M. Pápai, N. E. Henriksen, A. O. Dohn, K. B. Møller, *J. Chem. Phys. C* **122**, 7100–7119 (2018).
22. A. O. Dohn, D. Selli, G. Fazio, L. Ferraro, J. Mortensen, *et al.*, *Molecules* **23**, 2958 (2018).
23. Y.-J. Zhang, A. Khorshidi, G. Kastlunger, A. A. Peterson, *J. Chem. Phys.* **148**, 241740 (2018).

24. H. Lin, D. G. Truhlar, *Theor. Chem. Acc.* **117**, 185 (2006).
25. S. Pezeshki, H. Lin, *Mol. Sim.* **41**, 168–189 (2015).
26. K. Sneskov, T. Schwabe, O. Christiansen, J. Kongsted, *Phys. Chem. Chem. Phys.* **13**, 18551–18560 (2011).
27. U. N. Morzan, D. J. A. de Armiño, N. O. Foglia, F. Ramírez, M. C. G. Lebrero, *et al.*, *Chem. Rev.* **118**, 4071–4113 (2018).
28. A. Warshel, M. Levitt, *J. Mol. Biol.* **103**, 227–249 (1976).
29. G. A. Cisneros, K. T. Wikfeldt, L. Ojamäe, J. Lu, Y. Xu, *et al.*, *Chem. Rev.* **116**, 7501–7528 (2016).
30. H. Yu, W. F. Van Gunsteren, *Comput. Phys. Commun.* **172**, 69–85 (2005).
31. P. E. Lopes, B. Roux, A. D. MacKerell, *Theoretical Chemistry Accounts* **124**, 11–28 (2009).
32. C. J. Burnham, S. S. Xantheas, *J. Chem. Phys.* **116**, 1479–1492 (2002).
33. G. S. Fanourgakis, S. S. Xantheas, *J. Phys. Chem. A* **110**, 4100–4106 (2006).
34. G. S. Fanourgakis, S. S. Xantheas, *J. Chem. Phys* **128**, 074506 (2008).
35. C. Burnham, D. Anick, P. Mankoo, G. Reiter, *J. Chem. Phys* **128**, 154519 (2008).
36. G. R. Medders, V. Babin, F. Paesani, *J. Chem. Theory Comput.* **9**, 1103–1114 (2013).
37. V. Babin, G. R. Medders, F. Paesani, *J. Phys. Chem. Lett.* **3**, 3765–3769 (2012).
38. V. Babin, C. Leforestier, F. Paesani, *J. Chem. Theory Comput.* **9**, 5395–5403 (2013).
39. V. Babin, G. R. Medders, F. Paesani, *J. Chem. Theory Comput.* **10**, 1599–1607 (2014).
40. G. R. Medders, V. Babin, F. Paesani, *J. Chem. Theory Comput.* **10**, 2906–2910 (2014).
41. M. A. Thompson, G. K. Schenter, *J. Phys. Chem.* **99**, 6374–6386 (1995).
42. M. A. Thompson, *J. Phys. Chem.* **100**, 14492–14507 (1996).
43. R. A. Bryce, R. Buesnel, I. H. Hillier, N. A. Burton, *Chem. Phys. Lett.* **279**, 367–371 (1997).
44. F. Lipparini, V. Barone, *J. Chem. Theory Comput.* **7**, 3711–3724 (2011).
45. E. Boulanger, W. Thiel, *J. Chem. Theory Comput.* **8**, 4527–4538 (2012).
46. Z. Lu, Y. Zhang, *J. Chem. Theory Comput.* **4**, 1237–1248 (2008).
47. N. M. Thellamurege, D. Si, F. Cui, H. Zhu, R. Lai, *et al.*, *J. Comput. Chem.* **34**, 2816–2833 (2013).
48. E. G. Kratz, A. R. Walker, L. Lagardère, F. Lipparini, J.-P. Piquemal, *et al.*, *J. Comput. Chem.* **37**, 1019–1029 (2016).
49. J. Dziedzic, Y. Mao, Y. Shao, J. Ponder, T. Head-Gordon, *et al.*, *J. Chem. Phys.* **145**, 124106 (2016).
50. A. S. P. Gomes, C. R. Jacob, *Ann. Rep. Prog. Chem., Sect. C: Phys. Chem.* **108**, 222–277 (2012).

51. P. Söderhjelm, C. Husberg, A. Strambi, M. Olivucci, U. Ryde, *J. Chem. Theory Comput.* **5**, 649–658 (2009).
52. K. Sneskov, T. Schwabe, J. Kongsted, O. Christiansen, *J. Chem. Phys* **134**, 03B608 (2011).
53. S. Caprasecca, S. Jurinovich, L. Viani, C. Curutchet, B. Mennucci, *J. Chem. Theory Comput.* **10**, 1588–1598 (2014).
54. J. Kongsted, A. Osted, K. V. Mikkelsen, O. Christiansen, *Mol. Phys.* **100**, 1813–1828 (2002).
55. Q. Zeng, W. Liang, *J. Chem. Phys.* **143**, 134104 (2015).
56. D. Loco, É. Polack, S. Caprasecca, L. Lagardère, F. Lipparini, *et al.*, *J. Chem. Theor. Comput.* **12**, 3654–3661 (2016).
57. D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, B. Mennucci, *et al.*, *J. Chem. Theory Comput.* **13**, 4025–4033 (2017).
58. L. Jensen, P. T. van Duijnen, J. G. Snijders, *J. Chem. Phys.* **119**, 3800–3809 (2003).
59. A. H. Steindal, K. Ruud, L. Frediani, K. Aidas, J. Kongsted, *J. Phys. Chem. B* **115**, 3027–3037 (2011).
60. C. B. Nielsen, O. Christiansen, K. V. Mikkelsen, J. Kongsted, *J. Chem. Phys.* **126**, 154112 (2007).
61. J. M. Olsen, K. Aidas, J. Kongsted, *J. Chem. Theory Comput.* **6**, 3721–3734 (2010).
62. F. Lipparini, C. Cappelli, V. Barone, *J. Chem. Theory Comput.* **8**, 4153–4165 (2012).
63. C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes, *et al.*, *J. Chem. Theory Comput.* **5**, 1838–1848 (2009).
64. N. H. List, J. M. H. Olsen, J. Kongsted, *Phys. Chem. Chem. Phys.* **18**, 20234–20250 (2016).
65. M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, *et al.*, *J. Chem. Phys.* **138**, 244103 (2013).
66. C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes, *et al.*, *J. Chem. Theory Comput.* **5**, 1838–1848 (2009).
67. K. Visscher, W. Swope, D. Geerke, *Molecules* **23**, 3131 (2018).
68. D. Hršak, J. M. H. Olsen, J. Kongsted, *J. Chem. Theory Comput.*, acs.jctc.7b01153 (2018).
69. M. F. S. J. Menger, S. Caprasecca, B. Mennucci, *J. Chem. Theory Comput.* **13**, 3778–3786 (2017).
70. Y. Mao, Y. Shao, J. Dziedzic, C.-K. Skylaris, T. Head-Gordon, *et al.*, *J. Chem. Theory Comput.* **13**, 1963–1979 (2017).
71. J. Dziedzic, T. Head-Gordon, M. Head-Gordon, C.-K. Skylaris, *J. Chem. Phys.* **150**, 074103 (2019).

72. E. O. Jónsson, A. O. Dohn, H. Jónsson, *J. Chem. Theory Comput.* **15**, 6562–6577 (2019).
73. A. O. Dohn, E. O. Jónsson, H. Jónsson, *J. Chem. Theory Comput.* **15**, 6578–6587 (2019).
74. H. Partridge, D. W. Schwenke, *J. Chem. Phys.* **106**, 4618–4639 (1997).
75. S. Habershon, T. E. Markland, D. E. Manolopoulos, *J. Chem. Phys.* **131**, 024501 (2009).
76. H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, *et al.*, *J. Chem. Phys.* **120**, 9665–9678 (2004).
77. J. Abascal, E Sanz, R García Fernández, C Vega, *J. Chem. Phys.* **122**, 234511 (2005).
78. J. L. Abascal, C. Vega, *J. Chem. Phys* **123**, 234505 (2005).
79. D. M. Bates, G. S. Tschumper, *J. Phys. Chem. A* **113**, 3555–3559 (2009).
80. B. Temelso, K. A. Archer, G. C. Shields, *J. Phys. Chem. A* **115**, 12034–12046 (2011).
81. F. Lipparini, L. Lagardère, B. Stamm, E. Cancès, M. Schnieders, *et al.*, *J. Chem. Theory Comput.* **10**, 1638–1651 (2014).
82. A. Stone, *The Theory of Intermolecular Forces* (OUP Oxford, 2013).
83. B. Thole, *Chem. Phys.* **59**, 341–350 (1981).
84. M. Masia, M. Probst, R. Rey, *J. Chem. Phys.* **123**, 164505 (2005).
85. M. Masia, M. Probst, R. Rey, *Chem. Phys. Lett.* **420**, 267–270 (2006).
86. C. J. Burnham, J. Li, S. S. Xantheas, M. Leslie, *J. Chem. Phys.* **110**, 4566–4581 (1999).
87. A. J. Stone, *J. Phys. Chem. A* **115**, 7017–7027 (2011).
88. J. Sala, E. Guàrdia, M. Masia, *J. Chem. Phys.* **133**, 234101 (2010).
89. K. T. Wikfeldt, E. R. Batista, F. D. Vila, H. Jónsson, *Phys. Chem. Chem. Phys.* **15**, 16542 (2013).
90. P. E. Wormer, H. Hettema, *J. Chem. Phys* **97**, 5592–5606 (1992).
91. K. Tang, J. P. Toennies, *J. Chem. Phys* **80**, 3726–3741 (1984).
92. O. Rodrigues', *J. Math. Pures Appl.* **5**, 380–440 (1840).
93. F. Neese, *WIREs Computational Molecular Science* **2**, 73–78 (2012).
94. F. Neese, *WIREs Computational Molecular Science* **8**, e1327 (2018).
95. B. Huron, J. Malrieu, P Rancurel, *J. Chem. Phys.* **58**, 5745–5759 (1973).
96. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, *et al.*, *Nat. Meth.* **17**, 261–272 (2020).
97. *Unified SCME*, 2020.
98. S. R. Bahn, K. W. Jacobsen, *Comput. Sci. Eng.* **4**, 55 (2002).
99. A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, *et al.*, *J. Phys. Condens. Matter* **29**, 273002 (2017).
100. N. Bjerrum, *Science* **115**, 385–390 (1952).

101. K. Röttger, A. Endriss, J. Ihringer, S. Doyle, W. F. Kuhs, *Acta Crys. Sec. B* **50**, 644–648 (1994).
102. E. Whalley, *Trans. Faraday Soc.* **53**, 1578 (1957).
103. P. V. Hobbs, *Ice physics* (Oxford university press, 2010).
104. E. Whalley, *"Trans. Faraday Soc."* **54**, 1613 (1958).
105. E. Whalley, *J. Chem. Phys.* **81**, 4087–4092 (1984).
106. S. Rasti, J. Meyer, *J. Chem. Phys.* **150**, 234504 (2019).
107. K. Parlinski, Z. Q. Li, Y. Kawazoe, *Phys. Rev. Lett.* **78**, 4063–4066 (1997).
108. A. Togo, I. Tanaka, *Scripta Mater.* **108**, 1–5 (2015).
109. P. Vinet, J. R. Smith, J. Ferrante, J. H. Rose, *Phys. Rev. B* **35**, 1945–1953 (1987).
110. W. Kabsch, *Acta Cryst. Sec. A* **32**, 922–923 (1976).
111. J. C. Howard, J. L. Gray, A. J. Hardwick, L. T. Nguyen, G. S. Tschumper, *J. Chem. Theory Comput.* **10**, 5426–5435 (2014).
112. J. C. Howard, G. S. Tschumper, *J. Chem. Theory Comput.* **11**, 2126–2136 (2015).
113. J. Reimers, R. Watts, M. Klein, *Chem. Phys.* **64**, 95–114 (1982).
114. J. Reimers, R. Watts, *Chem. Phys.* **85**, 83–112 (1984).
115. M. A. Suhm, R. O. Watts, *Mol. Phys.* **73**, 463–469 (1991).
116. O. Loboda, F. Ingrosso, M. F. Ruiz-López, H. Reis, C. Millot, *J. Comp. Chem.* **37**, 2125–2132 (2016).
117. J. L. Abascal, C. Vega, *J. Phys. Chem. C* **111**, 15811–15822 (2007).