



**HAL**  
open science

## A molecular mechanics implementation of the cyclic cluster model

Juan Diego Samaniego-Rojas, Robin Gaumard, José Alexandre, Tzonka Mineva, Gerald Geudtner, Andreas Köster

► **To cite this version:**

Juan Diego Samaniego-Rojas, Robin Gaumard, José Alexandre, Tzonka Mineva, Gerald Geudtner, et al.. A molecular mechanics implementation of the cyclic cluster model. *Zeitschrift fur Naturforschung B*, 2024, 79 (4), pp.201-213. 10.1515/znb-2023-0090 . hal-04567111

**HAL Id: hal-04567111**

**<https://hal.umontpellier.fr/hal-04567111v1>**

Submitted on 18 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Juan Diego Samaniego-Rojas\*, Robin Gaumard, José Alejandro, Tzonka Mineva, Gerald Geudtner and Andreas M. Köster\*

# A molecular mechanics implementation of the cyclic cluster model

<https://doi.org/10.1515/znb-2023-0090>

Received October 25, 2023; accepted November 20, 2023

**Abstract:** The implementation of the cyclic cluster model (CCM) for molecular mechanics is presented in the framework of the computational chemistry program `DEMON2K`. Because the CCM is particularly well-suited for the description of periodic systems with defects, it can be used for periodic QM/MM approaches where the non-periodic QM part is treated as a defect in a periodic MM surrounding. To this end, we present here the explicit formulae for the evaluation of the Ewald sum and its first- and second-order derivatives as implemented in `DEMON2K`. The outlined implementation was tested in molecular dynamics (MD) simulations and periodic structure optimization calculations. MD simulations of an argon system were carried out using the Nosé-Hoover chain (NHC) thermostat and the Martyna-Tobias-Klein (MTK) barostat to control the temperature and pressure of the system, respectively. For the validation of CCM structure optimization a set of molecular crystals were optimized using the Ewald method for the evaluation of the electrostatic interactions. Two optimization procedures for the determination of the atomic positions and CCM cell parameters were tested. Our results show that the simultaneous optimization of the atomic positions and cell parameters is most efficient.

**Keywords:** cyclic cluster model; molecular mechanics; optimization

**Dedicated to** Professor Thomas Bredow of the University of Bonn on the occasion of his 60th birthday.

**\*Corresponding authors: Juan Diego Samaniego-Rojas and Andreas M. Köster**, Departamento de Química, CINVESTAV, Av. Instituto Politécnico Nacional 2508, A.P. 14-740, Ciudad de México, C.P. 07360, Mexico, E-mail: [jdiego@cinvestav.mx](mailto:jdiego@cinvestav.mx) (J.D. Samaniego-Rojas), [akoster@cinvestav.mx](mailto:akoster@cinvestav.mx) (A.M. Köster)

**Robin Gaumard and Tzonka Mineva**, ICGM, Université Montpellier, CNRS, ENSCM, Montpellier, France, E-mail: [robin.gaumard@enscm.fr](mailto:robin.gaumard@enscm.fr) (R. Gaumard), [tzonka.mineva@enscm.fr](mailto:tzonka.mineva@enscm.fr) (T. Mineva)

**José Alejandro**, Departamento de Química, UAM-Iztapalapa, San Rafael Atlixco 186, A.P. 55-534, Ciudad de México, C.P. 09340, Mexico, E-mail: [jra@xanum.uam.mx](mailto:jra@xanum.uam.mx)

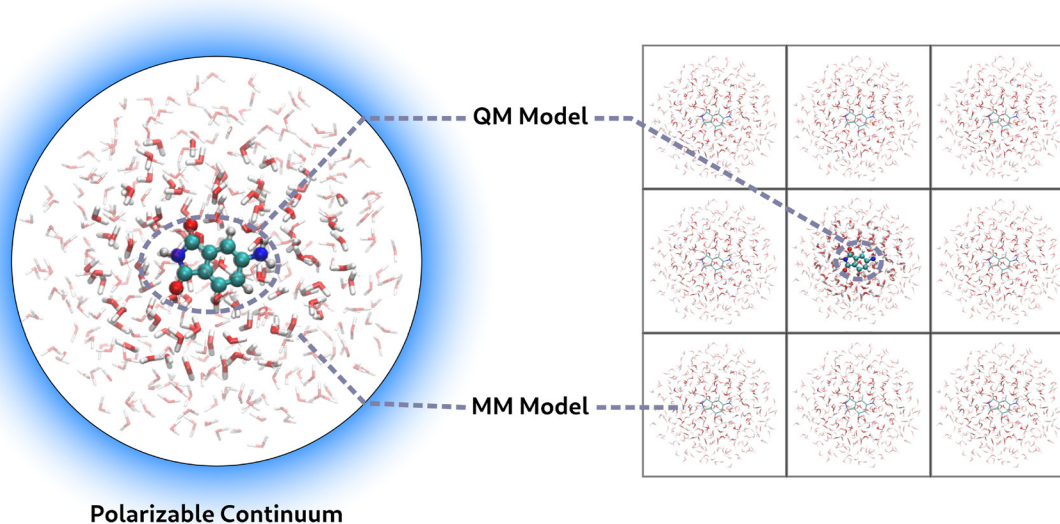
**Gerald Geudtner**, Departamento de Química, CINVESTAV, Av. Instituto Politécnico Nacional 2508, A.P. 14-740, Ciudad de México, C.P. 07360, Mexico, E-mail: [geudtner@cinvestav.mx](mailto:geudtner@cinvestav.mx)

## 1 Introduction

With the rise of hybrid quantum mechanical/molecular mechanical (QM/MM) schemes [1, 2] new atomistic approaches for the simulation of condensed matter systems have become available. In comparison to pure MM simulations, QM/MM calculations permit the incorporation of chemical bond breaking and forming in the QM region. Therefore, they are very well suited for the study of chemical reactions in solutions, on surfaces or in the bulk. For all these applications it is beneficial to have QM regions as large as possible. Thus, the QM method for such QM/MM applications must not only be accurate and reliable but also computationally efficient. A possible choice in this respect is auxiliary density functional theory (ADFT) [3, 4] as implemented in `DEMON2K` [5]. This is one of the main motivations behind the QM/MM implementation efforts in `DEMON2K` over the last decade [6, 7]. For a comprehensive overview we refer the interested reader to [8].

Although polarizable force fields are implemented in some versions of `DEMON2K` [7], we restrict ourselves for the here described development to non-polarizable force fields. With this setup for the QM/MM methodology, we now need appropriate models for the simulations. The droplet model, Figure 1 (left), consisting of a QM and an MM region embedded into a polarizable continuum has gained much popularity in `DEMON2K` QM/MM calculations. It is well suited for the simulation of solutions where QM molecules, free or solvated, are embedded in MM solvents [7, 9]. However, it is not suitable for the description of solids or for NPT molecular dynamics (MD) simulations.

To overcome this limitation we are developing at Cinvestav the QM/MM cyclic cluster model (CCM) as depicted in Figure 1 (right). In this model, the QM region, represented by the ball-and-stick molecules in Figure 1, is a QM *defect* in a periodic MM-CCM system. The CCM is an appropriate model for systems where the periodicity has been lifted by local defects because it avoids direct interactions between them [10]. The computational structure of the CCM we use in this work for MM systems is closely related to the quasimolecular large unit cell model of Dobrotvorskii and Evarestov [11]. A detailed comparison of the two models is given in the literature [12, 13]. Although the CCM has been implemented at the semiempirical [14–17], *ab initio* [18], and density-functional [19, 20] levels of theory, we are not aware of any



**Figure 1:** Droplet (left) and CCM (right) models for QM/MM solution systems. A ball-and-stick representation is used for the QM molecules and a stick-only representation for MM molecules.

implementation in the MM framework. Therefore, we present here such an implementation in `DEMON2K` as the first step towards a QM/MM-CCM methodology.

This work is organized as follows. In Section 2, we present the theory and implementation details of the CCM scheme for the calculation of the energy and energy derivatives within the MM branch of the `DEMON2K` program. In the next section, we describe our Ewald sum implementation for the efficient evaluation of long-range Coulomb interactions. This is an essential part for the appropriate description of polar systems, and it represents the energy component that requires most attention for its use in structure optimization methods and MD simulations. Thus, we present explicit formulae for the evaluation of analytic first- and second-order derivatives of these sums. The computational methodology is presented in Section 4. Finally, in Section 5, the here developed MM-CCM methodology is validated for molecular dynamics and optimization applications. To this end, MD simulations in the isobaric-isothermal (NPT) ensemble of liquid argon and structure optimizations of molecular crystals with MM-CCM are presented and discussed. Final conclusions are given in the last section of the article.

## 2 MM-CCM implementation

### 2.1 Molecular mechanics

The MM energy terms can be classified in two categories depending on whether the interacting atoms are bonded or

not. Therefore, the information about the molecular connectivity is required. In `DEMON2K`, this is either given by the user in the input, or automatically generated from the distances between the MM atoms. The resulting MM energy is given by:

$$E^{\text{MM}} = E^{\text{bond}} + E^{\text{bend}} + E^{\text{tors}} + E^{\text{itor}} + E^{\text{vdW}} + E^{\text{QQ}} \quad (1)$$

The first four terms represent the bond stretching, angle bending, torsion and out-of-plane (or improper) torsion energies. The first two terms in eq. (1) are harmonic potentials. The following two are periodic functions. For the van der Waals and the point charge Coulomb energy terms, the well-known 12-6 Lennard-Jones and Coulomb potentials are used:

$$E^{\text{vdW}} = \frac{1}{2} \sum_{A \neq B} 4\epsilon_{AB} \left[ \left( \frac{\sigma_{AB}}{|\mathbf{A} - \mathbf{B}|} \right)^{12} - \left( \frac{\sigma_{AB}}{|\mathbf{A} - \mathbf{B}|} \right)^6 \right] \quad (2)$$

$$E^{\text{QQ}} = \frac{1}{2} \sum_{A \neq B} \frac{q_A q_B}{|\mathbf{A} - \mathbf{B}|} \quad (3)$$

Throughout this work, capital latin letters,  $A, B, \dots$  denote atoms. The corresponding bold letters are the atomic position vectors.

As for the van der Waals interactions between MM atoms, the coupling of the QM and MM subsystems in the additive QM/MM scheme implemented in `DEMON2K` is also modelled by eq. (2). In that case, the  $A$  and  $B$  indices represent QM and MM atoms, respectively. For the calculation of the QM/MM  $\epsilon_{AB}$  and  $\sigma_{AB}$  combination rules of atomic parameters, *i.e.*  $\epsilon_A$  (QM atom) and  $\epsilon_B$  (MM atom) as well as  $\sigma_A$

(QM atom) and  $\sigma_B$  (MM atom), are used [8]. Therefore, in QM/MM calculations such force field parameters are required for MM and QM atoms. In this work, these atomic parameters are taken from the OPLS force field [21, 22], whereas diatomic parameters are obtained through combination rules which are given by a geometric average by default. Thus, the epsilon parameter for the interaction between two atoms is given by:

$$\epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B} \quad (4)$$

A similar formula is used for the  $\sigma_{AB}$  parameter.

Molecular dynamics and local optimization procedures for the time propagation and structure relaxation of molecular systems, respectively, require first- and second-order analytic derivatives of all terms in eq. (1). Explicit formulae for the implementation of this energy and its derivatives in DEMON2K can be found in [8]. They can be used alone or in combination with QM gradients and second-derivatives for structure optimizations, MD simulations or molecular property calculations. Simulations in the canonical (NVT) ensemble are carried out using the extended system approach with the Nosé-Hoover chain (NHC) thermostat [23, 24]. More recently, the pressure-controlling algorithm from Martyna, Tobias and Klein (MTK) [25] has been implemented in DEMON2K, allowing simulations in the NPT ensemble, too. The necessary volume definition for the barostat is provided by the CCM cluster.

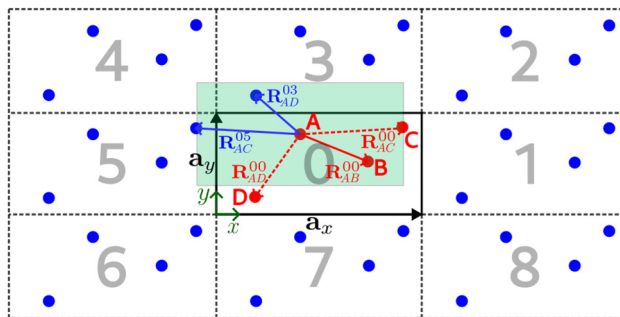
## 2.2 The cyclic cluster model

In DEMON2K, periodicity is introduced in the form of the cyclic cluster model (CCM) in which cyclic boundary conditions are directly applied to a cluster. Thus, the system remains finite. The CCM cluster must be a portion of a system that can be used to generate its periodicity. In principle, a primitive unit cell (PUC) can be used as CCM cluster, but for physically meaningful results a large unit cell (LUC) or supercell is usually considered. Therefore, electronic structure methods developed for molecular quantum chemistry, i.e. for finite systems, can be used in CCM calculations. The same applies to MM methods.

For the case of systems with defects, as mentioned above, direct defect-defect interactions are excluded in the CCM by construction. Consider Figure 2, where a two-dimensional system with four atoms is shown. The central cell, with solid black lines standing for the cell vectors  $\mathbf{a}_x$  and  $\mathbf{a}_y$ , represents the cluster with *real* atoms denoted by red dots. The replicas around the central cell contain translations of these atoms, known as *virtual* atoms in CCM, depicted by blue dots. As is shown in the figure, the interactions for atom A in the cluster

are limited to a region (shown in light green) formed by the restriction that the projection of the distance vector between two atoms onto an edge of the cell is not allowed to exceed half the length of that edge. Three cases are possible for the surrounding atoms: (i) they are inside the interaction region; (ii) they are on its edge; or (iii) they are outside the interaction region. These are the cases of the B, C and D atoms in Figure 2, respectively. No modification of the interatomic vector is required in the first case. However, for the second and third cases, these vectors need to be translated by the same translation vector as that of the cell containing the interacting virtual atom. In addition, for the particular case of the AC interaction, we observe that the real atom in cell 0 and the virtual atom in cell 5 (indicated by the superscript in the interatomic distance vectors in the figure) both fulfill the interaction restriction. In order to avoid double counting of interactions, it is possible to neglect one of them. However, this will cause a dipole moment in ionic systems and lead to unphysical results. To overcome this problem, we consider all the interactions and then take an average as suggested in [15]. By imposing these restrictions, the interaction region around each atom takes the shape of a Wigner-Seitz cell (WSC) generated from the (large) unit cell. As a result, an atom cannot interact with one of its translationally equivalent atoms and, therefore, direct defect-defect interactions are avoided.

Eq. (1) can be directly used in CCM by simply modifying the interatomic distance vectors. For the case of ionic systems, where long-range Coulomb interactions need to be calculated, the expression in eq. (3) will be replaced by the corresponding Ewald sum, as discussed in Section 3, for its efficient calculation. Besides this change, the expressions for



**Figure 2:** Two-dimensional CCM cluster (zero cell) with cell vectors  $\mathbf{a}_x$  and  $\mathbf{a}_y$  containing four atoms (red dots). The CCM cluster replicas 1 to 8 with virtual atoms (blue dots) are indicated by dashed lines. The light green area marks the interaction region for atom A. Interatomic distance vectors are denoted by the letter  $\mathbf{R}$  with a subindex indicating the start and end atoms and a superindex indicating the cell to which these atoms belong. Dashed interaction vectors are used to indicate that these vectors need to be modified. See text for details.

the energy, gradients and second-order energy derivatives of eq. (1) can be directly used for MM-CCM systems. Only the derivatives with respect to the cell parameters need to be added as outlined in the following subsection.

### 2.3 Energy derivatives

In the following, we assume that each edge of the cell lies along one of the Cartesian coordinate axes. Therefore, the corresponding cell vectors in our setup have only one non-vanishing component. For this reason, they will be addressed using only one index. Furthermore, from here on we will use the term cell parameter to refer to the cell vectors only component. Note that these cell parameters are also the lengths of the cell. For the calculation of CCM energy derivatives, we first define the relationship between the atomic position vectors and the cell parameters of the CCM cluster. This can be expressed as follows:

$$\mathbf{A} = f_{Ax}\mathbf{a}_x + f_{Ay}\mathbf{a}_y + f_{Az}\mathbf{a}_z \quad (5)$$

In eq. (5),  $f_{A\lambda}$  ( $\lambda = x, y, z$ ) represent the components of the atomic position vector  $\mathbf{A}$  in the basis of the cell vectors  $\mathbf{a}_x$ ,  $\mathbf{a}_y$  and  $\mathbf{a}_z$ , also known as *fractional coordinates*. Therefore, the derivative of the atomic position vector with respect to a cell parameter is given by the corresponding fractional coordinate. In our setup, the fractional coordinates are simply the ratio  $f_{A\lambda} = A_\lambda/a_\lambda$ , where  $a_\lambda$  is the magnitude of  $\mathbf{a}_\lambda$  and  $A_\lambda$  the corresponding component of the atomic position vector. From this expression, one can obtain the derivative with respect to an atomic coordinate in terms of a cell parameter and *vice versa*. Therefore, we can expand energy derivatives with respect to a cell parameter by the calculated atomic gradient component and the corresponding fractional coordinate:

$$\frac{\partial E}{\partial a_\lambda} = \sum_A \frac{\partial E}{\partial A_\lambda} \frac{\partial A_\lambda}{\partial a_\lambda} = \sum_A \frac{\partial E}{\partial A_\lambda} f_{A\lambda} \quad (6)$$

In our derivatives formulae Greek letters  $\lambda, \eta, \dots$  denote the Cartesian coordinates  $x, y$  and  $z$ . Second-order derivatives can be immediately obtained from the previous equation by recognizing the first-order differential operator with respect to the cell parameter  $a_\lambda$  in it and building the second-order as:

$$\frac{\partial^2 E}{\partial a_\lambda \partial a_\eta} = \sum_{A,B} \frac{\partial^2 E}{\partial A_\lambda \partial B_\eta} f_{A\lambda} f_{B\eta} \quad (7)$$

Finally, it is also possible to obtain mixed second-order derivatives where the differentiation is with respect to one atomic coordinate and one cell parameter. For this purpose,

we can start from eq. (6) and derive it with respect to an atomic coordinate. This yields:

$$\frac{\partial^2 E}{\partial A_\lambda \partial a_\eta} = \sum_B \frac{\partial^2 E}{\partial A_\lambda \partial B_\eta} f_{B\eta} \quad (8)$$

Equations (6)–(8) can be applied to all terms in eq. (1) as long as their dependency on the cell parameters is only through the atomic positions. We will see in the next section that this is not the case for the Ewald part of the energy.

## 3 The Ewald summation method

In order to extend the CCM to polar systems, long-range Coulomb interactions must be added [16]. Unfortunately, eq. (3) is only conditionally convergent when applied to periodic (infinite) systems. Therefore, the summation order must be defined first. A possibility is adding replicas in a spherical way according to their distances to the central unit cell. This definition makes it possible to transform the original sum in eq. (3) into two rapidly and totally convergent sums using the method of Ewald [26–32]. This changes the MM energy expression to:

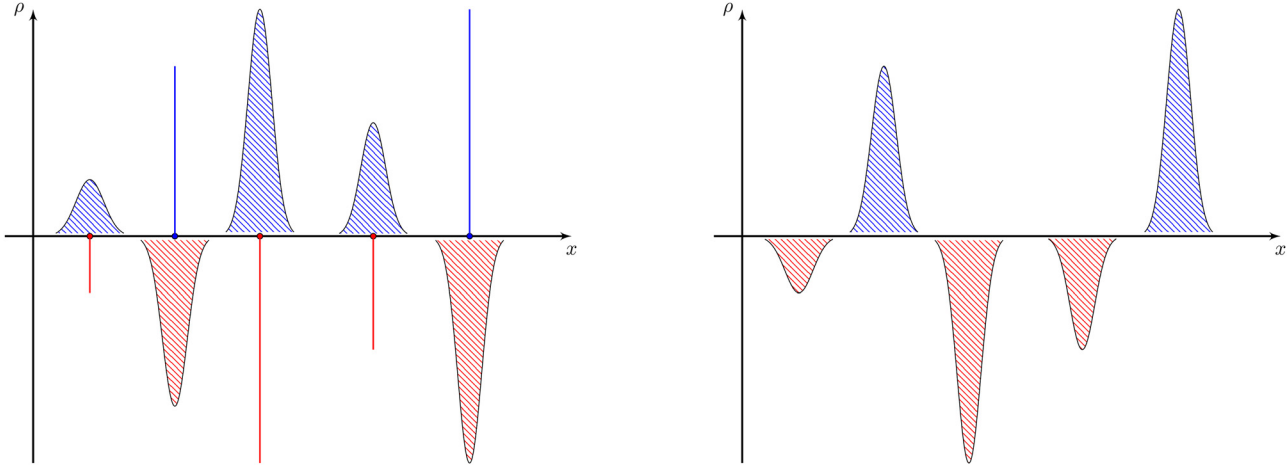
$$E^{\text{MM}} = E^{\text{bond}} + E^{\text{bend}} + E^{\text{tors}} + E^{\text{itor}} + E^{\text{vdW}} + E^{\text{Ew}} \quad (9)$$

The main idea of this approach is to place a charge distribution at the location of every point charge  $q_A$ , but with opposite sign, see Figure 3, left.

The potential arising from the point charges and the charge distributions can be rapidly calculated in direct space. A second set of compensating charge distributions needs to be added to recuperate the initial point charge distribution (see Figure 3, right). This potential is then evaluated in reciprocal space. This yields for the Ewald energy expression:

$$\begin{aligned} E^{\text{Ew}} &= \frac{1}{2} \sum_{A \neq B} \sum_{\mathbf{g}} \frac{q_A q_B}{|\mathbf{A} - \mathbf{B} + \mathbf{g}|} \operatorname{erfc}(\alpha|\mathbf{A} - \mathbf{B} + \mathbf{g}|) \\ &+ \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \mathcal{A}_k Q(\mathbf{k}) Q^*(\mathbf{k}) - \frac{\alpha}{\sqrt{\pi}} \sum_A q_A^2 - \sum_{A,B}^{\text{intra}} \frac{q_A q_B}{|\mathbf{A} - \mathbf{B}|} \\ \text{with } \mathcal{A}_k &= \frac{e^{-k^2/4\alpha^2}}{k^2}; \quad Q(\mathbf{k}) = \sum_A q_A e^{-i\mathbf{k} \cdot \mathbf{A}}; \quad Q^*(\mathbf{k}) = \sum_A q_A e^{i\mathbf{k} \cdot \mathbf{A}} \end{aligned} \quad (10)$$

The first three terms in eq. (10) represent the direct ( $E^{\text{dir}}$ ), reciprocal ( $E^{\text{rec}}$ ), and self-interaction ( $E^{\text{self}}$ ) part of the energy, respectively. The last term accounts for the intramolecular interactions, which are only allowed when the atoms are separated by more than three bonds. Thus, in this sum,  $A$  and  $B$  are the end atoms of either a bond, an angle or



**Figure 3:** One dimensional system with the charge density distributions used in the Ewald method. Point charges and shielding charge distributions (left) give rise to the direct part of the potential, while the compensating charge distributions (right) generate the part of the potential evaluated in reciprocal space.

a dihedral angle. Note that in  $E^{\text{dir}}$ , the sum should not include the term  $A = B$  for  $\mathbf{g} = \mathbf{0}$  to avoid divergence. The  $\mathbf{g}$  and  $\mathbf{k}$  vectors in the direct and reciprocal space, respectively, are given by:

$$\mathbf{g} = m_x \mathbf{a}_x + m_y \mathbf{a}_y + m_z \mathbf{a}_z \quad (11)$$

$$\mathbf{k} = n_x \mathbf{b}_x + n_y \mathbf{b}_y + n_z \mathbf{b}_z \quad (12)$$

Here  $m_\lambda$  and  $n_\lambda$  are integer numbers. The reciprocal vectors,  $\mathbf{b}_\lambda$ , form a lattice in reciprocal space and fulfill the condition  $\mathbf{a}_\lambda \cdot \mathbf{b}_\eta = 2\pi\delta_{\lambda\eta}$ . The number of  $\mathbf{g}$  and  $\mathbf{k}$  vectors needed for the convergence of eq. (10) is strongly dependent from the  $\alpha$  parameter. With  $\alpha = \sqrt{\pi}/V^{1/3}$ , where  $V$  is the volume of the cell, the direct and reciprocal sums converge at the same rate [33]. In DEMON2K, this parameter is calculated as  $\alpha = \kappa/V^{1/3}$  where  $\kappa = 6$  is chosen by default to achieve convergence of  $E^{\text{dir}}$  using only the central unit cell. A different value of  $\alpha$ , as well as the maximum values for the  $m_\lambda$  and  $n_\lambda$  integer numbers in Eqs. (11) and (12), can be given in the input. Note that the introduction of the Ewald sum in eq. (9) in QM/MM methods requires point charges for the replicas of the QM system as depicted in Figure 1, right. To this end, we use the force field charges of the corresponding MM atoms. Therefore, the necessary force field parameters for a QM atom extend to  $\epsilon_A$ ,  $\sigma_A$  and  $q_A$  in this approach.

Depending on the system setup, other corrections must be added to the Ewald energy. Because eq. (10) is derived under the assumption that the system is surrounded by a good conductor, a correction term needs to be added to compensate for the dipolar layer formed on the surface of the sphere arising from the summation definition. This term is:

$$E^{\text{dip}} = \frac{2\pi}{3V} \left| \sum_A q_A \mathbf{A} \right|^2 \quad (13)$$

Another correction is the addition of a background charge for charged systems [34]. This is given by:

$$E^{\text{chg}} = \frac{\pi}{3V\alpha^2} \left| \sum_A q_A \right|^2 \quad (14)$$

With this, the Ewald energy containing all its terms and corrections is given by the expression:

$$E^{\text{Ew}} = E^{\text{dir}} + E^{\text{rec}} - E^{\text{self}} - E^{\text{intra}} + E^{\text{dip}} - E^{\text{chg}} \quad (15)$$

First derivatives for the Ewald sums with respect to the cell parameters have been reported previously [35], although without the results for the correction terms. In the following subsections we present all the terms for completeness and as a basis for obtaining the second-order derivatives. For this same reason, the derivatives with respect to atomic coordinates are also given. The sum over the direct space vectors,  $\mathbf{g}$ , is omitted for compactness of the resulting equations.

### 3.1 First derivatives

Note that in eq. (15), the self-interaction and charge correction terms do not depend on atomic coordinates. Thus, the gradients of  $E^{\text{Ew}}$ , i.e. the energy derivatives with respect to atomic coordinates, consist of four terms only. These are:

$$\begin{aligned} \frac{\partial E^{\text{dir}}}{\partial A_\lambda} = & - \sum_B \frac{q_A q_B}{|\mathbf{A} - \mathbf{B}|^2} \left[ \frac{\text{erfc}(\alpha|\mathbf{A} - \mathbf{B}|)}{|\mathbf{A} - \mathbf{B}|} + \frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2|\mathbf{A} - \mathbf{B}|^2} \right] \\ & \times (A_\lambda - B_\lambda) \end{aligned} \quad (16)$$

$$\frac{\partial E^{\text{rec}}}{\partial A_\lambda} = \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \mathcal{A}_k [Q(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{A}} - Q^*(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{A}}] i q_A k_\lambda \quad (17)$$

$$\frac{\partial E^{\text{intra}}}{\partial A_\lambda} = - \sum_B^{\text{intra}} \frac{q_A q_B}{|\mathbf{A} - \mathbf{B}|^3} (A_\lambda - B_\lambda) \quad (18)$$

$$\frac{\partial E^{\text{dip}}}{\partial A_\lambda} = \frac{4\pi}{3V} q_A \sum_B q_B B_\lambda \quad (19)$$

Analytic gradients of the direct and reciprocal sums have been reported previously in [16], where the derivations were performed with respect to components of the interatomic vectors. Within QM/MM methods, analytic derivatives of eq. (10) with respect to atomic coordinate components can be found in [36].

Gradients with respect to the cell parameters differ from those of the other MM contributions in eq. (9) since the dependency of  $E^{\text{Ew}}$  on the cell parameters is not only through the atomic positions, but also through the volume, the  $\alpha$  parameter and the  $\mathbf{k}$  vectors. In general, the derivative of the Ewald energy with respect to a cell parameter component can be written as:

$$\begin{aligned} \frac{\partial E^{\text{Ew}}}{\partial a_\lambda} &= \sum_A \frac{\partial E^{\text{Ew}}}{\partial A_\lambda} \frac{\partial A_\lambda}{\partial a_\lambda} + \frac{\partial E^{\text{Ew}}}{\partial V} \frac{\partial V}{\partial a_\lambda} + \frac{\partial E^{\text{Ew}}}{\partial \alpha} \frac{\partial \alpha}{\partial a_\lambda} \\ &+ \sum_{\mathbf{k} \neq 0} \frac{\partial E^{\text{Ew}}}{\partial k_\lambda} \frac{\partial k_\lambda}{\partial a_\lambda} \end{aligned} \quad (20)$$

By applying the chain rule, the cell parameter derivatives of the different terms take the following explicit forms:

$$\begin{aligned} \frac{\partial E^{\text{dir}}}{\partial a_\lambda} &= \frac{1}{2a_\lambda} \sum_{A,B} q_A q_B \frac{2\alpha}{3\sqrt{\pi}} e^{-\alpha^2|\mathbf{A}-\mathbf{B}|^2} \\ &- \frac{1}{2} \sum_{A,B} \frac{q_A q_B}{|\mathbf{A} - \mathbf{B}|^2} \left[ \frac{\text{erfc}(\alpha|\mathbf{A} - \mathbf{B}|)}{|\mathbf{A} - \mathbf{B}|} + \frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2|\mathbf{A}-\mathbf{B}|^2} \right] \\ &(A_\lambda - B_\lambda) (f_{A\lambda} - f_{B\lambda}) \end{aligned} \quad (21)$$

$$\frac{\partial E^{\text{rec}}}{\partial a_\lambda} = - \frac{2\pi}{V a_\lambda} \sum_{\mathbf{k} \neq 0} \mathcal{A}_k Q(\mathbf{k}) Q^*(\mathbf{k}) \left( 1 + \frac{\mathbf{k}^2}{6\alpha^2} - \frac{k_\lambda^2}{2\alpha^2} - \frac{2k_\lambda^2}{\mathbf{k}^2} \right) \quad (22)$$

$$\frac{\partial E^{\text{self}}}{\partial a_\lambda} = - \frac{\alpha}{3\sqrt{\pi} a_\lambda} \sum_A q_A^2 \quad (23)$$

$$\frac{\partial E^{\text{dip}}}{\partial a_\lambda} = - \frac{2\pi}{3V a_\lambda} \left| \sum_A q_A \mathbf{A} \right|^2 + \frac{4\pi}{3V} \sum_{A,B} q_A q_B f_{A\lambda} B_\lambda \quad (24)$$

$$\frac{\partial E^{\text{chg}}}{\partial a_\lambda} = - \frac{\pi}{6V \alpha^2 a_\lambda} \left| \sum_A q_A \right|^2 \quad (25)$$

Note that the fractional coordinates,  $f_{A\lambda} - f_{B\lambda}$ , can be written as the ratio of the corresponding atomic coordinates and the cell parameter  $a_\lambda$ , and implemented in this way, too. Nonetheless, it is important to keep track of the fractional coordinates for the following derivatives, in which they are also kept constant. The derivatives of  $E^{\text{intra}}$  are not given explicitly since they can be obtained using eq. (6). Thus, only the derivatives with respect to atomic coordinates are needed for this term.

### 3.2 Second derivatives

For structure optimization and characterization, second-order energy derivatives are needed. As already explained for the gradients only four terms in eq. (15) have second-order non-vanishing derivatives with respect to atomic coordinates. These are given by:

$$\begin{aligned} \frac{\partial^2 E^{\text{dir}}}{\partial A_\lambda \partial B_\eta} &= \frac{(A_\lambda - B_\lambda)(A_\eta - B_\eta)}{|\mathbf{A} - \mathbf{B}|^4} q_A q_B \left[ \frac{3 \text{erfc}(\alpha|\mathbf{A} - \mathbf{B}|)}{|\mathbf{A} - \mathbf{B}|} \right. \\ &+ \left. \frac{2\alpha}{\sqrt{\pi}} (3 + 2\alpha^2|\mathbf{A} - \mathbf{B}|^2) e^{-\alpha^2|\mathbf{A}-\mathbf{B}|^2} \right] \\ &- \frac{\delta_{\lambda\eta}}{|\mathbf{A} - \mathbf{B}|^2} q_A q_B \left[ \frac{\text{erfc}(\alpha|\mathbf{A} - \mathbf{B}|)}{|\mathbf{A} - \mathbf{B}|} + \frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2|\mathbf{A}-\mathbf{B}|^2} \right] \end{aligned} \quad (26)$$

$$\frac{\partial^2 E^{\text{rec}}}{\partial A_\lambda \partial B_\eta} = \frac{2\pi}{V} \sum_{\mathbf{k} \neq 0} \mathcal{A}_k q_A q_B [e^{-i\mathbf{k}\cdot\mathbf{A}} e^{i\mathbf{k}\cdot\mathbf{B}} + e^{-i\mathbf{k}\cdot\mathbf{B}} e^{i\mathbf{k}\cdot\mathbf{A}}] k_\lambda k_\eta \quad (27)$$

$$\frac{\partial^2 E^{\text{intra}}}{\partial A_\lambda \partial B_\eta} = - \frac{3q_A q_B}{|\mathbf{A} - \mathbf{B}|^5} (A_\lambda - B_\lambda)(A_\eta - B_\eta) + \delta_{\lambda\eta} \frac{q_A q_B}{|\mathbf{A} - \mathbf{B}|^3} \quad (28)$$

$$\frac{\partial^2 E^{\text{dip}}}{\partial A_\lambda \partial B_\eta} = \frac{4\pi \delta_{\lambda\eta}}{3V} q_A q_B \quad (29)$$

Analytic second-order derivatives, as well as analytic gradients, of the direct and reciprocal terms with respect to atomic coordinate components have been reported in the literature [37]. In [38], derivatives of the dipole term are also given.

Second-order energy derivatives with respect to the cell parameters are arising for all terms of eq. (15). The formulae for the intramolecular part are not given explicitly as they can be calculated from eq. (7). For the direct sum we find:

$$\begin{aligned}
\frac{\partial^2 E^{\text{dir}}}{\partial a_\lambda \partial a_\eta} &= \frac{1}{2} \sum_{A,B} \frac{q_A q_B}{|\mathbf{A} - \mathbf{B}|^2} \left( 2\alpha^2 + \frac{3}{|\mathbf{A} - \mathbf{B}|^2} \right) \frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2 |\mathbf{A} - \mathbf{B}|^2} (A_\lambda - B_\lambda) (A_\eta - B_\eta) (f_{A\lambda} - f_{B\lambda}) (f_{A\eta} - f_{B\eta}) \\
&+ \frac{1}{2} \sum_{A,B} q_A q_B \frac{3 \operatorname{erfc}(\alpha |\mathbf{A} - \mathbf{B}|)}{|\mathbf{A} - \mathbf{B}|^5} (A_\lambda - B_\lambda) (A_\eta - B_\eta) (f_{A\lambda} - f_{B\lambda}) (f_{A\eta} - f_{B\eta}) \\
&+ \frac{1}{2} \sum_{A,B} \frac{q_A q_B}{a_\lambda a_\eta} \frac{4\alpha^3}{3\sqrt{\pi}} \left[ \frac{1}{3} |\mathbf{A} - \mathbf{B}|^2 - \frac{1}{6\alpha^2} \right] e^{-\alpha^2 |\mathbf{A} - \mathbf{B}|^2} \\
&- \frac{1}{2} \sum_{A,B} q_A q_B \frac{4\alpha^3}{3\sqrt{\pi}} \left[ \frac{(A_\lambda - B_\lambda)(f_{A\lambda} - f_{B\lambda})}{a_\eta} + \frac{(A_\eta - B_\eta)(f_{A\eta} - f_{B\eta})}{a_\lambda} \right] e^{-\alpha^2 |\mathbf{A} - \mathbf{B}|^2} \\
&- \frac{\delta_{\lambda\eta}}{2} \sum_{A,B} \frac{q_A q_B}{a_\lambda a_\eta} \frac{2\alpha}{3\sqrt{\pi}} e^{-\alpha^2 |\mathbf{A} - \mathbf{B}|^2} \\
&- \frac{\delta_{\lambda\eta}}{2} \sum_{A,B} \frac{q_A q_B}{|\mathbf{A} - \mathbf{B}|^2} \left[ \frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2 |\mathbf{A} - \mathbf{B}|^2} + \frac{\operatorname{erfc}(\alpha |\mathbf{A} - \mathbf{B}|)}{|\mathbf{A} - \mathbf{B}|} \right] (f_{A\lambda} - f_{B\lambda}) (f_{A\eta} - f_{B\eta}) \tag{30}
\end{aligned}$$

For the reciprocal energy term the result is:

$$\begin{aligned}
\frac{\partial^2 E^{\text{rec}}}{\partial a_\lambda \partial a_\eta} &= \frac{2\pi}{V a_\lambda a_\eta} \sum_{\mathbf{k} \neq 0} \frac{e^{-\mathbf{k}^2/4a^2}}{\mathbf{k}^2} Q(\mathbf{k}) Q^*(\mathbf{k}) \left( 1 + \frac{\mathbf{k}^2}{6a^2} - \frac{k_\lambda^2}{2a^2} - \frac{2k_\lambda^2}{\mathbf{k}^2} \right) \left( 1 + \frac{\mathbf{k}^2}{6a^2} - \frac{k_\eta^2}{2a^2} - \frac{2k_\eta^2}{\mathbf{k}^2} \right) \\
&- \frac{2\pi}{V a_\lambda a_\eta} \sum_{\mathbf{k} \neq 0} \frac{e^{-\mathbf{k}^2/4a^2}}{\mathbf{k}^2} Q(\mathbf{k}) Q^*(\mathbf{k}) \left[ \frac{1}{3a^2} \left( \frac{\mathbf{k}^2}{3} - k_\lambda^2 - k_\eta^2 \right) - \frac{4k_\lambda^2 k_\eta^2}{\mathbf{k}^4} \right] \\
&+ \frac{2\pi \delta_{\lambda\eta}}{V a_\lambda a_\eta} \sum_{\mathbf{k} \neq 0} \frac{e^{-\mathbf{k}^2/4a^2}}{\mathbf{k}^2} Q(\mathbf{k}) Q^*(\mathbf{k}) \left( 1 + \frac{\mathbf{k}^2}{6a^2} - \frac{3k_\lambda^2}{2a^2} - \frac{6k_\lambda^2}{\mathbf{k}^2} \right) \tag{31}
\end{aligned}$$

And finally, for the self-interaction and correction terms:

$$\frac{\partial^2 E^{\text{self}}}{\partial a_\lambda \partial a_\eta} = \left( \frac{1}{3} + \delta_{\lambda\eta} \right) \frac{\alpha}{3\sqrt{\pi} a_\lambda a_\eta} \sum_A q_A^2 \tag{32}$$

$$\begin{aligned}
\frac{\partial^2 E^{\text{dip}}}{\partial a_\lambda \partial a_\eta} &= \frac{2\pi(1 + \delta_{\lambda\eta})}{3V a_\lambda a_\eta} \left| \sum_A q_A \mathbf{A} \right|^2 \\
&- \frac{4\pi}{3V} \sum_{A,B} q_A q_B \left( \frac{f_{A\lambda} B_\lambda}{a_\eta} - \delta_{\lambda\eta} f_{A\lambda} f_{B\eta} \right) \tag{33}
\end{aligned}$$

$$\frac{\partial^2 E^{\text{chg}}}{\partial a_\lambda \partial a_\eta} = \left( \frac{1}{3} + \delta_{\lambda\eta} \right) \frac{\pi}{6V a^2 a_\lambda a_\eta} \left| \sum_A q_A \right|^2 \tag{34}$$

### 3.3 Mixed second derivatives

For the calculation of the full Hessian matrix, mixed second-order derivatives with respect to atomic coordinates and cell parameters are needed, too. To this end, we start from equations (21)–(25) and differentiate them with respect to an atomic coordinate. As explained in the derivation of the gradients, neither the self-interaction nor the charge correction have any dependency on the atomic coordinates. Therefore, only the direct, reciprocal and dipole terms contribute to these derivatives. For the direct energy term we have:



$$\begin{aligned}
\frac{\partial^2 E^{\text{dir}}}{\partial A_\lambda \partial a_\eta} &= \sum_{B \neq A} \frac{3q_A q_B}{|\mathbf{A} - \mathbf{B}|^4} \left[ \frac{\text{erfc}(\alpha|\mathbf{A} - \mathbf{B}|)}{|\mathbf{A} - \mathbf{B}|} + \frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2|\mathbf{A} - \mathbf{B}|^2} \right] (A_\lambda - B_\lambda) (A_\eta - B_\eta) (f_{A\eta} - f_{B\eta}) \\
&+ \sum_{B \neq A} \frac{q_A q_B}{|\mathbf{A} - \mathbf{B}|^2} \frac{4\alpha^3}{\sqrt{\pi}} e^{-\alpha^2|\mathbf{A} - \mathbf{B}|^2} \left[ (A_\eta - B_\eta) (f_{A\eta} - f_{B\eta}) - \frac{|\mathbf{A} - \mathbf{B}|^2}{3a_\eta} \right] (A_\lambda - B_\lambda) \\
&- \sum_{B \neq A} \frac{\delta_{\lambda\eta} q_A q_B}{|\mathbf{A} - \mathbf{B}|^2} \left[ \frac{\text{erfc}(\alpha|\mathbf{A} - \mathbf{B}|)}{|\mathbf{A} - \mathbf{B}|} + \frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2|\mathbf{A} - \mathbf{B}|^2} \right] (f_{A\eta} - f_{B\eta})
\end{aligned} \quad (35)$$

The differentiation of eq. (22) with respect to an atomic coordinate results in:

$$\frac{\partial^2 E^{\text{rec}}}{\partial A_\lambda \partial a_\eta} = \frac{i2\pi q_A}{V a_\eta} \sum_{\mathbf{k} \neq 0} \mathcal{A}_k \left( \frac{k_\eta^2}{2\alpha^2} + \frac{2k_\eta^2}{\mathbf{k}^2} - \frac{\mathbf{k}^2}{6\alpha^2} - 1 \right) [Q(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{A}} - Q^*(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{A}}] k_\lambda \quad (36)$$

And finally, for the dipole correction:

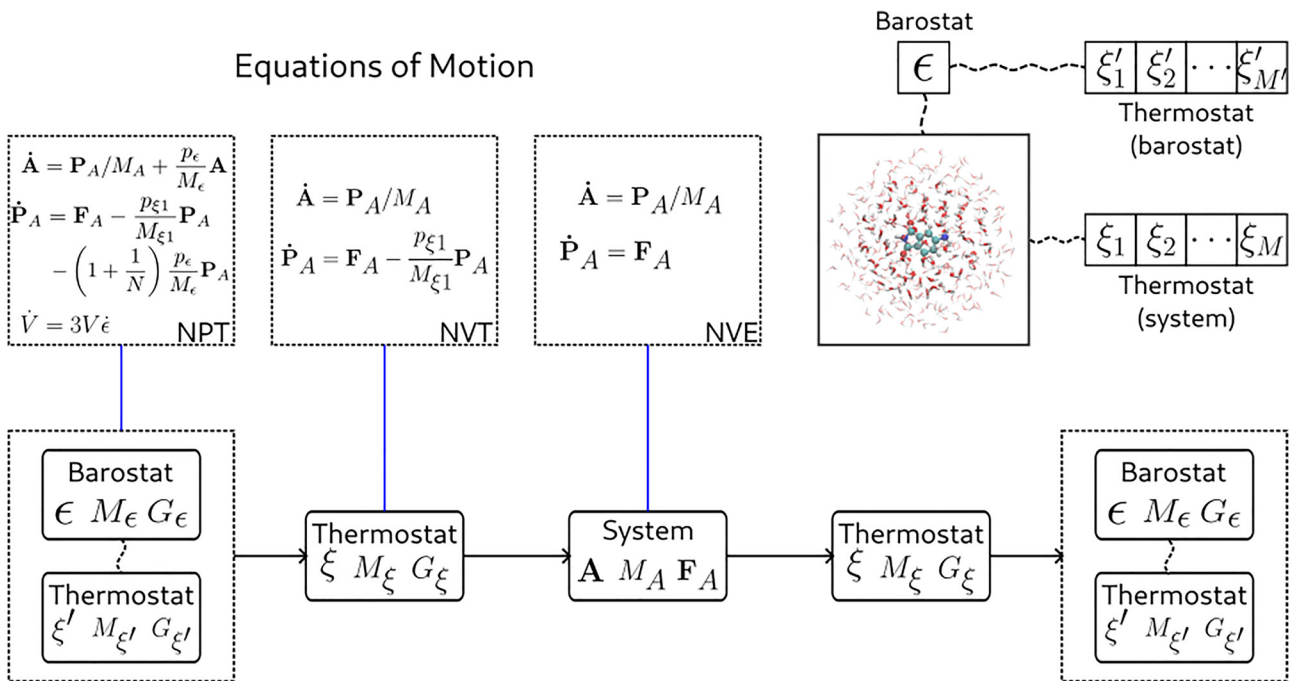
$$\frac{\partial^2 E^{\text{dip}}}{\partial A_\lambda \partial a_\eta} = -\frac{4\pi}{3V a_\eta} q_A \sum_B q_B B_\lambda + \frac{4\pi \delta_{\lambda\eta}}{3V} q_A \sum_B q_B f_{B\eta} \quad (37)$$

With these derivations, we can calculate the gradient and Hessian for the optimization of periodic systems. Note that the use of complex algebra is restricted to the calculation of the reciprocal energy term and its derivatives. Of course, all energy derivatives are real in nature as it can be shown by substituting the complex exponentials by cosine and sine functions. The use of complex algebra in obtaining

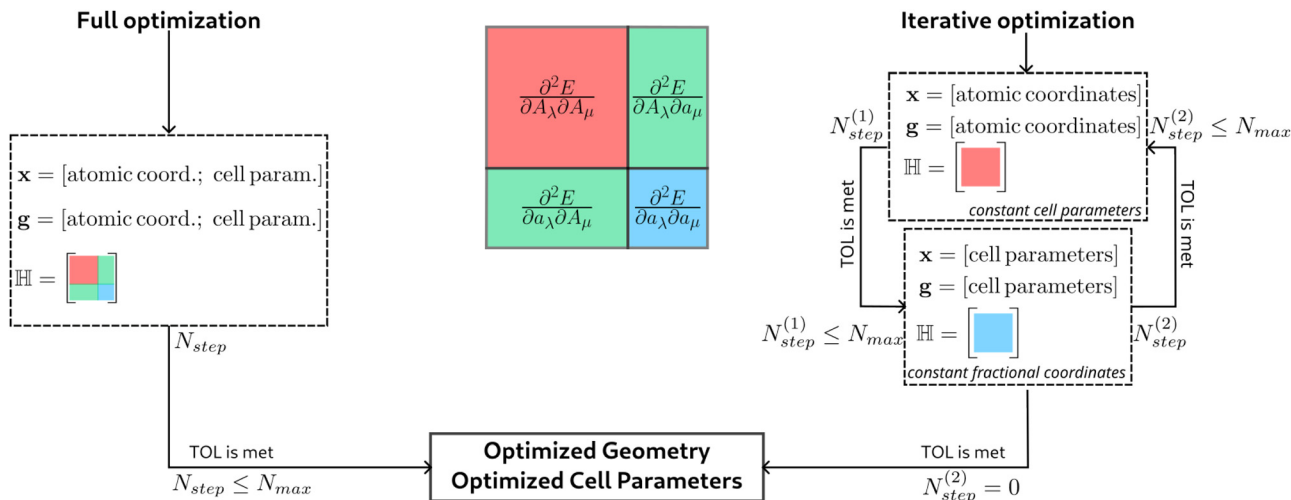
$Q(\mathbf{k})$ , and the terms derived from it, permits a more efficient calculation of eq. (10) and its derivatives by reducing a double sum over the atoms to just one [29].

## 4 Computational methodology

The incorporation of temperature and pressure into an MD simulation is based on the idea of modifying the equations of motion (EOM) by coupling the system to controlling devices. For temperature, the average value of the kinetic energy is driven towards a target value using the Nosé-Hoover chain (NHC) thermostat allowing simulations in the NVT ensemble. A pressure-controlling device can also be coupled to the volume of the system to drive its fluctuations towards the desired average pressure value. The new general MD step



**Figure 4:** General flowchart of the new deMon2k MD step. A thermostat is used for the system and another for the barostat. The coupling of the thermostat and barostat to the particle system is done through the modification of the equations of motion for the atomic positions ( $\mathbf{A}$ ) and momenta ( $\mathbf{P}_A$ ). A schematic representation of this coupling is shown in the upper right corner. A dot above a quantity indicates differentiation with respect to time.



**Figure 5:** Diagram of the full and iterative optimization procedures. Here,  $\mathbf{x}$ ,  $\mathbf{g}$  and  $\mathbb{H}$  represent the optimization variables, their gradients and their second derivatives, respectively.

implemented in DEMON2K is shown in Figure 4. Here the particle system collects all atom and cell parameter coordinates. The barostat and thermostats are composed of *fictitious* particles, each with a position and mass, and subject to forces, represented by the three symbols inside each box. Only the equations of motion for the particle system are given in Figure 4. The equations for the time propagation of the positions and momenta of the barostat and thermostats are not shown for brevity.

To validate our new CCM MD implementation we run NPT simulations of 864 argon atoms inside a cubic box with an initial length of 34.98 Å. Nosé-Hoover chains consisting of three thermostats were coupled to this system and to the barostat. The time step ( $\Delta t$ ) was set to 5.0 fs. A first test of the equivalence between the different ensembles was conducted by 50 ps simulations. For the specific case of the barostat implementation, simulations of 125 ps were carried out under different temperatures and pressures, too. The strength of the coupling of the thermostat and barostat to the system is set by their masses ( $M_\xi, M_\epsilon$  in Figure 4). These are proportional to characteristic time scales for the particles and barostat,  $\tau_p$  and  $\tau_b$ . As suggested in [39], we set them proportional to the time step to  $\tau_p = 20 \Delta t$  and  $\tau_b = 100 \Delta t$ .

The default structure minimization algorithm in DEMON2K is based on *quasi*-Newton methods in which the analytic Hessian or some approximation to it is used in conjunction with an update algorithm [40]. External degrees of freedom, particularly translational degrees of freedom for periodic systems, need to be removed from the gradient vector and the Hessian matrix. In DEMON2K, this is done

through a projector matrix as outlined in [9]. For the validation of the CCM structure optimization, we optimized the molecular structure and cell parameters of molecular crystals without any symmetry restriction using Cartesian coordinates. In this way, the crystal structure was allowed to change into a different spatial symmetry group during the optimization.

Throughout this work the convergence tolerance for the root mean square (RMS) of the forces was set to  $3.0 \times 10^{-4}$  a.u. and to  $4.5 \times 10^{-4}$  a.u. for the maximum force in all optimizations. The maximum number of optimization steps was set to 1000. Two different optimization procedures were tested. The first one, which we name full optimization, optimizes both kinds of degrees of freedom, namely the atomic coordinates and the cell parameters, simultaneously. The second one, named iterative optimization, optimizes atomic coordinates and cell parameters independently from each other. Figure 5 shows schematic diagrams for both approaches. Note that the mixed second-order derivatives are never used in the iterative scheme. These coupling terms represent the greatest difference between the two approaches. As starting Hessian unit matrices were used in both approaches. For the full optimization a calculated start Hessian was tested, too.

## 5 Results

As a first test of the suitability of the implemented NVT and NPT algorithms, the equilibrium properties obtained from

**Table 1:** Average property values obtained from MD simulations of the argon system described in the text. The total energy (last column) is different for each simulation due to the inclusion of the thermostats and barostat fictitious particles.

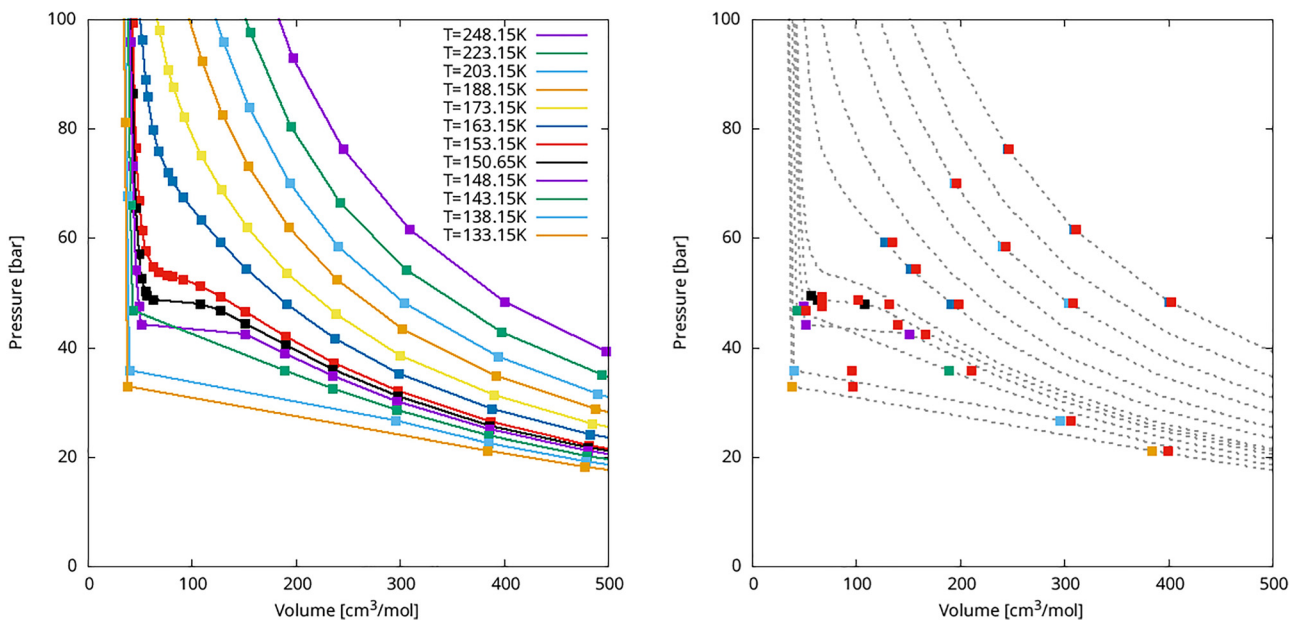
Simulation	Property				$\langle E^{\text{Tot}} \rangle$ (a.u.)
	$\langle T \rangle$ (K)	$\langle P \rangle$ (bar)	$\langle \text{Box length} \rangle$ (Å)	$\langle E^{\text{Pot}} \rangle$ (a.u.)	
NVE	$283 \pm 4$	$2752 \pm 86$	$34.98 \pm 0.00$	$-1.4242 \pm 0.02$	$-0.2640 \pm 3 \times 10^{-5}$
NVT	$283 \pm 8$	$2743 \pm 122$	$34.98 \pm 0.00$	$-1.4256 \pm 0.02$	$-0.2558 \pm 3 \times 10^{-5}$
NPT	$283 \pm 8$	$2751 \pm 150$	$34.99 \pm 7.19$	$-1.4231 \pm 0.02$	$2.4460 \pm 3 \times 10^{-5}$

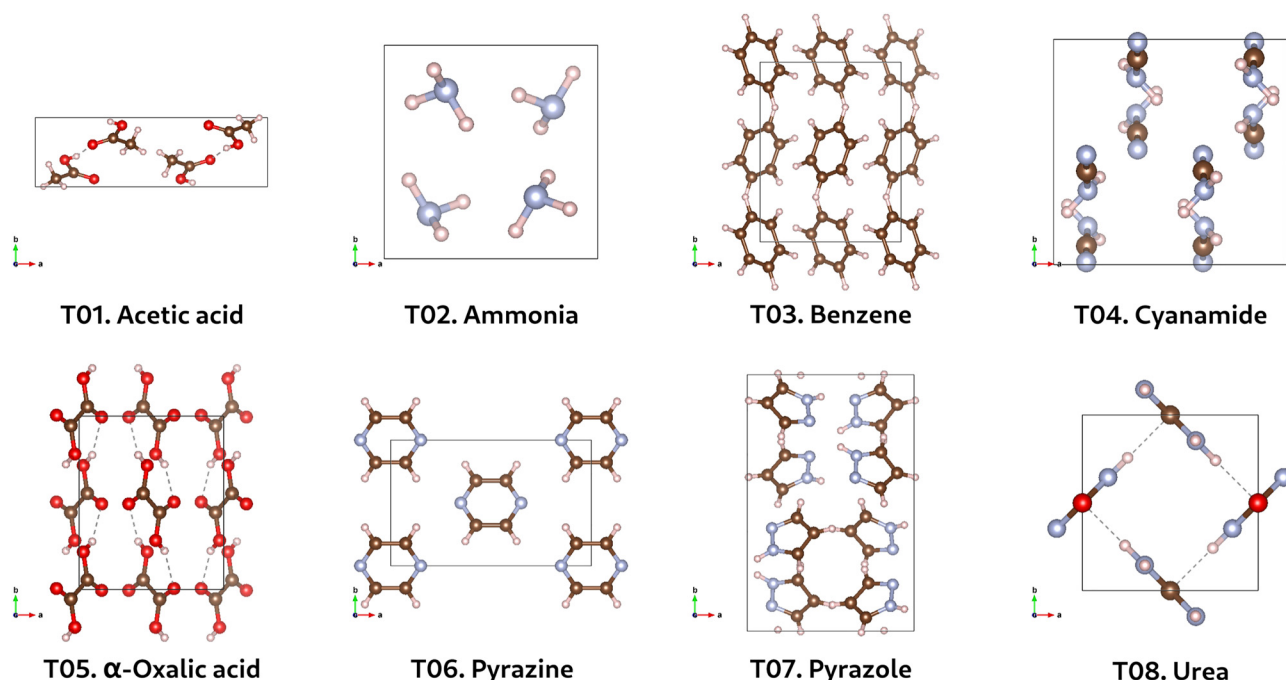
the NVE, NVT and NPT simulations of the 864 argon atom system were compared. The same system setup as described in the computational methodology section was used to this end. In these simulations we recorded 50 ps trajectories. First, an NVE simulation was performed. Average temperature and pressure values were obtained from this simulation and used for the generation of trajectories from NVT and NPT runs. The results are summarized in Table 1. The agreement between the different ensembles, especially between the NVE and NPT, indicates that the implemented constant-pressure algorithm is able to describe this system accurately while conserving the total energy to a reasonable degree.

Further MD simulations of the argon system in the NPT ensemble were performed generating trajectories of 125 ps. The temperature and pressure values were set according to the experimental data reported in [41]. The selected volume and pressure values from the experimental isotherms shown in Figure 6 (left) are depicted on the graph on the

right part of the picture with colored dots. These pressure values, together with the corresponding temperatures, were used as the target conditions for the simulations. From these, the final average volume was calculated and used to plot the results shown in Figure 6 (right) by red squares. In general, the volume is overestimated by the Lennard-Jones model of argon. Good agreement with experiment is found for high temperatures, whereas larger discrepancies appear when the pressure is increased along the lower temperature curves. This is in agreement with the ideal gas model.

To test the CCM optimization procedures described in the previous section, a set of molecular crystals consisting of eight crystals with orthorhombic unit cells, see Figure 7, are optimized. These systems are acetic acid, ammonia, benzene, cyanamide,  $\alpha$ -oxalic acid, pyrazine, pyrazole and urea. The starting molecular structures and cell parameters were taken from references [42–49] and correspond to the experimentally determined crystalline structures which, at room temperature and under ambient pressure, do not

**Figure 6:** Experimental (left) isotherms of argon. Some points (right) were picked from these isotherms. The red squares represent the calculated volumes for each  $T, P$  pair chosen.



**Figure 7:** Test set for optimization consisting of eight orthorhombic molecular crystals. Intramolecular and intermolecular interactions were modelled according to eq. (9).

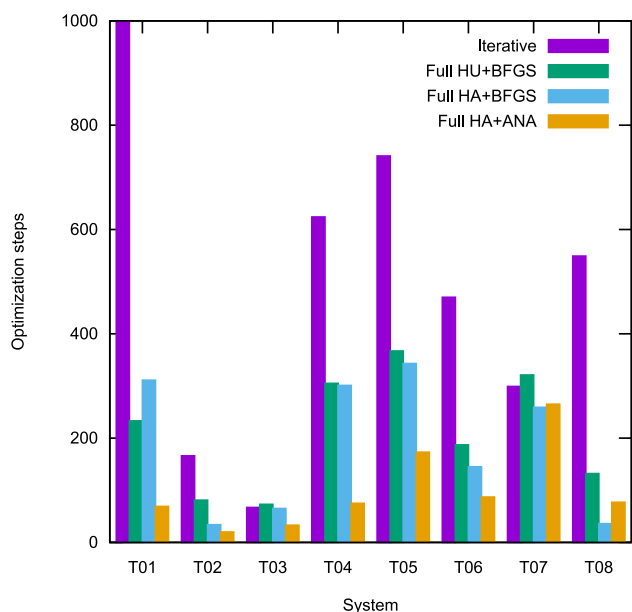
present polymorphism. The intermolecular interactions in these systems include electrostatic, hydrogen-bond and  $\pi$ -stacking, among others. In our MM treatment though, only electrostatic and van der Waals interactions are considered. In Table 2 the results from the optimization of all systems are compared between the two approaches depicted in Figure 5. As starting Hessian a unit matrix was used for this comparison. Given the differences between the two optimization procedures, we cannot expect to obtain the same results for each system. Nonetheless, as it can be seen from Table 2, the optimized cell parameters do not differ by more than 1.0 Å between the two approaches, except for system T01 which did not converge within 1000 steps using the iterative approach. The change in volume

and the decrease in energy follow the same trend in both approaches, with a general tendency towards lower volumes indicating a denser molecular packing in these systems.

Full optimizations were performed for all systems using the analytic second-order derivatives matrix as starting Hessian, too. Two optimization algorithms were tested. One *quasi-Newton* method based on the BFGS update formula [40], and a Newton restricted step method that calculates second-order derivatives in each step. The number of optimization steps to reach convergence is shown in Figure 8. The use of the analytic Hessian accounts for a reduction in the number of steps, although not as large as when switching from the iterative to the full optimization procedure. An exception is system T08 for which the use of the analytically

**Table 2:** Cell parameters (Å), change in volume (%) and decrease in total energies (a.u.) for the optimized test systems with a starting unit Hessian.

System	Initial parameters			Optimized (iter)			$\Delta V$	$\Delta E$	Optimized (full)			$\Delta V$	$\Delta E$
T01	13.1510	3.9230	5.7620	13.5913	3.7616	5.4262	-6.68	-0.2206	15.1289	3.5722	4.9161	-10.63	-0.2411
T02	5.1305	5.1305	5.1305	4.9188	4.9188	4.9188	-11.88	-0.0135	4.9604	4.9604	4.9604	-9.62	-0.0150
T03	7.3900	9.4200	6.8100	7.3643	9.3771	6.7705	-1.38	-0.0089	7.3959	9.2472	6.7210	-3.04	-0.0090
T04	6.8560	6.6280	9.1470	7.2886	6.6199	9.8996	+14.92	-0.3503	7.2984	6.3968	10.1086	+13.54	-0.3529
T05	6.5480	7.8440	6.0860	6.6886	7.2784	6.9125	+7.65	-0.0785	6.2959	6.1029	9.0912	+11.75	-0.1141
T06	9.3250	5.8500	3.7330	9.4463	6.5187	3.0927	-6.48	-0.1141	9.2727	6.8726	3.0935	-3.19	-0.1145
T07	8.1900	12.5880	6.7730	8.1068	13.4625	6.2754	-1.92	-0.4230	7.5627	14.1918	6.3021	-3.13	-0.4344
T08	5.5650	5.5650	4.6840	5.2247	5.2247	5.1262	-3.53	-0.0110	5.1654	5.1654	5.2918	-2.67	-0.0112



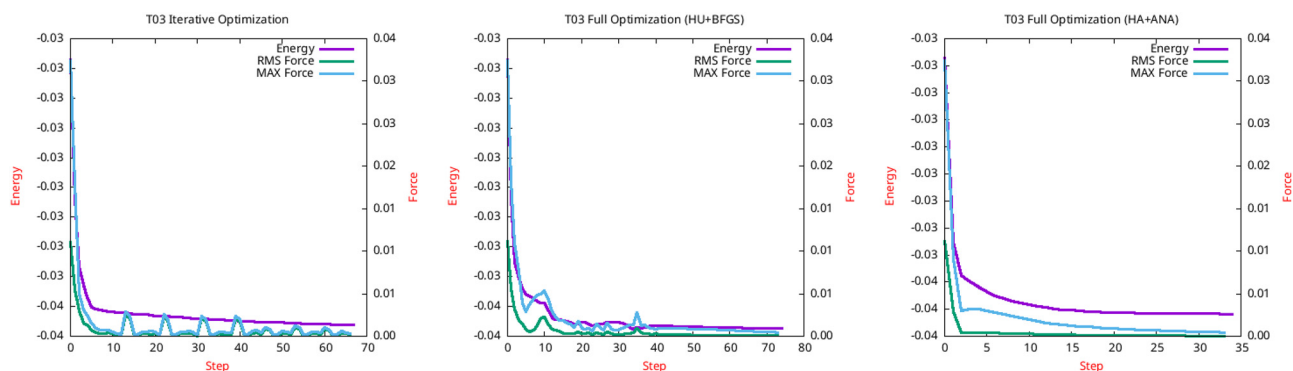
**Figure 8:** Number of optimization steps for the iterative and full optimization procedures. HU indicates that the starting Hessian is the unit matrix, whereas HA means that the start Hessian was analytically calculated. In the *quasi*-Newton optimization the Hessian was updated with the BFGS formula, whereas it was analytically (ANA) calculated in the corresponding Newton method.

calculated Hessian at each step of the optimization required a larger number of steps. Closer inspection of the final structure reveals a minimum lowest in energy by 0.0145 a.u., suggesting that a different optimized structure had been reached. Finally, to illustrate the differences in the behaviour of the energy and forces between the two optimization procedures, we depict the energy, as well as the RMS and maximum forces, for the optimization of system T03 in Figure 9. During the iterative optimization (left), oscillations in the force curves can be observed. They arise because of the independent optimization of the two degrees of freedom.

When one is optimized, the other is left constant, and thus having minimized one kind of forces, the other kind are not necessarily small. These oscillations are not present in the full optimization (middle), in particular near convergence. Besides the general decrease in the number of optimization steps, a qualitative improvement can be seen (right) when using the analytic Hessian in each optimization step. This can be due to the different nature of the atomic coordinates and cell parameters, which demands a more precise treatment when the two are considered together, since, as for now, the full optimization implemented in DEMON2K uses the same step lengths and tolerances for both degrees of freedom.

## 6 Conclusions

The cyclic cluster model has been implemented in the MM branch of the DEMON2K program for orthorhombic systems. For ionic systems, we use the Ewald method, for the calculation of Coulomb interactions. The corresponding analytic first- and second-order energy derivatives with respect to both the atomic coordinates and cell parameters have been derived and implemented. Molecular dynamics simulations have been conducted in different ensembles, using the NHC thermostat and the MTK barostat. The simulations show a good total energy conservation and correspondence between the different methodologies. The argon simulations show that the implemented equations of motion are suitable for constant pressure and temperature simulations. As expected, the Lennard-Jones potential with the OPLS force field parameters give best results near the ideal gas conditions for this system. Two CCM optimization approaches have been presented, of which the full optimization shows to be more convenient in terms of the number of optimization steps over the iterative approach,



**Figure 9:** Profiles of the energy and the RMS and maximum forces from the optimization of system T03 where oscillations in the forces are observed for the iterative procedure (left). Smoother energy and force curves are obtained with the use of the analytic Hessian in each step of the full optimization (right). See text for further details.

with a considerable improvement in this respect even with a starting unit Hessian. In most cases, the use of the analytic second-order derivatives while optimizing both degrees of freedom at the same time further reduces the steps to reach convergence. The extension of this work to QM/MM is currently under development in our laboratories.

**Acknowledgments:** This work was supported by the CONACYT project A1-S-11929 and the ECOS project 321168. JDSR gratefully acknowledges CONACYT for his PhD fellowship 490270.

**Research ethics:** Not applicable.

**Author contributions:** The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** The authors state no conflict of interest.

**Research funding:** This work was supported by the CONACYT 490270; CONACYT A1-S-11929; ECOS 321168.

**Data availability:** The raw data and computational program can be obtained on reasonable requests from the corresponding authors.

## References

- Warshel A., Karplus M. *J. Am. Chem. Soc.* 1972, 94, 5612–5625.
- Warshel A., Levitt M. *J. Mol. Biol.* 1976, 103, 227–249.
- Köster A. M., Reveles J. U., del Campo J. M. *J. Chem. Phys.* 2004, 121, 3417–3424.
- Calaminici P., Alvarez-Ibarra A., Cruz-Olvera D., Domínguez-Soria V.-D., Flores-Moreno R., Gamboa G. U., Geudtner G., Goursot A., Mejía-Rodríguez D., Salahub D. R., Zuniga-Gutierrez B., Köster A. M. Auxiliary density functional theory: from molecules to nanostructures. In *Handbook of Computational Chemistry*; Leszczynski J., Kaczmarek-Kedziera A., Puzyn T., Papadopoulos M. G., Reis H., Shukla K. M., Eds. Springer International Publishing: Cham (Switzerland), 2017; pp. 795–860.
- Geudtner G., Calaminici P., Carmona-Espíndola J., del Campo J. M., Domínguez-Soria V. D., Moreno R. F., Gamboa G. U., Goursot A., Köster A. M., Reveles J. U., Mineva T., Vásquez-Pérez J. M., Vela A., Zúñiga-Gutiérrez B., Salahub D. R. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* 2012, 2, 548–555.
- Salahub D. R., Noskov S. Y., Lev B., Zhang R., Ngo V., Goursot A., Calaminici P., Köster A. M., Alvarez-Ibarra A., Mejía-Rodríguez D., Řezáč J., Cailliez F., de la Lande A. *Molecules* 2015, 20, 4780–4812.
- de la Lande A., Alvarez-Ibarra A., Hasnaoui K., Cailliez F., Wu X., Mineva T., Cuny J., Calaminici P., López-Sosa L., Geudtner G., Navizet I., García Iriepa C., Salahub D. R., Köster A. M. *Molecules* 2019, 24, 1653.
- Samaniego-Rojas J. D., Hernández-Segura L. I., López-Sosa L., Delgado-Venegas R. I., Gomez B., Lambry J. C., de la Lande A., Mineva T., Alejandro J., Zúñiga-Gutiérrez B. A., Flores-Moreno R., Calaminici P., Geudtner G., Köster A. M. QM/MM with Auxiliary DFT in deMon2k. In *Multiscale Dynamics Simulations: Nano and Nano-bio Systems in Complex Environments*; Salahub D. R., Wei D., Eds. The Royal Society of Chemistry: London, 2021; pp. 1–54.
- López-Sosa L., Calaminici P., Köster A. M. *J. Comput. Chem.* 2023, 44, 2358–2368.
- Deák P. *Acta Phys. Acad. Sci. Hung.* 1981, 50, 247–262.
- Dobrotvorskii A. M., Evarestov E. A. *Phys. Status Solidi B* 1974, 66, 83–91.
- Deák P. *Phys. Status Solidi B* 2000, 217, 9–21.
- Evarestov R. A., Petrashen M. I., Ledovskaya E. M. *Phys. Status Solidi B* 1975, 68, 453–461.
- Noga J., Baňacký P., Biskupič S., Boča R., Pelikán P., Svrček M., Zajac A. *J. Comput. Chem.* 1999, 20, 253–261.
- Bredow T., Geudtner G., Jug K. *J. Comput. Chem.* 2001, 22, 89–101.
- Janetzko F., Bredow T., Jug K. *J. Chem. Phys.* 2002, 116, 8994–9004.
- Janetzko F., Evarestov R. A., Bredow T., Jug K. *Phys. Status Solidi B* 2004, 241, 1032–1040.
- Peintinger M. F., Bredow T. *J. Comput. Chem.* 2014, 35, 839–846.
- Miró J., Deák P., Ewels C. P., Jones R. *J. Phys.: Condens. Matter* 1997, 9, 9555–9562.
- Janetzko F., Köster A. M., Salahub D. R. *J. Chem. Phys.* 2008, 128, 024102.
- Jorgensen W. L., Tirado-Rives J. *J. Am. Chem. Soc.* 1988, 110, 1657–1666.
- Dodda L. S., Cabeza de Vaca I., Tirado-Rives J., Jorgensen W. L. *Nucleic Acids Res.* 2017, 45, W331–W336.
- Nosé S. *J. Chem. Phys.* 1984, 81, 511–519.
- Martyna G. J., Klein M. L., Tuckerman M. *J. Chem. Phys.* 1992, 97, 2635–2643.
- Martyna G. J., Tobias D. J., Klein M. L. *J. Chem. Phys.* 1994, 101, 4177–4189.
- Ewald P. P. *Ann. Phys.* 1921, 369, 253–287.
- de Leeuw S. W., Perram J. W., Smith E. R. *Proc. R. Soc. A* 1980, 373, 27–56.
- Toukmaji A. Y., Board J. A. *Comput. Phys. Commun.* 1996, 95, 73–92.
- Allen M. P., Tildesley D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 2017.
- Hammonds K. D., Heyes D. M. *J. Chem. Phys.* 2022, 157, 074108.
- Heyes D. M. *Phys. Rev. B* 1994, 49, 755–764.
- Brown D., Neyertz S. *Mol. Phys.* 1995, 84, 577–595.
- Nijboer B., De Wette F. *Physica* 1957, 23, 309–321.
- Hub J. S., de Groot B. L., Grubmüller H., Groenhof G. *J. Chem. Theory Comput.* 2014, 10, 381–390.
- Doll K., Dovesi R., Orlando R. *Theor. Chem. Acc.* 2004, 112, 394–402.
- Holden Z. C., Rana B., Herbert J. M. *J. Chem. Phys.* 2019, 150, 144115.
- Wells B. A., Chaffee A. L. *J. Chem. Theory Comput.* 2015, 11, 3684–3695.
- Simmonett A. C., Brooks B. R. *J. Chem. Phys.* 2021, 154, 104101.
- Tuckerman M. E., Alejandro J., López-Rendón R., Jochim A. L., Martyna G. J. *J. Phys. A* 2006, 39, 5629–5651.
- Reveles J. U., Köster A. M. *J. Comput. Chem.* 2004, 25, 1109–1116.
- Michels A., Levelt J. M., De Graaff W. *Physica* 1958, 24, 659–671.
- Boese R., Bläser D., Latz R., Bäumen A. *Acta Crystallogr.* 1999, C55, IUC9900001.
- Boese R., Niederprüm N., Bläser D., Maulitz A., Antipin M. Y., Mallinson P. R. *J. Phys. Chem. B* 1997, 101, 5794–5799.
- Bacon G. E., Curry N. A., Wilson S. A., Spence R. *Proc. R. Soc. A* 1964, 279, 98–110.
- Denner L., Luger P., Buschmann J. *Acta Crystallogr.* 1988, C44, 1979–1981.
- Derissen J. L., Smith P. H. *Acta Crystallogr.* 1974, B30, 2240–2242.
- de With G., Harkema S., Feil D. *Acta Crystallogr.* 1976, B32, 3178–3185.
- La Cour T., Rasmussen S. E. *Acta Chem. Scand.* 1973, 27, 1845–1854.
- Zavodnik V., Stash A., Tsirelson V., de Vries R., Feil D. *Acta Crystallogr.* 1999, B55, 45–54.