

QM/MM with Auxiliary DFT

Juan D Samaniego-Rojas, Luis-Ignacio Hernández-Segura, Luis López-Sosa, Rogelio I Delgado-Venegas, Badhin Gomez, Jean-Christophe Lambry, Aurelien de la Lande, Tzonka Mineva, José Alejandre, Bernardo A Zúñiga-Gutiérrez, et al.

▶ **To cite this version:**

Juan D Samaniego-Rojas, Luis-Ignacio Hernández-Segura, Luis López-Sosa, Rogelio I Delgado-Venegas, Badhin Gomez, et al.. QM/MM with Auxiliary DFT. Dennis R. Salahub; Dongqing Wei. Multiscale Dynamics Simulations: Nano and Nano-bio Systems in Complex Environments, 22, RSC, pp.1 - 54, 2021, Theoretical and Computational Chemistry Series, 978-1-83916-467-5. 10.1039/9781839164668-00001 . hal-03452193

HAL Id: hal-03452193
https://hal.science/hal-03452193

Submitted on 1 Aug 2022

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.

Published on 24 September 2021 on https://pubs.rsc.org | doi:10.1039/9781839164668-00001

QM/MM with Auxiliary DFT in deMon2k

Juan D. Samaniego-Rojas[a], Luis-Ignacio Hernández-Segura[a], Luis López-Sosa[a], Rogelio I. Delgado-Venegas[b], Badhin Gomez[c], Jean-Christophe Lambry[d], Aurelien de la Lande[e], Tzonka Mineva[f], José Alejandre[g], Bernardo A. Zúñiga-Gutiérrez[h], Roberto Flores-Moreno[h], Patrizia Calaminici[a], Gerald Geudtner[a], Andreas M. Köster[a]

[a] Departamento de Química, Cinvestav, Avenida Instituto Politécnico Nacional 2508, A.P. 14-740, CDMX, C.P. 07360, México

[b] Department of Chemistry, University of Calgary, 2500 University Dr. NW, Calgary, Alberta, Canada, T2N 1N4

[c] Departamento de Farmacía, Bioquímica y Biotecnología, Universidad Católica de Santa María, Arequipa, Perú

[d] Laboratoire d'Optique et Biosciences, Ecole polytechnique, Institut polytechnique de Paris, CNRS, INSERM, Route de Saclay, 91128 Palaiseau cedex, France

[e] Laboratoire de Chimie Physique, CNRS, Université Paris Sud, Université Paris Saclay, 15 avenue Jean Perrin, 91405 Orsay, France

[f] ICGM, Univ. Montpellier, CNRS, ENSCM, Montpellier, France

[g] Departamento de Química, UAM-Iztapalapa, San Rafael Atlixco 186, A.P. 55-534, CDMX, C.P. 09340, México

[h] Departamento de Química, Universidad de Guadalajara, CUCEI, Blvd. Marcelino García Barragán 1421, C.P. 44430 Guadalajara, Jalisco, México

Abstract

This chapter describes the theoretical background of the Quantum Mechanical/Molecular Mechanical (QM/MM) implementation in deMon2k within the framework of Auxiliary Density Functional Theory (ADFT). It aims to give the reader an overview of the current state-of-the-art of this QM/MM implementation and perspectives for its future development. To this end, we first derive the ADFT working equations for the QM and QM/MM energy and gradient expressions. Based on the joined QM/MM gradient expression, we present algorithms for QM/MM structure optimizations, transition state searches and molecular dynamics simulations. The use of auxiliary density perturbation theory (ADPT) in the framework of QM/MM is discussed on illustrative example implementations including analytic second-order ADFT energy derivatives, nuclear magnetic resonance chemical shift calculations and excited state calculations by time-dependent ADFT. The chapter closes with the description of a transformation program to generate deMon2k QM/MM inputs.

1. Introduction

Over the last two decades, many Quantum Mechanical/Molecular Mechanical (QM/MM) implementations have emerged based on the pioneering works of Warshel and Karplus [1] and Warshel and Levitt [2] published in the seventies. Common to these approaches is the possibility to describe a relatively small part of a complex system at a more or less sophisticated QM level of theory whereas a computationally efficient MM force field is employed for the remaining part of the system. The fundamental idea behind these QM/MM methodologies is to balance the accuracy and reliability of the simulation with the size of the underlying model. Naturally, the interface between the QM and MM regions is most delicate and defines to a large extend the achievable accuracy and reliability of the overall simulation. Therefore, it is common practice to extend the QM region as much as possible in order to diminish these interface effects. To this end, a promising first-principle QM approach is Kohn-Sham Density Functional Theory (DFT) [3,4]. Although its computational demand is similar to the well-known Hartree-Fock method, it incorporates electron correlation through the used exchange-correlation functional. In the Linear Combination of Gaussian Type Orbital (LCGTO) approximation, Kohn-Sham DFT possesses a formal N_{bas}^4 scaling, N_{bas} being the number of basis functions in the calculation, due to the expansion of the two-electron Coulomb repulsion into four-center electron repulsion integrals (ERIs). Already in very early molecular LCGTO-DFT Kohn-Sham implementations, the calculation of the four-center ERIs was identified as a critical computational bottleneck. To overcome this bottleneck, Dunlap and co-workers [5,6] introduced the variational fitting of the Coulomb potential. Due to its variational nature, this approximation yields also a variational energy expression free of four-center ERIs. Note, however, that the variational bound shifts below the corresponding four-center ERI Kohn-Sham bound. For relative energies, this shift in the variational bound has no effect. Thus, Kohn-Sham LCGTO-DFT approaches incorporating the variational fitting of the Coulomb

potential yield the same relative energies as their four-center ERI counterparts at a considerably reduced computational cost. The same holds for so-called resolution-of-the-identity (RI) approaches [7] because RI is a direct result of the variational fitting of the Coulomb potential and, therefore, possesses the same variational bound [8]. To avoid four-center ERI calculations for hybrid functionals, the variational fitting of the Fock potentials [9-15] was introduced in deMon2k [16], too. The resulting four-center ERI free Kohn-Sham energy model is one of the first-principle QM approaches available in deMon2k for QM/MM simulations. It is named Density Fitted DFT (DF-DFT).

The computational bottleneck in medium sized (50 to 500 atoms) DF-DFT calculations is the numerical integration of the exchange-correlation energy and its functional derivatives. Although, this computational bottleneck has been identified at the same time as the one for the two-electron Coulomb repulsion, it has received much less attention. Early LCGTO-DFT implementations like deMon-KS [17] or DGAUS [18] used the semi-numerical fit of the exchange-correlation potential with auxiliary functions as proposed by Sambe and Felton [19] to circumvent this bottleneck. This approach gives reasonable energies but is prune to numerical instabilities in energy derivatives [20]. To overcome this problem the direct use of the (linear scaling) approximated density for the calculation of the exchange-correlation potential was suggested [21-24]. In the framework of deMon2k this has led to the development of Auxiliary Density Functional Theory (ADFT) [25] which is another first-principle energy model available for deMon2k QM/MM simulations. Because ADFT yields relative energies of the same quality as Kohn-Sham, albeit with a significant reduction in computational demand, we will focus in the following on ADFT as first-principle QM approach in QM/MM calculations. With the now available MPI parallelized ADFT implementation in deMon2k large QM regions with 500 to 5000 atoms have become feasible. In

these calculations, the computational bottleneck has shifted to linear algebra tasks such as matrix multiplications and matrix diagonalizations. In this respect, it is important to note that the variational fitting of the Coulomb potential as well as the use of the approximated density for the calculation of the exchange-correlation contributions introduce additional linear algebra tasks in form of solving inhomogeneous linear equation systems. To avoid corresponding computational bottlenecks we have adapted Krylov subspace methods, such as MINRES [26] and Eirola-Nevanlinna [27], for the density fitting [28] and perturbation theory [29] calculations in deMon2k.

This chapter is organized as follows. In the next section, we derive the QM/MM energy and gradient working equations implemented in deMon2k. Section 3 describes algorithms for QM/MM structure optimizations and transition state searches based on the trust region method for local optimization. In the next section, the propagator equations for QM/MM molecular dynamics simulations in various closed system ensembles using factorized Liouville operators are presented. Section 5 introduces auxiliary density perturbation theory (ADPT) for the calculation of analytic second-order ADFT energy derivatives. Their extension to QM/MM, currently under development in the deMon developers' community, is discussed, too. The following section describes deMon2k magnetic shielding and excited states calculations within the QM/MM framework. These are typical examples for QM/MM electronic property calculations employing perturbation theory. Section 7 presents the new QM/MM Input Builder (QIB) for deMon2k. A résumé of the current state-of-the-art ADFT QM/MM implementation in deMon2k is given in the last section of this book chapter.

2. Energies and gradients

The QM LCGTO Kohn-Sham energy expression is given by:

$$E^{QM} = \sum_{\mu,\nu} P_{\mu\nu}\, H_{\mu\nu}^{QM} + \frac{1}{2}\sum_{\mu,\nu}\sum_{\sigma,\tau} P_{\mu\nu}\, P_{\sigma\tau}\, \langle \mu\nu \| \sigma\tau \rangle + E_{xc}[\rho] \tag{1}$$

In our notation Greek letters, here μ, ν, σ and τ, denote (contracted) atom centered Gaussian type orbitals. For clarity of presentation, we restrict ourselves to closed-shell systems. The closed-shell density matrix elements, $P_{\mu\nu}$, are calculated as:

$$P_{\mu\nu} = 2\sum_{i}^{occ} c_{\mu i}\, c_{\nu i} \tag{2}$$

The upper limit in Eq. (2), occ, denotes the number of doubly occupied molecular orbitals (MOs) in the QM system and $c_{\mu i}$ are the corresponding MO coefficients. For an isolated QM system, the first term in Eq. (1) represents the mono-electronic core energy. The QM core matrix elements are given in atomic units, which we use throughout this chapter, by:

$$H_{\mu\nu}^{QM} = -\frac{1}{2}\langle \mu | \boldsymbol{\nabla}^2 | \nu \rangle + \sum_{C}^{QM} \left\langle \mu \left| \frac{Z_C}{|\boldsymbol{r} - \boldsymbol{C}|} \right| \nu \right\rangle \tag{3}$$

They collect the kinetic energy and nuclear attraction of the electrons. The second term in Eq. (1) describes the Coulomb repulsion between the electrons. It includes the four-center two-electron integrals responsible for the N_{bas}^4 scaling of the method:

$$\langle \mu\nu \| \sigma\tau \rangle = \iint \frac{\mu(\boldsymbol{r}_1)\, \nu(\boldsymbol{r}_1)\, \sigma(\boldsymbol{r}_2)\, \tau(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\, d\boldsymbol{r}_1\, d\boldsymbol{r}_2 \tag{4}$$

In the here used short-hand ERI notation $\|$ represents the two-electron Coulomb operator $1/|\boldsymbol{r}_1 - \boldsymbol{r}_2|$. It also separates functions from electron 1, on the left, from those of electron 2, on the right. In the following, we will use the same notation for two- and three-center ERIs. The last

term in Eq. (1) denotes the exchange-correlation energy functional evaluated with the Kohn-Sham density and its derivatives. In LCGTO the Kohn-Sham density can be expressed as:

$$\rho(\boldsymbol{r}) = \sum_{\mu,\nu} P_{\mu\nu}\, \mu(\boldsymbol{r})\, \nu(\boldsymbol{r}) \tag{5}$$

As already mentioned, the two-electron Coulomb repulsion in Eq. (1) introduces a formal N_{bas}^4 scaling due to the four-center ERIs. To circumvent this computational bottleneck, we employ the variational fitting of the Coulomb potential based on the following second-order energy error functional [5,6]:

$$\varepsilon_2^H = \frac{1}{2} \iint \frac{[\rho(\boldsymbol{r}_1) - \tilde{\rho}(\boldsymbol{r}_1)]\ [\rho(\boldsymbol{r}_2) - \tilde{\rho}(\boldsymbol{r}_2)]}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\, d\boldsymbol{r}_1\, d\boldsymbol{r}_2 \tag{6}$$

The here appearing approximated electronic density, $\tilde{\rho}(\boldsymbol{r})$, also named auxiliary density, is calculated as a linear combination of auxiliary functions which we denote by Latin letters:

$$\tilde{\rho}(\boldsymbol{r}) = \sum_{\bar{k}} x_{\bar{k}}\, \bar{k}(\boldsymbol{r}) \tag{7}$$

In deMon2k atom centered primitive Hermite Gaussian functions, indicated by a bar, are used as auxiliary functions [30]. Inserting this expansion of $\tilde{\rho}(\boldsymbol{r})$ and the LCGTO expansion of the Kohn-Sham density, Eq. (5), into Eq. (6) yields:

$$\begin{aligned} \varepsilon_2^H &= \frac{1}{2}\ \langle\rho\|\rho\rangle - \langle\rho\|\tilde{\rho}\rangle + \frac{1}{2}\ \langle\tilde{\rho}\|\tilde{\rho}\rangle \\ &= \frac{1}{2} \sum_{\mu,\nu} \sum_{\sigma,\tau} P_{\mu\nu}\, P_{\sigma\tau}\ \langle\mu\nu\|\sigma\tau\rangle - \sum_{\mu,\nu} \sum_{\bar{k}} P_{\mu\nu}\ \langle\mu\nu\|\bar{k}\rangle\ x_{\bar{k}} + \frac{1}{2} \sum_{\bar{k},\bar{l}} x_{\bar{k}}\, x_{\bar{l}}\, \langle\bar{k}\|\bar{l}\rangle \end{aligned} \tag{8}$$

It is straightforward to show that ε_2^H is semi-positive definite [31]. Thus, we obtain from Eq. (8):

$$\frac{1}{2} \sum_{\mu,\nu} \sum_{\sigma,\tau} P_{\mu\nu}\, P_{\sigma\tau}\ \langle\mu\nu\|\sigma\tau\rangle \geq \sum_{\mu,\nu} \sum_{\bar{k}} P_{\mu\nu}\ \langle\mu\nu\|\bar{k}\rangle\ x_{\bar{k}} - \frac{1}{2} \sum_{\bar{k},\bar{l}} x_{\bar{k}}\, x_{\bar{l}}\, \langle\bar{k}\|\bar{l}\rangle \tag{9}$$

Inserting this inequality into the Kohn-Sham energy expression, Eq. (1), yields the QM DF-DFT energy expression:

$$E^{QM} = \sum_{\mu,\nu} P_{\mu\nu}\, H_{\mu\nu}^{QM} + \sum_{\mu,\nu}\sum_{\bar{k}} P_{\mu\nu}\, \langle \mu\nu \| \bar{k} \rangle\; x_{\bar{k}} - \frac{1}{2}\sum_{\bar{k},\bar{l}} x_{\bar{k}}\, x_{\bar{l}}\, \langle \bar{k} \| \bar{l} \rangle + E_{xc}[\rho] \tag{10}$$

The Coulomb fitting coefficients, $x_{\bar{k}}$, in Eq. (10) are calculated by minimizing the second-order error functional of Eq. (6), which is accompanied by an energy maximization [32]:

$$\frac{\partial \varepsilon_2^H}{\partial x_{\bar{k}}} = -\sum_{\mu,\nu} P_{\mu\nu}\, \langle \mu\nu \| \bar{k} \rangle + \sum_{\bar{l}} x_{\bar{l}}\, \langle \bar{l} \| \bar{k} \rangle \equiv 0 \quad \forall\, \bar{k} \tag{11}$$

Collecting these equations for all auxiliary functions yields the inhomogeneous linear equation system,

$$\boldsymbol{G}\, \boldsymbol{x} = \boldsymbol{J}\,, \tag{12}$$

which must be solved in each self-consistent field (SCF) iteration. In Eq. (12) $\boldsymbol{G}$ and $\boldsymbol{J}$ are the Coulomb matrix and vector, respectively, given by

$$\boldsymbol{G} = \begin{pmatrix} \langle \bar{1} \| \bar{1} \rangle & \langle \bar{1} \| \bar{2} \rangle & \dots & \langle \bar{1} \| \bar{m} \rangle \\ \langle \bar{2} \| \bar{1} \rangle & \langle \bar{2} \| \bar{2} \rangle & \dots & \langle \bar{2} \| \bar{m} \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \bar{m} \| \bar{1} \rangle & \langle \bar{m} \| \bar{2} \rangle & \dots & \langle \bar{m} \| \bar{m} \rangle \end{pmatrix} \text{ and } \boldsymbol{J} = \begin{pmatrix} \langle \rho \| \bar{1} \rangle \\ \langle \rho \| \bar{2} \rangle \\ \vdots \\ \langle \rho \| \bar{m} \rangle \end{pmatrix}.$$

As the QM Kohn-Sham energy, the QM DF-DFT energy is variational, due to the variational fitting, but with a lower energy bound. For relative energies this shift of the variational bound has no effect.

To obtain the ADFT energy expression, only the functional argument of E_{xc} must be changed to the auxiliary density. Thus, the QM ADFT energy is given by [22]:

$$E^{QM} = \sum_{\mu,\nu} P_{\mu\nu}\, H_{\mu\nu}^{QM} + \sum_{\mu,\nu}\sum_{\bar{k}} P_{\mu\nu}\, \langle \mu\nu \| \bar{k} \rangle\; x_{\bar{k}} - \frac{1}{2}\sum_{\bar{k},\bar{l}} x_{\bar{k}}\, x_{\bar{l}}\, \langle \bar{k} \| \bar{l} \rangle + E_{xc}[\tilde{\rho}] \tag{13}$$

To demonstrate that also the ADFT QM energy is variational, we now calculate ADFT Kohn-Sham matrix elements as derivatives of the energy with respect to density matrix elements:

$$K_{\mu\nu} = \frac{\partial E^{QM}}{\partial P_{\mu\nu}} = H_{\mu\nu}^{QM} + \sum_{\bar{k}} \langle \mu\nu \| \bar{k} \rangle \; x_{\bar{k}} + \frac{\partial E_{xc}[\tilde{\rho}]}{\partial P_{\mu\nu}} \tag{14}$$

For the derivative of the ADFT exchange-correlation energy, $E_{xc}[\tilde{\rho}]$, assuming the local density approximation (LDA) for clarity of discussion, follows:

$$\frac{\partial E_{xc}[\tilde{\rho}]}{\partial P_{\mu\nu}} = \int \frac{\delta E_{xc}[\tilde{\rho}]}{\delta \tilde{\rho}(\boldsymbol{r})} \frac{\partial \tilde{\rho}(\boldsymbol{r})}{\partial P_{\mu\nu}} \, d\boldsymbol{r} \tag{15}$$

To proceed, we now define the ADFT exchange-correlation potential, $v_{xc}[\tilde{\rho}]$, as the functional derivative of the ADFT exchange-correlation energy with respect to the auxiliary density:

$$v_{xc}[\tilde{\rho}] \equiv \frac{\delta E_{xc}[\tilde{\rho}]}{\delta \tilde{\rho}(\boldsymbol{r})} \tag{16}$$

The differentiation of the auxiliary density in Eq. (15) can be expressed with the formal solution of Eq. (12),

$$x_{\bar{k}} = \sum_{\bar{l}} G_{\bar{k}\bar{l}}^{-1} \langle \bar{l} \| \mu\nu \rangle \, P_{\mu\nu} \, ,$$

as:

$$\frac{\partial \tilde{\rho}(\boldsymbol{r})}{\partial P_{\mu\nu}} = \sum_{\bar{k},\bar{l}} \bar{k}(\boldsymbol{r}) \, G_{\bar{k}\bar{l}}^{-1} \langle \bar{l} \| \mu\nu \rangle \tag{17}$$

Inserting Eq. (16) and (17) into Eq. (15) yields for the derivative of the ADFT exchange-correlation energy:

$$\frac{\partial E_{xc}[\tilde{\rho}]}{\partial P_{\mu\nu}} = \sum_{\bar{k},\bar{l}} \langle \mu\nu \| \bar{k} \rangle \, G_{\bar{k}\bar{l}}^{-1} \langle \bar{l} | v_{xc}[\tilde{\rho}] \rangle \tag{18}$$

Thus, the ADFT Kohn-Sham matrix elements can be straightforwardly calculated by the variation of the corresponding energy. This demonstrates the variational nature of the ADFT energy

expression in Eq. (13), albeit with its own variational bound. As for the DF-DFT energy, relative ADFT energies are in good agreement with their Kohn-Sham counterparts. Therefore, ADFT yields relative energies that are almost indistinguishable from their Kohn-Sham counterparts ($\lesssim 0.1$ kcal/mol) with significant reduction in the computational scaling and demand. For the explicit formulation of the ADFT Kohn-Sham matrix elements it is convenient to introduce the exchange-correlation coefficients as:

$$z_{\bar{k}} \equiv \sum_{\bar{l}} G^{-1}_{\bar{k}\bar{l}} \left\langle \bar{l} \middle| v_{xc}[\tilde{\rho}] \right\rangle \tag{19}$$

Eq. (19) can be reformulated as an inhomogeneous linear equation system analog to the Coulomb fitting equation system in Eq. (12):

$$\boldsymbol{G}\,\boldsymbol{z} = \boldsymbol{L} \tag{20}$$

In Eq. (20) $\boldsymbol{L}$ denotes the exchange-correlation vector with elements $L_{\bar{k}} = \left\langle \bar{k} \middle| v_{xc}[\tilde{\rho}] \right\rangle$. Note that $\boldsymbol{z}$ is spin polarized. In the most recent deMon2k versions the linear equation systems of Eqs. (12) and (20) are solved by the Krylov subspace method MINRES [28]. As a result, the computational overhead associated to the calculation of Coulomb and exchange-correlation coefficients is negligible even for systems with hundreds of thousands of auxiliary functions. Thus, the use of extended auxiliary function sets such as GEN-A2* [33] presents no computational challenge. With the Coulomb and exchange-correlation coefficients at hand, the ADFT Kohn-Sham matrix can be expressed as:

$$K_{\mu\nu} = \frac{\partial E^{QM}}{\partial P_{\mu\nu}} = H^{QM}_{\mu\nu} + \sum_{\bar{k}} \left\langle \mu\nu \middle\| \bar{k} \right\rangle \left(x_{\bar{k}} + z_{\bar{k}}\right) \tag{21}$$

Note that the ADFT Kohn-Sham matrix is free of density matrix elements, which greatly simplifies data handling in parallel applications. This is one reason for the excellent parallel scaling of ADFT.

Whereas it is rather straightforward to extend the here outlined LDA ADFT discussion to functionals of the generalized gradient approximation (GGA), the computationally efficient implementation of hybrid functionals into ADFT (and DF-DFT) is less obvious. To facilitate our discussion, we use the following generic ADFT hybrid functional QM energy expression:

$$E^{QM} = \sum_{\mu,\nu} P_{\mu\nu}\, H_{\mu\nu}^{QM} + \sum_{\mu,\nu}\sum_{\bar{k}} P_{\mu\nu}\, \langle \mu\nu \| \bar{k} \rangle\, x_{\bar{k}} - \frac{1}{2}\sum_{\bar{k},\bar{l}} x_{\bar{k}}\, x_{\bar{l}}\, \langle \bar{k} \| \bar{l} \rangle + \alpha\, E_x^F + E_{xc}[\tilde{\rho}] \qquad (22)$$

Here α represents the mixing parameter for the Fock exchange energy, E_x^F, of the hybrid functional. The corresponding counterpart $(1-\alpha)$ for the DFT exchange energy is absorbed into the ADFT exchange-correlation energy functional and will be of no further concern to the following discussion. For a more detailed discussion of the structure of global and range-separated hybrid functionals in ADFT, we refer the interested reader to [15,34-36]. The computational drawback arises from the Fock exchange energy,

$$E_x^F = -\sum_{\mu,\nu}\sum_{\sigma,\tau} P_{\mu\nu}\, P_{\sigma\tau}\, \langle \mu\sigma \| \nu\tau \rangle\,,$$

which contains four-center ERIs and, therefore, will jeopardize the computational benefits of ADFT. To overcome this problem, we implemented the variational fitting of the Fock potential [13,15] in deMon2k. To this end, we express the Fock exchange energy in terms of occupied MOs,

$$\psi_i(\boldsymbol{r}) = \sum_{\mu} c_{\mu i}\, \mu(\boldsymbol{r})\,,$$

as

$$E_x^F = -\sum_{i,j}^{occ} \iint \frac{\psi_i(\boldsymbol{r}_1)\, \psi_j(\boldsymbol{r}_1)\, \psi_i(\boldsymbol{r}_2)\, \psi_j(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\, d\boldsymbol{r}_1\, d\boldsymbol{r}_2 \qquad (23)$$

For the variational Fock potential fitting we now introduce MO transition densities, $\rho_{ij}(\boldsymbol{r})$, and corresponding approximated transition densities, $\tilde{\rho}_{ij}(\boldsymbol{r})$:

$$\rho_{ij}(\boldsymbol{r}) = \psi_i(\boldsymbol{r})\psi_j(\boldsymbol{r}) = \sum_{\mu,\nu} c_{\mu i}\, c_{\nu j}\, \mu(\boldsymbol{r})\, \nu(\boldsymbol{r}) \tag{24}$$

$$\tilde{\rho}_{ij}(\boldsymbol{r}) = \sum_{\bar{k}} x_{\bar{k}ij}\, \bar{k}(\boldsymbol{r}) \tag{25}$$

With these MO transition densities the variational fitting of the Fock potential can be based on the following second-order error functional [13,15]:

$$\varepsilon_2^F = -\sum_{i,j}^{occ} \iint \frac{\left[\rho_{ij}(\boldsymbol{r}_1) - \tilde{\rho}_{ij}(\boldsymbol{r}_1)\right]\left[\rho_{ij}(\boldsymbol{r}_2) - \tilde{\rho}_{ij}(\boldsymbol{r}_2)\right]}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\, d\boldsymbol{r}_1\, d\boldsymbol{r}_2 \tag{26}$$

Expanding the transition densities in Eq. (26) yields:

$$\varepsilon_2^F = -\sum_{i,j}^{occ} \langle \rho_{ij} \| \rho_{ij} \rangle + 2 \sum_{i,j}^{occ} \langle \rho_{ij} \| \tilde{\rho}_{ij} \rangle - \sum_{i,j}^{occ} \langle \tilde{\rho}_{ij} \| \tilde{\rho}_{ij} \rangle$$

$$= -\frac{1}{4} \sum_{\mu,\nu} \sum_{\sigma,\tau} P_{\mu\nu}\, P_{\sigma\tau}\, \langle \mu\sigma \| \nu\tau \rangle + 2 \sum_{i,j}^{occ} \sum_{\bar{k}} \langle \psi_i\, \psi_j \| \bar{k} \rangle\, x_{\bar{k}ij} - \sum_{i,j}^{occ} \sum_{\bar{k},\bar{l}} x_{\bar{k}ij}\, x_{\bar{l}ij}\, \langle \bar{k} \| \bar{l} \rangle$$

Because ε_2^F is (termwise) semi-negative definite [35] the following inequality holds:

$$-\frac{1}{4} \sum_{\mu,\nu} \sum_{\sigma,\tau} P_{\mu\nu}\, P_{\sigma\tau}\, \langle \mu\sigma \| \nu\tau \rangle \leq -2 \sum_{i,j}^{occ} \sum_{\bar{k}} \langle \psi_i\, \psi_j \| \bar{k} \rangle\, x_{\bar{k}ij} + \sum_{i,j}^{occ} \sum_{\bar{k},\bar{l}} x_{\bar{k}ij}\, x_{\bar{l}ij}\, \langle \bar{k} \| \bar{l} \rangle \tag{27}$$

The Fock exchange fitting coefficients are obtained by the maximization of the negative fitting error,

$$\frac{\partial \varepsilon_2^F}{\partial x_{\bar{k}ij}} = -2 \sum_{i,j}^{occ} \langle \psi_i\, \psi_j \| \bar{k} \rangle + 2 \sum_{\bar{l}} x_{\bar{l}ij}\, \langle \bar{l} \| \bar{k} \rangle \equiv 0 \quad \forall\, \bar{k} \wedge i,j \in occ, \tag{28}$$

which is accompanied by the minimization of the approximated Fock exchange energy. This suggests that the simultaneous variational fitting of Coulomb and Fock potentials will benefit from systematic error compensation if the same auxiliary function set is used for both fits. In fact, if the variational nature of both potential fits are preserved, the resulting four-center ERI free total

Hartree-Fock energies for molecules with thousands of basis functions are nearly indistinguishable ($<$ 1 kcal/mol) from their four-center ERI counterparts! Corresponding binding energy differences between four- and three-center Hartree-Fock calculations are below 10^{-2} kcal/mol. From Eq. (28) the following linear equation systems,

$$\boldsymbol{G}\,\boldsymbol{x}_{ij} = \boldsymbol{J}_{ij} \;\; \forall \;\; i,j \in occ, \tag{29}$$

with

$$J_{\bar{k}ij} = \langle \psi_i \, \psi_j \| \bar{k} \rangle \tag{30}$$

are obtained. Therefore, the Fock exchange fitting coefficients are given by:

$$x_{\bar{k}ij} = \sum_{\bar{l}} G^{-1}_{\bar{k}\bar{l}} \, \langle \bar{l} \, \| \, \psi_i \, \psi_j \rangle \tag{31}$$

Inserting these Fock exchange fitting coefficients into Eq. (27) and the resulting approximation for the Fock exchange energy into Eq. (22) yields the following generic ADFT hybrid functional energy expression free of four-center ERIs:

$$E^{QM} = \sum_{\mu,\nu} P_{\mu\nu} \, H^{QM}_{\mu\nu} + \sum_{\mu,\nu} \sum_{\bar{k}} P_{\mu\nu} \, \langle \mu\nu \| \bar{k} \rangle \;\; x_{\bar{k}} - \frac{1}{2} \sum_{\bar{k},\bar{l}} x_{\bar{k}} \, x_{\bar{l}} \, \langle \bar{k} \| \bar{l} \rangle - \alpha \sum_{i,j}^{occ} \sum_{\bar{k},\bar{l}} \langle \psi_i \, \psi_j \| \bar{k} \rangle \;\; G^{-1}_{\bar{k}\bar{l}} \;\; \langle \bar{l} \, \| \, \psi_i \, \psi_j \rangle + E_{xc}[\tilde{\rho}] \tag{32}$$

Direct implementation of Eq. (32) leads to an algorithm that scales as $\mathcal{O}(N_{aux} \times N_{bas}^2 \times N_{occ})$ with N_{aux} and N_{occ} being the number of auxiliary functions and occupied MOs, respectively. Such an algorithm is useful only when $\mathcal{O}(N_{bas}) \gg \mathcal{O}(N_{occ})$ [10,11]. Note, however, that Eq. (32) is invariant under orthogonal transformation of the MOs, because the density matrix is invariant under such transformations. Thus, any set of MOs obtained by an orthogonal transformation of the canonical MOs (CMOs) can be used in Eq. (32). In particular, the CMOs (see Fig.1a) can be

transformed into spatially localized MOs (LMOs) by minimizing or maximizing, an appropriated functional (Fig 1b). This facilitates efficient screening in the integral transformation from the atomic to the molecular representation.

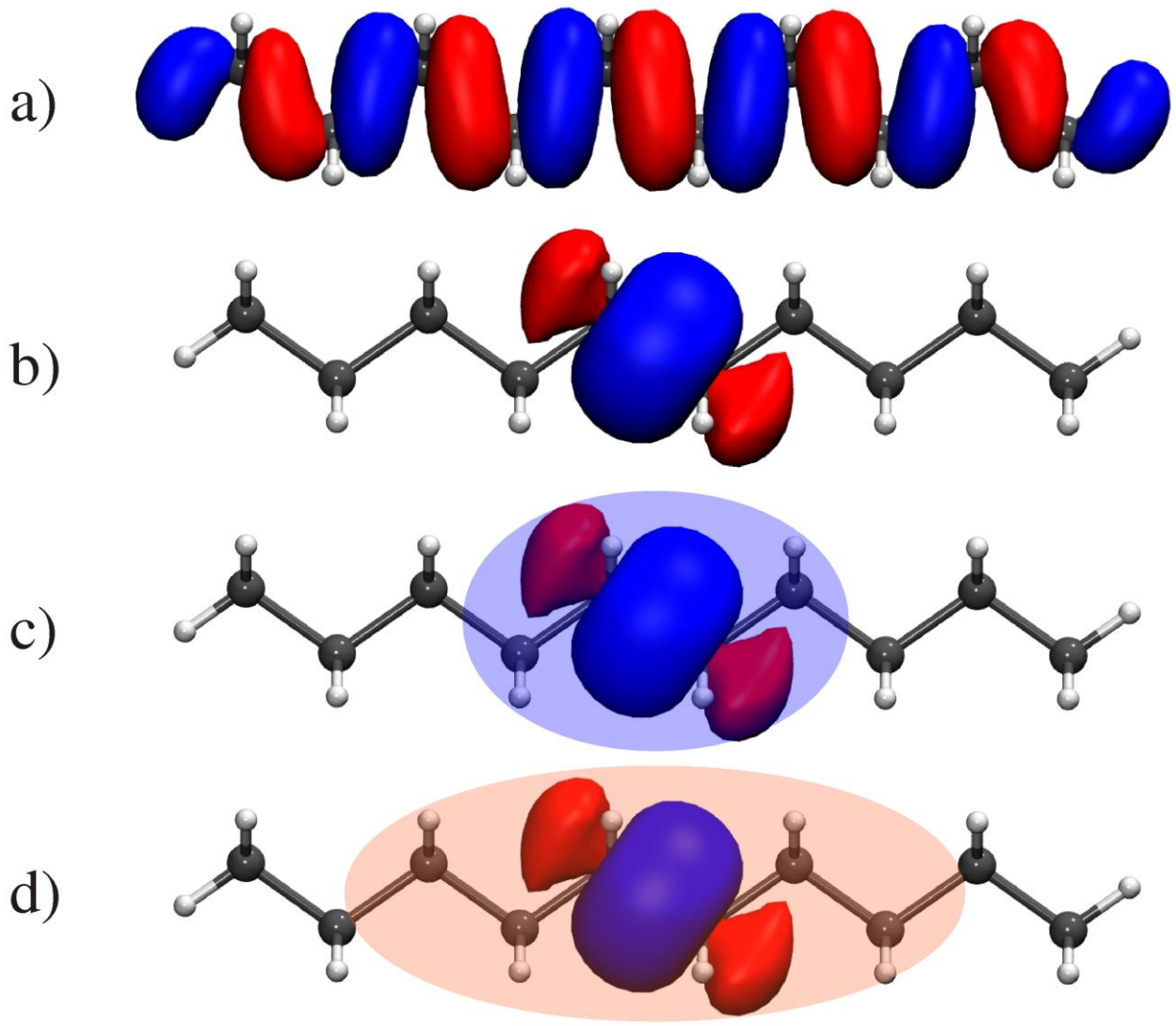


Figure 1: Canonical (a) and localized (b) molecular orbitals of n-$C_{10}H_{22}$.The atomic centers for the local auxiliary function set (c) and local basis set (d) are indicated by blue and red regions, respectively.

The recently developed variational fitting of the Fock potential [13] based on Foster-Boys localization explores this possibility by utilizing LMOs in Eq. (32). As a result, it shows an improved computational performance for the calculation of variational fitted Fock exchange energies. In this implementation, a two step localization is used. First an incomplete Cholesky

decomposition [37] is performed followed by a tighter Foster-Boys localization [38,39]. The molecular orbital localization permits a variational fitting of Fock exchange by defining fitting domains around each LMO. In particular, we define fitting domains around each localized MO in terms of atomic centers. To this end, we calculate the atomic Löwdin population for each atom A in a given LMO, $\acute{\psi}_i$, according to

$$n_{iA} = \sum_{\mu \in A} \sum_{\nu} S_{\mu\nu}^{1/2} \, c_{\nu i} \, .$$

After ordering these atomic populations from the largest to the smallest, we sum them up until a threshold value (default is 0.9995 for final energy calculation) is reached. All atoms that contribute to this sum define the atomic domain for the given LMO. The auxiliary functions on these atoms define the local auxiliary function set given by the blue region in Fig. 1c. The corresponding local basis set is defined by all basis functions of the domain atoms and augmented by basis functions from neighboring atoms with significant overlap (by default $\geq 10^{-6}$) into the domain. This is the red region in Fig. 1d. Because of this localization, each LMO has a particular Coulomb matrix. The computational cost for computing all these local $\boldsymbol{G}$ and $\boldsymbol{G^{-1}}$ matrices [40] for all occupied LMOs is in larger systems more than overcompensated by the reduced dimensionality of these matrices. Thus, the final QM energy expression used for ADFT QM/MM calculation has the form:

$$E^{QM} = \sum_{\mu,\nu} P_{\mu\nu} \, H_{\mu\nu} + \sum_{\mu,\nu} \sum_{\bar{k}} P_{\mu\nu} \, \langle \mu\nu \| \bar{k} \rangle \, x_{\bar{k}} - \frac{1}{2} \sum_{\bar{k},\bar{l}} x_{\bar{k}} \, x_{\bar{l}} \, \langle \bar{k} \| \bar{l} \rangle - \frac{\alpha}{2} \sum_{\mu,\nu} \sum_{i}^{occ} \sum_{\bar{k},\bar{l}} P_{\mu\nu} \, \langle \mu \, \acute{\psi}_i \| \bar{k} \rangle \, G_{\bar{k}\bar{l}}^{-1} \, \langle \bar{l} \, \| \, \acute{\psi}_i \, \nu \rangle + E_{xc}[\tilde{\rho}] \tag{33}$$

Note that in QM/MM calculations, the core Hamilton matrix elements in Eq. (33) are augmented by the electrostatic embedding from the MM region. They are given by:

$$H_{\mu\nu} = H_{\mu\nu}^{QM} - \sum_{D}^{MM} \langle \mu\nu | \hat{\mathcal{A}}_D(\mathbf{0}) \rangle \, Q_D \tag{34}$$

In Eq. (34) $H_{\mu\nu}^{QM}$ are the QM core matrix elements given by Eq. (3) and Q_D are the MM atomic charges. For generality of discussion, we introduce the nuclear attraction type operator $\hat{\mathcal{A}}_D$ defined as [41]:

$$\hat{\mathcal{A}}_D(\boldsymbol{k}) = \left(\frac{\partial}{\partial D_x}\right)^{k_x} \left(\frac{\partial}{\partial D_y}\right)^{k_y} \left(\frac{\partial}{\partial D_z}\right)^{k_z} \frac{1}{|\boldsymbol{r} - \boldsymbol{D}|} \tag{35}$$

This general definition allows inclusion of higher point moments on MM atoms, e.g. for polarizable force fields [42]. Work in this direction is currently under development within the deMon developers' community. However, for clarity of presentation we restrict ourselves here to MM point charges, i.e. $k_x = k_y = k_z = 0$ in Eq. (35). Although the core Hamilton matrix elements are only calculated once for each geometry at the beginning of the SCF, they can become computational demanding for larger number of MM atoms if standard nuclear attraction integral recurrence relations [43] are used. Due to the slow decay of Coulomb interactions an efficient screening of these integrals is not possible. However, it is possible to separate the QM/MM electrostatic interaction integrals in Eq. (34) into near- and far-field integrals [44]. The far-field QM/MM electrostatic interaction integrals are obtained by the following asymptotic expansion [45]:

$$\langle \alpha\beta | \hat{\mathcal{A}}_D(\mathbf{0}) \rangle^A \sim \sum_{m_x} \sum_{m_y} \sum_{m_z} \frac{(-1)^m}{m_x! \, m_y! \, m_z!} \left(\frac{\partial}{\partial D_x}\right)^{k_x} \left(\frac{\partial}{\partial D_y}\right)^{k_y} \left(\frac{\partial}{\partial D_z}\right)^{k_z} \frac{1}{|\boldsymbol{A} - \boldsymbol{D}|} \langle \alpha + m | \beta \rangle$$

The superscript A indicates the center of the asymptotic expansion. There are always two expansions, one for the center where atomic orbital α is located and the other for the center where the atomic orbital β is located. The extension of this asymptotic expansion to higher moments is

possible as shown in the double asymptotic expansion of ERIs [46]. Thus, QM/MM core Hamilton matrix elements in deMon2k are calculated as (only the asymptotic expansion with respect to center A is explicitly shown):

$$H_{\mu\nu} = H_{\mu\nu}^{QM} - \sum_{D}^{Near} \langle \mu\nu | \hat{\mathcal{A}}_D(\mathbf{0}) \rangle \, Q_D$$

$$- \sum_{m_x} \sum_{m_y} \sum_{m_z} \frac{(-1)^m}{m_x! \, m_y! \, m_z!} \langle \alpha + m | \beta \rangle \sum_{D}^{Far} T_{AD}(\boldsymbol{m}) \, Q_D \tag{36}$$

with

$$T_{AD}(\boldsymbol{m}) = \left(\frac{\partial}{\partial D_x} \right)^{m_x} \left(\frac{\partial}{\partial D_y} \right)^{m_y} \left(\frac{\partial}{\partial D_z} \right)^{m_z} \frac{1}{|\boldsymbol{A} - \boldsymbol{D}|} \tag{37}$$

In Eq. (36) the sum limits $Near$ and Far refer to MM atoms in the near- and far-field of the QM region. There are two major advantages associated to the above asymptotic expansion of the electrostatic MM embedding. First, nuclear attraction type integrals are substituted by overlap type integrals, $\langle \alpha + m | \beta \rangle$, which avoids the calculation of the incomplete gamma function. Second, the far-field loop over MM atoms is factored out and, therefore, scales only with the number of MM atoms independent from the electronic expansion of the QM region. As a result, the MM embedding in deMon2k QM/MM SCF calculations produces only a small computational overhead in the range of a few percent compared to the corresponding QM calculation [44]. The QM/MM modified core matrix elements enter the corresponding ADFT Kohn-Sham matrix according to Eq. (21) where $H_{\mu\nu}^{QM}$ is substituted by $H_{\mu\nu}$ from Eq. (34). In this way, the MM embedding alters directly the SCF.

In practical QM/MM applications the user has to choose the density functional approximation (DFA), the basis set, the auxiliary functions set and other specifications in the here outlined methodological frameworks of DF-DFT and ADFT. The DFAs available for QM/MM calculations in deMon2k are LDA, GGA and meta-GGA (only within DF-DFT) as well as global

and range-separated hybrids. By default, the LDA is used in form of Dirac exchange [47] and VWN correlation [48]. Because in ADFT LDA and GGA performance is similar, we recommend GGA functionals for QM/MM calculations. In particular, PBE [49] in combination with the GGA optimized DZVP-GGA basis and GEN-A2* auxiliary function set [33] yields reliable optimized structure parameters and fair qualitative binding energies throughout the periodic table. Note that for heavier elements scalar relativistic effects must be taken into account either by effective core potentials (ECPs) or by model core potentials (MCPs) [50]. The ECP approach can also be used for the introduction of capping potentials [51] in QM/MM calculations. We also recommend the PBE functional for QM/MM calculation because analytic ADFT kernel formulas [52] are implemented in deMon2k. This greatly facilitates QM/MM response calculations with Auxiliary Density Perturbation Theory (ADPT). For most other GGA functionals the ADFT kernels are calculated by finite-differences [53]. If van der Waals interaction are important, e.g. in QM/MM structure optimizations, the DF-DFT and ADFT LDA or GGA energy expressions must be augmented by empirical dispersion terms [54-56] because these interactions are not captured by those functionals. Although ADFT hybrid calculations show similar serial performance as corresponding GGA runs, we do not recommend them for initial QM/MM structure optimizations due to their less than ideal parallel scaling [40]. Instead, we advocate composite approaches [57] consisting of GGA structure optimization and single-point hybrid energy calculations, e.g. with PBE0 [58,59]. At this point, it is important to note that basis set convergence for hybrid functionals is much slower than for GGA functionals. To this end, we have made good experience with augmented Dunning type basis sets [60]. The handling of the three-center ERIs and the grid accuracy for the numerical integration of the exchange-correlation energy and potential are important for the computational performance of QM/MM calculations, too. By default, the direct SCF setting is used in deMon2k. In this approach the three-center ERIs are repeatedly calculated

(twice) in each SCF iteration. If sufficient random access memory (RAM) is available, the use of the so-called mixed SCF approach [61,62] is very beneficial for the computational performance. The tightening of the adaptive grid accuracy [63,64] beyond the default value of 10^{-5} a.u. will increase computational demand. Thus, care is advised. For linear algebra tasks, we recommend the use of MINRES for the density fitting and the ScaLAPACK extension of deMon2k for matrix diagonalization.

With QM energy models for QM/MM calculations at hand, we now express the QM/MM energy as:

$$E^{QM/MM} = E^{QM} + E^{MM} + E^{QMMM} + E^{NN} \tag{38}$$

In Eq. (38) E^{MM}, E^{QMMM} and E^{NN} denote the MM energy, the QMMM interaction energy and the nuclear repulsion energy for the QM/MM system, respectively. All three terms are classical and, therefore, have no dependency from the wave function and electronic density. In deMon2k the following MM energy expression is implemented:

$$E^{MM} = E^{bond} + E^{Urey} + E^{bend} + E^{tors} + E^{itors} + E^{vdW} + E^{QQ} \tag{39}$$

The first five terms correspond to bonded contributions, whereas the last two are non-bonded terms. For simplicity of discussion, we will use generic formulas for the individual terms, i.e. for bond stretching, Urey-Bradley interaction, angle bending and so on. The analytic forms of the generic bond stretching, Urey-Bradley interaction, angle bending and proper torsion are given by:

$$E^{bond} = k_b(d - d_0)^2 \tag{40}$$

$$E^{Urey} = k_l(l - l_0)^2 \tag{41}$$

$$E^{bend} = k_\vartheta(\vartheta - \vartheta_0)^2 \tag{42}$$

$$E^{tors} = \sum_n V_n[1 + \cos(n\varphi - \varphi_n)] \tag{43}$$

In the first equation, the distance is calculated as $d = |\boldsymbol{A} - \boldsymbol{B}|$. The equilibrium distance d_0 and the force constant k_b are force field parameters for the specific bond. In the Urey term (see Fig. 2), l represents the distance between two atoms, with the difference that these atoms are at the ends of an angle. The equilibrium distance l_0 and force constant k_l are given by the force field, too.

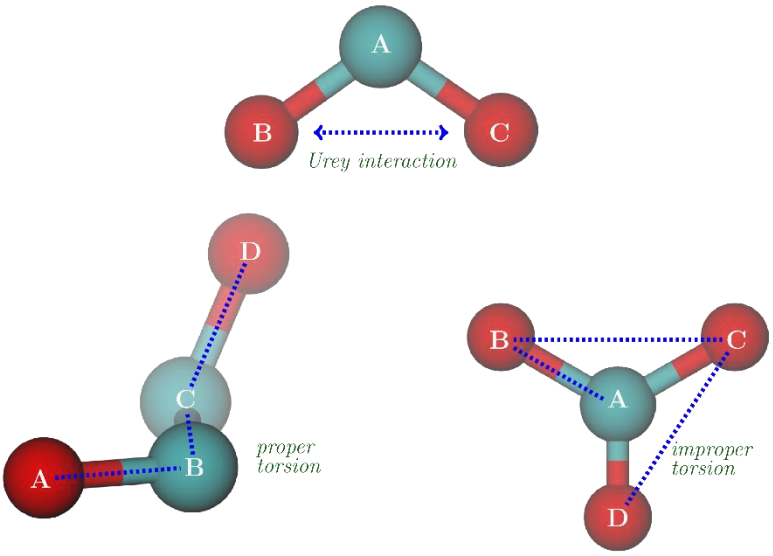


Figure 2: Urey-Bradley interaction (top) as modeled by a spring-like system formed by the B and C atoms. Bottom: Definition of a proper torsion (left) over consecutive bond connected atoms and definition of an improper torsion (right) where consecutive bond connectivity is not mandatory.

Also the angle bending in Eq. (42) is modeled by a harmonic potential. The corresponding force constant, k_ϑ, and equilibrium angle, ϑ_0, are force field parameters for the bending. Eq. (43) is used for the calculation of both proper and improper torsion as depicted in Fig. 2. The difference between proper and improper torsions is not in their mathematical formulation, but rather on how they are defined: Proper torsions are defined over atoms successively connected, such that whenever going from one atom to the following, there is always a bond between them. On the other hand, for

improper torsions this connectivity is not mandatory. This serves various purposes like restricting three atoms into a plane, e.g. to describe a sp^2 carbon binding. Fig. 2 shows, besides the Urey interactions in the top, typical torsion definitions in the bottom. Similar to the rest of the bonded terms the parameters needed for the calculations of the torsion contributions are given by the force field dataset (FFDS) file. The number of expansion coefficients, V_n, and phase factors, φ_n, in Eq. (43) is four for proper torsions, and three for the improper torsions. All here discussed bonded MM interactions require molecular connectivity information, which is either provided by the user in the deMon2k input file [65] (see also Section 7) or automatically generated based on the distances between MM atoms. The latter is particularly useful for molecular assemblies, e.g. water clusters.

For the non-bonded interactions, the energy expressions used are the well-known van der Waals (aka Lennard Jones) and point charge Coulomb potentials:

$$E^{vdW} = 4\sum_{A>B}^{MM} \epsilon_{AB}\left[\left(\frac{\sigma_{AB}}{|\boldsymbol{A}-\boldsymbol{B}|}\right)^{12} - \left(\frac{\sigma_{AB}}{|\boldsymbol{A}-\boldsymbol{B}|}\right)^{6}\right] \tag{44}$$

$$E^{QQ} = \sum_{A>B}^{MM} \frac{Q_A Q_B}{|\boldsymbol{A}-\boldsymbol{B}|} \tag{45}$$

As for the bonded terms, the atomic parameters for atoms A and B in Eqs. (44) and (45) are defined by the force field. To obtain the diatomic parameters ϵ_{AB} and σ_{AB}, a combination rule, which can be either an arithmetic or a geometric average, is used as it is shown below for the epsilon parameter:

$$\epsilon_{AB} = \frac{1}{2}(\epsilon_A + \epsilon_B)$$

$$\epsilon_{AB} = \sqrt{\epsilon_A \cdot \epsilon_B}$$

By default in deMon2k, an arithmetic mean is used to obtain σ_{AB}, whereas the geometric mean is used for the calculation of ϵ_{AB}. The force field parameters are read from the FFDS file. By default,

the FFDS file for OPLS-AA [66] is used. Alternatively, the user can select the family of AMBER force-field parameters [67], which are available in a common format with corresponding FFDS files, too.

The last two terms of Eq. (38) are the mechanical QM/MM interaction and the nuclear repulsion. Both have simple analytic formulas in terms of QM and MM atomic positions. The mechanical QM/MM interaction energy, E^{QMMM}, is modeled in deMon2k by a Lennard-Jones potential:

$$E^{QMMM} = 4 \sum_{A}^{QM} \sum_{D}^{MM} \epsilon_{AD} \left[\left(\frac{\sigma_{AD}}{|\boldsymbol{A} - \boldsymbol{D}|} \right)^{12} - \left(\frac{\sigma_{AD}}{|\boldsymbol{A} - \boldsymbol{D}|} \right)^{6} \right] \tag{46}$$

This is the same analytic expression as of Eq. (44) for the van der Waals interactions between MM atoms. The difference is that in Eq. (46) one of the atoms is a QM atom and the other is a MM atom. The diatomic parameters ϵ_{AD} and σ_{AD} are again calculated by the above described combination rules from atomic parameters. To this end, MM atomic ϵ and σ parameters are assigned to the QM atoms by the QM/MM keyword [65] in deMon2k. The last term in Eq. (38) is the electrostatic repulsion between the QM point charge nuclei with themselves and with the MM atomic point charges. It has the following form:

$$E^{NN} = \sum_{A>B}^{QM} \frac{Z_A Z_B}{|\boldsymbol{A} - \boldsymbol{B}|} + \sum_{A}^{QM} \sum_{D}^{MM} \frac{Z_A Q_D}{|\boldsymbol{A} - \boldsymbol{D}|} \tag{47}$$

The here appearing MM atomic point charges, Q_D, are again taken from the FFDS file. Fig. 3 depicts in the top the energy modules of deMon2k that can contribute to QM/MM calculations. Once the QM/MM energy is calculated, which implies a converged SCF solution for the QM part, gradient components are calculated by differentiation of Eq. (38) with respect to atomic coordinates $A_\lambda, \lambda = x, y, z$:

$$\frac{\partial E^{QM/MM}}{\partial A_\lambda} = \frac{\partial E^{QM}}{\partial A_\lambda} + \frac{\partial E^{MM}}{\partial A_\lambda} + \frac{\partial E^{QMMM}}{\partial A_\lambda} + \frac{\partial E^{NN}}{\partial A_\lambda} \tag{48}$$

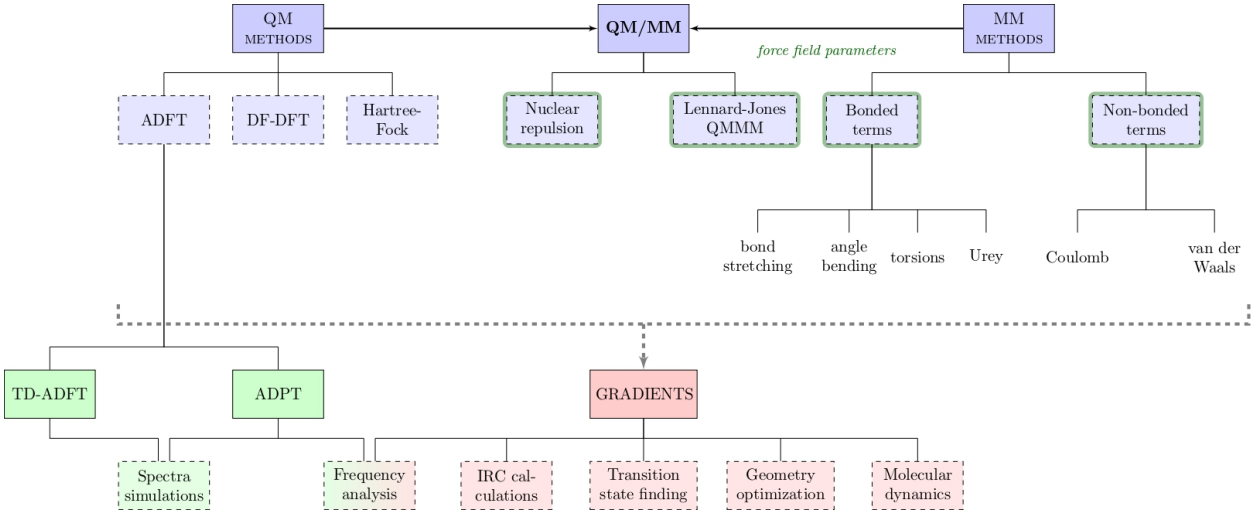


Figure 3: Diagram of deMon2k energy modules for QM/MM calculations and their relation to gradient and property calculations (see also Section 5 and 6).

As Eq. (48) shows, all energy terms of a QM/MM calculation contribute to the corresponding gradient. From the gradient on, QM and MM atoms are treated on equal footing in deMon2k. Thus, the following utility modules for structure optimization and chemical reaction modeling (see Section 3) as well as for molecular dynamics simulation (see Section 4) are the same for QM, MM and QM/MM calculations. On the other hand, the analytic higher order SCF energy derivatives calculated in the TD-ADFT and ADPT modules of Fig. 3 are directly linked to the ADFT module and, therefore, are only available for the QM part of a QM/MM calculation (see Section 6). The QM/MM frequency analysis is linked to the ADFT as well as MM module as outlined in Section 5.

The first term in Eq. (48) is the generic QM energy derivative. For ADFT [15,22,35,36] QM/MM calculations this term is obtained by differentiation of Eq. (33) with respect to a QM atomic coordinate:

$$\frac{\partial E^{QM}}{\partial A_\lambda} = -\sum_{\mu,\nu} W_{\mu\nu}\, S_{\mu\nu}^{(A_\lambda)} + \sum_{\mu,\nu} P_{\mu\nu}\, H_{\mu\nu}^{(A_\lambda)} + \sum_{\mu,\nu}\sum_{\bar{k}} P_{\mu\nu}\, \langle \mu\nu \| \bar{k} \rangle^{(A_\lambda)} (x_{\bar{k}} + z_{\bar{k}}) +$$

$$\sum_{\bar{k},\bar{l}} x_{\bar{k}}\, G_{\bar{k}\bar{l}}^{(A_\lambda)} \left(\frac{1}{2}\, x_{\bar{l}} + z_{\bar{l}}\right) + \sum_{\bar{k}} x_{\bar{k}}\, \langle \bar{k}^{(A_\lambda)} | v_{xc}[\tilde{\rho}] \rangle + \alpha \sum_{\bar{k},\bar{l}} \Gamma_{\bar{k}\bar{l}}\, G_{\bar{k}\bar{l}}^{(A_\lambda)} -$$

$$\alpha \sum_{i}^{occ} \sum_{\bar{k}} \sum_{\sigma} \sum_{\mu,\nu} P_{\mu\nu}\, x_{\bar{k}\mu i}\, \langle \bar{l} \| \nu\sigma \rangle^{(A_\lambda)} c_{\sigma i} \tag{49}$$

The superscripts denote differentiation of molecular integrals or functions with respect to the QM atomic coordinate. Eq. (49) consists of three contributions. The first one is the Pulay force [68] given by the first term of Eq. (49). It collects all derivatives of the MO coefficients. Thus, no response calculation is needed for the gradient evaluation as expected from Wigner's $2n + 1$ rule [69,70]. The second contribution arises from the QM/MM core (second term in Eq. 49), Coulomb (third and fourth term in Eq. 49), and exchange-correlation (fifth term in Eq. 49), derivatives. Together with the Pulay forces, these contributions define the QM ADFT energy derivatives for LDA and GGA. The last two terms in Eq. (49) define the third contribution given by the derivative of the three-center ERI Fock exchange energy. The closed-shell energy weighted density matrix elements appearing in the Pulay term are given by:

$$W_{\mu\nu} = 2 \sum_{i}^{occ} \varepsilon_i\, c_{\mu i}\, c_{\nu i} \tag{50}$$

Here ε_i refers to the canonical MO energies as obtained from the converged SCF. The $\Gamma_{\bar{k}\bar{l}}$ and $x_{\bar{k}\mu i}$ coefficients are calculated as:

$$\Gamma_{\bar{k}\bar{l}} = \frac{1}{2}\sum_{i}^{occ}\sum_{\mu,\nu} x_{\bar{k}\mu i}\, P_{\mu\nu}\, x_{\bar{l}\nu i} \tag{51}$$

$$x_{\bar{k}\mu i} = \sum_{\bar{l}} G^{-1}_{\bar{k}\bar{l}}\ \langle \bar{l} \,\|\, \mu\, \dot{\psi}_i \rangle \tag{52}$$

The QM energy inside the QM/MM energy expression, Eq. (38), depends, through the core matrix elements defined in Eq. (34), from the atomic coordinates of the MM atoms, too. Therefore, the corresponding gradient component for a MM atom coordinate D_λ is given by:

$$\frac{\partial E^{QM}}{\partial D_\lambda} = -\sum_{\mu,\nu}\sum_{D}^{MM} \langle \mu\nu \big| \hat{\mathcal{A}}_D(1_\lambda) \rangle\, Q_D \tag{53}$$

We note that for these derivative integrals (electric field integrals) asymptotic expansions [31] analog to the core matrix elements are available in deMon2k, too. The DF-DFT gradients for QM/MM calculations differ from the here derived ADFT gradients in the term for the exchange-correlation in Eq. (49).

The second term in Eq. (48) is the MM energy derivative. It only depends from the atomic coordinates of the MM atoms. As the MM energy, the MM derivatives can be partitioned into bonded and non-bonded derivatives. For the derivatives of the generic bond stretching and Urey-Bradley terms with respect to the MM atomic coordinate A_λ follows:

$$\frac{\partial E^{bond}}{\partial A_\lambda} = 2k_b(d_0 - d)\frac{A_\lambda - B_\lambda}{d} \tag{54}$$

$$\frac{\partial E^{Urey}}{\partial A_\lambda} = 2k_l(l - l_0)\frac{A_\lambda - B_\lambda}{l} \tag{55}$$

From the corresponding differentiation of the angle bending energy, we find:

$$\frac{\partial E^{bend}}{\partial A_\lambda} = 2k_\vartheta(\vartheta - \vartheta_0)\ \frac{\partial \vartheta}{\partial A_\lambda} \tag{56}$$

$$\frac{\partial \vartheta}{\partial A_\lambda} = \frac{\cos\vartheta}{\sin\vartheta}\frac{(A_\lambda - B_\lambda)}{|\boldsymbol{A}-\boldsymbol{B}|^2} - \frac{(C_\lambda - B_\lambda)}{|\boldsymbol{A}-\boldsymbol{B}||\boldsymbol{C}-\boldsymbol{B}|\sin\vartheta} \tag{57}$$

$$\frac{\partial \vartheta}{\partial C_\lambda} = \frac{\cos\vartheta}{\sin\vartheta}\frac{(C_\lambda - B_\lambda)}{|\boldsymbol{C}-\boldsymbol{B}|^2} - \frac{(A_\lambda - B_\lambda)}{|\boldsymbol{A}-\boldsymbol{B}||\boldsymbol{C}-\boldsymbol{B}|\sin\vartheta} \tag{58}$$

Here A_λ and C_λ refer to atomic coordinates of the end atoms of the $A-B-C$ angle. The angle derivative for the central B atom is calculated from the translation invariance as:

$$\frac{\partial \vartheta}{\partial B_\lambda} = -\frac{\partial \vartheta}{\partial A_\lambda} - \frac{\partial \vartheta}{\partial C_\lambda} \tag{59}$$

Due to their definition, both proper and improper torsions share the same differentiation formulas, and thus we write the equations referring to the proper torsions only. We start by writing the gradient of Eq. (43) in terms of the derivative of the dihedral angle φ:

$$\frac{\partial E^{tors}}{\partial A_\lambda} = -\sum_n n\, V_n \sin(n\varphi - \varphi_n)\frac{\partial \varphi}{\partial A_\lambda} \tag{60}$$

As for the case of the bending energy gradients, the derivatives of the dihedral angle depend on the relative position of the atom in the dihedral. The four different cases are:

$$\frac{\partial \varphi}{\partial A_\lambda} = -\frac{P_\lambda}{|\boldsymbol{A}-\boldsymbol{B}|\sin^2\vartheta_{ABC}} \tag{61}$$

$$\frac{\partial \varphi}{\partial D_\lambda} = -\frac{Q_\lambda}{|\boldsymbol{C}-\boldsymbol{D}|\sin^2\vartheta_{BCD}} \tag{62}$$

$$\frac{\partial \varphi}{\partial B_\lambda} = \left(\frac{|\boldsymbol{A}-\boldsymbol{B}|\cos\vartheta_{ABC}}{|\boldsymbol{B}-\boldsymbol{C}|} - 1\right)\frac{\partial \varphi}{\partial A_\lambda} - \left(\frac{|\boldsymbol{C}-\boldsymbol{D}|\cos\vartheta_{BCD}}{|\boldsymbol{B}-\boldsymbol{C}|}\right)\frac{\partial \varphi}{\partial D_\lambda} \tag{63}$$

$$\frac{\partial \varphi}{\partial C_\lambda} = \left(\frac{|\boldsymbol{C}-\boldsymbol{D}|\cos\vartheta_{BCD}}{|\boldsymbol{B}-\boldsymbol{C}|} - 1\right)\frac{\partial \varphi}{\partial D_\lambda} - \left(\frac{|\boldsymbol{A}-\boldsymbol{B}|\cos\vartheta_{ABC}}{|\boldsymbol{B}-\boldsymbol{C}|}\right)\frac{\partial \varphi}{\partial A_\lambda} \tag{64}$$

The vectors $\boldsymbol{P}$ and $\boldsymbol{Q}$ are perpendicular to the planes (ABC and BCD, respectively) that define the dihedral angle φ. They are given by the following cross products:

$$\boldsymbol{P} = (\boldsymbol{B} - \boldsymbol{A}) \times (\boldsymbol{C} - \boldsymbol{B})$$

$$\boldsymbol{Q} = (\boldsymbol{C} - \boldsymbol{D}) \times (\boldsymbol{B} - \boldsymbol{C})$$

Finally, the energy derivatives of the non-bonded MM interactions are comprised by the gradients of the van der Waals and Coulomb potentials. These are given by:

$$\frac{\partial E^{vdW}}{\partial A_\lambda} = -48 \sum_{B \neq A}^{MM} \epsilon_{AB} \left[\left(\frac{\sigma_{AB}}{|\boldsymbol{A} - \boldsymbol{B}|} \right)^{12} - \frac{1}{2} \left(\frac{\sigma_{AB}}{|\boldsymbol{A} - \boldsymbol{B}|} \right)^{6} \right] \frac{A_\lambda - B_\lambda}{|\boldsymbol{A} - \boldsymbol{B}|^2} \tag{65}$$

$$\frac{\partial E^{QQ}}{\partial A_\lambda} = - \sum_{B \neq A}^{MM} \frac{A_\lambda - B_\lambda}{|\boldsymbol{A} - \boldsymbol{B}|^3} \, Q_A Q_B \tag{66}$$

For the QM/MM mechanical interaction, we find similar results to Eq. (65) since the same model is used for both the E^{QMMM} and the E^{vdW} energy terms. Nonetheless, for E^{QMMM} we have derivatives with respect to both QM and MM atomic coordinates:

$$\frac{\partial E^{QMMM}}{\partial A_\lambda} = -48 \sum_{D}^{MM} \epsilon_{AD} \left[\left(\frac{\sigma_{AD}}{|\boldsymbol{A} - \boldsymbol{D}|} \right)^{12} - \frac{1}{2} \left(\frac{\sigma_{AD}}{|\boldsymbol{A} - \boldsymbol{D}|} \right)^{6} \right] \frac{A_\lambda - D_\lambda}{|\boldsymbol{A} - \boldsymbol{D}|^2} \tag{67}$$

$$\frac{\partial E^{QMMM}}{\partial D_\lambda} = -48 \sum_{A}^{QM} \epsilon_{AD} \left[\left(\frac{\sigma_{AD}}{|\boldsymbol{A} - \boldsymbol{D}|} \right)^{12} - \frac{1}{2} \left(\frac{\sigma_{AD}}{|\boldsymbol{A} - \boldsymbol{D}|} \right)^{6} \right] \frac{A_\lambda - D_\lambda}{|\boldsymbol{A} - \boldsymbol{D}|^2} \tag{68}$$

In these gradients, the QM and MM atomic coordinates are given by A_λ and D_λ, respectively. The same convention is used for the gradients of the repulsion between the QM point charge nuclei with themselves and with the MM atomic point charges, Eq. (47). These gradients are given by:

$$\frac{\partial E^{NN}}{\partial A_\lambda} = - \sum_{A \neq B}^{QM} \frac{A_\lambda - B_\lambda}{|\boldsymbol{A} - \boldsymbol{B}|^3} \, Z_A Z_B - \sum_{D}^{MM} \frac{A_\lambda - D_\lambda}{|\boldsymbol{A} - \boldsymbol{D}|^3} Z_A Q_D \tag{69}$$

$$\frac{\partial E^{NN}}{\partial D_\lambda} = \sum_{A}^{QM} \frac{A_\lambda - D_\lambda}{|\boldsymbol{A} - \boldsymbol{D}|^3} \, Z_A Q_D \tag{70}$$

With these final equations, all terms needed for the calculation of the QM/MM energy gradients are at hand for the different methodologies implemented in deMon2k, as depicted in Fig. 3.

3. Local structure minimization and transition state search

The Potential Energy Surface (PES) of molecular systems is a multidimensional function arising from the Born-Oppenheimer approximation [71] that possesses different kinds of stationary points. At these points, the first energy derivatives with respect to the nuclear coordinates, i.e. the energy gradients, are zero. For a molecule with *M* atoms the dimension of the PES is given by its internal degrees of freedom, which are either $3\,M-5$ or $3\,M-6$ depending if the molecule is linear or not, respectively. The different types of stationary points are indicated on the PES model illustrated in Figure 4.

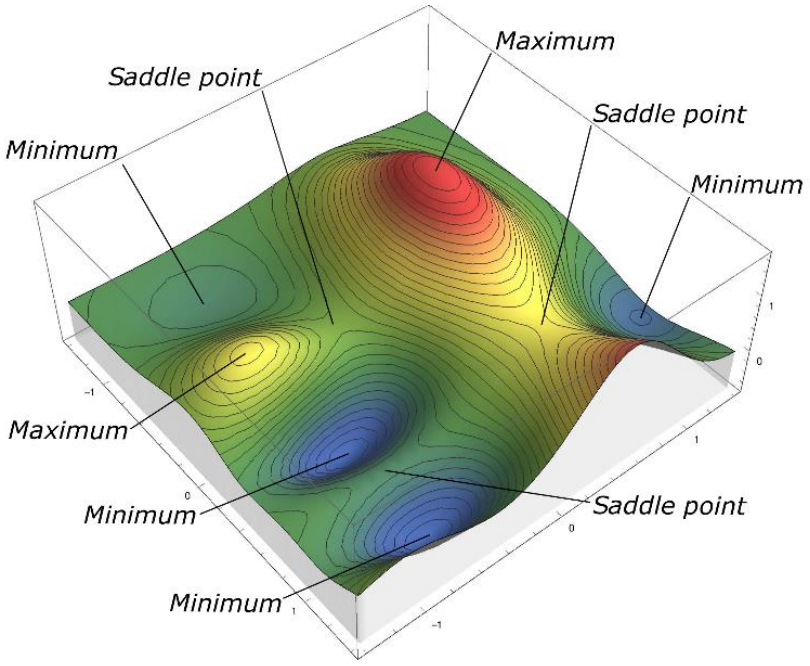


Figure 4: Stationary points on a PES model.

As shown in this figure, different stationary points such as minima, maxima and saddle points can exist on a PES. The number of stationary points of the PES increases exponentially as the number of atoms increases. Fortunately, from a chemical point of view only minima and first order saddle

points are of interest because they correspond to stable minimum structures and transition states, respectively. Nevertheless, the exploration of the full PES remains a formidable task for many molecular systems.

On the other hand, the PES can be efficiently explored in local regions employing numerical optimization techniques. For quantum chemical structure minimization, quasi-Newton methods [72-84] are commonly applied. They are particular efficient for QM methods because only energy and gradient information are needed. Because in deMon2k QM/MM calculations the QM and MM gradients are joined together (see Fig. 3) a quasi-Newton trust region method (TRM) [72,73] is employed for local QM/MM structure optimizations by default. To this end, the PES is expanded around a reference molecular structure by a second-order Taylor series. This quadratic PES expansion is given by:

$$q(\boldsymbol{p}) = E_0 + \mathbf{g}^T \boldsymbol{p} + \frac{1}{2} \boldsymbol{p}^T \boldsymbol{B} \boldsymbol{p} \tag{71}$$

In Eq. (71) **g** is the QM/MM gradient vector and $\boldsymbol{B}$ the corresponding approximation of the Hessian matrix. Both quantities are evaluated at the reference structure with energy E_0. In order to find a minimum the quadratic expansion is optimized with respect to the step direction $\boldsymbol{p}$. To ensure that the model function $q(\boldsymbol{p})$ is a good approximation of the PES, we restrict the step size within a trust region defined by a hypersurface with radius h around the reference structure. Thus, the following constrained minimization problem is obtained:

$$\min_{\boldsymbol{p}} q(\boldsymbol{p}) \quad subject\ to \quad |\boldsymbol{p}| \leq h \tag{72}$$

For the solution of Eq. (72) the following Lagrange function is introduced [85]:

$$L(\boldsymbol{p}, \lambda) = E_0 + \mathbf{g}^T \boldsymbol{p} + \frac{1}{2} \boldsymbol{p}^T \boldsymbol{B} \boldsymbol{p} + \frac{1}{2} \lambda (\boldsymbol{p}^T \boldsymbol{p} - h^2) \tag{73}$$

From the stationary condition of this Lagrange function

$$\frac{\partial L(\boldsymbol{p},\lambda)}{\partial p_A} = 0 \qquad \forall\, A \;\in \mathrm{QM/MM} \tag{74}$$

the step direction $\boldsymbol{p}$ is defined as:

$$\boldsymbol{p} = -(\boldsymbol{B} + \lambda \boldsymbol{E})^{-1}\,\mathbf{g} \tag{75}$$

For the determination of the Lagrange multiplier λ that multiplies the identity matrix in Eq. (75) standard techniques from the literature [73,85-87] are used. Similar, for the update of the (approximated) Hessian matrix the BFGS update formula, named after its inventors Broyden [88], Fletcher [89], Goldfarb [90] and Shanno [91], is used:

$$\boldsymbol{B}^{I+1} = \boldsymbol{B}^{I} - \frac{\boldsymbol{B}^{I}\boldsymbol{\delta}\widetilde{\boldsymbol{\delta}}\boldsymbol{B}^{I}}{\widetilde{\boldsymbol{\delta}}\boldsymbol{B}^{I}\boldsymbol{\delta}} + \frac{\boldsymbol{\gamma}\widetilde{\boldsymbol{\gamma}}}{\widetilde{\boldsymbol{\gamma}}\boldsymbol{\delta}} \tag{76}$$

In Eq. (76) $\boldsymbol{\gamma} = \boldsymbol{g}^{I+1} - \boldsymbol{g}^{I}$ and $\boldsymbol{\delta} = \boldsymbol{x}^{I+1} - \boldsymbol{x}^{I}$ holds. The superscripts I and $I+1$ refer to optimization steps. Thus, QM/MM structure optimizations in deMon2k follow the same path as pure QM structure optimizations. The advantage is the robustness of the underlying algorithms. The disadvantage of the TRM in particular is the need for the diagonalization of the (approximated) Hessian matrix. This limits its applicability to QM/MM system sizes with up to around 10,000 atoms. For larger systems, alternative step methods free of second-order energy derivatives such as steepest descent (see keyword STEPTYPE [65]) are available in deMon2k, too.

Besides the excellent performance of the TRM for molecular structure optimizations, it is also well suited for the search of transition states on the PES. In general, a transition state search is more complicate than the just discussed structure minimization due to the (first-order) saddle point nature (see Fig. 4) of transition states as depicted in Fig. 5.

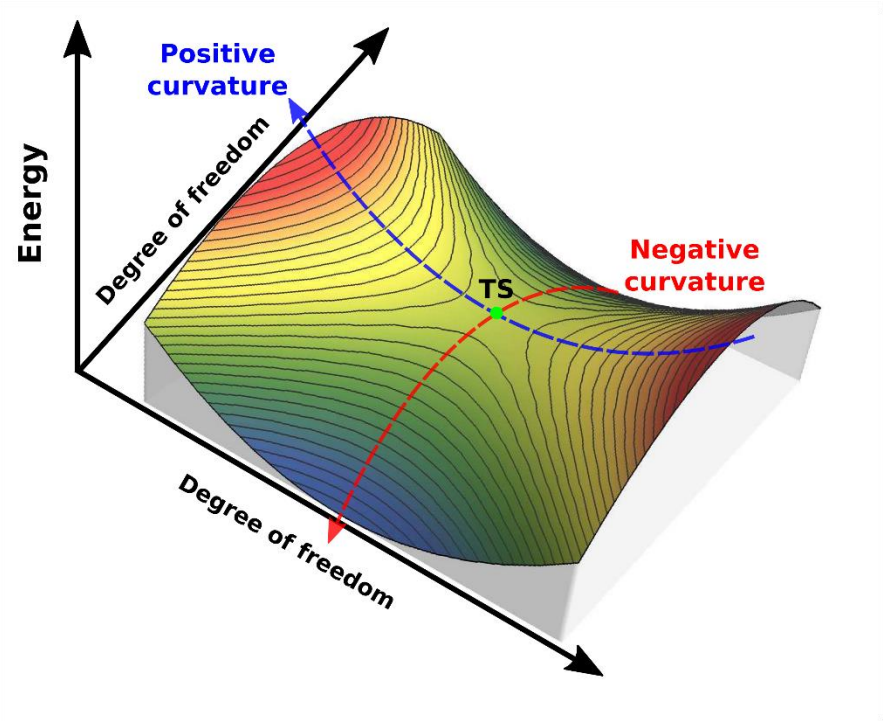


Figure 5: Transition state (TS) nature on the model PES.

However, combining the TRM with eigenvector following in the so-called uphill trust region method [92-94] yields a robust local transition state search algorithm. The corresponding step formula is given by:

$$\boldsymbol{p} = -(\boldsymbol{B} + \lambda \boldsymbol{M})^{-1}\,\mathbf{g} \tag{77}$$

In Eq. (77) $\boldsymbol{M}$ is the so-called mode matrix defined as:

$$\boldsymbol{M} = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \tag{78}$$

The position of the negative entry in the mode matrix $\boldsymbol{M}$ defines the Hessian matrix mode, which is followed uphill. From a computational point of view, it is desirable that the initial Hessian matrix for the transition state search has the right eigenvalue spectrum, i.e. one negative eigenvalue and all others positive. This can be achieved by an appropriate start structure and the explicit calculation

of the initial Hessian matrix. To this end, the hierarchical transition state search algorithm [94] is currently extended for QM/MM calculations in one of our laboratories. In transition state searches the unbiased Powell-symmetric-Broyden [95] update formula is used in deMon2k by default:

$$\boldsymbol{B}^{I+1} = \boldsymbol{B}^{I} + \frac{(\boldsymbol{\gamma} - \boldsymbol{B}^{I}\boldsymbol{\delta})\widetilde{\boldsymbol{\delta}} + \boldsymbol{\delta}(\boldsymbol{\gamma} - \boldsymbol{B}^{I}\boldsymbol{\delta})^{\boldsymbol{T}}}{\widetilde{\boldsymbol{\delta}}\boldsymbol{\delta}} - \frac{\left[\widetilde{\boldsymbol{\delta}}(\boldsymbol{\gamma} - \boldsymbol{B}^{I}\boldsymbol{\delta})\right]\boldsymbol{\delta}\widetilde{\boldsymbol{\delta}}}{\left(\widetilde{\boldsymbol{\delta}}\boldsymbol{\delta}\right)^{2}} \tag{79}$$

4. Molecular dynamics

Whereas the QM/MM molecular structure optimization in deMon2k utilizes algorithms from QM calculations, the QM/MM molecular dynamics simulations in deMon2k uses MM inspired algorithms. To this end, we assume classical equations of motion for the QM nuclei and MM atoms and introduce the concepts of ensembles for closed systems. Based on this approximation microcanonical (NVE) and canonical (NVT) QM/MM molecular dynamics (MD) simulations can be performed with deMon2k. The implementation of isothermal-isobaric (NPT) QM/MM MD simulations is currently under development in the deMon developers' community.

For the conservation of temperature and pressure during MD simulations, it is necessary to include thermostats and barostats that can regulate these properties. For this purpose, the reformulation of the equations of motion in terms of the Liouville operator [96] simplifies the application and implementation of the propagation equations. We start our discussion by writing the Newton equations for a system of M particles consisting of QM nuclei and MM atoms:

$$\dot{\boldsymbol{R}}_A = \boldsymbol{P}_A / m_A \tag{80}$$

$$\dot{\boldsymbol{P}}_A = \boldsymbol{F}_A \tag{81}$$

In these equations, the index A addresses QM nuclei and MM atoms with masses m_A. Their positions and momenta are labeled by the symbols $\boldsymbol{R}_A$ and $\boldsymbol{P}_A$. Dots indicate corresponding time derivatives. We now rewrite Eq. (80) and (81) as:

$$\dot{\boldsymbol{R}}_A = iL\boldsymbol{R}_A \tag{82}$$

$$\dot{\boldsymbol{P}}_A = iL\boldsymbol{P}_A \tag{83}$$

The Liouville operator used in these expressions is given by:

$$iL = \sum_A^M \left(\dot{\boldsymbol{R}}_A \frac{\partial}{\partial \boldsymbol{R}_A} + \dot{\boldsymbol{P}}_A \frac{\partial}{\partial \boldsymbol{P}_A} \right) \tag{84}$$

A more compact and general form of this operator is found if we collect the positions and momenta of all particles in the system into a single variable $\boldsymbol{x} = (\boldsymbol{R}^M, \boldsymbol{P}^M)$, known as the phase-space vector. With this vector, Eq. (84) can be written as $iL = \dot{\boldsymbol{x}} \cdot \nabla_{\boldsymbol{x}}$, where $\nabla_{\boldsymbol{x}}$ collects the differentiation operators from all variables in $\boldsymbol{x}$. Therefore, the evolution in time of a system is given by:

$$\boldsymbol{x}(t) = e^{iLt}\boldsymbol{x}(0) \tag{85}$$

The exponential term in this equation is known as the classical propagator. Its form depends on the equations governing the progress of the system and, as such, the analytic solution of Eq. (85) becomes very difficult as the system grows in complexity. Therefore, a numerical scheme is used for its implementation, namely by the Trotter factorization [97] of the classical propagator:

$$e^{iL\Delta t} = e^{(iL_1 + iL_2)\Delta t} \approx e^{iL_2\Delta t/2} e^{iL_1\Delta t} e^{iL_2\Delta t/2} \tag{86}$$

Here, the operator of Eq. (84) has been arbitrarily separated in two terms, iL_1 and iL_2, which propagate, individually, positions and momenta. Their forms are:

$$iL_1 = \sum_A^M \dot{\boldsymbol{R}}_A \frac{\partial}{\partial \boldsymbol{R}_A} \tag{87}$$

$$iL_2 = \sum_A^M \dot{\boldsymbol{P}}_A \frac{\partial}{\partial \boldsymbol{P}_A} \tag{88}$$

As a result, the propagator for NVE simulations reads:

$$e^{iL\Delta t} = \prod_A^M \exp\left(\frac{\Delta t}{2m_A} \boldsymbol{F}_A \frac{\partial}{\partial \dot{\boldsymbol{R}}_A}\right) \prod_A^M \exp\left(\Delta t \dot{\boldsymbol{R}}_A \frac{\partial}{\partial \boldsymbol{R}_A}\right) \prod_A^M \exp\left(\frac{\Delta t}{2m_A} \boldsymbol{F}_A \frac{\partial}{\partial \dot{\boldsymbol{R}}_A}\right) \tag{89}$$

The application of this operator to the phase-space vector of a system results in the well-known velocity Verlet algorithm [98]. The advantage of using this formulation resides on the direct translation between the resulting working equations and the programming statements.

For the generation of the NVT ensemble, a variety of different thermostats [99] are available in deMon2k. All of them work by controlling the kinetic energy fluctuations of the system by changing the particle velocities. The extended phase-space approach of the Nosé-Hoover Chain (NHC) [100,101] thermostat, by which additional variables are included into the simulation, provides a methodology to simulate the canonical ensemble using the following equations of motion:

$$\dot{\boldsymbol{R}}_A = \boldsymbol{P}_A / m_A \tag{90}$$

$$\dot{\boldsymbol{P}}_A = \boldsymbol{F}_A - \frac{P_{\eta_1}}{Q_1} \boldsymbol{P}_A \tag{91}$$

$$\dot{\eta}_k = P_{\eta_k} / Q_k \tag{92}$$

$$\dot{P}_{\eta_k} = G_k - \frac{P_{\eta_{k+1}}}{Q_{k+1}} P_{\eta_k} \tag{93}$$

$$\dot{P}_{\eta_N} = G_N \tag{94}$$

In these equations the k index loops over the N thermostats, each having a *position* η_k with no units, a *mass* Q_k with units of energy times time squared, and *momentum* P_{η_k} in units of energy times time. Only the first link of the thermostat chain couples to the system as indicated by Eq. (91). The $k \geq 2$ thermostats will regulate the variables of the previous link in the chain. The forces driving the evolution of the thermostats' dynamical variables are given by:

$$G_1 = \sum_A^M m_A \dot{\boldsymbol{R}}_A - 3Mk_BT \tag{95}$$

$$G_k = \frac{P_{\eta_{k-1}}^2}{Q_{k-1}} - k_BT \tag{96}$$

Here, k_B is the Boltzmann constant and T the desired macroscopic temperature. The resulting propagator builds upon that of Eq. (89), with the difference that the new factorization will include a third term iL_T accounting for the presence of the thermostat variables η_k and P_{η_k}. This third term has the form:

$$iL_T = \sum_{k=1}^{N}\left[\frac{P_{\eta_k}}{Q_k}\frac{\partial}{\partial \eta_k} + G_k\frac{\partial}{\partial P_{\eta_k}}\right] - \frac{P_{\eta_1}}{Q_1}\sum_A^M \boldsymbol{P}_A \cdot \frac{\partial}{\partial \boldsymbol{P}_A} - \sum_{k=1}^{N-1}\frac{P_{\eta_{k+1}}}{Q_{k+1}}P_{\eta_k}\frac{\partial}{\partial P_{\eta_k}} \tag{97}$$

The corresponding propagator in its factorized form is then given by:

$$e^{iL\Delta t} = e^{(iL_1+iL_2+iL_T)\Delta t} = e^{iL_T\Delta t/2}\left[e^{iL_2\Delta t/2}e^{iL_1\Delta t}e^{iL_2\Delta t/2}\right]e^{iL_T\Delta t/2} \tag{98}$$

This is essentially the same propagator as for the NVE case, the only difference being the update of the thermostat variables before and after the application of the velocity Verlet algorithm. The NHC algorithm within the formalism of the Liouville operator is implemented in the current development version of the deMon2k software.

Within the NVE and NVT ensembles dynamics of solute-solvent systems can effectively be performed using QM/MM, capped by a continuum solvent model (PCM). The PCM model is based on the Onsager approximation, implemented in the self-consistent reaction field DFT method [67,102-105]. Within the QM/MM scheme, the reaction field energy reads:

$$E^{RF} = -\frac{1}{2}\, g \left(\sum_{A}^{QM} \sum_{D}^{MM} Z_A \boldsymbol{A} + Q_D \boldsymbol{D} \right)^2 + g \sum_{A}^{QM} \sum_{D}^{MM} (Z_A \boldsymbol{A} + Q_D \boldsymbol{D}) \sum_{\mu,\nu} P_{\mu\nu} \langle \mu | \boldsymbol{r} | \nu \rangle - \frac{1}{2} g \sum_{\mu,\nu} \sum_{\sigma,\tau} P_{\mu\nu} P_{\sigma\tau} \langle \mu | \boldsymbol{r} | \nu \rangle \langle \sigma | \boldsymbol{r} | \tau \rangle \tag{99}$$

In Eq. (99) g is the response function of the dielectric solvent medium with respect to the potential of the QM/MM system. For a spherical cavity with radius a, and the solvent medium with a dielectric constant ε, the response function is:

$$g = \frac{1}{a^3} \frac{2\,(\varepsilon - 1)}{(2\varepsilon + 1)} \tag{100}$$

The reaction-field energy is added to the QM/MM energy in Eq. (38) and the gradient of the reaction-field with respect to the atomic coordinates $\frac{\partial E^{RF}}{\partial A_\lambda}$is added to the QM/MM energy gradient in Eq. (48).

For QM/MM simulations at constant temperature and pressure, the Martyna, Tobias, Klein (MTK) [106,107] barostat is currently extended for QM/MM applications in deMon2k. In this algorithm, the volume of the system is introduced as a dynamical variable together with the corresponding momentum. When used in combination with the NHC thermostat, the isobaric-isothermic NPT ensemble can be sampled. The equations of motion for the particle system and barostat are:

$$\dot{\boldsymbol{R}}_A = \frac{\boldsymbol{P}_A}{m_A} + \frac{P_\epsilon}{W}\boldsymbol{R}_A \tag{101}$$

$$\dot{\boldsymbol{P}}_A = \boldsymbol{F}_A - \frac{P_{\eta_1}}{Q_1}\boldsymbol{P}_A - \left(1 + \frac{1}{M}\right)\frac{P_\epsilon}{W}\boldsymbol{P}_A \tag{102}$$

$$\dot{V} = \frac{3V}{W}P_\epsilon \tag{103}$$

$$\dot{P}_\epsilon = G_\epsilon - \frac{P_{\xi_1}}{Q_1'}P_\epsilon \tag{104}$$

The terms labeled with η belong to the thermostat coupled to the particles, whereas those labeled with ξ correspond to the thermostat coupled to the barostat, whose variables are labeled with ϵ. Eqs. (92) to (94) describe the dynamics of both thermostats, although the ξ thermostat is subject to the forces:

$$G_1' = \frac{P_\epsilon^2}{W} - k_B T \tag{105}$$

$$G_k' = \frac{P_{\xi_{k-1}}^2}{Q_{k-1}'} - k_B T \tag{106}$$

The force driving the barostat, G_ϵ, is given by:

$$G_\epsilon = \left(1 + \frac{1}{M}\right)\sum_A^M \frac{\boldsymbol{P}_A^2}{m_A} + \sum_A^M \boldsymbol{R}_A \cdot \boldsymbol{F}_A - 3VP_{ext} \tag{107}$$

The external pressure, P_{ext}, is the target value for the simulation. The corresponding propagator resembles that of Eq. (98) with the addition of the terms from the barostat and the second thermostat. However, the presence of the barostat changes how positions and velocities of the particles are updated within the (modified) velocity Verlet algorithm. It now reads:

$$\boldsymbol{R}_A(t + \Delta t) = \boldsymbol{R}_A(t)e^{v_\epsilon \Delta t} + \Delta t \dot{\boldsymbol{R}}_A(t)e^{v_\epsilon \Delta t/2}\frac{\sinh(v_\epsilon \Delta t/2)}{v_\epsilon \Delta t/2} \tag{108}$$

$$\dot{R}_A(t+\Delta t/2) = \dot{R}_A(t)e^{-\frac{\alpha v_\epsilon \Delta t}{2}} + \frac{\Delta t}{2m_A}F_A e^{-\alpha v_\epsilon \Delta t/4}\frac{\sinh(\alpha v_\epsilon \Delta t/4)}{\alpha v_\epsilon \Delta t/4} \quad (109)$$

Here $\alpha = 1 + 1/M$. The changes in the cell containing the system are considered isotropic and retrieved from the instantaneous volume.

From this discussion, it is obvious that a volume definition is needed for the simulation of systems under constant pressure. To this end, we introduce periodic boundary conditions in form of the Cyclic Cluster Model (CCM) [108-110]. In this model, the molecules are placed inside a (large) unit cell, which defines the size of the system during the calculation, e.g. the number of atoms. This unit cell is then translated along its axes in all directions without increasing the system size for the calculation. Such a scheme is shown in Fig. 6 for a two-dimension unit cell containing the atoms A0 , B0 and C0. The dashed objects are created by translation of the central unit cell.

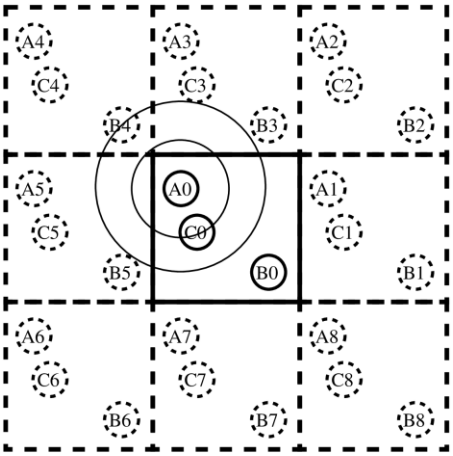


Figure 6: Translation scheme and interactions of atom A in a CCM calculation.

In the CCM framework, the interaction between the atoms A and B, for example, may not be calculated between A0 and B0. In case of the scheme shown in Fig. 6, the calculated interaction will be between A0 and B4. The atom B4 is chosen because it has from all B atoms the smallest distance to atom A0. On the other hand, the calculated interaction for atoms A and C is between A0

and C0 because from all C atoms C0 has the smallest distance to atom A0. If there are several combinations, which have the same distance, all will be calculated and an average is used. As in a finite system, the interactions are calculated in real space. What is different is the relative orientation of the atoms to each other. For a structure like a regular lattice of argon atoms, this procedure creates for each atom an identical and symmetric environment. Therefore, the CCM scheme is an alternative realization of periodic boundary conditions. Fig. 7 compares the convergence of the cohesive energy of solid Lennard-Jones argon with increasing cluster size between the free cluster model (FCM) and CCM. As this figure shows CCM clusters reach the converged cohesive energy of 8.42 kJ/mol at much smaller cluster sizes, i.e. at smaller average coordination numbers, than FCM clusters.

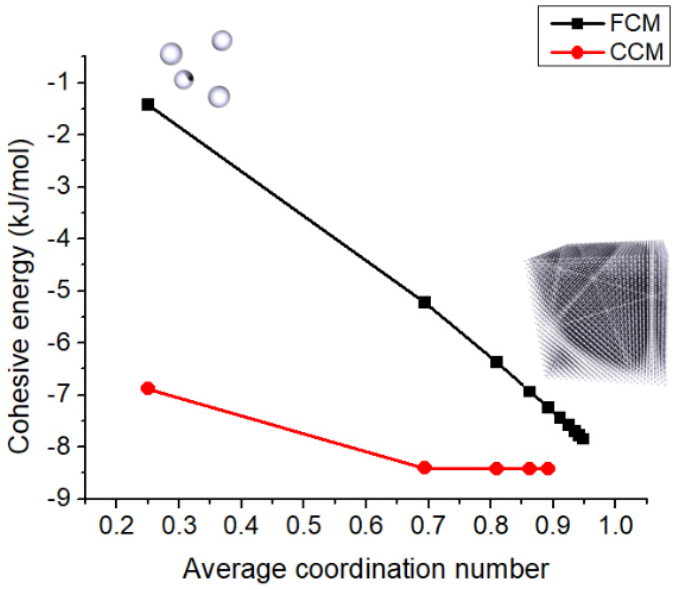


Figure 7: Cohesive energy of solid argon as a function of cluster size calculated using the Free Cluster Model (FCM) and the Cyclic Cluster Model (CCM).

Another characteristic of CCM is that the atoms inside the system cannot leave it due to the equality between atoms and their translations. Thus, if during a BOMD simulation an atom would leave the unit cell, a translation would bring it back inside. This is so because the projection of the distance

vector between two atoms onto an edge of the unit cell is not allowed to exceed half of the length of that edge, effectively creating an interaction region of this size around each atom. Only atoms within this region generate interactions with the atom in the center. The shape of this region is equal to a Wigner-Seitz cell (WSC) constructed from the unit cell. With this, one can describe the CCM as exchanging summations over all atoms in the (large) unit cell by summations over all atoms in a WSC. A further consequence of the use of WSCs is that there are no direct interactions of an atom with its equivalent in a translated unit cell, meaning that direct defect-defect interactions are not existing. This is particularly useful in a QM/MM solvation model for the treatment of infinitely diluted systems in which the solute is seen as a defect in the periodicity of the solvent environment.

In order to extend CCM to polar systems long-range Coulomb interactions must be added. To this end, the lattice sum algorithm of Ewald in its classical [111] formulation is available in deMon2k. The target quantity for this method is the electrostatic potential generated by the classical point charges of the MM system given by:

$$E = \frac{1}{2}\sum_{C}^{MM}\sum_{D}^{MM}\sum_{|\boldsymbol{m}|=0}^{\infty}\frac{Q_C Q_D}{|\boldsymbol{C}-\boldsymbol{D}+\boldsymbol{m}L|} \tag{110}$$

For sake of simplicity, it has been assumed that the particles are inside a cubic box of length L. The vector $\boldsymbol{m}$ of integer components works, together with L, as a translation operator of the reference cell. Its norm is represented by $|\boldsymbol{m}|$. Care must be taken in Eq. (110) for $\boldsymbol{m} = (0,0,0)$ to avoid self-interaction. In the classical Ewald method, the calculation of Eq. (110) for neutral systems is carried out by means of two rapidly and absolutely convergent series evaluated in real and reciprocal space [112]. The starting point of this method is the total potential generated by the point charges of the particles inside the original cell and its images, namely:

$$\phi(\boldsymbol{r}) = \sum_{C}^{MM} \sum_{|\boldsymbol{m}|=0}^{\infty} \frac{Q_C}{|\boldsymbol{r} - \boldsymbol{C} + \boldsymbol{m}L|} \tag{111}$$

The main idea behind the Ewald method is to use Gaussian charge distributions of the same magnitude but opposite sign of the point charges to shield them from their environment, as it is shown in Fig. 8 (left). These distributions are of the form:

$$\rho_C(\boldsymbol{r}) = \frac{Q_C \alpha^3}{\sqrt{\pi^3}} e^{-\alpha^2(\boldsymbol{r}-\boldsymbol{C})^2} \tag{112}$$

The parameter α is the width of the distributions. It also has the effect of biasing the total potential energy towards one of the two main contributions in which it is divided. Of course, this energy is invariant with respect to the value of α.

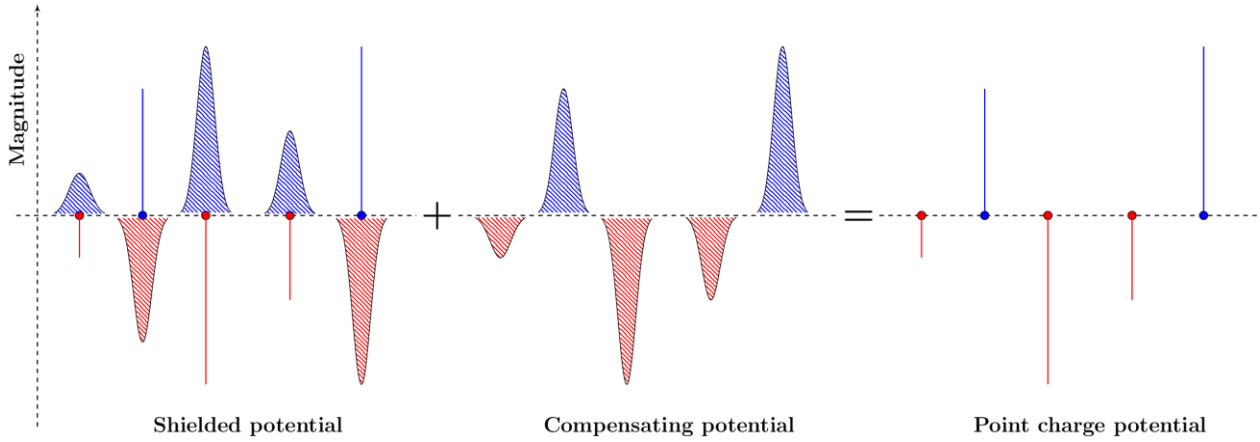


Figure 8: One-dimensional point and Gaussian charge distributions. The original potential (right-hand side) is obtained by adding the shielded and compensating potentials.

Due to the presence of the Gaussian charge distributions, the point-charge potential becomes short-range and can be calculated directly in real space. On the other hand, a second set of Gaussian distributions with the same sign and magnitude of the original point charge distributions is added to cancel the first set of distributions. Thus, the potential from the point charges is split into two parts, and the energy expression of Eq. (110) becomes:

$$E = \frac{1}{2}\sum_{C}^{MM}\sum_{D}^{MM}\sum_{|\boldsymbol{m}|=0}^{\infty}\frac{Q_C Q_D}{|\boldsymbol{C}-\boldsymbol{D}+\boldsymbol{m}L|}\,\mathrm{erfc}(\alpha|\boldsymbol{C}-\boldsymbol{D}+\boldsymbol{m}L|) \; + \\ \frac{4\pi}{V}\sum_{C}^{MM}\sum_{D}^{MM}\sum_{|\boldsymbol{k}|\neq 0}^{\infty}\frac{Q_C Q_D}{k^2}e^{i\boldsymbol{k}\cdot(\boldsymbol{C}-\boldsymbol{D})}e^{-k^2/4\alpha^2} - \frac{\alpha}{\sqrt{\pi}}\sum_{C}^{MM}Q_C^2 \qquad (113)$$

The first term in this equation is a sum in the manner of the Coulomb potential calculated over the original and image cells ($\boldsymbol{m}$ vectors), weighted by the complimentary error function $\mathrm{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}\mathrm{d}u$. This function approaches zero as its argument grows, improving the convergence of the sum. The second term in Eq. (113) is a sum arising from the compensating distributions and is calculated over the reciprocal vectors $\boldsymbol{k} = \frac{2\pi}{L}\boldsymbol{m}$. The last term in Eq. (113) accounts for self-interaction.

In practice, finite limits for the $\boldsymbol{m}$ and $\boldsymbol{k}$ sums are needed. These limits are chosen as a compromise of the desired accuracy and the computational demand for the calculation. Usually, the sum over $\boldsymbol{m}$ is carried out in the original cell only ($|\boldsymbol{m}| = 0$), whereas the $\boldsymbol{k}$ sum is performed over around 300 vectors [112, 113]. This is achieved by tuning the α parameter towards *large* values. For $\alpha = \sqrt{\pi}/L$ [114] both sums converge at the same rate. A value of $\alpha = 6/L$ is used in deMon2k. To illustrate the performance of this implementation we compare MM CCM water calculations augmented by long-range Coulomb lattice sums according to Eq. (113) against deMon2k embedding and periodic Tinker [115] calculations. To this end, a large unit cell (CCM cluster) of four water molecules as depicted in Fig. 9 (left) is used. For the deMon2k embedding calculations performed with Eq. (110), the system was replicated around the original cell following both a cubic and a spherical pattern (Fig. 9, right). The convergence of the energy of these systems towards the value calculated with Eq. (113) of -22.40 kcal/mol is shown in Fig. 10 (left).

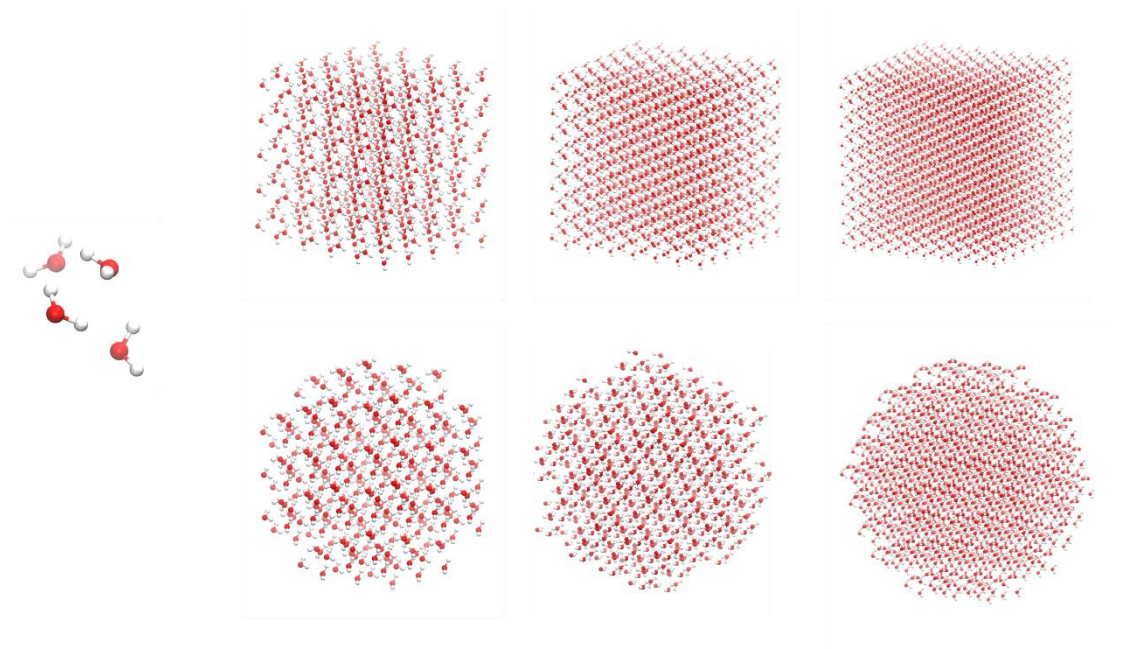


Figure 9: CCM cluster of four water molecules (left) together with its cubic (right, up) and spherical (right, down) embedding supercells of increasing size.

The norms of the corresponding forces per atom for this system are shown in the right part of this figure.

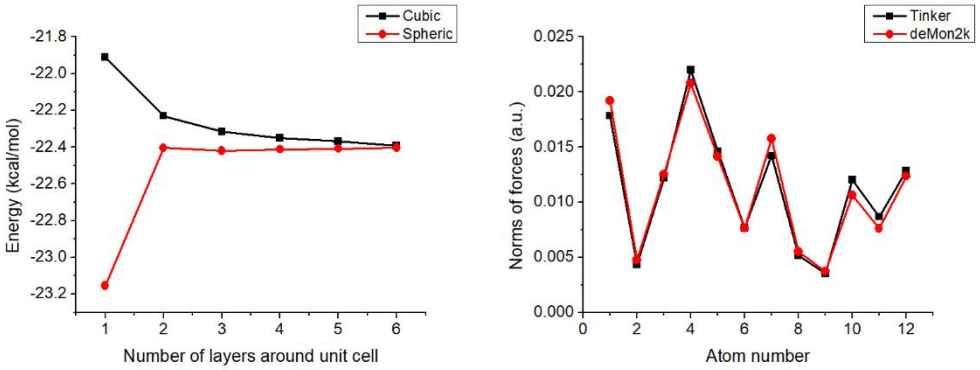


Figure 10: Energy *convergence of deMon2k embedding calculations towards the Ewald energy result (left) and comparison of the deMon2k CCM force norms against periodic Tinker calculations (right).*

The Ewald method as described above is implemented in deMon2k for MM systems under CCM periodic boundary conditions. Nonetheless, in the case of QM/MM systems, interactions between the QM electron density and the periodic point charges need to be considered. This same problem has been recently investigated in the literature [116], however, within a different working framework to the one discussed here. In QM/MM systems the potential energy has the following form:

$$E = \sum_{C}^{MM} \sum_{|\boldsymbol{m}|=0}^{\infty} \int \frac{Q_C \rho(\boldsymbol{r})}{|\boldsymbol{r} - \boldsymbol{C} + \boldsymbol{m}L|} \mathrm{d}\boldsymbol{r} = \int \phi(\boldsymbol{r}) \rho(\boldsymbol{r}) \mathrm{d}\boldsymbol{r} \tag{114}$$

The potential $\phi(\boldsymbol{r})$ is given by Eq. (111). As it was commented before, this potential is splits into two contributions by the Ewald method. Therefore, Eq. (114) is partitioned into two integrals. Their form, when considering that the real part of the potential is calculated in the original cell only, are:

$$E_{(1)} = \sum_{C}^{MM} \sum_{a,b} Q_C P_{ab} \int a(\boldsymbol{r}) b(\boldsymbol{r}) \frac{\mathrm{erfc}(\alpha|\boldsymbol{r} - \boldsymbol{C}|)}{|\boldsymbol{r} - \boldsymbol{C}|} \, \mathrm{d}\boldsymbol{r} \tag{115}$$

$$E_{(2)} = \frac{4\pi}{V} \sum_{C}^{MM} \sum_{|\boldsymbol{k}|\neq 0}^{\infty} \sum_{a,b} \frac{Q_C}{k^2} P_{ab} e^{-k^2/4\alpha^2} \int a(\boldsymbol{r}) b(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{C})} \, \mathrm{d}\boldsymbol{r} \tag{116}$$

Notice that there is no need for a self-interaction term, which only appears in Eq. (113) due to the point-like nature of the MM charges. An efficient methodology for the evaluation of these integrals is currently under development in one of our research groups. The availability of these integrals and their gradients will make it possible to use the MTK barostat inside QM/MM calculations and thus, fully atomistic simulations under ambient conditions will become possible.

5. Second-order energy derivatives

The calculation of second-order energy derivatives with respect to nuclear displacements in form of Hessian matrix elements is essential for the characterization of optimized molecular structures. It is also of great importance for transition state searches (see Section 3) and structure optimizations with soft internal degrees of freedom that are common to QM/MM systems. The dimensionality of the Hessian matrix is $9M^2$, with M being the number of atoms in the system. Although the first-order electronic energy derivatives are relatively easy to calculate, second-order derivatives are less straightforward. The simplest way to calculate second-order derivatives is to follow a finite-difference method where typically two first-order derivatives are required for each atomic coordinate. This method is available for QM/MM calculations in deMon2k but requires significant computational resources and is prone to numerical instabilities. Thus, an analytic second-order energy derivative formulation for QM/MM is very desirable.

Similar to the calculation of the QM/MM gradient components, after a converged SCF solution for the QM part, Hessian matrix components of a molecular system are obtained by differentiation of Eq. (48) once again with respect to atomic coordinates B_η, with $\eta = x, y, z$:

$$\frac{\partial^2 E^{QM/MM}}{\partial A_\lambda \partial B_\eta} = \frac{\partial^2 E^{QM}}{\partial A_\lambda \partial B_\eta} + \frac{\partial^2 E^{MM}}{\partial A_\lambda \partial B_\eta} + \frac{\partial^2 E^{QMMM}}{\partial A_\lambda \partial B_\eta} + \frac{\partial^2 E^{NN}}{\partial A_\lambda \partial B_\eta} \tag{117}$$

The first term in Eq. (117) corresponds to the generic QM second-order derivative. For an ADFT QM-only calculation, this term is calculated by differentiation of Eq. (13), with respect to a QM atom coordinate A_λ and a QM atom coordinate B_η, and the second, third and fourth term of Eq. (117) will vanish. According to Wigner's $2n + 1$ rule [69,70], the resulting expression contains the first-order response density matrix as well as the response Coulomb and exchange-correlation

fitting coefficients. Based on this, the expression of the second-order ADFT energy derivatives can be partitioned in two contributions as [117]:

$$\frac{\partial^2 E^{QM}}{\partial A_\lambda \partial B_\eta} = \frac{\partial^2 E^{QM}_{(1)}}{\partial A_\lambda \partial B_\eta} + \frac{\partial^2 E^{QM}_{(2)}}{\partial A_\lambda \partial B_\eta} \tag{118}$$

The first contribution of Eq. (118), is completely independent from the perturbed density matrix and fitting coefficients and has the form:

$$\begin{aligned}\frac{\partial E^{QM}_{(1)}}{\partial A_\lambda \partial B_\eta} = & -\sum_{\mu,\nu} W_{\mu\nu}\, S^{(A_\lambda,B_\eta)}_{\mu\nu} + \sum_{\mu,\nu} P_{\mu\nu}\, H^{(A_\lambda,B_\eta)}_{\mu\nu} + \\ & \sum_{\mu,\nu}\sum_{\bar{k}} P_{\mu\nu} \left\langle \mu\nu \middle\| \bar{k} \right\rangle^{(A_\lambda,B_\eta)} (x_{\bar{k}} + z_{\bar{k}}) - \sum_{\bar{k},\bar{l}} G_{\bar{k}\bar{l}}{}^{(A_\lambda,B_\eta)}\, x_{\bar{k}} \left(\frac{1}{2} x_{\bar{l}} + z_{\bar{l}}\right) + \\ & \sum_{\bar{k}} x_{\bar{k}} \left\langle \bar{k}^{(A_\lambda,B_\eta)} \middle| v_{xc}[\tilde{\rho}] \right\rangle + \sum_{\bar{k},\bar{l}} x_{\bar{k}} \left\langle \bar{k}^{(A_\lambda)} \middle| f_{xc}[\tilde{\rho}] \middle| \bar{l}^{(B_\eta)} \right\rangle \end{aligned} \tag{119}$$

Here, $f_{xc}[\tilde{\rho}]$, denotes the ADFT exchange-correlation kernel, defined as the functional derivative of the ADFT exchange-correlation potential. In the case of pure density functionals, the exchange-correlation kernel is defined as [52,118]:

$$\frac{\delta v_{xc}[\tilde{\rho}](\boldsymbol{r})}{\delta \tilde{\rho}(\boldsymbol{r}')} = \delta(\boldsymbol{r} - \boldsymbol{r}') \frac{d v_{xc}[\tilde{\rho}](\boldsymbol{r})}{d \tilde{\rho}(\boldsymbol{r}')} = f_{xc}[\tilde{\rho}](\boldsymbol{r}) \tag{120}$$

The analytic molecular integral derivatives in Eq. (119) are calculated over the corresponding integral recurrence relations [30,119], whereas numerical integration of the exchange-correlation potential and kernel are evaluated with the same adaptive grid that is used for the SCF [63,64]. The exchange-correlation kernel is calculated either from analytic expressions for each functional [52] or by finite-differences if only analytic exchange-correlation potential expressions are available [53]. The density, $P_{\mu\nu}$, and energy-weighted density, $W_{\mu\nu}$, matrix elements are calculated from the

converged canonical MO coefficients and energies. The Coulomb fitting coefficients, $x_{\bar{k}}$, and exchange-correlation fitting coefficients, $z_{\bar{k}}$, are also taken directly from the previously converged single-point energy calculation. The second contribution to the QM second-order ADFT energy derivative in Eq. (118), contains only terms that depend on first-order response matrix or vector elements. It is given by:

$$\begin{aligned}\frac{\partial E_{(2)}^{QM}}{\partial A_\lambda \partial B_\eta} = & -\sum_{\mu,\nu} W_{\mu\nu}^{(B_\eta)} S_{\mu\nu}^{(A_\lambda)} + \sum_{\mu,\nu} P_{\mu\nu}^{(B_\eta)} H_{\mu\nu}^{(A_\lambda)} + \\ & \sum_{\mu,\nu}\sum_{\bar{k}} P_{\mu\nu}^{(B_\eta)} \langle \mu\nu \| \bar{k} \rangle^{(A_\lambda)} (x_{\bar{k}} + z_{\bar{k}}) + \\ & \sum_{\mu,\nu}\sum_{\bar{k}} P_{\mu\nu} \langle \mu\nu \| \bar{k} \rangle^{(A_\lambda)} \left(x_{\bar{k}}^{(B_\eta)} + z_{\bar{k}}^{(B_\eta)} \right) - \sum_{\bar{k},\bar{l}} G_{\bar{k}\bar{l}}{}^{(A_\lambda)} (x_{\bar{k}} + z_{\bar{k}})\, x_{\bar{l}}^{(B_\eta)} - \\ & \sum_{\bar{k},\bar{l}} G_{\bar{k}\bar{l}}{}^{(A_\lambda)} z_{\bar{k}}^{(B_\eta)}\, x_{\bar{l}} + \sum_{\bar{k}} x_{\bar{k}}^{(B_\eta)} \langle \bar{k}^{(A_\lambda)} | v_{xc}[\tilde{\rho}] \rangle + \\ & \sum_{\bar{k},\bar{l}} x_{\bar{k}} \langle \bar{k}^{(A_\lambda)} | f_{xc}[\tilde{\rho}] | \bar{l} \rangle\, x_{\bar{l}}^{(B_\eta)}\end{aligned} \tag{121}$$

Again, the here appearing molecular integral derivatives can be straightforwardly calculated by integral recurrence relations. For the calculation of the perturbed matrix and vector elements, we employ auxiliary density perturbation theory (ADPT). This methodology is an adaptation of McWeeny's self-consistent perturbation (SCP) theory [120-124] to ADFT. It allows the calculation of the first-order response density matrix elements with respect to QM atom displacements by the following expression:

$$P_{\mu\nu}^{(B_\eta)} = 2 \sum_{i}^{occ} \sum_{a}^{uno} \frac{\mathcal{K}_{ia}^{(B_\eta)} - \epsilon_i \mathcal{S}_{ia}^{(B_\eta)}}{\epsilon_i - \epsilon_a} \left(c_{\mu i} c_{\nu a} + c_{\mu a} c_{\nu i} \right) - \frac{1}{2} \sum_{\sigma,\tau} P_{\mu\sigma} S_{\sigma\tau}^{(B_\eta)} P_{\tau\mu} \tag{122}$$

In Eq. (122) $\mathcal{K}_{ia}^{(B_\eta)}$ and $\mathcal{S}_{ia}^{(B_\eta)}$ denote first-order response Kohn-Sham and overlap matrix elements in molecular orbital representation, which are defined by:

$$\mathcal{K}_{ia}^{(B_\eta)} \equiv \sum_{\sigma,\tau} c_{\sigma\tau} K_{\sigma\tau}^{(B_\eta)} c_{\tau a} \tag{123}$$

and

$$S_{ia}^{(B_\eta)} \equiv \sum_{\sigma,\tau} c_{\sigma\tau} S_{\sigma\tau}^{(B_\eta)} c_{\tau a} \tag{124}$$

In conventional Kohn-Sham DFT, due to the fact that the perturbed Kohn-Sham matrix depends on the perturbed density matrix, Eq. (122) has to be solved iteratively to achieve self-consistency [120]. This iterative evaluation can become a serious computational bottleneck if convergence is slow. In contrast, in the ADFT framework, the corresponding Kohn-Sham matrix only depends on the fitted density [125]. As a result, McWeeny's SCP method can be reformulated as a non-iterative approach in terms of the response of the fitting coefficients. The resulting ADPT equations yield coupled-perturbed Kohn-Sham (CPKS) analog equation systems, albeit of significantly reduced dimensions. It reads [29,125,126]:

$$\left[\frac{1}{4}\boldsymbol{G} - \boldsymbol{A}(\boldsymbol{E}+\boldsymbol{F})\right]\boldsymbol{x}^{(B_\eta)} = \boldsymbol{b}^{(B_\eta)} \tag{125}$$

Here $\boldsymbol{G}$ denotes the Coulomb matrix, Eq. (12). The Coulomb response matrix, $\boldsymbol{A}$, and the kermel matrix, $\boldsymbol{F}$, are given by:

$$A_{\bar{k}\bar{l}} = \sum_{i}^{occ}\sum_{a}^{uno} \frac{\langle \bar{k}\|ia\rangle\langle ia\|\bar{l}\rangle}{\epsilon_i - \epsilon_a} \tag{126}$$

and:

$$F_{\bar{k}\bar{m}} = \sum_{\bar{l}} G_{\bar{k}\bar{l}}^{-1}\langle \bar{l}|f_{xc}[\tilde{\rho}]|\bar{m}\rangle \tag{127}$$

The linear equation system of Eq. (125) keeps the same form irrespective of the nature of the perturbation parameter (electric, magnetic, QM or MM nuclear displacement). The essential

difference arises from the perturbation vector elements, $b_{\bar{k}}^{(B_\eta)}$, which are characteristic to the specific perturbation. So far, ADPT has been successfully applied for the calculation of static and dynamic polarizabilities [29,125,126], first hyperpolarizabilities [127], Fukui functions [128,129] and second-order nuclear displacements [117]. For the latter the perturbation vector becomes much more complicated due to the perturbation dependency of basis and auxiliary functions. In fact, in this case the calculation of these vector elements can become a computational bottleneck because they are needed for each individual perturbation.

To solve the ADPT equation system, Eq. (125), we can calculate the inverse of the response matrix, $\boldsymbol{R} = \frac{1}{4}\boldsymbol{G} - \boldsymbol{A}(\boldsymbol{E} + \boldsymbol{F})$ and then solve the equation system analytical. This is possible because the dimension of the ADPT response matrix $(N_{aux} \times N_{aux})$ is much smaller than the $(N_{occ} \times N_{uno}) \times (N_{occ} \times N_{uno})$ CPKS dimension [125]. An important consideration is that $\boldsymbol{R}$ is perturbation-independent. This is clearly an advantage, because regardless of the number of perturbations (or perturbation types), $\boldsymbol{R}$ remains unaltered. Thus, a single inversion is needed and $\boldsymbol{R}^{-1}$ can be stored on disk and read when required. However, as system size grows, the response matrix $\boldsymbol{R}$ tends to become ill-conditioned, and therefore, prone to numerical instabilities during the matrix inversion. Therefore, a robust and efficient algorithm for solving large indefinite non-symmetric systems of linear equations is needed [130]. In deMon2k we use the Eirola-Nevanlinna (EN) algorithm [27,29,131]. As the evaluation of analytic QM second energy derivatives requires the solution of three times the number of QM atoms ADPT equation systems, the pre-calculation of the Coulomb response matrix, $\boldsymbol{A}$, and the kernel matrix, $\boldsymbol{F}$, is convenient. However, the calculation of the Coulomb response matrix introduces an overall $\mathcal{O}(N_{aux}^4)$ computational complexity [125]. Despite its unfavorable scaling, the explicit evaluation of $\boldsymbol{A}$ and $\boldsymbol{F}$ is usually computationally advantageous for second-order energy derivatives because these matrices have to

be calculated only once. In analogy to the conventional SCF approach, this implementation for the solution of the ADPT equation systems is called the conventional EN (Con-EN) method [117]. A drawback of this implementation, besides its high order formal scaling, is its random access memory (RAM) demand. Thus, it is only suitable for small systems with up to a few hundred of atoms. To overcome this limitation, a direct implementation of the EN algorithm (Dir-EN) is available in deMon2k, too. The advantage of this method is that the explicit computation of the Coulomb response matrix and kernel matrix is avoided. Instead, the action of these matrices on a trial vector $\boldsymbol{p}$ is calculated in the iterative Dir-EN procedure. This reduces the overall $\mathcal{O}(N_{aux}^4)$ complexity to $\mathcal{O}(N_{aux}^3)$ [29] by breaking down the matrix vector multiplication into a series of sub-steps. This allows efficient ERI screening and double asymptotic expansions in the action calculations with the Coulomb response matrix. Similar, numerical integration screening techniques can be used for the action calculation with the kernel matrix. Besides the reduction in computational complexity, the Dir-EN implementation has also a reduced RAM demand. It requires about 2 N_{bas}^2 matrices and a few N_{aux} vectors, compared to the $N_{bas} \times N_{occ} \times N_{uno}$ RAM demand for the explicit Coulomb response matrix calculation in the Con-EN implementation. This permits ADPT calculations on systems with thousand and more atoms [29]. As for the direct SCF algorithm implemented in deMon2k, the Dir-EN implementation requires the repetitive calculation of three-center ERIs in each iteration step of the EN algorithm. For many perturbations, as needed for the calculation of analytic second-order energy-derivatives, the repetitive ERI calculation is usually the most demanding computational task in the Dir-EN method. Fortunately, three-center ERI calculation is well developed in deMon2k [30,46,119] and also very well parallelized [132,133].

Once the ADPT equation systems are solved and the perturbed Coulomb fitting coefficients are obtained, the calculation of the remaining response matrix and vector elements is straightforward. First, the perturbed exchange-correlation fitting coefficients, $z_{\bar{k}}^{(B_\eta)}$, are obtained according to [117]:

$$z_{\bar{k}}^{(B_\eta)} = \sum_{\bar{l}} G_{\bar{k}\bar{l}}^{-1} \langle \bar{k}^{(B_\eta)} | v_{xc}[\tilde{\rho}] \rangle - \sum_{\bar{l},\bar{m}} G_{\bar{k}\bar{l}}^{-1} G_{\bar{l}\bar{m}}^{(B_\eta)} z_{\bar{m}} + \sum_{\bar{l},\bar{m}} G_{\bar{k}\bar{l}}^{-1} \langle \bar{l} | f_{xc}[\tilde{\rho}] | \bar{m} \rangle \, x_{\bar{m}}^{(B_\eta)} + \sum_{\bar{l},\bar{m}} G_{\bar{k}\bar{l}}^{-1} \langle \bar{l} | f_{xc}[\tilde{\rho}] | \bar{m}^{(B_\eta)} \rangle \, x_{\bar{m}} \tag{128}$$

After that, the perturbed Kohn-Sham matrix elements, which are obtained by direct differentiation of Eq. (21) with respect to a QM nuclear coordinate, are calculated with the perturbed Coulomb and exchange-correlation fitting coefficients:

$$K_{\mu\nu}^{(B_\eta)} = H_{\mu\nu}^{(B_\eta)} + \sum_{\bar{k}} \langle \mu\nu \| \bar{k} \rangle^{(B_\eta)} \, (x_{\bar{k}} + z_{\bar{k}}) + \sum_{\bar{k}} \langle \mu\nu \| \bar{k} \rangle \, \left(x_{\bar{k}}^{(B_\eta)} + z_{\bar{k}}^{(B_\eta)} \right) \tag{129}$$

Afterwards, the perturbed Kohn-Sham matrix elements are transformed into MO representation using Eq. (123), and the perturbed density matrix is calculated according to Eq. (122). Once the perturbed density and Kohn-Sham matrices are available the perturbed energy-weighted density matrix $\boldsymbol{W}^{(B_\eta)}$ is calculated as:

$$W_{\mu\nu}^{(B_\eta)} = \frac{1}{2} \sum_{\sigma,\tau} \left(P_{\mu\sigma}^{(B_\eta)} K_{\sigma\tau} P_{\tau\nu} + P_{\mu\sigma} K_{\sigma\tau}^{(B_\eta)} P_{\tau\nu} + P_{\mu\sigma} K_{\sigma\tau} P_{\tau\nu}^{(B_\eta)} \right) \tag{130}$$

With all perturbed matrix and vector elements at hand, the corresponding contributions of the perturbation-dependent second derivatives, Eq. (121), can be calculated. This concludes the calculation of analytic second-order ADFT energy derivatives with respect to nuclear displacements. For the simulation of infrared (IR) and Raman spectra within the double harmonic

approximation [134,135] second- and third-order mixed energy derivatives with respect to QM atomic coordinates and components of an external electric field are needed, too. They are implemented as analytical and semi-numerical derivatives in deMon2k [136]. Finally, we note that the here outlined analytic second-order ADFT energy derivative calculation is currently extended to effective and model core potentials employing half-numerical integral evaluation [50].

Although ADPT within ADFT can provide first-principle response properties of systems up to around thousand atoms [117], a frequency analysis of a protein with many thousand atoms in water is still computationally unfeasible with a first-principle QM-only methodology. For such systems with many thousands of atoms the QM/MM methodology in deMon2k can be used [42,44]. In analogy to the QM/MM energy expression, Eq. (38), second-order QM/MM energy derivatives can be expressed as:

$$\frac{\partial^2 E^{QM/MM}}{\partial A_\lambda \partial B_\eta} = \frac{\partial^2 E^{QM}}{\partial A_\lambda \partial B_\eta} + \frac{\partial^2 E^{MM}}{\partial A_\lambda \partial B_\eta} + \frac{\partial^2 E^{QMMM}}{\partial A_\lambda \partial B_\eta} + \frac{\partial^2 E^{NN}}{\partial A_\lambda \partial B_\eta} \tag{131}$$

The last three terms of the right-hand side of Eq. (131) are analytic derivatives of the corresponding gradient terms discussed in Section 2. Therefore, they are straightforward to calculate and will not be further discussed here. The complete Hessian matrix, schematically depicted in Fig. 11, can be separated into four regions, labeled as QM-QM, QM-MM, MM-QM and MM-MM in Fig. 11. By symmetry, the QM-MM and MM-QM blocks are identical. In the QM-QM block, second ADFT energy derivatives with respect to QM atomic coordinates are collected. They are calculated as outlined above. Note, however, that they include according to Eq. (119) second derivatives of the QM/MM core Hamilton matrix elements given by Eq. (36). The elements of the QM-QM block also contain analytic second derivatives of E^{QMMM} and E^{NN} with respect to QM atomic coordinates, which are straightforward to calculate. The QM-MM block elements contain

contributions from the same three terms, namely E^{QM}, E^{QMMM} and E^{NN}. Again, we discuss only the terms arising from the (mixed) second derivatives of E^{QM}. Their calculation is simpler because the basis and auxiliary functions of the QM atoms are independent from the coordinates of the MM atoms and the MM atoms themselves carry none of these functions. Taking this into account, the evaluation of Eq. (118) for a QM atomic coordinate A_λ and a MM atomic coordinate D_η yields:

$$\frac{\partial^2 E^{QM}}{\partial D_\eta \partial A_\lambda} = \sum_{\mu,\nu} P_{\mu\nu}\, H_{\mu\nu}^{(D_\eta, A_\lambda)} + \sum_{\mu,\nu} P_{\mu\nu}^{(A_\lambda)}\, H_{\mu\nu}^{(D_\eta)} \tag{132}$$

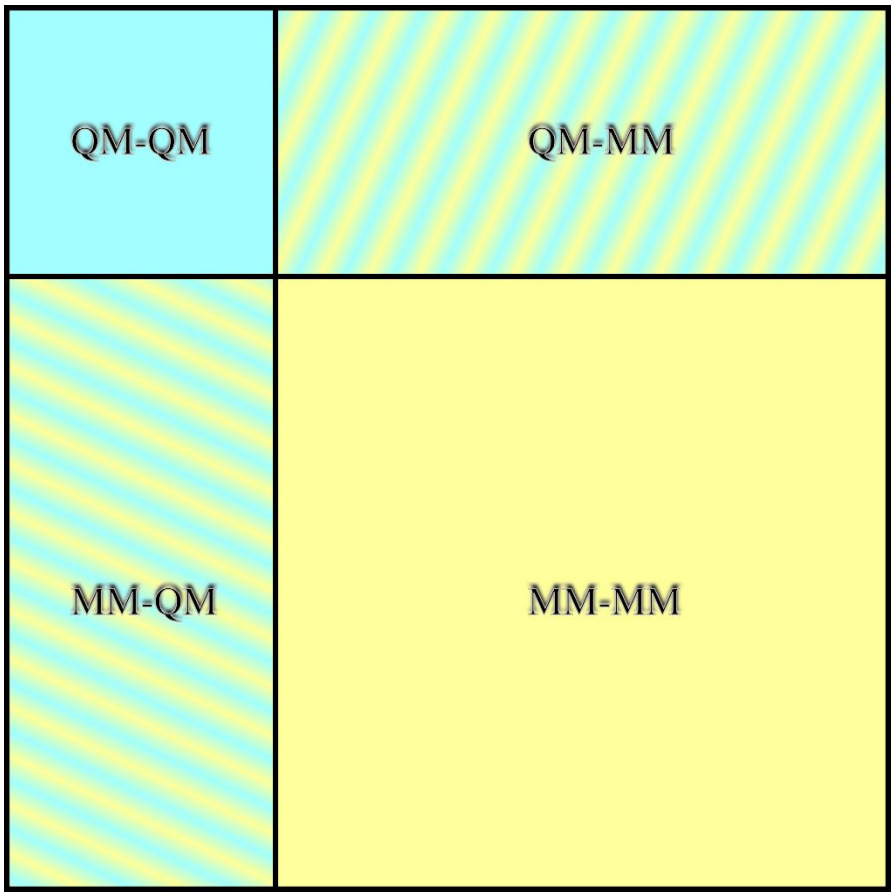


Figure 11: Hessian matrix structure for a QM/MM system.

Note that the perturbed density matrix elements in Eq. (132) are the same as the once needed for the elements in the QM-QM block. The perturbed core matrix elements in Eq. (129) are given by (without employing asymptotic expansions)

$$H_{\mu\nu}^{(D_\eta)} = -\langle\mu\nu\big|\hat{\mathcal{A}}_D\big(1_{\boldsymbol{\eta}}\big)\rangle\, Q_D \tag{133}$$

and

$$H_{\mu\nu}^{(D_\eta,A_\lambda)} = -\langle\mu\nu\big|\hat{\mathcal{A}}_D\big(1_{\boldsymbol{\eta}}\big)\rangle^{(A_\lambda)}\, Q_D. \tag{134}$$

Finally the MM-MM block elements contain contributions from all four analytic second derivatives on the right-hand side of Eq. (131). Again the E^{MM}, E^{QMMM} and E^{NN} second derivatives with respect to MM atomic coordinates are straightforward to calculate. For the E^{QM} analytic second derivatives with respect to MM atomic coordinates D_λ and D'_η follows:

$$\frac{\partial^2 E^{QM}}{\partial D_\lambda \partial D'_\eta} = \sum_{\mu,\nu} P_{\mu\nu}\, H_{\mu\nu}^{(D_\lambda,D'_\eta)} + \sum_{\mu,\nu} P_{\mu\nu}^{(D'_\eta)}\, H_{\mu\nu}^{(D_\lambda)} \tag{135}$$

with

$$H_{\mu\nu}^{(D_\lambda,D'_\eta)} = \begin{cases} \langle\mu\nu\big|\hat{\mathcal{A}}_D\big(\mathbf{1}_{\boldsymbol{\lambda}} + \mathbf{1}_{\eta}\big)\rangle & if\ D = D' \\ 0 & else \end{cases} \tag{136}$$

Eq. (135) implies the calculation of perturbed density matrix elements, $P_{\mu\nu}^{(D'_\eta)}$, for each corresponding MM degree of freedom. This is expected to be the principal bottleneck for the analytical Hessian calculation of a QM/MM system because there are usually much more MM atoms than QM atoms. Therefore, the ADPT equations must be solved for *all* MM degrees of freedom, too. Fortunately, the response equation system will have important simplifications as basis and auxiliary functions are independent from MM atom coordinates. These simplifications have been already explored in the framework of conventional DFT/MM calculations using CPKS

response equations [137,138]. Despite these improvements, such calculations are still prohibitive for large systems, including enzymes. Thus, a derivation of second-order ADFT energy derivatives within the QM/MM framework is expected to be a useful tool to evaluate frequencies and infrared intensities of complex systems with many thousands of atoms. Work in this direction is currently under development within the deMon developers' community.

6. QM/MM Magnetic shielding and excited states calculations

The calculation of magnetic properties such as magnetic shieldings and magnetizabilities of very large systems containing thousands of atoms, like the ones including explicit solvent molecules, can be efficiently handled by QM/MM methods. In the particular case of ADFT QM/MM magnetic property calculations [139-143], QM regions with around 1,000 atoms and 15,000 thousand basis functions can be used.The ADFT magnetic shielding tensor elements, $\sigma_{\lambda\eta}$, are defined as the second-order ADFT energy derivative with respect to a component of the magnetic field, $\mathcal{H}_{\lambda}$, and a component of the nuclear magnetic moment, μ_{C_η}:

$$\sigma_{\lambda\eta} = \frac{\partial^2 E^{QM}}{\partial \mathcal{H}_{\lambda} \partial \mu_{C_\eta}} \tag{137}$$

In a QM/MM calculation, only the QM atoms possess nuclear magnetic moments. Therefore, Eq. (137) is only defined for the QM atoms. Furthermore, the other terms of the QM/MM energy expression, Eq. (38), do not contribute to this mixed second energy derivative. Thus, expanding Eq. (137) yields for the shielding tensor elements:

$$\sigma_{\lambda\eta} = \sum_{a,b} P_{ab}^{(\lambda)} H_{ab}^{(\eta)} + \sum_{a,b} P_{ab} H_{ab}^{(\lambda\eta)} \tag{138}$$

In this equation a and b index Gauge-Including Atomic Orbitals (GIAOs) [139,144] centered at the QM nuclei A and B. The GIAOs, $\phi_a(\boldsymbol{r})$, are obtained by multiplying (contracted) GTOs with the so-called London factor:

$$\phi_a(\boldsymbol{r}) = a(\boldsymbol{r})\, e^{-\frac{i}{2}\boldsymbol{\mathcal{H}}\,(A-\boldsymbol{G})\times\boldsymbol{r}} \tag{139}$$

Here $a(\boldsymbol{r})$ is the field independent atomic orbital in form of a (contracted) GTO, and $\boldsymbol{A}$ and $\boldsymbol{G}$ are the position vectors of the QM nucleus A and the gauge, respectively. The use of GIAOs effectively resolves the gauge origin problem in magnetic shielding calculations. The perturbed core matrix elements in Eq. (138) remain unaltered in QM/MM shielding calculations compared to their QM-only counterparts. Their efficient calculation is described in [139]. Thus, the essential difference between QM/MM and QM shielding calculations arises from the calculation of the perturbed density matrix elements with respect to the external magnetic field components $\mathcal{H}_\lambda$. In ADFT GIAO, using LDA or GGA [145], they are given according to McWeeny's SCP theory [120-124] by:

$$P_{ab}^{(\lambda)} = 2\sum_{o}^{occ}\sum_{u}^{uno}\frac{\mathcal{K}_{ou}^{(\lambda)} - \epsilon_o \mathcal{S}_{ou}^{(\lambda)}}{\epsilon_o - \epsilon_u}\,(c_{ao}c_{bu} - c_{au}c_{bo}) - \frac{1}{2}\sum_{c,d} P_{ac}S_{cd}^{(\lambda)}P_{db} \tag{140}$$

Here, o and u denote occupied and unoccupied MOs, respectively. The minus sign in Eq. (140) compared to Eq. (122) arises from the imaginary perturbation dependency of the GIAO basis according to Eq. (139). The Kohn-Sham and overlap matrix elements in molecular orbital representation, $\mathcal{K}_{ou}^{(\lambda)}$ and $\mathcal{S}_{ou}^{(\lambda)}$, are calculated according to Eq. (123) and (124). The underlying perturbed matrix elements are given by [145]:

$$S_{ab}^{(\lambda)} = \frac{i}{2}\,\langle a|[(\boldsymbol{A}-\boldsymbol{B})\times\boldsymbol{r}]_\lambda|b\rangle \tag{141}$$

and

$$K_{ab}^{(\lambda)} = \frac{i}{2}\left\langle a\middle|[(\boldsymbol{A}-\boldsymbol{B})\times\boldsymbol{r}]_\lambda\,\hat{H}\middle|b\right\rangle - \frac{i}{2}\left\langle a|[(\boldsymbol{r}-\boldsymbol{B})\times\boldsymbol{\nabla}]_\lambda|b\right\rangle + \frac{i}{2}\sum_{\bar{k}}\left\langle ab[(\boldsymbol{A}-\boldsymbol{B})\times\boldsymbol{r}]_\lambda \parallel \bar{k}\right\rangle\,(x_{\bar{k}}+z_{\bar{k}}) \tag{142}$$

Because in QM/MM calculations the one-electron core Hamilton operator, $\hat{H}$, is augmented by the electrostatic embedding potential from the MM environment, Eq. (34), we can write Eq. (142) as:

$$K_{ab}^{(\lambda)} = K_{ab}^{QM^{(\lambda)}} - \frac{i}{2}\sum_{D}^{MM}\left\langle ab[(\boldsymbol{A}-\boldsymbol{B})\times\boldsymbol{r}]_\lambda|\,\hat{\mathcal{A}}_D(\boldsymbol{0})\right\rangle Q_D \tag{143}$$

Thus, the last term in Eq. (143) must be additionally computed for $\lambda = x, y, z$ in QM/MM magnetic shielding calculations. Efficient recurrence relations are available in deMon2k for these modified nuclear attraction type integrals. Therefore, the computational efficiency of the ADFT-GIAO methodology is preserved in QM/MM calculations.

Another example for QM linear response in the framework of ADFT QM/MM is the calculation of excited states. To this end, we compute excited states properties through Time-Dependent ADFT (TD-ADFT) [146]. In this formalism, the QM/MM excited state energy, E^*, is calculated as a single-electron excitation in the QM region from QM/MM the ground-state, Eq. (38):

$$E^* = E^{QM/MM} + \omega \tag{144}$$

For a given number of excited states, these excitation energies, ω, are obtained by solving an equation system that resembles that of the Random-Phase Approximation (RPA) [147]:

$$\begin{pmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{B} & \boldsymbol{A}\end{pmatrix}\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{pmatrix} = \omega\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & -\boldsymbol{I}\end{pmatrix}\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{pmatrix} \tag{145}$$

Both matrices, ***A*** and ***B,*** contain the Coulomb and Kernel response of the Kohn-Sham matrix, to a change in the electronic density. These matrices are free of any explicit contribution of the MM environment. For pure functionals, they are given by

$$A_{ia,jb} = \delta_{ij}\delta_{ab}(\epsilon_a - \epsilon_i) + \sum_{\bar{k},\bar{l}} \langle \psi_i\, \psi_a \| \bar{k} \rangle\, G^{-1}_{\bar{k}\bar{l}} \langle \bar{l} \| \psi_j\, \psi_b \rangle +$$

$$\sum_{\bar{m},\bar{l}} \sum_{\bar{k},\bar{l}} \langle \psi_i\, \psi_a \| \bar{k} \rangle\, G^{-1}_{\bar{k}\bar{l}} \langle \bar{l} | f_{xc}[\tilde{\rho}] | \bar{m} \rangle G^{-1}_{\bar{m}\bar{n}}\, \langle \bar{n} \| \psi_j\, \psi_b \rangle \tag{146}$$

and

$$B_{ia,jb} = \sum_{\bar{k},\bar{l}} \langle \psi_i\, \psi_a \| \bar{k} \rangle \mathrm{G}^{-1}_{\bar{k}\bar{l}} \langle \bar{l} \| \psi_b\, \psi_j \rangle +$$

$$\sum_{\bar{m},\bar{l}} \sum_{\bar{k},\bar{l}} \langle \psi_i\, \psi_a \| \bar{k} \rangle\, G^{-1}_{\bar{k}\bar{l}} \langle \bar{l} | f_{xc}[\tilde{\rho}] | \bar{m} \rangle G^{-1}_{\bar{m}\bar{n}}\, \langle \bar{n} \| \psi_b\, \psi_j \rangle\,. \tag{147}$$

As can be seen from Eq. (146) the $\boldsymbol{A}$ matrix has, in the diagonal, the energy differences between unoccupied, ϵ_a, and occupied, ϵ_i, canonical MO energies. This term, which is also known as single particle transition energy, is a few orders of magnitude bigger than the other contributions and therefore dominant in the calculation of the excitation energies. In this diagonal term the use of MM embedding contributes implicitly by shifting these molecular orbital energies, resulting in calculated excitation energies that are in good agreement with experiments, even in polar solutions [148,149].

The equation system in Eq. (145) can be reduced to the following eigenvalue problem

$$\boldsymbol{\Omega F} = \omega^2 \boldsymbol{F} \tag{148}$$

with

$$\boldsymbol{\Omega} = (\boldsymbol{A} - \boldsymbol{B})^{1/2}(\boldsymbol{A} + \boldsymbol{B})(\boldsymbol{A} - \boldsymbol{B})^{1/2} \tag{149}$$

and

$$\boldsymbol{F} = (\boldsymbol{A} - \boldsymbol{B})^{-1/2}(\boldsymbol{X} + \boldsymbol{Y}). \tag{150}$$

The calculation of $(\boldsymbol{A}-\boldsymbol{B})^{-1/2}$ is trivial if no Fock exchange is requested, which is the case of the TD-ADFT implementation in deMon2k [150,151]. Another simplification of the TD-ADFT equation system is the so-called Tamm-Dancoff approximation (TDA) [152] which is obtained by setting $\boldsymbol{B}=\boldsymbol{0}$ as:

$$\boldsymbol{AX}=\omega\boldsymbol{X} \tag{151}$$

Although it is an approximation, it has been shown that using the TDA yields the same or in some cases even better accuracy than the full equation system [152].

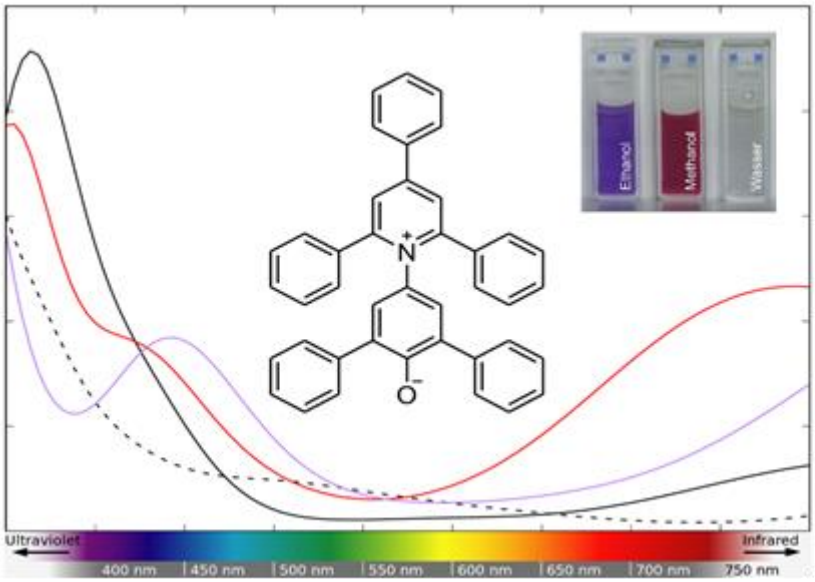


Fig. 1 Visible energy window of the calculated Reichardt's dye spectra in vacuum (dashed black), water (solid black), methanol (red).and ethanol (violet). Solutions of the molecule in water, methanol and ethanol is shown in the top right insert (taken from https://commons.wikimedia.org/wiki/File:Solvatochromismusbetain.jpg).

Regardless of the type of TD-ADFT, namely Eq. (148) or Eq. (151), the TD-ADFT matrix, $\boldsymbol{A}$ or $\boldsymbol{\Omega}$, is of dimension $(occ\ x\ uno)^2$, being occ and uno the number of occupied and unoccupied MOs, respectively. The diagonalization, or just storage, of this matrix can become computationally demanding even for very small systems (~20 atoms). Therefore, to avoid storage and explicit diagonalization of the full TD-ADFT matrix, the Davidson method [153,154], which is a direct diagonalization procedure, is implemented in deMon2k. With this method, it is possible to solve the TD-ADFT equation system for a given number of lowest excitation energies. As an example,

the calculated TD-ADFT visible spectrum of the Reichardt's dye at the PBE/DZVP-GGA/GEN-A2* level is shown in Fig. 12. For these calculations the molecular structures were optimized in vacuum, water, methanol and ethanol. All solvent molecules where calculated at the MM level, accounting for a total of around 1500 atoms. The OPLS-AA force field was used in all QM/MM calculations. As Fig. 12 shows the experimental observed solvent shifts in the excitation spectra are qualitatively correct reproduced by our QM/MM calculations.

7. Transformation program for deMon2k input generation

A major practical obstacle to run QM/MM simulations for complex molecular systems as those encountered in biology or in materials science is to prepare input files. For most systems, deMon2k requires to specify for each MM atom, its atom type, as defined in the FFDS file, and importantly its connectivity, i.e. the list of other MM atoms to which it is bonded. Setting-up a QM/MM simulation by hand is thus a tedious task that turns impossible as soon as the system of interest contains more than a few tens of atoms. Actually, in practice, one has generally prepared and investigated the system at the MM level before switching to QM/MM, to investigate, for example, a chemical reaction energy profile or to evaluate electronic properties at a QM level of theory. Highly optimized packages for classical molecular simulations like NAMD, LAMMPS and GROMACS to name a few, are readily available to generate the topology of the systems and to set-up classical MM simulations from atom geometries in standard formats (e.g. the Protein Data Bank PDB format). A powerful strategy is therefore to translate the information contained in the topology files provided by one's favorite MM package to a deMon2k compatible input file. To this end, we have devised a dedicated program, called QIB (for QM/MM Input Builder). QIB is currently capable of converting topology files from the AMBER package, CHARMM and GROMACS to

the deMon2k format. QIB also permits to create a QM/MM input file from scratch, provided a reference PDB file. Another essential capability of QIB is to take care of a series of technical but important details appearing in QM/MM simulations.

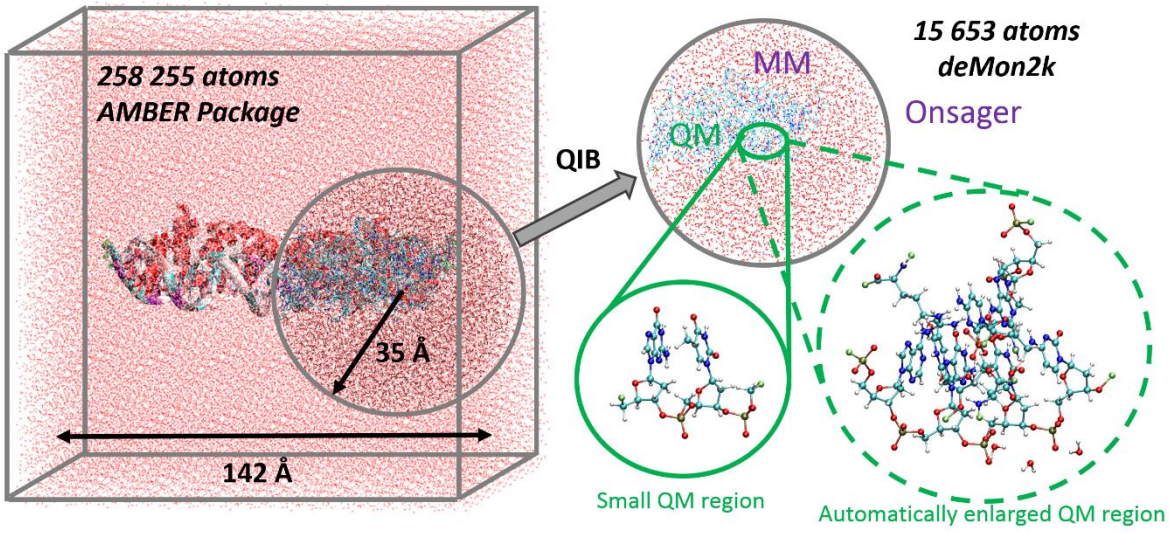


Figure 13: Pictogram for the automatically generation of deMon2k input files from standard MM package topology and geometry files by QIB for QM/MM simulations. The program generates a minimal deMon2k input file and a xyz file with the geometry of the QM region encompassing link atoms.

To illustrate the capabilities of QIB, we take the example of a QM/MM simulation of a DNA/protein complex shown in Fig. 13. The latter has been solvated in a of 142 Å edge box of TIP3P water molecules and contains 258 550 atoms. QIB extracts important information from the AMBER prmtop file such as the atom names, the residue pointers and the connectivity. In a second step, it identifies for each atom the corresponding atom type within the deMon2k FFDS file. Together with a coordinate file that can be provided under various formats, this is all the information needed to prepare the geometry block in deMon2k. As currently deMon2k does not implement periodic boundary conditions but can account for remote electrostatic effect with the Onsager polarizable model (see Section 4), it is advisable to prune the system to fit within spheres, typically with a 30 to 40 Å radius. Using the information on the residues to which each atom belongs, the pruning can be operated keeping the molecular integrity of each residue. In the present case, this leads to a molecular system encompassing around 15,600 atoms. For QM/MM

simulations, the user can provide an atom selection that defines a primary QM region. Eventually, the latter is enlarged by a distance criterion, as illustrated in Fig. 13. This option is useful to include solvation waters close to the primary QM region, or to simply generate a QM/MM set-up without the need to identify all atom indexes defining the QM region. A geometry file for the QM region is produced for checking. Note that QIB automatically handles the case of covalent bonds appearing across the QM/MM border. It further suggests a list of atom types for QM atoms in order to calculate Lennard-Jones interactions with MM atoms. Other options are available in QIB, and example of QM/MM simulations with the molecular systems depicted in Fig. 13 are reported in Chapter four of this book.

8. Résumé

In this chapter, we tried to give an overview over the current state-of-the-art of the QM/MM implementation in deMon2k and the corresponding input support program QIB. As this chapter shows the QM/MM implementation in deMon2k is a community effort and as such has various perspectives. Although most basic computational tasks such as QM/MM structure optimizations, molecular dynamics simulations and property calculations are now readily available within the QM/MM module of deMon2k, several larger implementation efforts are still under way. We tried to give a snapshot of this situation in this chapter. As a community, our aim is to provide a QM/MM approach that is easy to use and yet powerful enough to address challenging problems in chemistry, materials science and biology. Because our developments are in the framework of deMon2k, they will be general accessible, free of charge, to the academic community.

Acknowledgements

Support from the SENER-CONACyT project 280158 and the CONACyT project A1-S-11929 is gratefully acknowledged. For the QM/MM deMon2k program development computational

resources for the CONACyT infrastructure project GIC-268251 and the SEP-Cinvestav project No. 65 are used.

References

[1] A. Warshel and M. Karplus, *J. Am. Chem. Soc.*, 1972, **94**, 5612.

[2] A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.

[3] P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **B864**, 136.

[4] W. Kohn and L.J. Sham, *Phys. Rev.*, 1965, **A1133** 140.

[5] B.I. Dunlap, J.W.D. Connolly and J.R. Sabin, *J. Chem. Phys.*, 1979, **71**, 4993.

[6] J.W. Mintmire and B.I. Dunlap, *Phys. Rev. A.*, 1982, **25**, 88.

[7] O. Vahtras, J. Almlöf and M.W. Feyereisen, *Chem. Phys. Lett.*, 1993, **213**, 514.

[8] B.I. Dunlap, N. Rösch and S.B. Trickey, *Mol. Phys.*, 2010, **108**, 3167.

[9] H. Früchtl, R.A. Kendall, R.J. Harrison and K.G. Dyall, *Int. J. Quant. Chem.*, 1997, **64**, 63.

[10] S. Hamel, M.E. Casida and D.R. Salahub, *J. Chem. Phys.*, 2001, **114**, 7342.

[11] F. Weigend, *Phys. Chem. Chem. Phys.*, 2002, **4**, 4285.

[12] R. Polly, H.-J. Werner, F.R. Manby and P.J. Knowles, *Mol. Phys.*, 2004, **102**, 2311.

[13] D. Mejía-Rodríguez and A.M. Köster, *J. Chem. Phys.*, 2014, **141**, 124114.

[14] S.F. Manzer, E. Epifanovsky and M. Head-Gordon, *J. Chem. Theory Comput.*, 2015, **11**, 518.

[15] F.A. Delesma, G. Geudtner, D. Mejía-Rodríguez, P. Calaminici and A.M. Köster, *J. Chem. Theory Comput.*, 2018, **14**, 5608.

[16] A.M. Köster, G. Geudtner, A. Alvarez-Ibarra, P. Calaminici, M.E. Casida, J. Carmona-Espíndola, F.A. Delesma, R. Delgado-Venegas, V.D. Domínguez, R. Flores-Moreno, G.U. Gamboa, A. Goursot, T. Heine, A. Ipatov, A. de la Lande, F. Janetzko, J.M. del Campo, N. Pedroza Montero, L.G.M. Pettersson, D. Mejía-Rodríguez, J. Reveles, J. Vásquez-Pérez, A. Vela, B. Zúñiga-Gutiérrez and D.R. Salahub, deMon2k; Version 6.1.5; The deMon developers: Cinvestav, Mexico City, 2020.
See also: http://www.demon-software.com.

[17] M.E. Casida, C. Daul, A. Goursot, A.M. Köster, L.G.M. Pettersson, E. Proynov, A. St-Amant, D.R. Salahub, H. Duarte, N. Godbout, J. Guan, C. Jamorski, M. Leboeuf, V.G. Malkin, O.L. Malkina, F. Sim and A. Vela, deMon-KS Version 3.4, deMon Software, Montréal (1996).

[18] J. Andzelm and E. Wimmer, *J. Chem. Phys.*, 1992, **96**, 1280.

[19] H. Sambe and R.H. Felton, *J. Chem. Phys.*, 1975, **62**, 1122.

[20] D.R. Salahub, J. Weber, A. Goursot, A.M. Köster and A. Vela, Applied Density Functional Theory and the deMon Code, 1964 - 2004, in *Theory and Applications of the Computational Chemistry: The First 40 Years*, Editors: C.E. Dykstra, G. Frenking, K.S. Kim, G. Scuseria, Elsevier, Amsterdam, 2005

[21] D.N. Laikov, *Chem. Phys. Lett.*, 1997, **281**, 151.

[22] A.M. Köster, J.U. Reveles and J.M. del Campo, *J. Chem. Phys.*, 2004, **121**, 3417.

[23] U. Birkenheuer, A.B. Gordienko, V.A. Nasluzov, M.K. Fuchs-Rohr and N. Rösch, *Int. J. Quant. Chem.*, 2005, **102**, 743.

[24] A.V. Bienvenu and G. Knizia, *J. Chem. Theory Comput.*, 2018, **14**, 1297.

[25] P. Calaminici, A. Alvarez-Ibarra, D. Cruz-Olvera, V.D. Dominguez-Soria, R. Flores-Moreno, G.U. Gamboa, G. Geudtner, A. Goursot, D. Mejía-Rodríguez, D.R. Salahub, B. Zuniga-Gutierrez and A.M. Köster, Auxiliary Density Functional Theory: From Molecules to Nanostructures, in *Handbook of Computational Chemistry*, Editors: J. Leszczynski, A. Kaczmarek-Kedziera, T. Puzyn, M. G. Papadopoulos, H. Reis and M.K. Shukla, Springer International Publishing: Chapter 18, pp. 795, 2nd edition, 2017.

[26] S.-C.T. Choi, C.C. Paige and M.A. Saunders, *SIAM J. Sci. Comput.*, 2011, **33**, 1810.

[27] T. Eirola and O. Nevanlinna, *Linear Algebra and its Applications*, 1989, **121**, 511.

[28] J.N. Pedroza-Montero, J.L. Morales, G. Geudtner, A. Álvarez-Ibarra, P. Calaminici and A.M. Köster, *J. Chem. Theory Comput.*, 2020, **16**, 2965

[29] D. Mejía-Rodríguez, R.I. Delgado-Venegas, P. Calaminici and A.M. Köster, *J. Chem. Theory Comput.*, 2015, **11**, 1493.

[30] A.M. Köster, *J. Chem. Phys.*, 2003, **118**, 9943.

[31] A. Álvarez-Ibarra, *Asymptotic Expansion of Molecular Integrals in Self-Consistent Auxiliary Density Functional Methods*, Ph.D. thesis, Cinvestav, 2013.

[32] A.M. Köster, J.M. del Campo, F. Janetzko and B. Zuniga-Gutierrez, *J. Chem. Phys.*, 2009, **130**, 114106.

[33] P. Calaminici, F. Janetzko, A.M. Köster, R. Mejia-Olvera and B. Zuniga-Gutierrez, *J. Chem. Phys.*, 2007, **126**, 044108.

[34] D. Mejía-Rodríguez, X. Huang, J.M. del Campo and A.M. Köster, *Adv. Quantum Chem.*, 2015, **71**, 41.

[35] D. Mejía-Rodríguez, *Low-Order Scaling Methods for Auxiliary Density Functional Theory*, Ph.D. thesis, Cinvestav, 2015.

[36] F.A. Delesma-Díaz, *Range-Separated Hybrid Functionals in Auxiliary Density Functional Theory*, Ph.D. thesis, Cinvestav, 2020.

[37] F. Aquilante, T. Pedersen, A. Merés and H. Koch, *J. Chem. Phys.*, 2006, **125**, 174101.

[38] S. F. Boys, *Rev. Mod. Phys.*, 1960, **32**, 296.

[39] J. M. Foster and S. F. Boys, *Rev. Mod. Phys.*, 1960, **32**, 300.

[40] J.N. Pedroza-Montero, F.A. Delesma, J.L. Morales, P. Calaminici and A.M. Köster, *J. Chem. Phys.*, 2020, **153**, 134112.

[41] S. Obara and A. Saika, *J. Chem. Phys.* 1986, **84**, 3963.

[42] A. de la Lande, A. Alvarez-Ibarra, K. Hasnaoui, F. Cailliez, X. Wu, T. Mineva, J. Cuny, P. Calaminici, L. López-Sosa, G. Geudtner, I. Navizet, C. Garcia Iriepa, D.R. Salahub and A.M. Köster, *Molecules*, 2019, **24**, 1653.

[43] C. Quintanar, R. Caballero and A.M. Köster, *Int. J. Quantum Chem.*, 2004, **96**, 483.

[44] D.R. Salahub, S.Y. Noskov, B. Lev, R. Zhang, V. Ngo, A. Goursot, P. Calaminici, A.M. Köster, A. Alvarez-Ibarra, D. Mejía-Rodríguez, J. Řezáč, F. Cailliez and A. de la Lande, *Molecules*, 2015, **20**, 4780.

[45] A. Alvarez-Ibarra, A.M. Köster, R. Zhang and D.R. Salahub, *J. Chem. Theory Comput.*, 2012, **8**, 4232.

[46] A. Alvarez-Ibarra and A.M. Köster, *J. Chem. Phys.*, 2013, **139**, 024102.

[47] P.A.M. Dirac, *Proc. Cambr. Phil. Soc.*, 1930, **26**, 376.

[48] S.H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 1980, **58**, 1200.

[49] J.P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865; ibid, 1997, **78**, 1396(E).

[50] R. Flores-Moreno, R.J. Alvarez-Mendez, A. Vela and A.M. Köster, *J. Comput. Chem.*, 2006, **27**, 1009.

[51] N. Jardillier and A. Goursot, *Chem. Phys. Lett.*, 2008, **454**, 65.

[52] B. Zuniga-Gutierrez and A.M. Köster, *Mol. Phys.*, 2016, **114**, 1026.

[53] S.V. Shedge, J. Carmona-Espíndola, S. Pal and A.M. Köster, *J. Phys. Chem. A*, 2010, **114**, 2357.

[54] A. Goursot, T. Mineva, R. Kevorkyants and D. Talbi, *J. Chem. Theory Comput.*, 2007, **3**, 755.

[55] P. Jurečka, J. Černý, P. Hobza and D.R. Salahub, *J. Comput. Chem.*, 2007, **28**, 555.

[56] S. Grimme, J. Anthony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

[57] S.E. Pérez-Figueroa, P. Calaminici and A.M. Köster, *J. Phys. Chem. A*, 2019, **123**, 4565.

[58] J.P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982.

[59] C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158.

[60] T.H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007.

[61] A. Alvarez-Ibarra and A.M. Köster, *Mol. Phys.*, 2015, **113**, 3128.

[62] A. Álvarez-Ibarra, P. Calaminici, A. Goursot, C.Z. Gómez-Castro, R. Grande-Aztatzi, T. Mineva, D.R. Salahub, J.M. Vásquez-Pérez, A. Vela, B. Zuniga-Gutierrez and A.M. Köster, First Principles Computational Biochemistry with deMon2k in *Frontiers in Computational Chemistry*, Volume 2, Zaheer-Ul_Haq and J.D. Madura (Eds), Bentham Science Publishers (eBook), pp 281, 2015.

[63] M. Krack and A.M. Köster, *J. Chem. Phys.*, 1998, **108**, 3226.

[64] A.M. Köster, R. Flores-Moreno and J.U. Reveles, *J. Chem. Phys.*, 2004, **121**, 681.

[65] A.M. Köster, G. Geudtner, G.U. Gamboa, A. Alvarez-Ibarra, P. Calaminici, R. Flores-Moreno, A. Goursot, A. de la Lande, D. Mejia-Rodriguez, T. Mineva, L.G.M. Pettersson, J.M. Vasquez-Perez, B. Zuniga-Gutierrez, S.B. Trickey and D.R. Salahub, *The deMon2k Users' Guide, Version 5.0*, Cinvestav, Mexico-City, 2018.
See also: http://www.demon-software.com.

[66] W.L. Jorgensen, D.S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*,2015, **118**, 11225.

[67] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, Proteins, 2006, 65, 712.

[68] P. Pulay, *Mol. Phys.*, 1969, **17**, 197.

[69] S.T. Epstein, *The Variation Method in Quantum Chemistry*, Academic Press London, 1974.

[70] A. Dalgarno and A.L. Stewart, *Proc. Roy. Soc. (London) A*, 1956, **238**, 269.

[71] M. Born and J.R. Oppenheimer, *Ann. Phys.*, 1927, **389**, 457.

[72] R. Fletcher, *Practical methods of optimization*, John Wiley & Sons, Chichester, N.Y. 1980.

[73] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.

[74] H.B. Schlegel, *J. Comput. Chem.*, 1982, **3**, 214.

[75] K.P. Lawley, *Advances in Chemical Physics: Ab Initio Methods in Quantum Chemistry Part I*, John Wiley & Sons, Chichester, N.Y., 1987.

[76] D.R. Yarkony, *in Modern Electronic Structure Theory: Part I*, World Scientific, Singapore, 1995.

[77] P. von R. Schleyer, *in Encyclopedia of Computational Chemistry*, Wiley, New York, 1998.

[78] Ö. Farkas and H.B. Schlegel, *J. Chem. Phys.*,1998, **109**, 7100.

[79] Ö. Farkas and H.B. Schlegel, *J. Chem. Phys.*, 1999, **111**, 10806.

[80] Ö. Farkas and H.B. Schlegel, *Phys. Chem. Chem. Phys.*, 2002, **4**, 11.

[81] H.B. Schlegel, *J. Comput. Chem.*, 2003, **24**, 1514.

[82] F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Chichester, NY, 2007.

[83] H.B. Schlegel, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 790.

[84] E.G. Lewars, *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, Springer-Verlag, Dordrecht, 2011.

[85] J.M. del Campo and A.M. Köster, *Croatica Chimica Acta*, 2009, **82**, 283.

[86] K. Levenberg, *Quart. Appl. Math.*, 1944, **2**, 164.

[87] D.W. Marquardt, *SIAM J.*, 1963, **11**, 431.

[88] C.G. Broyden, *J. Inst. Math. Appl.*, 1970, **6**, 76.

[89] R. Fletcher, *Comput. J.*, 1970, **13**, 317.

[90] D. Goldfarb, *Math. Comput.*, 1970, **24**, 23.

[91] D.F. Shanno, *Math. Comput.*, 1970, **24**, 647.

[92] P. Culot, G. Dive, V.H. Nguyen and J.M. Ghuysen, *Theor. Chim. Acta*, 1992, **82**, 189.

[93] J.M. Bofill, *J. Comput. Chem.*, 1994, **15**, 1.

[94] J.M. del Campo and A.M. Köster, *J. Chem. Phys.*, 2008, **129**, 024107.

[95] M.J.D. Powell, *Math. Program*, 1971, **1**, 26.

[96] M.E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, Oxford, 2010.

[97] H.F. Trotter, *Proc. Am. Math. Soc.*, 1959, **10**, 545.

[98] W.C. Swope, H.C. Anderson, P.H. Berens and K.R. Wilson, *J. Chem. Phys.*, 1982, **76**, 637.

[99] G.U. Gamboa, J.M. Vázquez-Pérez, P. Calaminici and A.M. Köster, *Int. J. Quantum Chem.*, 2010, **110**, 2172.

[100] S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511.

[101] G.J. Martyna, M. Klein and M. Tuckerman, *J. Chem. Phys*. 1992, **97**, 2635.

[102] W.D. Cornell, P. Cieplak, C.I. Bayly, I R. Gould, K.M. Merz, Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell and P.A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179.

[103] T. Mineva and N. Russo, *Int. J. Quantum Chem.*, 1997, **61**, 665.

[104] T.E. Cheatham III, P. Cieplak and P.A. Kollman, *J. Biomol. Struct. Dyn.*, 1999, **16**, 845.

[105] J. Wang, P. Cieplak and P.A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049.

[106] G.J. Martyna, D. Tobias and M. Klein, *J. Chem. Phys*., 1994, **101**, 4177.

[107] M. Tuckerman, J. Alejandre, R. López-Rendón and A.L. Jochim, *J. Phys. A: Math. Gen.*, 2006, **39**, 5629.

[108] T. Bredow, G. Geudtner and K. Jug, *J. Comput. Chem.*, 2001, 22, 89.

[109] F. Janetzko, A.M. Köster and D.R. Salahub, *J. Chem. Phys.*, 2008, **128**, 024102.

[110] F. Janetzko, T. Bredow, G. Geudtner and A.M. Köster, *J. Comput. Chem.*, 2008, **29**, 2295.

[111] S.W. de Leeuw, J.W. Perram and E.R. Smith, *Proc. R. Soc. Lond. A*, 1980, **373**, 27.

[112] A.Y. Toukmaji and J.A. Board Jr., *Comput. Phys. Commun.*, 1996, **95**, 73.

[113] M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*, 2nd edition, Oxford University Press, Oxford, 2017.

[114] B. Nijboer and F. De Wette, *Physica*, 1957, **23**, 309.

[115] J.A. Rackers, M.L. Laury, C. Lu, Z. Wang, L. Lagardère, J.-P. Piquemal, P. Ren and J.W. Ponder, *J. Chem. Theory Comput.*, 2018, **14**, 5273.

[116] Y. Kawashima, K. Ishimura and M. Shiga, *J. Chem. Phys.*, 2019, **150**, 124103.

[117] R. I. Delgado-Venegas, D. Mejía-Rodríguez, R. Flores-Moreno, P. Calaminici and A. M. Köster, J. Chem. Phys., 2016, **145**, 224103.

[118] R. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York, 1989.

[119] A. M. Köster, *J. Chem. Phys.*, 1996, **104**, 4114.

[120] G. Diercksen and R. McWeeny, *J. Chem. Phys.*, 1966, **44**, 3554.

[121] R. McWeeny and G. Diercksen, *J. Chem. Phys.*, 1968, **49**, 4852.

[122] J.L. Dodds, R. McWeeny, W.T. Raynes and J.P. Riley., *Mol. Phys.*, 1977, **33**, 611.

[123] J.L. Dodds, R. McWeeny and A.J. Sadlej, *Mol. Phys.*, 1977, **34**, 1779.

[124] R. McWeeny, *Methods of Molecular Quantum Mechanics*, 2nd edition, Academic Press, London, 2001.

[125] R. Flores-Moreno and A. M. Köster, *J. Chem. Phys.*, 2008, **128**, 134105.

[126] J. Carmona-Espíndola, R. Flores-Moreno and A. M. Köster, *J. Chem. Phys.*, 2010, **133**, 084102.

[127] J. Carmona-Espíndola, R. Flores-Moreno and A. M. Köster, *Int. J. Quantum Chem.*, 2012, **112**, 3461.

[128] R. Flores-Moreno, J. Melin, J. V. Ortiz and G. Merino, *J. Chem. Phys.*, 2008, **129**, 224105.

[129] R. Flores-Moreno, *J. Chem. Theory Comput.*, 2010, **6**, 48.

[130] A. Neumaier, *SIAM Review*, 1998, **40**, 636.

[131] U.M. Yang and K.A. Gallivan, *Applied Numerical Mathematics*, 1995, **19**, 287.

[132] G. Geudtner, F. Janetzko, A. M. Köster, A. Vela and P. Calaminici, *J. Comp. Chem.*, 2006, **27**, 483.

[133] G. Geudtner, P. Calaminici, J. Carmona-Espíndola, J.M. del Campo, V.D. Dominguez-Soria, R. Flores-Moreno, G.U. Gamboa, A. Goursot, A.M. Köster, J.U. Reveles, T. Mineva, J.M.

Vasquez-Perez, A. Vela, B. Zuniga-Gutierrez and D.R. Salahub, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2012, **2**, 548.

[134] E.B. Wilson, J.C. Decius and P.C. Cross, *Molecular vibrations: The theory of infrared and Raman vibrational spectra*, Dover, New York, 1980.

[135] J. Neugebauer, M. Reiher, C. Kind and B.A. Hess, *J. Comp. Chem.* , 2002, **23**, 895.

[136] R.I. Delgado-Venegas, P. Calaminici and A.M. Köster, *Mol. Phys.*, 2019, **117**, 1367.

[137] Q. Cui and M. Karplus, *J. Chem. Phys.*, 2000, **112**, 1133.

[138] Q. Cui and M. Karplus, *J. Phys. Chem. B*, 2000, **104**, 3721.

[139] B. Zuniga-Gutierrez, G. Geudtner and A.M. Köster, *J. Chem. Phys.*, 2011, **134**, 124108.

[140] B. Zuniga-Gutierrez, G. Geudtner and A. M. Köster, *J. Chem. Phys.*, 2012, **137**, 094113.

[141] B. Zuniga-Gutierrez, M. Camacho-Gonzalez, P. Simon-Bastida, A. Bendana-Castillo, P. Calaminici, and A.M. Köster, *J. Phys. Chem. A*, 2014, **119**, 1469.

[142] B. Zuniga-Gutierrez, M. Camacho-Gonzalez, P. Simon-Bastida, A. Bendana-Castillo, P. Calaminici and A.M. Köster, *J. Chem. Phys.*, 2015, **143**, 104103.

[143] B. Zuniga-Gutierrez, R. Flores-Moreno, V. Medel-Juarez, A. Varona and H.N. González-Ramírez, *J. Chem. Phys.*, 2020, **152**, 014105.

[144] R. Ditchfield, *J. Chem. Phys.*, 1972, **56**, 5688.

[145] B.A. Zuniga-Gutierrez, *Magnetic Shielding, Spin-Spin Coupling Constant and Magnetizability Tensors Calculated with Auxiliary Density Perturbation Theory*, Ph.D. thesis, Cinvestav, 2011.

[146] A. Ipatov, A. Fouqueau, C. Perez del Valle, F. Cordova, M.E. Casida, A.M. Köster, A. Vela, *J. Mol. Structure: THEOCHEM*, 2006, **762**, 179.

[147] M.E. Casida, Time-dependent density functional response theory for molecules, in *Recent Advances in Density Functional Methods (Part I)*, Editor: D. P. Chong, World Scientific: Singapore: Chapter 5, pp 155, 1995.

[148] D. Jacquemin, E.A. Perpète, G.E. Scuseria, I. Ciofini and C. Adamo, *J. Chem. Theory. Comput.*, 2008, **4**, 123.

[149] Y. Shao, Y. Mei, D. Sundholm and V.R.I. Kaila, *J. Chem. Theory Comput.*, 2020, **16**, 587.

[150] C. Jamorski, M.E. Casida and D.R. Salahub, *J. Chem. Phys.*, 1996, **104**, 5134.

[151] J. Carmona-Espíndola and A.M. Köster, *Can. J. Chem.*, 2013, **91**, 795.

[152] S. Hirata and M. Head-Gordon, *Chem. Phys. Lett.*, 1999, **314**, 291.

[153] E. Davidson, *J. Comput. Phys.*, 1975, **17**, 87.

[154] G. Cisneros, M. Berrondo, C.F. Bunge, *Comp. Chem.*, 1986, **10**, 281.