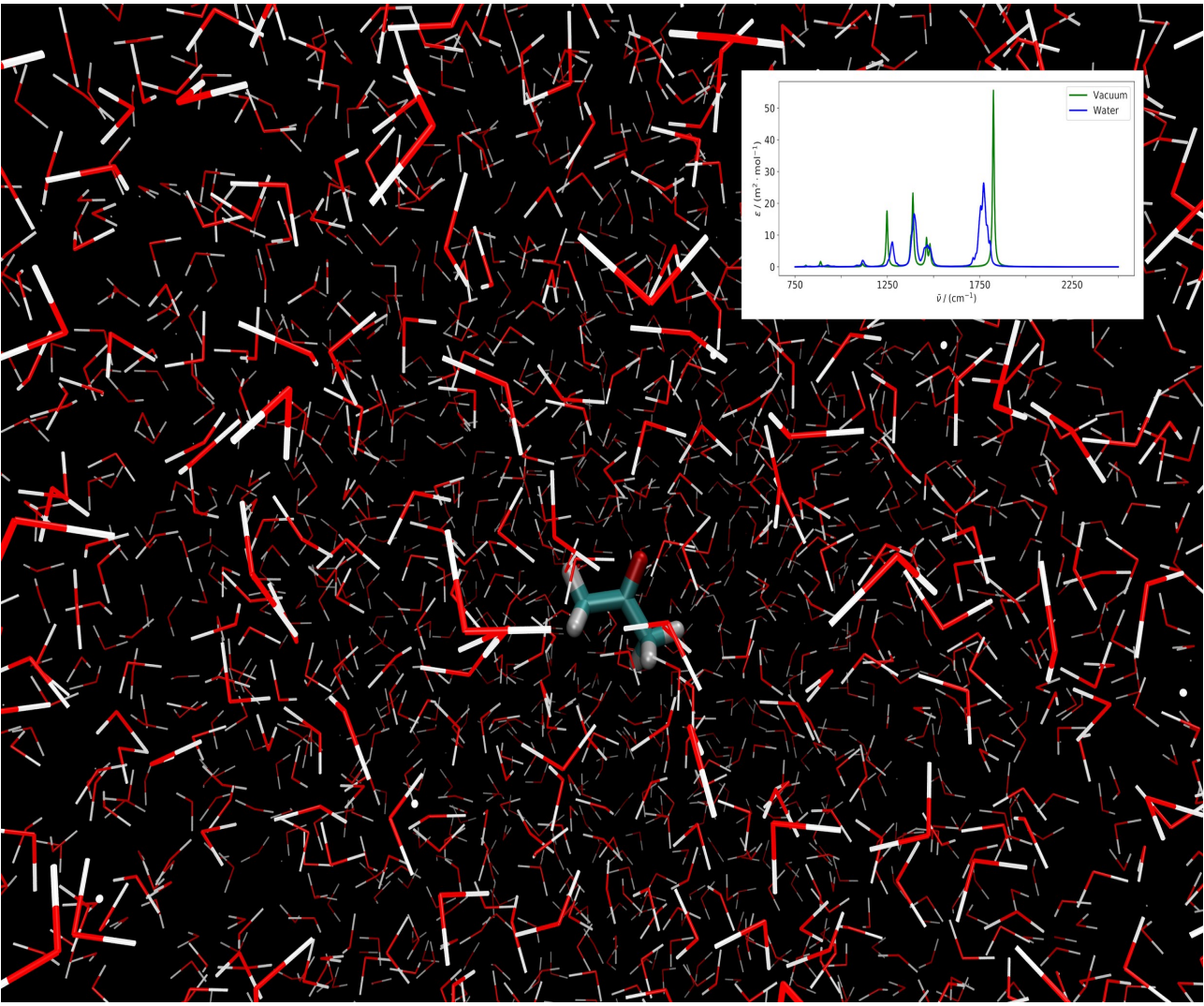
UiT The Arctic University of Norway

Faculty of Science and Technology, Department of Chemistry

Vibrational spectroscopy beyond the harmonic approximation with the Polarizable Embedding model

Karen Oda Hjorth Minde Dundas

A dissertation for the degree of Philosophiae Doctor, December 2020



Abstract

In order to efficiently and accurately calculate vibrational properties for solvated systems, a theoretical framework for combining response theory with the Polarizable Embedding model (PE) has been derived and implemented, and is presented in this thesis. An open-ended recursive formalism is utilized through the implementation in OpenRSP, allowing energy-derivatives to be calculated analytically up to arbitrary order. In this way, errors associated with numerical differentiation are avoided, and calculations of properties of higher order can be performed without the need for additional implementation. The PE model is a focused embedding model, and includes solvent effects through both static and instantaneous interaction energies between a central molecular region and a surrounding environment. The central region is modelled using quantum-mechanical methods (and therefore is commonly referred to as the QM region), whereas the environment is treated through classical multipoles and polarizabilities. The multipoles and polarizabilities are placed on so-called sites, typically located on the atoms in the environment. This retains a discrete atomistic model while still considerably reducing the overall computational cost compared to a full quantum mechanical description. The method allows the user to automatically calculate the multipoles and polarizabilities, thereby avoiding any need for using predetermined parameters. In addition to its efficiency, the PE method is thus both accurate and flexible. In this thesis, a new combination of the PE method with the open-ended response framework is presented, in addition to the implementation done in order to calculate various vibrational frequencies and intensities from the calculated energy derivatives. The theoretical background for derivatives of the PE energy is outlined, and software for the calculation of these has been developed

through a combination of the LSDalton, OpenRSP and FraME programs. A new software package, SpectroscPy, has been developed in order to calculate spectroscopic frequencies and intensities, for either vacuous or solvated systems. The current version of the program allows the user to produce IR, Raman and hyper-Raman spectra. The implementation can be extended to better model repulsion interactions between the QM region and the environment through the Polarizable Density Embedding method. Newer versions of the implementation should also contain the functionality necessary to perform calculations on biomolecules, as it for the moment only allows systems where no covalent bonds need to be cut in the partition of the QM region from the environment, and/or between the environmental fragments. Once magnetic derivatives are available through OpenRSP, implementation of vibrational-rotational spectroscopic properties is also a route that should be investigated.

Contents

List of Publications

This thesis is based on the following scientific publications.

Articles:

I **K. O. H. M. Dundas**, M. T. P. Beerepoot, M. Ringholm, S. Reine, R. Bast, N. H. List, J. Kongsted, K. Ruud and J. M. H. Olsen, "Harmonic Infrared and Raman Spectra in Molecular Environments using the Polarizable Embedding Model", *J. Chem. Theory Comput.*, (submitted).

II **K. O. H. M. Dundas**, S. S. Reine, B. Gao, K. Ruud, J. M. H. Olsen, M. Ringholm, "Analytic High-order Geometric Derivatives with Polarizable Embedding in a Response Theory Framework", (manuscript).

III J. M. H. Olsen, *et al.*, "Dalton Project: A Python platform for molecular- and electronic-structure simulations of complex systems", *J. Chem. Phys.* **152** (2020), 214115.

Software:

IV **K. O. H. M. Dundas**, M. Ringholm, Y. Coraton, B. Ofstad, "SpectroscPy: The Python Package for Vibrational Spectroscopy" (2020), https://doi.org/10.5281/zenodo.3712927

V J. M. H. Olsen, **K. O. H. M. Dundas**, M. Ringholm, "FraME: Fortran library for Fragment-based Multiscale Embedding" (2020), https://gitlab.com/FraME-projects/FraME

VI Bakken, *et al.*, "LSDalton - An electronic structure program" (2018), git@gitlab.com:dalton/lsdalton.git

Acknowledgements

My work with this PhD has been a process that has gone on over the duration of many years, as for everyone else attempting to earn this degree. Looking back at this time, it has been an eventful journey of learning and growth, both with regards to my subject, but also as a human still in the early years of adulthood. This journey is not one I have trodden alone, and I am therefore happy that I get this space to give a nod to some of those that have shared it with me.

When I first arrived in Tromsø, I was greeted by a number of people that I have come to know and appreciate, even though I at first struggled with the fact that an overwhelming number of them had a first name starting with the letter "M". The first was my main supervisor Kenneth Ruud, who by offering me this position has given me the opportunity to delve into the fascinating field that the rest of this thesis is about. His wealth of knowledge is something that I will be continue to be amazed by, and am grateful to have gotten to learn from. In addition to Kenneth, I have had a set of excellent co-supervisors, each with their own scientific specialities and general abilities. This team consists of Maarten T. P. Beerepoot, Magnus Ringholm and Jógvan Magnus Haugaard Olsen. I have learned so much from each of you, both scientifically and about the world within research in general. I would also like to thank my colleagues and friends Marius, Michal and Magnar, who in each their separate ways have been very important to me. Additionally, my co-authors that have not yet been mentioned, namely Simen Reine, Bin Gao, Radovan Bast, Nanna Holmgaard List and Jacob Kongsted, also receive my thanks.

My gratitude goes to the Norwegian Supercomputing Program (NOTUR) for providing invaluable computational resources, the Norwegian Research Council (NFR) for funding my research through the MultiDim project (grant nr. 250743), as well as for their grant of a Center of Excellence to the Hylleraas center, to which I have been associated. I would also like to thank John Kendrick and Julien Bloino, who have both gone out of their way when it comes to discussing scientific topics posed by an unknown PhD student.

Outside campus, I would like to especially thank my good friends, Hanna, Manuel and Jenny. Thanks for still being there even though I have

been in a thesis-haze these last few months. And then of course, my dear parents, Siv and Bjørn. This is the second time I thank you in a thesis, and your support through both of these times in my life has been completely crucial. Even though you are miles away, I don't know what I would do without you. Then finally, to my husband Steinar, you have been running a household alone the last half year, and allowed me to come home at strange hours, but always had dinner ready. You know what you mean to me, and the support you have given me through this time.

Abbreviations

The following abbreviations are all defined as they appear in the text, but are summarized here for reference.

AO	Atomic Orbital
CC	Coupled Cluster
CI	Configuration Interaction
DFT	Density Functional Theory
DVPT2	Deperturbed Second-Order Vibrational Perturbation Theory
FDE	Frozen Density Embedding
FQ	Fluctuating Charges
FWHM	Full Width at Half Maximum
GVPT2	Generalized Second-Order Vibrational Perturbation Theory
HF	Hartree-Fock
IR	Infrared
KS–DFT	Kohn-Sham Density Functional Theory
LJ	Lennard-Jones
MCSCF	Multi-Configurational SCF
MD	Molecular Dynamics
MFCC	Molecular Fractionation with Conjugate Caps
MM	Molecular Mechanics
MO	Molecular Orbital
PCM	Polarizable Continuum Model
PDE	Polarizable Density Embedding
PE	Polarizable Embedding
QM	Quantum Mechanic
QM/MM	Quantum Mechanic/Molecular Mechanic
ROA	Raman Optical Activity
SCF	Self-Consistent Field
SEP	Solvent Embedding Potential
SI	International System of Units
VCD	Vibrational Circular Dichroism
VPT2	Second-order Vibrational Perturbation Theory

Introduction

For more than 200 years, the laws of Newton were believed to hold for all matter, but as the discovery of the building blocks of matter progressed in the early years of the 20th century, it became clear that when the system or particle studied was small enough, or had a high enough speed, other rules applied.[1–3] The discoveries made by the pioneers of quantum mechanics (QM) revolutionized the way we look at nature, and provided us with ground-breaking insights such as that all matter has both wave- and particle-like properties, and that speed and position cannot be simultaneously determined, through no fault of the measuring equipment.[1,4–7]

When studying chemistry on a microscopic level, it is necessary to incorporate the insights from QM, leading us to the field of quantum chemistry, which can be viewed as the tool for calculating chemically relevant properties of electrons, protons and neutrons in molecular environments. Although it is possible to accurately predict the behaviour of extremely small systems, such as the hydrogen atom, through analytic formulas, the complexity of chemically interesting systems makes quantum chemistry primarily a field concerned with the development of methods and approximations. This is because even for even slightly larger systems, the number of particles and the number of interactions between them make it impossible to solve the equations used for the hydrogen atom.[8] Most of the work happening in quantum chemistry today therefore involves making and improving approximations, and approximations of approximations etc.[2], that lead to methods for finding useful information about chemical systems within reasonable time frames and requiring reasonable amounts of computer power.[8]

One of the earliest of such methods was the Hartree-Fock (HF)[9,10] method that approximated these interactions by treating all other electrons than the one considered as an averaged surrounding field, and (as is common

in most QM methods) all the protons in each atom as one combined nucleus with a positive charge equal to the sum of the proton-charges. Although useful in certain cases, for most chemical problems, this approximation is too severe, giving results deviating from experimental results. Many different routes have been taken towards better models, where the Coupled Cluster (CC)[11] is in many cases the most accurate among the computationally feasible methods, while the Density Functional theory (DFT)[12] is widely popular because of its efficiency.

For small gaseous systems these QM methods work well, but interesting chemical systems are usually in liquid or solid bulk form, with a high number of molecules with strong interactions between them. Both the molecular geometry and other properties can be strongly influenced when it is surrounded by an environment, and including solvation effects is therefore of high importance.[13,14] At the same time, the increase in the number of particles and the number of interactions between them makes the use of QM methods practically impossible. It has therefore become necessary to develop methods that allow us to keep the a high degree of the accuracy of the QM methods, while at the same time not drastically increasing the computational burden when an environment is added, as compared to the isolated molecule. This leads us to the area of focused embedding methods, which exploits the fact that chemically relevant properties often are to some degree localized, which means that we can identify which area of the system contribute the most to the phenomenon we are studying. When studying a solute in the presence of a solvent, the molecular property will be localized to the solute, and is thus treated with a higher level of sophistication than its environment.

While the solute will typically be treated with a QM method, there exist many strategies when it comes to the environment. A popular strategy is to treat the molecule as existing inside a cavity in a continuous dielectric medium, as in the Polarizable Continuum Model (PCM).[15] Continuum models have proven to be highly efficient, and in many cases accurate, however, they are not able to model specific interactions between the environment and the molecule, such as for instance hydrogen bonds.[13] Discrete atomistic models have therefore become popular, and chief among them is the Quantum Mechanic/Molecular Mechanic (QM/MM)[16] method, as illustrated in Figure 1, which keeps the position and to some degree the

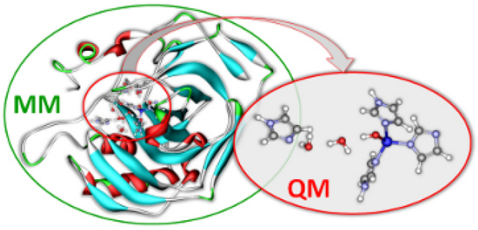


Figure 1. In QM/MM, the part of the system responsible for the chemical property observed is treated with QM methods, while the bulk is treated with MM methods.[18]

properties of the environmental atoms, but treats them using Molecular Mechanics (MM)[17], or classical laws.

Extensive work has been done within the field of QM/MM during the years, and in 2013 Warshel, Levitt and Karplus were awarded the Nobel prize in chemistry for their work on the subject[19,20]. There are many different ways to go about setting up a QM/MM method,[21] where one of the differences between these strategies is which molecular parameters are chosen to model the environmental atoms, and how these are obtained. This determines which interactions between the core and environmental regions are accounted for, and to what degree of accuracy. Electrostatic interactions are found by including permanent multipoles in the environment, while polarization effects are modelled by induced multipoles through distributed polarizabilites. In this work we focus on the method developed by Kongsted and co-workers, called the Polarizable Embedding (PE)[22] model. This method is based on both electrostatic and induced interactions between the core and the environment, and has the ability to calculate the necessary parameters for a given solvent through its implementations in PyFraME,[23] using the LoProp partitioning approach[24]. This makes it highly flexible with regards to choice of molecular environment, as it is independent of any pre-calculated parameters.

The PE method has proven a useful tool for both electronic[25,26] and magnetic[27,28] properties, and in the studies presented in this thesis, we extend the approach to the vibrational domain. Combined with the formalism expressing response theory as quasi-energy derivatives,[29,30] we explore harmonic and anharmonic vibrational properties of arbitrary order. There already exist a number of studies up to fourth-order properties related to

vibrational spectroscopy calculated with PE or PE-like methods.[22,25,31–36] What is unique to our approach is that it is analytical and open-ended with regards to order of energy derivatives, making any further implementation when proceeding to higher orders unnecessary, in addition to avoiding any errors introduced by numerical differentiation. This is possible through the recursive algorithms implemented in OpenRSP and developed by Ringholm and coworkers.[30,37]

As already noted, the focus of this thesis is on vibrational spectroscopy, which requires energy- and quasi-energy derivatives with respect to distortions of nuclear coordinates, called geometric derivatives, as well as derivatives with respect to electric field strengths experienced in the presence of an electric field. Through second-order geometric and up to fourth-order mixed electric and geometric derivatives, we are able to simulate vibrational spectroscopies such as Infrared (IR), Raman and Hyper-Raman within the Harmonic approximation.[38] Anharmonic effects are included through the second-order vibrational perturbation theory (VPT2).[39–44] This requires fourth-order geometric and up to sixth-order mixed electric and geometric derivatives, properties that have not previously been calculated with the PE model.

The aim of this doctoral thesis has been to develop the theory and computational tools for calculating vibrational spectroscopic properties with PE. In addition to serving as a background for the included papers, the introductory chapters are made with the hope that they will be useful for anyone entering the field and wishing to further develop or use the PE method on (vibrational) molecular property calculations. These chapters are therefore written to fit the knowledge level of someone having completed a Master's degree in theoretical chemistry.

The thesis includes three papers, **Paper I - III**, as well as creation of or involvement in three software programs. **Paper I** shows the application of the procedure on harmonic IR and Raman spectroscopy of acetone in three different solvents, while **Paper II** demonstrates the novel implementation of analytic higher-order geometric derivatives through anharmonic corrections of IR spectra, on two different molecular systems. **Paper III** was the first demonstration of parts of our software, although only on harmonic IR and Raman for vacuous benzene, as one of many contributions to the creation a Python platform created to make the Dalton[45,46] and LSDalton[45,47]

programs more accessible to the general user.

The rest of the text is organized as follows. The basic formulas and ideas of quantum mechanics are briefly described in Chapter 1. Chapter 2 covers a detailed description of the theory necessary for vibrational spectroscopy calculations, while Chapter 3 covers the Polarizable Embedding model, both basic theory and with the application to vibrational spectroscopies. An overview of the software developed and used is found in Chapter 4, followed by a summary of the publications making up this work in Chapter 5 and finally an outlook towards future developments and applications of this work in Chapter 6.

Chapter 1

Quantum mechanics

This part is a recapitulation of quantum chemistry concepts that it will be necessary to be familiar with in order to follow the more specific developments in later chapters. The more experienced quantum chemist is welcome to skip ahead to the more in-depth parts, as this chapter is specifically tailored to a beginning PhD student. For a further background in quantum chemistry, the reader is referred to the textbooks by Atkins and Friedman[1,8] and Helgaker, Jørgensen and Olsen[48]. These works have also been influential on this first chapter.

Central to quantum chemistry is the solution of the Schrödinger equation,[1]

$$\hat{H}\Psi = E\Psi \tag{1.0.1}$$

describing how the energy E of a system can be calculated as the eigenvalue of the Hamiltonian operator $\hat{H}$, with the wavefunction Ψ of the system being the eigenfunction. As mentioned in the introduction, accurate solutions to the Schrödinger equation is possible only for systems with very few particles, and developing strategies to solve it for larger systems can be said to have been the primary target of the many methods and approximations that have been developed within the field of quantum chemistry since its inception. Expressing the form of both $\hat{H}$ and Ψ is key to these developments.

We start by looking at $\hat{H}$ and some of its common approximations. A typical way of writing it is as the sum of operators for the kinetic $\hat{T}$ and potential $\hat{V}$ energies within the molecular system,

$$\begin{aligned}\hat{H}^{\text{mol}} =& \hat{T} + \hat{V} \\ =& -\frac{1}{2}\sum_{i}^{N_{\text{el}}} \nabla_i^2 - \frac{1}{2}\sum_{A}^{N_{\text{nuc}}} \frac{1}{m_A}\nabla_A^2 \\ &+ \frac{1}{2}\sum_{i}^{N_{\text{el}}}\sum_{j\neq i}^{N_{\text{el}}} \frac{1}{r_{ij}} - \sum_{i}^{N_{\text{el}}}\sum_{A}^{N_{\text{nuc}}} \frac{Z_A}{r_{iA}} + \frac{1}{2}\sum_{A}^{N_{\text{nuc}}}\sum_{B\neq A}^{N_{\text{nuc}}} \frac{Z_A Z_B}{r_{AB}}\end{aligned} \tag{1.0.2}$$

where the first two terms belong to $\hat{T}$ and the last three to $\hat{V}$. Note that this formula is written in so-called atomic units, where the elementary constants e (the elementary charge), $\hbar$ (the reduced Planck's constant), a_0 (the Bohr radius) and m_e (the electron mass) are all defined as equal to 1. The summation indices i and j run over the number of electrons N_{el} and A and B over the number of nuclei N_{nuc} in the system. Z_A is the charge of nucleus A and $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ where $\mathbf{r}_i$ is the position of particle i. It should be noted that Eq. (1.0.2) is valid only for the non-relativistic time-independent case.

This work, as is common in quantum chemistry, will keep to the Born-Oppenheimer approximation,[1,49] assuming that the nuclei are so much heavier than the electrons that they can by comparison be considered as stationary. In the electronic Hamiltonian the second term is therefore omitted and the fifth term is viewed as a constant nuclear repulsion term, h_{nuc}. The total molecular Hamiltonian, $\hat{H}^{\text{mol}}$, can thus be written as

$$\hat{H}^{\text{mol}} = -\frac{1}{2}\sum_{A}^{N_{\text{nuc}}} \frac{1}{m_A}\nabla_A^2 + \hat{H}^{\text{el}} \tag{1.0.3}$$

Although nuclear motion is central to this work, it is at this point sensible to look closer at the electronic Hamiltonian, $\hat{H}^{\text{el}}$, which notably also includes the nuclear repulsion term, h_{nuc}. Sorting the remaining terms based on their order of dependence on electronic position gives the electronic Hamiltonian as

$$\hat{H}^{\text{el}} = \hat{h} + \hat{g} + h_{nuc} \tag{1.0.4}$$

where the one-electron term $\hat{h}$ is given by

$$\hat{h} = -\frac{1}{2}\sum_i \nabla_i^2 - \sum_{iA}\frac{Z_A}{r_{iA}} \tag{1.0.5}$$

and the two-electron term $\hat{g}$ by

$$\hat{g} = \frac{1}{2}\sum_{i,j\neq i}\frac{1}{r_{ij}} \tag{1.0.6}$$

The wavefunction Ψ can be a function of time, position and the quantum-mechanical quantity of spin. The formalism presented in this work, however, is time-independent, and Ψ will therefore also be assumed independent of time, and consequently a function only of space and spin, $\Psi(\mathbf{r}, \mathbf{m}_s)$. It is common to choose Ψ as a combination of one-electron functions φ_k centered on the molecules, molecular orbitals (MOs). Like Ψ, φ_k is dependent on both position and space, and are therefore called spin-orbitals. Each can be factored into a spatial and a spin part [48]

$$\varphi_k(\mathbf{r}, \mathbf{m}_s) = \psi_k(\mathbf{r})\sigma_k(\mathbf{m}_s) \tag{1.0.7}$$

For electrons, the spin-quantum number m_s can be only $\frac{1}{2}$ or $-\frac{1}{2}$, and $\sigma(\mathbf{m}_s)$ therefore only has two states, α and β. For closed-shell systems, it is common to require that pairs of electrons share spatial orbitals ψ_k, but have opposite spin-states. Instead of writing the set of spin-orbitals as $\varphi_1, \varphi_2, ..., \varphi_{N_\mathrm{el}}$, it can thus be written as $\varphi_{1,\alpha}, \varphi_{1,\beta}, ..., \varphi_{\frac{1}{2}N_\mathrm{el},\beta}$.

A basic rule of quantum chemistry is the Pauli principle, stating that the total wavefunction Ψ must be anti-symmetric upon exchange of any of its labels, or in other words, it must change sign upon permutation of two identical fermions. [8] An example of this would be if two electrons switched orbital. To ensure this antisymmetry, the total wavefunction Ψ is typically a combination of the MOs φ_k through a Slater determinant

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{N_\mathrm{el}}) = \left(\frac{1}{N_\mathrm{el}!}\right)^{\frac{1}{2}} \begin{vmatrix} \varphi_{1,\alpha}(\mathbf{r}_1) & \varphi_{1,\beta}(\mathbf{r}_1) & \cdots & \varphi_{\frac{1}{2}N_\mathrm{el},\beta}(\mathbf{r}_1) \\ \varphi_{1,\alpha}(\mathbf{r}_2) & \varphi_{1,\beta}(\mathbf{r}_2) & \cdots & \varphi_{\frac{1}{2}N_\mathrm{el},\beta}(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{1,\alpha}(\mathbf{r}_{N_\mathrm{el}}) & \varphi_{1,\beta}(\mathbf{r}_{N_\mathrm{el}}) & \cdots & \varphi_{\frac{1}{2}N_\mathrm{el},\beta}(\mathbf{r}_{N_\mathrm{el}}) \end{vmatrix} \tag{1.0.8}$$

where N_{el} is the number of electrons and $\mathbf{r}_i$ is the position of electron i. We will in this work keep to the restricted formulation, where two electrons occupy each spatial orbital, and will from this point on therefore no longer concern ourselves with the spin degree of freedom, looking only at spatial orbitals with space for two electrons.

In many cases it is practical to further expand these MOs in terms of orbitals centered on the atomic nuclei, atomic orbitals (AOs). This is typically done through a linear combination

$$\psi_i = \sum_\mu C_{\mu i} \chi_\mu \tag{1.0.9}$$

where the summation index μ runs over the number of atomic orbitals, $C_{\mu i}$ are the so-called MO coefficients and χ_μ are the atomic orbitals. The orbital coefficients are related to the density matrix D by the following relation

$$D_{\mu\nu} = 2 \sum_i C^*_{\mu i} C_{\nu i} \tag{1.0.10}$$

which is a quantity of high importance to be reencountered in Chapter 3. It can be seen as the electronic density in the region of overlap between χ_μ and χ_ν.[8] The factor 2 follows from the previously discussed restricted formalism, with each of the spatial orbitals being doubly occupied. In an unrestricted formalism, Eq. (1.0.10) would not have a factor 2, but the sum i would run over the twice as many MOs. The atomic orbitals χ_μ make up what is called a set of basis functions centered on each atomic position. The choice made of how to represent these functions is one that has big impact on calculations, and care should therefore be taken in this decision[48]. Any further discussion on the topic of the wavefunction is, however, outside the scope of this work, and we now move on to briefly discuss a few of the most important QM methods.

Among the terms in the electronic Hamiltonian in Eq. (1.0.4), the one that is the most problematic to calculate is $\hat{g}$, even though it has a seemingly simple and elegant form. The sheer number of interactions to be calculated, as each electron interacts with every single one of the others, is in itself a big hurdle. Additionally, this term couples different electrons, meaning that the separation of the wavefunction into one-electron functions, or orbitals, is an

approximation. In order to solve the Schrödinger equation for one orbital, it is necessary to know the $\frac{1}{2}N_{\text{el}} - 1$ other ones. The first method often introduced that presents a simplified version of the two-electron term is the Hartree-Fock (HF)[9,10] method, which approaches the problem iteratively. Considering one electron at a time, all the other electrons are modeled as a continuous electron cloud treated averagely. The method then loops through all of the electrons, and the procedure is repeated until convergence is reached. Such schemes for iteratively solving the Schrödinger equation are called Self-Consistent Field (SCF) methods,[8] of which HF is an example. It is often said about HF that it disregards Coulomb electron correlation. This points to the fact that no instantaneous interactions between individual electrons are accounted for. Although this makes the HF method in most cases unsuitable for accurate calculations, it is in many cases still used for comparison or as a preparatory step before more accurate methods are employed due to its efficient nature.

The Coulomb electron correlation is typically grouped into two contributions, nondynamical and dynamical correlation.[50] Nondynamical correlation is related to the breakdown of the single-determinant approximation, as a consequence of the existence of several important electronic configurations. Dynamical correlation on the other hand, is due to instantaneous electron-electron interactions, requiring a description of cusps in the $\frac{1}{r_{ij}}$ term in order to avoid singularities as $r_{ij} \to 0$. There have been devised many different solutions for alleviating the lack of electron correlation in HF. A typical way to account for nondynamical correlation is by including more configurations than the HF configuration in the wavefunction through methods such as the Multi-configurational SCF method (MCSCF). Dynamical correlation can be included by adding excited states to the wavefunction,[8,50] a philosophy that has led to well-known methods such as the Configuration Interaction (CI) and Coupled Cluster (CC) methods, where excited states are included linearly and exponentially, respectively. The method used in this work, however, is called Density Functional Theory (DFT) and takes a different approach to the correlation problem. DFT originates from a theorem formulated by Hohenberg and Kohn in 1964,[12] proving that it is possible to determine any ground-state electronic property as a functional of the electron density ρ. Among the many ways to formulate this theory is Kohn-Sham DFT (KS–DFT),[51] where the ground-state electronic energy

functional E_v is formulated as

$$E_v[\rho] = h_{nuc} + \int v(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + T_s[\rho] + J[\rho] + E_{\text{xc}}[\rho] \tag{1.0.11}$$

where h_{nuc} is still the nuclear repulsion, and the second term is the interaction between electrons and nuclei and therefore closely related to the third term in Eq. (1.0.2), with

$$v(\mathbf{r}) = -\sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{r}_A|} \tag{1.0.12}$$

The third term, T_s is the kinetic energy of an imagined case with non-interacting electrons, and J is the classical part of the electron-electron potential, the Coulomb energy. These are subsets of the first and third terms in Eq. (1.0.2).

The final term in Eq. (1.0.11), the exchange-correlation energy functional E_{xc}, approximates the parts of the kinetic energy T and electron-electron interaction energy V_{ee} that are not included in T_s and J. This is the correlation energy, the contribution to the kinetic energy caused by interactions between the electrons, and the exchange energy, a non-classical contribution caused by the requirement for an antisymmetric wavefunction.

Many different strategies have been devised to model E_{xc}, but common among them is that they are all approximations, each with their strengths and their weaknesses. In the context of DFT functionals, "Jacobs Ladder"[52] is often mentioned, a hierarchy over some of the most important types of functionals. The functionals in "Jacobs Ladder" range from the local density approximation which is an integral of a function only of the density, to include more and more sophisticated terms such as density gradients (generalized gradient approximations), higher-order density derivatives or even exactly computed exchange form Kohn-Sham orbitals (Hybrid orbitals).[53] As DFT functionals are not the primary focus of this thesis, this is not a subject that will be discussed in any more detail here, except noting that it is always important before starting a study to make an informed choice as to which functional is best suited. The system upon which calculations are performed as well as the property being calculated are both influenced, possibly in different ways, by the choice of functional.

In this Chapter, the fundamentals of quantum chemistry have been covered to such an extent as to prepare the reader for further developments beyond calculating the ground-state energy with QM methods. The next chapter, Chapter 2, introduces what happens when a system is exposed to an external electromagnetic field. This links theoretical to experimental chemistry as this is the foundation of spectroscopic techniques, and calculating the properties observed in such situations is essential for theoretically reproducing results from these experimental techniques, such as for instance Infrared and Raman spectroscopies.

Chapter 2

Vibrational spectroscopy

Having in Chapter 1 been concerned only with the ground-state electronic energy, it is now time to shift the focus to properties arising when a system is subject to an external electromagnetic field, as is for instance the case when considering spectroscopic processes. These molecular properties include for instance the dipole moment or polarizability, and can be obtained as derivatives of the time-averaged quasienergy as covered in detail in Thorvaldsen *et al.*,[29] and in a fashion more relevant to this work in Ringholm's thesis.[54] The details of this formalism are outside the scope of this work and we will here only focus on the the results necessary for our needs.

The books by Wilson *et al.*[38] and Norman *et al.*[55] are both good sources on vibrational spectroscopy. These two works provide a large part of the background for the theory both behind vibrational properties in general, as well as for the spectroscopic techniques covered in this chapter, namely Infrared, Raman and hyper-Raman. The commonly used harmonic approximation is also covered here, as well as how to go beyond the harmonic approximation by including anharmonic corrections.

2.1 Molecular properties

Almost all molecular properties necessary for vibrational spectroscopy can be expressed as derivatives of the energy.[1] If our system is exposed to an

[1] All properties involving non-static external fields should be evaluated as derivatives of the aforementioned time-averaged Quasi-energy, instead of simply the energy. However, as the derivations are analogous, we will not distinguish between the two in this thesis.

external field, we can identify the electric properties from a Taylor expansion of the molecular energy $E^{\text{mol}} = \langle\Psi|\hat{H}^{\text{mol}}|\Psi\rangle$ in terms of the electric field strength $\mathbf{F}$,

$$\begin{aligned} E^{\text{mol}}(\mathbf{r},\mathbf{R},\mathbf{F}) =& E_0^{\text{mol}}(\mathbf{r},\mathbf{R}) + \left.\frac{\partial E^{\text{mol}}(\mathbf{r},\mathbf{R},\mathbf{F})}{\partial \mathbf{F}}\right|_{\mathbf{F}=\mathbf{0}} \mathbf{F} \\ &+ \frac{1}{2}\left.\frac{\partial^2 E^{\text{mol}}(\mathbf{r},\mathbf{R},\mathbf{F})}{\partial \mathbf{F}^2}\right|_{\mathbf{F}=\mathbf{0}} \mathbf{F}^2 \\ &+ \frac{1}{6}\left.\frac{\partial^3 E^{\text{mol}}(\mathbf{r},\mathbf{R},\mathbf{F})}{\partial \mathbf{F}^3}\right|_{\mathbf{F}=\mathbf{0}} \mathbf{F}^3 + \cdots \end{aligned} \tag{2.1.1}$$

where the subscript $\mathbf{F} = 0$ denotes that the derivatives are evaluated at zero field strengths. The molecular Hamiltonian in this case has been perturbed by the electric field, and can be written as $\hat{H}^{\text{mol}} = \hat{H}_0^{\text{mol}} + \hat{H}^{\text{int}}$, where $\hat{H}_0^{\text{mol}}$ is the unperturbed molecular Hamiltonian, as described in Eq. (1.0.3), and $\hat{H}^{\text{int}}$ is the interaction operator between the field and the molecule. In the electric dipole approximation, for example, $\hat{H}^{\text{int}} = -\hat{\mu}F$, but it can also depend on higher-order interactions. The zeroth-order term, is thus the molecular energy in the absence of an electric field, and is given by $E_0^{\text{mol}}(\mathbf{r},\mathbf{R}) = \langle\Psi|\hat{H}_0^{\text{mol}}|\Psi\rangle$.

An alternative expansion of the energy can be done in terms of polarization properties, such as the dipole moment μ, polarizability α, hyper polarizability β, etc.,

$$\begin{aligned} E^{\text{mol}}(\mathbf{r},\mathbf{R},\mathbf{F}) =& E_0^{\text{mol}}(\mathbf{r},\mathbf{R}) - \boldsymbol{\mu}(\mathbf{r},\mathbf{R},\mathbf{F})\mathbf{F} - \frac{1}{2}\boldsymbol{\alpha}(\mathbf{r},\mathbf{R},\mathbf{F})\mathbf{F}^2 \\ &- \frac{1}{6}\boldsymbol{\beta}(\mathbf{r},\mathbf{R},\mathbf{F})\mathbf{F}^3 + \cdots \end{aligned} \tag{2.1.2}$$

Comparing the phenomenological expansion in Eq. (2.1.2) with the Taylor expansion in Eq. (2.1.1), we can identify the dipole moment $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu}(\mathbf{r},\mathbf{R},\mathbf{F}) = -\left.\frac{\partial E^{\text{mol}}(\mathbf{r},\mathbf{R},\mathbf{F})}{\partial \mathbf{F}}\right|_{\mathbf{F}=\mathbf{0}} \tag{2.1.3}$$

and the polarizability $\boldsymbol{\alpha}$ as

$$\boldsymbol{\alpha}(\mathbf{r},\mathbf{R},\mathbf{F}) = -\frac{\partial^2 E^{\mathrm{mol}}(\mathbf{r},\mathbf{R},\mathbf{F})}{\partial \mathbf{F}^2}\bigg|_{\mathbf{F}=0} \tag{2.1.4}$$

Above, we discussed the molecular energy in relation to electric fields, providing electric molecular properties. Another class of properties central to vibrational spectroscopy is connected to the geometry of the molecule. A molecular vibration involves an alteration of the positions of the involved nuclei, and we will now investigate the most important properties related to such motions. As a consequence of the Born-Oppenheimer approximation, the nuclei can be seen to move in an effective potential field generated by the electrons.[55] The molecular energy in the absence of an electric field is given by Eq. (1.0.3), which is separated into the nuclear kinetic term ($\hat{T}_N = -\frac{1}{2}\sum_A \frac{\nabla_A^2}{m_A}$) and the electronic potential term $\hat{H}^{\mathrm{el}}$. In the presence of an electric field, we from here on include the interaction energy into the electronic potential, which results in a total Hamiltonian $\hat{H}^{\mathrm{mol}} = \hat{T}_N + \hat{H}^{\mathrm{el}}$. As a consequence of the Born-Oppenheimer approximation, the electronic and vibrational wavefunctions are separable, and the electronic wavefunctions can be seen to be unaffected by nuclear displacements ($\nabla_A|\Psi^{\mathrm{el}}\rangle \approx 0$), resulting in

$$\hat{T}_N\Psi = \Psi^{\mathrm{el}}\left(\hat{T}_N\Psi^{\mathrm{vib}}\right) \tag{2.1.5}$$

where Ψ^{el} and Ψ^{vib} are the electronic and nuclear wavefunctions, respectively. Both of the wavefunctions are associated with a specific electronic state. It is therefore possible to write the molecular energy E^{mol} for a specific electronic state as

$$\begin{aligned}
E^{\mathrm{mol}} =& \langle\Psi^{\mathrm{vib}}\Psi^{\mathrm{el}}|\hat{T}_N + \hat{H}^{\mathrm{el}}|\Psi^{\mathrm{el}}\Psi^{\mathrm{vib}}\rangle \\
=& \langle\Psi^{\mathrm{vib}}|\hat{T}_N|\Psi^{\mathrm{vib}}\rangle\langle\Psi^{\mathrm{el}}|\Psi^{\mathrm{el}}\rangle + \langle\Psi^{\mathrm{vib}}\Psi^{\mathrm{el}}|\hat{H}^{\mathrm{el}}|\Psi^{\mathrm{el}}\Psi^{\mathrm{vib}}\rangle \\
=& \langle\Psi^{\mathrm{vib}}|\hat{T}_N|\Psi^{\mathrm{vib}}\rangle + \langle\Psi^{\mathrm{vib}}|V^{\mathrm{el}}|\Psi^{\mathrm{vib}}\rangle \\
=& \langle\Psi^{\mathrm{vib}}|\hat{T}_N + V^{\mathrm{el}}|\Psi^{\mathrm{vib}}\rangle
\end{aligned} \tag{2.1.6}$$

where we have in the third equality integrated over all electronic degrees of freedom, in order to obtain an energy-operator independent of electronic coordinates, $\hat{H}^{\mathrm{vib}} = \hat{T}_N + V^{\mathrm{el}}$.

The vibrational electronic potential has a complex dependency on the nuclear coordinates, but close to the equilibrium, it can be expressed as a Taylor expansion in terms of the displacement of the nuclear coordinates from their equilibrium position,

$$\begin{aligned} V^{\mathrm{el}} =& V_0^{\mathrm{el}} + \sum_{\alpha=x,y,z} \sum_{n}^{N} \left.\frac{\partial V^{\mathrm{el}}}{\partial \Delta\alpha_n}\right|_{\Delta\alpha_n=0} \Delta\alpha_n \\ &+ \frac{1}{2} \sum_{\alpha,\beta=x,y,z} \sum_{n,m}^{N} \left.\frac{\partial^2 V^{\mathrm{el}}}{\partial \Delta\alpha_n \partial \Delta\beta_m}\right|_{\Delta\alpha_n,\Delta\beta_m=0} \Delta\alpha_n \Delta\beta_m \\ &+ \frac{1}{6} \sum_{\alpha,\beta,\gamma=x,y,z} \sum_{n,m,r}^{N} \left.\frac{\partial^3 V^{\mathrm{el}}}{\partial \Delta\alpha_n \partial \Delta\beta_m \partial \Delta\gamma_r}\right|_{\Delta\alpha_n,\Delta\beta_m,\Delta\gamma_r=0} \\ &\times \Delta\alpha_n \Delta\beta_m \Delta\gamma_r + \cdots \end{aligned} \tag{2.1.7}$$

where α, β and γ here denote one of the three Cartesian coordinates, and the summation indices n, m and r run over the number of nuclei N. The symbol $\Delta\alpha_n$ is the the displacement of nucleus n from its equilibrium position along Cartesian coordinate α, and subscripts such as $\Delta\alpha_n = 0$ denote that the derivatives are evaluated at zero displacement, or in other words, at the equilibrium geometry. The zeroth-order term, V_0^{el}, is here the electronic potential at equilibrium geometry. The choice of symbols for Cartesian components might seem somewhat confusing as the same symbols have also recently been, and will in the future be, used to represent polarization properties. Fortunately, we will not have to deal with this confusion for long, as we now introduce a new, mass-weighted set of $3N$ coordinates

$$\begin{aligned} &q_1 = \sqrt{m_1}\Delta x_1, q_2 = \sqrt{m_1}\Delta y_1, q_3 = \sqrt{m_1}\Delta z_1, q_4 = \sqrt{m_2}\Delta x_2, \\ &\ldots, q_{3N} = \sqrt{m_N}\Delta z_N \end{aligned} \tag{2.1.8}$$

where m_1 is the mass of nucleus 1 etc. The Taylor expansion of the electronic potential in Eq. (2.1.7) can then be reformulated to

$$V^{\mathrm{el}} = V_0^{\mathrm{el}} + \sum_{i=1}^{3N} \left.\frac{\partial V^{\mathrm{el}}}{\partial q_i}\right|_{q_i=0} q_i + \frac{1}{2}\sum_{i,j=}^{3N} \left.\frac{\partial^2 V^{\mathrm{el}}}{\partial q_i \partial q_j}\right|_{q_i,q_j=0} q_i q_j + \frac{1}{6}\sum_{i,j,k}^{3N} \left.\frac{\partial^3 V^{\mathrm{el}}}{\partial q_i \partial q_j \partial q_k}\right|_{q_i,q_j,q_k=0} q_i q_j q_k + \cdots \tag{2.1.9}$$

Alternatively, the geometric Taylor expansion Eq. (2.1.9) can be expressed in terms of geometric properties,

$$V^{\mathrm{el}} = V_0^{\mathrm{el}} + \sum_{i=1}^{3N} g_i q_i + \frac{1}{2}\sum_{i,j=1}^{3N} H_{ij} q_i q_j + \frac{1}{6}\sum_{i,j,k=1}^{3N} C_{ijk} q_i q_j q_k + \cdots \tag{2.1.10}$$

where the i'th component of the molecular gradient g_i is given by

$$g_i = \left.\frac{\partial V^{\mathrm{el}}}{\partial q_i}\right|_{q_i=0} \tag{2.1.11}$$

the ij'th component of the molecular Hessian H_{ij} by

$$H_{ij} = \left.\frac{\partial^2 V^{\mathrm{el}}}{\partial q_i \partial q_j}\right|_{q_i,q_j=0} \tag{2.1.12}$$

and the ijk'th component of the cubic force field C_{ijk} by

$$C_{ijk} = \left.\frac{\partial^3 V^{\mathrm{el}}}{\partial q_i \partial q_j \partial q_k}\right|_{q_i,q_j,q_k=0} \tag{2.1.13}$$

The application of these properties, and more like them, in the context of vibrational spectroscopy, will be explored in the following sections.

2.2 The Harmonic approximation

One of the most central approximations in the theory of vibrational motion is the Harmonic approximation. It builds on the fact that close to the

nuclear equilibrium position, the potential energy surface created by the electrons has close to a parabolic shape, and that Eq (2.1.9) therefore in this region can be safely truncated after the second-order term. The zeroth-order term, V_0^{el}, is a constant that can be set to zero, and since the molecular gradient is evaluated at equilibrium, it is also zero. The harmonic molecular vibrational Hamiltonian is thus given by

$$\begin{aligned}\hat{H}^{\mathrm{harm,mol}} =& \hat{T}_N + \frac{1}{2}\sum_{i,j=1}^{3N} H_{ij} q_i q_j \\ =& -\frac{1}{2}\sum_{i=1}^{3N}\frac{\partial^2}{\partial q_i^2} + \frac{1}{2}\sum_{i,j=1}^{3N} H_{ij} q_i q_j \end{aligned} \tag{2.2.1}$$

As explained through slightly different paths in Wilson *et al.*[38] and Norman *et al.*[55], the Schrödinger equation within this harmonic Hamiltonian is most easily solved by diagonalizing the Hamiltonian. In practice, this means that through an eigenvalue analysis of the molecular Hessian $\mathbf{H}$, we obtain the angular vibrational frequencies ω_i of the molecule as eigenvalues, as well as the transformation matrix to a new set of normal coordinates Q_i as the eigenvectors. The normal coordinates diagonalize both the nuclear kinetic operator and the molecular Hessian. Of the set of $3N$ normal modes, three belong to translational motion and three (two for linear molecules) belong to rotation. Each of the $3N-6$ (or $3N-5$ for linear molecules) remaining normal coordinates Q_i belong to a specific vibrational mode of the molecule, with corresponding angular frequencies ω_i. In the discussion of molecular vibrations, we will operate only with this set of $3N-6$ vibrational normal coordinates Q_i. The new, diagonalized vibrational Hamiltonian is given by

$$\hat{H}^{\mathrm{harm,mol}} = -\frac{1}{2}\sum_{i=1}^{3N-6}\frac{\partial^2}{\partial Q_i^2} + \frac{1}{2}\sum_{i=1}^{3N-6} H_{ii} Q_i^2 \tag{2.2.2}$$

Just as a molecule can be electronically excited by an external electric field in the ultraviolet or visible spectrum, it can be excited into a heightened vibrational state by an external electric field in the infrared region of the spectrum. Having established the vibrational Hamiltonian, we turn our attention to the vibrational wavefunctions that for vibrationally excited state n of normal mode i in the harmonic approximation are given by[1]

$$\psi_{n_i}(Q_i) = N_{n,i} H_n(\xi_i) e^{-\xi_i^2/2} \tag{2.2.3}$$

where $N_{n,i}$ is the normalization constant, H_n is the n'th-order Hermite polynomial (a set of polynomials that are part of the solutions to harmonic motion) and ξ_i is a frequency-scaled coordinate, $\xi_i = \sqrt{\frac{\omega_i}{\hbar}} Q_i$. The normalization constant is given by

$$N_{n,i} = \left[\left(\frac{\omega_i}{\hbar \pi} \right)^{1/2} \frac{1}{2^n (n!)} \right]^{1/2} \tag{2.2.4}$$

For reasons that will become apparent, in the harmonic approximation, the only allowed transitions from the ground state are to the first excited state. The zeroth- and first-order Hermite polynomials are therefore of special importance, and are given by

$$H_0(\xi_i) = 1; H_1(\xi_i) = 2\xi_i. \tag{2.2.5}$$

As stated by Fermi's golden rule,[8] the rate of transition to a vibrationally excited state is dependent on the transition moment of the interaction operator, $\langle 0 | H^{\text{int}} | n \rangle$, where n is the collected excited state for all the normal modes. This has a strong relation to the intensity of spectroscopic processed, as the intensity is quadratically dependent on the rate of transition. In the dipole approximation, this means that the intensity is proportional to the square of the transition dipole moment $\langle 0 | \mu | n \rangle$. This is the case for Infrared absorption, but the nonlinear terms in Eq. (2.1.2) are the determining factors in scattering processes such as Raman and Hyper-Raman. We will now derive the transition dipole moment in the harmonic dipole approximation, but note that the derivations for other polarization properties is identical.

We express the dipole moment as a Taylor expansion with respect to the normal coordinates Q_i

$$\boldsymbol{\mu} = \boldsymbol{\mu}_0 + \sum_{i=1}^{3N-6} \frac{\partial \boldsymbol{\mu}}{\partial Q_i} \bigg|_{Q_i=0} Q_i + \frac{1}{2} \sum_{i,j=1}^{3N-6} \frac{\partial^2 \boldsymbol{\mu}}{\partial Q_i \partial Q_j} \bigg|_{Q_i,Q_j=0} Q_i Q_j + \cdots \tag{2.2.6}$$

Truncating after the first-order term leads to the *double* Harmonic approximation. The zeroth-order term, $\boldsymbol{\mu}_0$, is the dipole moment of the vibrational ground state, and the subscript $Q_i = 0$ denotes that the derivative is evaluated at equilibrium geometry, or zero displacement. In practice, such as in quantum chemistry codes, property-derivatives such as the ones of the dipole moment seen in the Taylor expansion are typically first calculated as derivatives of Cartesian coordinates and subsequently transformed to normal coordinate basis with the transition matrix found as the eigenvector in the Hessian analysis.

The zeroth-order contribution the transition dipole moment is $\langle 0|\boldsymbol{\mu}_0|n\rangle = \boldsymbol{\mu}_0\langle 0|n\rangle$, which because of orthogonality between different vibrational states becomes zero for all excited states. The harmonic transition dipole moment belonging to Cartesian coordinate α can thus be written as its single remaining term,

$$\begin{aligned}\langle 0|\mu_\alpha|n\rangle =& \langle 0|\sum_{i=1}^{3N-6}\left.\frac{\partial\mu_\alpha}{\partial Q_i}\right|_{Q_i=0} Q_i|n\rangle \\ =& \sum_{i=1}^{3N-6}\left.\frac{\partial\mu_\alpha}{\partial Q_i}\right|_{Q_i=0}\langle 0|Q_i|n\rangle \end{aligned} \tag{2.2.7}$$

As there are $3N-6$ vibrational normal modes that can all be excited, the total wavefunction for state n can be formulated as a product of the wavefunctions belonging to each of these modes, excited or not, so that $|n\rangle = |n_a\rangle|n_b\rangle\cdots|n_i\rangle\cdots|n_{3N-6}\rangle$.[38] This is possible as the nuclei are not fermions, and therefore not subject to the Pauli antisymmetry requirement. The subscripts here denote the normal mode. All these wavefunctions are given by Hermite polynomials, as in Eq. (2.2.3).

As the dipole gradient is considered a constant, the task is now to solve $\langle 0|Q_i|n\rangle$. Using a property of the Hermite polynomials, that

$$\int_{-\infty}^{\infty}\psi_{n_i'}(\xi_i)\psi_{n_i}(\xi_i)dQ_i = \delta_{n_i'n_i} \tag{2.2.8}$$

and that only $|n_i\rangle$ has a dependence on Q_i, it follows that the only non-vanishing integral for $\langle 0|Q_i|n\rangle$ is when all $|n_j\rangle = |0_j\rangle$ except $|n_i\rangle$, giving

$$\begin{aligned}\langle 0|Q_i|n\rangle =&\langle 0_a|0_a\rangle\langle 0_b|0_b\rangle\cdots\langle 0_i|Q_i|n_i\rangle\cdots = \langle 0_i|Q_i|n_i\rangle \\ =&\int_{-\infty}^{\infty} N_{0,i}H_0(\xi_i)e^{-\xi_i^2/2}Q_iN_{n,i}H_n(\xi_i)e^{-\xi_i^2/2}dQ_i \end{aligned} \tag{2.2.9}$$

Using Eq. (2.2.5), we see that $H_0(\xi_i)Q_i = \sqrt{\frac{\hbar}{4\omega_i}}H_1(\xi_i)$, resulting in

$$\begin{aligned}\langle 0|Q_i|n\rangle =&\int_{-\infty}^{\infty} N_{0,i}\left(\frac{\hbar}{4\omega_i}\right)^{1/2}H_1(\xi_i)e^{-\xi_i^2/2}N_{n,i}H_n(\xi_i)e^{-\xi_i^2/2}dQ_i \\ =&\frac{N_{0,i}}{N_{1,i}}\left(\frac{\hbar}{4\omega_i}\right)^{1/2}\int_{-\infty}^{\infty}\psi_1(Q_i)\psi_n(Q_i)dQ_i \\ =&\frac{N_{0,i}}{N_{1,i}}\left(\frac{\hbar}{4\omega_i}\right)^{1/2}\delta_{1n}\end{aligned} \tag{2.2.10}$$

Consequently, the only non-zero case is when $n = 1$, which means that

$$\langle 0|Q_i|n\rangle =\frac{N_{0,i}}{N_{1,i}}\left(\frac{\hbar}{4\omega_i}\right)^{1/2} = \sqrt{\frac{\hbar}{2\omega_i}} \tag{2.2.11}$$

Inserting this into Eq. (2.2.7), we end up with the harmonic approximation for component α of the dipole transition moment, given by

$$\langle 0|\mu_\alpha|n\rangle = \sum_{i=1}^{3N-6}\sqrt{\frac{\hbar}{2\omega_i}}\frac{\partial\mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0} \tag{2.2.12}$$

and for a specific mode i by

$$\langle 0|\mu_\alpha|n_i\rangle =\sqrt{\frac{\hbar}{2\omega_i}}\frac{\partial\mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0} \tag{2.2.13}$$

We will now, in the next few sections, explore the application of such polarization transition moments in vibrational spectroscopic techniques.

2.3 Infrared absorption

With the harmonic approximation for molecular vibrations in place, we can use this to explore the theory behind the specific spectroscopic techniques applied in this work. The first one is infrared absorption, a method commonly used for determining the structural identification of a sample. It is not within the scope of this work to delve into the experimental aspects of this method, but very briefly, a beam of electromagnetic light of a certain intensity I_0 and with a frequency covering a given range within the infrared spectrum is pointed at a sample. Photons in the infrared spectrum have the ability to excite molecules out of the vibrational ground state, causing some of the photons to be absorbed by the sample. The intensity of the ray of light exiting the sample I will therefore be different than that of the one entering (I_0) and vary as a function of frequency, and I is therefore picked up by a detector and plotted over the frequency range. This provides characteristic spectra for different molecular compounds. A schematic illustration of an IR spectrometer can be seen in Fig. 2.1.

The relationship between the incoming and outgoing intensities is given by Beer-Lambert's law[55]

$$I = I_0 e^{-N\sigma l} = I_0 \cdot 10^{-\varepsilon c l} \tag{2.3.1}$$

where N is the number of molecules per volume, σ is the absorption cross section, l is path length, ε is the molar decadic attenuated coefficient (a measure of the amount of photons absorbed) and c is the molar concentration. Manipulating Eq. (2.3.1) and using that $N = N_A c$, where N_A is Avogadro's constant, we get

$$\varepsilon = \frac{N_A \sigma}{\ln(10)} \tag{2.3.2}$$

In computational chemistry, one of the most commonly calculated quantity is the molar decadic attenuated coefficient.

The absorption cross section is given by

$$\sigma(\omega) = \frac{\pi\omega}{3\hbar\varepsilon_0 c} \sum_{\alpha=x,y,z} |\langle 0|\mu_\alpha|n\rangle|^2 \sum_{i}^{3N-6} f(\omega; \omega_i, \gamma_i) \tag{2.3.3}$$

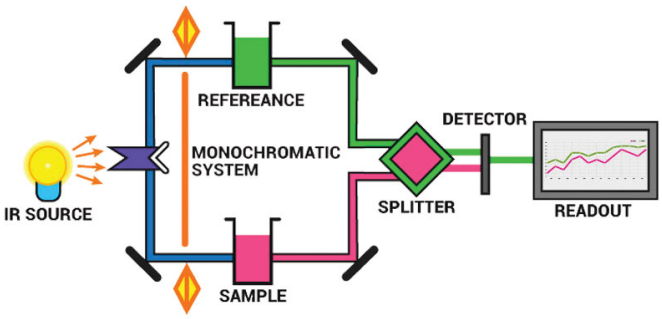


Figure 2.1. Schematic illustration of an IR spectrometer.[56] Light is transmitted from a light source, and filtered to only contain the desired range of wavelengths in a monochromator. The beam then passes through a sample which at certain wavelengths will absorb some of it. The light that has not been absorbed continues to the detector, and the difference between the intensity of the light before and after entering the sample is registered. In order to eliminate influence from the experimental setup, the IR spectrum from the sample is typically compared to that of a reference. If the sample is a water solution, a good reference would typically be pure water. The splitter in the figure ensures that the detector recieves alternating signals from the beam that has passed through the reference and that which has passed through the sample.

where ω is the angular frequency as a variable over the whole spectral range, ε_0 is the vacuum permittivity and $f(\omega;\omega_i,\gamma_i)$ is a lineshape function, included in order to account for homogenous broadening effects, such as the finite lifetime of the excited states. We use the Cauchy distribution,[55]

$$f(\omega;\omega_i,\gamma_i) = \frac{1}{\pi}\left[\frac{\gamma_i}{(\omega_i-\omega)^2+\gamma_i^2}\right] \tag{2.3.4}$$

to model the lineshape function, where γ_i is a damping factor associated with the life expectancy of the excited vibrational state. It is related to the more common term, the "full width at half maximum" (FWHM) by $2\gamma_i = \text{FWHM}$.

In cases where the harmonic approximation is applied, we insert Eq. (2.2.13) into Eq. (2.3.3), giving

$$\begin{aligned}\sigma_i(\omega) =& \frac{\pi\omega}{3\hbar\varepsilon_0 c}\sum_{\alpha=x,y,z}\left(\sqrt{\frac{\hbar}{2\omega_i}}\frac{\partial\mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0}\right)^2 f(\omega;\omega_i,\gamma_i)\\ =& \frac{\pi\omega}{6\varepsilon_0 c\omega_i}\sum_{\alpha=x,y,z}\left(\frac{\partial\mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0}\right)^2 f(\omega;\omega_i,\gamma_i)\end{aligned} \tag{2.3.5}$$

for mode i. The lineshape function causes σ to sharply peak at the transition frequencies, making $\omega \approx \omega_i$ a reasonable approximation,[55] giving the most commonly used form of the harmonic absorption cross section

$$\sigma_i(\omega) = \frac{\pi}{6\varepsilon_0 c}\sum_{\alpha=x,y,z}\left(\frac{\partial\mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0}\right)^2 f(\omega;\omega_i,\gamma_i) \tag{2.3.6}$$

Inserting this into Eq. (2.3.2) gives us the molar decadic attenuated coefficient for mode i

$$\varepsilon_i(\omega) = \frac{N_A\pi}{6\ln(10)\varepsilon_0 c}\sum_{\alpha=x,y,z}\left(\frac{\partial\mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0}\right)^2 f(\omega;\omega_i,\gamma_i) \tag{2.3.7}$$

2.3.1 Considerations regarding IR units

Before continuing to other types of vibrational spectroscopies, some aspects regarding the units of IR spectroscopies should be addressed, as there is

quite a lot of ambiguity in the way values are reported. This is especially true when it comes to the use of the word "intensity" in relation to computationally calculated quantities. From Beer-Lambert's law in Eq. (2.3.1), it is clear that reporting ε is not the same as reporting I, although they are closely related. In this work, the terms intensity or intensity-related property will be used when specific units of the values are not relevant, or as a collective name for all the different units. When reporting specific values, however, the exact physical quantity and its units should always be stated.

The three most common units used for infrared absorption are $\mathrm{L} \cdot \mathrm{mol}^{-1}\mathrm{cm}^{-1}$, $(\mathrm{D/Å})^2 \cdot \mathrm{amu}^{-1}$ and $\mathrm{km} \cdot \mathrm{mol}^{-1}$. Starting with the first, we see that it is closely related to the SI unit $\mathrm{m}^2 \cdot \mathrm{mol}^{-1}$, which are the units of ε. Moving on to the second choice of units, we note that D is a commonly used unit for dipole moment, Å is a length measure and amu is $\frac{1}{12}$'th of the mass of a carbon atom. The SI equivalent of $(\mathrm{D/Å})^2 \cdot \mathrm{amu}^{-1}$ is thus $\mathrm{C}^2 \cdot \mathrm{kg}^{-1}$, and this unit is used when intensities are reported as

$$\sum_{\alpha=x,y,z} \left(\frac{\partial \mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0} \right)^2 \tag{2.3.8}$$

The last unit, km/mol, is less easily recognizable from the equations above, and is in many places simply said to originate from the conversion $1(\mathrm{D/Å})^2 \cdot \mathrm{amu}^{-1} = 42.2561\ \mathrm{km} \cdot \mathrm{mol}^{-1}$. As the origin of this conversion is non-trivial, I will here describe the steps some in detail. When looking for the origins of this conversion, it is necessary first to note that ε is the molar *decadic* attenuated coefficient (10-based), which is related to the the molar *Napierian* (e-based)[57] attenuated coefficient ε_N by the relation $\varepsilon = \varepsilon_N / \ln(10)$, giving

$$\varepsilon_{N,i}(\omega) = \frac{\pi N_A}{6\varepsilon_0 c} \sum_{\alpha=x,y,z} \left(\frac{\partial \mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0} \right)^2 f(\omega; \omega_i, \gamma_i) \tag{2.3.9}$$

The next step is integrating over the spectral range in terms of wavenumbers $\bar{v}$

$$A_i = \int \varepsilon_{N,i}(\bar{v}) d\bar{v} \tag{2.3.10}$$

This requires a change of variables in ε_N, but is easily achieved as $\omega =$

$2\pi c\bar{\nu} \Rightarrow d\bar{\nu} = \frac{1}{2\pi c} d\omega$. Additionally, since $f(\omega;\omega_i,\gamma_i)$ is a distribution function,[55]

$$\int f(\omega;\omega_i,\gamma_i)d\omega = 1 \tag{2.3.11}$$

Inserting this into Eq. (2.3.10) gives

$$\begin{aligned} A_i &= \int \frac{\pi N_A}{6\varepsilon_0 c} \sum_{\alpha=x,y,z} \left(\frac{\partial \mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0} \right)^2 f(\omega;\omega_i,\gamma_i)d\bar{\nu} \\ &= \frac{\pi N_A}{6\varepsilon_0 c} \sum_{\alpha=x,y,z} \left(\frac{\partial \mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0} \right)^2 \int f(\omega;\omega_i,\gamma_i)d\bar{\nu} \\ &= \frac{\pi N_A}{6\varepsilon_0 c} \sum_{\alpha=x,y,z} \left(\frac{\partial \mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0} \right)^2 \frac{1}{2\pi c} \int f(\omega;\omega_i,\gamma_i)d\omega \\ &= \frac{N_A}{12c^2\varepsilon_0} \sum_{\alpha=x,y,z} \left(\frac{\partial \mu_\alpha}{\partial Q_i}\bigg|_{Q_i=0} \right)^2 \end{aligned} \tag{2.3.12}$$

which has units $\mathrm{m} \cdot \mathrm{mol}^{-1}$, the SI equivalent of $\mathrm{km} \cdot \mathrm{mol}^{-1}$. To compare the result in Eq. (2.3.12) with the previously mentioned factor of 42.2561, combine the constants placed before the sum into a common prefactor and convert the dipole gradient from $\mathrm{D} \cdot \mathrm{Å}^{-1} \cdot \mathrm{amu}^{-\frac{1}{2}}$ to the SI units $\mathrm{C} \cdot \mathrm{kg}^{-\frac{1}{2}}$.

2.4 Scattering spectroscopies

The theory behind infrared spectroscopy, a technique which falls into the category of absorption spectroscopies, was covered in the previous sections. In these types of spectroscopies the change in intensity of a beam of light before it enters the sample and after it exits it is monitored, but the frequencies of the incoming and outgoing light are identical. The spectroscopies described in the present section, however, fall into the category of scattering spectroscopy. Here, the photon entering the sample need not have the same energy as the one exiting it, and these spectroscopic techniques are thus examples of inelastic scattering processes.[1,58] In Raman spectroscopy, a photon of angular frequency ω_{in} within the visible spectrum is absorbed

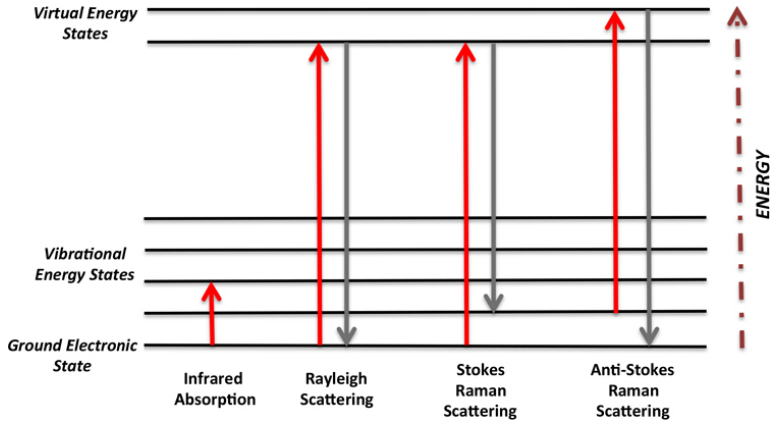


Figure 2.2. The various types of Raman scattering processes, as well as IR absorption spectroscopy.[59]

into the sample, exciting the system to a virtual electronic state. The sample then emits another photon with angular frequency ω_{out} which is close to, but not identical to ω_{in}. The difference between the angular frequencies of the exiting and entering photons, $|\omega_{\text{in}} - \omega_{\text{out}}| = \omega_i$ corresponds to a vibrational normal mode. Typically $\omega_{\text{in}} - \omega_{\text{out}}$ is positive, heightening the vibrational state of the molecule, and being characterized as Stokes scattering, but it can also be negative, which lowers the vibrational state, and is characterized as anti-Stokes scattering. The (less interesting) case where $\omega_{\text{in}} = \omega_{\text{out}}$ is called Rayleigh scattering, and is an example of elastic scattering. An illustration of the various types of Raman scattering processes can be seen in Fig. 2.2.

In Raman, the number of entering and exiting photons is the same, while for hyper-Raman, two photons enter the sample. These two photons combined cause an excitation to a virtual electronic state, and the process from there is identical to that of Raman, with one photon being emitted.[60] The vibrational motion observed thus has an angular frequency of $|2\omega_{\text{in}} - \omega_{\text{out}}| = \omega_i$. An illustration of hyper-Raman compared to Raman scattering can be seen in Fig. 2.3. Although this effect is much weaker than that of IR or Raman, it is still useful. A molecule might for instance be hyper-Raman active when it is not IR or Raman active.

Infrared intensity is closely related to the dipole transition moment, which is an energy-derivative of first order in field strength. Raman and

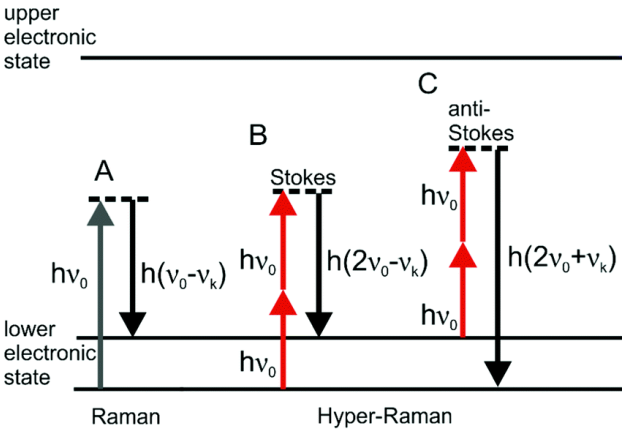


Figure 2.3. Hyper-Raman compared to Raman scattering. A) is Raman Stokes scattering, while B) and C) are hyper-Raman Stokes and anti-Stokes scattering, respectively.[61]

hyper-Raman spectroscopies, however, depend on second- and third-order properties, namely the transition polarizability and hyperpolarizability, respectively. The intensity-like quantity most commonly calculated is the absolute differential Raman scattering cross section ($\partial\sigma/\partial\Omega$) and is given by[42,62]

$$\frac{\partial\sigma}{\partial\Omega} = \frac{\hbar\left(\omega_{\text{in}} - \omega_i\right)^4}{2(2\pi c)^4\omega_i\left(1 - \exp\left(-\frac{\hbar\omega_i}{kT}\right)\right)}\left(k_a a_i^2 + k_b b_i^2\right) f(\omega;\omega_i,\gamma_i) \tag{2.4.1}$$

where Ω is the solid angle, k the Boltzmann constant, T the temperature in Kelvin and the Raman invariants a_i and b_i^2 are given by

$$a_i = \frac{1}{3}\sum_{\alpha=x,y,z}\left.\frac{\partial\alpha_{\alpha\alpha}}{\partial Q_i}\right|_{Q_i=0} \tag{2.4.2}$$

and

$$b_i^2 = \sum_{\alpha=x,y,z} \sum_{\beta\neq\alpha} \left(\frac{1}{2} \left(\left.\frac{\partial \alpha_{\alpha\alpha}}{\partial Q_i}\right|_{Q_i=0} - \left.\frac{\partial \alpha_{\beta\beta}}{\partial Q_i}\right|_{Q_i=0} \right)^2 + 3 \left(\left.\frac{\partial \alpha_{\alpha\beta}}{\partial Q_i}\right|_{Q_i=0} \right)^2 \right) \tag{2.4.3}$$

where $\alpha_{\alpha\beta}$ is the $\alpha\beta$'th Cartesian component of the polarizability $\boldsymbol{\alpha}$. The dependence on the polarizability gradients and no higher-order terms follow from the truncation of the Taylor expansion to first order in the harmonic approximation, as with the dipole moment in IR. There are three different choices for the coefficients k_a and k_b, and they are determined by the experimental setup. The origins of the various values of these coefficients can be studied in detail in Wilson, Decius and Cross,[38] but briefly explained, the choice $k_a = 0, k_b = 6$ belongs to the case where the incident light is polarized parallel to the direction of observation and perpendicular to the direction of propagation, and $k_a = 45, k_b = 7$ to the case where the incident light is polarized perpendicular both to the direction of observation and the direction of propagation. The subset of the last case which has light polarized parallel to the electric field, gives $k_a = 45, k_b = 4$. When reporting Raman results, it is always important to specify which polarization has been used.

The unit of $\partial\sigma/\partial\Omega$, as in Eq. (2.4.1), is $\mathrm{C}^4 \cdot \mathrm{J}^{-1} \cdot \mathrm{kg}^{-1} \cdot \mathrm{s}^3 \cdot \mathrm{m}^{-2}$. Other common units such as for instance $\text{Å}^4 \cdot \mathrm{amu}^{-1}$ are derived from only the polarizability gradient part $(k_a a_i^2 + k_b b_i^2)$ of Eq. (2.4.1).

For hyper-Raman spectra, we can similarly define a measure of the intensity through a scattering cross section related property, σ. In the same way as with Raman scattering, the experimental setup plays influences the outgoing intensity. Without going into details, there are two options for hyper-Raman, the vertically and horizontally polarized cases. The vertically polarized σ is given by[63]

$$\sigma_i^{VV} = \frac{\hbar(2\omega_{\mathrm{in}} - \omega_i)^4}{2\omega_i \left(1 - \exp\left(-\frac{\hbar\omega_i}{kT}\right)\right)} b_{\alpha\alpha\alpha,i}^2 \tag{2.4.4}$$

and the horizontal by

$$\sigma_i^{HV} = \frac{\hbar(2\omega_{\text{in}} - \omega_i)^4}{2\omega_i \left(1 - \exp\left(-\frac{\hbar\omega_i}{kT}\right)\right)} b^2_{\beta\alpha\alpha,i} \tag{2.4.5}$$

Comparing these two equations with the Raman equivalent, Eq. (2.4.1), hyper-Raman has a factor 2 in front of the incident frequency ω_{in}, while Raman does not. This originates from the previously stated fact that for hyper-Raman two incoming photons combine to cause one scattering process, while for Raman, only one incoming photon is involved.

Defining

$$\beta^i_{\alpha\beta\gamma} \equiv \left.\frac{\partial\beta_{\alpha\beta\gamma}}{\partial Q_i}\right|_{Q_i=0}, \tag{2.4.6}$$

where $\beta_{\alpha\beta\gamma}$ is element $\alpha\beta\gamma$ of the hyperpolarizability $\boldsymbol{\beta}$, the quantities $b^2_{\alpha\alpha\alpha,i}$ and $b^2_{\beta\alpha\alpha,i}$ are given by

$$\begin{aligned} b^2_{\alpha\alpha\alpha,i} =& \frac{1}{7}\sum_{\alpha}(\beta^i_{\alpha\alpha\alpha})^2 \\ &+ \frac{1}{35}\sum_{\alpha,\beta\neq\alpha}\Big(4(\beta^i_{\alpha\alpha\beta})^2 + 2\beta^i_{\alpha\alpha\alpha}\beta^i_{\alpha\beta\beta} + 4\beta^i_{\beta\alpha\alpha}\beta^i_{\alpha\alpha\beta} \\ &\qquad + 4\beta^i_{\alpha\alpha\alpha}\beta^i_{\beta\beta\alpha} + (\beta^i_{\beta\alpha\alpha})^2\Big) \\ &+ \frac{1}{105}\sum_{\alpha,\beta\neq\alpha,\gamma\neq\beta\neq\alpha}\Big(4\,\beta^i_{\alpha\alpha\beta}\beta^i_{\beta\gamma\gamma} + \beta^i_{\beta\alpha\alpha}\beta^i_{\beta\gamma\gamma} + 4\beta^i_{\alpha\alpha\beta}\beta^i_{\gamma\gamma\beta} \\ &\qquad + 2(\beta^i_{\alpha\beta\gamma})^2 + 4\beta^i_{\alpha\beta\gamma}\beta^i_{\beta\alpha\gamma}\Big) \end{aligned} \tag{2.4.7}$$

and

$$\begin{aligned} b^2_{\beta\alpha\alpha,i} =& \frac{1}{35}\sum_{\alpha}(\beta^i_{\alpha\alpha\alpha})^2 \\ &+ \frac{1}{105}\sum_{\alpha,\beta\neq\alpha}\Big(4\beta^i_{\alpha\alpha\alpha}\beta^i_{\alpha\beta\beta} + 8(\beta^i_{\alpha\alpha\beta})^2 - 6\beta^i_{\alpha\alpha\alpha}\beta^i_{\beta\beta\alpha} \\ &\qquad + 9\,(\beta^i_{\alpha\beta\beta})^2 - 6\beta^i_{\alpha\alpha\beta}\beta^i_{\beta\alpha\alpha}\Big) \\ &+ \frac{1}{105}\sum_{\alpha,\beta\neq\alpha,\gamma\neq\beta\neq\alpha}\Big(3\beta^i_{\alpha\beta\beta}\beta^i_{\alpha\gamma\gamma} - 2\beta^i_{\alpha\alpha\gamma}\beta^i_{\beta\beta\gamma} - 2\beta^i_{\alpha\alpha\beta}\beta^i_{\beta\gamma\gamma} \\ &\qquad + 6(\beta^i_{\alpha\beta\gamma})^2 - 2\beta^i_{\alpha\beta\gamma}\beta^i_{\beta\alpha\gamma}\Big) \end{aligned} \tag{2.4.8}$$

More details on experimental setups and thus a more thorough explanation the vertical and horizontal cases are outside the scope of this work. The reader is referred to Quinet *et al.*[63]

2.5 Anharmonic corrections

The above discussion is all within the framework of the harmonic approximation. This is based on the truncation of the Taylor expansion of the energy after second order of differentiation with respect to normal coordinates, using the Hessian to find vibrational frequencies. Polarization properties, such as the dipole moment, are truncated after first order. The harmonic approximation models the potential energy as a parabolic curve, while the actual shape is in fact more similar to that of the Morse potential,[1] shown in Fig. 2.4. The harmonic approximation is in other words quite good close to the equilibrium geometry, but breaks down quickly further away from the minimum. In order to better approximate the Morse potential, different methods have been developed, for instance by applying scaling factors.[64,65] However, a more accurate route (and the one chosen in this work), is to through perturbation theory include higher-order terms from the geometric Taylor expansion of the energy and polarization properties. This makes the vibrational wavefunctions slightly more complicated than those presented for the harmonic approximation in Eq. (2.2.3), although they are still based on Hermite polynomials.[66] To include these anharmonic terms in the vibrational frequencies, second-order vibrational perturbation theory (VPT2)[39,40,67,68] is often employed. Over the years, a variety of extensions and refinements to this theory have been developed[41–44], but in the works included here, we model the vibrational frequency with the general vibrational second-order perturbation theory (GVPT2), and the transition moments with a the deperturbed vibrational second-order perturbation theory (DVPT2).[42] These three theories will be explained in more detail in the next sections, but a short explanation of how they relate to each other is useful at this point. VPT2 provides the equations for the corrected frequencies and transition moments, however, many of the terms involved in these equations have denominators that in certain cases cause divergencies. This happens when certain harmonic frequencies, or specific combinations of them, are close to each other, a phenomenon called resonance. In VPT2 this is not considered,

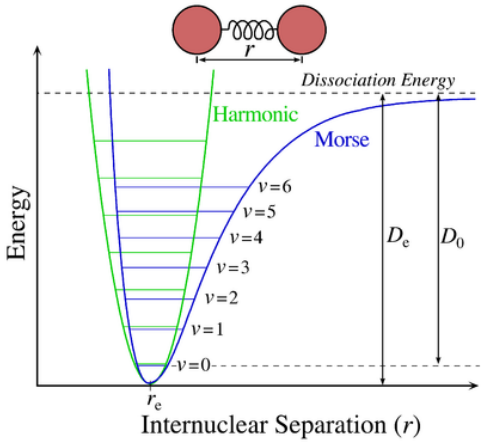


Figure 2.4. Morse potential energy curve (blue) and the potential energy curve of a harmonic oscillator (green), as a function of internuclear separation. Here, r_e is the equilibrium geometry, D_e is the dissociation energy, while D_0 is the actual energy required to completely break the bond. They are closely related, but not identical since the energy of the lowest vibrational level, $\nu = 0$, is nonzero. In addition to having different shapes, the harmonic and Morse potential curves differ in that while the harmonic oscillator has evenly spaced energy levels, this is not the case for the Morse potential.[69]

and the values are calculated directly from the equations, but in DVPT2 and GVPT2 these resonances are identified and the corresponding terms are then removed. This gives the DVPT2 results, but GVPT2 goes one step further, creating a new eigenvalue problem for the resonant terms and solving it to get further corrected results. The three methods can therefore be seen to build on each other, with VPT2 the most basic and GVPT2 the most refined. The following section summarizes key concepts and results, but for detailed derivations, refer to the original works.[39,42–44]

2.5.1 VPT2 theory

Before studying the theories in detail, a small clarification of the names might be helpful. VPT2 is called second-order, even though they utilize terms of up to fourth-order geometrical energy derivatives and third-order property derivatives. This is meaningful as these are perturbation theories made to correct the harmonic approximation, and consequently, the

zeroth-order term in the perturbation expansion is in fact based on the geometric second-order energy derivatives and first-order polarization property derivatives. When a second-order perturbative approach is added on top of this, fourth-order energy derivatives and third-order property derivatives are needed.

When now proceeding to tackle the corrected terms, we start with the corrected transition moments as expressed in the VPT2 formulation. The transition integral of a property $\boldsymbol{P}$ from state 0 to state n is defined as

$$\langle \boldsymbol{P} \rangle_{0n} = \frac{\langle \psi_0 | \boldsymbol{P} | \psi_n \rangle}{\sqrt{\langle \psi_0 | \psi_0 \rangle \langle \psi_n | \psi_n \rangle}} \tag{2.5.1}$$

where ψ are vibrational wave functions. We now employ perturbation theory to expand both ψ and $\boldsymbol{P}$ as

$$|\psi_k\rangle = |\psi_k^{(0)}\rangle + \lambda |\psi_k^{(1)}\rangle + \lambda^2 |\psi_k^{(2)}\rangle + \cdots \tag{2.5.2}$$

$$\boldsymbol{P} = \boldsymbol{P}^{(0)} + \lambda \boldsymbol{P}^{(1)} + \lambda^2 \boldsymbol{P}^{(2)} + \cdots \tag{2.5.3}$$

and truncate after second order. As already stated, $|\psi_k^{(0)}\rangle$ are the solutions to the eigenvalue problem in the double harmonic approximation. Inserting these expansions into Eq. (2.5.1) will give the formulation for the transition moment, but as Bloino and Barone[42] note, the resulting equation involves several thousand components, and we therefore here only repeat their final result. Defining α as a Cartesian component, for instance a Cartesian coordinate or an element of the polarizability tensor, and $1i$ as the corrected fundamental state, we get the following VPT2 result for component α of the corrected fundamental transition mode of property $\boldsymbol{P}$, here represented by the dipole moment $\boldsymbol{\mu}$, but general also for other electric polarization properties

$$
\begin{aligned}
\langle\mu^{\alpha}\rangle_{0,1i} =& \frac{1}{\sqrt{2}}\mu_i^{\alpha} + \frac{1}{4\sqrt{2}}\sum_j \mu_{ijj}^{\alpha} - \frac{1}{8\sqrt{2}}\sum_{jk}\left[k_{ijkk}\mu_j^{\alpha}\left(\frac{1}{w_i+w_j} - \frac{1-\delta_{ij}}{w_i-w_j}\right)\right. \\
&\left. + k_{ijk}\mu_{jk}^{\alpha}\left(\frac{1}{w_i+w_j+w_k} - \frac{1}{w_i-w_j-w_k}\right) + \frac{2}{w_j}k_{jkk}\mu_{ij}^{\alpha}\right] \\
&+ \frac{1}{2\sqrt{2}}\sum_{jk}\left(\sum_{\tau} B_{eq}^{\tau}\zeta_{ik}^{\tau}\zeta_{jk}^{\tau}\right)\mu_j^{\alpha}\left[\frac{\sqrt{w_i w_j}}{w_k}\left(\frac{1}{w_i+w_j} + \frac{1-\delta_{ij}}{w_i-w_j}\right)\right. \\
&\left. - \frac{w_k}{\sqrt{w_i w j}}\left(\frac{1}{w_i+w_j} - \frac{1-\delta_{ij}}{w_i-w_j}\right)\right] \\
&+ \frac{1}{16\sqrt{2}}\sum_{jkl}\left(k_{ikl}k_{jkl}\mu_j^{\alpha}\left[(1-\delta_{ij})(1-\delta_{ik})(1-\delta_{il})\right.\right. \\
&\times \left(\frac{1}{(w_i+w_j)(w_j+w_k+w_l)} - \frac{1}{(w_i-w_j)(w_j+w_k+w_l)}\right. \\
&+ \frac{1}{(w_i+w_k+w_l)(w_j+w_k+w_l)} - \frac{1}{(w_i-w_k-w_l)(w_j+w_k+w_l)} \\
&\left. + \frac{1}{(w_i-w_j)(w_i-w_k-w_l)} + \frac{1}{(w_i+w_j)(w_i+w_k+w_l)}\right) \\
&+ \delta_{ij}(1+\delta_{ik})(1-\delta_{il})\left(\frac{1}{2w_i(w_i+w_k+w_l)} - \frac{1}{2w_i(w_i-w_k-w_l)}\right. \\
&\left. + \frac{1}{2(w_i+w_k+w_l)^2} - \frac{1}{2(w_i-w_k-w_l)^2}\right) \\
&+ (1-\delta_{ij})(1-\delta_{ik})\delta_{il}\left(\frac{1}{w_k(w_i+w_j)} + \frac{2}{(2w_i+w_k)(w_i+w_j)}\right. \\
&+ \frac{3}{(w_i+w_j)(w_i+w_j+w_k)} + \frac{1}{(w_i-w_j)(w_i-w_j-w_k)} \\
&- \frac{2}{(w_i-w_j)(w_i+w_j+w_k)} - \frac{3}{w_k(w_i-w_j)} \\
&- \frac{1}{w_k(w_i-w_j-w_k)} + \frac{2}{(2w_i+w_k)(w_i+w_j+w_k)} \\
&\left.\left. + \frac{3}{w_k(w_i+w_j+w_k)}\right)\right]
\end{aligned}
$$

continued... (2.5.4)

$$
\begin{aligned}
\text{continued...} &+ k_{ijk}k_{llk}\mu_j^\alpha \left[\frac{\delta_{ij}}{w_i w_k}\left(1+\frac{2\delta_{ik}\delta_{il}}{9}\right)\right. \\
&+ (1-\delta_{ij})(1-\delta_{ik})(1-\delta_{il})\left(\frac{1}{(w_i+w_j)(w_i+w_j+w_k)}\right. \\
&+ \frac{1}{w_k(w_i+w_j)} - \frac{1}{w_k(w_i-w_j)} + \frac{1}{(w_i-w_j)(w_i-w_j-w_k)} \\
&+ \left.\frac{1}{w_k(w_i+w_j+w_k)} - \frac{1}{w_k(w_i-w_j-w_k)}\right) \\
&+ \delta_{ik}(1-\delta_{ij})\left((1+\delta_{il})\left[\frac{1}{(2w_i+w_j)(w_i+w_j)} - \frac{1}{w_i(w_i-w_j)}\right.\right. \\
&+ \left.\frac{1}{w_i(2w_i+w_j)}\right] \\
&+ \delta_{il}\left[\frac{1}{3w_i(w_i+w_j)} + \frac{1}{3w_i(2w_i+w_j)} - \frac{1}{(w_i-w_j)(2w_i+w_j)}\right] \\
&+ \left.\left.\frac{1}{w_i(w_i+w_j)} - \frac{1}{w_j(w_i-w_j)} + \frac{1}{w_i w_j}\right)\right]\Bigg)
\end{aligned}
$$

The notation $\boldsymbol{\mu}_i$, $\boldsymbol{\mu}_{ij}$ denote $\boldsymbol{\mu}$ differentiated with respect to $\hat{Q}_i$ and with respect to $\hat{Q}_i$ and $\hat{Q}_j$, respectively. The coordinate basis $\hat{Q}_i$ is dimensionless and related to the Q_i basis through

$$\hat{Q}_i = \sqrt{\frac{\omega_i}{\hbar}} Q_i \tag{2.5.5}$$

where ω_i is, as before, the harmonic angular frequency of normal mode i. From here on, the coordinate basis $\hat{Q}_i$ will be referred to as the reduced normal coordinate basis. If $\boldsymbol{T}$ is the transformation from Cartesian coordinates to Q_i, the double transformation matrix from Cartesian basis to reduced coordinate basis is given by

$$\hat{T}_{\alpha,i} = \sqrt{\frac{\hbar}{\omega_i}} T_{\alpha,i} \tag{2.5.6}$$

In a similar fashion, k_{ijk} and k_{ijkl} are the third- and fourth-order derivatives of the potential energy, B^τ_{eq} is the diagonal inertia tensor of the molecule at equilibrium geometry and ζ^τ_{ij} are Coriolis constants. The vibrational modes can couple to the molecular rotation through the Coriolis coupling,

Table 2.1. Conversion between the various frequency-related properties.

	Angular frequency ω [1/s]	Frequency v [1/s]	Wavenumber $\bar{v}$ [1/m]	Energy w [J]
ω [1/s]	1	$\omega = 2\pi v$	$\omega = 2\pi c\bar{v}$	$\omega = 2\pi w/h$
v [1/s]	$v = \omega/2\pi$	1	$v = c\bar{v}$	$v = w/h$
$\bar{v}$ [1/m]	$\bar{v} = \omega/2\pi c$	$\bar{v} = v/c$	1	$\bar{v} = w/hc$
w [J]	$w = h\omega/2\pi$	$w = hv$	$w = hc\bar{v}$	1

and may lead to intensity borrowing as well as changes to the vibrational energy levels. Such effects are expected to most prominent for small, highly symmetrical molecules, and to be of lower importance for solvated systems, and are neglected in the current implementation. In Eq. (2.5.4), w is the vibrational frequencies in energy-units. An overview of how the different frequency-related properties relate to each other can be found in Table 2.1. Note that the equations presented here, hold only for electric polarization properties. A general formulation that includes also magnetic polarization properties can be found in Ref. 44, which corrects a minor mistake found in the original work (Ref. 42).

With VPT2, in addition to the singly excited fundamental bands, doubly excited features are also observed. These are categorized as either overtones, where a state corresponding to one mode has been doubly excited, or combination bands, where states from two different modes each have been excited once. The non-vanishing nature of transitions to these states when using methods that go beyond the harmonic approximation can also be explained through a similar line of argumentation as that given in Section 2.2, but with the exception that the Taylor expansions is not truncated after first order. When including the second-order term, integrals of the type $\langle 0_i|Q_i^2|2_i\rangle$ and $\langle 0_i|Q_i|1_i\rangle\langle 0_j|Q_j|1_j\rangle$ appear and can be non-zero. A transition to overtone state 2_i is therefore allowed in the VPT2 approximation. The VPT2 approach has also been extended in order to be able to model three-quanta transitions,[44] but this phenomenon has not been included in this thesis.

The overtone transition moments $\langle\mu^\alpha\rangle_{0,2_i}$ and combination band transition moments $\langle\mu^\alpha\rangle_{0,1_i1_j}$ are given by

$$\langle\mu^{\alpha}\rangle_{0,(1+\delta ij)_i(1-\delta_{ij})_j} = \left(\sqrt{2}+(1-\sqrt{2})\delta_{ij}\right)\left[\frac{1}{2\sqrt{2}}\mu^{\alpha}_{ij} + \frac{1}{4\sqrt{2}}\sum_k k_{ijk}\mu^{\alpha}_k\left(\frac{1}{w_i+w_j-w_k}-\frac{1}{w_i+w_j+w_k}\right)\right] \tag{2.5.7}$$

Having formulated the anharmonically corrected transition moments, we will now similarly look at the corrected frequencies. Using the previously mentioned the VPT2 approach, the corrected fundamental, overtone- and combotone-frequencies in energy units, w_{1_i}, w_{2_i} and $w_{1_i1_j}$, respectively, are given by

$$w_{1_i} = w_i + 2X_{ii} + \frac{1}{2}\sum_{j\neq i} X_{ij}, \tag{2.5.8}$$

$$w_{2_i} = 2w_{1_i} + 2X_{ii} \tag{2.5.9}$$

and

$$w_{1_i1_j} = w_{1_i} + w_{1_j} + X_{ij} \tag{2.5.10}$$

where X_{ii} and X_{ij} are given by

$$X_{ii} = \frac{k_{iiii}}{16} - \sum_k \frac{k^2_{iik}}{32}\left(\frac{4}{w_k}+\frac{1}{2w_i+w_k}-\frac{1}{2w_i-w_k}\right) \tag{2.5.11}$$

and

$$\begin{aligned} X_{ij} =& \frac{k_{iijj}}{4} - \frac{1}{4}\sum_k \frac{k_{iik}k_{jjk}}{w_k} + \sum_\tau B^{\tau}_{eq}\left(\zeta^{\tau}_{ij}\right)^2\left(\frac{w_i}{w_j}+\frac{w_j}{w_i}\right) \\ &- \sum_k \frac{k^2_{ijk}}{8}\left(\frac{1}{w_i+w_j+w_k}+\frac{1}{-w_i+w_j+w_k}+\frac{1}{w_i-w_j+w_k}\right. \\ &\left. - \frac{1}{w_i+w_j-w_k}\right) \end{aligned} \tag{2.5.12}$$

2.5.2 Spectroscopic intensities for both the harmonic and anharmonic case

In Section 2.5.1 the VPT2 formulation of anharmonic corrected transition moments and frequencies was presented. A key point in that section was the fact that these properties were derived in a reduced coordinate basis. The intensity-equations outlined in Sections 2.3 and 2.4, however, are formulated in non-reduced coordinates. A strategy to get around this difference, and also make the equations hold both in the harmonic and anharmonic case, is to reformulate the intensity-equations for IR, Raman and hyper-Raman in terms of transition moments, as will be done in this section.

The harmonic IR molar decadic attenuated coefficient is given in Eq. (2.3.7), and the harmonic transition dipole moment for mode i in Eq. (2.2.13). When Eq. (2.2.13) is inserted into Eq. (2.3.7), we get the molar decadic attenuated coefficient in terms of transition dipole moment, given by

$$\varepsilon_i = \frac{\pi N_A \omega_i}{3\ln(10)\varepsilon_0 c\hbar} \sum_{\alpha=x,y,z} |\langle 0|\mu_\alpha|n_i\rangle|^2 f(\omega;\omega_i,\gamma_i) \tag{2.5.13}$$

Alternatively, this could have been found by combining Eqs. (2.3.2) and (2.3.3), and is not restricted to the harmonic approximation.

Similarly the generalized form of the integrated Napierian attenuated coefficient for mode i in Eq. (2.3.12) becomes

$$A_i = \frac{N_A \omega_i}{6\varepsilon_0 c^2 \hbar} \sum_{\alpha=x,y,z} \langle 0|\mu_\alpha|n_i\rangle|^2 \tag{2.5.14}$$

The harmonic Raman absolute differential scattering coefficient is given in Eq. (2.4.1). In a similar fashion as with IR, we wish to formulate our equations in terms of the transition polarizability moment. Therefore, we define the transition moment equivalents of a_i and b_i^2 as

$$\langle a_i\rangle \equiv \frac{1}{3}\sum_a \langle 0|\alpha_{\alpha\alpha}|n_i\rangle \tag{2.5.15}$$

and

$$\langle b_i^2\rangle \equiv \sum_\alpha \sum_{\beta\neq a} \left(\frac{1}{2}\left(\langle 0|\alpha_{\alpha\alpha}|n_i\rangle - \langle 0|\alpha_{\beta\beta}|n_i\rangle\right)^2 + 3|\langle 0|\alpha_{\alpha\beta}|n_i\rangle|^2\right) \tag{2.5.16}$$

Using that the harmonic polarizability transition moment is given by

$$\langle 0|\alpha_{\alpha\beta}|n_i\rangle = \sqrt{\frac{\hbar}{2\omega_i}}\frac{\partial\alpha_{\alpha\beta}}{\partial Q_i}\bigg|_{Q_i=0} \tag{2.5.17}$$

this allows us to formulate a generalized Raman absolute differentiated scattering cross section as

$$\begin{aligned}\frac{\partial\sigma}{\partial\Omega} =& \frac{\hbar\left(\omega_{\text{in}}-\omega_i\right)^4}{2(2\pi c)^4\omega_i\left(1-\exp\left(-\frac{\hbar\omega_i}{kT}\right)\right)}\frac{2\omega_i}{\hbar}\left(k_a\langle a_i\rangle^2+k_b\langle b_i^2\rangle\right)f(\omega;\omega_i,\gamma_i)\\ =& \frac{\left(\omega_{\text{in}}-\omega_i\right)^4}{(2\pi c)^4\left(1-\exp\left(-\frac{\hbar\omega_i}{kT}\right)\right)}\left(k_a\langle a_i\rangle^2+k_b\langle b_i^2\rangle\right)f(\omega;\omega_i,\gamma_i)\end{aligned} \tag{2.5.18}$$

For hyper-Raman, we will look only at the vertically polarized (VV) form found in Eq. (2.4.4), but the strategy will be identical for the horizontally polarized (HV) terms. Defining

$$\begin{aligned}\langle b^2_{\alpha\alpha\alpha,i}\rangle =& \frac{1}{7}\sum_\alpha\langle 0|\beta_{\alpha\alpha\alpha}|n_i\rangle^2\\ &+\frac{1}{35}\sum_{\alpha,\beta\neq\alpha}\Big(\ 4\langle 0|\beta_{\alpha\alpha\beta}|n_i\rangle^2+2\langle 0|\beta_{\alpha\alpha\alpha}|n_i\rangle\langle 0|\beta_{\alpha\beta\beta}|n_i\rangle\\ &\qquad+4\langle 0|\beta_{\beta\alpha\alpha}|n_i\rangle\langle 0|\beta_{\alpha\alpha\beta}|n_i\rangle\\ &\qquad+4\langle 0|\beta_{\alpha\alpha\alpha}|n_i\rangle\langle 0|\beta_{\beta\beta\alpha}|n_i\rangle\\ &\qquad+\langle 0|\beta_{\beta\alpha\alpha}|n_i\rangle^2\Big)\\ &+\frac{1}{105}\sum_{\alpha,\beta\neq\alpha,\gamma\neq\beta\neq\alpha}\Big(\ 4\langle 0|\beta_{\alpha\alpha\beta}|n_i\rangle\langle 0|\beta_{\beta\gamma\gamma}|n_i\rangle\\ &\qquad+\langle 0|\beta_{\beta\alpha\alpha}|n_i\rangle\langle 0|\beta_{\beta\gamma\gamma}|n_i\rangle\\ &\qquad+4\langle 0|\beta_{\alpha\alpha\beta}|n_i\rangle\langle 0|\beta_{\gamma\gamma\beta}|n_i\rangle+2\langle 0|\beta_{\alpha\beta\gamma}|n_i\rangle^2\\ &\qquad+4\ \langle 0|\beta_{\alpha\beta\gamma}|n_i\rangle\langle 0|\beta_{\beta\alpha\gamma}|n_i\rangle\Big)\end{aligned} \tag{2.5.19}$$

results in a generalized scattering cross section given by

$$\sigma_i^{VV} = \frac{(2\omega_{\text{in}}-\omega_i)^4}{[1-\exp(-\hbar\omega_i/kT)]}\langle b^2_{\alpha\alpha\alpha,i}\rangle \tag{2.5.20}$$

2.5.3 Identification and treatment of divergencies caused by resonance

Before finishing the sections on anharmonic corrections, the subject of resonances should be further explored. As previously mentioned, resonances are caused when a combination of harmonic frequencies in a term's denominator approach zero, causing divergence. There are terms with such possible divergencies in the equations for both the corrected vibrational energies and for the corrected transition moments, and their identification and treatment are covered in this section.

For the frequencies, the possible source of divergencies exist in the $\mathbf{X}$ matrix. For X_{ii} there is a source for divergence in the last term of Eq. (2.5.11) if $2w_i$ approaches w_k, and for X_{ij} all the three last terms in Eq. (2.5.12) can cause divergence if the sum of two energies get too close to the last. In the case of the last term, this would be if $w_i + w_j$ got too close to w_k. This type of divergence, where the sum of two energies closely correspond to a third is called a Fermi resonance. As we have just seen, the corrected energies can be affected by this phenomenon, and looking at Eqs. (2.5.4) and (2.5.7) it is apparent that it can also affect the corrected transition moments.

In order to avoid such divergencies, it is important to identify Fermi resonances. This is done by a double test as suggested by Martin, Bloino and Barone,[39,40,42]

$$\begin{aligned} |2\bar{v}_i - \bar{v}_k| \leq 200 \text{ cm}^{-1}; \quad |\Delta_{ik}| \geq 1 \text{ cm}^{-1} \\ |\bar{v}_i + \bar{v}_j - \bar{v}_k| \leq 200 \text{ cm}^{-1}; \quad |\Delta_{ijk}| \geq 1 \text{ cm}^{-1} \end{aligned} \tag{2.5.21}$$

where the Martin parameters Δ_{ik} and Δ_{ijk} are given by

$$\Delta_{ik} = \frac{k_{iik}^4}{25600 \cdot h^4 c^4 (2\bar{v}_i - \bar{v}_k)^3} \tag{2.5.22}$$

and

$$\Delta_{ijk} = \frac{k_{ijk}^4}{6400 \cdot h^4 c^4 (\bar{v}_i + \bar{v}_j - \bar{v}_k)^3} \tag{2.5.23}$$

If these tests are positive, then a Fermi resonance has been identified. There is a choice regarding how to proceed with the corrected energies. A common and efficient treatment has been to simply discard the affected terms, corresponding to the DVPT2 method, and is what has in this work be done for the transition moments. The GVPT2 method, which we use for frequencies, goes one step further. After having discarded the terms for which a resonance is identified through the DVPT2 method, a matrix is built with the DVPT2 frequencies in energy units on the diagonal, as well as scaled cubic force-constants on the off-diagonal elements belonging to the resonances. An eigenvalue analysis is then performed on this matrix, and the GVPT2 frequencies are found from the resulting eigenvalues.[39,40] Although not applied here, a method for a similar treatment of transition moments can be found in Vázquez and Stanton[70].

In addition to the Fermi resonances that can be found in all the equations for anharmonic corrections, there is another cause for divergence, called 1-1 resonance. This is when two energies come too close to each other, $w_i \approx w_j$, and can be seen to affect quite a few terms in the fundamental transition moment, Eq. (2.5.4). A tree-fold test is performed, where[43]

$$|\bar{v}_i - \bar{v}_j| \leq 100 \text{ cm}^{-1} \tag{2.5.24}$$

has to be true, and at least one of

$$|K| \geq 1 \text{ cm}^{-1} \tag{2.5.25}$$

and

$$|K|/(\bar{v}_i - \bar{v}_j)^2 \geq 1 \text{ cm} \tag{2.5.26}$$

has to be true in order for a 1-1 resonance to be identified. K is a measure of the coupling in the term, for instance k_{ijkk} in the third term in Eq. (2.5.4). In this example, $K = k_{ijkk}/(100hc)$. In the term containing the Coriolis coupling, $K = \sum_\tau B^\tau_{eq}\zeta^\tau_{ik}\zeta^\tau_{jk}/(100hc)$, and for terms with a product of cubic terms, $K = k_{ijk}k_{lmn}/(100hc)^2$. Note that in this last case, the unit cm^{-1} in Eq. (2.5.25) and 1cm in Eq. (2.5.26) should be exchanged with cm^{-2} and no unit, respectively. When a 1-1 resonance is identified, the corresponding

term is discarded. Over the years, different versions of the 1-1 resonance tests have been developed and used[42,43], but the one laid out here, and implemented in the SpectroscPy code[71], is in accordance with the work by Bloino *et al.* in Ref 43.

In this chapter, various types of vibrational spectroscopy as well as various levels of approximations of these have been covered. All of this theory has been developed with the aim to reproduce experimental results. It is therefore important that the computational cost is not so high that although the theory is in place, no interesting systems can be studied for practical reasons. Chapter 3 covers the Polarizable Embedding model which, as other QM/MM models, combines the accuracy of QM methods with the efficiency of MM methods, allowing theoretical studies to be made with systems larger than only a few small molecules.

Chapter 3

The Polarizable Embedding method

The subject of vibrational spectroscopy was covered in detail in Chapter 2, and we now turn our attention to the second main subject of this thesis, how to calculate these properties for realistic systems. If our system is a solution or a large biomolecule, optimizing the geometry or calculating energy derivatives with a QM method is far too demanding. One solution to this problem could be to isolate the part of the total system that we deem most important to the vibrational motion producing the spectrum, and calculate properties for this small part alone. Although such vacuum calculations can be highly informative, the removal of a larger part of the system often affects the results, in a direction away from experimental values, as the solvent can strongly impact both the structure and molecular properties of the solvent. Most dramatically, if the isolated part previously was connected to a larger structure through covalent bonds, as in many biomolecules, cutting these covalent bonds and removing the environment will have a large impact. Although weaker, a solvated system would experience similar effects because of the removal of electrostatic and Van der Waals interactions.

Being able to include the environmental effects, while not dramatically increasing the computational cost has been the subject of many developments. As mentioned in the introduction to this thesis, focused embedding methods have been developed for this purpose. By choosing a QM/MM strategy, we can combine the accuracy of QM methods with the low cost

of classical methods, while simultaneously being able to account for specific interactions between atoms in the core and environment that are lost in continuum models. In this work, the sole focus is on the Polarizable Embedding[22,72] (PE) method. For a broader perspective on focused embedding, QM/MM and their use in spectroscopic calculations, the reader is referred to a list of reviews.[13,21,73,74]

3.1 PE ground state formalism

As Polarizable Embedding falls into the category of focused embedding methods, the point of view is a specific part of the system, the core, in the presence of the rest of the system, the environment. Although it is possible, the total energy of the system is therefore not calculated in a focused embedding method, only the energy of the core and the interaction energy between the core and the environment. The energy of the core is calculated with a QM method, typically with KS–DFT, although development has also been done with the Coupled Cluster and Multi-Configuration SCF methods.[75,76] The environment is divided into fragments (for a solvent one per molecule) containing sites (one for each atom). In order to model both permanent and induced electrostatic interactions, the environment is constructed by placing an expansion of multipoles and polarizabilities on each atomic site. These parameters can either be taken from standard potentials or be explicitly calculated by separate QM calculations on each environmental molecular fragment.[77] This strategy ensures that PE on the one hand is completely independent of any pre-parametrization and can therefore be employed on any system, while it on the other hand preserves a high degree of the molecular structure of the environment. Parts of Chapter 3 might closely resemble parts of the articles in this work, but are included in order for this introduction to provide a complete overview. It also relies heavily on Olsen's thesis[78] and earlier papers from our group.[25,33]

The combined energy E of the core, calculated with DFT, and the interaction energy is expressed as

$$E(\mathbf{D}) = E_{\mathrm{DFT}}(\mathbf{D}) + E_{\mathrm{PE}}(\mathbf{D}) \tag{3.1.1}$$

where $\mathbf{D}$ is the density matrix of the core region in the atomic orbital

(AO) basis, E_{DFT} is the KS–DFT energy of the core region and E_{PE} is the interaction energy between the core region and the atomic sites in the environment. E_{DFT} is the same quantity as in Eq. (1.0.11), although the version used in our PE approach is formulated in terms of the AO basis,[25]

$$E_{\mathrm{DFT}}(\mathbf{D}) \overset{\mathrm{Tr}}{=} \mathbf{hD} + \frac{1}{2}\mathbf{G}^{\gamma}(\mathbf{D})\mathbf{D} + E_{\mathrm{xc}}[\rho(\mathbf{D})] + h_{\mathrm{nuc}} \tag{3.1.2}$$

where $\overset{\mathrm{Tr}}{=}$ denotes that the trace should be taken of all the terms on the right-hand side of the equation, $\mathbf{hD}$ contains the one-electron terms, $\mathbf{G}^{\gamma}(\mathbf{D})\mathbf{D}$ the two-electron terms, $E_{\mathrm{xc}}[\rho(\mathbf{D})]$ is the exchange-correlation energy and h_{nuc} is the nuclear repulsion energy. The reader is referred to the brief discussion on DFT in Chapter 1, but beyond that, this subject is outside the scope of this work.

The second term in Eq. (3.1.1), $E_{\mathrm{PE}}(\mathbf{D})$, is a sum of the permanent and induced electrostatic, as well as non-electrostatic, interactions between the core region and the environment, and is written as

$$E_{\mathrm{PE}}(\mathbf{D}) = E_{\mathrm{PE}}^{\mathrm{es}}(\mathbf{D}) + E_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D}) + E_{\mathrm{PE}}^{\mathrm{rep}} + E_{\mathrm{PE}}^{\mathrm{disp}} \tag{3.1.3}$$

where $E_{\mathrm{PE}}^{\mathrm{es}}(\mathbf{D})$ is the electrostatic, $E_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})$ the induction, $E_{\mathrm{PE}}^{\mathrm{rep}}$ the repulsion and $E_{\mathrm{PE}}^{\mathrm{disp}}$ the dispersion contributions to the embedding energy. Not that the terms "electrostatic" and "induction" might be a bit misleading, as they refer to permanent electrostatic and induced electrostatic interactions, respectively, and therefore both being of an electrostatic nature. However, this terminology has manifested itself within PE theory, and will therefore be used throughout the remainder of this work.

The electrostatic energy contains the interactions of the electrons and nuclei in the core region with the permanent multipoles in the environment

$$\begin{aligned} E_{\mathrm{PE}}^{\mathrm{es}}(\mathbf{D}) = &- \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{\mu\nu\in c} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) D_{\mu\nu} \\ &+ \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{n\in c}^{N_{\mathrm{nuc}}} T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n) Z_n \\ &\overset{\mathrm{Tr}}{=} \mathbf{h}_{\mathrm{PE}}^{\mathrm{es}}\mathbf{D} + h_{\mathrm{PE}}^{\mathrm{es}} \end{aligned} \tag{3.1.4}$$

The summation index a runs over the number of fragments N_{frag} in the environment, and summation index s runs over the S_a sites within fragment a. The absolute value of the multi-index α runs from zero to the highest order of multipole K_s present on site s. The multi-index notation is one where both the order of the multipole and its Cartesian orientation is expressed through $\alpha = (\alpha_x, \alpha_y, \alpha_z)$. A charge is in this notation given by $\alpha = (0,0,0)$, with $|\alpha| = 0$ while the y-component of a dipole moment is given by $\alpha = (0,1,0)$, with $|\alpha| = 1$.[78,79] The factorial $\alpha! = \alpha_x!\alpha_y!\alpha_z!$, and $M_s^{[\alpha]}$ is the multipole on site s of multipole multi-index α. The summation indices μ and ν run over the atomic orbitals in the core region c, and the summation index n runs over the N_{nuc} nuclei in the core region. As in Chapter 1, the charge of nucleus n is Z_n and $D_{\mu\nu}$ is the $\mu\nu$'th element of the density matrix. The interaction between the electrons and environment site s is modelled using a one-electron interaction integral $t_{\mu\nu}^{[\alpha]}$

$$t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) = \int \chi_\mu(\mathbf{r};\mathbf{r}_n)\partial_{\mathbf{r}}^{\alpha}\frac{1}{|\mathbf{r}-\mathbf{r}_s|}\chi_\nu(\mathbf{r};\mathbf{r}_n)d\mathbf{r} \tag{3.1.5}$$

where $\mathbf{r}_s$ is the position of site s, χ is an atomic orbital, $\mathbf{r}$ and $\mathbf{r}_n$ are the positions of the electrons and nuclei, respectively. The operator $\partial_{\mathbf{r}}^{\alpha}$ is a multi-index partial derivative operator that in general is defined as

$$\partial_{\mathbf{r}}^{\alpha} \equiv \frac{\partial^{|\alpha|}}{\partial^{\alpha_x}\partial^{\alpha_y}\partial^{\alpha_z}} \tag{3.1.6}$$

The subscript $\mathbf{r}$ indicates that it is the electronic coordinate that is subject to differentiation. The last term in Eq. (3.1.5) is the interaction tensor between site s and nucleus n, $T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n)$, defined generally by

$$T^{[\alpha]}(\mathbf{r}_i, \mathbf{r}_j) \equiv \partial_{\mathbf{r}_j}^{\alpha}\frac{1}{|\mathbf{r}_j - \mathbf{r}_i|} \tag{3.1.7}$$

The second term in Eq. (3.1.3) is the induction energy, originating from the induced multipoles in the environment interacting with the potential generated by the core, as well as from the other environment sites. The induction energy is given by

$$E_{\text{PE}}^{\text{ind}} = \frac{1}{2}\bar{\mathcal{M}}^T\mathcal{V} \tag{3.1.8}$$

where the matrices $\bar{\mathcal{M}}$ and $\mathcal{V}$ contain the induced multipoles and the potentials as blocks of polytensors, one for each site in the environment

$$\bar{\mathcal{M}} = \begin{pmatrix} \bar{\mathbf{M}}_1 \\ \bar{\mathbf{M}}_2 \\ \vdots \\ \bar{\mathbf{M}}_S \end{pmatrix}, \mathcal{V} = \begin{pmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_S \end{pmatrix} \tag{3.1.9}$$

Each of the elements $\mathbf{V}_s$ is a polytensor,[80] containing all the orders of $|\alpha|$, at site s

$$\mathbf{V}_s = \begin{pmatrix} \mathbf{V}_s^{(0)} \\ \mathbf{V}_s^{(1)} \\ \frac{1}{2}\mathbf{V}_s^{(2)} \\ \vdots \\ \frac{1}{K!}\mathbf{V}_s^{(K)} \end{pmatrix} \tag{3.1.10}$$

where $\mathbf{V}_s^{(i)}$ contains all the combinations of multi-indices α for which $|\alpha| = i$. The induced multipoles are expressed through an interaction matrix $\mathcal{A}$, and the potentials,

$$\bar{\mathcal{M}} = -\mathcal{A}^{-1}\mathcal{V} \tag{3.1.11}$$

where $\mathcal{A}$ is composed of the polytensors containing localized polarizabilites and interaction tensors, giving

$$\mathcal{A} = \begin{pmatrix} \mathbf{P}_1^{-1} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1S} \\ \mathbf{T}_{21} & \mathbf{P}_2^{-1} & \cdots & \mathbf{T}_{2S} \\ \vdots & & \ddots & \vdots \\ \mathbf{T}_{S1} & \cdots & \cdots & \mathbf{P}_S^{-1} \end{pmatrix} \tag{3.1.12}$$

The polarizabilities $\mathbf{P}_s$ are defined as two-dimensional polytensors

$$\mathbf{P}_s = \begin{pmatrix} \mathbf{P}_s^{(0,0)} & \mathbf{P}_s^{(0,1)} & \cdots & \mathbf{P}_s^{(0,L)} \\ \mathbf{P}_s^{(1,0)} & \mathbf{P}_s^{(1,1)} & \cdots & \mathbf{P}_s^{(1,L)} \\ \vdots & & \ddots & \vdots \\ \mathbf{P}_s^{(K,0)} & \cdots & \cdots & \mathbf{P}_s^{(K,L)} \end{pmatrix} \tag{3.1.13}$$

and the interaction tensors $\mathbf{T}_{ss'}$ are similarly

$$\mathbf{T}_{ss'} = \begin{pmatrix} \mathbf{T}_{ss'}^{(0,0)} & -\mathbf{T}_{ss'}^{(0,1)} & \cdots & (-1)^L\mathbf{T}_{ss'}^{(0,L)} \\ \mathbf{T}_{ss'}^{(1,0)} & -\mathbf{T}_{ss'}^{(1,1)} & \cdots & (-1)^L\mathbf{T}_{ss'}^{(1,L)} \\ \vdots & & \ddots & \vdots \\ \mathbf{T}_{ss'}^{(K,0)} & \cdots & \cdots & (-1)^L\mathbf{T}_{ss'}^{(K,L)} \end{pmatrix} \tag{3.1.14}$$

In this formulation, the two numbers in parenthesis denote which type of polytensors interact through the polarizability or interaction tensor. For instance, $\mathbf{P}_s^{(0,1)}$ is the polarizability corresponding to a potential polytensor with a multi-index with an absolute value of 0 interacting with one with a multi-index with an absolute value of 1. The case of $\mathbf{P}_s^{(1,1)}$, corresponds to the situation where a field interacts with another field through the polarizability.

Defining a matrix $\mathcal{B} = \mathcal{A}^{-1}$ and using the fact that $\mathcal{A}$ is symmetric ($\mathbf{B}_{s's} = \mathbf{B}_{ss'}$), the induction energy can be formulated as

$$\begin{aligned} E_{\mathrm{PE}}^{\mathrm{ind}} =& \frac{1}{2}\bar{\mathcal{M}}^T\mathcal{V} = \frac{1}{2}\sum_a^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\bar{\mathbf{M}}_s^T\mathbf{V}_s = -\frac{1}{2}\sum_a^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_b^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\mathbf{V}_{s'}^T\mathbf{B}_{s's}\mathbf{V}_s \\ =& -\frac{1}{2}\sum_a^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_b^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{i=0}^{K_s}\sum_{j=0}^{K_{s'}}\frac{1}{i!j!}(\mathbf{V}_{s'}^{(j)})^T\mathbf{B}_{s's}^{(j,i)}\mathbf{V}_s^{(i)} \\ =& -\frac{1}{2}\sum_a^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_b^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}V_{s'}^{[\beta]}B_{s's}^{[\beta,\alpha]}V_s^{[\alpha]} \end{aligned} \tag{3.1.15}$$

The potential experienced a site s of a specific mulit-index α, $V_s^{[\alpha]}$, is generated by the electrons and nuclei in the core as well as the multipoles in the $N_{\mathrm{frag}} - 1$ other fragments. It can therefore be formulated in terms of its origins,

$$V^{[\alpha]}(\mathbf{D}, \mathbf{r}_s) = V^{\mathrm{el},[\alpha]}(\mathbf{D}, \mathbf{r}_s) + V^{\mathrm{nuc},[\alpha]}(\mathbf{r}_s) + V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s) \tag{3.1.16}$$

where we have adopted a dependency based notation, $V_s^{[\alpha]} \rightarrow V^{[\alpha]}(\mathbf{D}, \mathbf{r}_s)$. The potential generated by the electrons in the core, $V^{\mathrm{el},[\alpha]}(\mathbf{D}, \mathbf{r}_s)$, is given by

$$V^{\mathrm{el},[\alpha]}(\mathbf{D}, \mathbf{r}_s) = -\sum_{\mu\nu} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) D_{\mu\nu} \tag{3.1.17}$$

The second term in Eq. (3.1.16), $V^{\mathrm{nuc},[\alpha]}(\mathbf{r}_s)$, is the potential generated by the nuclei in the core, given by

$$V^{\mathrm{nuc},[\alpha]}(\mathbf{r}_s) = \sum_{n \in c}^{N_{\mathrm{nuc}}} T^{[\alpha]}(\mathbf{r}_n, \mathbf{r}_s) Z_n \tag{3.1.18}$$

and the third, $V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s)$, is the potential generated by the multipoles in the other fragments, given by

$$V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s) = \sum_{d \neq a}^{N_{\mathrm{frag}}} \sum_{s'' \in d}^{S_d} \sum_{|\gamma|=0}^{K_{s''}} \frac{(-1)^{|\gamma|}}{\gamma!} T^{[\gamma+\alpha]}(\mathbf{r}_s, \mathbf{r}_{s''}) M_{s''}^{[\gamma]} \tag{3.1.19}$$

Inserting the potential expansion of the potential in Eq. (3.1.16), as well as the equations for the fields from each source (Eqs. (3.1.17), (3.1.18) and (3.1.19)), into the right-hand side potential in Eq. (3.1.15), results in

$$
\begin{aligned}
E_{\mathrm{PE}}^{\mathrm{ind}} = & -\frac{1}{2}\sum_{a}^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}V^{[\beta]}(\mathbf{D},\mathbf{r}_{s'})B_{s's}^{[\beta,\alpha]} \\
& \times\left[-\sum_{\mu\nu}t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s)D_{\mu\nu}\right] \\
& -\frac{1}{2}\sum_{a}^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}V^{[\beta]}(\mathbf{D},\mathbf{r}_{s'})B_{s's}^{[\beta,\alpha]} \\
& \times\left[\sum_{n\in c}^{N_{\mathrm{nuc}}}T^{[\alpha]}(\mathbf{r}_n,\mathbf{r}_s)Z_n\right] \\
& -\frac{1}{2}\sum_{a}^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}V^{[\beta]}(\mathbf{D},\mathbf{r}_{s'})B_{s's}^{[\beta,\alpha]} \\
& \times\left[\sum_{d\neq b}^{N_{\mathrm{frag}}}\sum_{s''\in d}^{S_d}\sum_{|\gamma|=0}^{K_{s''}}\frac{(-1)^{|\gamma|}}{\gamma!}T^{[\gamma+\alpha]}(\mathbf{r}_s,\mathbf{r}_{s''})M_{s''}^{[\gamma]}\right] \qquad (3.1.20)
\end{aligned}
$$

which when compared with the Eq. (3.1.4) for the electrostatic energy, shows certain similarities in terms of dependency on the interaction integrals, density matrices, interaction tensors and nuclear charge, in the first two terms, that were not apparent in the unexpanded forms in Eqs. (3.1.8) and (3.1.15). The main difference for these terms, is that while the electrostatic energy involves permanent multipoles, the induction energy involves induced multipoles through $V^{[\beta]}(\mathbf{D},\mathbf{r}_{s'})B_{s's}^{[\beta,\alpha]}$.

The working equations for the induction energy used in the implemented code, however, are attained by a full expansion of the potentials both on the left-hand and the right-hand side of $\mathbf{B}$ in Eq. (3.1.15), in terms of the origin of the potential, giving

$$
\begin{aligned}
E_{\mathrm{PE}}^{\mathrm{ind}} &= -\frac{1}{2}\sum_{a}^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}V^{[\beta]}(\mathbf{D},\mathbf{r}_{s'})B_{s's}^{[\beta,\alpha]}V^{[\alpha]}(\mathbf{D},\mathbf{r}_s) \\
&= -\frac{1}{2}\sum_{a}^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}\Big[V^{\mathrm{el},[\beta]}(\mathbf{D},\mathbf{r}_{s'}) \\
&\quad\times B_{s's}^{[\beta,\alpha]}V^{\mathrm{el},[\alpha]}(\mathbf{D},\mathbf{r}_s) \\
&\quad+2\,(V^{\mathrm{nuc},[\beta]}(\mathbf{r}_{s'})+V^{\mathrm{mul},[\beta]}(\mathbf{r}_{s'}))B_{s's}^{[\beta,\alpha]}V^{\mathrm{el},[\alpha]}(\mathbf{D},\mathbf{r}_s) \\
&\quad+(V^{\mathrm{nuc},[\beta]}(\mathbf{r}_{s'})+V^{\mathrm{mul},[\beta]}(\mathbf{r}_{s'}))B_{s's}^{[\beta,\alpha]}(V^{\mathrm{nuc},[\alpha]}(\mathbf{r}_s)+V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s))\Big] \\
&\overset{\mathrm{Tr}}{=}\frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D}+\mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}}\mathbf{D}+h_{\mathrm{PE}}^{\mathrm{ind}}
\end{aligned}
\tag{3.1.21}
$$

where $\frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D}$ is the interaction energy between the induced dipole moments caused by electrons in the core, and the potential generated by the same electrons. Similarly $\mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}}\mathbf{D}$ is the interaction energy between the induced dipole moments caused by the electrons in the core and the potential generated by the nuclei in the core and the multipoles on the other environmental sites, and $h_{\mathrm{PE}}^{\mathrm{ind}}$ is the interaction energy between the induced dipole moments caused by the nuclei in the core and the other multipoles, and the potential generated by the nuclei and multipoles. The terms $\frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D}$, $\mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}}\mathbf{D}$ and $h_{\mathrm{PE}}^{\mathrm{ind}}$ are second, first and zeroth order in density matrix dependence, respectively. The advantage of formulating the induction energy in this way can be seen by comparing with the energy of the core region, Eq. (3.1.2), where the terms are also formulated based on their dependence on the density matrix (which is also the case for the electrostatic energy in Eq. (3.1.4)). This formulation thus facilitates an analogous handling by OpenRSP[37] of the PE terms as the one that is already implemented for the QM terms. Another note on the implementation is that even though the induced multipoles are here written out as a product of a potential and an interaction matrix, these matrix multiplications are in fact solved in an iterative fashion.

Finally, having now covered both the electrostatic and induced terms in Eq. (3.1.3), the remaining terms are the non-electrostatic repulsion $E_{\mathrm{PE}}^{\mathrm{rep}}$ and dispersion $E_{\mathrm{PE}}^{\mathrm{disp}}$ terms. The non-electrostatic repulsion is included in order to alleviate problems caused by the fact that the wavefunction of the total

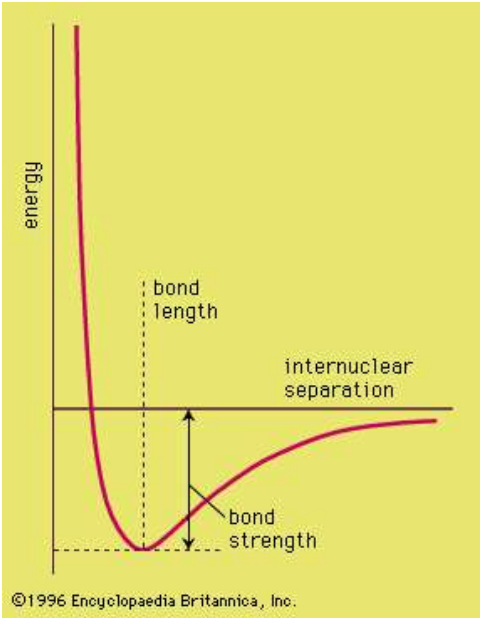


Figure 3.1. Potential energy as a function of intermolecular separation.[81]

system (including both the QM region and the environment) has not been antisymmetrized, and is therefore often called Pauli or exchange repulsion. Dispersion interactions are long-range attractive interactions between two non-permanent bodies,[1] as a result of the fluctuations in the electron cloud around the nuclei. The repulsion interaction and the dispersion interaction can be modelled together as the $1/r^{12}$ and $1/r^6$ terms of the Lennard-Jones (LJ) potential, respectively, in the PE model given by[33]

$$E_{\mathrm{PE}}^{\mathrm{LJ}} = 4 \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a} \sum_{n \in c}^{N_{\mathrm{nuc}}} \varepsilon_{sn} \left[\sigma_{sn}^{12} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \right)^{12} - \sigma_{sn}^{6} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s} \right)^{6} \right] \quad (3.1.22)$$

where the parameters ε_{sn} and σ_{sn} conceptually originate from the shape of an intermolecular potential energy curve, as can be seen in Fig. 3.1 as the bond strength and bond length, respectively. In practice, however, these parameters are typically found for a given atom or molecule from a database, such as for instance the Compound Database http://virtualchemistry.org/.[82,83]

Although for the most part a cheap and accurate enough approximation,

the LJ potential has in certain cases proven insufficient,[33] as it has no explicit dependence on the density matrix. This is in particular the case when looking at electronic excited states that are of a diffuse nature,[84] where core wavefunctions can end up extending unphysically far into the environment due to a too weak repulsion, a phenomenon commonly referred to as electron-spillout. This is because one of the fundamental assumptions for the PE model is that the core and environment are non-overlapping. The core wavefunction is therefore antisymmetrized in isolation,[26] breaking the Pauli antisymmetry principle for cases where there is in fact overlap between the two areas. Use of too diffuse basis functions is therefore problematic with PE. A solution to this problem has been presented through the Polarizable Density Embedding (PDE) model,[34,85] where the innermost layer of the environment is modeled with polarizabilities and frozen densities instead of a multipole expansion. This strategy of keeping the densities in certain parts of the system frozen, and others "thawed", is something that PDE has in common with other QM/QM-methods, such as the Frozen Density Embedding (FDE) approach[86] and subsystem DFT[73]. However, the modelling of polarization of the inner region through polarizabilities, as well as keeping the majority of the environment classical, is computationally advantageous compared to QM/QM-strategies that require iterative freeze-and-thaw procedures, involving multiple calculations of environmental densities.[26] PDE alleviates the problem of electron-spillout by improving the description of the overlap between the QM- and environmental wavefunctions through a repulsion operator, following the strategy of Huzinaga and Cantu.[87] Another strategy for improving the dispersion and repulsion interactions were presented by Curutchet *et al.*[88] through a density-dependent LJ-like potential. However, at the present time, no investigation has been performed on the subject of the effect that these extensions would have in the context of vibrational spectroscopy.

3.2 Vibrational spectroscopy with Polarizable Embedding

Having now established the ground-state formalism of the PE model, it is time to place it in the context of vibrational spectroscopy. In Chapter

2, one of the main points was that vibrational frequencies and intensities depend upon various types of energy derivatives. For instance, harmonic IR intensities depend on the dipole moment gradient, which is an energy derivate bilinear in electronic field strength and in nuclear displacement. When employing the PE model, our energy is the sum of the QM energy of the core region and the interaction energy between the core and the environment. In addition to calculating energy derivatives of the QM energy, it is therefore also necessary to calculate them for the PE energy. Much work has been done in our group regarding how to accurately and efficiently calculate the QM energy derivatives,[29,30,54] and the sole focus of this part will therefore be upon the differentiated PE terms. The reader should, however, note that the framework into which these derivatives are inserted is the same as for the differentiated QM terms, and is referred to Ringholm *et al.*[30] and Steindal *et al.*[25] for details.

Many of the contributions to the PE energy, presented in Section 3.1, contain products of up to four terms that depend on nuclear position. When determining the geometric derivatives of these, the General Leibniz rule,[89] a generalized formula of the product rule[90] is applied. Defining $d^{(n)}$ as a general differentiation operator of n'th order, the n'th-order derivative of the product $u \cdot v$ is given by

$$d^{(n)}(u \cdot v) = \sum_{i=0}^{n} \binom{n}{i} d^{(n-i)}u \cdot d^{(i)}v \tag{3.2.1}$$

To extend to a product of three terms instead of two, replace v with $s \cdot t$, giving

$$d^{(n)}(u \cdot s \cdot t) = \sum_{i=0}^{n} \binom{n}{i} d^{(n-i)}u \cdot \sum_{j=0}^{i} \binom{i}{j} d^{(i-j)}s \cdot d^{(j)}t \tag{3.2.2}$$

and to fourth order by further replacing u with $q \cdot r$

$$d^{(n)}(u \cdot s \cdot t) = \sum_{i=0}^{n} \binom{n}{i} \sum_{k=0}^{n-i} \binom{n-i}{k} d^{(n-i-k)}q \cdot d^{(k)}r \cdot \sum_{j=0}^{i} \binom{i}{j} d^{(i-j)}s \cdot d^{(j)}t \tag{3.2.3}$$

In contrast to geometric perturbation, the only PE terms that are non-vanishing upon electric perturbations are the density matrices. This means that terms independent of the density matrix vanish upon differentiation with respect to electric field strength, while those first-order in the density matrix become

$$d_f^{(n)}(u \cdot D_{\mu\nu}) = u \cdot d_f^{(n)} D_{\mu\nu} \tag{3.2.4}$$

where $d_f^{(n)}$ is an n'th-order electric perturbation and u is any term independent of the density matrix. For terms second order in density matrix dependence, Eq. (3.2.1) is used,

$$d_f^{(n)}(u \cdot D_{\mu\nu} \cdot D_{\rho\sigma}) = u \cdot \sum_{i=0}^{n} \binom{n}{i} d_f^{(n-i)} D_{\mu\nu} \cdot d_f^{(i)} D_{\rho\sigma} \tag{3.2.5}$$

is employed.

3.2.1 Differentiated electrostatic terms

With the tools for handling arbitrary orders of geometric and electric perturbations for product-terms established, it is time to consider the effect on the individual terms in E_{PE}, starting with the electrostatic terms in Eq. (3.1.4),

$$\mathrm{Tr}\, \mathbf{h}_{\mathrm{PE}}^{\mathrm{es}} \mathbf{D} = -\sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{\mu\nu \in c} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) D_{\mu\nu} \tag{3.2.6}$$

which has nuclear dependence in $t_{\mu\nu}$ through the atomic orbitals as well as in $D_{\mu\nu}$. Defining $d_g^{|\gamma|}$ as a $|\gamma|$'th-order geometric differentiation operator, we get the $|\gamma|$'th-order geometric derivative

$$\begin{aligned} d_g^{(|\gamma|)} \mathrm{Tr}\, \mathbf{h}_{\mathrm{PE}}^{\mathrm{es}} \mathbf{D} = &-\sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \\ &\times \sum_{\mu\nu \in c} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} d_g^{(|\gamma|-|\delta|)} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot d_g^{(|\delta|)} D_{\mu\nu} \end{aligned} \tag{3.2.7}$$

Being dependent on the density matrix to first order, a q'th-order electric perturbation gives

$$d_f^{(q)} \mathrm{Tr}\, \mathbf{h}_{\mathrm{PE}}^{\mathrm{es}} \mathbf{D} = - \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{\mu\nu \in c} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot d_f^{(q)} D_{\mu\nu}, \quad (3.2.8)$$

and a mixed geometric and electric perturbation yields

$$\begin{aligned} d_g^{(|\gamma|)} d_f^{(q)} \mathrm{Tr}\, \mathbf{h}_{\mathrm{PE}}^{\mathrm{es}} \mathbf{D} = & - \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \\ & \times \sum_{\mu\nu \in c} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} d_g^{(|\gamma|-|\delta|)} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot d_g^{(|\delta|)} d_f^{(q)} D_{\mu\nu} \end{aligned} \quad (3.2.9)$$

The second electrostatic term, $h_{\mathrm{PE}}^{\mathrm{es}}$ is given by

$$h_{\mathrm{PE}}^{\mathrm{es}} = \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{n \in c}^{N_{\mathrm{nuc}}} T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n) Z_n \quad (3.2.10)$$

and contains no dependence on the density matrix. It will therefore vanish upon electric perturbation. Additionally, the only term depending on nuclear coordinates is $T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n)$, which is given by Eq. (3.1.7). Thus,

$$d_g^{|\gamma|} h_{\mathrm{PE}}^{\mathrm{es}} = \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{n \in c}^{N_{\mathrm{nuc}}} d_g^{|\gamma|} T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n) Z_n \quad (3.2.11)$$

where $T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n)$ depends only on a single nucleus, and will therefore vanish when differentiated with regards to the coordinates of two different nuclei, giving

$$d_g^{|\gamma|} T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n) = \partial_{\mathbf{r}_n}^{\gamma} T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n) = \partial_{\mathbf{r}_n}^{\gamma} \partial_{\mathbf{r}_n}^{\alpha} \frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} = T^{[\alpha+\gamma]}(\mathbf{r}_s, \mathbf{r}_n) \quad (3.2.12)$$

on the condition that $d_g^{|\gamma|}$ corresponds to a differentiation with respect only to $\mathbf{r}_n$, and zero otherwise.

3.2.2 Differentiated induction terms

Moving on to the induction energy, we see that it is a product of a higher number of terms that are non-vanishing upon differentiation than the electrostatic contribution. This causes a non-trivial complication of the differentiated expressions, because a higher order of the product-rule has to be employed. Starting with each of the terms in Eq. (3.1.21), and inserting Eqs. (3.1.17) and (3.1.18) (it is unnecessary to insert (3.1.19) as V^{mul} depends neither on nuclear positions nor on the density matrix), we get the fully expanded equations

$$\begin{aligned}\operatorname{Tr} \frac{1}{2}\mathbf{G}_{\text{PE}}^{\text{ind}}(\mathbf{D})\mathbf{D} = &-\frac{1}{2}\sum_{a}^{N_{\text{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\text{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}\\ &\times\left(\sum_{\mu\nu}t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'})D_{\mu\nu}\right)B_{s's}^{[\beta,\alpha]}\sum_{\rho\sigma}t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s)D_{\rho\sigma},\end{aligned} \tag{3.2.13}$$

$$\begin{aligned}\operatorname{Tr} \mathbf{h}_{\text{PE}}^{\text{ind}}\mathbf{D} = &\sum_{a}^{N_{\text{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\text{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}\\ &\times\left(\sum_{n\in c}^{N_{\text{nuc}}}T^{[\beta]}(\mathbf{r}_n,\mathbf{r}_{s'})Z_n+V^{\text{mul},[\beta]}(\mathbf{r}_{s'})\right)\\ &\times B_{s's}^{[\beta,\alpha]}\sum_{\mu\nu}t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s)D_{\mu\nu}\end{aligned} \tag{3.2.14}$$

and

$$\begin{aligned}h_{\text{PE}}^{\text{ind}} = &-\frac{1}{2}\sum_{a}^{N_{\text{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\text{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}\\ &\times\left(\sum_{n\in c}^{N_{\text{nuc}}}T^{[\beta]}(\mathbf{r}_n,\mathbf{r}_{s'})Z_n+V^{\text{mul},[\beta]}(\mathbf{r}_{s'})\right)\\ &\times B_{s's}^{[\beta,\alpha]}\left(\sum_{m\in c}^{N_{\text{nuc}}}T^{[\alpha]}(\mathbf{r}_m,\mathbf{r}_s)Z_m+V^{\text{mul},[\alpha]}(\mathbf{r}_s)\right)\end{aligned} \tag{3.2.15}$$

Using Eq. (3.2.3) for arbitrary-order derivatives of a product of four terms, the $|\gamma|$'th-order geometric derivative of $\mathbf{Tr}\frac{1}{2}\mathbf{G}_{\text{PE}}^{\text{ind}}(\mathbf{D})\mathbf{D}$ is

$$
\begin{aligned}
d_g^{(|\gamma|)}\mathrm{Tr}\,\frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D} = -\frac{1}{2}\sum_{a}^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}\sum_{|\delta|=0}^{|\gamma|}\binom{|\gamma|}{|\delta|} \\
\times\sum_{|\varepsilon|=0}^{|\gamma|-|\delta|}\binom{|\gamma|-|\delta|}{|\varepsilon|} \\
\times\left(\sum_{\mu\nu} d_g^{(|\gamma|-|\delta|-|\varepsilon|)}t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'})\cdot d_g^{(|\varepsilon|)}D_{\mu\nu}\right) \\
\times B_{s's}^{[\beta,\alpha]}\sum_{|\zeta|=0}^{|\delta|}\binom{|\delta|}{|\zeta|}\sum_{\rho\sigma} d_g^{(|\delta|-|\zeta|)}t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s)\cdot d_g^{(|\zeta|)}D_{\rho\sigma}
\end{aligned}
\tag{3.2.16}
$$

As $\mathrm{Tr}\,\frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D}$ is second order in density dependence, we use Eq. (3.2.4) in order to find electric perturbations of q'th order to be

$$
\begin{aligned}
d_f^{(q)}\mathrm{Tr}\,\frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D} = -\frac{1}{2}\sum_{a}^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!} \\
\times\sum_{i=0}^{q}\binom{q}{i}\left(\sum_{\mu\nu} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'})\cdot d_f^{(q-i)}D_{\mu\nu}\right) \\
\times B_{s's}^{[\beta,\alpha]}\sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s)\cdot d_f^{(i)}D_{\rho\sigma}
\end{aligned}
\tag{3.2.17}
$$

Mixed geometric and electronic perturbations lead to

$$
\begin{aligned}
d_g^{(|\gamma|)} d_f^{(q)} \mathrm{Tr}\, \frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D} = &-\frac{1}{2}\sum_{a}^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!} \\
&\times \sum_{|\delta|=0}^{|\gamma|}\binom{|\gamma|}{|\delta|}\sum_{|\varepsilon|=0}^{|\gamma|-|\delta|}\binom{|\gamma|-|\delta|}{|\varepsilon|}\sum_{i=0}^{q}\binom{q}{i} \\
&\times \left(\sum_{\mu\nu} d_g^{(|\gamma|-|\delta|-|\varepsilon|)} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) d_g^{(|\varepsilon|)} d_f^{(q-i)} D_{\mu\nu}\right) \\
&\times B_{s's}^{[\beta,\alpha]} \sum_{|\zeta|=0}^{|\delta|}\binom{|\delta|}{|\zeta|} \\
&\times \sum_{\rho\sigma} d_g^{(|\delta|-|\zeta|)} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s)\cdot d_g^{(|\zeta|)} d_f^{(i)} D_{\rho\sigma} \qquad (3.2.18)
\end{aligned}
$$

The second term in Eq. (3.1.21), $\mathrm{Tr}\,\mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}}\mathbf{D}$, is a product of three terms, and Eq. (3.2.2) is therefore used to find its $|\gamma|$'th-order geometric derivative

$$
\begin{aligned}
d_g^{(|\gamma|)} \mathrm{Tr}\, \mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}}\mathbf{D} = &\sum_{a}^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_{b}^{N_{\mathrm{frag}}}\sum_{s'\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_{s'}}\frac{1}{\alpha!\beta!}\sum_{|\delta|=0}^{|\gamma|}\binom{|\gamma|}{|\delta|} \\
&\times \left(\sum_{n\in c} d_g^{(|\gamma|-|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n \right. \\
&\left. + \delta_{0(|\gamma|-|\delta|)} V^{\mathrm{mul},[\beta]}(\mathbf{r}_{s'})\right) \\
&\times B_{s's}^{[\beta,\alpha]} \sum_{|\zeta|=0}^{|\delta|}\binom{|\delta|}{|\zeta|}\sum_{\rho\sigma} d_g^{(|\delta|-|\zeta|)} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_g^{(|\zeta|)} D_{\rho\sigma}. \qquad (3.2.19)
\end{aligned}
$$

where $\partial_{\mathbf{r}_i}^{\beta} T^{[\alpha]}(\mathbf{r}_i, \mathbf{r}_j) = (-1)^{|\beta|} T^{[\alpha+\beta]}(\mathbf{r}_i, \mathbf{r}_j)$[78] means that

$$
d_g^{(|\gamma|)} T^{[\alpha]}(\mathbf{r}_n, \mathbf{r}_{s'}) = (-1)^{|\gamma|} T^{[\alpha+\gamma]}(\mathbf{r}_n, \mathbf{r}_{s'}) \qquad (3.2.20)
$$

on the condition that $d_g^{(|\gamma|)}$ corresponds to a differentiation with respect to $\mathbf{r}_n$, and zero otherwise. The Kronecker delta $\delta_{0(|\gamma|-|\delta|)}$ denotes that V^{mul} vanishes if $|\gamma|-|\delta|$ is different from zero, in other words if it is differentiated with respect to nuclear position. Since $\mathrm{Tr}\,\mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}}\mathbf{D}$ is first order in dependence on the density matrix, q'th-order electric perturbations will simply give

$$d_f^{(q)} \mathrm{Tr}\, \mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}} \mathbf{D} = \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{s'\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_{s'}} \frac{1}{\alpha!\beta!} \times \left(\sum_{n\in c}^{N_{\mathrm{nuc}}} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n + V^{\mathrm{mul},[\beta]}(\mathbf{r}_{s'}) \right) \times B_{s's}^{[\beta,\alpha]} \sum_{\mu\nu} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot d_f^{(q)} D_{\mu\nu} \tag{3.2.21}$$

and a mix between geometric and electronic perturbations gives

$$d_g^{(|\gamma|)} d_f^{(q)} \mathrm{Tr}\, \mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}} \mathbf{D} = \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{s'\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_{s'}} \frac{1}{\alpha!\beta!} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} \times \left(\sum_{n\in c} d_g^{(|\gamma|-|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n + \delta_{0(|\gamma|-|\delta|)} V^{\mathrm{mul},[\beta]}(\mathbf{r}_{s'}) \right) \times B_{s's}^{[\beta,\alpha]} \sum_{|\zeta|=0}^{|\delta|} \binom{|\delta|}{|\zeta|} \sum_{\rho\sigma} d_g^{(|\delta|-|\zeta|)} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) \cdot d_g^{(|\zeta|)} d_f^{(q)} D_{\rho\sigma} \tag{3.2.22}$$

For the last term in Eq. (3.1.21), $h_{\mathrm{PE}}^{\mathrm{ind}}$ as given in Eq. (3.2.15), is a product of only two terms. Therefore, with Eq. (3.2.1), the $|\gamma|$'th-order geometric derivative is given by

$$d_g^{(|\gamma|)} h_{\mathrm{PE}}^{\mathrm{ind}} = -\frac{1}{2} \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{s'\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_{s'}} \frac{1}{\alpha!\beta!} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} \times \left(\sum_{n\in c} d_g^{(|\gamma|-|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n + \delta_{0(|\gamma|-|\delta|)} V^{\mathrm{mul},[\beta]}(\mathbf{r}_{s'}) \right) \times B_{s's}^{[\beta,\alpha]} \left(\sum_{m\in c} d_g^{(|\delta|)} T^{[\alpha]}(\mathbf{r}_m, \mathbf{r}_s) Z_m + \delta_{0|\delta|} V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s) \right) \tag{3.2.23}$$

As $h_{\mathrm{PE}}^{\mathrm{ind}}$ contains no dependence on the density matrix, it will vanish upon any electric perturbation.

3.2.3 Differentiated Lennard-Jones potential

The very last contribution to the PE energy is the Lennard-Jones potential given in Eq. (3.1.22), modelling the dispersion and repulsion energies by

$$E_{\mathrm{PE}}^{\mathrm{LJ}} = 4 \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a} \sum_{n\in c}^{N_{\mathrm{nuc}}} \varepsilon_{sn} \left[\sigma_{sn}^{12} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \right)^{12} - \sigma_{sn}^{6} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \right)^{6} \right] \tag{3.2.24}$$

Using Eq. (3.1.7), this can be reformulated to

$$E_{\mathrm{PE}}^{\mathrm{LJ}} = 4 \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a} \sum_{n\in c}^{N_{\mathrm{nuc}}} \varepsilon_{sn} \left[\sigma_{sn}^{12} \left(T^{[(0,0,0)]}(\mathbf{r}_s, \mathbf{r}_n) \right)^{12} - \sigma_{sn}^{6} \left(T^{[(0,0,0)]}(\mathbf{r}_s, \mathbf{r}_n) \right)^{6} \right] \tag{3.2.25}$$

It is here necessary to employ a generalized version of the chain rule[90], namely the multivariate version of Faà di Bruno's formula.[91] It states that if you have a function $f(y)$ where $y = g(x_1, x_2, \ldots, x_n)$, the n'th order derivative of f is

$$\frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} f(y) = \sum_{\pi\in\Pi} f^{(|\pi|)}(y) \prod_{B\in\pi} \frac{\partial^{|B|} y}{\prod_{j\in B} \partial x_j} \tag{3.2.26}$$

where the variables $x_1, x_2, \ldots, x_n$ do not all have to be distinct. In the case of geometric derivatives, the differentiation will be with regards to the x, y or z component of the position of a specific nucleus, and each of $x_1, x_2, \ldots, x_n$ therefore has to be one of these. The summation index π here runs over all Π possible partitions of the set $1, 2, \ldots, n$, which is all the ways you can group these values into non-empty subsets, where each value is included exactly once.[92] The summation index B runs over each block within partition π, and j runs over all elements in block B. For a given order n, the number of partitions is given by

$$A_n = \sum_{k=0}^{n-1} \binom{n-1}{k} A_k \tag{3.2.27}$$

To make this clearer, take the example of a third-order derivative. In this case, all the possible partitions of the set $1, 2, 3$ is

$$\Pi = (\{1\},\{2\},\{3\}),(\{1,2\},\{3\}),(\{1,3\},\{2\}),(\{1\},\{2,3\}),(\{1,2,3\}) \tag{3.2.28}$$

π runs over each of these five partitions. For the partition $\pi = (\{1,2\},\{3\})$, $|\pi| = 2$ and B runs over each of these two blocs. For the block $B = \{1,2\}$, $|B| = 2$ and j runs over each of these two elements.

If we now define the following functions

$$g_{LJ} = T^{[(0,0,0)]}(\mathbf{r}_s, \mathbf{r}_n) = \frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \tag{3.2.29}$$

$$f_{LJ}(g_{LJ}) = \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|}\right)^{12} = (g_{LJ})^{12} \tag{3.2.30}$$

$$h_{LJ}(g_{LJ}) = \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|}\right)^{6} = (g_{LJ})^{6} \tag{3.2.31}$$

the LJ potential can be written as

$$E_{\mathrm{PE}}^{\mathrm{LJ}} = 4 \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{n\in c}^{N_{\mathrm{nuc}}} \varepsilon_{sn} \left[\sigma_{sn}^{12} f_{LJ}(g_{LJ}) - \sigma_{sn}^{6} h_{LJ}(g_{LJ})\right] \tag{3.2.32}$$

If we then employ Faà di Bruno's formula, Eq. (3.2.26), and Eq. (3.2.12), we can formulate the $x_1 x_2 \cdots x_{|\gamma|}$-component of the $|\gamma|$'th order derivative with respect to nucleus n of the LJ potential is

$$\left(d_g^{|\gamma|} E_{\mathrm{PE}}^{\mathrm{LJ}}\right)_{x_1 x_2 \cdots x_{|\gamma|}, n} = 4 \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \varepsilon_{sn} \left[\sigma_{sn}^{12} \sum_{\pi\in\Pi} f_{LJ}^{(|\pi|)} \prod_{B\in\pi} \frac{\partial^{|B|} g_{LJ}}{\prod_{j\in B} \partial x_j} - \sigma_{sn}^{6} \sum_{\pi\in\Pi} h_{LJ}^{(|\pi|)} \prod_{B\in\pi} \frac{\partial^{|B|} g_{LJ}}{\prod_{j\in B} \partial x_j}\right] \tag{3.2.33}$$

where

$$f_{LJ}^{(k)} = \frac{\partial^k}{\partial g_{LJ}^k} f_{LJ}(g_{LJ}) = \begin{cases} \frac{12!}{(12-k)!} g_{LJ}^{12-k}, & \text{if } k < 13 \\ 0, & \text{otherwise} \end{cases} \tag{3.2.34}$$

$$h_{LJ}^{(k)} = \frac{\partial^k}{\partial g_{LJ}^k} h_{LJ}(g_{LJ}) = \begin{cases} \frac{6!}{(6-k)!} g_{LJ}^{6-k}, & \text{if } k < 7 \\ 0, & \text{otherwise} \end{cases} \tag{3.2.35}$$

and

$$\frac{\partial^k}{\partial x_1 \partial x_2 \cdots \partial x_k} g_{LJ}(x_1, x_2, \ldots, x_k) = T^{[(k_x,k_y,k_z)]}(\mathbf{r}_s, \mathbf{r}_n) \tag{3.2.36}$$

where k_x is the number of the variables that belong to Cartesian component x of nucleus n, and thus, $k_x + k_y + k_z = k$. The LJ potential will vanish upon any perturbation of more than one nucleus at a time.

Finally, since $E_{\text{PE}}^{\text{LJ}}$ has no dependence on the density matrix, all terms will vanish upon electric perturbation.

3.2.4 Application example: harmonic IR spectroscopy with PE

In order to better understand the significance of the differentiated terms derived to arbitrary order in the previous section, the current section is an application example related to vibrational spectroscopy where the terms necessary in order to produce an IR spectrum with PE are derived. The QM terms involved will not be covered. As described in Section 2.1, the harmonic vibrational frequencies are found through an eigenvalue analysis of the molecular Hessian, and Section 2.3 shows how IR intensity-related properties depend upon the dipole gradient. To calculate an IR spectrum, it is therefore necessary to calculate the second-order geometric derivative, as well as the first-order electric in combination with first-order geometric derivative of the PE energy. This produces the following terms.

3.2.4.1 Second-order geometric derivative: molecular Hessian

Throughout this section, $|\delta| = 1$ and $|\gamma| = 2$.

$$
\begin{aligned}
d_g^{(2)} E_{\mathrm{PE}}^{\mathrm{es}} = &- \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{\mu\nu \in c} \\
&\times \left(d_g^{(2)} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot D_{\mu\nu} + 2 d_g^{(1)} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot d_g^{(1)} D_{\mu\nu} + t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) d_g^{(2)} D_{\mu\nu} \right) \\
&+ \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{n \in c}^{N_{\mathrm{nuc}}} d_g^{(|\gamma|)} T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n) Z_n
\end{aligned} \tag{3.2.37}
$$

$$
\begin{aligned}
d_g^{(2)} E_{\mathrm{PE}}^{\mathrm{ind}} = &- \sum_{a}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{s' \in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_{s'}} \frac{1}{\alpha! \beta!} \\
&\times \Bigg[\left(\sum_{\mu\nu} d_g^{(2)} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) \cdot D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) D_{\rho\sigma} \\
&+ 2 \left(\sum_{\mu\nu} d_g^{(1)} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) d_g^{(1)} D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) D_{\rho\sigma} \\
&+ \left(\sum_{\mu\nu} d_g^{(1)} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} d_g^{(1)} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) D_{\rho\sigma} \\
&+ 2 \left(\sum_{\mu\nu} d_g^{(1)} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_g^{(1)} D_{\rho\sigma} \\
&+ \left(\sum_{\mu\nu} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) d_g^{(2)} D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) D_{\rho\sigma} \\
&+ \left(\sum_{\mu\nu} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) d_g^{(1)} D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_g^{(1)} D_{\rho\sigma} \Bigg] \\
&\text{continued...}
\end{aligned} \tag{3.2.38}
$$

continued... =

$$
\begin{aligned}
&+ \sum_{a}^{N_{\text{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\text{frag}}} \sum_{s'\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_{s'}} \frac{1}{\alpha!\beta!} \\
&\times \Bigg[\left(\sum_{n\in c} d^{(|\gamma|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n \right) B^{[\beta,\alpha]}_{s's} \sum_{\rho\sigma} t^{[\alpha]}_{\rho\sigma}(\mathbf{r}_s) D_{\rho\sigma} \\
&+ 2 \left(\sum_{n\in c} d_g^{(|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n \right) B^{[\beta,\alpha]}_{s's} \sum_{\rho\sigma} d_g^{(1)} t^{[\alpha]}_{\rho\sigma}(\mathbf{r}_s) D_{\rho\sigma} \\
&+ 2 \left(\sum_{n\in c} d_g^{(|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n \right) B^{[\beta,\alpha]}_{s's} \sum_{\rho\sigma} t^{[\alpha]}_{\rho\sigma}(\mathbf{r}_s) d_g^{(1)} D_{\rho\sigma} \\
&+ \left(\sum_{n\in c} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n + V^{\text{mul},[\beta]}(\mathbf{r}_{s'}) \right) \\
&\times B^{[\beta,\alpha]}_{s's} \sum_{\rho\sigma} d_g^{(2)} t^{[\alpha]}_{\rho\sigma}(\mathbf{r}_s) \cdot D_{\rho\sigma} \\
&+ 2 \left(\sum_{n\in c} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n + V^{\text{mul},[\beta]}(\mathbf{r}_{s'}) \right) \\
&\times B^{[\beta,\alpha]}_{s's} \sum_{\rho\sigma} d_g^{(1)} t^{[\alpha]}_{\rho\sigma}(\mathbf{r}_s) \cdot d_g^{(1)} D_{\rho\sigma} \\
&+ \left(\sum_{n\in c} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n + V^{\text{mul},[\beta]}(\mathbf{r}_{s'}) \right) \\
&\times B^{[\beta,\alpha]}_{s's} \sum_{\rho\sigma} t^{[\alpha]}_{\rho\sigma}(\mathbf{r}_s) d_g^{(2)} D_{\rho\sigma} \Bigg]
\end{aligned}
$$

continued...

$$
\begin{aligned}
\text{continued...} = - & \sum_{a}^{N_{\text{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\text{frag}}} \sum_{s'\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_{s'}} \frac{1}{\alpha!\beta!} \\
& \times \left[\left(\sum_{n\in c} d_g^{(|\gamma|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n \right) \right. \\
& \times B_{s's}^{[\beta,\alpha]} \left(\sum_{m\in c} T^{[\alpha]}(\mathbf{r}_m, \mathbf{r}_s) Z_m + V^{\text{mul},[\alpha]}(\mathbf{r}_s) \right) \Bigg) \\
& + \left(\sum_{n\in c} d_g^{(|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n \right) \\
& \times \left. B_{s's}^{[\beta,\alpha]} \left(\sum_{m\in c} d_g^{(|\delta|)} T^{[\alpha]}(\mathbf{r}_m, \mathbf{r}_s) Z_m \right) \right]
\end{aligned}
$$

using the fact that the summation runs over all sites within all fragments, which means that the equation is symmetric with respect to α and β.

$$
\begin{aligned}
\left(d_g^{(2)} E_{\text{PE}}^{\text{LJ}} \right)_{xy,n} = & 4 \sum_{a}^{N_{\text{frag}}} \sum_{s\in a}^{S_a} \varepsilon_{sn} \\
& \times \left[132\sigma_{sn}^{12} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \right)^{10} T_x^{[\delta]}(\mathbf{r}_s, \mathbf{r}_n) T_y^{[\delta]}(\mathbf{r}_s, \mathbf{r}_n) \right. \\
& + 12\sigma_{sn}^{12} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \right)^{11} T_{xy}^{[\gamma]}(\mathbf{r}_s, \mathbf{r}_n) \\
& - 30\sigma_{sn}^{6} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \right)^{4} T_x^{[\delta]}(\mathbf{r}_s, \mathbf{r}_n) T_y^{[\delta]}(\mathbf{r}_s, \mathbf{r}_n) \\
& \left. - 6\sigma_{sn}^{6} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \right)^{5} T_{xy}^{[\gamma]}(\mathbf{r}_s, \mathbf{r}_n) \right]
\end{aligned} \tag{3.2.39}
$$

3.2.4.2 First-order geometric and first-order electric derivative: electronic dipole gradient

In this section, $|\delta| = 1$.

$$d_g^{(1)} d_f^{(1)} E_{\mathrm{PE}}^{\mathrm{es}} = - \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \times \sum_{\mu\nu \in c} \left(d_g^{(1)} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot d_f^{(1)} D_{\mu\nu} + t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot d_g^{(1)} d_f^{(1)} D_{\mu\nu} \right) \tag{3.2.40}$$

$$\begin{aligned}
d_g^{(1)} d_f^{(1)} E_{\mathrm{PE}}^{\mathrm{ind}} = & - \sum_{a}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{s' \in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_{s'}} \frac{1}{\alpha! \beta!} \\
& \times \left[\left(\sum_{\mu\nu} d_g^{(1)} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) \cdot d_f^{(1)} D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) D_{\rho\sigma} \right. \\
& + \left(\sum_{\mu\nu} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) d_g^{(1)} d_f^{(1)} D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) D_{\rho\sigma} \\
& + \left(\sum_{\mu\nu} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) d_f^{(1)} D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} d_g^{(1)} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) \cdot D_{\rho\sigma} \\
& \left. + \left(\sum_{\mu\nu} t_{\mu\nu}^{[\beta]}(\mathbf{r}_{s'}) d_f^{(1)} D_{\mu\nu} \right) B_{s's}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_g^{(1)} D_{\rho\sigma} \right] \\
& + \sum_{a}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{s' \in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_{s'}} \frac{1}{\alpha! \beta!} \\
& \times \left[\left(\sum_{n \in c}^{N_{\mathrm{nuc}}} d_g^{(|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n \right) B_{s's}^{[\beta,\alpha]} \sum_{\mu\nu} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot d_f^{(1)} D_{\mu\nu} \right. \\
& + \left(\sum_{n \in c}^{N_{\mathrm{nuc}}} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n + V^{\mathrm{mul},[\beta]}(\mathbf{r}_{s'}) \right) \\
& \times B_{s's}^{[\beta,\alpha]} \sum_{\mu\nu} d_g^{(1)} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) \cdot d_f^{(1)} D_{\mu\nu} \\
& + \left(\sum_{n \in c}^{N_{\mathrm{nuc}}} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_{s'}) Z_n + V^{\mathrm{mul},[\beta]}(\mathbf{r}_{s'}) \right) \\
& \left. \times B_{s's}^{[\beta,\alpha]} \sum_{\mu\nu} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) d_g^{(1)} d_f^{(1)} D_{\mu\nu} \right]
\end{aligned} \tag{3.2.41}$$

All PE terms involved in the Hessian and dipole gradient have now been expressed, and both harmonic vibrational frequencies and infrared

intensities can therefore be calculated. How this is done in practice and how to calculate a vibrational spectrum with PE will be covered in Chapter 4.

Chapter 4

From start to finish in a vibrational PE calculation: Software and practical considerations

In previous chapters, the theory behind vibrational spectroscopies and the Polarizable Embedding model have been covered in detail. This chapter covers how such calculations are actually performed within our framework of software applications, organized in the order of the workflow displayed in Fig. 4.1. First aspects connected to the preparation of structures (steps 1 - 4)) are described; second, the main calculations and the programs involved (steps 5 and 6); and third, the post processing of these results (steps 7 and END). Throughout this section, the focus will be on solvated (or other weakly bound) systems as this is the topic of all the papers included in this work, but note that the overall procedure will be applicable also for biomolecules with a few minor alterations.

4.1 Preparation of molecular structures

Before getting into the real number-crunching parts of a PE calculation, there are a number of pieces that need to be put in place. The first task is to create geometries for the isolated solute molecule. This can be done with

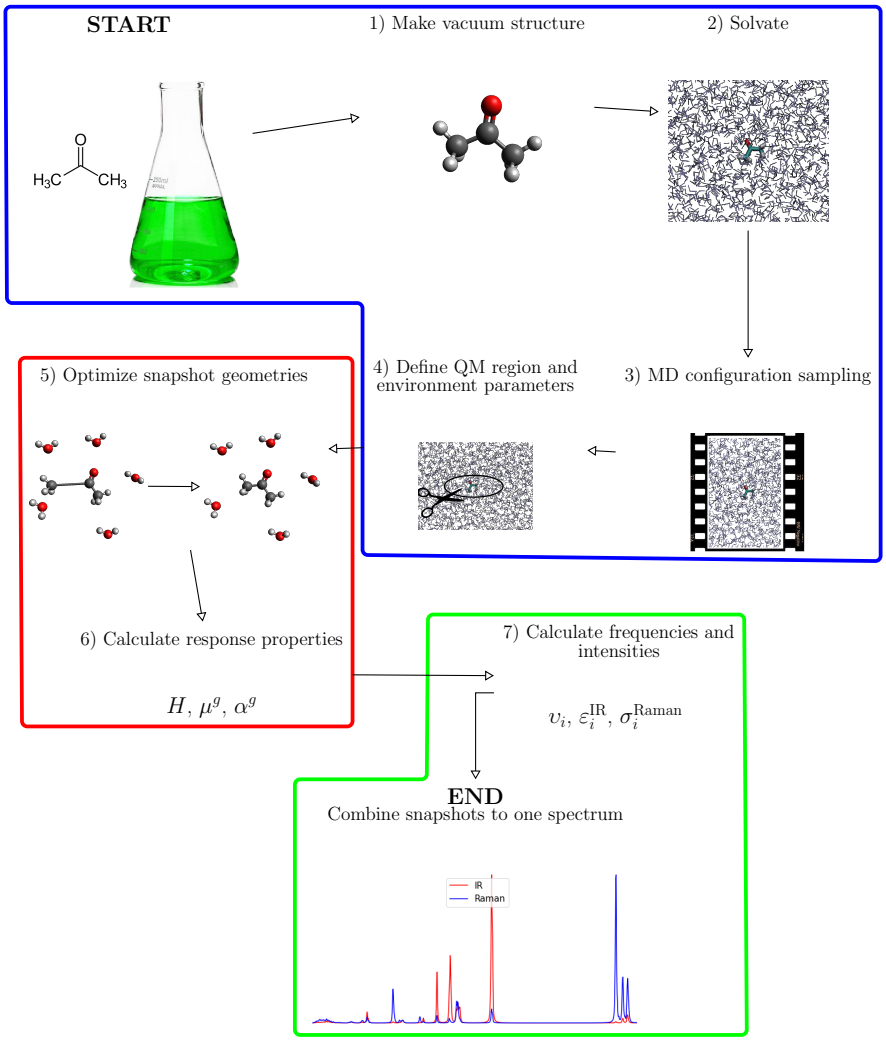


Figure 4.1. Workflow from concept idea to finished spectrum in a vibrational spectroscopic calculation for a solvated system with PE. All images are purely illustrational. The steps inside the blue box are described in Section 4.1, those inside the red box in Section 4.2 and those inside the green box in Section 4.3.

a number of visualization programs, among them Avogadro.[93,94] The GROMACS[95–97] program is then used to add solvent molecules and equilibrate the system.

At this point, a set of coordinates have been created with one specific configuration of the solvent molecules and the solute. There are, however, a staggering amount of distinct and possible configurations. As the solvent molecules interact with the solute through intermolecular interactions that often depend on the distance between particles (see for instance Eqs. (3.1.5), (3.1.7) and (3.1.22)), the choice of configuration will have a direct impact on the results of the calculations. In order to statistically sample the various possible solvent configurations, a Molecular Dynamics (MD)[17] simulation is run. The interactions between the particles are classically calculated and these forces cause the molecules to alter shape and position as a function of time. This produces a trajectory over time from which snapshots can be extracted, giving a set of slightly different configurations that together mimic an ever-changing realistic bulk solution. Any further description of MD is outside the scope of this work, and the interested reader is referred to Leach's textbook[17] for further details.

Having now obtained a set of configurations, the next step is to define the QM part and the environment, and provide the information in a format readable by the programs we will use for our property calculations. This is performed by PyFraME,[23] a program specifically developed for this task. The solute molecule is identified and defined as the QM region. It is also possible to include the innermost layer of solvent molecules in the QM region as this can improve the accuracy of the calculations.[25] For the part defined as the environment, PyFraME then defines each solvent molecule as a fragment, and each atom as a site. Multipoles and polarizabilities are placed on each atom, either by explicitly calculating these through a calculation on each fragment using the LoProp partitioning approach[24], or by using a set of pre-calculated averaged solvent embedding potentials[77] (SEP) parameters. While the first is accurate, the second is cheap. A compromise between accuracy and computational cost, where the innermost environment layers have explicitly calculated parameters, while the outer layers use SEP parameters, can therefore in many cases be a good option.

In order to decide how many molecules to include in the QM region, we look for convergence in the results while increasing the size of the QM region.

Similar studies should be made before deciding at what point to switch from explicit to SEP parameters and how large the total system must be on order to simulate a bulk liquid. For a thorough overview of PyFraME's functionality, the reader is referred to Steinmann *et al.*[32] This paper summarizes some of these choices just mentioned for the implementation in Dalton[45], but these considerations hold also for the LSDalton/FraME implementation (to be described in the next section). A choice should also be made regarding the number of snapshots to include in order to sample the natural variation within a liquid. In order to do this, we look for a repetition of the results after a certain number of snapshots, due to recurring solvent-solute conformations. In **Paper I**, we performed an analysis on the basis of 250 snapshots, but the results from this investigation are somewhat inconclusive, and it can be argued that the sample size was not of sufficient size. In this paper, we also investigated the deviation of the results based on the environmental size compared to a reference size of a 16Å radius from the center of mass of our QM region, finding relatively small errors at a radius of 12Å. It could, however, be pointed out that a larger reference size might have been more appropriate. We therefore cannot at present provide any rule of thumb as to the appropriate number of snapshots or the size of the environment that should be included, but aim to in the future perform a more thorough investigation in order to get a better understanding of the effects of these, and other, parameters for vibrational spectroscopies. Additionally, these numbers may vary from system to system, and we therefore encourage the reader to perform independent analyses for their particular case.

4.2 Geometry optimization and property calculation

As explained in Chapter 2, the theory employed in this work is a good approximation only for small displacements from the equilibrium geometry of the QM region. This is in particular true for the harmonic approximation, but as the higher-order terms added perturbatively to include anharmonic effects are also evaluated at the equilibrium geometry, we will also encounter problems when using second-order vibrational perturbation theories if our

geometry is too far away from equilibrium. Both with and without anharmonic corrections, a large displacement will also result in the appearance of forces on the nuclei acting to restore the equilibrium geometry of the molecule, causing a contamination of all terms in the Taylor expansion with contributions that are not related to vibrational motion. Any molecular conformation evaluated at a non-zero temperature, which is the case when they are obtained from MD, will inevitably deviate from the equilibrium geometry. In a study of how different solvent models influence Raman Optical Activity (ROA) spectra, Hopmann *et al.*[14] show that when neglecting to optimize the geometry of the QM region in a structure obtained from an MD simulation, the resulting spectrum is reduced to numerical noise. An alternative could be to employ a QM/MM-MD model, where the intramolecular forces within the QM region are better represented. However, as this is a costly strategy, we employ a cheaper method, as suggested by Kjellgren *et al.*,[98], at some expense in terms of the accuracy. This method involves performing a geometry optimization solely on the QM region, keeping the environment frozen. In our work, this is performed by a combination of the PE library FraME[99] and the quantum chemistry package LSDalton.[45,47] An overview of how these are connected to each other can be seen in Fig. 4.2

Also shown in Fig. 4.2 is how, after the geometry optimization, a property calculation is performed by a combination of LSDalton, FraME and the general, recursive and open-ended response code OpenRSP.[30,37] The individual perturbed terms are calculated by FraME and LSDalton, while OpenRSP monitors which ones need to be calculated and combines them to the total output which is printed from LSDalton. OpenRSP is recursive, in a way that makes it completely open-ended when it comes to order of differentiation. It also keeps the amount of calculated terms to a minimum by applying a high degree of non-redundancy and utilizing the (k, n)-rule[30] which significantly reduces the number of terms needed. Another advantage of this approach is that all energy derivatives are calculated analytically, avoiding introduction of errors by numerical differentiation, which can be expected to get progressively more severe as the order of differentiation is increased.

An illustration of the non-redundancy provided by OpenRSP can be shown with the example of a second-order geometric perturbation. For

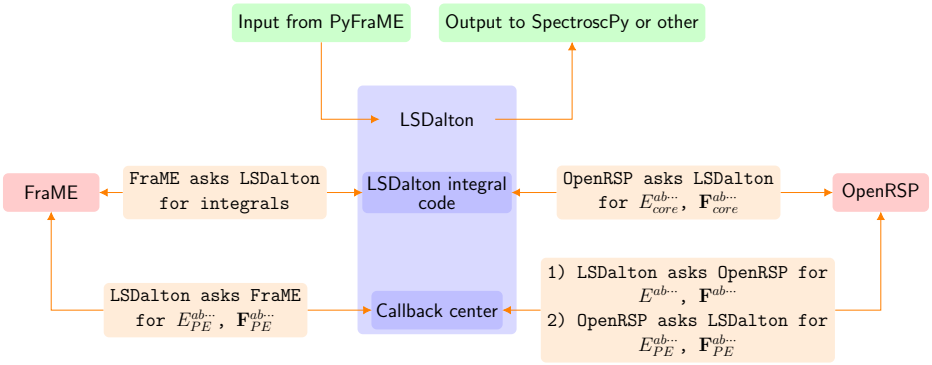


Figure 4.2. Overview of how the software used in PE property calculations are connected. Input generated by PyFraME is sent in to LSDalton which delegates work depending on what type of calculation is requested. If it is 1) a geometry optimization, LSDalton performs the optimization, but asks FraME for PE contributions to the energy and Fock matrix. FraME uses the high-order differentiated integrals[100] (HODI) module within LSDalton as well as internal integral routines in LSDalton to calculate the integrals in the electronic electrostatic term and in the electronic potential. The PE contributions are then returned to LSDalton which finishes the geometry optimization and writes output geometry. If the requested calculation is 2) a property calculation, LSDalton asks OpenRSP for the needed energy-derivatives. OpenRSP keeps track of all the terms generated and asks LSDalton to calculate the individual perturbed terms for the QM contributions, and through the callback function in LSDalton requests individual perturbed terms for the PE contributions. Both QM- and PE contributions are returned to OpenRSP where they are collected and sent back to LSDalton. A tensor file containing the energy derivative is written, to be postprocessed by SpectroscPy. Note also that, although not illustrated here, perturbed density matrices are found by a procedure, managed by OpenRSP, which includes solving the so-called response equations[101] with LSDalton functionality.

a molecule with N atoms, a first-order geometric perturbation generates $3N$ derivatives, each corresponding to moving one of the N atoms in one of the 3 Cartesian directions. A second-order geometric perturbation therefore generates $(3N)^2$ derivatives, but the non-redundancy lies in recognizing that it is not necessary to calculate all of these. For instance, moving an atom first in the x-direction and then in the y-direction gives an identical result as first moving it in the y-direction and then in the x-direction. This symmetry allows OpenRSP to calculate a much smaller number of derivatives, given by

$$\binom{n+m-1}{m} = \frac{(n+m-1)!}{m!\,(n-1)!} \tag{4.2.1}$$

where m is the order of perturbation and n is the number of derivatives a first-order perturbation of the current type generates. For an example of a second-order geometric perturbation on a molecule containing four atoms, OpenRSP calculates only

$$\frac{(3\cdot 4+2-1)!}{2!\,(3\cdot 4-1)!} = 78 \tag{4.2.2}$$

derivatives, as opposed to the redundant case where $(3N)^2 = 12^2 = 144$ derivatives would have been calculated.

4.3 Postprocessing

With the energy derivatives calculated, all the components needed to calculate vibrational frequencies and intensities are in place. The Python package SpectroscPy[71] reads the output from the LSDalton calculation, and performs an eigenanalysis on the molecular Hessian, producing the harmonic vibrational frequencies and the transformation matrix from Cartesian to normal coordinates. After transforming all energy derivatives to the normal-coordinate basis, the theory presented in Chapter 2 is used to calculate intensities, in addition to, if requested, finding the anharmonic corrections to frequencies and intensities.

We recall from Section 4.2 that a geometry optimization and property calculation has been performed for every single snapshot collected from the

MD trajectory. Each of these configurations will produce slightly different energy derivatives and consequently slightly different frequencies and intensities. When presented in a table, an average of the values is given, but when creating our spectrum we wish to preserve the variation in the values in order to be as close as possible to reproducing the real spectrum. The peaks in an experimental spectrum will for instance never be symmetric because of the bulk nature of the sample, and the statistical sampling helps capture key aspects of the actual shape by including different solvent conformations through the different snapshots.

A linear combination of all the discrete frequency and intensity points belonging to mode k of snapshot i, $\{\omega_{ik}, I_{ik}\}$, is performed to get a continuous curve $S(\omega)$ defined as

$$S(\omega) = \frac{1}{n_{\text{snap}}} \sum_{i}^{n_{\text{snap}}} \sum_{k}^{n_{\text{mod}}} I_{ik} f(\omega, \omega_{ik}, \gamma_{ik}) \tag{4.3.1}$$

where the summation index i runs over the n_{snap} number of snapshots, the summation index k runs over the n_{mod} number of vibrational modes, I_{ik} is the intensity belonging to vibrational mode i of snapshot k. The Cauchy distribution f is defined as in Eq. (2.3.4), as a function of the frequency ω and parametrically of ω_{ik} and γ_{ik}. The broadening factor γ_{ik} is typically given as a single constant for all snapshots and normal modes, so that $\gamma_{ik} = \gamma$.

When comparing with for instance Eq. (2.3.7), notice that the molar attenuate IR coefficient already has a lineshape function at the end. The table values I_{ik} are in this case given by everything to the left of the lineshape function. In cases with integrated intensity-related properties, such as the Integrated Napierian attenuated coefficient in Eq. (2.3.12), the addition of a lineshape function at the end becomes somewhat more artificial. This is, however, what is done, and in order to provide comparability, also how we do it here.

Before concluding this section on the various software packages needed for a vibrational spectroscopic PE calculation, the recently released DaltonProject Python platform[102] should also be mentioned. The platform presents an alternative to running the programs in Sections 4.2 and 4.3 manually, one after the other. Instead, the DaltonProject links all the programs together, providing a user friendly, albeit somewhat less flexible setup.

Chapter 5

Summary of Papers

5.1 Paper I: Harmonic Infrared and Raman Spectra in Molecular Environments using the Polarizable Embedding Model

This paper is the first study of analytically calculated second-order geometric properties with the PE model. Acetone in various solvents has been investigated, in terms of harmonic frequencies and IR and Raman intensities, which requires energy-derivatives up to second order with regards to both electric and geometric perturbation. This paper is the culmination of many years of efforts, involving a number of different contributors and pieces functionality. This includes implementation in LSDalton of integrals necessary for PE contributions, making DFT functionality available with OpenRSP, and developing the necessary background theory. More directly related to this paper, a great work has been done with the implementation of FraME, which makes PE available both directly with LSDalton and in combination with OpenRSP. The SpectroscPy package was also first developed for this paper, and is a combination of renovated old scripts as well as novel functionality, that has all been put together into one package. My contribution to this paper has been in the interface between OpenRSP and FraME, in the parts of FraME related to geometric derivatives, as well as being the main developer of the SpectroscPy package, though not of the old scripts. I have also, with help, run the calculations and been integral in the writing of this paper.

5.2 Paper II: Analytic High-order Geometric Derivatives with Polarizable Embedding in a Response Theory Framework

The work included in this paper is an extension of the functionality and formulation presented in Paper I, to arbitrary order of electric and geometric properties, at the Hartree-Fock level of theory. Through our application on IR spectroscopy with anharmonic corrections through VPT2, we demonstrate the use of geometric derivatives up to fourth order as well as mixed geometric and electronic, where all derivatives are calculated analytically. Both solvated acetonitrile systems and endohedral fullerene have been investigated in this paper. In addition to implementation of arbitrary-order derivatives in FraME, implementation of GVPT2 frequencies and DVPT2 intensities has been done in SpectroscPy. An open-ended formulation of the PE equations has also been developed. My contributions to this paper have been the following. In FraME, I have been involved in the open-ended implementation together with the other authors, and in SpectroscPy, I have rewritten the aforementioned scripts as well as creating new functionality and flexibility for VPT2. I have produced the entirety of the results, though with some help, written large parts of the first draft to this article, and been intimately involved in the further work with the text. The formulation of the differentiated equations have been derived by myself, with guidance and feedback.

5.3 Paper III: Dalton Project: A Python platform for molecular- and electronic-structure simulations of complex systems

The paper is the result of an effort to increase automatization, customization and integration of different functionality, as well as user friendliness for the Dalton program family. A Python shell is created around Dalton and LSDalton and various other related program packages in order to in a straight-forward manner extract the desired quantum chemical values. A variety of different functionality is now available through this Python platform, and in addition to presenting the general aspects of the new platform, the

paper also provides six application examples. In this paper I am the main contributor to the application example ”D. Open-ended response theory for electric and geometric perturbations: Infrared and Raman spectroscopy with the OpenRSP and SpectroscPy modules“ where we demonstrate how vibrational properties can be extracted from the platform and also for the first time publish results provided through the Python pip-downloadable program package SpectroscPy.

Chapter 6

Conclusions & Outlook

In this thesis and its associated articles, we have shown how vibrational spectra can be calculated for solvated systems. Using a recursive and non-redundant strategy for the calculation of energy derivatives, and the PE method to include solvation effects, ensures results that are accurate, and once further work on optimization of our code has been done, also efficiently obtained. This work demonstrates the applications currently available, namely Infrared, Raman and Hyper-Raman spectroscopies, with or without anharmonic corrections, and with or without solvent effects. Although this setup is already highly useful, there are many possible extensions that could be implemented in order to improve accuracy or to add new applications. In this chapter I will cover some of these possible routes, starting with strategies for improving the accuracy of the methods used, and then by discussing possible new applications.

Currently, PE models dispersion and non-electrostatic repulsion effects through the LJ potential, which is not always sufficient. For instance when looking at electronically excited states, a repulsion modelled by the LJ potential, or not at all, can give unphysical results. Although electronic absorption is outside the scope of this work, it would still be interesting to investigate whether LJ is a sufficient model for dispersion and repulsion effects in the context of vibrational spectroscopy. This could be done by implementing and comparing with Curuchet's density-dependent LJ-like potential[88]. Another natural extension would be to implement PDE in the current framework, and thus include repulsion through a density-dependent repulsion operator. PDE is already available through other program pack-

ages[85], but not yet in FraME, which is also where any implementation of Curuchet's potential would also be performed, if desired.

On the side of vibrational spectroscopy, a possible improvement towards higher accuracy could be made by implementing the full GVPT2 for the anharmonic corrections, by adding variational solutions also to the DVPT2 transition moments. Furthermore, Eq. (2.5.4), which is used to calculate the anharmonic corrected transition moments, is still not fully general.[103] In cases where terms are removed because of identified 1-1 resonances, some cancellation of terms appearing in the derivation of Eq. (2.5.4) no longer hold. The actual numerical implications from this are, however, estimated to be small. In spite of this, the correct equation should be implemented in SpectroscPy in the future.

It would also be interesting to look into the possibilities of exchanging DFT with other QM methods in FraME, such as investigated by Sneskov *et al.*[75] with CC, or by Hedegård *et al.*[76] with MCSCF. As done by Hršak *et al.*[104], it is alternatively possible to calculate the fragment parameters, and in the case of PDE, frozen densities, with a different QM method than is used in the subsequent calculations. Taking this one step further, one could also consider taking the PE and PDE methods in a multilevel direction, for instance having different levels of theory layered within the QM region. This could be pursued in various directions, taking inspiration from FDE,[105] subsystem DFT[73] and Wave-Function Theory-in-DFT.[106] Another strategy, which has already been explored both for PE[107,108] and other methods,[35,109–112] is the combination of QM/MM methods with an outer solvation layer modelled by a continuum model. This reduces the number of specific solvent-solute interactions, while still retaining the most important ones close to the QM region. It is therefore a strategy that lowers the computational cost, while still being able to model key properties. Another benefit of including an outer PCM shell, is that while PE is modelled with a finite system-size, PCM does not have this limitation, and is in that respect closer to a bulk description. Within the framework of fluctuating charges/PCM (FQ/PCM), vibrational properties have already been investigated,[35,112] however, this is an as of yet untouched subject within PE (and PDE), and one that should be investigated.

On the application side, a natural path given the work within our group would be to facilitate the calculation of other vibrational or chiroptical spec-

troscopies such as Vibrational Circular Dichroism (VCD) and Raman Optical Activity with PE. This would require OpenRSP to calculate magnetic energy derivatives, FraME to provide the necessary PE contributions and SpectroscPy to calculate frequencies and intensities for these spectroscopies. A slightly more general implementation of the anharmonic corrections would also be necessary as it is now tailored for electric polarization properties only.

This work exclusively deals with solvated systems, or on other systems that do not involve any covalent bonds between the QM region and the environment, or between environmental fragments. The theory of the PE method, however, is applicable also to systems where this is not the case, such as in many biomolecules. For such systems, how to cut these bonds without too severely altering the vibrational motions, is a challenge that requires some extra care to be taken in comparison with the case where only weaker interactions are severed. Possible solutions might be to include more of the environment into the core region than that which strictly contributes to the vibrational effect (although this might in certain cases make in necessary to cut more covalent bonds), or to place artificially heavy atoms on the border between the two regions, in order to prevent too extreme deviations away from equilibrium positions. The method of Molecular fractionation with conjugate caps[113] (MFCC), where covalent cuts are capped with appropriate atoms or atomic groups, is already in use for determining fragment parameters in PyFraME[32], and could be implemented in FraME for property calculations.

In summary, there are many possible and interesting developments that could be explored both in the direction of improving the methods used and when it comes to which applications could be made possible. As most of them are contained to a single or few areas of the theory and software, they should be very realistic to obtain, and could even be explored simultaneously. One should, however, not underestimate the power of what is already in place, which should be able to provide interesting insights for many different applications.

Bibliography

[1] P. W. Atkins and J. De Paula, *Atkins' physical chemistry* (Oxford university press, 2010).

[2] P. A. M. Dirac, "Quantum mechanics of many-electron systems," Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **123**, 714 (1929).

[3] P. A. M. Dirac, "The Quantum Theory of the Electron," Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **117**, 610 (1928).

[4] M. Born and P. Jordan, "Zur quantenmechanik," Zeitschrift für Physik **34**, 858 (1925).

[5] N. Bohr, "Das Quantenpostulat und die neuere Entwicklung der Atomistik," Naturwissenschaften **16**, 245 (1928).

[6] W. Heisenberg, "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," Z. Phys. **41**, 172 (1927).

[7] E. Schrödinger, "Quantisierung als eigenwertproblem," Annalen der physik **385**, 437 (1926).

[8] P. W. Atkins and R. S. Friedman, *Molecular quantum mechanics* (Oxford university press, 2011).

[9] D. R. Hartree and W. Hartree, "Self-consistent field, with exchange, for beryllium," Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences **150**, 9 (1935).

[10] J. C. Slater, "Note on Hartree's method," Phys. Rev. **35**, 210 (1930).

[11] J. Čížek, "On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods," J. Chem. Phys. **45**, 4256 (1966).

[12] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," Phys. Rev. **136**, B864 (1964).

[13] T. Giovannini, F. Egidi, and C. Cappelli, "Molecular spectroscopy of aqueous solutions: a theoretical perspective," Chemical Society Reviews (2020).

[14] K. H. Hopmann, K. Ruud, M. Pecul, A. Kudelski, M. Dracinsky, and P. Bour, "Explicit versus implicit solvent modeling of raman optical activity spectra," J. Phys. Chem. B **115**, 4128 (2011).

[15] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," Chem. Rev. **105**, 2999 (2005).

[16] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," J. Mol. Biol. **103**, 227 (1976).

[17] A. R. Leach, *Molecular modelling: principles and applications* (Pearson education, 2001).

[18] M. Fyta, "Multiscale, hybrid, and coarse-grained methods," in *Computational Approaches in Physics*, 2053-2571 (Morgan & Claypool Publishers, 2016) pp. 7–1 to 7–17.

[19] A. R. Fersht, "Profile of Martin Karplus, Michael Levitt, and Arieh Warshel, 2013 Nobel laureates in chemistry," Proceedings of the National Academy of Sciences **110**, 19656 (2013).

[20] H. Hodak, "The Nobel Prize in chemistry 2013 for the development of multiscale models of complex chemical systems: a tribute to Martin Karplus, Michael Levitt and Arieh Warshel." J. Mol. Biol. **426**, 1 (2013).

[21] U. N. Morzan, D. J. Alonso de Armiño, N. O. Foglia, F. Ramírez, M. C. González Lebrero, D. A. Scherlis, and D. A. Estrin, "Spectroscopy in complex environments from QM–MM simulations," Chem. Rev. **118**, 4071 (2018).

[22] J. M. Olsen, K. Aidas, and J. Kongsted, "Excited states in solution through polarizable embedding," J. Chem. Theory Comput. **6**, 3721 (2010).

[23] J. M. H. Olsen, "PyFraME: Python tools for Fragment-based Multiscale Embedding (version 0.1.1)," (218), https://doi.org/10.5281/zenodo.1443314.

[24] L. Gagliardi, R. Lindh, and G. Karlström, "Local properties of quantum chemical systems: The LoProp approach," J. Chem. Phys. **121**, 4494 (2004).

[25] A. H. Steindal, M. T. P. Beerepoot, M. Ringholm, N. H. List, K. Ruud, J. Kongsted, and J. M. H. Olsen, "Open-ended response theory with polarizable embedding: multiphoton absorption in biomolecular systems," Phys. Chem. Chem. Phys. **18**, 28339 (2016).

[26] N. H. List, J. M. H. Olsen, and J. Kongsted, "Excited states in large molecular systems through polarizable embedding," Phys. Chem. Chem. Phys. **18**, 20234 (2016).

[27] C. Steinmann, J. M. H. Olsen, and J. Kongsted, "Nuclear magnetic shielding constants from quantum mechanical/molecular mechanical calculations using polarizable embedding: Role of the embedding potential," J. Chem. Theory Comput. **10**, 981 (2014).

[28] P. Reinholdt, M. S. Nørby, and J. Kongsted, "Modeling of Magnetic Circular Dichroism and UV/vis absorption spectra using Fluctuating Charges or Polarizable Embedding within a resonant-convergent response theory formalism," J. Chem. Theory Comput. **14**, 6391 (2018).

[29] A. J. Thorvaldsen, K. Ruud, K. Kristensen, P. Jørgensen, and S. Coriani, "A density matrix-based quasienergy formulation of the Kohn–Sham density functional response theory using perturbation-and time-dependent basis sets," J. Chem. Phys. **129**, 214108 (2008).

[30] M. Ringholm, D. Jonsson, and K. Ruud, "A general, recursive, and open-ended response code," J. Comput. Chem. **35**, 622 (2014).

[31] N. H. List, H. J. r. Jensen, and J. Kongsted, "Local electric fields and molecular properties in heterogeneous environments through polarizable embedding," Phys. Chem. Chem. Phys. **18**, 10070 (2016).

[32] C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted, and J. M. H. Olsen, "Response properties of embedded molecules through the polarizable embedding model," Int. J. Quantum Chem. **119**, e25717 (2019).

[33] N. H. List, M. T. P. Beerepoot, J. M. H. Olsen, B. Gao, K. Ruud, H. J. r. Jensen, and J. Kongsted, "Molecular quantum mechanical gradients within the polarizable embedding approach—Application to the internal vibrational Stark shift of acetophenone," J. Chem. Phys. **142**, 034119 (2015).

[34] L. J. Nåbo, J. M. H. Olsen, N. H. List, L. M. Solanko, D. Wüstner, and J. Kongsted, "Embedding beyond electrostatics—The role of wave function confinement," J. of Chem. Phys. **145**, 104102 (2016).

[35] F. Lipparini, C. Cappelli, G. Scalmani, N. De Mitri, and V. Barone, "Analytical first and second derivatives for a fully polarizable QM/classical hamiltonian," J. Chem. Theory Comput. **8**, 4270 (2012).

[36] T. Giovannini, M. Olszówka, F. Egidi, J. R. Cheeseman, G. Scalmani, and C. Cappelli, "Polarizable Embedding Approach for the Analytical Calculation of Raman and Raman Optical Activity Spectra of Solvated Systems," J. Chem. Theory Comput. **13**, 4421 (2017).

[37] R. Bast, D. H. Friese, B. Gao, D. J. Jonsson, S. Reine, M. Ringholm, K. Ruud, and A. J. Thorvaldsen, "OpenRSP," (2018), https://doi.org/10.5281/zenodo.1491928.

[38] E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular vibrations: the theory of infrared and Raman vibrational spectra* (Courier Corporation, 1955).

[39] V. Barone, "Anharmonic vibrational properties by a fully automated second-order perturbative approach," J. Chem. Phys **122**, 014108 (2005).

[40] J. M. L. Martin, T. J. Lee, P. R. Taylor, and J. François, "The anharmonic force field of ethylene, C2H4, by means of accurate ab initio calculations," J. Chem. Phys. **103**, 2589 (1995).

[41] J. Bloino, M. Biczysko, and V. Barone, "General perturbative approach for spectroscopy, thermodynamics, and kinetics: Methodological background and benchmark studies," J. Chem. Theory Comput. **8**, 1015 (2012).

[42] J. Bloino and V. Barone, "A second-order perturbation theory route to vibrational averages and transition properties of molecules: General formulation and application to infrared and vibrational circular dichroism spectroscopies," J. Chem. Phys. **136**, 124108 (2012).

[43] J. Bloino, M. Biczysko, and V. Barone, "Anharmonic effects on vibrational spectra intensities: infrared, Raman, vibrational circular dichroism, and Raman optical activity," J. Phys. Chem. A **119**, 11862 (2015).

[44] J. Bloino, "A VPT2 Route to Near-Infrared Spectroscopy: The Role of Mechanical and Electrical Anharmonicity," J. Phys. Chem. A **119**, 5269 (2015).

[45] K. Aidas *et al.*, "The Dalton quantum chemistry program system," WIREs Comput. Mol. Sci. **4**, 269 (2014).

[46] Dalton: a molecular electronic structure program (v2018.2), **2019**. See https://daltonprogram.org.

[47] S. Coriani *et al.*, "LSDALTON, a linear scaling molecular electronic structure program, Release Dalton2018 (2018), see http://daltonprogram. org," (2018).

[48] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular electronic-structure theory* (John Wiley & Sons, 2014).

[49] M. Born and R. Oppenheimer, "Zur quantentheorie der molekeln," Ann. Phys. **389**, 457 (1927).

[50] S. Reine and T. Saue, *European Summerschool in Quantum Chemistry Book III* (ESQC committee, 2017).

[51] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. **140**, A1133 (1965).

[52] J. P. Perdew and K. Schmidt, "Jacob's ladder of density functional approximations for the exchange-correlation energy," in *AIP Conference Proceedings*, Vol. 577 (AIP, 2001) pp. 1–20.

[53] S. Reine and T. Saue, *European Summerschool in Quantum Chemistry Book II* (ESQC committee, 2017).

[54] M. Ringholm, *Recursive calculation of high-order molecular response properties*, Ph.D. thesis (2013).

[55] P. Norman, K. Ruud, and T. Saue, *Principles and practices of molecular properties: Theory, modeling, and simulations* (Wiley Online Library, 2018).

[56] S. Aryal, "Infrared (IR) Spectroscopy," (2018), https://microbenotes.com/infrared-ir-spectroscopy/.

[57] S. E. Braslavsky, "Glossary of terms used in photochemistry, (IUPAC Recommendations 2006)," Pure and Applied Chemistry **79**, 293 (2007).

[58] A. D. McNaught, A. Wilkinson, *et al.*, *Compendium of chemical terminology*, Vol. 1669 (Blackwell Science Oxford, 1997).

[59] M. J. Baker, C. S. Hughes, and K. A. Hollywood, "Raman spectroscopy," in *Biophotonics: Vibrational Spectroscopic Diagnostics*, 2053-2571 (Morgan & Claypool Publishers, 2016) pp. 3–1 to 3–13.

[60] R. Paschotta *et al.*, "article on 'hyper Raman scattering'," in *Encyclopedia of laser physics and technology*, Vol. 1 (Wiley Online Library, 2008).

[61] F. Madzharova, Z. Heiner, and J. Kneipp, "Surface enhanced hyper Raman scattering (SEHRS) and its applications," Chem. Soc. Rev. **46**, 3980 (2017).

[62] M. J. Colles and J. E. Griffiths, "Relative and absolute Raman scattering cross sections in liquids," J. Chem. Phys. **56**, 3384 (1972).

[63] O. Quinet, B. Champagne, and V. Rodriguez, "Experimental and theoretical investigation of the Raman and hyper-Raman spectra of acetonitrile and its derivatives," J. Chem. Phys. **124**, 244312 (2006).

[64] P. Pulay, G. Fogarasi, G. Pongor, J. E. Boggs, and A. Vargha, "Combination of theoretical ab initio and experimental information to obtain reliable harmonic force constants. scaled quantum mechanical (qm) force fields for glyoxal, acrolein, butadiene, formaldehyde, and ethylene," J. Am. Chem. Soc. **105**, 7037 (1983).

[65] J. Baker, A. A. Jarzecki, and P. Pulay, "Direct scaling of primitive valence force constants: an alternative approach to scaled quantum mechanical force fields," J. Phys. Chem. A **102**, 1412 (1998).

[66] D. M. Bishop and B. Kirtman, "A perturbation method for calculating vibrational dynamic dipole polarizabilities and hyperpolarizabilities," J. Chem. Phys. **95**, 2646 (1991).

[67] H. H. Nielsen, "The Vibration-Rotation Energies of Molecules," Rev. Mod. Phys. **23**, 90 (1951).

[68] D. Papoušek and M. R. Aliev, *Molecular Vibrational-Rotational Spectra* (Elsevier, Amsterdam, NL, 1981).

[69] Wikipedia, "Morse potential," (2020), https://en.wikipedia.org/wiki/Morse_potential.

[70] J. Vázquez and J. F. Stanton, "Treatment of Fermi resonance effects on transition moments in vibrational perturbation theory," Mol. Phys. **105**, 101 (2007).

[71] K. O. H. M. Dundas, M. Ringholm, Y. Cornaton, and B. Ofstad, "SpectroscPy: The Python Package for Vibrational Spectroscopy," (2019), https://doi.org/10.5281/zenodo.3712927.

[72] J. M. H. Olsen and J. Kongsted, "Molecular properties through polarizable embedding," in *Advances in Quantum Chemistry*, Vol. 61 (Elsevier, 2011) pp. 107–143.

[73] C. R. Jacob and J. Neugebauer, "Subsystem density-functional theory," WIREs Comput. Mol. Sci. **4**, 325 (2014).

[74] L. O. Jones, M. A. Mosquera, G. C. Schatz, and M. A. Ratner, "Embedding Methods for Quantum Chemistry: Applications from Materials to Life Sciences," J. Am. Chem. Soc. **142**, 3281 (2020).

[75] K. Sneskov, T. Schwabe, J. Kongsted, and O. Christiansen, "The polarizable embedding coupled cluster method," J. Chem. Phys. **134**, 03B608 (2011).

[76] E. D. Hedegård, N. H. List, H. J. r. Jensen, and J. Kongsted, "The multi-configuration self-consistent field method within a polarizable embedded framework," J. Chem. Phys. **139**, 044101 (2013).

[77] M. T. P. Beerepoot, A. H. Steindal, N. H. List, J. Kongsted, and J. M. H. Olsen, "Averaged solvent embedding potential parameters for multiscale modeling of molecular properties," J. Chem. Theory Comput. **12**, 1684 (2016).

[78] J. M. H. Olsen, *Development of Quantum Chemical Methods towards Rationalization and Optimal Design of Optically Active Proteins*, Ph.D. thesis (2012).

[79] X. Saint Raymond, "Elementary introduction to the theory of pseudodifferential operators," Boca Raton (1991).

[80] J. Applequist, J. R. Carl, and K. Fung, "Atom dipole interaction model for molecular polarizability. application to polyatomic molecules and determination of atom polarizabilities," J. Am. Chem. Soc. **94**, 2952 (1972).

[81] A. Augustyn, W. Hosch, G. Lotha, *et al.*, *Encyclopædia Britannica* (Encyclopædia Britannica, inc., 2019).

[82] C. Caleman, P. J. van Maaren, M. Hong, J. S. Hub, L. T. Costa, and D. van der Spoel, "Force field benchmark of organic liquids: density,

enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant," J. Chem. Theory Comput. **8**, 61 (2012).

[83] D. van der Spoel, P. J. van Maaren, and C. Caleman, "GROMACS molecule & liquid database," Bioinformatics **28**, 752 (2012).

[84] G. Fradelos and T. A. Wesołowski, "Importance of the Intermolecular Pauli Repulsion in Embedding Calculations for Molecular Properties: The Case of Excitation Energies for a Chromophore in Hydrogen-Bonded Environments," J. Phys. Chem. A **115**, 10018 (2011).

[85] J. M. H. Olsen, C. Steinmann, K. Ruud, and J. Kongsted, "Polarizable density embedding: a new QM/QM/MM-based computational strategy," J. Phys. Chem. A **119**, 5344 (2015).

[86] T. A. Wesolowski and A. Warshel, "Frozen density functional approach for ab initio calculations of solvated molecules," J. Phys. Chem. **97**, 8050 (1993).

[87] S. Huzinaga and A. A. Cantu, "Theory of separability of many-electron systems," J. Chem. Phys. **55**, 5543 (1971).

[88] C. Curutchet, L. Cupellini, J. Kongsted, S. Corni, L. Frediani, A. H. Steindal, C. A. Guido, G. Scalmani, and B. Mennucci, "Density-dependent formulation of dispersion–repulsion interactions in hybrid multiscale quantum/molecular mechanics (QM/MM) models," J. Chem. Theory Comput. **14**, 1671 (2018).

[89] P. J. Olver, *Applications of Lie groups to differential equations*, Vol. 107 (Springer Science & Business Media, 2000) pp. 318–319.

[90] R. A. Adams and C. Essex, *Calculus: a complete course. Vol. 7* (Pearson Canada 7th, 2010) p. 109 and 115.

[91] M. Hardy, "Combinatorics of partial derivatives," Electronic Journal of Combinatorics (2006).

[92] R. A. Brualdi, "Introductory combinatorics, forth edition," (2004).

[93] M. Hanwell *et al.*, "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform," J. Cheminformatics **4**, 17 (2012).

[94] "Avogadro: an open-source molecular builder and visualization tool. Version 1.1.1." http://avogadro.cc/.

[95] D. van der Spoel *et al.*, "GROMACS: fast, flexible, and free," J. Comput. Chem. **26**, 1701 (2005).

[96] B. Hess *et al.*, "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation," J. Chem. Theory Comput. **4**, 435 (2008).

[97] M. Abraham *et al.*, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1**, 19 (2015).

[98] E. R. Kjellgren, J. M. H. Olsen, and J. Kongsted, "Importance of Accurate Structures for Quantum Chemistry Embedding Methods: Which Strategy Is Better?" J. Chem. Theory Comput. **14**, 4309 (2018).

[99] J. M. H. Olsen, K. O. H. M. Dundas, and M. Ringholm, FraME: Fortran library for Fragment-based Multiscale Embedding (development version), **2020**. Available at https://gitlab.com/FraME-projects/FraME.

[100] S. Reine, E. Tellgren, and T. Helgaker, "A unified scheme for the calculation of differentiated and undifferentiated molecular integrals over solid-harmonic Gaussians," Phys. Chem. Chem. Phys. **9**, 4771 (2007).

[101] S. Coriani, S. Høst, B. Jansík, L. Thøgersen, J. Olsen, P. Jørgensen, S. Reine, F. Pawłowski, T. Helgaker, and P. Sałek, "Linear-scaling implementation of molecular response theory in self-consistent field electronic-structure theory," J. Chem. Phys, **126**, 154108 (2007).

[102] J. M. H. Olsen *et al.*, "Dalton Project: A Python platform for molecular-and electronic-structure simulations of complex systems," J. Chem. Phys. **152**, 214115 (2020).

[103] J. Bloino, "Personal communication," (2020).

[104] D. Hršak, J. M. H. Olsen, and J. Kongsted, "Polarizable density embedding coupled cluster method," J. Chem. Theory Comput. **14**, 1351 (2018).

[105] T. A. Wesolowski and J. Weber, "Kohn-Sham equations with constrained electron density: an iterative evaluation of the ground-state electron density of interacting molecules," Chemical physics letters **248**, 71 (1996).

[106] A. S. P. Gomes, C. R. Jacob, and L. Visscher, "Calculation of local excitations in large systems by embedding wave-function theory in density-functional theory," Phys. Chem. Chem. Phys. **10**, 5353 (2008).

[107] A. H. Steindal, K. Ruud, L. Frediani, K. Aidas, and J. Kongsted, "Excitation energies in solution: the fully polarizable qm/mm/pcm method," J. Phys. Chem. B **115**, 3027 (2011).

[108] M. S. Nørby, C. Steinmann, J. M. H. Olsen, H. Li, and J. Kongsted, "Computational approach for studying optical properties of dna systems in solution," Journal of chemical theory and computation **12**, 5050 (2016).

[109] F. Lipparini, C. Cappelli, and V. Barone, "Linear response theory and electronic transition energies for a fully polarizable qm/classical hamiltonian," J. Chem. Theory Comput. **8**, 4153 (2012).

[110] S. Caprasecca, C. Curutchet, and B. Mennucci, "Toward a unified modeling of environment and bridge-mediated contributions to electronic energy transfer: a fully polarizable qm/mm/pcm approach," J. Chem. Theory Comput. **8**, 4462 (2012).

[111] H. Li and M. S. Gordon, "Polarization energy gradients in combined quantum mechanics, effective fragment potential, and polarizable continuum model calculations," J. Chem. Phys. **126**, 124112 (2007).

[112] T. Giovannini, M. Olszowka, and C. Cappelli, "Effective fully polarizable QM/MM approach to model vibrational circular dichroism spectra of systems in aqueous solution," J. Chem. Theory Comput. **12**, 5483 (2016).

[113] D. W. Zhang and J. Z. H. Zhang, "Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy," J. Chem. Phys. **119**, 3599 (2003).

Paper I

Harmonic Infrared and Raman Spectra in Molecular Environments using the Polarizable Embedding Model

K. O. H. M. Dundas, M. T. P. Beerepoot, M. Ringholm, S. Reine,
R. Bast, N. H. List, J. Kongsted, K. Ruud and J. M. H. Olsen
submitted to *J. Chem. Theory Comput.*

Harmonic Infrared and Raman Spectra in Molecular Environments using the Polarizable Embedding Model

Karen Oda Hjorth Minde Dundas,[†] Maarten T. P. Beerepoot,[†] Magnus Ringholm,[†] Simen Reine,[‡] Radovan Bast,[¶] Nanna Holmgaard List,[§,‖] Jacob Kongsted,[⊥] Kenneth Ruud,[*,†] and Jógvan Magnus Haugaard Olsen[*,#]

†*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway*

‡*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, N-0315 Oslo, Norway*

¶*Department of Information Technology, UiT The Arctic University of Norway, N-9037 Tromsø, Norway*

§*Department of Chemistry and The PULSE Institute, Stanford University, Stanford, CA 94305, USA*

‖*SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA*

⊥*Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK–5230 Odense M, Denmark*

#*Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark*

E-mail: kenneth.ruud@uit.no; jmho@chem.au.dk

Abstract

We present a fully analytic approach to calculate infrared (IR) and Raman spectra of molecules embedded in complex molecular environments, modeled by the fragment-based polarizable embedding (PE) model. We provide the theory for the calculation of analytic second-order geometric derivatives of molecular energies and first-order geometric derivatives of electric dipole moments and dipole–dipole polarizabilities within the PE model. The derivatives are implemented using a general open-ended response theory framework, thus allowing for an extension to higher-order derivatives. The embedding-potential parameters used to describe the environment in the PE model are derived through first-principles calculations, allowing a wide variety of systems to be modeled, including solvents, proteins, and other large and complex molecular environments. Here we present proof-of-principle calculations of IR and Raman spectra of acetone in different solvents. This work is an important step towards calculating accurate vibrational spectra in realistic molecular environments.

1 Introduction

Vibrational spectroscopy, in particular infrared (IR) absorption and Raman scattering, is one of the most important spectroscopic methods for elucidating molecular structure.[1] Many vibrational bands primarily signify the presence of certain chemical bonds and functional groups. However, the so-called fingerprint region, located between 1500 and 500 cm^{-1}, has in addition a spectral pattern that is often unique or near-unique for any given molecule, typically containing bands whose corresponding vibrational motions involve the backbone of the molecular structure. Whereas the use of databases of vibrational spectra of known compounds was essential in facilitating structural characterization of molecules in earlier days, this has in more recent years been complemented with a direct comparison to spectra obtained from quantum-chemical calculations.[2–4]

In the harmonic approximation, vibrational normal modes and their energy levels—determining the position of spectral peaks—are obtained from the second-order geometric

derivatives of the molecular energy with respect to nuclear displacements. Spectral intensities, on the other hand, are found from the normal-mode displacement gradient of the relevant polarization properties, which for IR absorption is the electric dipole moment and for Raman scattering the electric dipole–dipole polarizability.[5] From a computational perspective, an added challenge in the calculation of vibrational properties compared to, for instance, properties involving only electric–dipole perturbations,[6] is the dependence of the basis functions on the position of the nuclei.[2,4] The theory and implementations of analytic first-[7] and second-order[8] geometric derivatives of molecular energies were presented already in the late 1960s and 1970s, respectively. These developments, and in particular the analytical calculation of second-order geometric derivatives at the level of density-functional theory (DFT),[9–11] have today made quantum-chemical calculations an integral part of structural characterizations of molecules using vibrational spectroscopy. At the electron-correlated levels of theory, coupled-cluster methods now allow vibrational frequencies to be obtained with an accuracy that rivals that of even highly accurate experimental studies.[12–14] In addition, computationally elaborate schemes have been developed that allow anharmonicities to be efficiently calculated also at the DFT level of theory.[3,15,16] More recently, IR and Raman spectra have also been calculated from Car–Parrinello molecular dynamics simulations, in which anharmonic effects as well as broadening of peaks due to solvent interactions are automatically included.[17–19]

Vibrational spectroscopy is also an important tool to understand molecular bonding and the interaction of molecules with their surroundings. Even small inter- or intramolecular interactions may change the strength of particular bonds and thus the corresponding vibrational frequencies. Hydrogen bonding may have a particularly significant effect on bond strengths as well as on the coupling to other nuclei through its strongly directional nature.[20,21] Thus, vibrational spectroscopy is well suited to study subtle interactions even in large biomolecular aggregates.[22] This calls for computational methods that can model the effects of a surrounding environment.

Solvent effects are nowadays routinely included, for instance, through the use of continuum solvation models in calculations of vibrational spectra.[23] In particular, the polarizable continuum model (PCM)[24,25] is a popular approach used in quantum-chemical studies of solvated systems. However, specific intermolecular interactions are not described by such models. Moreover, modeling highly heterogeneous environments, such as proteins and other typical biomolecular systems, is problematic within a continuum approach. Quantum mechanics/molecular mechanics (QM/MM) methods, pioneered by Warshel and Levitt,[26] is an appealing alternative. In QM/MM, the chemically interesting part of a system is treated at a quantum-mechanical level of theory and the surroundings are treated using a classical molecular-mechanics force field. Unlike in the PCM, such QM/MM methods retain the atomistic structure of the environment, which is important for describing directional and structural effects on molecular properties. A wide range of different QM/MM methodologies has been developed.[27–29] They can be divided into three main classes depending on the level of approximation for the quantum–classical interactions, namely mechanical, electrostatic, and polarized embedding. In mechanical embedding, the interactions between the quantum and classical parts are described purely classically. In this approximation, there are only corrections to the energy and indirect geometric effects from the environment. The quantities needed for simulating IR and Raman spectra can thus be computed using the same approaches as for pure QM and MM. The next level of complexity is electrostatic embedding, where the electronic density of the quantum part is directly polarized by the charge distribution of the classical part, i.e. by the embedding potential. This is achieved through an embedding-potential operator that contains the electrostatic potential from the partial point charges (or more generally by the permanent multipoles) describing charge distribution of the classical part. Finally, in polarized embedding, the classical part is described by a polarizable potential that allows for mutual polarization between the quantum and classical parts.

Second-order geometric derivatives within an electrostatic-embedding QM/MM approach

were presented by Cui and Karplus.[30] They used the full Hessian to perform a vibrational analysis of the entire system, i.e., including both the QM and MM subsystems. The full vibrational analysis of such large systems may lead to computational bottlenecks in solving the vibrational eigenvalue problem due to the large matrices that would need to be diagonalized, as well as due to the high density of vibrational states. Various approaches have been proposed to deal with these challenges.[31–36] Li and Jensen[37] applied a partial Hessian vibrational analysis (PHVA)[38,39] to the effective fragment potential[40,41] method, which is a polarized embedding approach, using numerical differentiation to determine the Hessian for the quantum part. The PHVA approximation in a QM/MM setting usually implies that only the QM–QM block of the full Hessian is used, thus ignoring the MM–MM, QM–MM, and MM–QM blocks. More recently, Lipparini et al.[42] presented analytic second-order derivatives for a polarized-embedding approach based on fluctuating charges[43,44] and used this to compute IR spectra and later also Raman spectra,[45] in both cases within the PHVA approximation. Giovannini et al.[46] derived and implemented second-order derivatives to their extended fluctuating charges and dipoles model.

In this work, we present the theory and implementation of fully analytic first- and second-order geometric derivatives of energies and first-order geometric derivatives of dipole moments and dipole–dipole polarizabilities in the framework of the polarizable embedding (PE) model and within the PHVA approximation.[47,48] This work builds on our previous work on analytic first-order geometric derivatives of the PE energy.[49] The PE model can be characterized as fragment-based classical embedding akin to QM/MM, with the difference that it focuses solely on the central quantum part. In this model, the environment is represented by fragment-based distributed multipoles and polarizabilities. It can be used to model complex systems, such as solute–solvent systems and large biomolecules such as proteins and nucleic acids, as well as other large molecular systems that are amenable to fragmentation.[50] To split large molecules into smaller fragments, the molecular fractionation with conjugate caps (MFCC)[51,52] procedure can be used. The environment is treated classically, however, the

parameters (multipoles and polarizabilities) are derived from first-principles calculations on each individual fragment in the environment. This approach has been shown to yield highly accurate embedding potentials.[53–57]

The theory and implementation presented here build on earlier work that provide us with a flexible framework for the calculation of frequency-dependent molecular properties of arbitrary order for perturbation-dependent basis sets.[2,58–63] The additional contributions arising from the PE model for the calculation of molecular Hessians as well as dipole and polarizability gradients have been implemented so as to be used together with the general open-ended framework of OpenRSP.[58,59,64] The stage is thus set for extensions to higher-order geometric derivatives. Furthermore, the theory has been formulated in terms of the atomic-orbital (AO) density matrix, making the approach agnostic to the exact parametrization of the self-consistent field (SCF) wave function.[65]

The implementation is demonstrated through proof-of-principle calculations on acetone in various solvents. Acetone was selected as the model system because it has a rather simple vibrational spectrum and its semi-polar nature makes it soluble in both polar and non-polar solvents.

In the following, we first present the key quantities needed to calculate the additional contributions from the PE model to the molecular Hessian and the dipole and polarizability gradients in Sec. 2. In Sec. 3, we provide the computational details before we in Sec. 4 present our results for the computed IR and Raman spectra. We end the paper in Sec. 5 with some concluding remarks.

2 Theory

In this section, we first present a brief summary of the theory for IR and Raman spectroscopy in the double-harmonic approximation. Within this approximation, the necessary components are the molecular Hessian and first-order derivatives of the dipole moment and

polarizability with respect to nuclear displacements. The calculation of these properties at quantum-mechanical levels of theory is well established for molecular systems *in vacuo*. The reader is referred to relevant literature for details, see *e.g.* Ref. 6. Here, we focus on the contributions that arise when a molecule is embedded in a polarizable environment—specifically, when this environment is described by the PE model. The theoretical foundation for the PE model and its formulation within quantum-mechanical response theory has been extensively covered in earlier works,[47,48,62,66–68] and we therefore only present the basic equations of the PE model here. In the last part of this section, we present the contributions from the PE model to the second-order geometric derivatives of the energy as well as to the first-order geometric derivatives of the dipole moment and polarizability. The equations will be expressed in an AO SCF formulation, following earlier works.[58,59,62]

2.1 Vibrational frequencies and IR & Raman intensities

The harmonic approximation is frequently employed when describing vibrational wave functions and builds on a Taylor expansion of the energy E in terms of a set of mass-weighted nuclear Cartesian displacement coordinates relative to the equilibrium geometry

$$E(\mathbf{q}) = E_0 + \sum_{i=1}^{3N_{\text{nuc}}} \left.\frac{\partial E}{\partial q_i}\right|_{\mathbf{q}=0} q_i + \frac{1}{2} \sum_{i,j=1}^{3N_{\text{nuc}}} \left.\frac{\partial^2 E}{\partial q_i \partial q_j}\right|_{\mathbf{q}=0} q_i q_j + \cdots \tag{1}$$

The displacement coordinates are given by

$$\mathbf{q}^{\mathsf{T}} = \left(\sqrt{m_1}\Delta x_1 \ \sqrt{m_1}\Delta y_1 \ \sqrt{m_1}\Delta z_1 \ \sqrt{m_2}\Delta x_2 \ \ldots \ \sqrt{m_{N_{\text{nuc}}}}\Delta z_{N_{\text{nuc}}} \right) \tag{2}$$

where m_n is the mass of nucleus n and Δx_n, Δy_n, and Δz_n are the nuclear displacements from the equilibrium geometry of the Cartesian coordinates of nucleus n. The sums in eq (1) thus run over all Cartesian coordinates of the molecular geometry and the subscript $\mathbf{q} = 0$ denotes that the derivatives are evaluated at the equilibrium geometry. The first term on the right-

hand side of the equation is the energy at the equilibrium geometry that does not depend on the displacement coordinates and is therefore not important in the further analysis in this work. The second term contains the mass-weighted molecular gradient $\partial E/\partial q_i$, which is zero at the equilibrium geometry, and the last term contains the mass-weighted molecular Hessian $\partial^2 E/\partial q_i \partial q_j$. Through an eigenanalysis of the Hessian,[5] one can obtain the normal-mode frequencies from the eigenvalues, whereas the eigenvectors correspond to a transformation matrix that defines each normal coordinate Q_I in terms of Cartesian displacements. Three of the eigenvectors correspond to the overall translation of the system and three eigenvectors (two for linear molecules) correspond to the overall rotation of the system.

IR intensities are often reported in terms of the *molar decadic attenuation coefficient*, ε, which has units $\mathrm{m}^2 \cdot \mathrm{mol}^{-1}$. To facilitate comparisons to other works, we summarize the commonly used units for reporting IR intensities in Table 1. Within the double-harmonic approximation, ε for vibrational mode I is obtained from the equation

$$\varepsilon_I(\bar{\nu}) = \frac{N_\mathrm{A}}{12\ln(10)\varepsilon_0 c^2}\left[\sum_{\alpha=x,y,z}\left(\left.\frac{\partial\mu_\alpha}{\partial Q_I}\right|_{\mathbf{Q}=0}\right)^2\right] f(\bar{\nu};\bar{\nu}_I,\gamma_I) \tag{3}$$

where N_A is the Avogadro constant, c is the speed of light, ε_0 is the vacuum permittivity, and μ_α is a Cartesian component of the electric dipole moment. The lineshape function $f(\bar{\nu};\bar{\nu}_I,\gamma_I))$ is introduced to take into account homogeneous broadening effects, such as the finite lifetime of the excited vibrational states. In this work, we use a Cauchy distribution with a damping factor γ_I, so that[6]

$$f(\bar{\nu};\bar{\nu}_I,\gamma_I) = \frac{1}{\pi}\left[\frac{\gamma_I}{(\bar{\nu}_I-\bar{\nu})^2+\gamma_I^2}\right] \tag{4}$$

where $\bar{\nu}$ is the wavenumber of the incident radiation and $\bar{\nu}_I$ is the wavenumber associated with vibrational mode I. The lineshape function broadens the peaks with a half width at half maximum (HWHM) of the peak associated with mode I being γ_I. The dipole moment

gradient can also be expressed as a mixed energy derivative

$$\left.\frac{\partial \mu_a}{\partial Q_I}\right|_{\mathbf{Q}=0} = \left.\frac{\partial^2 E}{\partial F_\alpha \partial Q_I}\right|_{\mathbf{F}=0,\mathbf{Q}=0} \tag{5}$$

where F_α is the α'th component of the electric-field strength and the derivative is evaluated at zero field strength and at the equilibrium geometry.

Table 1: Units of Most Commonly Reported IR Intensities[6,69]

Property	Unit	Origin
–	$\mathrm{C}^2 \cdot \mathrm{kg}^{-1}$	$\mu_i^Q = \sum_{\alpha=x,y,z} \left(\frac{\partial \mu_\alpha}{\partial Q_i}\big|_{\mathbf{Q}=0}\right)^2$
–	$\mathrm{D}^2 \cdot \mathrm{Å}^{-2} \cdot \mathrm{amu}^{-1}$	$1.4924 \cdot 10^{12} \cdot \mu_i^Q$
Molar decadic attenuation coefficient	$\mathrm{m}^2 \cdot \mathrm{mol}^{-1}$	$\varepsilon_i = \frac{N_\mathrm{A}}{12 \ln(10) \varepsilon_0 c^2} \mu_i^Q f(\bar{\nu}; \bar{\nu}_i, \gamma_i)$
Molar decadic attenuation coefficient	$\mathrm{L} \cdot \mathrm{mol}^{-1} \cdot \mathrm{cm}^{-1}$	$10 \cdot \varepsilon_i$
Napierian integrated molecular attenuation coefficient	$\mathrm{m} \cdot \mathrm{mol}^{-1}$	$A_i = \ln(10) \int \varepsilon_i \mathrm{d}\bar{\nu} = \frac{N_\mathrm{A}}{12 c^2 \varepsilon_0} \mu_i^Q$
Napierian integrated molecular attenuation coefficient	$\mathrm{km} \cdot \mathrm{mol}^{-1}$	$10^{-3} \cdot A_i$

Just as for IR, there are several commonly used ways to report Raman intensities, but most are related to the *absolute differential Raman scattering cross section*, σ', with units $\mathrm{C}^4 \cdot \mathrm{s}^2 \cdot \mathrm{J}^{-1} \cdot \mathrm{m}^{-2} \cdot \mathrm{kg}^{-1}$. Some of the most commonly used Raman units are reported in Table 2. In the double-harmonic approximation and at temperature T, σ' is computed as[70,71]

$$\sigma'_I(\bar{\nu}) = \frac{\partial \sigma_I(\bar{\nu})}{\partial \Omega} = \frac{h(\bar{\nu}_0 - \bar{\nu}_I)^4}{16\pi^3 c^2 \bar{\nu}_I \left(1 - \exp\left(-\frac{hc\bar{\nu}_I}{kT}\right)\right)} \left(45a_I^2 + 7b_I^2\right) f(\bar{\nu}; \bar{\nu}_I, \gamma_I) \tag{6}$$

where σ is the total scattering cross section, Ω is the solid angle, $\bar{\nu}_0$ is the wavenumber of the incident light, and k is the Boltzmann constant. The constants 45 and 7 stem from the fact that we evaluate σ' for an experimental setup where the light entering the sample is polarized perpendicular to the direction of observation and its propagation.[5] Other choices of combination coefficients belong to other experimental setups. The Raman invariants a_I

and b_I^2 are given by[72]

$$a_I = \frac{1}{3} \sum_{\alpha=x,y,z} \left. \frac{\partial \alpha_{\alpha\alpha}}{\partial Q_I} \right|_{\mathbf{Q}=0} \tag{7}$$

and

$$b_I^2 = \sum_{\alpha=x,y,z} \sum_{\beta \neq \alpha} \left(\frac{1}{2} \left(\left. \frac{\partial \alpha_{\alpha\alpha}}{\partial Q_I} \right|_{\mathbf{Q}=0} - \left. \frac{\partial \alpha_{\beta\beta}}{\partial Q_I} \right|_{\mathbf{Q}=0} \right)^2 + 3 \left(\left. \frac{\partial \alpha_{\alpha\beta}}{\partial Q_I} \right|_{\mathbf{Q}=0} \right)^2 \right) \tag{8}$$

respectively, where $\alpha_{\alpha\beta}$ is the component of the electric dipole–dipole polarizability corresponding to Cartesian coordinates α and β. As the frequency-dependent polarizability gradient involves a frequency-dependent electric field, it cannot directly be represented as an energy derivative, and instead a quasienergy, $\tilde{E}$, (which reduces to the energy in the absence of a frequency-dependent electric field) derivative is used[4,6,73,74]

$$\left. \frac{\partial \alpha_{\alpha\beta}}{\partial Q_I} \right|_{\mathbf{Q}=0} = \left. \frac{\partial^3 \tilde{E}}{\partial F_\alpha \partial F_\beta \partial Q_I} \right|_{\mathbf{F}=0,\mathbf{Q}=0} \tag{9}$$

where the derivative is evaluated at zero field strengths and at the equilibrium geometry.

Table 2: Units of Most Commonly Reported Raman Intensities[70,75]

Property	Unit	Origin[a]
–	$\mathrm{C}^4 \cdot \mathrm{m}^2 \cdot \mathrm{J}^{-2} \cdot \mathrm{kg}^{-1}$	$\alpha_i^Q = k_a a_i^2 + k_b b_i^2$
–	$\mathrm{m}^4 \cdot \mathrm{kg}^{-1}$	$\left(\frac{1}{4\pi\varepsilon_0}\right)^2 \alpha_i^Q$
–	$\mathrm{Å}^4 \cdot \mathrm{amu}^{-1}$	$1.3413 \cdot 10^{33} \cdot \alpha_i^Q$
Absolute differential scattering cross section	$\mathrm{C}^4 \cdot \mathrm{s}^2 \cdot \mathrm{J}^{-1} \cdot \mathrm{m}^{-2} \cdot \mathrm{kg}^{-1}$	$\sigma' = \frac{h(\bar{\nu}_0 - \bar{\nu}_i)^4}{16\pi^3 c^2 \bar{\nu}_i \left(1 - \exp\left(-\frac{hc\bar{\nu}_i}{kT}\right)\right)} \alpha_i^Q f(\bar{\nu}; \bar{\nu}_i, \gamma_i)$

[a] Values of combination coefficients k_a and k_b depend on the experimental setup.[5] We use $k_a = 45$ and $k_b = 7$ in the present work.

2.2 Polarizable embedding

The PE model is an atomistic classical scheme for efficiently and accurately including complex environments in quantum-mechanical calculations. The total system is split into a core quantum region, which is described by a quantum-mechanical method, and its environment,

whose effects on the core part are described effectively through an embedding potential. The environment is further partitioned into computationally manageable fragments. In case of solvents typically consists of the individual solvent molecules, while a fragmentation procedure is used for more complex environments.[50,67] For each fragment, a quantum-mechanical calculation is performed, producing a set of electric multipoles and polarizabilities that are distributed to a number of sites within the fragment, usually the atomic centers. Alternatively, the multipoles and polarizabilities can be taken from existing pre-parametrized potentials that have been derived for proteins,[57] a series of solvents,[55] and a few lipids.[76]

The energy of a quantum region in the presence of an environment can be separated into two contributions,

$$E(\mathbf{D}) = E_{\mathrm{QM}}(\mathbf{D}) + E_{\mathrm{PE}}(\mathbf{D}) \tag{10}$$

where $E_{\mathrm{QM}}(\mathbf{D})$ is the energy of the quantum region, $E_{\mathrm{PE}}(\mathbf{D})$ is the embedding energy that describes the interactions between the quantum region and the environment, and $\mathbf{D}$ is the AO density matrix. In this work, Kohn–Sham density-functional theory (KS-DFT) is used for the quantum region, thus $E_{\mathrm{QM}}(\mathbf{D}) = E_{\mathrm{DFT}}(\mathbf{D})$. The KS-DFT energy is given by

$$E_{\mathrm{DFT}}(\mathbf{D}) \overset{\mathrm{Tr}}{=} \mathbf{hD} + \frac{1}{2}\mathbf{G}^{\gamma}(\mathbf{D})\mathbf{D} + E_{\mathrm{xc}}\left[\rho(\mathbf{D})\right] + h_{\mathrm{nuc}} \tag{11}$$

where $\overset{\mathrm{Tr}}{=}$ indicates that the trace is taken of each term on the right-hand side, $\mathbf{h}$ contains the one-electron terms (kinetic energy and electron–nuclear attraction), $\mathbf{G}^{\gamma}(\mathbf{D})$ contains the two-electron terms (electronic Coulomb and fractional exchange interactions), $E_{\mathrm{xc}}\left[\rho(\mathbf{D})\right]$ is the exchange–correlation contribution as a (nonlinear) functional of the density, and h_{nuc} is the nuclear–nuclear interaction energy. We will not go into further details about these terms here, but we note that the dependence of the individual contributions to the energy on the AO density matrix is either independent (h_{nuc}), linear ($\mathbf{hD}$), quadratic ($\mathbf{G}(\mathbf{D})\mathbf{D}$) or nontrivial ($E_{\mathrm{xc}}\left[\rho(\mathbf{D})\right]$), and that this separation of terms into orders of density-matrix dependence is used by OpenRSP. The contributions from the PE model to be presented

in the following can also be grouped into zeroth-, first-, and second-order density-matrix dependence. We have chosen to do so in this work to align our implementation with the corresponding interfaces to OpenRSP.

The PE energy can be written as

$$E_{\mathrm{PE}}(\mathbf{D}) = E_{\mathrm{es}}(\mathbf{D}) + E_{\mathrm{ind}}(\mathbf{D}) + E_{\mathrm{LJ}} \tag{12}$$

where $E_{\mathrm{es}}(\mathbf{D})$ is the electrostatic energy from the interaction between the permanent multipoles in the environment and the electrons and nuclei in the quantum region, $E_{\mathrm{ind}}(\mathbf{D})$ is the induction energy resulting from the polarization of the environment modeled by induced multipoles, and E_{LJ} is the energy due to non-electrostatic repulsion and dispersion interactions modeled by a 6-12 Lennard-Jones (LJ) potential.

In the following, we will present the electrostatic, induction, and LJ energies, where we will make use of a multi-index notation.[77] A multi-index is denoted by α, β, etc., and consists of three indices associated with the three Cartesian coordinates (i.e., $\alpha = (\alpha_x, \alpha_y, \alpha_z)$). The addition and subtraction of multi-indices is performed component-wise, i.e., $\alpha \pm \beta = (\alpha_x \pm \beta_x, \alpha_y \pm \beta_y, \alpha_z \pm \beta_z)$. The absolute value of a multi-index is defined as $|\alpha| = \alpha_x + \alpha_y + \alpha_z$, and the factorial as $\alpha! = \alpha_x! \alpha_y! \alpha_z!$. The multi-index power is given by $\mathbf{R}^\alpha = R^{\alpha_x} R^{\alpha_y} R^{\alpha_z}$. A partial derivative is written as $\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial^{\alpha_x} \partial^{\alpha_y} \partial^{\alpha_z}}$. Summing over the absolute value of a multi-index implicitly includes a sum over all possible multi-indices for each of the absolute values in the sum, e.g., $\sum_{|\alpha|=0}^{1} \alpha = (0,0,0) + (1,0,0) + (0,1,0) + (0,0,1)$. A Cartesian component of a tensor is specified with a multi-index in square brackets, e.g., $T^{[\alpha]}$.

The electrostatic energy describes the interactions between the electrons and nuclei in the quantum region and the permanent multipoles in the environment. Using the multi-index

notation, we can write it as

$$
\begin{aligned}
E_{\text{es}}(\mathbf{D}) = &- \sum_{a=1}^{N_{\text{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{\mu\nu} t_{\mu\nu}^{[\alpha]}(\mathbf{R}_s) D_{\mu\nu} \\
&+ \sum_{a=1}^{N_{\text{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{n=1}^{N_{\text{nuc}}} T^{[\alpha]}(\mathbf{R}_s, \mathbf{R}_n) Z_n \\
&\overset{\text{Tr}}{=} \mathbf{h}_{\text{es}} \mathbf{D} + h_{\text{es}}
\end{aligned}
\tag{13}
$$

where N_{frag} is the number of fragments in the environment, S_a is the number of sites in fragment a, K_s is the maximum order of the multipoles on site s in fragment a, $M_s^{[\alpha]}$ is a component of a Cartesian multipole on site s, μ and ν are indices of the AOs belonging to the quantum part, $t_{\mu\nu}^{[\alpha]}(\mathbf{R}_s)$ is the $\mu\nu$'th element of the $|\alpha|$'th-order derivative of a one-electron electrostatic-potential integral, $D_{\mu\nu}$ is the $\mu\nu$'th element of the AO density matrix, $T^{[\alpha]}(\mathbf{R}_s, \mathbf{R}_n)$ is a component of a Cartesian interaction tensor involving the positions of site s and nucleus n, N_{nuc} is the number of nuclei in the quantum region, and Z_n is the charge of nucleus n. An interaction tensor is generally defined as

$$
T^{[\alpha]}(\mathbf{R}_i, \mathbf{R}_j) = \partial_{\mathbf{R}_j}^{\alpha} \frac{1}{|\mathbf{R}_j - \mathbf{R}_i|}
\tag{14}
$$

where the subscript on the multi-index partial-derivative operator denotes the coordinate that the derivative is taken with respect to. The $t_{\mu\nu}^{[\alpha]}(\mathbf{R}_s)$ integrals can be defined in terms of interaction tensors as

$$
t_{\mu\nu}^{[\alpha]}(\mathbf{R}_s) = \int \chi_\mu(\mathbf{r}; \mathbf{R}_n) T^{[\alpha]}(\mathbf{R}_s, \mathbf{r}) \chi_\nu(\mathbf{r}; \mathbf{R}_n) \mathrm{d}\mathbf{r}
\tag{15}
$$

where $\mathbf{r}$ is the electron coordinate and $\chi_\mu(\mathbf{r}; \mathbf{R}_n)$ and $\chi_\nu(\mathbf{r}; \mathbf{R}_n)$ are AOs with a parametric dependence on the nuclear coordinates. The multipole–electron part of the interaction energy depends linearly on the density matrix while the multipole–nuclear interaction is a scalar that does not depend on the density matrix, as shown in the last equality of eq (13).

The second energy term in eq (12) is the induction energy, which is the result of the polarization of the environment. The polarization is modeled using polarizabilities that give rise to induced multipoles describing the response of a given fragment to the fields from the electrons and nuclei in the quantum part as well as the permanent multipoles in the environment. The induction energy can be formulated in terms of a generalized classical linear-response matrix of Cartesian polytensors (which are defined as a set of Cartesian tensors in a sequence of increasing rank) [78]

$$\mathcal{A} = \begin{pmatrix} \mathbf{P}_1^{-1} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1S} \\ \mathbf{T}_{21} & \mathbf{P}_2^{-1} & \cdots & \mathbf{T}_{2S} \\ \vdots & & \ddots & \vdots \\ \mathbf{T}_{S1} & \cdots & \cdots & \mathbf{P}_S^{-1} \end{pmatrix} \tag{16}$$

whose diagonal blocks contain inverse Cartesian polytensors that themselves consist of the multipole–multipole polarizabilities of a given site while the off-diagonal blocks hold the corresponding polytensors that consist of interaction tensors which describe the interaction between polarizable sites. The induction energy can then be written as

$$E_{\text{ind}} = \frac{1}{2}\bar{\mathcal{M}}\mathcal{V} \tag{17}$$

where $\bar{\mathcal{M}}$ is a matrix containing polytensors of the induced multipoles and $\mathcal{V}$ is a matrix that consists of polytensors that contain the derivatives of the electrostatic potential from the electrons, nuclei, and permanent multipoles at the polarizable sites. The induced multipoles can be determined by solving the matrix equation

$$\bar{\mathcal{M}} = -\mathcal{A}^{-1}\mathcal{V} = -\mathcal{B}\mathcal{V} \tag{18}$$

In practice, the matrix equation is never solved explicitly, since the linear-response matrix quickly becomes too large for environments with many sites, and instead an iterative solver

is used. Using the multi-index notation, the induction energy can be written as

$$
\begin{aligned}
E_{\mathrm{ind}}(\mathbf{D}) &= \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \sum_{b=1}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\beta|=0}^{K_t} \frac{-1}{2\alpha!\beta!} V^{[\alpha]}(\mathbf{D},\mathbf{R}_s) B_{st}^{[\alpha+\beta]} V^{[\beta]}(\mathbf{D},\mathbf{R}_t) \\
&= \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \sum_{b=1}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\beta|=0}^{K_t} \frac{-1}{\alpha!\beta!} \Bigg[\frac{1}{2} V_{\mathrm{e}}^{[\alpha]}(\mathbf{D},\mathbf{R}_s) B_{st}^{[\alpha+\beta]} V_{\mathrm{e}}^{[\beta]}(\mathbf{D},\mathbf{R}_t) \\
&\quad + \big(V_{\mathrm{n}}^{[\alpha]}(\mathbf{R}_s) + V_{\mathrm{m}}^{[\alpha]}(\mathbf{R}_s)\big) B_{st}^{[\alpha+\beta]} V_{\mathrm{e}}^{[\beta]}(\mathbf{D},\mathbf{R}_t) \\
&\quad + \frac{1}{2} \big(V_{\mathrm{n}}^{[\alpha]}(\mathbf{R}_s) + V_{\mathrm{m}}^{[\alpha]}(\mathbf{R}_s)\big) B_{st}^{[\alpha+\beta]} \big(V_{\mathrm{n}}^{[\beta]}(\mathbf{R}_t) + V_{\mathrm{m}}^{[\beta]}(\mathbf{R}_t)\big) \Bigg] \\
&\overset{\mathrm{Tr}}{=} \frac{1}{2} \mathbf{G}_{\mathrm{ind}}(\mathbf{D})\mathbf{D} + \mathbf{h}_{\mathrm{ind}}\mathbf{D} + h_{\mathrm{ind}}
\end{aligned}
\tag{19}
$$

where $V^{[\alpha]}(\mathbf{D},\mathbf{R}_s)$ is a component of the $|\alpha|$'th-order derivative of the electrostatic potential and $B_{st}^{[\alpha+\beta]}$ is a component of the st'th block of the inverse of the linear-response matrix in eq (16). In the second equality, we expand the energy in terms of derivatives of the electrostatic potentials from the electrons

$$
V_{\mathrm{e}}^{[\alpha]}(\mathbf{D},\mathbf{R}_s) \overset{\mathrm{Tr}}{=} -\mathbf{t}^{[\alpha]}(\mathbf{R}_s)\mathbf{D} \tag{20}
$$

nuclei

$$
V_{\mathrm{n}}^{[\alpha]}(\mathbf{R}_s) = \sum_{n=1}^{N_{\mathrm{nuc}}} T^{[\alpha]}(\mathbf{R}_n,\mathbf{R}_s) Z_n \tag{21}
$$

and permanent multipoles

$$
V_{\mathrm{m}}^{[\alpha]}(\mathbf{R}_s) = \sum_{b\neq a}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\beta|=0}^{K_t} \frac{(-1)^{|\alpha+\beta|}}{\alpha!} T^{[\alpha+\beta]}(\mathbf{R}_t,\mathbf{R}_s) M_t^{[\beta]} \tag{22}
$$

and collect terms that depend on the density to second, first, and zeroth order, respectively, which are then given in matrix form in the last equality. The sum over fragments in the multipole electrostatic potential (eq (22)) excludes the fragment that contains site s, here assumed to be fragment a.

Finally, the last term in eq (12) is the LJ potential energy, which effectively describes non-electrostatic repulsion and dispersion. It is given by

$$E_{\mathrm{LJ}} = 4 \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_{n=1}^{N_{\mathrm{nuc}}} \varepsilon_{sn} \left[\sigma_{sn}^{12} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{12} - \sigma_{sn}^{6} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{6} \right] \tag{23}$$

where Lorentz–Berthelot rules are used to combine parameters, i.e., $\sigma_{sn} = \frac{1}{2}(\sigma_s + \sigma_n)$ and $\varepsilon_{sn} = \sqrt{\varepsilon_s \varepsilon_n}$. Here, σ_s and ε_s are LJ parameters of atoms in the environment and σ_n and ε_n are LJ parameters of the atoms in the quantum region. The LJ potential energy is thus purely classical and independent of the density matrix.

The PE energy gives rise to Fock-matrix contributions that are found by minimizing the energy with respect to variations of the electron density

$$\mathbf{F}_{\mathrm{PE}} = \mathbf{h}_{\mathrm{es}} + \mathbf{G}_{\mathrm{ind}}(\mathbf{D}) + \mathbf{h}_{\mathrm{ind}} \tag{24}$$

2.3 Derivatives of the PE energy

In this section, we present the additional contributions to the geometric derivatives of the energy, dipole, and polarizability, that arise for a molecule embedded in a polarizable environment described by the PE model. These, and all other contributions, i.e., those for a molecule in vacuum, are considered in the framework of a density-matrix-based quasienergy formulation (see, e.g., works by Thorvaldsen et al.[58] and Ringholm, Jonsson, and Ruud[59] for details). In this approach, properties are determined as derivatives of the quasienergy Lagrangian, which up to third order can be written as[58]

$$L^{a_1} \overset{\{\mathrm{Tr}\}_T}{=} \mathcal{E}^{0,a_1} - \mathbf{S}^{a_1}\mathbf{W} \tag{25}$$

$$L^{a_1 a_2} \overset{\{\mathrm{Tr}\}_T}{=} \mathcal{E}^{0,a_1 a_2} + \boldsymbol{\mathcal{E}}^{1,a_1}\mathbf{D}^{a_2} - \mathbf{S}^{a_1 a_2}\mathbf{W} - \mathbf{S}^{a_1}\mathbf{W}^{a_2} \tag{26}$$

$$L^{a_1a_2a_3} \overset{\{\mathrm{Tr}\}_T}{=} \boldsymbol{\mathcal{E}}^{0,a_1a_2a_3} + \boldsymbol{\mathcal{E}}^{1,a_1a_2}\mathbf{D}^{a_3} + \boldsymbol{\mathcal{E}}^{1,a_1a_3}\mathbf{D}^{a_2} + \boldsymbol{\mathcal{E}}^{1,a_1}\mathbf{D}^{a_2a_3} + \boldsymbol{\mathcal{E}}^{2,a_1}\mathbf{D}^{a_2}\mathbf{D}^{a_3}$$
$$- \mathbf{S}^{a_1a_2a_3}\mathbf{W} - \mathbf{S}^{a_1a_2}\mathbf{W}^{a_3} - \mathbf{S}^{a_1a_3}\mathbf{W}^{a_2} - \mathbf{S}^{a_1}\mathbf{W}^{a_2a_3} \tag{27}$$

where $\overset{\{\mathrm{Tr}\}_T}{=}$ means that a trace and time-average of each term on the right-hand side is taken, $\mathcal{E}$ is the quasienergy, $\mathbf{D}$ is the density matrix, $\mathbf{S}$ is the overlap matrix, and $\mathbf{W}$ is the energy-weighted density matrix

$$\mathbf{W} = \mathbf{DFD} \tag{28}$$

The superscripts a_1, a_2, and a_3 denote derivatives with respect to given perturbations (either geometric or electric dipole perturbations in this work) with associated frequencies ω_{a_1}, ω_{a_2}, and ω_{a_3}, respectively. The notation employed here for quasienergy derivatives of n'th order is defined as

$$\boldsymbol{\mathcal{E}}^{m,a_1a_2\cdots a_n} = \frac{\partial^{m+n}\mathcal{E}}{(\partial\mathbf{D}^{\mathsf{T}})^m \partial\varepsilon_{a_1}\partial\varepsilon_{a_2}\cdots\partial\varepsilon_{a_n}} \tag{29}$$

where ε_{a_1}, ε_{a_2}, ..., ε_{a_n} are the strengths associated with perturbations a_1, a_2, ..., a_n, respectively. The quasienergy derivatives are expressed using the $n+1$ rule where only n'th-order derivatives of the density matrix are needed to calculate a property of order $n+1$. Derivatives of the density matrix with respect to the perturbation designated as a_1 are not present as a consequence of the application of the time-averaged Hellmann–Feynman theorem in the derivation of the quasienergy gradient (eq (25)). We again refer to Thorvaldsen et al. [58] for further details concerning the approach. Finally, we note that the quasienergy derivatives reduce to standard energy derivatives for time-independent properties.

In the following, we use superscripts g_1 and g_2 to denote a derivative with respect to an arbitrary Cartesian component of an arbitrary nuclear coordinate, and superscripts f_1 and f_2 to denote a derivative with respect to an arbitrary Cartesian component of the external field. For the molecular properties treated in this work, the contributions from the interactions between the quantum region and its environment are found by taking the relevant derivatives of the interaction energies according to the forms indicated in eqs (25)–(27). These

expressions contain perturbed density and Fock matrices, the latter as part of the perturbed energy-weighted density matrix. The calculation of perturbed density and Fock matrices also entails the evaluation of contributions stemming from the derivatives of the PE Fock matrix (eq (24)) and contributions to the electronic Hessian when solving the response equations. We refer to previous work[58,59] for details about the general method used to obtain perturbed density and Fock matrices. We note, however, that the additional contributions to the perturbed density and Fock matrices arising from the interaction between the quantum region and its environment are included in the following presentation, although only derivatives of the energy are explicitly addressed.

We begin with the derivatives of the electrostatic interaction energy (eq (13)). Here there is a dependence on nuclear positions in the nuclear–multipole part and through the AOs in the electrostatic-potential integrals (eq (15)) that appear in the electron–multipole part. The contributions from the electrostatic interactions to the geometric gradient and Hessian are

$$L_{\mathrm{es}}^{g_1} \overset{\mathrm{Tr}}{=} \mathbf{h}_{\mathrm{es}}^{g_1}\mathbf{D} + h_{\mathrm{es}}^{g_1} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{h}_{\mathrm{es}}\mathbf{D} \tag{30}$$

$$\begin{aligned} L_{\mathrm{es}}^{g_1g_2} \overset{\mathrm{Tr}}{=}\ & \mathbf{h}_{\mathrm{es}}^{g_1g_2}\mathbf{D} + h_{\mathrm{es}}^{g_1g_2} + \mathbf{h}_{\mathrm{es}}^{g_1}\mathbf{D}^{g_2} - \mathbf{S}^{g_1g_2}\mathbf{D}\mathbf{h}_{\mathrm{es}}\mathbf{D} \\ & - \mathbf{S}^{g_1}\mathbf{D}^{g_2}\mathbf{h}_{\mathrm{es}}\mathbf{D} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{h}_{\mathrm{es}}\mathbf{D}^{g_2} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{h}_{\mathrm{es}}^{g_2}\mathbf{D} \end{aligned} \tag{31}$$

and the contributions to the dipole and polarizability gradients are

$$L_{\mathrm{es}}^{g_1f_1} \overset{\mathrm{Tr}}{=} \mathbf{h}_{\mathrm{es}}^{g_1}\mathbf{D}^{f_1} - \mathbf{S}^{g_1}\mathbf{D}^{f_1}\mathbf{h}_{\mathrm{es}}\mathbf{D} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{h}_{\mathrm{es}}\mathbf{D}^{f_1} \tag{32}$$

$$L_{\mathrm{es}}^{g_1f_1f_2} \overset{\mathrm{Tr}}{=} \mathbf{h}_{\mathrm{es}}^{g_1}\mathbf{D}^{f_1f_2} - \mathbf{S}^{g_1}\mathbf{D}^{f_1}\mathbf{h}_{\mathrm{es}}\mathbf{D}^{f_2} - \mathbf{S}^{g_1}\mathbf{D}^{f_2}\mathbf{h}_{\mathrm{es}}\mathbf{D}^{f_1} \tag{33}$$

Since we do not consider local field effects in this work, $\mathbf{h}_{\mathrm{es}}$ and h_{es} are independent of the external field. The first- and second-order geometric derivatives of the $\mathbf{h}_{\mathrm{es}}$ matrix that appear

in eqs (30)–(33) are given by

$$\mathbf{h}_{\rm es}^{g_1} = \sum_{a=1}^{N_{\rm frag}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \mathbf{t}^{[\alpha],g_1}(\mathbf{R}_s) \tag{34}$$

$$\mathbf{h}_{\rm es}^{g_1g_2} = \sum_{a=1}^{N_{\rm frag}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \mathbf{t}^{[\alpha],g_1g_2}(\mathbf{R}_s) \tag{35}$$

and the derivatives of $h_{\rm es}$ are

$$h_{\rm es}^{g_1} = \sum_{a=1}^{N_{\rm frag}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{n=1}^{N_{\rm nuc}} T^{[\alpha],g_1}(\mathbf{R}_s,\mathbf{R}_n) Z_n \tag{36}$$

$$h_{\rm es}^{g_1g_2} = \sum_{a=1}^{N_{\rm frag}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{n=1}^{N_{\rm nuc}} T^{[\alpha],g_1g_2}(\mathbf{R}_s,\mathbf{R}_n) Z_n \tag{37}$$

We next consider the induction energy (eq (19)) where there is a dependence on nuclear positions through the nuclear and electronic electrostatic potentials. The geometric gradient and Hessian of the induction energy are given by

$$L_{\rm ind}^{g_1} \overset{\rm Tr}{=} \frac{1}{2}\mathbf{G}_{\rm PE}^{{\rm ind},g_1}(\mathbf{D})\mathbf{D} + \mathbf{h}_{\rm ind}^{g_1}\mathbf{D} + h_{\rm ind}^{g_1} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{G}_{\rm ind}(\mathbf{D})\mathbf{D} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{h}_{\rm ind}\mathbf{D} \tag{38}$$

$$\begin{aligned} L_{\rm ind}^{g_1g_2} \overset{\rm Tr}{=} & \frac{1}{2}\mathbf{G}_{\rm ind}^{g_1g_2}(\mathbf{D})\mathbf{D} + \mathbf{h}_{\rm ind}^{g_1g_2}\mathbf{D} + h_{\rm ind}^{g_1g_2} + \mathbf{G}_{\rm ind}^{g_1}(\mathbf{D})\mathbf{D}^{g_2} + \mathbf{h}_{\rm ind}^{g_1}\mathbf{D}^{g_2} \\ & - \mathbf{S}^{g_1g_2}\mathbf{D}\mathbf{G}_{\rm ind}\mathbf{D} - \mathbf{S}^{g_1g_2}\mathbf{D}\mathbf{h}_{\rm ind}\mathbf{D} - \mathbf{S}^{g_1}\mathbf{D}^{g_2}\mathbf{G}_{\rm ind}(\mathbf{D})\mathbf{D} - \mathbf{S}^{g_1}\mathbf{D}^{g_2}\mathbf{h}_{\rm ind}\mathbf{D} \\ & - \mathbf{S}^{g_1}\mathbf{D}\mathbf{G}_{\rm ind}(\mathbf{D})\mathbf{D}^{g_2} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{h}_{\rm ind}\mathbf{D}^{g_2} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{G}_{\rm ind}^{g_2}(\mathbf{D})\mathbf{D} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{h}_{\rm ind}^{g_2}\mathbf{D} \end{aligned} \tag{39}$$

and the contributions to the dipole and polarizability gradients are given by

$$\begin{aligned} L_{\rm ind}^{g_1f_1} \overset{\rm Tr}{=} & \mathbf{G}_{\rm ind}^{g_1}(\mathbf{D})\mathbf{D}^{f_1} + \mathbf{h}_{\rm ind}^{g_1}\mathbf{D}^{f_1} - \mathbf{S}^{g_1}\mathbf{D}^{f_1}\mathbf{G}_{\rm ind}(\mathbf{D})\mathbf{D} - \mathbf{S}^{g_1}\mathbf{D}^{f_1}\mathbf{h}_{\rm ind}\mathbf{D} \\ & - \mathbf{S}^{g_1}\mathbf{D}\mathbf{G}_{\rm ind}(\mathbf{D})\mathbf{D}^{f_1} - \mathbf{S}^{g_1}\mathbf{D}\mathbf{h}_{\rm ind}\mathbf{D}^{f_1} \end{aligned} \tag{40}$$

$$L_{\text{ind}}^{g_1 f_1 f_2} \overset{\text{Tr}}{=} \mathbf{G}_{\text{ind}}^{g_1}(\mathbf{D})\mathbf{D}^{f_1 f_2} + \mathbf{h}_{\text{ind}}^{g_1}\mathbf{D}^{f_1 f_2} + \mathbf{G}_{\text{ind}}^{g_1}(\mathbf{D}^{f_1})\mathbf{D}^{f_2} - \mathbf{S}^{g_1}\mathbf{D}^{f_1}\mathbf{G}_{\text{ind}}(\mathbf{D})\mathbf{D}^{f_2}$$
$$- \mathbf{S}^{g_1}\mathbf{D}^{f_1}\mathbf{h}_{\text{ind}}\mathbf{D}^{f_2} - \mathbf{S}^{g_1}\mathbf{D}^{f_2}\mathbf{G}_{\text{ind}}(\mathbf{D})\mathbf{D}^{f_1} - \mathbf{S}^{g_1}\mathbf{D}^{f_2}\mathbf{h}_{\text{ind}}\mathbf{D}^{f_1} \tag{41}$$

Inserting the expression for the electronic electrostatic potential (eq (20)) allows us to write the first- and second-order geometric derivatives of the $\mathbf{G}_{\text{ind}}(\mathbf{D})$ matrix as

$$\mathbf{G}_{\text{ind}}^{g_1}(\mathbf{D}) = \sum_{a=1}^{N_{\text{frag}}}\sum_{s\in a}^{S_a}\sum_{|\alpha|=0}^{K_s}\sum_{b=1}^{N_{\text{frag}}}\sum_{t\in b}^{S_b}\sum_{|\beta|=0}^{K_t}\frac{-1}{\alpha!\beta!}\Big(\text{Tr}\big[\mathbf{t}^{[\alpha],g_1}(\mathbf{R}_s)\mathbf{D}\big]\, B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta]}(\mathbf{R}_t)$$
$$+ \text{Tr}\big[\mathbf{t}^{[\alpha]}(\mathbf{R}_s)\mathbf{D}\big]\, B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta],g_1}(\mathbf{R}_t)\Big) \tag{42}$$

$$\mathbf{G}_{\text{ind}}^{g_1 g_2}(\mathbf{D}) = \sum_{a=1}^{N_{\text{frag}}}\sum_{s\in a}^{S_a}\sum_{|\alpha|=0}^{K_s}\sum_{b=1}^{N_{\text{frag}}}\sum_{t\in b}^{S_b}\sum_{|\beta|=0}^{K_t}\frac{-1}{\alpha!\beta!}\Big(\text{Tr}\big[\mathbf{t}^{[\alpha],g_1 g_2}(\mathbf{R}_s)\mathbf{D}\big]\, B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta]}(\mathbf{R}_t)$$
$$+ \text{Tr}\big[\mathbf{t}^{[\alpha],g_1}(\mathbf{R}_s)\mathbf{D}\big]\, B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta],g_2}(\mathbf{R}_t) + \text{Tr}\big[\mathbf{t}^{[\alpha],g_2}(\mathbf{R}_s)\mathbf{D}\big]\, B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta],g_1}(\mathbf{R}_t)$$
$$+ \text{Tr}\big[\mathbf{t}^{[\alpha]}(\mathbf{R}_s)\mathbf{D}\big]\, B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta],g_1 g_2}(\mathbf{R}_t)\Big) \tag{43}$$

$$\mathbf{G}_{\text{ind}}^{g_1}(\mathbf{D}^{f_1}) = \sum_{a=1}^{N_{\text{frag}}}\sum_{s\in a}^{S_a}\sum_{|\alpha|=0}^{K_s}\sum_{b=1}^{N_{\text{frag}}}\sum_{t\in b}^{S_b}\sum_{|\beta|=0}^{K_t}\frac{-1}{\alpha!\beta!}\Big(\text{Tr}\big[\mathbf{t}^{[\alpha],g_1}(\mathbf{R}_s)\mathbf{D}^{f_1}\big]\, B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta]}(\mathbf{R}_t)$$
$$+ \text{Tr}\big[\mathbf{t}^{[\alpha]}(\mathbf{R}_s)\mathbf{D}^{f_1}\big]\, B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta],g_1}(\mathbf{R}_t)\Big) \tag{44}$$

The first- and second-order derivatives of the $\mathbf{h}_{\text{ind}}$ matrix are given by

$$\mathbf{h}_{\text{ind}}^{g_1} = \sum_{a=1}^{N_{\text{frag}}}\sum_{s\in a}^{S_a}\sum_{|\alpha|=0}^{K_s}\sum_{b=1}^{N_{\text{frag}}}\sum_{t\in b}^{S_b}\sum_{|\beta|=0}^{K_t}\frac{-1}{\alpha!\beta!}\Big(V_{\text{n}}^{[\alpha],g_1}(\mathbf{R}_s)B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta]}(\mathbf{R}_t)$$
$$+ \big(V_{\text{n}}^{[\alpha]}(\mathbf{R}_s) + V_{\text{m}}^{[\alpha]}(\mathbf{R}_s)\big)B_{st}^{[\alpha+\beta]}\mathbf{t}^{[\beta],g_1}(\mathbf{R}_t)\Big) \tag{45}$$

$$\mathbf{h}_{\text{ind}}^{g_1g_2} = \sum_{a=1}^{N_{\text{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \sum_{b=1}^{N_{\text{frag}}} \sum_{t\in b}^{S_b} \sum_{|\beta|=0}^{K_t} \frac{-1}{\alpha!\beta!} \Bigg(V_{\text{n}}^{[\alpha],g_1g_2}(\mathbf{R}_s) B_{st}^{[\alpha+\beta]} \mathbf{t}^{[\beta]}(\mathbf{R}_t) + V_{\text{n}}^{[\alpha],g_1}(\mathbf{R}_s) B_{st}^{[\alpha+\beta]} \mathbf{t}^{[\beta],g_2}(\mathbf{R}_t) + V_{\text{n}}^{[\alpha],g_2}(\mathbf{R}_s) B_{st}^{[\alpha+\beta]} \mathbf{t}^{[\beta],g_1}(\mathbf{R}_t) + \big(V_{\text{n}}^{[\alpha]}(\mathbf{R}_s) + V_{\text{m}}^{[\alpha]}(\mathbf{R}_s)\big) B_{st}^{[\alpha+\beta]} \mathbf{t}^{[\beta],g_1g_2}(\mathbf{R}_t) \Bigg) \tag{46}$$

The electrostatic potentials from the nuclei and multipoles that appear here are defined in eqs (21) and (22), respectively. Geometric derivatives of the electrostatic potentials from the multipoles disappear as they do not depend on nuclear positions, while the gradient and Hessian of the nuclear electrostatic potential are given by

$$V_{\text{n}}^{[\alpha],g_1}(\mathbf{R}_s) = \sum_{n=1}^{N_{\text{nuc}}} T^{[\alpha],g_1}(\mathbf{R}_n, \mathbf{R}_s) Z_n \tag{47}$$

$$V_{\text{n}}^{[\alpha],g_1g_2}(\mathbf{R}_s) = \sum_{n=1}^{N_{\text{nuc}}} T^{[\alpha],g_1g_2}(\mathbf{R}_n, \mathbf{R}_s) Z_n \tag{48}$$

The last part of the geometric derivatives of the induction energy is h_{ind}, which depends on nuclear positions through the nuclear electrostatic potential. The first- and second-order derivatives of this term are

$$h_{\text{ind}}^{g_1} = \sum_{a=1}^{N_{\text{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \sum_{b=1}^{N_{\text{frag}}} \sum_{t\in b}^{S_b} \sum_{|\beta|=0}^{K_t} \frac{-1}{\alpha!\beta!} V_{\text{n}}^{[\alpha],g_1}(\mathbf{R}_s) B_{st}^{[\alpha+\beta]} \big(V_{\text{n}}^{[\beta]}(\mathbf{R}_t) + V_{\text{m}}^{[\beta]}(\mathbf{R}_t)\big) \tag{49}$$

$$h_{\text{ind}}^{g_1g_2} = \sum_{a=1}^{N_{\text{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \sum_{b=1}^{N_{\text{frag}}} \sum_{t\in b}^{S_b} \sum_{|\beta|=0}^{K_t} \frac{-1}{\alpha!\beta!} \Bigg(V_{\text{n}}^{[\alpha],g_1g_2}(\mathbf{R}_s) B_{st}^{[\alpha+\beta]} \big(V_{\text{n}}^{[\beta]}(\mathbf{R}_t) + V_{\text{m}}^{[\beta]}(\mathbf{R}_t)\big) + V_{\text{n}}^{[\alpha],g_1}(\mathbf{R}_s) B_{st}^{[\alpha+\beta]} V_{\text{n}}^{[\beta],g_2}(\mathbf{R}_t) \Bigg) \tag{50}$$

where the geometric derivatives of the nuclear electrostatic potential are given in eqs (47) and (48).

Finally, there is the LJ potential energy (eq (23)) that only contributes to the geometric

gradient and Hessian because it neither depends on the external field nor the density matrix. The contributions to the geometric gradient and Hessian are

$$L_{\mathrm{LJ}}^{g_1} = 4 \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{n=1}^{N_{nuc}} \varepsilon_{sn} \Bigg[12\sigma_{sn}^{12} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{11} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{g_1} - 6\sigma_{sn}^{6} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{5} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{g_1} \Bigg] \tag{51}$$

$$\begin{aligned} L_{\mathrm{LJ}}^{g_1 g_2} = 4 \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{n=1}^{N_{nuc}} \varepsilon_{sn} \Bigg[& 132\sigma_{sn}^{12} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{10} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{g_1} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{g_2} \\ & + 12\sigma_{sn}^{12} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{11} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{g_1 g_2} \\ & - 30\sigma_{sn}^{6} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{4} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{g_1} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{g_2} \\ & - 6\sigma_{sn}^{6} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{5} \left(\frac{1}{|\mathbf{R}_n - \mathbf{R}_s|} \right)^{g_1 g_2} \Bigg] \end{aligned} \tag{52}$$

3 Computational Details

The properties needed to simulate harmonic IR and Raman spectra were calculated for acetone in three different solvents, namely water, chloroform, and acetone. To simulate IR and Raman spectra of solute–solvent systems, it is necessary to adequately sample the configurational space. In this work, we sampled structures for all three systems from classical molecular dynamics (MD) simulations. The partial Hessian and first-order dipole and polarizability derivatives were then calculated for each structure. The final spectra were obtained by convolution of the spectra of each structure. Spectra of acetone in vacuum and acetone in the three solvents using the PCM were also computed for comparison. In the following, we provide the details for each step. All input and output files, as well as scripts used to run the calculations and to extract data, have been deposited on Zenodo.[79]

3.1 Generation of structures

Classical MD simulations were performed using GROMACS 2019.3[80–82] employing the OPLS-AA force field.[83] The OPLS-AA topology for acetone and chloroform were taken from the GROMACS molecule and liquid database at virtualchemistry.org.[84,85] The TIP3P potential[86] was used for water. Initial cubic boxes of $60 \times 60 \times 60$ Å were created and the system was then minimized with 100 steps of steepest descent and 1000 steps of conjugate gradient (5000 in case of chloroform solvent). An equilibration protocol containing both NPT and NVT ensembles was performed. For water and acetone solvents, an initial simulation in the NPT ensemble was run for 0.5 ns, followed by a 2 ns simulation in the NVT ensemble. Because of difficulties with the equilibration of acetone in chloroform (see ref. 79 for details), these two steps were preceded by two additional equilibration steps, consisting of a 0.0001 ps NVT simulation and a 0.05 ps NPT simulation, with time-steps of 0.01 and 0.1 fs, respectively. Initial velocities were taken from a Maxwell distribution at 298 K. All simulations were performed with periodic boundary conditions, the leap-frog integrator, and a time step of 1 fs (for all except the aforementioned additional steps for acetone in chloroform). Non-bonded interactions were cut off at 15 Å and electrostatic interactions beyond the cut-off were treated using the smooth particle-mesh Ewald[87] method. The Berendsen thermostat (298 K) and barostat (1 bar) were used with a coupling constant of 0.5 ps to maintain the temperature and pressure in the NPT equilibration.[88] The velocity-rescaling thermostat[89] with a coupling constant of 0.5 ps was used to maintain the temperature at 298 K in the NVT simulations. After the equilibration steps, a 10 ns NVT production run was performed. We then extracted 250 snapshots at 10 ps intervals from the first 2.5 ns of the final trajectory.

3.2 Geometry optimization and property calculations

For each structure extracted from the MD trajectory, the geometry of the central acetone molecule was first optimized in the presence of rigid solvent molecules. The partial Hessian

and first-order dipole and polarizability derivatives were then calculated using the optimized structures. The 250 equidistant snapshots that were extracted from the MD trajectory were used to perform a preliminary analysis of the convergence with respect to sample size (see Section A.1). Based on this analysis, and taking into account the computational cost, we found that 50 equidistant snapshots is adequate for our purposes, which is to demonstrate our implementation through proof-of-principle calculations. We note here that with 50 snapshots, we could produce Raman spectra that are well converged with respect to the number of snapshots, whereas comparatively larger errors were observed for some IR-active modes, in particular the carbonyl stretching mode (see Figure A1).

The effects from the solvent were modeled by embedding potentials produced using PyFraME.[90] The solvent was extracted using a center-of-mass distance criterion, i.e., solvent molecules with their center of mass within the cut-off distance from the center of mass of the central acetone were included. We used a cut-off distance of 12 Å which results in adequate accuracy (see Section A.2). For each solvent molecule in the solvent shell, atom-centered multipoles and polarizabilities were derived using the LoProp scheme.[91,92] For this a calculation using the Dalton program[93,94] is performed employing the B3LYP[95–99] exchange–correlation functional and a recontracted version of 6-31+G*[100–102] (called loprop-6-31+G* in Dalton). LJ parameters were taken from the OPLS-AA force field.

All geometry optimizations were performed at the PBE0[103–106]/pcseg-2[107] level of theory. The PBE0 functional was chosen based on its accuracy in the modelling of molecular geometries.[105] The pcseg-2 basis set was chosen as is has shown to give good results with DFT for both molecular structures and vibrational properties.[108] Additional support for the choice of triple-zeta basis was found through a convergence analysis that showed it to be a good compromise between accuracy and computational cost compared to its double- and quadruple-zeta counterparts (see Section A.3). The LSDalton program[75,93] was used for optimizations in vacuum and in solvent utilizing the FraME library[109] for the environment contributions. These optimizations used a fine integration grid, an initial numerical Hes-

sian, and Baker convergence criteria.[110] A few snapshots were discarded at this stage due to convergence issues. Therefore 49, 48, and 47 snapshots for water, acetone, and chloroform solvents, respectively, were used in the subsequent property calculations. Geometry optimizations utilizing the PCM were performed with Gaussian 16[111] using the pcseg-2 basis set obtained from the Basis Set Exchange.[112] The geometry optimizations using Gaussian were performed with a tight SCF threshold (SCF=VeryTight) and a fine integration grid (Int=SuperFine). To accompany the PCM-based structures, we also performed a geometry optimization in vacuum using Gaussian with the same settings.

The partial Hessian and first-order dipole and polarizability derivatives of acetone in vacuum and in environments described by the PE model were calculated using LSDalton, FraME, and OpenRSP.[58,59,64] Gaussian was used for the PCM-based calculations and its accompanying vacuum calculations. The same settings were used for the property calculations as for the geometry optimizations. The frequency-dependent polarizability derivatives were calculated using an input wavelength of 514.5 nm. This corresponds to an argon laser that has been used in Raman experiments on aqueous acetone.[113] The energy derivatives and molecular geometry were used by the vibrational spectroscopy package SpectroscPy[114] to perform a Hessian eigenvalue analysis to obtain the harmonic vibrational frequencies and normal coordinates, and to calculate the IR and Raman intensities. Raman intensities were calculated at 298 K. IR and Raman spectra were generated by combining the individual spectra of each structure into a single spectrum. Specifically, for IR we use

$$\varepsilon(\bar{\nu}) = \frac{1}{N}\sum_{j=1}^{N}\sum_{I=1}^{N_Q}\varepsilon_I^j(\bar{\nu}) \tag{53}$$

where N is the number of snapshots, N_Q is the number of vibrational modes, and $\varepsilon_I^j(\bar{\nu})$ is the molar decadic attenuation coefficient of the I'th vibrational mode in snapshot j (eq (3)). For Raman, we similarly use

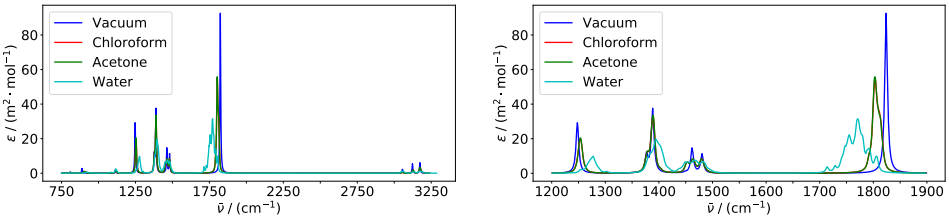
$$\sigma'(\bar{\nu}) = \frac{1}{N}\sum_{j=1}^{N}\sum_{I=1}^{N_Q}\sigma_I'^j(\bar{\nu}) \tag{54}$$

where $\sigma_I^{\prime j}(\bar{\nu})$ is the absolute differential scattering cross section of the I'th vibrational mode in snapshot j (eq (6)). The Cauchy distribution was used in both the IR and Raman cases as a basis for a lineshape function with a HWHM of 3.0 cm^{-1} for all modes (see eq (4)).

As mentioned in Section 2.1, the harmonic vibrational frequencies are found from an eigenanalysis of the molecular Hessian in mass-weighted Cartesian coordinates. This produces $3N$ frequencies and the corresponding normal modes, but not all of these are vibrational, as six of these (five for linear molecules) describe an overall translation and rotation of the molecule. In order to distinguish between low-frequency vibrational modes and the translational and rotational modes, it is common to project out translation and rotation from the Hessian. However, this approach cannot be used here, since we use the PHVA approximation. Moreover, the core molecule is embedded in a rigid solvent cage and is therefore no longer free to move around in space. This will inevitably introduce errors in our calculations. Low-frequency modes are especially susceptible to contamination by translational and rotational motions. Visual inspection of an arbitrarily chosen snapshot indicated that the six modes with lowest energy do not correspond to purely translational and rotational motion, and simultaneously that additional low-frequency modes also show some extent of globalized motion. Simply removing the six modes of lowest frequency is therefore not a good choice for the embedded systems. Instead, we identified from the visual inspection a cut-off at 750 cm^{-1} above which the modes have only negligible contamination of translational and rotational motion. We do not consider or discuss the normal modes with lower frequencies due to these impurities. By comparing the frequencies obtained with and without projecting out translation and rotation, we estimate that the error in the remaining vibrational modes is only a few cm^{-1} for the localized higher-frequency modes and never exceeds 10 cm^{-1} on average for any mode (see Figure S1 in the Supporting Information).

4 Results and Discussion

Harmonic IR and Raman spectra of acetone in vacuum and in water, chloroform, and acetone solutions are presented in Figures 1 and 2, respectively. Averaged vibrational wavenumbers and associated IR and Raman intensities are tabulated in the Supporting Information.



(a) Wavenumber range: 750–3250 cm^{-1} (b) Wavenumber range: 1200–1900 cm^{-1}

Figure 1: IR spectra of acetone in various environments modeled by the PE model. Spectra are based on averages over all snapshots. Calculations were performed using PE-PBE0/pcseg-2 with acetone embedded in a 12 Å solvent shell. A half width at half maximum value of 3.0 cm^{-1} was used to broaden individual peaks. Only modes above 750 cm^{-1} are included. Panel (b) displays the part the spectrum with highest IR absorption.

The three strongest peaks in the IR spectra presented in Figure 1 can be assigned[115] to the carbonyl stretching mode (around 1800 cm^{-1}), the symmetric methyl deformation (umbrella) mode (around 1390 cm^{-1}), and the asymmetric C–C stretching mode (around 1260 cm^{-1}). We will limit the following discussion to these three peaks. It is worth noting that the spectra for acetone in chloroform and acetone solutions are virtually overlapping, suggesting no significant differences in solute–solvent structure and dynamics for these two solvents. The effect of hydrogen bonding between the acetone solute and water solvent is evident from the -53 cm^{-1} shift of the carbonyl stretching mode relative to vacuum, whereas the shift is -21 cm^{-1} in chloroform and acetone solvents. The C–C stretching mode is shifted to higher wavenumbers by the solvents, though less in magnitude. Indeed, this shift is $+29$ cm^{-1} in water and only $+6$ cm^{-1} in chloroform and acetone. The wavenumber of the methyl umbrella mode is only slightly shifted by the water solvent ($+6$ cm^{-1}) while it is unaffected by the chloroform and acetone solvents. These shifts correlate well with the change in the

bond lengths that are presented in Table 3.

Table 3: Bond Lengths (in Å) in Acetone in Vacuum and Different Solvents Modeled by the PE Model and the PCM

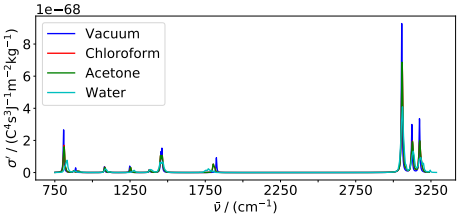
	Vacuum	Chloroform		Acetone		Water	
Bond		PE[a]	PCM	PE[a]	PCM	PE[a]	PCM
C=O	1.206	1.210 (0.001)	1.211	1.210 (0.001)	1.213	1.221 (0.005)	1.213
C–C	1.507	1.503 (0.002)	1.503	1.503 (0.002)	1.502	1.493 (0.006)	1.501
C–H	1.093	1.091 (0.003)	1.093	1.091 (0.003)	1.092	1.091 (0.003)	1.092
C–H'	1.087	1.091 (0.003)	1.087	1.091 (0.003)	1.087	1.091 (0.003)	1.087

[a] average over all snapshots with standard deviations in parentheses.

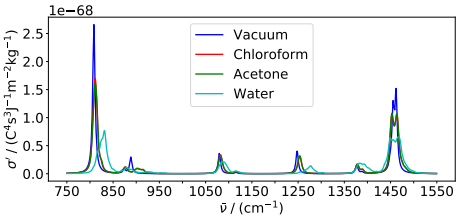
Acetone in aqueous solution forms hydrogen bonds with two water molecules on average, which results in an elongation of the carbonyl bond and a subsequent shift of the carbonyl stretch to lower wavenumbers. The C–C bonds, on the other hand, are contracted, which results in a shift of the C–C stretching frequency to higher wavenumbers. The methyl umbrella mode can be linked to the H–C–C bond angles, which vary only slightly in the presence of a solvent and are always between 109 and 111 degrees.

The configurational variety in the snapshots extracted from the MD simulation causes an inhomogeneous broadening. Even though the broadening of the peaks in the spectrum is in part determined by the chosen broadening factor, a comparison between the different solvents can be made. The most substantial broadening in the IR spectrum (Figure 1) is observed for the carbonyl stretch in water. Correspondingly, the standard deviations associated with the calculated wavenumber and IR intensity are 20 cm^{-1} and 23 $km \cdot mol^{-1}$, respectively. This is roughly three times larger than the standard deviations in chloroform and acetone, which are 6 cm^{-1} and 8 $km \cdot mol^{-1}$ for both solvents. The broadening of the carbonyl stretching mode in water can in part be attributed to the strong hydrogen bonding solvent. In contrast, the weaker dipole–dipole interactions between the acetone solute and chloroform and acetone solvent molecules result in smaller shifts and less pronounced broadening. The fine structure of the carbonyl stretching peak is most likely due to limited sampling.

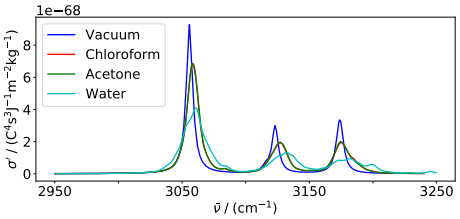
Raman spectra calculated with an input wavelength of 514.5 nm are shown in Figure 2.



(a) Wavenumber range: 750–3250 cm^{-1}



(b) Wavenumber range: 750–1550 cm^{-1}



(c) Wavenumber range: 2950–3250 cm^{-1}

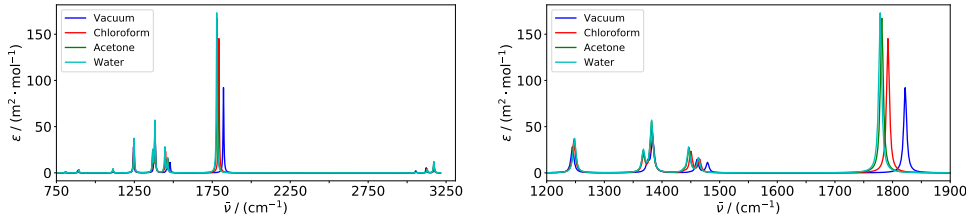
Figure 2: Raman spectra of acetone in various environments modeled by the PE model. Spectra are based on averages over all snapshots. Calculations were performed using PE-PBE0/pcseg-2 with acetone embedded in a 12 Å solvent shell and using an input wavelength of 514.5 nm. A half-width at half-maximum value of 3.0 cm^{-1} was used to broaden individual peaks. Only modes above 750 cm^{-1} are included. Panels (b) and (c) display the parts of spectrum with highest Raman activity.

The strongest peaks in the Raman spectrum can be assigned[115] to the symmetric and asymmetric C–H stretching modes (above 3000 cm^{-1}), the symmetric C–C stretch (around 800 cm^{-1}), and the asymmetric methyl deformation modes (around 1450 cm^{-1}). The spectra for the chloroform and acetone solutions are overlapping also for Raman scattering. Solvent effects are most apparent by the +23 cm^{-1} shift and substantial broadening (standard deviation of 16 cm^{-1}) of the C–C symmetric stretch in water. The corresponding shift in acetone and chloroform is only minor (+3 cm^{-1}). This is in agreement with the shortening of the C–C bond, which is 0.014 Å in water, 0.004 Å in chloroform, and 0.005 Å in acetone (Table 3). The frequency of the methyl deformation modes are virtually unchanged when adding a solvent. The symmetric C–H stretch is shifted by +5 cm^{-1} in water and +2 cm^{-1} in acetone and chloroform. The asymmetric C–H stretches are shifted by +9 cm^{-1} in water,

+4 cm^{-1} in acetone, and +1 cm^{-1} in chloroform. The broadening of these peaks is in part due to larger separation of the two modes underlying each of the peaks. In the case of the highest-frequency band in water, however, there is also a large spread of the wavenumbers of both underlying modes, with standard deviations of 17 and 13 cm^{-1}, respectively.

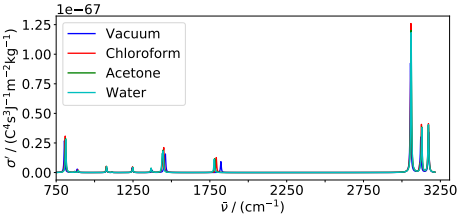
A question that naturally arises is whether the additional computational cost of the configurational sampling associated with the PE model is reasonable compared to using a continuum solvation model. To answer this question we calculated IR and Raman spectra using the PCM. Before comparing the spectra, we will briefly examine the effect on the geometry of the acetone solute. We note that the differences in bond lengths (Table 3) are small and may be of the same order as numerical errors, such as those introduced by the tessellation of the molecular cavity in the PCM. Addition of a solvent through PCM also leads to a slight elongation of the C=O bond and a slight shortening of the C–C bond and virtually no effect on the C–H bond length. The solvent effect on acetone bond lengths in chloroform, acetone, and water is very similar using the PCM. In other words, the larger solvent shift in water found in the PE calculations is not reproduced by the PCM. This reflects the lack of specific interactions (hydrogen bonds) in the PCM.

IR and Raman spectra for acetone in vacuum and in the presence of solvents modeled by the PCM are shown in Figures 3 and 4.

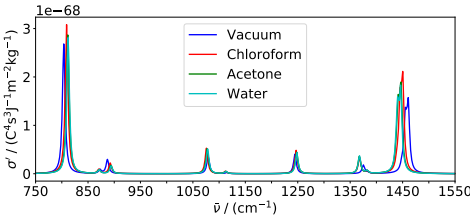


(a) Wavenumber range: 750–3250 cm^{-1} (b) Wavenumber range: 1200–1900 cm^{-1}

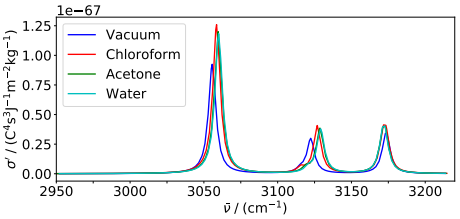
Figure 3: IR spectra of acetone in various environments modeled by the PCM. Calculations were performed using PCM-PBE0/pcseg-2. A half width at half maximum value of 3.0 cm^{-1} was used to broaden individual peaks. Only modes above 750 cm^{-1} are included. Panel (b) displays the part of the spectrum with highest IR absorption.



(a) Wavenumber range: 750–3250 cm^{-1}



(b) Wavenumber range: 750–1550 cm^{-1}



(c) Wavenumber range: 2950–3220 cm^{-1}

Figure 4: Raman spectra of acetone in various environments modeled by the PCM. Calculations were performed using PCM-PBE0/pcseg-2 using an input wavelength of 514.5 nm. A half width at half maximum value of 3.0 cm^{-1} was used to broaden individual peaks. Only modes above 750 cm^{-1} are included. Panels (b) and (c) display the parts of the spectrum with highest Raman activity.

When comparing these results with the corresponding spectra obtained using PE to model solvent effects (Figures 1 and 2), there are two substantial qualitative differences. First, the PCM is not able to reproduce the inhomogeneous broadening, due to lack of explicit configurational sampling. These effects are substantial in the spectra calculated with PE, especially for water. Second, the effect of acetone solvent is very similar to the effect of chloroform when modeled by the PE model, but similar to the effect of water when modeled by the PCM. The same trend is observed in the bond lengths (Table 3).

Solvents shifts of IR- and Raman-active modes modeled by the PCM are qualitatively similar to those modeled by the PE model, though there are some differences. The carbonyl stretching mode is shifted by −43 cm^{-1} from vacuum to water using the PCM, which is 10 cm^{-1} less than using the PE model. The acetone solvent shift of the carbonyl stretch is similar (−41 cm^{-1}) to the water solvent shift using the PCM, whereas it is only −21

cm^{-1} using the PE model. For the C–C stretching mode, the PCM predicts comparatively small shifts for all solvents, whereas the PE model predicts a much larger shift in water. Indeed, the asymmetric and symmetric modes are shifted +3 and +8 cm^{-1} with the PCM and +29 and +23 cm^{-1} with the PE model in water, +3 and +7 cm^{-1} with the PCM and +6 and +3 cm^{-1} with the PE model in acetone, and +2 and +5 cm^{-1} with the PCM and +6 and +3 cm^{-1} with the PE model in chloroform. The opposite behavior is observed for the asymmetric methyl deformation mode. None of the solvents cause a shift of this mode using the PE model, whereas the solvents shifts using the PCM model are -15 cm^{-1} in water and acetone and slightly less in chloroform.

The intensity of a peak is measured as the integral of the area under the peak and directly comparing heights between PE and PCM spectra can therefore be misleading. It is more sensible to compare PCM intensities to PE intensities that are averaged over the snapshots (see the Supporting information). The most prominent change in intensity upon solution is exhibited by the IR intensity of the carbonyl stretch in water, with an increase of 88 % with PCM and 51 % with PE. In general, changes in intensities upon solution are more pronounced using the PCM than using the PE model. Indeed, PCM gives larger intensities than PE for all modes except the symmetric C–H stretch, where the intensity is lowered by 22 % with PCM and by 10 % with PE.

5 Conclusions

We have presented the theory for the calculation of harmonic IR and Raman spectra of embedded molecules using the PE model to describe environment effects. The derived first- and second-order geometric derivatives of the energy and first-order geometric derivative of the dipole and polarizability are fully analytic and have been implemented in a general open-ended framework, thus facilitating extensions to higher-order geometric derivatives.

The implementation is illustrated through proof-of-principle calculations of IR and Ra-

man spectra for acetone in three different solvents, namely water, acetone, and chloroform. As expected, we observe that the presence of a solvent has a substantial effect on the IR and Raman spectra. This can be observed as frequency shifts, changes in intensities, broadening as well as alterations of the shape of the peaks. The effects of hydrogen bonding between the acetone solute and water as a solvent are evident especially from substantial shift and broadening of the carbonyl stretching mode in the IR spectrum and the C–C symmetric stretching mode in the Raman spectrum. These specific solute–solvent effects on the IR and Raman spectra can only be modeled with an atomistic description of the molecular environment. Apart from these specific interactions, comparison of calculations with the PCM and the PE model show qualitatively similar solvent effects, but in general larger frequency shifts with the PE model and larger intensity changes with the PCM.

This work is the first step towards modeling accurate vibrational spectra in realistic molecular environments. An extension of the present work to higher-order geometric derivatives is in progress. This will allow us to include second-order anharmonic effects through the calculation of cubic and quartic force fields. Moreover, the combination of the current implementation with higher-order electric derivatives[62] will enable the calculation of, for example, hyperpolarizability gradients and thus hyper-Raman spectroscopy. We will also explore the incorporation of local field effects through an extension of the effective external field model.[116,117]

A Convergence Analyses

We performed a series of convergence analyses with the aim of determining the basis set, size of the environment, and number of snapshots, that give an accurate representation of the investigated systems at a reasonable computational cost. For these analyses, we use the acetone-in-water system, since the aqueous environment was found to give the largest solvent effects of the solvents investigated here. To evaluate which basis set and cut-off radius to

use, we inspect the absolute error relative to a reference value which is the largest basis set and cut-off radius used. To determine the number of snapshots to include, we investigate the mean and maximum absolute errors of moving averages using samples of different sizes relative to an average value obtained using 250 snapshots. For a sample size S and a set of properties p_1, p_2, ..., p_N, where N is the total number of snapshots, the sample average can be defined as

$$\bar{p}_S^j = \frac{1}{S} \sum_{i=1+j}^{S+j} p_i \tag{55}$$

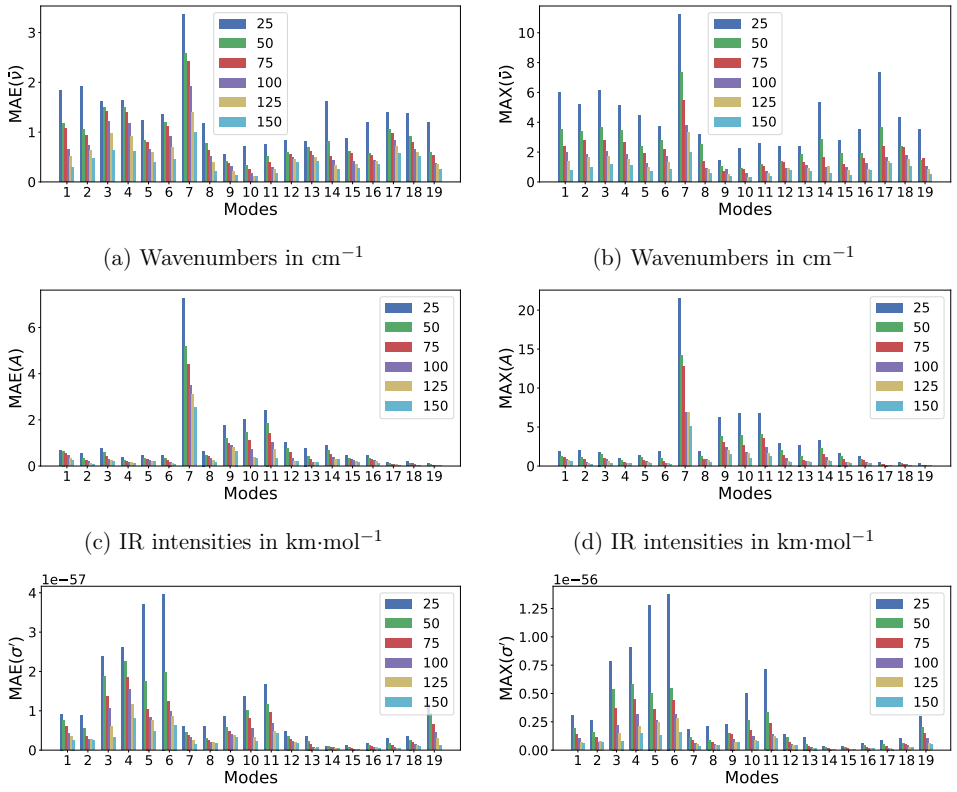
where j is the sample index, e.g., for $j = 0$, the sample includes properties p_1 to p_S, for $j = 1$ it includes properties p_2 to p_{S+1}, etc. For a given S, $\bar{p}_S^j$ can only be determined for $j \leq N - S$. The mean absolute error (MAE) for a given sample size is then found as

$$(\mathrm{MAE})_S = \frac{1}{N_S} \sum_j |\bar{p}_S^j - \bar{p}_{250}| \tag{56}$$

where N_S is the number of samples of size S and $\bar{p}_{250}$ is the global average, i.e., the average value across all snapshots. The maximum absolute error (MAX) for a given sample size is determined as the sample average that is furthest from the global average.

A.1 Convergence with sample size

In this section, we investigate the convergence with respect to sample size. The purpose is to determine how many snapshots that are needed to reach an error that does not add substantially to the errors introduced by the choice of basis set and system size. For this analysis, we consider the mean and maximum absolute errors of moving averages calculated for increasing sample sizes relative to a sample size of 250 snapshots. This gives an indication of the error that can be expected from sampling a number of snapshots consecutively from an MD trajectory. Due to the large number of snapshots, these calculations were performed at the HF/pcseg-1 level of theory in a 12 Å solvent shell of water. The results are presented in Figure A1. We observe a rather slow but steady convergence as the sample size is increased.



(a) Wavenumbers in cm^{-1}

(b) Wavenumbers in cm^{-1}

(c) IR intensities in $\mathrm{km \cdot mol}^{-1}$

(d) IR intensities in $\mathrm{km \cdot mol}^{-1}$

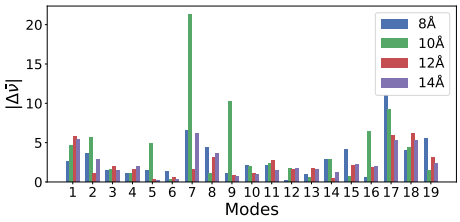
(e) Raman intensities in $\mathrm{C}^4 \cdot \mathrm{s}^2 \cdot \mathrm{J}^{-1} \cdot \mathrm{m}^{-2} \cdot \mathrm{kg}^{-1}$

(f) Raman intensities in $\mathrm{C}^4 \cdot \mathrm{s}^2 \cdot \mathrm{J}^{-1} \cdot \mathrm{m}^{-2} \cdot \mathrm{kg}^{-1}$

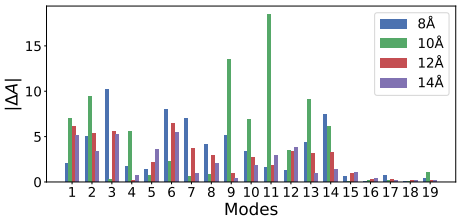
Figure A1: Convergence with respect to the sample size. Plots show mean and maximum absolute errors (MAE & MAX) of moving averages of a given sample size for all vibrational modes above 750 cm^{-1} compared to a sample size of 250. All calculations were performed using HF/pcseg-1 with acetone embedded in a 12 Å water shell.

The carbonyl stretching mode (no. 7 in Figure A1) has the largest error both in terms of wavenumbers and IR intensity, but has a very low Raman cross section. Even with 150 snapshots, the mean absolute errors for this mode are 1.0 cm^{-1} and 2.5 $\mathrm{km \cdot mol}^{-1}$ for wavenumbers and IR intensity, respectively, and the maximum absolute errors are 2.0 cm^{-1} and 5.0 $\mathrm{km \cdot mol}^{-1}$, which is of the same order as the basis-set error (see Section A.3). The Raman intensities, on the other hand, are reasonably well converged with a sample size of about 75 snapshots with mean and maximum absolute errors below 1.9×10^{-57} and 5.0×10^{-57}

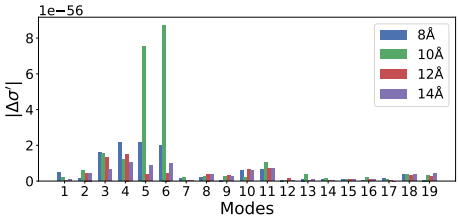
$C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$, respectively, which is well below the largest basis set error (see Section A.3). The convergence of Raman spectra with sample size have been studied previously in the context of Raman optical activity and much more simplistic QM/MM modeling,[118] where it was concluded that in view of the expected experimental errors in Raman intensities, 50 snapshots were required to give reliable Raman intensities. On this basis and considering the computational cost, we will use a sample size of 50 to calculate IR and Raman spectra, keeping in mind that this may result in comparatively large errors for wavenumbers and IR intensities for some of the modes, and in particular for the carbonyl stretching mode.



(a) Wavenumbers in cm^{-1}



(b) IR intensities in $km \cdot mol^{-1}$



(c) Raman intensities in $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$

Figure A2: Convergence with respect to size of the molecular environment. Plots show absolute errors of wavenumbers and associated IR and Raman intensities for all vibrational modes above 750 cm^{-1} compared to a 16 Å solvent shell. All calculations were performed using PBE0/pcseg-2 on a single snapshot of acetone embedded in water.

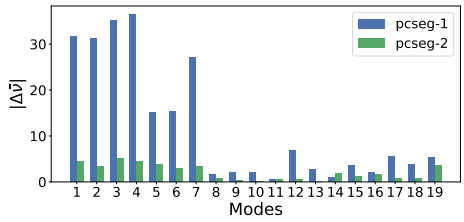
A.2 Convergence with size of molecular environment

To determine a suitable size of the solvent environment, calculations were performed for a single snapshot of aqueous acetone with a solvent shell radius ranging from 8 to 16 Å. The

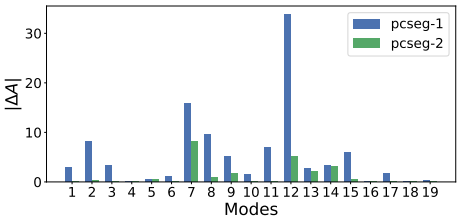
PBE0 functional was used together with the pcseg-2 basis set in all calculations. As can be seen from the absolute errors presented in Figure A2, none of the properties converge smoothly with system size. In fact, there appears to be an oscillating behavior for some of the modes. From a visual inspection, it is found that the main reason for this oscillating behavior is the change in the position of the acetone molecule within the solvent cavity, whereas the geometry of acetone remains fairly similar for all environments. In our current implementation, the computational cost grows rather steeply with increasing system size, mandating the need to balance cost to errors due to truncation of the size of the environment. Given this limitation, we find that a system size of 12 Å has residual errors that are of the same order as the errors due to our use of the pcseg-2 basis set (see Section A.3). Indeed, RMSDs relative to 16 Å decrease from 6.5 cm^{-1}, 6.8 $km{\cdot}mol^{-1}$, and $2.7{\times}10^{-56}$ $C^4{\cdot}s^2{\cdot}J^{-1}{\cdot}m^{-2}{\cdot}kg^{-1}$ for 10 Å to 2.9 cm^{-1}, 3.3 $km{\cdot}mol^{-1}$, and $5.6{\times}10^{-57}$ $C^4{\cdot}s^2{\cdot}J^{-1}{\cdot}m^{-2}{\cdot}kg^{-1}$ for 12 Å. Increasing the system size to 14 Å does not improve the overall error (RMSDs relative to 16 Å are 3.0 cm^{-1}, 2.8 $km{\cdot}mol^{-1}$, and $5.0{\times}10^{-57}$ $C^4{\cdot}s^2{\cdot}J^{-1}{\cdot}m^{-2}{\cdot}kg^{-1}$).

A.3 Convergence with basis set

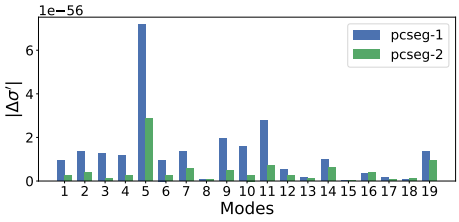
In order to identify the most accurate and cost-efficient basis set to be used in the calculation of the vibrational properties, vibrational frequencies and associated IR and Raman intensities were calculated for a single snapshot of aqueous acetone using PBE0 with three different basis sets: pcseg-1, pcseg-2, and pcseg-3[107] that are of double-, triple-, and quadruple-zeta quality, respectively. The frequency-dependent polarizability derivatives were evaluated at a wavelength of 514.5 nm. The size of the environment was arbitrarily set to include solvent molecules with a center of mass within a 10 Å radius from the center of mass of the acetone molecule. The results can be seen in Figure A3 as absolute errors compared to results obtained using pcseg-3. Wavenumbers (Figure A3a) are off by up to 37 cm^{-1} for the pcseg-1 basis set. In contrast, the largest errors obtained using pcseg-2 are about 3–5 cm^{-1} and are mostly associated with the high-frequency modes, i.e., C–H and C=O stretching



(a) Wavenumbers in cm^{-1}



(b) IR intensities in $km \cdot mol^{-1}$



(c) Raman intensities in $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$

Figure A3: Convergence with respect to basis set. Plots show absolute errors of wavenumbers and associated IR and Raman intensities for all vibrational modes above 750 cm^{-1} compared to pcseg-3. All calculations were performed using PBE0 on a single snapshot of acetone embedded in a 10 Å water shell.

modes.[115] In terms of intensities, using pcseg-1 results in errors that are generally below 10 $km \cdot mol^{-1}$ for IR (Figure A3b) and below 3×10^{-56} $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ for Raman (Figure A3c). Using pcseg-2 results in errors that are generally below 4 $km \cdot mol^{-1}$ for IR and below 0.7×10^{-56} $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ for Raman. A few vibrational modes dominate, showing larger errors for both basis sets. Specifically, the IR intensities of the C=O stretch and symmetric CH_3 deformation modes are off by about 8.1 and 5.2 km mol^{-1}, respectively, whereas in the Raman case the symmetric C–H and C–C stretching modes are off by 2.9×10^{-56} and 0.9×10^{-56} $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$, respectively, for the pcseg-2 basis set. These are the modes with the largests intensities and absorption cross sections. Overall, using pcseg-2 results in root-mean-squared deviations (RMSDs) of 2.7 cm^{-1}, 2.4 $km \cdot mol^{-1}$, and 0.77×10^{-56} $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$, whereas pcseg-1 results in RMSDs of 17.6 cm^{-1}, 9.5 $km \cdot mol^{-1}$, and 2.0×10^{-56} $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$, both compared to pcseg-3. Our results clearly show that

pcseg-1 is not adequate for accurate calculations of IR and Raman intensities. Taking into account that the relative computational cost of pcseg-2 compared to pcseg-3 is approximately one fourth, we conclude that pcseg-2 is a good compromise between accuracy and cost.

Acknowledgement

This work has received support from the Norwegian Supercomputing Program (NOTUR) through a grant of computer time (Grant No. NN4654K). Financial support is acknowledged from the Research Council of Norway through its Centres of Excellence scheme (Grant No. 262695) and a research grant (Grant No. 250743). M.R. acknowledges financial support from the Research Council of Norway and MSCA COFUND (Grant No. 274918). J.M.H.O. acknowledges financial support from VILLUM FONDEN (Grant No. 29478).

Supporting Information Available

References

(1) *Handbook of Vibrational Spectroscopy*; Wiley, 2002.

(2) Bast, R.; Ekström, U.; Gao, B.; Helgaker, T.; Ruud, K.; Thorvaldsen, A. J. The ab initio calculation of molecular electric, magnetic and geometric properties. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2627–2651.

(3) Barone, V.; Baiardi, A.; Biczysko, M.; Bloino, J.; Cappelli, C.; Lipparini, F. Implementation and validation of a multi-purpose virtual spectrometer for large systems in complex environments. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12404–12422.

(4) Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent advances in wave function-based methods of molecular-property calculations. *Chem. Rev.* **2012**, *112*, 543–631.

(5) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular vibrations: the theory of infrared and Raman vibrational spectra*; Courier Corporation, 1955.

(6) Norman, P.; Ruud, K.; Saue, T. *Principles and practices of molecular properties: Theory, modeling, and simulations*; Wiley Online Library, 2018.

(7) Pulay, P. Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules: I. Theory. *Mol. Phys.* **1969**, *17*, 197–204.

(8) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. Electron correlation theories and their application to the study of simple reaction potential surfaces. *Int. J. Quantum Chem.* **1978**, *14*, 545–560.

(9) Komornicki, A.; Fitzgerald, G. Molecular gradients and hessians implemented in density functional theory. *J. Chem. Phys.* **1993**, *98*, 1398–1421.

(10) Johnson, B. G.; Frisch, M. J. Analytic second derivatives of the gradient-corrected density functional energy. Effect of quadrature weight derivatives. *Chem. Phys. Lett.* **1993**, *216*, 133–140.

(11) Deglmann, P.; Furche, F.; Ahlrichs, R. An efficient implementation of second analytical derivatives for density functional methods. *Chem. Phys. Lett.* **2002**, *362*, 511–518.

(12) Morgan, W. J.; Matthews, D. A.; Ringholm, M.; Agarwal, J.; Gong, J. Z.; Ruud, K.; Allen, W. D.; Stanton, J. F.; Schaefer III, H. F. Geometric Energy Derivatives at the Complete Basis Set Limit: Application to the Equilibrium Structure and Molecular Force Field of Formaldehyde. *J. Chem. Theory Comput.* **2018**, *14*, 1333–1350.

(13) Gauss, J.; Stanton, J. F. Analytic first and second derivatives for the CCSDT-n (n=1-3) models: a first step towards the efficient calculation of CCSDT properties. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2047–2060.

(14) Stanton, J. F.; Gauss, J. Analytic second derivatives in high-order many-body perturbation and coupled-cluster theories: Computational considerations and applications. *Int. Rev. Phys. Chem.* **2000**, *19*, 61–95.

(15) Barone, V. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys.* **2005**, *122*, 014108.

(16) Barone, V.; Bloino, J.; Guido, C. A.; Lipparini, F. A fully automated implementation of VPT2 Infrared intensities. *Chem. Phys. Lett.* **2010**, *496*, 157–161.

(17) Gaigeot, M.-P.; Martinez, M.; Vuilleumier, R. Infrared spectroscopy in the gas and liquid phase from first principle molecular dynamics simulations: application to small peptides. *Mol. Phys.* **2007**, *105*, 2857–2878.

(18) Durlak, P.; Latajka, Z. Car-Parrinello molecular dynamics and density functional theory simulations of infrared spectra for acetic acid monomers and cyclic dimers. *Chem. Phys. Lett.* **2009**, *477*, 249–254.

(19) Thomas, M.; Brehm, M.; Fligg, R.; Voehringer, P.; Kirchner, B. Computing vibrational spectra from ab initio molecular dynamics. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6608–6622.

(20) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press, 1997.

(21) Grabowski, S. J. In *Hydrogen Bonding—New Insights*; Leszczynski, J., Ed.; Challenges and advances in computational chemistry and physics; Springer: Dordrecht, 2006; Vol. 3.

(22) Fried, S. D.; Boxer, S. G. Measuring electric fields and noncovalent interactions using the vibrational Stark effect. *Acc. Chem. Res.* **2015**, *48*, 998–1006.

(23) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3094.

(24) Miertus, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum– a utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* **1981**, *55*, 117–129.

(25) Cances, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

(26) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(27) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

(28) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev.* **2015**, *115*, 6217–6263.

(29) Morzan, U. N.; Alonso de Armiño, D. J.; Foglia, N. O.; Ramírez, F.; González Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in Complex Environments from QM–MM Simulations. *Chem. Rev.* **2018**, *118*, 4071–4113.

(30) Cui, Q.; Karplus, M. Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations. *J. Chem. Phys.* **2000**, *112*, 1133–1149.

(31) A., G.; Woodcock III, H. L.; Larkin, J. D.; Miller, B. T.; Shao, Y.; Kong, J.; Van Neck, D.; Van Speybroeck, V.; Waroquier, M.; Brooks, B. R. Efficient Calculation of QM/MM Frequencies with the Mobile Block Hessian. *J. Chem. Theory Comp.* **2011**, *7*, 496–514.

(32) Schwinn, K.; Ferré, N.; Huix-Rotllant, M. Efficient Analytic Second Derivative of Electrostatic Embedding QM/MM Energy: Normal Mode Analysis of Plant Cryptochrome. *J. Chem. Theory Comp.* **2020**, *16*, 3816–3824.

(33) Bouř, P.; Sopková, J.; Bednárová, L.; Maloň, P.; Keiderling, T. A. Transfer of Molecular Property Tensors in Cartesian Coordinates: A New Algorithm for Simulation of Vibrational Spectra. *J. Comp. Chem.* **1997**, *18*, 646–659.

(34) Reiher, M.; Neugebauer, J. A mode-selective quantum-chemical method for tracking molecular vibrations applied to functionalized carbon nanotubes. *J. Chem. Phys.* **2003**, *118*, 1634–1641.

(35) Herrmann, C.; Neugebauer, J.; Reiher, M. QM/MM Vibrational Mode Tracking. *J. Comp. Chem.* **2008**, *29*, 2460–2470.

(36) Kiewisch, K.; Neugebauer, J.; Reiher, M. Selective calculation of high-intensity vibrations in molecular resonance Raman spectra. *J. Chem. Phys.* **2008**, *129*, 204103.

(37) Li, H.; Jensen, J. H. Partial Hessian vibrational analysis: the localization of the molecular vibrational energy and entropy. *Theor. Chem. Acc.* **2002**, *107*, 211–219.

(38) Jin, S.; Head, J. D. Theoretical investigation of molecular water adsorption on the Al(111) surface. *Surf. Sci.* **1994**, *318*, 204–216.

(39) Calvin, M. D.; Head, J. D.; Jin, S. Theoretically modelling the water bilayer on the Al(111) surface using cluster calculations. *Surf. Sci.* **1996**, *345*, 161–172.

(40) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J Chem. Phys.* **1996**, *105*, 1968–1986.

(41) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.;

Stevens, W. J. The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry. *J. Phys. Chem. A* **2001**, *105*, 293–307.

(42) Lipparini, F.; Cappelli, C.; Scalmani, G.; De Mitri, N.; Barone, V. Analytical first and second derivatives for a fully polarizable QM/classical hamiltonian. *J. Chem. Theory Comput.* **2012**, *8*, 4270–4278.

(43) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

(44) Bryce, R. A.; Buesnel, R.; Hillier, I. H.; Burton, N. A. A solvation model using a hybrid quantum mechanical/molecular mechanical potential with fluctuating solvent charges. *Chem. Phys. Lett.* **1997**, *279*, 367–371.

(45) Giovannini, T.; Olszówka, M.; Egidi, F.; Cheeseman, J. R.; Scalmani, G.; Cappelli, C. Polarizable Embedding Approach for the Analytical Calculation of Raman and Raman Optical Activity Spectra of Solvated Systems. *J. Chem. Theory Comput.* **2017**, *13*, 4421–4435.

(46) Giovannini, T.; Grazioli, L.; Ambrosetti, M.; Cappelli, C. Calculation of IR Spectra with a Fully Polarizable QM/MM Approach Based on Fluctuating Charges and Fluctuating Dipoles. *J. Chem. Theory Comput.* **2019**, *15*, 5495–5507.

(47) Olsen, J. M.; Aidas, K.; Kongsted, J. Excited states in solution through polarizable embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(48) Olsen, J. M. H.; Kongsted, J. *Adv. Quantum Chem.*; Elsevier, 2011; pp 107–143.

(49) List, N.; Beerepoot, M. T. P.; Olsen, J. M. H.; Gao, B.; Ruud, K.; Jensen, H. J. r.; Kongsted, J. Molecular quantum mechanical gradients within the polarizable embedding approach—Application to the internal vibrational Stark shift of acetophenone. *J. Chem. Phys.* **2015**, *142*, 034119.

(50) Steinmann, C.; Reinholdt, P.; Nørby, M. S.; Kongsted, J.; Olsen, J. M. H. Response properties of embedded molecules through the polarizable embedding model. *Int. J. Quantum Chem.* **2019**, *119*, e25717.

(51) Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599–3605.

(52) Söderhjelm, P.; Ryde, U. How Accurate Can a Force Field Become? a Polarizable Multipole Model Combined with Fragment-wise Quantum-mechanical Calculations. *J. Phys. Chem. A* **2008**, *113*, 617–627.

(53) Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. Solvation Effects on Electronic Transitions: Exploring the Performance of Advanced Solvent Potentials in Polarizable Embedding Calculations. *J. Chem. Theory Comput.* **2011**, *7*, 2209–2217.

(54) Olsen, J. M. H.; List, N. H.; Kristensen, K.; Kongsted, J. Accuracy of Protein Embedding Potentials: An Analysis in Terms of Electrostatic Potentials. *J. Chem. Theory Comput.* **2015**, *11*, 1832–1842.

(55) Beerepoot, M. T. P.; Steindal, A. H.; List, N. H.; Kongsted, J.; Olsen, J. M. H. Averaged Solvent Embedding Potential Parameters for Multiscale Modeling of Molecular Properties. *J. Chem. Theory Comput.* **2016**, *12*, 1684–1695.

(56) Nørby, M. S.; Steinmann, C.; Olsen, J. M. H.; Li, H.; Kongsted, J. Computational Approach for Studying Optical Properties of DNA Systems in Solution. *J. Chem. Theory Comput.* **2016**, *12*, 5050–5057.

(57) Reinholdt, P.; Kjellgren, E. R.; Steinmann, C.; Olsen, J. M. H. Cost-Effective Potential for Accurate Polarizable Embedding Calculations in Protein Environments. *J. Chem. Theory Comput.* **2019**, *16*, 1162–1174.

(58) Thorvaldsen, A.; Ruud, K.; Kristensen, K.; Jørgensen, P.; Coriani, S. A density matrix-based quasienergy formulation of the Kohn–Sham density functional response theory using perturbation-and time-dependent basis sets. *J. Chem. Phys.* **2008**, *129*, 214108.

(59) Ringholm, M.; Jonsson, D.; Ruud, K. A general, recursive, and open-ended response code. *J. Comput. Chem.* **2014**, *35*, 622–633.

(60) Cornaton, Y.; Ringholm, M.; Louant, O.; Ruud, K. Analytic calculations of anharmonic infrared and Raman vibrational spectra. *Phys. Chem. Chem. Phys.* **2016**, *18*, 4201–4215.

(61) Cornaton, Y.; Ringholm, M.; Ruud, K. Complete analytic anharmonic hyper-Raman scattering spectra. *Phys. Chem. Chem. Phys.* **2016**, *18*, 22331–22342.

(62) Steindal, A. H.; Beerepoot, M. T. P.; Ringholm, M.; List, N. H.; Ruud, K.; Kongsted, J.; Olsen, J. M. H. Open-ended response theory with polarizable embedding: multiphoton absorption in biomolecular systems. *Phys. Chem. Chem. Phys.* **2016**, *18*, 28339–28352.

(63) Di Remigio, R.; Beerepoot, M. T. P.; Cornaton, Y.; Ringholm, M.; Steindal, A. H.; Ruud, K.; Frediani, L. Open-ended formulation of self-consistent field response theory with the polarizable continuum model for solvation. *Phys. Chem. Chem. Phys.* **2017**, *19*, 366–379.

(64) Bast, R.; Friese, D. H.; Gao, B.; Jonsson, D.; Ringholm, M.; Reine, S.; Ruud, K. OpenRSP: an open-ended response property library (version 1.0.0), **2020**, DOI: 10.5281/zenodo.3923836. See `https://www.openrsp.org/`.

(65) Bast, R.; Thorvaldsen, A. J.; Ringholm, M.; Ruud, K. Atomic orbital-based cubic response theory for one-, two- and four-component relativistic self-consistent field models. *Chem. Phys.* **2011**, *356*, 177.

(66) Olsen, J. M. H. Development of Quantum Chemical Methods towards Rationalization and Optimal Design of Photoactive Proteins. Ph.D. thesis, 2012; University of Southern Denmark, Faculty of Science, https://doi.org/10.6084/m9.figshare.156852.

(67) List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited states in large molecular systems through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20234–20250.

(68) List, N. H.; Norman, P.; Kongsted, J.; Jensen, H. J. A. A quantum-mechanical perspective on linear response theory within polarizable embedding. *J. Chem. Phys.* **2017**, *146*, 234101.

(69) Kendrick, J.; Burnett, A. D. PDielec: The calculation of infrared and terahertz absorption for powdered crystals. *J. Comput. Chem.* **2016**, *37*, 1491–1504.

(70) Colles, M. J.; Griffiths, J. E. Relative and absolute Raman scattering cross sections in liquids. *J. Chem. Phys.* **1972**, *56*, 3384–3391.

(71) Bloino, J.; Barone, V. A second-order perturbation theory route to vibrational averages and transition properties of molecules: General formulation and application to infrared and vibrational circular dichroism spectroscopies. *J. Chem. Phys.* **2012**, *136*, 124108.

(72) Barron, L. D. *Molecular light scattering and optical activity*; Cambridge University Press, 2009.

(73) Christiansen, O.; Jørgensen, P.; Hättig, C. Response functions from Fourier component variational perturbation theory applied to a time-averaged quasienergy. *Int. J. Quantum Chem.* **1998**, *68*, 1–52.

(74) Coriani, S.; Kjaergaard, T.; Jørgensen, P.; Ruud, K.; Huh, J.; Berger, R. An Atomic-Orbital-Based Lagrangian Approach for Calculating Geometric Gradients of Linear Response Properties. *J. Chem. Theory Comput.* **2010**, *6*, 1028–1047.

(75) LSDalton: a linear scaling molecular electronic structure program (v2020.0-dev), **2020**. See `https://daltonprogram.org`.

(76) Witzke, S.; List, N. H.; Olsen, J. M. H.; Steinmann, C.; Petersen, M.; Beerepoot, M. T. P.; Kongsted, J. An averaged polarizable potential for multiscale modeling in phospholipid membranes. *J. Comput. Chem.* **2017**, *38*, 601–611.

(77) Saint Raymond, X. *Elementary introduction to the theory of pseudodifferential operators*; CRC Press: Boca Raton, 1991.

(78) Applequist, J. Cartesian polytensors. *J. Math. Phys.* **1983**, *24*, 736–741.

(79) Dundas, K. O. H. M.; Beerepoot, M. T. P.; Ringholm, M.; Olsen, J. M. H. Dataset for the "Harmonic infrared and Raman spectra using a polarizable embedding model" . 2020; `https://doi.org/10.5281/zenodo.4271303`.

(80) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(81) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(82) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(83) Jorgensen, W. L.; Tirado-Rives, J. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6665–6670.

(84) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force field benchmark of organic liquids: Density, enthalpy of vaporization, heat ca-

pacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *J. Chem. Theory Comput.* **2012**, *8*, 61–74.

(85) van der Spoel, D.; van Maaren, P. J.; Caleman, C. GROMACS molecule & liquid database. *Bioinformatics* **2012**, *28*, 752–753.

(86) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(87) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(88) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(89) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(90) Olsen, J. M. H. PyFraME: Python framework for Fragment-based Multiscale Embedding (version 0.2.0), **2018**, DOI: 10.5281/zenodo.1443314. Available at `https://gitlab.com/FraME-projects/PyFraME`.

(91) Gagliardi, L.; Lindh, R.; Karlström, G. Local properties of quantum chemical systems: The LoProp approach. *J. Chem. Phys.* **2004**, *121*, 4494–4500.

(92) Vahtras, O. LoProp for Dalton (version 0.2.2), **2017**. Available at `https://github.com/vahtras/loprop`.

(93) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.;

Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansík, B.; Jensen, H. J. Aa.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.; Packer, M. J.; Pawlowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V. V.; Sałek, P.; Samson, C. C. M.; de Merás, A. S.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. The Dalton quantum chemistry program system. *WIREs Comput. Mol. Sci.* **2014**, *4*, 269–284.

(94) Dalton: a molecular electronic structure program (v2018.2), **2019**. See `https://daltonprogram.org`.

(95) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys* **1993**, *98*, 1372–1377.

(96) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys* **1993**, *98*, 5648–5652.

(97) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(98) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(99) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation

of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(100) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(101) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654–3665.

(102) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li-F. *J. Comput. Chem.* **1983**, *4*, 294–301.

(103) Perdew, J. P.; Burke, K. Ernzerhof, m., 1997. Errata: Generalized gradient approximation made simple. *Phys. Rev. Lett. 78*, 1396.

(104) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(105) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(106) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew–Burke–Ernzerhof exchange-correlation functional. *J. Chem. Phys.* **1999**, *110*, 5029–5036.

(107) Jensen, F. Unifying general and segmented contracted basis sets. Segmented polarization consistent basis sets. *J. Chem. Theory Comput.* **2014**, *10*, 1074–1085.

(108) Boussessi, R.; Ceselin, G.; Tasinato, N.; Barone, V. DFT meets the segmented polarization consistent basis sets: Performances in the computation of molecular struc-

tures, rotational and vibrational spectroscopic properties. *J. Mol. Struct.* **2020**, *1208*, 127886.

(109) Olsen, J. M. H.; Dundas, K. O. H. M.; Ringholm, M. FraME: Fortran library for Fragment-based Multiscale Embedding (development version), **2020**. Available at `https://gitlab.com/FraME-projects/FraME`.

(110) Baker, J. Techniques for geometry optimization: A comparison of Cartesian and natural internal coordinates. *J. Comput. Chem.* **1993**, *14*, 1085–1100.

(111) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16 Revision A.03. 2016; Gaussian Inc. Wallingford CT.

(112) Pritchard, B. P.; Altarawy, D.; Didier, B.; Gibson, T. D.; Windus, T. L. New basis set exchange: An open, up-to-date resource for the molecular sciences community. *J. Chem. Inf. Model* **2019**, *59*, 4814–4820.

(113) Idrissi, A.; Longelin, S.; Sokolić, F. Study of Aqueous Acetone Solution at Various

Concentrations: Low-Frequency Raman and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2001**, *105*, 6004–6009.

(114) Dundas, K. O. H. M.; Ringholm, M.; Cornaton, Y.; Ofstad, B. SpectroscPy: The Python Package for Vibrational Spectroscopy. **2020**. `https://doi.org/10.5281/zenodo.4099152`.

(115) Shimanouchi, T. Tables of Molecular Vibrational Frequencies Consolidated (National Bureau of Standards, Washington, DC, 1972). *Vol. I* 1–160.

(116) Cammi, R.; Cappelli, C.; Corni, S.; Tomasi, J. On the Calculation of Infrared Intensities in Solution within the Polarizable Continuum Model. *J. Phys. Chem. A* **2000**, *104*, 9874–9879.

(117) List, N. H.; Jensen, H. J. A.; Kongsted, J. Local electric fields and molecular properties in heterogeneous environments through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 10070–10080.

(118) Hopmann, K. H.; Ruud, K.; Pecul, M.; Kudeski, A.; Dračinsky, M.; Bouř, P. Explicit versus Implicit Solvent Modeling of Raman Optical Activity Spectra. *J. Phys. Chem. B* **2011**, *115*, 4128–4137.

Supporting Information: Harmonic Infrared and Raman Spectra in Molecular Environments using the Polarizable Embedding Model

Karen Oda Hjorth Minde Dundas,[†] Maarten T. P. Beerepoot,[†] Magnus Ringholm,[†] Simen Reine,[‡] Radovan Bast,[¶] Nanna Holmgaard List,[§,‖] Jacob Kongsted,[⊥] Kenneth Ruud,[*,†] and Jógvan Magnus Haugaard Olsen[*,#]

†*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, UiT The Arctic University of Norway, N-9037 Tromsø, Norway*

‡*Hylleraas Centre for Quantum Molecular Sciences, Department of Chemistry, University of Oslo, N-0315 Oslo, Norway*

¶*Department of Information Technology, UiT The Arctic University of Norway, N-9037 Tromsø, Norway*

§*Department of Chemistry and The PULSE Institute, Stanford University, Stanford, CA 94305, USA*

‖*SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA*

⊥*Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK–5230 Odense M, Denmark*

#*Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark*

E-mail: kenneth.ruud@uit.no; jmho@chem.au.dk

Table S1: Averaged Wavenumbers and Associated IR and Raman Intensities of Acetone in Water Modelled by the PE Model[a]

Mode	Wavenumber[b]	IR[c]	Raman ($\lambda_0 = 0$ nm)[d]	Raman ($\lambda_0 = 514.5$ nm)[e]
1	3193(16.7)	4.561(4.4)	44.281 (5.3)	33.956 (3.6)
2	3176(13.0)	5.534(4.3)	38.883 (5.8)	30.618 (4.4)
3	3136(8.4)	3.124(2.5)	46.876 (7.5)	39.472 (6.4)
4	3124(7.6)	2.363(2.2)	33.462 (8.9)	28.716 (7.5)
5	3063(7.9)	4.293(3.7)	159.816 (29.6)	148.469 (28.0)
6	3051(6.6)	3.958(3.1)	41.747 (29.7)	39.882 (28.1)
7	1769(19.8)	303.181(22.7)	1.307 (0.2)	18.107 (1.7)
8	1486(6.9)	24.907(5.5)	0.138 (0.1)	3.236 (2.3)
9	1470(4.9)	23.022(14.1)	0.677 (0.2)	17.014 (4.8)
10	1458(4.6)	12.528(11.7)	0.440 (0.2)	11.444 (5.0)
11	1448(5.1)	22.367(14.8)	0.598 (0.2)	16.111 (5.4)
12	1399(7.7)	99.709(8.6)	0.136 (0.1)	4.582 (2.7)
13	1383(5.7)	34.325(9.1)	0.169 (0.1)	6.188 (2.1)
14	1276(9.9)	52.941(9.5)	0.114 (0.0)	6.129 (1.0)
15	1119(5.6)	10.961(2.8)	0.015 (0.0)	1.369 (0.6)
16	1088(6.2)	2.285(2.2)	0.103 (0.0)	9.578 (1.5)
17	931(16.2)	4.037(1.2)	0.015 (0.0)	2.182 (1.0)
18	881(9.4)	1.034(0.7)	0.023 (0.0)	4.847 (2.8)
19	828(8.6)	0.983(0.5)	0.121 (0.0)	34.538 (5.0)

[a] standard deviation in parentheses; [b] given in units of cm^{-1}; [c] given in units of $km \cdot mol^{-1}$; [d] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ $[\times 10^{-60}]$; [e] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ $[\times 10^{-57}]$.

Table S2: Averaged Wavenumbers and Associated IR and Raman Intensities of Acetone in Acetone Modelled by the PE Model[a]

Mode	Wavenumber[b]	IR[c]	Raman ($\lambda_0 = 0$ nm)[d]	Raman ($\lambda_0 = 514.5$ nm)[e]
1	3181(6.2)	5.338(1.5)	48.406(2.6)	37.958(2.0)
2	3173(4.1)	8.960(2.6)	43.443(2.6)	34.700(2.1)
3	3130(6.1)	5.638(1.7)	56.105(7.8)	47.970(6.9)
4	3121(3.7)	1.685(1.0)	24.196(9.5)	21.001(8.1)
5	3060(4.7)	1.623(1.1)	190.095(20.8)	176.771(19.7)
6	3053(2.9)	0.479(0.5)	19.731(20.2)	18.842(19.0)
7	1806(6.0)	247.751(8.1)	1.567(0.1)	19.581(0.8)
8	1481(3.5)	24.251(1.8)	0.059(0.0)	1.401(1.2)
9	1465(3.4)	22.735(10.2)	0.878(0.1)	22.294(3.1)
10	1455(3.0)	9.205(8.5)	0.591(0.2)	15.329(5.7)
11	1451(1.8)	8.889(7.5)	0.446(0.2)	11.818(6.4)
12	1389(2.7)	91.660(2.9)	0.043(0.0)	1.346(0.3)
13	1378(2.6)	26.580(3.2)	0.099(0.0)	3.715(0.6)
14	1254(2.9)	63.157(3.6)	0.134(0.0)	8.005(0.6)
15	1116(2.5)	5.982(0.8)	0.009(0.0)	0.881(0.2)
16	1083(3.5)	0.132(0.1)	0.094(0.0)	8.827(0.7)
17	906(6.7)	5.707(0.8)	0.024(0.0)	4.211(0.9)
18	876(4.4)	0.543(0.4)	0.015(0.0)	3.193(1.2)
19	812(4.1)	1.547(0.3)	0.151(0.0)	47.481(3.1)

[a] standard deviation in parentheses; [b] given in units of cm^{-1}; [c] given in units of $km \cdot mol^{-1}$; [d] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ $[\times 10^{-60}]$; [e] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ $[\times 10^{-57}]$.

Table S3: Averaged Wavenumbers and Associated IR and Raman Intensities of Acetone in Chloroform Modelled by the PE Model[a]

Mode	Wavenumber[b]	IR[c]	Raman ($\lambda_0 = 0$ nm)[d]	Raman ($\lambda_0 = 514.5$ nm)[e]
1	3181(6.3)	5.420(1.7)	48.503(2.8)	38.025(2.2)
2	3173(4.2)	8.721(2.5)	43.437(2.5)	34.684(2.0)
3	3123(6.3)	5.608(1.6)	55.379(7.0)	47.352(6.2)
4	3121(3.8)	1.819(1.1)	25.008(8.8)	21.702(7.5)
5	3060(4.8)	1.654(1.1)	190.269(21.0)	176.977(19.9)
6	3053(2.9)	0.462(0.4)	19.545(20.7)	18.658(19.4)
7	1806(6.1)	247.500(8.0)	1.568(0.1)	19.591(0.8)
8	1481(3.3)	24.193(1.9)	0.061(0.0)	1.431(1.2)
9	1466(3.3)	21.915(9.7)	0.868(0.1)	22.023(3.1)
10	1456(3.1)	9.627(8.0)	0.593(0.2)	15.365(5.4)
11	1451(1.8)	9.145(7.4)	0.451(0.2)	11.959(6.2)
12	1389(2.8)	91.587(3.0)	0.043(0.0)	1.328(0.4)
13	1378(2.7)	26.531(3.0)	0.099(0.0)	3.712(0.6)
14	1254(3.0)	63.353(3.7)	0.134(0.0)	7.977(0.5)
15	1116(2.4)	6.007(0.7)	0.009(0.0)	0.896(0.2)
16	1082(3.1)	0.131(0.1)	0.094(0.0)	8.819(0.7)
17	906(6.8)	5.812(0.8)	0.024(0.0)	4.245(0.9)
18	876(4.5)	0.524(0.4)	0.016(0.0)	3.294(1.2)
19	812(3.8)	1.555(0.3)	0.151(0.0)	47.513(3.1)

[a] standard deviation in parentheses; [b] given in units of cm^{-1}; [c] given in units of $km \cdot mol^{-1}$; [d] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ $[\times 10^{-60}]$; [e] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ $[\times 10^{-57}]$.

Table S4: Wavenumbers and Associated IR and Raman Intensities of Acetone in Vacuum

Mode	Wavenumber[a]	IR[b]	Raman ($\lambda_0 = 0$ nm)[c]	Raman ($\lambda_0 = 514.5$ nm)[d]
1	3175	4.496	45.151	35.311
2	3173	9.558	35.165	28.219
3	3123	11.851	60.138	51.772
4	3116	0.000	6.743	6.144
5	3056	4.817	177.032	164.416
6	3050	0.556	0.646	0.813
7	1824	200.895	1.390	16.493
8	1481	23.805	0.001	0.006
9	1462	31.168	0.922	23.635
10	1455	0.000	0.737	18.653
11	1449	0.737	0.076	2.149
12	1388	80.149	0.019	0.501
13	1377	21.572	0.073	2.842
14	1248	63.346	0.117	7.148
15	1116	3.473	0.004	0.427
16	1080	0.041	0.069	6.421
17	888	5.988	0.026	5.167
18	874	0.000	0.008	1.400
19	808	1.630	0.147	47.177

[a] given in units of cm^{-1}; [b] given in units of $km \cdot mol^{-1}$; [c] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-60}$]; [d] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-57}$].

Table S5: Wavenumbers and Associated IR and Raman Intensities of Acetone in Water Modelled by the PCM

Mode	Wavenumber[a]	IR[b]	Raman ($\lambda_0 = 0$ nm)[c]	Raman ($\lambda_0 = 514.5$ nm)[d]
1	3173	8.333	81.628	43.498
2	3171	20.217	69.652	35.484
3	3129	8.581	129.258	65.803
4	3121	0.000	10.321	7.343
5	3060	1.185	376.677	209.725
6	3052	0.653	0.490	0.846
7	1779	375.709	2.584	20.111
8	1463	33.958	0.000	0.006
9	1446	59.376	1.685	27.546
10	1440	0.000	1.157	20.731
11	1438	3.565	0.179	2.873
12	1382	121.601	0.048	1.007
13	1368	50.516	0.294	6.463
14	1248	81.062	0.192	7.728
15	1113	9.859	0.024	0.776
16	1078	0.058	0.159	9.117
17	894	7.404	0.019	2.819
18	871	0.000	0.011	1.339
19	812	2.403	0.212	49.909

[a] given in units of cm^{-1}; [b] given in units of $km \cdot mol^{-1}$; [c] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-60}$]; [d] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-57}$].

Table S6: Wavenumbers and Associated IR and Raman Intensities of Acetone in Acetone Modelled by the PCM

Mode	Wavenumber[a]	IR[b]	Raman ($\lambda_0 = 0$ nm)[c]	Raman ($\lambda_0 = 514.5$ nm)[d]
1	3173	7.912	79.175	44.125
2	3171	19.369	66.795	35.899
3	3129	8.750	123.709	66.988
4	3119	0.000	10.104	7.400
5	3060	1.391	359.405	212.864
6	3051	0.586	0.580	0.910
7	1781	362.994	2.512	20.651
8	1463	33.059	0.000	0.007
9	1447	57.434	1.655	28.468
10	1441	0.000	1.132	21.157
11	1438	3.497	0.169	2.903
12	1382	118.933	0.046	0.992
13	1367	47.286	0.264	6.269
14	1248	80.024	0.187	7.920
15	1113	9.163	0.023	0.797
16	1077	0.058	0.147	9.169
17	894	7.280	0.020	3.015
18	871	0.000	0.011	1.383
19	811	2.297	0.209	50.913

[a] given in units of cm^{-1}; [b] given in units of $km \cdot mol^{-1}$; [c] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-60}$]; [d] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-57}$].

Table S7: Wavenumbers and Associated IR and Raman Intensities of Acetone in Chloroform Modelled by the PCM

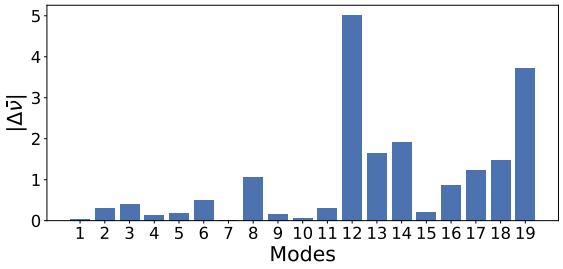
Mode	Wavenumber[a]	IR[b]	Raman ($\lambda_0 = 0$ nm)[c]	Raman ($\lambda_0 = 514.5$ nm)[d]
1	3173	6.571	69.779	46.116
2	3171	16.264	56.795	37.421
3	3127	9.443	103.910	71.013
4	3116	0.000	9.300	7.604
5	3059	2.252	299.749	223.322
6	3049	0.460	0.795	1.056
7	1792	315.271	2.212	22.504
8	1465	30.509	0.000	0.007
9	1450	48.920	1.479	31.352
10	1445	0.000	1.035	22.684
11	1439	2.602	0.137	3.109
12	1382	109.084	0.035	0.881
13	1368	38.450	0.191	5.851
14	1247	75.909	0.168	8.579
15	1112	6.796	0.015	0.857
16	1076	0.040	0.118	9.367
17	892	6.859	0.024	3.868
18	871	0.000	0.010	1.558
19	809	1.919	0.196	54.698

[a] given in units of cm^{-1}; [b] given in units of $km \cdot mol^{-1}$; [c] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-60}$]; [d] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-57}$].

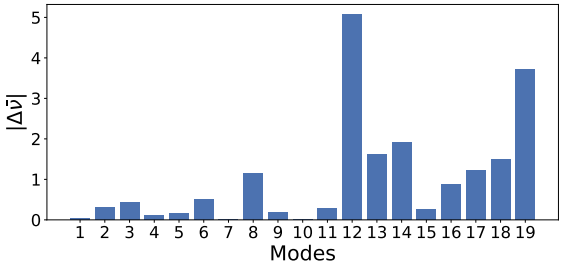
Table S8: Wavenumbers and Associated IR and Raman Intensities of Acetone in Vacuum calculated with Gaussian[S1]

Mode	Wavenumber[a]	IR[b]	Raman ($\lambda_0 = 0$ nm)[c]	Raman ($\lambda_0 = 514.5$ nm)[d]
1	3174	4.524	45.283	35.363
2	3173	9.559	35.149	28.147
3	3122	11.881	60.095	51.674
4	3116	0.000	6.742	6.130
5	3056	4.886	176.814	163.890
6	3049	0.562	0.649	0.814
7	1822	199.965	1.376	16.380
8	1479	23.771	0.001	0.005
9	1461	30.895	0.916	23.519
10	1455	0.000	0.737	18.627
11	1449	0.640	0.077	2.174
12	1384	82.688	0.015	0.398
13	1375	22.729	0.081	3.147
14	1245	60.844	0.117	7.169
15	1115	3.476	0.004	0.428
16	1079	0.036	0.069	6.416
17	887	5.888	0.026	5.168
18	873	0.000	0.008	1.385
19	804	1.520	0.145	47.607

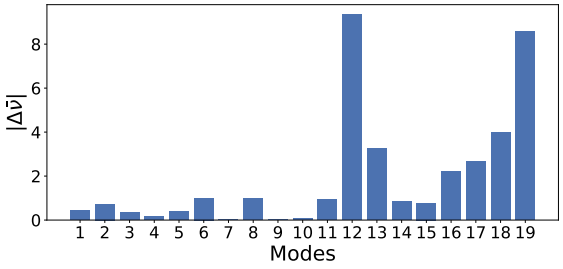
[a] given in units of cm^{-1}; [b] given in units of $km \cdot mol^{-1}$; [c] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-60}$]; [d] given in units of $C^4 \cdot s^2 \cdot J^{-1} \cdot m^{-2} \cdot kg^{-1}$ [$\times 10^{-57}$].



(a) Chloroform



(b) Acetonitrile



(c) Water

Figure S1: Absolute differences (in cm^{-1}) between average wavenumbers calculated with and without projecting out translation and rotation.

References

(S1) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Men-

nucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16 Revision A.03. 2016; Gaussian Inc. Wallingford CT.

Paper II

Analytic High-order Geometric Derivatives with Polarizable Embedding in a Response Theory Framework

K. O. H. M. Dundas, S. Reine, B. Gao, K. Ruud,
J. M. H. Olsen and M. Ringholm
(manuscript)

Analytic High-Order Geometric Derivatives with Polarizable Embedding in a Response Theory Framework

Karen O. H. M. Dundas, Simen S. Reine, Bin Gao,
Kenneth Ruud, Jógvan Magnus Haugaard Olsen, Magnus Ringholm

December 24, 2020

Abstract

We present an approach for the analytic calculation of arbitrary-order geometric derivatives of energies and electric dipole polarization properties in the polarizable embedding (PE) framework. The approach builds on a flexible response theory framework that can handle both frequency-dependent perturbations and perturbation-dependent basis sets at the self-consistent field level of theory, and in which the effects of the environment are described by the PE model. As a proof of principle, the approach is applied to calculate second-order vibrational perturbation theory corrections to vibrational energy levels and infrared intensities of acetonitrile in solution as well as for the $H_2O@C_{60}$ host-cage system.

1 Introduction

Spectroscopic data can be analyzed in terms of response functions.[1] They describe the change in the energy or an observable property of the system, such as the molecular dipole moment, when subjected to different kinds of perturbations such as electromagnetic fields, displacements of the atomic nuclei, or small terms in the Hamiltonian that are otherwise neglected when optimizing the electronic energy and molecular density. By analyzing the pole structure of the response function, excitation energies of the molecular system can be identified, and from a residue analysis of the response functions one can extract transition probabilities between different states in the molecular system.[1;2] The ability to calculate response properties by evaluating response functions for a wide range of perturbations therefore translates into a corresponding ability to calculate observable spectroscopic data, and thereby support or predict experimental observations.

The use of response theory in quantum chemistry for the calculation of properties of spectroscopic relevance has a long history, and historical accounts have been given in recent reviews.[2;3] One can base such calculations on finite-difference schemes or analytic techniques, or a combination of both.[4;5;6]. Our group has had a particular interest in the development of analytic approaches based on a flexible and open-ended response theory and a recursive algorithmic approach for its implementation[7;8;9]. This framework can in principle calculate any response property at the Hartree–Fock (HF) or density functional theory (DFT) level of theory given that necessary contributions, such as differentiated one- and two-electron integrals, exchange–correlation contributions, and solutions of so-called response equations, are available through external modules. Our approach is implemented in the program library OpenRSP with much of the same generality as the underlying theory.[10]

Response properties only involving perturbations arising from displacements of the atomic nuclei—so-called geometric perturbations—have several important areas of application. For example, the molecular Hessian can be used for the analysis of vibrational energy levels and normal modes in the harmonic approximation for the vibrational wavefunctions.[11] Using third- and semi-diagonal fourth-order geometric derivatives of the energy, anharmonic corrections to the harmonic vibrational en-

ergy levels can be calculated using vibrational perturbation theory (VPT) truncated at second order (VPT2),[12;13;14] which can be further refined into the generalized VPT2 (GVPT2) method[4;15] employed in this work. Higher-order VPT techniques which involve even higher-order geometric derivatives have also been investigated.[6]

Properties that in addition to geometric perturbations also involve perturbations related to interactions with an electromagnetic field, whose coupling is represented in the form of electric and magnetic multipole moments, can be used to describe scattering and absorption cross sections in spectroscopies that probe molecular vibrations. This includes infrared (IR) and Raman spectroscopy, or their chiroptical analogues vibrational circular dichroism (VCD) and Raman optical activity (ROA) spectroscopy respectively.[1]

IR and Raman spectroscopy can provide important information about molecular structure and bonding.[11;16] The IR and Raman absorption and scattering cross sections are related to the geometric derivatives of the molecular dipole moment and polarizability, respectively.[11] In the commonly employed double-harmonic approximation,[11] only the first-order geometric derivatives of these properties are needed, together with harmonic oscillator wavefunctions for describing the vibrational motion. VPT2 can be used to go beyond the double-harmonic approximation for spectral intensities,[15] and involve second- and third-order derivatives of the relevant polarization properties and third- and semi-diagonal fourth-order geometric derivatives of the energy. In addition to providing corrections to the single-quantum transition features described by the harmonic approximation, these expressions also give the leading-order contributions to two-quantum transitions (overtone and combination bands) and even three-quantum[17] features in the spectra.

We have previously presented fully analytic calculations of anharmonic IR and Raman spectra *in vacuo*.[18] However, most experiments are carried out in solution or on larger molecular complexes, and in these cases the effects of the surroundings on the molecule or subsystem of interest cannot *a priori* be neglected. At the same time, it is in many cases computationally intractable to treat the entire system at a fully quantum-mechanical level of theory. A common approach that strikes a good balance between computational cost and accuracy is to use focused models that treat the smaller region of interest in a fully quantum-mechanical manner (this region will henceforth be referred to as the QM region) and account for the effects of the surrounding environment by a less computationally demanding level of theory. A large number of different embedding methods have been developed, and we refer to a number of reviews on this subject.[19;20;21]

In continuum models, the environment is described by a polarizable dielectric continuous medium interacting with the solvated molecule, which is placed within a cavity of the surrounding medium. The polarizable continuum model (PCM)[20] is an example of one such method that has had considerable success owing to its reasonable computational cost and in general reliable results. Our open-ended response-theory formalism has been interfaced to a PCM module by Di Remigio *et al.*[22] for electric dipole polarization properties and their single residues, and was applied to the fully analytic calculation of multi-photon absorption (MPA) cross sections up to fifth order. The analytic calculation of geometric derivatives with PCM has been presented for first- and second-order derivatives by Cappelli *et al.*[23].

One limitation of continuum models is that they are not able to model specific interactions between the environment and the QM region, such as hydrogen bonds, whose effects may constitute important contributions in the description of molecular vibrations.[24;25] Atomistic modelling provides an improved description of such interactions, and this has led to the development of discrete embedding methods that retain the atomic structure of the environment while reducing the computational cost significantly compared to a full QM calculation. A well-established atomistic model is the quantum mechanics/molecular mechanics (QM/MM)[26] class of methods, where the central molecular region is treated with QM methods, while the environment is described by classical (molecular mechanics) force fields.

Different QM/MM approaches have been presented to model the environment and its interaction with the QM region,[27;28;21] differing in the types of interactions that are accounted for. With elec-

trostatic embedding methods, only the QM region is polarized by the environment typically through partial point charges in the environment. In polarized embedding methods, also the environment is polarized by the QM region, e.g., through the addition of atom-centered polarizabilities in the environment.

One way to model a polarizable environment is through fluctuating charges (FQ),[29;30] which, together with its extension to fluctuating dipoles, has been used for a variety of purposes.[31;32;33;34;35;36] Electrostatic and polarization interactions are described through a set of atom-centered charges (or dipoles) that depend on the atomic parameters of electronegativity and hardness. Despite the advantages of the FQ method, such as its accuracy and efficiency,[37] the parameterization of the environment is one of its key limitations, as it is based on fitting and therefore dependent on having a parameterization of the environment prior to calculations.

In the polarizable embedding (PE) method,[38;39] the parameterization problem is addressed by defining each environmental molecule as a fragment, and performing explicit QM calculations on each of these environmental fragments, using the LoProp partitioning approach to derive atom-centered multipoles and polarizabilities.[40] This makes the PE method able to be straightforwardly applied for any solvent, and even biomolecular system when used in combination with molecular fractionation with conjugate caps.[41;42] Another advantage of the PE method over the FQ method is that it can account for any anisotropy of the environment through higher-order multipoles and polarizabilities, though these effects are in general found to be small after averaging when an MD sampling is used as the basis for the calculations.[37]

The PE method is by now a well-established method that has been employed to a wide range of applications. Over the years, it has been used to calculate both electric,[43] optical[44;45] and magnetic[37;46] properties, and recently also harmonic vibrational properties and IR/Raman spectra[47] from the molecular Hessian and dipole and polarizability gradients. In this work, we present an open-ended formalism for the analytic calculation of arbitrary-order geometric derivatives of the energy (in the form of a quasienergy Lagrangian) and electric dipole polarization properties within the PE model. This allows for PE effects to be included fully analytically in properties used for the calculation of vibrational spectra including anharmonic corrections.

We here demonstrate this functionality with the calculation of anharmonic corrections to the vibrational energy levels and IR intensities of the endohedral fullerene system $H_2O@C_{60}$ and of acetonitrile in different solvents. This requires calculation of energy derivatives up to fourth order, involving both pure geometric perturbations and a mixture of geometric and electric-dipole perturbations.[15] In the present work, the underlying level of electronic structure theory is Hartree–Fock in lieu of further generalization of the local implementation of exchange–correlation functionality. This suffices as a proof-of-concept demonstration of the general theory and methodology.

The rest of this paper is organized as follows: In Section 2, we present the theory underlying our approach for the calculation of arbitrary-order geometric derivatives of PE effects. In Section 3, we present some more details of our implementation together with the computational details for our calculations. In Section 4 we present and discuss our results, before we in Section 5 give some concluding remarks and an outlook.

2 Theory

Concerning methodology related to the main topic of the present work: The calculation of molecular response properties for molecules in vacuum, using a density-matrix-based quasienergy formulation of Hartree–Fock or Kohn–Sham density functional response theory and recursive programming, has been described elsewhere.[7;8] We also refer to earlier work for an introduction to the PE model used in this work,[38;48;43] including our recent work on IR and Raman spectra calculated with this model,[47] and we will in this section only provide a brief recapitulation of the methodology in order to provide the basic theoretical foundation for the present work. We also summarize the main elements of vibrational perturbation theory used in this work.

2.1 Polarizable embedding

In the PE model, the system of interest is divided into a region to be treated at a quantum-mechanical level of theory (the QM region) and a molecular environment described at a lower level of theory. The environment is then further divided into partitions called fragments, where each fragment is typically, but not necessarily, one molecule in the environment. Within each fragment, we define a set of sites, positioned on the coordinates of each of the atoms.

We express the total energy E of our system as the sum of the energy of the QM region, which we in this work will treat at the Hartree–Fock level of theory and denote by E_{QM}, and the interaction energy between the QM region and the environment, the embedding energy, E_{PE},

$$E(\mathbf{D}) = E_{\mathrm{QM}}(\mathbf{D}) + E_{\mathrm{PE}}(\mathbf{D}), \tag{1}$$

where $\mathbf{D}$ is the density matrix of the QM region in the atomic-orbital basis. Limiting ourselves to Hartree–Fock theory and perturbations that do not simultaneously have an implicit dependence in the basis set and depend on the frequency of the applied field, we can write the QM energy $E_{\mathrm{QM}}(\mathbf{D})$ as[7]

$$E_{\mathrm{QM}}(\mathbf{D}) \overset{\mathrm{Tr}}{=} \frac{1}{2}\mathbf{G}(\mathbf{D})\mathbf{D} + \mathbf{h}\mathbf{D} + \mathbf{V}^t\mathbf{D} + h_{\mathrm{nuc}}, \tag{2}$$

where $\overset{\mathrm{Tr}}{=}$ indicates that the trace of all terms on the right-hand side should be taken, and where $\mathbf{G}$ represents the electronic Coulomb and exchange contributions, $\mathbf{h}$ is the one-electron contributions (kinetic energy and nuclear attraction), $\mathbf{V}^t$ is the matrix representation of coupling to an external field, and h_{nuc} is the nuclear repulsion energy.

The PE energy $E_{\mathrm{PE}}(\mathbf{D})$ is given by

$$E_{\mathrm{PE}}(\mathbf{D}) = E_{\mathrm{PE}}^{\mathrm{es}}(\mathbf{D}) + E_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D}) + E_{\mathrm{PE}}^{\mathrm{rep}} + E_{\mathrm{PE}}^{\mathrm{disp}}, \tag{3}$$

where $E_{\mathrm{PE}}^{\mathrm{es}}(\mathbf{D})$ accounts for the electrostatic, $E_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})$ for the polarization, $E_{\mathrm{PE}}^{\mathrm{rep}}$ for the non-electrostatic repulsion and $E_{\mathrm{PE}}^{\mathrm{disp}}$ for the dispersion contribution to the embedding energy.

The electrostatic contribution $E_{\mathrm{PE}}^{\mathrm{es}}$ originates from the static interactions between the electrons and nuclei of the quantum system with the permanent multipoles of the environment

$$\begin{aligned} E_{\mathrm{PE}}^{\mathrm{es}}(\mathbf{D}) = &- \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{\mu\nu} t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s) D_{\mu\nu} \\ &+ \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{n}^{N_{\mathrm{nuc}}} T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n) Z_n \\ &\overset{\mathrm{Tr}}{=} \mathbf{h}_{\mathrm{PE}}^{\mathrm{es}}\mathbf{D} + h_{\mathrm{PE}}^{\mathrm{es}}, \end{aligned} \tag{4}$$

where the summation index a runs over the N_{frag} number of environmental fragments, and where we introduced a notation that groups the terms into first- or zeroth-order dependence on the density matrix, as $\mathbf{h}_{\mathrm{PE}}^{\mathrm{es}}\mathbf{D}$ and $h_{\mathrm{PE}}^{\mathrm{es}}$, respectively. The summation index s runs over the S_a sites within fragment a. The absolute value of the multi-index α runs to the maximum multipole order K_s of site s. For a further description of these multi-indices, we refer to Refs. 44 and 47. The summation indices μ and ν run over the atomic orbitals in the QM region, with $D_{\mu\nu}$ being the $\mu\nu$ element of the density matrix. The summation index n runs over the N_{nuc} nuclei in the QM region and Z_n represents the nuclear charge of nucleus n. Each site s has a permanent multipole moment $M_s^{[\alpha]}$ of order $|\alpha|$, and $t_{\mu\nu}^{[\alpha]}(\mathbf{r}_s)$ is the one-electron electrostatic potential integral defined by

$$t^{[\alpha]}_{\mu\nu}(\mathbf{r}_s) = \int \chi_\mu(\mathbf{r};\mathbf{r}_n)\partial^\alpha_{\mathbf{r}} \frac{1}{|\mathbf{r}-\mathbf{r}_s|}\chi_\nu(\mathbf{r};\mathbf{r}_n)d\mathbf{r}. \tag{5}$$

where $\chi_\mu(\mathbf{r};\mathbf{r}_n)$ is atomic orbital μ, which has a parametric dependence on the nuclei. The interaction tensor between site s and nucleus n, $T^{[\alpha]}(\mathbf{r}_s,\mathbf{r}_n)$, can be expressed in the general form

$$T^{[\alpha]}(\mathbf{r}_i,\mathbf{r}_j) = \partial^\alpha_{\mathbf{r}_j} \frac{1}{|\mathbf{r}_j-\mathbf{r}_i|}. \tag{6}$$

As in our paper describing the implementation of harmonic IR and Raman spectra,[47] we formulate the induction energy in a general manner with respect to multipole order, so that

$$\begin{aligned} E^{\mathrm{ind}}_{\mathrm{PE}}(\mathbf{D}) =& \frac{1}{2}\bar{\mathcal{M}}^T\mathcal{V} \\ =& -\frac{1}{2}\sum_a^{N_{\mathrm{frag}}}\sum_{s\in a}^{S_a}\sum_b^{N_{\mathrm{frag}}}\sum_{t\in b}^{S_b}\sum_{|\alpha|=0}^{K_s}\sum_{|\beta|=0}^{K_t}\frac{1}{\alpha!\beta!}V^{[\beta]}(\mathbf{D},\mathbf{r}_t)B^{[\beta,\alpha]}_{ts}V^{[\alpha]}(\mathbf{D},\mathbf{r}_s), \end{aligned} \tag{7}$$

where the summation index b runs over the number of fragments N_{frag} in the environment, t over the S_b sites in fragment b, and $|\beta|$ runs over all the multi-indices up to the K_t maximum multipole-order of site t. The matrix $\bar{\mathcal{M}}$ contains a list of polytensors containing all the induced multipoles, one for each site, and $\mathcal{V}$ similarly contains the potentials. The induced multipoles are given by $\bar{\mathcal{M}}^T = -\mathcal{A}^{-1}\mathcal{V}$. The interaction matrix $B^{[\beta,\alpha]}_{ts}$ is an element of the matrix $\mathcal{B} = \mathcal{A}^{-1}$, where

$$\mathcal{A} = \begin{pmatrix} \mathbf{P}_1^{-1} & \mathbf{T}_{12} & \cdots & \mathbf{T}_{1S} \\ \mathbf{T}_{21} & \mathbf{P}_2^{-1} & \cdots & \mathbf{T}_{2S} \\ \vdots & & \ddots & \vdots \\ \mathbf{T}_{S1} & \cdots & \cdots & \mathbf{P}_S^{-1} \end{pmatrix}, \tag{8}$$

where $\mathbf{P}_s^{-1}$ is the inverse of the polytensor of polarizabilities on site s, $\mathbf{T}_{ts}$ is the polytensor of interaction tensors between sites t and s, and S is the total number of sites in the environment. The potential experienced on each site has its origin in either the electrons or the nuclei in the QM region, or the permanent multipoles on sites in the other fragments, and can therefore be expressed as

$$V^{[\alpha]}(\mathbf{D},\mathbf{r}_s) = V^{\mathrm{el},[\alpha]}(\mathbf{D},\mathbf{r}_s) + V^{\mathrm{nuc},[\alpha]}(\mathbf{r}_s) + V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s), \tag{9}$$

where the electronic potential is given by

$$V^{\mathrm{el},[\alpha]}(\mathbf{D},\mathbf{r}_s) = -\sum_{\mu\nu} t^{[\alpha]}_{\mu\nu}(\mathbf{r}_s)D_{\mu\nu}, \tag{10}$$

the nuclear potential by

$$V^{\mathrm{nuc},[\alpha]}(\mathbf{r}_s) = \sum_n^{N_{\mathrm{nuc}}} T^{[\alpha]}(\mathbf{r}_n,\mathbf{r}_s)Z_n, \tag{11}$$

and the potential from the multipoles in the other fragments by

$$V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s) = \sum_{d\neq a}^{N_{\mathrm{frag}}} \sum_{u\in d}^{S_d} \sum_{|\gamma|=0}^{K_u} \frac{(-1)^{|\gamma|}}{\gamma!} T^{[\gamma+\alpha]}(\mathbf{r}_s, \mathbf{r}_u) M_u^{[\gamma]}, \tag{12}$$

where the summation index d runs over the number of fragments N_{frag} in the environment, except fragment a, u runs over all the sites in fragment d and $|\gamma|$ runs over all the possible multi-indices up to the maximum multipole-order of K_u of site u. Inserting this into Eq. (7), we obtain an expression for the induction energy arranged by order of density matrix dependence as

$$\begin{aligned}
E_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D}) = &-\frac{1}{2} \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha!\beta!} \\
&\times \Big[V^{\mathrm{el},[\beta]}(\mathbf{D}, \mathbf{r}_t) B_{ts}^{[\beta,\alpha]} V^{\mathrm{el},[\alpha]}(\mathbf{D}, \mathbf{r}_s) \\
&+ 2\left(V^{\mathrm{nuc},[\beta]}(\mathbf{r}_t) + V^{\mathrm{mul},[\beta]}(\mathbf{r}_t)\right) B_{ts}^{[\beta,\alpha]} V^{\mathrm{el},[\alpha]}(\mathbf{D}, \mathbf{r}_s) \\
&+ \left(V^{\mathrm{nuc},[\beta]}(\mathbf{r}_t) + V^{\mathrm{mul},[\beta]}(\mathbf{r}_t)\right) B_{ts}^{[\beta,\alpha]} \left(V^{\mathrm{nuc},[\alpha]}(\mathbf{r}_s) + V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s)\right) \Big] \\
\overset{\mathrm{Tr}}{=} &\frac{1}{2} \mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D}) \mathbf{D} + \mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}} \mathbf{D} + h_{\mathrm{PE}}^{\mathrm{ind}},
\end{aligned} \tag{13}$$

where we in the last equality introduced a notation for grouping these contributions into orders of dependence on $\mathbf{D}$, where $\frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D}$, $\mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}}\mathbf{D}$ and $h_{\mathrm{PE}}^{\mathrm{ind}}$ collect terms with respectively second-, first-, and zeroth-order such dependence.

Finally, dispersion and repulsion terms will in this work be approximated by a 12-6 Lennard-Jones potential[43] denoted by $E_{\mathrm{PE}}^{\mathrm{LJ}}$, so that in this work, $E_{\mathrm{PE}}^{\mathrm{rep}} + E_{\mathrm{PE}}^{\mathrm{disp}} = E_{\mathrm{PE}}^{\mathrm{LJ}}$, where

$$E_{\mathrm{PE}}^{\mathrm{LJ}} = 4 \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{n}^{N_{nuc}} \varepsilon_{sn} \left[\sigma_{sn}^{12} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \right)^{12} - \sigma_{sn}^{6} \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|} \right)^{6} \right], \tag{14}$$

where $\varepsilon_{sn} = \sqrt{\varepsilon_s \varepsilon_n}$ and $\sigma_{sn} = \frac{1}{2}(\sigma_s + \sigma_n)$, where ε_s and ε_n are potential well depths and σ_s and σ_n are the distances beyond which the potential between the particles is zero, and where we note in passing that $E_{\mathrm{PE}}^{\mathrm{LJ}}$ is independent of the density matrix.

Finally, several of the PE interaction energy contributions presented above also lead to contributions to the Fock matrix $\mathbf{F}(\mathbf{D}) = \frac{\partial E(\mathbf{D})}{\partial \mathbf{D}^T}$. As for the energy in Eq. (1), we can consider the Fock matrix as a sum $\mathbf{F}(\mathbf{D}) = \mathbf{F}_{\mathrm{PE}}(\mathbf{D}) + \mathbf{F}_{\mathrm{QM}}(\mathbf{D})$, where, based on Eq. (2),

$$\mathbf{F}_{\mathrm{QM}}(\mathbf{D}) = \mathbf{G}(\mathbf{D}) + \mathbf{h} + \mathbf{V}^t, \tag{15}$$

and where the PE interaction Fock matrix $\mathbf{F}_{\mathrm{PE}}$ is given by

$$\mathbf{F}_{\mathrm{PE}}(\mathbf{D}) = \mathbf{G}_{\mathrm{ind}}(\mathbf{D}) + \mathbf{h}_{\mathrm{ind}} + \mathbf{h}_{\mathrm{es}}. \tag{16}$$

2.2 Geometric derivatives of polarizable embedding contributions

We will in this section develop expressions for the general-order geometric derivatives of the PE interaction terms presented in the preceding section, including combination derivatives that both involve geometric and electric dipole perturbations. Geometric derivatives are expressed in a Cartesian basis, and for a molecule with N atoms, there are $3N$ Cartesian components of the nuclear positions.

We note that many of the PE interaction terms contain products of terms that depend on nuclear positions: The highest-order dependence occurs in the $\mathrm{Tr}\frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D}$ term of the induction energy,

which is of fourth order in this regard. We therefore begin by considering some generalized differentiation formulae for product terms which can also be applied, where relevant, for differentiation with respect to electric dipole perturbations.

For a two-term product $(u \cdot v)$, a generalized differentiation of order n can be expressed through the differentiation operator $d^{(n)}$, using the general Leibniz rule[49], as

$$d^{(n)}(u \cdot v) = \sum_{k=0}^{n} \binom{n}{k} d^{(n-k)}u \cdot d^{(k)}v. \tag{17}$$

By substituting $v \leftarrow gf$ and $u \leftarrow hl$, we can extend this formulation to cases with products of three or four terms, so that

$$d^{(n)}(u \cdot g \cdot f) = \sum_{k=0}^{n} \binom{n}{k} d^{(n-k)}u \cdot \sum_{j=0}^{k} \binom{k}{j} d^{(k-j)}g \cdot d^{(j)}f \tag{18}$$

and

$$\begin{aligned} d^{(n)}(h \cdot l \cdot g \cdot f) = &\sum_{k=0}^{n} \binom{n}{k} \sum_{i=0}^{n-k} \binom{n-k}{i} d^{(n-k-i)}h \cdot d^{(i)}l \\ &\times \sum_{j=0}^{k} \binom{k}{j} d^{(k-j)}g \cdot d^{(j)}f, \end{aligned} \tag{19}$$

respectively.

Using Eqs. (17)-(19), we now consider the differentiation of the individual terms in $E_{\mathrm{PE}}(\mathbf{D})$. Starting with the density-matrix first-order electrostatic term, $\mathrm{Tr}\,\mathbf{h}^{\mathrm{es}}_{\mathrm{PE}}\mathbf{D}$ in Eq. (4), we see that it is a product of two terms that depend on the nuclear coordinates; $t^{[\alpha]}_{\mu\nu}(\mathbf{r}_s)$, through the basis sets, and $D_{\mu\nu}$. Defining $d_g^{(|\gamma|)}$ as a $|\gamma|$-fold geometric perturbation of multi-index γ and using Eq. (17), we get

$$\begin{aligned} d_g^{(|\gamma|)}\mathrm{Tr}\mathbf{h}^{\mathrm{es}}_{\mathrm{PE}}\mathbf{D} = &-\sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{\mu\nu} \\ &\times \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} d_g^{(|\gamma|-|\delta|)} t^{[\alpha]}_{\mu\nu}(\mathbf{r}_s) \cdot d_g^{(|\delta|)} D_{\mu\nu}. \end{aligned} \tag{20}$$

The only terms in the PE contributions that are dependent on electric perturbations are the density matrices. Therefore, defining $d_f^{(m)}$ as an m'th order electric perturbation, an m-fold electric derivative of $\mathrm{Tr}\mathbf{h}^{\mathrm{es}}_{\mathrm{PE}}\mathbf{D}$ becomes

$$d_f^{(m)}\mathrm{Tr}\mathbf{h}^{\mathrm{es}}_{\mathrm{PE}}\mathbf{D} = -\sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{\mu\nu} t^{[\alpha]}_{\mu\nu}(\mathbf{r}_s) \cdot d_f^{(m)} D_{\mu\nu}, \tag{21}$$

and a mixed geometric and electric derivative becomes

$$\begin{aligned} d_f^{(m)} d_g^{(|\gamma|)}\mathrm{Tr}\mathbf{h}^{\mathrm{es}}_{\mathrm{PE}}\mathbf{D} = &-\sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{\mu\nu} \\ &\times \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} d_g^{(|\gamma|-|\delta|)} t^{[\alpha]}_{\mu\nu}(\mathbf{r}_s) \cdot d_f^{(m)} d_g^{(|\delta|)} D_{\mu\nu}. \end{aligned} \tag{22}$$

The next term in Eq. (4), $h_{\mathrm{PE}}^{\mathrm{es}}$, is simpler because it only contains one term that depends on the nuclear coordinates, $T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n)$, and furthermore depends only on the coordinates of a single nucleus, n, which means that it will vanish upon differentiation with respect to geometric components that do not all belong to the same nucleus. The geometric derivative of $h_{\mathrm{PE}}^{\mathrm{es}}$ is given by

$$d_g^{(|\gamma|)} h_{\mathrm{PE}}^{\mathrm{es}} = \sum_{a=1}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{|\alpha|=0}^{K_s} \frac{(-1)^{|\alpha|}}{\alpha!} M_s^{[\alpha]} \sum_{n}^{N_{\mathrm{nuc}}} d_g^{(|\gamma|)} T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n) Z_n. \tag{23}$$

where $d_g^{[\gamma]} T^{[\alpha]}(\mathbf{r}_s, \mathbf{r}_n) = T^{[\alpha+\gamma]}(\mathbf{r}_s, \mathbf{r}_n)$, as long as the differentiation is with respect only to nucleus n, and zero otherwise. As this term has no dependence on density matrices, it will vanish upon electric perturbation.

Turning now to the induction terms, the term $\mathrm{Tr}\,\frac{1}{2}\mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D}$ is a product of four terms that all depend on nuclear coordinates. Using Eq. (19) gives

$$\begin{aligned}
d_g^{(|\gamma|)} \mathrm{Tr} \frac{1}{2} \mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D} = &-\frac{1}{2} \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha!\beta!} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} \\
&\times (d_g^{(|\gamma|-|\delta|)} V^{\mathrm{el},[\beta]}(\mathbf{D}, \mathbf{r}_t)) B_{ts}^{[\beta,\alpha]} d_g^{(|\delta|)} V^{\mathrm{el},[\alpha]}(\mathbf{D}, \mathbf{r}_s) \\
= &-\frac{1}{2} \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha!\beta!} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} \\
&\times \sum_{\mu\nu} \sum_{|\epsilon|=0}^{|\gamma|-|\delta|} \binom{|\gamma|-|\delta|}{|\epsilon|} (d_g^{(|\gamma|-|\delta|-|\epsilon|)} t_{\mu\nu}^{[\beta]}(\mathbf{r}_t) d_g^{(|\epsilon|)} D_{\mu\nu}) \\
&\times B_{ts}^{[\beta,\alpha]} \sum_{\rho\sigma} \sum_{|\zeta|=0}^{|\delta|} \binom{|\delta|}{|\zeta|} d_g^{(|\delta|-|\zeta|)} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_g^{(|\zeta|)} D_{\rho\sigma}.
\end{aligned} \tag{24}$$

upon a $|\gamma|$-fold geometric perturbation, where the summation indices ρ and σ run over the atomic orbitals in the QM region. Due to its second-order dependence on the density matrix, it becomes

$$\begin{aligned}
d_f^{(m)} \mathrm{Tr} \frac{1}{2} \mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D})\mathbf{D} = &-\frac{1}{2} \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha!\beta!} \\
&\times \sum_{l=0}^{m} \binom{m}{l} \sum_{\mu\nu} t_{\mu\nu}^{[\beta]}(\mathbf{r}_t) d_f^{(m-l)} D_{\mu\nu}) \\
&\times B_{ts}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_f^{(l)} D_{\rho\sigma}.
\end{aligned} \tag{25}$$

upon an m-fold electric perturbation, and the mixed electric dipole and geometric derivatives are

$$
\begin{aligned}
d_f^{(m)} d_g^{(|\gamma|)} \mathrm{Tr} \frac{1}{2} \mathbf{G}_{\mathrm{PE}}^{\mathrm{ind}}(\mathbf{D}) \mathbf{D} = & -\frac{1}{2} \sum_a^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_b^{N_{\mathrm{frag}}} \sum_{t \in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha! \beta!} \sum_{l=0}^{m} \binom{m}{l} \\
& \times \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} \sum_{\mu\nu} \sum_{|\epsilon|=0}^{|\gamma|-|\delta|} \binom{|\gamma|-|\delta|}{|\epsilon|} \\
& \times (d_g^{(|\gamma|-|\delta|-|\epsilon|)} t_{\mu\nu}^{[\beta]}(\mathbf{r}_t) d_f^{(m-l)} d_g^{(|\epsilon|)} D_{\mu\nu}) \\
& \times B_{ts}^{[\beta,\alpha]} \sum_{\rho\sigma} \sum_{|\zeta|=0}^{|\delta|} \binom{|\delta|}{|\zeta|} d_g^{(|\delta|-|\zeta|)} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_f^{(l)} d_g^{(|\zeta|)} D_{\rho\sigma}. \qquad (26)
\end{aligned}
$$

The second term of the induction energy, $\mathrm{Tr}\, \mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}} \mathbf{D}$, is a product of three terms. Using Eq. (18) and that the potential from the multipoles, $V^{\mathrm{mul},[\alpha]}$, always vanishes upon geometric differentiation, we have that

$$
\begin{aligned}
d_g^{(|\gamma|)} \mathrm{Tr} \mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}} \mathbf{D} = & \sum_a^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_b^{N_{\mathrm{frag}}} \sum_{t \in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha! \beta!} \\
& \times \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} d_g^{(|\gamma|-|\delta|)} (V^{\mathrm{nuc},[\beta]}(\mathbf{r}_t) + V^{\mathrm{mul},[\beta]}(\mathbf{r}_t)) B_{ts}^{[\beta,\alpha]} \\
& \times d_g^{(|\delta|)} V^{\mathrm{el},[\alpha]}(\mathbf{D}, \mathbf{r}_s) \\
= & \sum_a^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_b^{N_{\mathrm{frag}}} \sum_{t \in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha! \beta!} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} \\
& \times \left(\sum_n^{N_{\mathrm{nuc}}} d_g^{(|\gamma|-|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_t) Z_n + \delta_{0(|\gamma|-|\delta|)} V^{\mathrm{mul},[\beta]}(\mathbf{r}_t) \right) \\
& \times B_{ts}^{[\beta,\alpha]} \sum_{\rho\sigma} \sum_{j=0}^{|\delta|} \binom{|\delta|}{j} d_g^{(|\delta|-j)} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_g^{(j)} D_{\rho\sigma}. \qquad (27)
\end{aligned}
$$

where $d_g^{(|\beta|)} T^{[\alpha]}(\mathbf{r}_n, \mathbf{r}_s) = (-1)^{|\beta|} T^{[\alpha+\beta]}(\mathbf{r}_n, \mathbf{r}_s)$ if $d_g^{(|\beta|)}$ corresponds to a differentiation with respect to nucelus n only, and zero otherwise.

Due to its first-order dependence on $\mathbf{D}$, electric differentiation results in

$$
\begin{aligned}
d_f^{(m)} \mathrm{Tr} \mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}} \mathbf{D} = & \sum_a^{N_{\mathrm{frag}}} \sum_{s \in a}^{S_a} \sum_b^{N_{\mathrm{frag}}} \sum_{t \in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha! \beta!} \\
& \times \left(\sum_n^{N_{\mathrm{nuc}}} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_t) Z_n + V^{\mathrm{mul},[\beta]}(\mathbf{r}_t) \right) \\
& \times B_{ts}^{[\beta,\alpha]} \sum_{\rho\sigma} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_f^{(m)} D_{\rho\sigma}, \qquad (28)
\end{aligned}
$$

while mixed electric and geometric differentiation is given by

$$
\begin{aligned}
d_f^{(m)} d_g^{(|\gamma|)} \mathrm{Tr}\mathbf{h}_{\mathrm{PE}}^{\mathrm{ind}}\mathbf{D} = & \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha!\beta!} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} \\
& \times \left(\sum_{n}^{N_{\mathrm{nuc}}} d_g^{(|\gamma|-|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_t) Z_n + \delta_{0(|\gamma|-|\delta|)} V^{\mathrm{mul},[\beta]}(\mathbf{r}_t) \right) \\
& \times B_{ts}^{[\beta,\alpha]} \sum_{\rho\sigma} \sum_{j=0}^{|\delta|} \binom{|\delta|}{j} d_g^{(|\delta|-j)} t_{\rho\sigma}^{[\alpha]}(\mathbf{r}_s) d_f^{(m)} d_g^{(j)} D_{\rho\sigma}.
\end{aligned} \tag{29}
$$

All remaining contributions to the PE energy are independent of $\mathbf{D}$, and will therefore vanish upon differentiation with respect to any electric dipole perturbations.

The induction energy term $h_{\mathrm{PE}}^{\mathrm{ind}}$ is a product of two terms that depend on the nuclear coordinates. Using Eq. (17) and following the reasoning above, we can write the geometric derivatives of this term as

$$
\begin{aligned}
d_g^{(|\gamma|)} h_{\mathrm{PE}}^{\mathrm{ind}} = & -\frac{1}{2} \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha!\beta!} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} \\
& \times d_g^{(|\gamma|-|\delta|)} (V^{\mathrm{nuc},[\beta]}(\mathbf{r}_t) + V^{\mathrm{mul},[\beta]}(\mathbf{r}_t)) B_{ts}^{[\beta,\alpha]} \\
& \times d_g^{(|\delta|)} (V^{\mathrm{nuc},[\alpha]}(\mathbf{r}_s) + V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s)) \\
= & -\frac{1}{2} \sum_{a}^{N_{\mathrm{frag}}} \sum_{s\in a}^{S_a} \sum_{b}^{N_{\mathrm{frag}}} \sum_{t\in b}^{S_b} \sum_{|\alpha|=0}^{K_s} \sum_{|\beta|=0}^{K_t} \frac{1}{\alpha!\beta!} \sum_{|\delta|=0}^{|\gamma|} \binom{|\gamma|}{|\delta|} \\
& \times \left(\sum_{n}^{N_{\mathrm{nuc}}} d_g^{(|\gamma|-|\delta|)} T^{[\beta]}(\mathbf{r}_n, \mathbf{r}_t) Z_n + \delta_{0(|\gamma|-|\delta|)} V^{\mathrm{mul},[\beta]}(\mathbf{r}_t) \right) \\
& \times B_{ts}^{[\beta,\alpha]} \left(\sum_{m}^{N_{\mathrm{nuc}}} d_g^{(|\delta|)} T^{[\alpha]}(\mathbf{r}_m, \mathbf{r}_s) Z_m + \delta_{0|\delta|} V^{\mathrm{mul},[\alpha]}(\mathbf{r}_s) \right).
\end{aligned} \tag{30}
$$

where the summation index m runs over the N_{nuc} nuclei in the QM region.

Finally, the Lennard-Jones potential contributions can be written as functions of functions, on the form $\left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|}\right)^k$, where k is either 12 or 6, and it is therefore necessary to employ a multivariate version of the Faà di Bruno formula[50] which formulates the n'th order derivative of a function $f(g(x_1, \ldots, x_n))$ with respect to $x_1, \ldots, x_n$ as

$$
\frac{\partial^n}{\partial x_1 \cdot \partial x_2 \cdots \partial x_n} f(y) = \sum_{\pi\in\Pi} f^{(|\pi|)}(y) \prod_{B\in\pi} \frac{\partial^{|B|} y}{\prod_{j\in B} \partial x_j} \tag{31}
$$

where the variables x_i may all be different, but need not necessarily be so. In the present case, they are restricted to the three Cartesian coordinates, x, y or z, all belonging to the same nucleus n, as the LJ potential vanishes upon differentiation with respect to coordinates belonging to more than one nucleus. Note that the LJ potential also vanishes upon differentiation with respect to electric field strengths. The summation index π runs over the Π possible partitions of the set $K = \{1, 2, \ldots, n\}$, in other words, Π is the collection of all the distinct ways the values in K can be grouped into non-empty subsets, where each values is included exactly once.[51] The index B runs over each block within partition π, and the index j runs over all elements in block B. For example, for a third-order derivative, $K = \{1, 2, 3\}$, and

$$\Pi = ([1],[2],[3]),([1,2],[3]),([1,3],[2]),([1],[2,3]),([1,2,3]) \tag{32}$$

In this case, there are five partitions, the first of which is $\pi = ([1],[2],[3])$. This partition has three blocks, the first of which is $B = [1]$, which contains only one element, $j = 1$.

Defining

$$g_{LJ} = T^{[(0,0,0)]}(\mathbf{r}_s, \mathbf{r}_n) = \frac{1}{|\mathbf{r}_n - \mathbf{r}_s|}, \tag{33}$$

where $T^{[(0,0,0)]}$ is the zeroth-order interaction tensor, corresponding to a charge-charge interaction,

$$f_{LJ}(g_{LJ}) = \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|}\right)^{12} = (g_{LJ})^{12}, \tag{34}$$

and

$$h_{LJ}(g_{LJ}) = \left(\frac{1}{|\mathbf{r}_n - \mathbf{r}_s|}\right)^{6} = (g_{LJ})^{6}, \tag{35}$$

we can formulate a general expression for the $x_1 x_2 \cdots x_{|\gamma|}$ component of the $|\gamma|$'th-order derivative of the LJ potential with respect to nucleus n as

$$\begin{aligned}\left(d_g^{(|\gamma|)} E_{PE}^{LJ}\right)_{x_1 x_2 \cdots x_{|\gamma|}, n} = 4 \sum_{a}^{N_{\text{frag}}} \sum_{s \in a}^{S_a} \varepsilon_{sn} \Bigg[\sigma_{sn}^{12} \sum_{\pi \in \Pi} f_{LJ}^{(|\pi|)} \prod_{B \in \pi} \frac{\partial^{|B|} g_{LJ}}{\prod_{j \in B} \partial x_j} \\ - \sigma_{sn}^{6} \sum_{\pi \in \Pi} h_{LJ}^{(|\pi|)} \prod_{B \in \pi} \frac{\partial^{|B|} g_{LJ}}{\prod_{j \in B} \partial x_j} \Bigg],\end{aligned} \tag{36}$$

where

$$f_{LJ}^{(k)} = \frac{\partial^k}{\partial g_{LJ}^k} f_{LJ}(g_{LJ}) = \begin{cases} \frac{12!}{(12-k)!} g_{LJ}^{12-k}, & \text{if } k < 13 \\ 0, & \text{otherwise,} \end{cases} \tag{37}$$

$$h_{LJ}^{(k)} = \begin{cases} \frac{6!}{(6-k)!} g_{LJ}^{6-k}, & \text{if } k < 7 \\ 0, & \text{otherwise} \end{cases}, \tag{38}$$

and

$$\frac{\partial^k}{\partial x_1 \partial x_2 \cdots \partial x_k} g_{LJ}(x_1, x_2, \ldots, x_k) = T^{[(k_x, k_y, k_z)]}(\mathbf{r}_s, \mathbf{r}_n) \tag{39}$$

where k_x is the number of the variables $x_1, x_2, \ldots, x_k$ that belong to Cartesian coordinate x, etc., giving $k_x + k_y + k_z = k$. Note that the LJ potential will vanish upon any perturbation with respect to more than one nucleus at a time.

Finally, we remark that many of the expressions shown in this section for differentiated PE interaction contributions involve differentiated density matrices. The obtainment of these differentiated matrices and the evaluation of the full expressions for the response properties also entail the obtainment and use of differentiated Fock matrices when using the aforementioned open-ended response property approach. We refer to Refs. 7 and 8 for more details about these aspects of the general

response property calculation. The expressions needed to evaluate the required differentiated Fock matrix contributions on the right-hand side of Eq. (16) can be found from an analogous development to that presented above, and we do not show them in further detail in the present work.

Having carried out the development presented above, the open-ended expressions for derivatives of PE interaction contributions can be combined with our general framework for open-ended calculation of molecular response properties,[7;8] and our present implementation of this is described in some further detail in Section 3.

This extension of our earlier open-ended implementation of the PE model for electric dipole properties[44] to also include geometric derivatives allows us to calculate properties involving molecular vibrations. The necessary one- and two-electron integrals are obtained from the high-order differentiated integrals (HODIs) available in the LSDalton program package.[52] The HODIs have been implemented utilizing the unified scheme for calculation of differentiated and undifferentiated integrals of Reine, Tellgren and Helgaker.[53]

2.3 2nd-order vibrational perturbation theory

In this work, we apply the methodology presented in this paper to the calculation, including solvent/environment effects, of anharmonic IR intensities and vibrational energy levels, whose working expressions are obtained from second-order vibrational perturbation theory (VPT2),[4;12;13;15;54] which will be recapitulated in the following. VPT2 requires the calculation of second-, third- and fourth-order geometrical derivatives of the molecular energy as well as, for corrected IR intensities, first-, second- and third-order geometrical derivatives of the dipole moment. In the following, geometric derivatives are expressed in a normal-coordinate basis, obtained for isolated molecules through a vibrational analysis of the molecular Hessian.[11] However, in the embedded systems where the PE method has been used to model the environment, a partial Hessian vibrational analysis (PHVA)[55;56;57] is employed, using only the QM–QM block of the full Hessian, as done in our previous work.[47]. From the vibrational analysis we also obtain the harmonic vibrational frequencies.

The straightforward application of VPT2 may lead to problems with near-singularities related to degeneracies and accidental near-degeneracies by Fermi resonances.[12] In order to address this issue, we choose in this work to largely follow the so-called generalized VPT2 (GVPT2) model,[4] which has established itself as a well-performing approach to handle singularities due to Fermi resonances, taking terms deemed to be affected by such resonances out of the perturbation treatment and resolving the involved Fermi state polyads in a variational manner[4] (as was also carried out in Ref. 12); or the deperturbed VPT2 (DVPT2) model[15;58], where the affected terms are removed as in GVPT2 but no subsequent variational treatment is done.

In VPT2, using indices i and j for the N_{mod} vibrational modes, corrected vibrational fundamental frequencies ν_{1_i}, overtones ν_{2_i}, and combination bands $\nu_{1_i1_j}$, all in energy units, are given by

$$\nu_{1_i} = \omega_i + 2X_{ii} + \frac{1}{2}\sum_{j\neq i}^{N_{\text{mod}}} X_{ij}, \tag{40}$$

$$\nu_{2_i} = 2\nu_{1_i} + 2X_{ii}, \tag{41}$$

and

$$\nu_{1_i1_j} = \nu_{1_i} + \nu_{1_j} + X_{ij}, \tag{42}$$

where ω_i is the harmonic frequency of mode i, and where the anharmonicity matrices X_{ii} and X_{ij} are given by

$$X_{ii} = \frac{k_{iiii}}{16} - \sum_k \frac{k_{iik}^2}{32}\left(\frac{4}{\omega_k} + \frac{1}{2\omega_i+\omega_k} - \frac{1}{2\omega_i-\omega_k}\right), \tag{43}$$

and

$$X_{ij} = \frac{k_{iijj}}{4} - \frac{1}{4}\sum_k \frac{k_{iik}k_{jjk}}{\omega_k} - \sum_k \frac{k_{ijk}^2}{8}\left(\frac{1}{\omega_i+\omega_j+\omega_k} + \frac{1}{-\omega_i+\omega_j+\omega_k} + \frac{1}{\omega_i-\omega_j+\omega_k} - \frac{1}{\omega_i+\omega_j-\omega_k}\right), \tag{44}$$

where k_{ijk} and k_{ijkl} are the frequency-reduced third- and fourth-order geometric derivatives of the potential energy—that is, the cubic and quartic force constants. We have in the above expressions and in our implementation and calculations not included Coriolis interactions. Such interactions are most pronounced for small, highly symmetric molecules where it may lead to intensity borrowing and also changes in the vibrational energies. It is assumed to be less important in the case of solvated systems, although reports have been made of Coriolis effects in larger molecules in solution.[59]

From Eq. (43), we note that Fermi-resonance singularities or near-singularities can occur in the last term if $2\omega_i \approx \omega_k$, and similarly in Eq. (44) in the three last terms if any of $\omega_j + \omega_k \approx \omega_i$, $\omega_i + \omega_k \approx \omega_j$ or $\omega_i + \omega_j \approx \omega_k$ occur – particularly so if the associated state coupling as dictated by the corresponding cubic force constant is appreciably large. The decision about whether to regard a tuple of states as involved in a Fermi resonance in this context must necessarily be subject to a choice of threshold, and an established test is to look at a dual set of criteria, where both of[4;12;15]

$$|2\bar{\nu}_i - \bar{\nu}_k| \leq 200 \text{ cm}^{-1}; \quad |\Delta_{ik}| \geq 1 \text{ cm}^{-1}, \tag{45}$$

for the diagonal case, involving the first overtone of mode i and the fundamental k, or both of

$$|\bar{\nu}_i + \bar{\nu}_j - \bar{v}_k| \leq 200 \text{ cm}^{-1}; \quad |\Delta_{ijk}| \geq 1 \text{ cm}^{-1}, \tag{46}$$

for the off-diagonal case, involving a combination state of modes i and j and a fundamental k, must be satisfied in order for the involved states to be deemed as being involved in a Fermi resonance. In the above expressions, the so-called Martin parameters[4;12] Δ_{ik} and Δ_{ijk} are given by

$$\Delta_{ik} = \frac{k_{iik}^4}{25600 \cdot h^4c^4(2\bar{\nu}_i - \bar{\nu}_k)^3}, \tag{47}$$

and

$$\Delta_{ijk} = \frac{k_{ijk}^4}{6400 \cdot h^4c^4(\bar{\nu}_i + \bar{\nu}_j - \bar{v}_k)^3}, \tag{48}$$

where $\bar{\nu}_i$ is the harmonic wavenumber of mode i, *i.e.* corresponding to ω_i, but in units of cm^{-1} instead of energy units. The near-singular terms associated with the states involved in these resonances may be disregarded in the perturbation treatment. In GVPT2, an additional variational treatment is performed (after removal of the aforementioned terms) by solving a new eigenvalue problem for a coupling matrix containing the so-called deperturbed frequencies and cubic and quartic force constants.[4;12]

Infrared intensities are commonly represented by the molar decadic absorption coefficient ε,[1] which for an excitation from the ground state to an excited state $|n\rangle$ is given by

$$\varepsilon_n = \frac{\pi N_A \bar{\nu}_n}{3\ln(10)\varepsilon_0 \hbar c} \sum_{\alpha=x,y,z} |\langle \mu^\alpha \rangle_{0,n}|^2 f(\bar{\nu}; \bar{\nu}_n, \gamma_n), \tag{49}$$

where N_A is Avogadro's constant, ε_0 is the vacuum permittivity, c is the speed of light and $\hbar$ is the reduced Planck's constant. The wavenumber associated with the given state is $\bar{\nu}_n$. The transition dipole moment $\langle 0|\mu_\alpha|n\rangle$ from the ground state to $|n\rangle$, of Cartesian coordinate element α, is represented by $\langle \mu^\alpha \rangle_{0,n}$. The lineshape function f accounts for effects such as the finite lifetime of the excited state,

employs a broadening factor γ_n,[1] and is modelled by the Cauchy distribution. In this work, $|n\rangle$ is either a singly excited state $|1_i\rangle$ of a mode i, or a doubly excited combination state $|1_i 1_j\rangle$ of modes i and j or overtone state $|2_i\rangle$.

A formulation of anharmonic corrections to the vibrational energy levels has already been presented, and we will now show expressions for the anharmonically corrected transition dipole moment — to be used in Eq. (49) — obtained by the perturbation-theory-based approach developed by Barone, Bloino and co-workers.[15;17] The form of the working expressions for these corrections is unchanged upon inclusion of PE effects, and are given by

$$\begin{aligned}
\langle\mu^\alpha\rangle_{0,1i} =& \frac{1}{\sqrt{2}}\mu_i^\alpha + \frac{1}{4\sqrt{2}}\sum_j \mu_{ijj}^\alpha - \frac{1}{8\sqrt{2}}\sum_{jk}\left[k_{ijkk}\mu_j^\alpha\left(\frac{1}{\omega_i+\omega_j} - \frac{1-\delta_{ij}}{\omega_i-\omega_j}\right)\right. \\
&+ k_{ijk}\mu_{jk}^\alpha\left(\frac{1}{\omega_i+\omega_j+\omega_k} - \frac{1}{\omega_i-\omega_j-\omega_k}\right) + \left.\frac{2k_{jkk}\mu_{ij}^\alpha}{\omega_j}\right] \\
&+ \frac{1}{16\sqrt{2}}\sum_{jkl}\left(k_{ikl}k_{jkl}\mu_j^\alpha\left[\frac{4\omega_j(\omega_k+\omega_l)(1-\delta_{ij})(1-\delta_{ik})(1-\delta_{il})}{(\omega_i^2-\omega_j^2)(\omega_i^2-(\omega_k+\omega_l)^2)}\right.\right. \\
&+ \frac{(\omega_k-\omega_l)((\omega_k+\omega_l)^2-3\omega_i^2)}{\omega_i(\omega_i^2-(\omega_k+\omega_l)^2)^2}\delta_{ij}(1+\delta_{ik})(1-\delta_{il}) \\
&- \left.\frac{4\omega_j(4\omega_i+3\omega_k)}{\omega_k(\omega_i^2-\omega_j^2)(2\omega_i+\omega_k)}(1-\delta_{ij})(1-\delta_{ik})\delta_{il}\right] \\
&+ k_{ijk}k_{llk}\mu_j^\alpha\left[\frac{\delta_{ij}}{\omega_i\omega_k}\left(1+\frac{2\delta_{ik}\delta_{il}}{9}\right)\right. \\
&- \frac{4\omega_j}{\omega_k(\omega_i^2-\omega_j^2)}(1-\delta_{ij})(1-\delta_{ik})(1-\delta_{il}) \\
&- \left.\left.\frac{4\omega_j\delta_{ik}(1-\delta_{ij})}{\omega_i(\omega_i^2-\omega_j^2)}\left(1+\frac{2\delta_{il}}{3}\right)\right]\right)
\end{aligned} \tag{50}$$

for the corrected fundamental bands, and by

$$\begin{aligned}
\langle\mu^\alpha\rangle_{0,(1+\delta_{ij})_i(1-\delta_{ij})_j} =& \left(\sqrt{2}+(1-\sqrt{2})\delta_{ij}\right)\frac{1}{2\sqrt{2}} \\
&\times\left[\mu_{ij}^\alpha + \sum_k k_{ijk}\mu_k^\alpha\frac{\omega_k}{(\omega_i+\omega_j)^2-\omega_k^2}\right]
\end{aligned} \tag{51}$$

for overtone ($i=j$) and combination bands ($i\neq j$). In Eqs. (50) and (51), the only harmonic (*i. e.* not anharmonic) contribution is the first term on the right-hand side of Eq. (50). These equations are here formulated for the specific case of electric polarization properties, while the formulation in the original works is general also for their magnetic counterparts. As there is a minor error in the expression for the corrected fundamental transition moments in the original work (Ref. 15), we refer the reader to the more recent work in Ref. 17, where this has been corrected. We note from Eqs. (43) (44),(50), and (51), that the expressions to be evaluated involve the full cubic force constant tensor but only the semi-diagonal quartic force constants.

As was discussed concerning the corrections to vibrational energy levels, the IR intensity expressions can likewise be affected by Fermi resonances. These are identified in the same way as shown above for the energy levels, but are in the present work treated in the DVPT2 fashion, which entails simply removing the affected terms from the perturbation treatment. We note that a GVPT2 treatment for transition moments has been developed by Vazquez and Stanton[60].

In addition to Fermi resonances, transition moments for fundamentals (not overtone and combination bands) can be subject to another source of near-singularity, namely the so-called 1-1 resonances which occur in terms of the type $1/(\omega_i - \omega_j)$ (due to the condensed form of our expression, they are not all directly visible; see Ref. 17 for the factorized equations) when $\omega_i \approx \omega_j$. These resonances can be identified by a three-fold test, whereby a 1-1 resonance is deemed to occur if[58]

$$|\bar{\nu}_i - \bar{\nu}_j| \leq 100 \text{ cm}^{-1}, \tag{52}$$

is true, and at least one of

$$|K| \geq 10 \text{ cm}^{-1}, \tag{53}$$

and

$$|K|/(|\bar{\nu}_i - \bar{\nu}_j)^2 \geq 1 \text{ cm}, \tag{54}$$

is true. The factor K is here a measure of coupling strength — $k_{ijkk}/(100hc)$ for terms involving quartic force fields and $k_{ijk}k_{lmn}/(100hc)^2$ for terms containing a product of cubic force fields. In the case of a product of cubic force fields, the threshold in Eq. (53) is 10 cm^{-2} instead of 10 cm^{-1}, and in Eq. (54) 1 (dimensionless) instead of 1 cm.

We remark that from personal communication with Bloino,[61] we have been made aware that in cases where 1-1 resonances have been identified, and the affected near-singular terms have been removed, then certain term cancellations that were valid in the absence of such removal and were applied to produce the expression underlying Eq. (50) in the present work no longer hold. In order to achieve complete consistence in such a situation, a version of Eq. (50) which takes this into account should be used, but this has not been done in the present work.

3 Implementation and computational details

Implementation

Our present implementation of the theory presented above takes place in an environment of well-defined modules. The underlying host program is the LSDalton[52;62] program package, which also contains the HODI[53;63] module for perturbed one- and two-electron integrals. The identification of and overarching management of terms contributing to a response property is carried out by the OpenRSP open-ended response program library.[10] The calculation of PE contributions is managed and carried out by the FraME[64] library.

The general program structure for a response property calculation in this framework can be summarized as follows: LSDalton will invoke OpenRSP for the calculation of a given property or collection of properties, supplying the unperturbed Fock, density, and overlap matrices — $\mathbf{F}$, $\mathbf{D}$, and $\mathbf{S}$, respectively. Any PE contributions to these matrices are supplied by FraME. OpenRSP will then identify and manage the calculation of the necessary perturbed QM and PE contributions needed for calculating the perturbed $\mathbf{F}$ and $\mathbf{D}$ used in subsequent stages of the calculation, and then do the same for the contributions needed for the actual response property evaluation. OpenRSP identifies, obtains and assembles the various differentiated contributions, making use of callback routines to the host program, which in the present case is LSDalton, which in turn calls the appropriate modules for their calculation. Terms with a zeroth-, first-, and second-order dependence on the density matrix are grouped into separate callback categories, and there is also a callback category separation between differentiated Fock matrix and energy contributions. We note that the differentiated PE contributions to the Fock matrix and energy were similarly grouped into orders of density matrix dependence and are considered under the same callback categorizations.

Non-PE one- and two-electron integral contributions are calculated using the HODI module or other internal routines within the LSDalton package. PE contributions — entailing *e.g.* the calculation of undifferentiated or differentiated electrostatic-potential integrals, interaction tensors, Lennard-

Jones contributions, and any subsequent contraction with perturbed or unperturbed density matrices supplied as part of the invocation of a callback function — are obtained from FraME, which in turn calls HODI for any necessary differentiated integrals. OpenRSP also uses callback functionality for obtaining the nuclear repulsion contributions that are independent of the density matrix and for obtaining solutions to the so-called response equations.[65]

Both OpenRSP's and FraME's matrix operations, including their connection to LSDalton's native matrix functionality, are mediated by the use of the QcMatrix package[66], making such operations in the two former modules agnostic to any specific implementation of matrix structure and operations in the host program or elsewhere.

The overall structure of the programs and modules used for the calculation of energy derivatives are illustrated in Figure 1. We also note that in addition to what has been described in this current section and visualized in Figure 1, we use the independent Python Package for Vibrational Spectroscopy, SpectroscPy,[67] to obtain vibrational energy levels and intensities and to plot the calculated spectra.

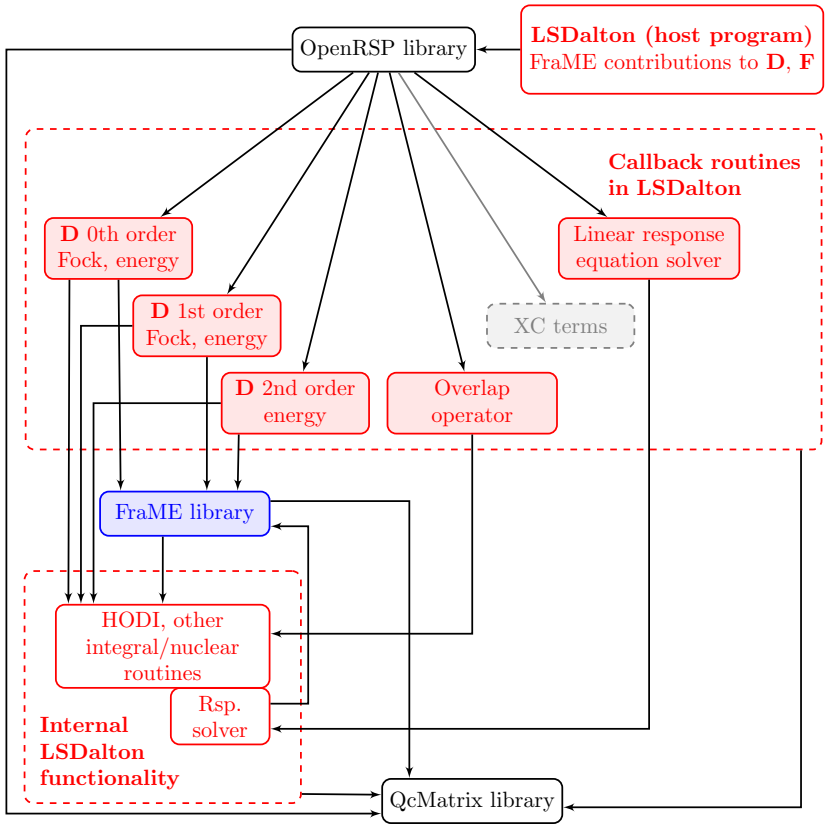


Figure 1: An overview of the program modules used in this work including the FraME library. Exchange–correlation (XC) functionality is not used in this work but is included here for completeness.

Water in fullerene

For the generation of the structure of the endohedral fullerene complex, coordinates for C_{60} were obtained from the Protein Data Bank[68] and a single water molecule was inserted using Avogadro[69;70]. The resulting structure was converted to a suitable format using the Molecular formats converter[71]. The PyFraME[72] package was then used to identify the water molecule as the QM region, and to calculate classical parameters for each atom in the fullerene (which was identified as the environment), using the implementation of the LoProp partitioning approach[40;73] in the Dalton program[62;74]. The calculations were performed using the B3LYP functional[75;76;77;78;79] and a recontracted version of the 6-31+G* basis set[80;81;82] (called loprop-6-31+G* in Dalton). LJ parameters were taken from the OPLS-AA force field.[83]

Geometry optimization was then performed on the encapsulated water molecule with a frozen environment using LSDalton[52;62] and FraME[64]. The structure was optimized with Hartree–Fock (HF) as the underlying electronic structure level of theory and the pcseg-2[84] basis set. The optimized endohedral fullerene complex is shown in Figure 2. Subsequent response calculations for properties needed to obtain anharmonic vibrational energies and IR spectral intensities were then performed, using the same packages in combination with the open-ended response theory framework OpenRSP.[10] Similar geometry optimizations and response calculations were also performed for water in vacuum for comparison. The Python Package for Vibrational Spectroscopy, SpectroscPy[67], was then used to perform the harmonic vibrational analysis and to calculate anharmonically corrected vibrational energy levels and IR spectra. A broadening factor of $\gamma_i = 3$ cm^{-1} for each peak was used in the visualization of the vibrational spectra. Concerning the vibrational analysis, for the water molecule in isolation, the rotational and translational modes were projected out, but for water in fullerene, this projection cannot be carried out, since we use PHVA. In order to address this, we instead define a harmonic wavenumber cut-off above which we from visual inspection judge the modes to be reasonably dominated by vibrational motion, and exclude all modes whose harmonic wavenumbers fall beneath this threshold from all subsequent analysis. In the case of endohedral fullerene, this cut-off was set at 1500 cm^{-1}.

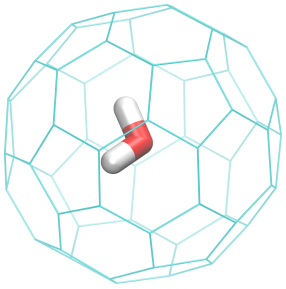


Figure 2: Optimized geometry of the endohedral fullerene complex with water.

Acetonitrile

The study of the solvated acentonitrile system was carried out using the three solvents chloroform, acetonitrile and water, in addition to a reference calculation for acetonitrile in vacuum. Classical MD simulations were performed with GROMACS 2018.3[85;86;87], using the optimized potential for liquid simulations all-atom (OPLS-AA) force field.[83] The OPLS-AA topology for acetonitrile and chloroform were taken from the GROMACS molecule and liquid database at virtualchemistry.org,[88;89] while the TIP3P potential[90] was used for water. An initial cubic box of 60Å × 60Å × 60Å containing a

single acetonitrile molecule was created, and solvent molecules were added. The structures were then minimized using 200 steepest descent steps and 40 000 conjugate gradient steps. An initial NVT equilibration was run for 0.5 ns with a time-step of 0.01 fs, followed by two NPT equilibrations, the first for 0.2 ns with a time step of 0.05 fs, and the second for 5 ns with a time step of 0.5 fs. A second NVT equilibration was then run for 1 ns before an NVT production simulation was run for 10 ns, both with a time step of 0.5 fs. All equilibrations and simulations were performed using the velocity rescaling[91] (298 K) thermostat and the Berendsen barostat[92] (1 bar), both with a coupling constant of 0.5 ps. Periodic boundary conditions, the leap-frog integrator and a cut-off at 15 Å were used. Outside the cut-off, electrostatic interactions were modelled using the Ewald smooth particle mesh.[93] Snapshots from the final trajectory were extracted at 10 ps intervals.

In this proof-of principle work, only 10 snapshots, every second among the first 20 snapshots, were selected for further calculations. In a similar procedure to what was described above for water in fullerene, the central acetonitrile molecule was defined as the QM region, and solvent parameters were calculated using PyFraME, Dalton and LoProp, for each of the selected snapshots. These structures were then optimized with LSDalton and FraME with HF as the underlying level of electronic structure theory and the pcseg-2 basis set, before response calculations were performed with the LSDalton/FraME/OpenRSP setup to obtain the necessary response properties for calculating anharmonic vibrational energy levels and IR spectral intensities. Due to convergence issues related to the geometry optimization, the analysis consists only of 9 snapshots in the case of acetonitrile solvent. Following the procedure outlined above for the water in fullerene calculations, a harmonic wavenumber cut-off, for the present calculations chosen at 400 cm^{-1}, was employed, and all modes whose harmonic wavenumbers fall beneath this threshold were excluded in all further analysis. The IR lineshapes from each of the snapshots within a specific solvent were broadened as described for the water in fullerene calculations, using the same broadening factor of $\gamma_i = 3$ cm^{-1}, and then linearly combined for the spectra presented in Section 4, while the table values presented in Tables 3-6 in the Supporting Information are the result of averaging over the snapshots. Similar geometry optimizations and response calculations were also performed for acetonitrile in vacuum for comparison, but in this case, the aforementioned harmonic wavenumber cut-off was not employed, and rotational and translational modes were instead projected out. All relevant data associated with the water in fullerene and acetonitrile in solvents calculations performed in this work can be found in the Dataset associated with this manuscript.[94]

4 Results and discussion

Water in fullerene

Infrared spectra for water in C_{60} and in vacuum are shown in Figure 3, and the corresponding numerical values can be found in the Supporting Information (SI). Bond lengths and bond angles are also tabulated in the SI. Combination bands are observed to be weak, and overtone features are so weak that they are not discernible in Figure 3. Both overtone features and combination bands are, however, presented in separate plots in the left and right panels in Figure 4, respectively.

The inclusion of anharmonic effects leads to a lowering of the vibrational frequencies of all the fundamental transitions, ranging from around -200 cm^{-1} for the asymmetric stretching mode (the most high-frequent of the fundamental modes in the region around 4000 cm^{-1}) to -70 cm^{-1} for the bending mode (in the region around 1700 cm^{-1}). The effect of anharmonic corrections to the intensities is much less dramatic, with a maximum change of 5 $km \cdot mol^{-1}$.

With the embedding of water in fullerene, we observe a negative frequency shift for all transitions with respect to the ground state, including overtones and combination states. The most prominent frequency shift for the (anharmonic) fundamentals is observed for the asymmetric O-H stretching mode, with a shift of -71 cm^{-1}. In general, the overtones display frequencies and frequency shifts of about twice those of the fundamentals. As for the fundamentals, the most prominent frequency

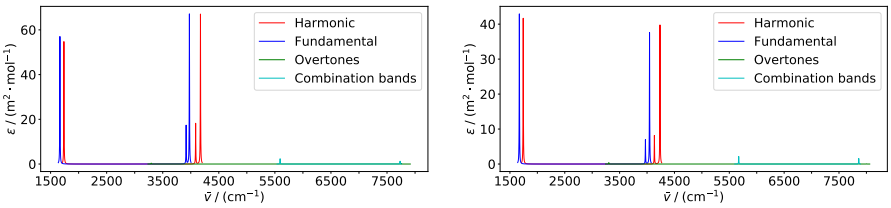


Figure 3: IR spectra for water in fullerene (left panel) and in vacuum (right panel). Double-harmonic approximation spectra in red; spectra with anharmonic corrections plotted separately for fundamental transitions (dark blue), overtones (green) and combination bands (cyan).

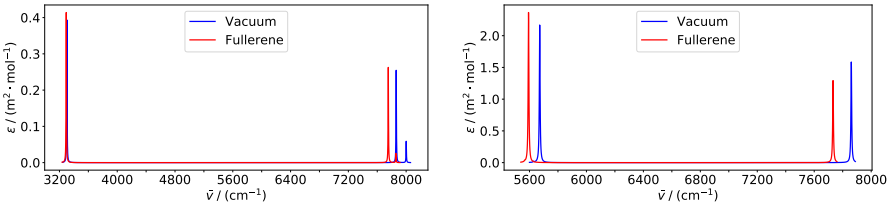


Figure 4: Close-up of Figure 3, separated into spectra for overtones (left panel) and combination bands (right panel).

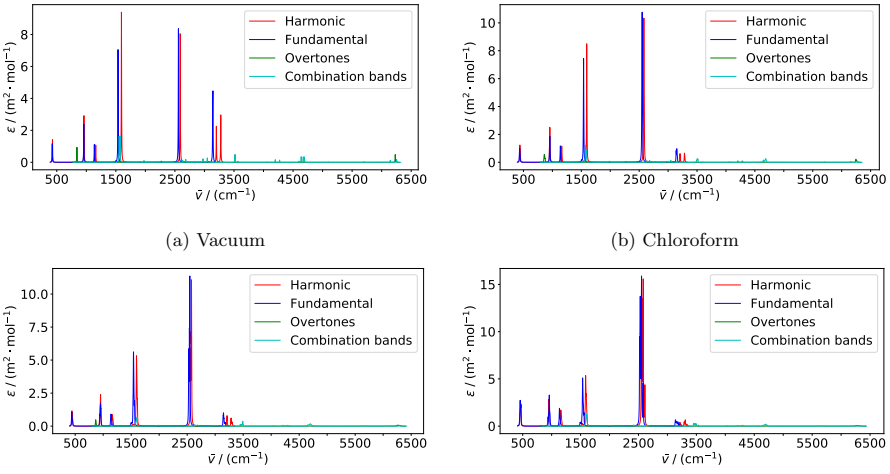
shift (-143 cm^{-1}) is displayed for the overtone of the asymmetric O-H stretching mode (the most high frequent in the region around 8000 cm^{-1}). Among the combination bands, the most prominent frequency shift takes place for the combination of the two stretching modes.

All fundamental transition features gain intensity when the water molecule is enclosed in fullerene. The largest increase is observed for the asymmetric O-H stretching mode, which is increased by 64 $km \cdot mol^{-1}$, while the highest relative change is observed for the symmetric stretching mode, which is increased by 149 %. An interesting observation is that for water in fullerene, there is a 1-1 resonance between the two stretching modes which is absent in vacuum. This affects the corrected intensities of the fundamentals, but the actual numerical implications are not large. Most overtone and combination band intensities are increased by less than 10 % upon embedding in fullerene. There are, however, two exceptions, namely the overtone of the asymmetric stretching mode, whose intensity is decreased by 57 % and the combination band of the two stretching modes, whose intensity is decreased by 18 %.

Acetonitrile

The harmonic and anharmonic IR spectra for acetonitrile in vacuum and the three different solvents are shown in Figure 5, demonstrating the effect of anharmonic corrections. Figure 6 shows spectra exclusively for the fundamental transitions, but with all solvents in the same plot, in order to better illustrate solvent effects for these transitions. Similar plots are shown for overtone features and combination bands in Figures 7 and 8, respectively. The corresponding numerical values are collected in the Supporting Information. Bond lengths and angles are shown in Table 1, and changes in these upon solvation are typically found to be small. In this section, all frequency shifts as a result of embedding in solvent are based on the positions of the peaks in the spectra, while values related to

intensities, the inclusion of anharmonic effects, and the frequency difference between individual states, are taken from the tables in the Supporting Information.



(a) Vacuum

(b) Chloroform

(c) Acetonitrile

(d) Water

Figure 5: Infrared spectra of acetonitrile in various environments. Double-harmonic approximation spectra in red; spectra with anharmonic corrections plotted separately for fundamental transitions (dark blue), overtones (green) and combination bands (cyan).

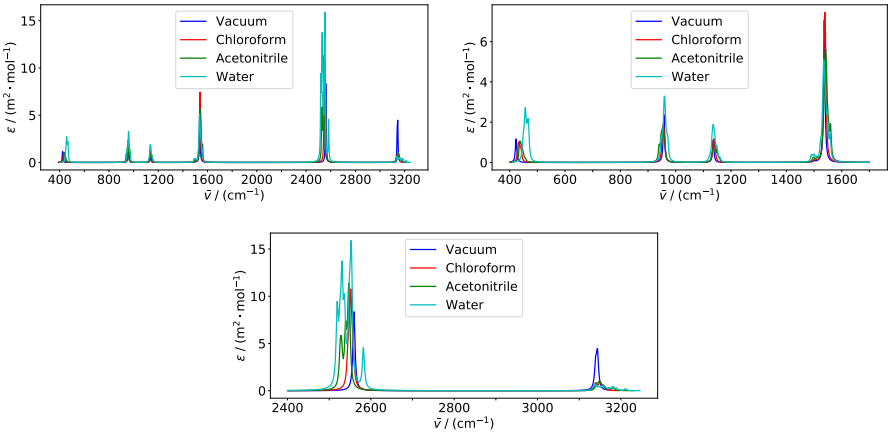


Figure 6: Anharmonic fundamental IR spectra for fundamental transitions in acetonitrile in vacuous, chloroform, acetonitrile and water environments. The lower and upper right panels are magnified versions of the upper left.

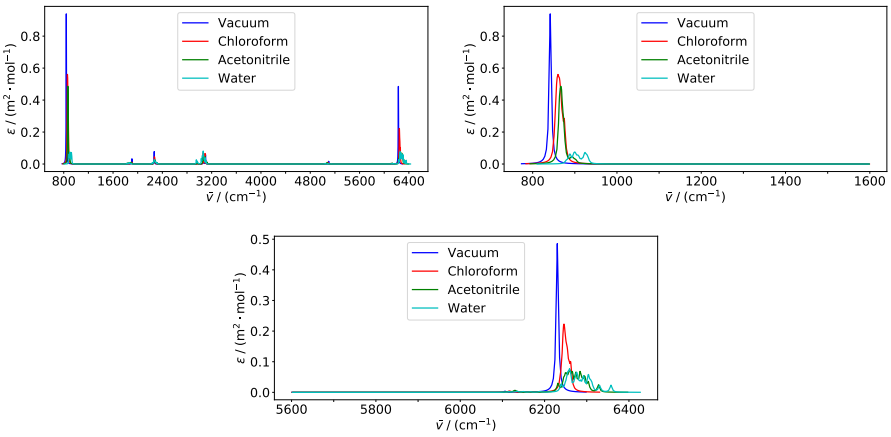


Figure 7: Overtone IR spectra for acetonitrile in vacuous, chloroform, acetonitrile and water environments. The lower and upper right panels are magnified versions of the upper left.

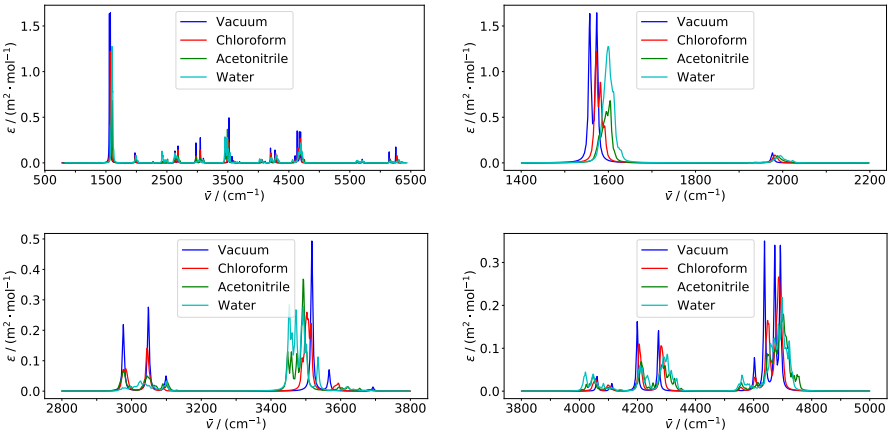


Figure 8: Combination band IR spectra for acetonitrile in vacuous, chloroform, acetonitrile and water environments. The lower and upper right panels are magnified versions of the upper left.

Anharmonic corrections lead in general to a lowering of the vibrational frequencies of the fundamental transitions in all solvents, the only exception being the C-C stretching mode in water and acetonitrile which displays a small shift towards higher frequencies when anharmonic effects are included. In contrast, no clear trend is seen for the anharmonic corrections to the intensities. There are six strong fundamental bands belonging to the following modes: Three C-H stretching modes in the region above 3000 cm^{-1}, the C-N stretching mode at around 2550 cm^{-1} and two H-C-H bending modes at around 1550 cm^{-1}. Two weaker CH_3 rocking modes can be found around 1550 cm^{-1}, a

Table 1: Bond lengths[a] and angles[b] for acetonitrile in various solvents[c] and vacuum

Feature	Vacuum	Chloroform	Acetonitrile	Water
Bond lengths				
C-C	1.464	1.464	1.465	1.462
C-N	1.126	1.126	1.127	1.127
C-H	1.080	1.079	1.079	1.079
Bond angles				
C-C-N	180.0	179.6	179.4	178.8
H-C-C	109.7	109.5	109.4	109.1

[a] given in units of Å; [b] given in units of degrees; [c] values are averaged over the snapshots.

C-C stretching mode at around 950 cm^{-1} and two weak N-C-C bending modes at around 450 cm^{-1}. There is also a very weak CH_3 umbrella mode in the region around 1500 cm^{-1}, but its intensity is so small as not to be discernible in any of the spectra.

For the isolated acetonitrile molecule, and all solvents except water, the asymmetric and symmetric C-H stretching modes are well separated in frequency in the harmonic approximation, resulting in two distinct peaks with the former being the most intense, bearing in mind that its intensity is a combination of those of the fundamental transitions for two asymmetric modes. However, when anharmonic corrections are included, the symmetric and asymmetric C-H stretching frequencies become more similar, resulting in a single, more intense, band. When solvents are included, we observe a different effect, where the frequency split between the symmetric stretch and the most high-frequent of the asymmetric stretching modes is increased, ranging from 16 cm^{-1} in chloroform to 38 cm^{-1} in water. This is due to the fact that whereas the frequency of the symmetric stretching mode is quite unchanged upon addition of solvent, the frequencies of both asymmetric modes, and in particular the most high-frequent one, are increased by up to 36 cm^{-1}, compared to vacuum.

Among the previously mentioned dominant fundamental peaks, the most prominent belong to the the C-H stretching modes, the C-N stretching mode and the H-C-C bending modes, which will be discussed in the following. The most intense peak in the fundamental IR spectrum belongs to the C-N stretching mode. The addition of a solvent leads to a negative frequency shift for this peak for all solvents, of up to -13 cm^{-1}. The frequency shift is most pronounced upon solvation in acetonitrile, which is in agreement with the fact that embedding in this solvent also leads to the largest elongation of the C-N bond. No clear trend in the solvent-induced frequency shifts is apparent for the other two prominent fundamental bands, but the frequency shifts are in general smaller than those for the C-N stretching mode, in line with the smaller changes observed in the corresponding bond lengths and angles upon solvation. A qualitative difference between acetonitrile in solvents and in vacuum can, however, be observed in the two H-C-H bending modes, whose frequencies in vacuum are very close, but are split by up to 9 cm^{-1} in water. The largest solvent effect on the intensities is observed for the C-N stretching mode fundamental transition in water, whose associated intensity is increased by more than a factor of 5 compared to vacuum. The other solvents also leads to a marked increase in the intensity associated with this transition — an increase by roughly a factor of 1 to 2 compared to vacuum. The intensity of the C-H stretching modes' fundamental transitions are on average reduced by approximately 50 % with the addition of solvent, while those for the H-C-H bending modes on average increase by 50 %.

Among the overtones, two peaks dominate, belonging to the two asymmetric C-H stretching overtones at 6300 cm^{-1} and the two N-C-C bends at 900 cm^{-1}. All overtone bands for the investigated systems are weak, with intensities lower than 2 $km \cdot mol^{-1}$. Both of these features display an increase in frequency when adding a solvent, by up to +58 cm^{-1} for the N-C-C bending mode in water. In contrast, only minor changes upon solvation are observed in the intensities for these two dominating peaks, except for the N-C-C overtone band, whose intensity increases by 50 % in chloroform, but decreases by 64 % in water, in both cases relative to vacuum. From Figure 7 it is clear that both of

the overtone features experience a high degree of broadening upon solvation, which to a large extent is due to the MD sampling, in particular for water. The broadening is further increased by a frequency split within each category of vibrational motion, in particular in water, where the two asymmetric C-H stretching modes are split by 29 cm^{-1} and the two N-C-C bending modes are split by 27 cm^{-1}.

There are three major combination bands, belonging to the combinations of the two methyl rocking modes with the two N-C-C bending modes (at 1600 cm^{-1}), the combination of the C-C stretch with the C-N stretch (at 3500 cm^{-1}) and the combination of each of the two H-C-H bending modes combined with each of the two asymmetric C-H stretching modes, mixing in the spectrum with the combination bands of the methyl umbrella and the asymmetric C-H stretch (at 4700 cm^{-1}). The most intense of these features, at 1600 cm^{-1}, is due to several overlapping bands, each with individual intensities of up to 2.5 $km \cdot mol^{-1}$ in water. All these combination bands show considerable frequency shifts upon solvation. The peaks at around 1600 and 4700 cm^{-1} both show positive frequency shifts of up to +63 cm^{-1} in acetonitrile compared to in vacuum. The peak at 3500 cm^{-1} experiences a negative frequency shift of up to -64 cm^{-1} with water compared to vacuum.

For the results presented in this section that involve solvation, we note that the extent of the conformational sampling employed and the size of solvent region included in this proof-of-concept work were quite limited, and may thus have produced results that are not fully representative of bulk solvation behavior. We therefore remark that the findings presented above must be viewed with this in mind, and that future applications of this methodology would likely benefit from increased attention to these topics.

5 Conclusions and outlook

We have presented an open-ended formulation of geometric derivatives of environment interaction contributions described by a polarizable embedding model. The strategy underlying the present formulation and implementation is aligned with an existing open-ended formulation of response theory and the methodology described in this paper is applicable for calculating polarizable embedding contributions to molecular properties corresponding to freely chosen combinations of geometric and electric dipole perturbations. We have presented proof-of-principle results that demonstrate this functionality by calculating anharmonically corrected vibrational energy levels and infrared spectra for water in a fullerene cage and acetonitrile in various solvents, comparing each set of calculations to their vacuum counterparts.

We observe that the encapsulation of a water molecule in fullerene shifts the vibrational wavenumbers of water to lower values. The same trend is observed upon introduction of anharmonic corrections, both for the lone and encapsulated water molecule. We also observe a change in intensity for overtone and combination band features upon encapsulation which in many cases is small, but in two cases were observed to decrease the intensity by 18% and 57%, respectively.

For acetonitrile, the results in solvent are based on limited MD sampling and using a rather small solvation shell. For a more representative description of solvent behavior, a larger solvation shell should be used and a more extensive conformational sampling should be performed, but we have nevertheless made some observations from our results, keeping these limitations in mind. We observe that anharmonic corrections in general lower the frequencies of the fundamental transitions, with the exception of that of the C-C stretching mode when solvated in water and acetonitrile, but no clear trend was observed for the anharmonic corrections to the intensities of the fundamental transitions. We observe that the introduction of solvents sometimes produces quite strong intensity changes in the strongest fundamental transition peaks, compared to the vacuum results. In two cases, for the solvents chloroform and water, we also observe significant intensity changes compared to vacuum for the N-C-C overtone band.

We have in the present work limited ourselves to anharmonic IR spectra, but the methodology developed here allows us to calculate properties that can be used to study a wide range of vibrational spectroscopies whose description uses geometric derivatives of electric dipole polarization properties,

such as hyper-Raman spectra[95;96;97;98] or other advanced vibrational spectroscopic processes like two-dimensional IR spectroscopy (see e.g. Ref. 99). Extension of the present methodology to also cover perturbations corresponding to magnetic dipoles or multipoles, or higher-order electric multipoles, would pave the way for applications in the area of chiroptical spectroscopy. Although this would entail additional development, much of the general framework presented here is also applicable in the calculation of these properties.

6 Acknowledgements

This work has received support from the Norwegian Supercomputing Program (NOTUR) through a grant of computer time (Grant No. NN4654K). Financial support is acknowledged from the Research Council of Norway through its Centres of Excellence scheme (Grant No: 262695) and a research grant (Grant No. 250743). J.M.H.O. acknowledges financial support from VILLUM FONDEN (Grant No. 29478). M.R. acknowledges financial support from the Research Council of Norway and MSCA COFUND (Grant No. 274918).

References

[1] P. Norman, K. Ruud, and T. Saue. *Principles and practices of molecular properties: Theory, modeling, and simulations*. Wiley Online Library, 2018.

[2] T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud. Recent advances in wave-function-based methods of molecular-property calculations. *Chem. Rev.*, 112:543–631, 2012.

[3] R. Bast, U. Ekström, B. Gao, T. Helgaker, K. Ruud, and A. J. Thorvaldsen. The ab initio calculation of molecular electric, magnetic and geometric properties. *Phys. Chem. Chem. Phys.*, 13:2627–2651, 2011.

[4] V. Barone. Anharmonic vibrational properties by a fully automated second-order perturbative approach. *J. Chem. Phys*, 122(1):014108, 2005.

[5] T.-C. Jagau, J. Gauss, and K. Ruud. Analytic evaluation of the dipole hessian matrix in coupled-cluster theory. *J. Chem. Phys.*, 139:154106, 2013.

[6] J. Z. Gong, D. A. Matthews, Changala. P. B., and J. F. Stanton. Fourth-order vibrational perturbation theory with the Watson Hamiltonian: Report of working equations and preliminary results. *J. Chem. Phys.*, 149(11):114102, 2018.

[7] A. J. Thorvaldsen, K. Ruud, K. Kristensen, P. Jørgensen, and S. Coriani. A density matrix-based quasienergy formulation of the kohn–sham density functional response theory using perturbation- and time-dependent basis sets. *J. Chem. Phys.*, 129(21):214108, 2008.

[8] M. Ringholm, D. J. Jonsson, and K. Ruud. A general, recursive, and open-ended response code. *J. Comput. Chem.*, 35(8):622–633, 2014.

[9] D. H. Friese, M. T. P. Beerepoot, M. Ringholm, and K. Ruud. Open-ended recursive approach for the calculation of multiphoton absorption matrix elements. *J. Chem. Theory Comput.*, 11:1129–1144, 2015.

[10] R. Bast, D. H. Friese, B. Gao, D. J. Jonsson, M. Ringholm, S. Reine, and K. Ruud. OpenRSP: an open-ended response property library (version 1.0.0), **2020**, DOI: 10.5281/zenodo.3923836. See https://www.openrsp.org/.

[11] E. B. Wilson, J. C. Decius, and P. C. Cross. *Molecular vibrations: the theory of infrared and Raman vibrational spectra*. Courier Corporation, 1955.

[12] J. M. L. Martin, T. J. Lee, P. R. Taylor, and J. François. The anharmonic force field of ethylene, C2H4, by means of accurate ab initio calculations. *J. Chem. Phys.*, 103(7):2589–2602, 1995.

[13] H. H. Nielsen. The Vibration-Rotation Energies of Molecules. *Rev. Mod. Phys.*, 23:90, 1951.

[14] D. Papoušek and M. R. Aliev. *Molecular Vibrational-Rotational Spectra*. Elsevier, Amsterdam, NL, 1981.

[15] J. Bloino and V. Barone. A second-order perturbation theory route to vibrational averages and transition properties of molecules: General formulation and application to infrared and vibrational circular dichroism spectroscopies. *J. Chem. Phys.*, 136(12):124108, 2012.

[16] *Handbook of Vibrational Spectroscopy*. Wiley, 2002.

[17] J. Bloino. A VPT2 Route to Near-Infrared Spectroscopy: The Role of Mechanical and Electrical Anharmonicity. *J. Phys. Chem. A*, 119(21):5269–5287, 2015.

[18] Y. Cornaton, M. Ringholm, O. Louant, and K. Ruud. Analytic calculations of anharmonic infrared and Raman vibrational spectra. *Phys. Chem. Chem. Phys.*, 18(5):4201–4215, 2016.

[19] André Severo Pereira Gomes and Christoph R. Jacob. Quantum-chemical embedding methods for treating local electronic excitations in complex chemical systems. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.*, 108:222, 2012.

[20] J. Tomasi, B. Mennucci, and R. Cammi. Quantum mechanical continuum solvation models. *Chem. Rev.*, 105:2999–3094, 2005.

[21] U. N. Morzan, D. J. A. de Armiño, N. O. Foglia, F. Ramírez, M. C. G. Lebrero, D. A. Scherlis, and D. A. Estrin. Spectroscopy in complex environments from QM–MM simulations. *Chem. Rev.*, 118:4071–4113, 2018.

[22] R. Di Remigio, M. T. P. Beerepoot, Y. Cornaton, M. Ringholm, A. H. Steindal, K. Ruud, and L. Frediani. Open-ended formulation of self-consistent field response theory with the polarizable continuum model for solvation. *Phys. Chem. Chem. Phys.*, 19(1):366–379, 2017.

[23] C. Cappelli, S. Corni, R. Cammi, B. Mennucci, and J. Tomasi. Nonequilibrium formulation of infrared frequencies and intensities in solution: Analytical evaluation within the polarizable continuum model. *J. Chem. Phys.*, 113(24):11270–11279, 2000.

[24] P. Bouř, D. Michalík, and J. Kapitán. Empirical solvent correction for multiple amide group vibrational modes. *J. Chem. Phys.*, 122(14):144501, 2005.

[25] T. Giovannini, F. Egidi, and C. Cappelli. Molecular spectroscopy of aqueous solutions: a theoretical perspective. *Chemical Society Reviews*, 2020.

[26] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, 103:227–249, 1976.

[27] Hans Martin Senn and Walter Thiel. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.*, 48(7):1198–1229, jan 2009.

[28] Elizabeth Brunk and Ursula Rothlisberger. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chem. Rev.*, 115(12):6217–6263, apr 2015.

[29] S. W. Rick, S. J. Stuart, and B. J. Berne. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.*, 101(7):6141–6156, October 1994.

[30] R. A. Bryce, R. Buesnel, I. H. Hillier, and N. A. Burton. A solvation model using a hybrid quantum mechanical/molecular mechanical potential with fluctuating solvent charges. *Chem. Phys. Lett.*, 279(5-6):367–371, November 1997.

[31] F. Lipparini, C. Cappelli, and V. Barone. Linear response theory and electronic transition energies for a fully polarizable QM/classical Hamiltonian. *Journal Chem. Theory Comput.*, 8(11):4153–4165, 2012.

[32] F. Lipparini, C. Cappelli, G. Scalmani, N. De Mitri, and V. Barone. Analytical first and second derivatives for a fully polarizable QM/classical hamiltonian. *J. Chem. Theory Comput.*, 8:4270–4278, 2012.

[33] T. Giovannini, M. Olszówka, F. Egidi, J. R. Cheeseman, G. Scalmani, and C. Cappelli. Polarizable embedding approach for the analytical calculation of raman and raman optical activity spectra of solvated systems. *J. Chem. Theory Comput.*, 13:4421–4435, 2017.

[34] R. Di Remigio, T. Giovannini, M. Ambrosetti, C. Cappelli, and L. Frediani. Fully polarizable QM/fluctuating charge approach to two-photon absorption of aqueous solutions. *J. Chem. Theory Comput.*, 15(7):4056–4068, 2019.

[35] T. Giovannini, L. Grazioli, M. Ambrosetti, and C. Cappelli. Calculation of IR spectra with a fully polarizable QM/MM approach based on fluctuating charges and fluctuating dipoles. *J. Chem. Theory Comput.*, 15(10):5495–5507, August 2019.

[36] T. Giovannini, A. Puglisi, M. Ambrosetti, and C. Cappelli. Polarizable QM/MM approach with fluctuating charges and fluctuating dipoles: The QM/FQFμ model. *J. Chem. Theory Comput.*, 15(4):2233–2245, 2019.

[37] P. Reinholdt, M. S. Nørby, and J. Kongsted. Modeling of Magnetic Circular Dichroism and UV/vis absorption spectra using Fluctuating Charges or Polarizable Embedding within a resonant-convergent response theory formalism. *J. Chem. Theory Comput.*, 14(12):6391–6404, 2018.

[38] J. M. Olsen, K. Aidas, and J. Kongsted. Excited states in solution through polarizable embedding. *J. Chem. Theory Comput.*, 6(12):3721–3734, 2010.

[39] C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted, and J. M. H. Olsen. Response properties of embedded molecules through the polarizable embedding model. *Int. J. of Quantum Chem.*, 119(1):e25717, 2019.

[40] L. Gagliardi, R. Lindh, and G. Karlström. Local properties of quantum chemical systems: The LoProp approach. *J. Chem. Phys.*, 121(10):4494–4500, 2004.

[41] D. W. Zhang and J. Z. H. Zhang. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein–molecule interaction energy. *J. Chem. Phys.*, 119(7):3599–3605, 2003.

[42] P. Söderhjelm and U. Ryde. How accurate can a force field become? A polarizable multipole model combined with fragment-wise quantum-mechanical calculations. *J. Phys. Chem. A*, 113(3):617–627, 2008.

[43] N. List, M. T. P. Beerepoot, J. M. H. Olsen, B. Gao, K. Ruud, H. J. Aa. Jensen, and J. Kongsted. Molecular quantum mechanical gradients within the polarizable embedding approach—application to the internal vibrational stark shift of acetophenone. *J. Chem. Phys.*, 142(3):034119, 2015.

[44] A. H. Steindal, M. T. P. Beerepoot, M. Ringholm, N. H. List, K. Ruud, J. Kongsted, and J. M. H. Olsen. Open-ended response theory with polarizable embedding: multiphoton absorption in biomolecular systems. *Phys. Chem. Chem. Phys.*, 18(40):28339–28352, 2016.

[45] N. H. List, J. M. H. Olsen, and J. Kongsted. Excited states in large molecular systems through polarizable embedding. *Phys. Chem. Chem. Phys.*, 18(30):20234–20250, 2016.

[46] C. Steinmann, J. M. H. Olsen, and J. Kongsted. Nuclear magnetic shielding constants from quantum mechanical/molecular mechanical calculations using polarizable embedding: role of the embedding potential. *J. Chem. Theory Comput.*, 10:981–988, 2014.

[47] K. O. H. M. Dundas, M. T. P. Beerepoot, M. Ringholm, S. Reine, R. Bast, N. H. List, J. Kongsted, K. Ruud, and J. M. H. Olsen. Harmonic Infrared and Raman Spectra in Molecular Environments using the Polarizable Embedding Model. J. Chem. Theory Comput., submitted.

[48] J. M. H. Olsen and J. Kongsted. Molecular properties through polarizable embedding. In *Adv. Quantum Chem.*, pages 107–143. Elsevier, 2011.

[49] P. J. Olver. *Applications of Lie groups to differential equations*, volume 107. Springer Science & Business Media, 2000.

[50] W. P. Johnson. The Curious History of Faà di Bruno's Formula. *Am. Math. Monthly*, 109(3):217–234, 2002.

[51] R. A. Brualdi. Introductory Combinatorics, forth edition, 2004.

[52] S. Coriani, T. Helgaker, S. Høst, B. Jansik, P. Jørgensen, J. Kauczor, T. Kjaergaard, K. Kristensen, J. Olsen, S. Reine, et al. Lsdalton, a linear scaling molecular electronic structure program, release dalton2018 (2018), see http://daltonprogram. org. 2018.

[53] S. Reine, E. Tellgren, and T. Helgaker. A unified scheme for the calculation of differentiated and undifferentiated molecular integrals over solid-harmonic Gaussians. *Phys. Chem. Chem. Phys.*, 9:4771–4779, 2007.

[54] J. Bloino, M. Biczysko, and V. Barone. General perturbative approach for spectroscopy, thermodynamics, and kinetics: Methodological background and benchmark studies. *J. Chem. Theory Comput.*, 8(3):1015–1036, 2012.

[55] H. Li and J. H. Jensen. Partial Hessian vibrational analysis: the localization of the molecular vibrational energy and entropy. *Theor. Chem. Acc.*, 107(4):211–219, April 2002.

[56] S. J and J. D. Head. Theoretical investigation of molecular water adsorption on the Al(111) surface. *Surf. Sci.*, 318(1-2):204–216, October 1994.

[57] M. D. Calvin, J. D. Head, and S. Jin. Theoretically modelling the water bilayer on the Al(111) surface using cluster calculations. *Surf. Sci.*, 345(1-2):161–172, January 1996.

[58] J. Bloino, M. Biczysko, and V. Barone. Anharmonic effects on vibrational spectra intensities: infrared, Raman, vibrational circular dichroism, and Raman optical activity. *J. Phys. Chem. A*, 119(49):11862–11874, 2015.

[59] N. C. Craig and S. V. Krasnoshchekov. Vibrational spectroscopy of tolane; Coriolis coupling between Raman-active modes of g symmetry. *Mol. Phys.*, 117(9-12):1059–1068, 2019.

[60] J. Vázquez and J. F. Stanton. Treatment of Fermi resonance effects on transition moments in vibrational perturbation theory. *Mol. Phys.*, 105(1):101–109, 2007.

[61] J. Bloino. Personal communication, 2020.

[62] K. Aidas et al. The Dalton quantum chemistry program system. *WIREs Comput. Mol. Sci.*, 4(3):269–284, 2014.

[63] J. M. H. Olsen et al. Dalton Project: A Python platform for molecular-and electronic-structure simulations of complex systems. *J. Chem. Phys.*, 152(21):214115, 2020.

[64] J. M. H. Olsen, K. O. H. M. Dundas, and M. Ringholm. FraME: Fortran library for Fragment-based Multiscale Embedding (development version), **2020**. Available at https://gitlab.com/FraME-projects/FraME.

[65] S. Coriani, S. Høst, B. Jansík, L. Thøgersen, J. Olsen, P. Jørgensen, S. Reine, F. Pawłowski, T. Helgaker, and P. Sałek. Linear-scaling implementation of molecular response theory in self-consistent field electronic-structure theory. *J. Chem. Phys,*, 126(15):154108, 2007.

[66] B. Gao. QcMatrix Version 0.1.0, 2015. https://gitlab.com/bingao/qcmatrix.

[67] K. O. H. M. Dundas, M. Ringholm, Y. Cornaton, and B. Ofstad. SpectroscPy: The Python Package for Vibrational Spectroscopy. **2020**. https://doi.org/10.5281/zenodo.4099152.

[68] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28(1):235–242, 2000. rcsb.org.

[69] Avogadro: an open-source molecular builder and visualization tool. version 1.1.1. http://avogadro.cc/.

[70] M. Hanwell and et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, 4(1):17, 2012.

[71] Molecular formats converter. https://www.webqc.org/molecularformatsconverter.php.

[72] J. M. H. Olsen. PyFraME: Python framework for Fragment-based Multiscale Embedding (version 0.2.0), **2018**, DOI: 10.5281/zenodo.1443314. Available at https://gitlab.com/FraME-projects/PyFraME.

[73] O. Vahtras. LoProp for Dalton (version 0.2.2), **2017**. Available at https://github.com/vahtras/loprop.

[74] Dalton: a molecular electronic structure program (v2018.2), **2019**. See https://daltonprogram.org.

[75] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, 1988.

[76] C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, 1988.

[77] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and Frisch M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98:11623, 1994.

[78] S. H. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58:1200–1211, 1980.

[79] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98(45):11623–11627, 1994.

[80] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self-consistent molecular orbital methods. XII. further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56(5):2257–2261, March 1972.

[81] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople. Self-consistent molecular orbital methods. XXIII. a polarization-type basis set for second-row elements. *J. Chem. Phys.*, 77(7):3654–3665, October 1982.

[82] T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. V. R. Schleyer. Efficient diffuse function-augmented basis sets for anion calculations. III. the 3-21+G basis set for first-row elements, Li-F. *J. Comput. Chem.*, 4(3):294–301, 1983.

[83] W. L. Jorgensen and J. Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. USA*, 102:6665–6670, 2005.

[84] F. Jensen. Unifying general and segmented contracted basis sets. segmented polarization consistent basis sets. *Journal of chemical theory and computation*, 10(3):1074–1085, 2014.

[85] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. *J. Comput. Chem.*, 26:1701–1718, 2005.

[86] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435–447, 2008.

[87] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[88] C. Caleman, P. J.. van Maaren, M. Hong, J. S. Hub, L. T. Costa, and D. van der Spoel. Force field benchmark of organic liquids: Density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *J. Chem. Theory Comput.*, 8:61–74, 2012.

[89] D. van der Spoel, P. J. van Maaren, and C. Caleman. Gromacs molecule & liquid database. *Bioinformatics*, 28:752–753, 2012.

[90] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[91] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:014101, 2007.

[92] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[93] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103:8577–8593, 1995.

[94] K. O. H. M. Dundas, M. Ringholm, and J. M. H. Olsen. Dataset for "Analytic high-order geometric derivatives with polarizable embedding in a response function framework", November 2020.

[95] J. C. Decius and J. E. Rauch. *at the Ohio State Symposium on molecular Structure and Spectroscopy, paper 48*, 1959.

[96] R. W. Terhune, P. P. Maker, and C. M. Savage. Measurements of Nonlinear Light Scattering. *Phys. Rev. Lett.*, 14:681, 1965.

[97] S. J. Cyvin, J. E. Rauch, and J. C. Decius. Theory of Hyper-Raman Effects (Nonlinear Inelastic Light Scattering): Selection Rules and Depolarization Ratios for the Second-Order Polarizability. *J. Chem. Phys.*, 43:4083, 1965.

[98] O. Quinet, B. Champagne, and V. Rodriguez. Experimental and theoretical investigation of the Raman and hyper-Raman spectra of acetonitrile and its derivatives. *J. Chem. Phys.*, 124:244312, 2006.

[99] Kyungwon Kwak, Sangyob Cha, Minhaeng Cho, and John C. Wright. Vibrational interactions of acetonitrile: Doubly vibrationally resonant ir–ir–visible four-wave-mixing spectroscopy. *J. Chem. Phys.*, 117(12):5675–5687, 2002.

Supplementary information: Analytic High-Order Geometric Derivatives with Polarizable Embedding in a Response Theory Framework

Karen O. H. M. Dundas, Simen S. Reine, Bin Gao,
Kenneth Ruud, Jógvan Magnus Haugaard Olsen, Magnus Ringholm

December 24, 2020

Table 1: Bond lengths[a] and angles[b] for water in fullerene and in vacuum.

Feature	Vacuum	Fullerene
O-H bond lengths		
O-H1	0.9395	0.9426
O-H2	0.9395	0.9422
H-O-H bond angles		
Angle	106.4	105.5

[a] given in units of Å; [b] given in units of degrees;

Table 2: Harmonic and anharmonic wavenumbers and IR intensities for vacuous water; transitions from ground state.

State	Wavenumbers[a]		IR[b]	
	Harmonic	Anharmonic	Harmonic	Anharmonic
Fundamental				
1	4236	4046	86.293	81.551
2	4134	3971	17.660	15.009
3	1742	1672	90.467	93.166
Overtone				
1	8472	8001	-	0.128
2	8268	7863	-	0.553
3	3484	3303	-	0.854
Combination				
2 1	8370	7861	-	3.441
3 1	5978	5672	-	4.705
3 2	5876	5628	-	0.029

[a] given in units of cm^{-1}; [b] given in units of $km \cdot mol^{-1}$

Table 3: Harmonic and anharmonic wavenumbers and IR intensities for water in fullerene; transitions from ground state.

State	Wavenumbers[a]		IR[b]	
	Harmonic	Anharmonic	Harmonic	Anharmonic
Fundamental				
1	4173	3975	145.518	145.624
2	4087	3917	39.421	37.368
3	1738	1666	118.845	123.790
Overtone				
1	8345	7858	-	0.055
2	8174	7753	-	0.570
3	3475	3294	-	0.899
Combination				
2 1	8260	7732	-	2.805
3 1	5910	5593	-	5.133
3 2	5825	5567	-	0.031

[a] given in units of cm^{-1}; [b] given in units of $km \cdot mol^{-1}$

Table 4: Harmonic and anharmonic wavenumbers and IR intensities for vacuous acetonitrile; transitions from ground state. Only transitions with intensities of 0.3 km·mol^{-1} or higher are shown.

State	Wavenumbers[a]		IR[b]	
	Harmonic	Anharmonic	Harmonic	Anharmonic
Fundamental				
1	3277	3144	3.224	3.996
2	3277	3144	3.220	3.991
3	3201	3139	4.876	5.425
4	2590	2560	17.470	18.159
5	1595	1537	10.205	7.682
6	1595	1537	10.194	7.649
8	1160	1136	1.177	1.204
9	1160	1136	1.175	1.202
10	960	957	6.333	5.153
11	428	423	1.535	1.268
12	428	423	1.534	1.267
Overtone				
1	6553	6230	-	0.527
2	6553	6230	-	0.527
11	856	842	-	1.019
12	856	842	-	1.019
Combination				
2 1	6553	6254	-	0.378
5 1	4871	4692	-	0.359
5 2	4871	4673	-	0.358
6 1	4871	4673	-	0.357
6 2	4871	4692	-	0.360
7 1	4811	4638	-	0.376
7 2	4811	4638	-	0.375
10 4	3550	3517	-	1.070
11 8	1588	1557	-	1.694
11 9	1587	1557	-	1.736
12 8	1587	1573	-	1.754
12 9	1587	1573	-	1.698

[a] given in units of cm^{-1}; [b] given in units of km·mol^{-1}

Table 5: Harmonic and anharmonic wavenumbers and IR intensities acetonitrile in chloroform; transitions from ground state. Only transitions with intensities of 0.3 km·mol^{-1} or higher are shown.

State	Wavenumbers[a]		IR[b]	
	Harmonic	Anharmonic	Harmonic	Anharmonic
Fundamental				
1	3287	3156	0.890	1.288
2	3283	3151	1.401	1.957
3	3206	3140	1.716	1.965
4	2581	2551	35.564	36.971
5	1595	1540	12.548	10.454
6	1592	1538	11.625	11.108
8	1164	1140	1.732	1.825
9	1160	1136	1.772	1.843
10	955	955	8.451	6.536
11	443	438	2.518	2.139
12	436	432	2.428	2.061
Overtone				
1	6575	6254	-	0.531
2	6565	6246	-	0.521
11	885	870	-	1.395
12	873	859	-	1.646
Combination				
2 1	6570	6273	-	0.358
5 1	4882	4687	-	0.371
5 2	4878	4692	-	0.381
6 1	4879	4694	-	0.389
6 2	4875	4677	-	0.409
7 1	4824	4652	-	0.362
7 2	4819	4646	-	0.350
10 4	3536	3505	-	1.575
11 8	1606	1584	-	1.130
11 9	1603	1579	-	0.896
12 8	1600	1577	-	1.572
12 9	1597	1573	-	1.771

[a] given in units of cm^{-1}; [b] given in units of km·mol^{-1}

Table 6: Harmonic and anharmonic wavenumbers and IR intensities acetonitrile in acetonitrile; transitions from ground state. Only transitions with intensities of 0.3 km·mol^{-1} or higher are shown.

State	Wavenumbers[a]		IR[b]	
	Harmonic	Anharmonic	Harmonic	Anharmonic
Fundamental				
1	3301	3175	1.304	1.239
2	3285	3156	1.690	2.017
3	3210	3146	2.536	2.902
4	2571	2541	49.365	51.144
5	1600	1547	13.966	14.057
6	1595	1541	12.966	12.711
7	1536	1503	0.674	0.976
8	1169	1145	2.410	2.620
9	1161	1137	2.188	2.295
10	949	951	9.626	8.946
11	449	446	2.606	2.330
12	443	439	2.863	2.505
Overtone				
1	6601	6287	-	0.473
2	6570	6259	-	0.478
11	899	885	-	0.697
12	887	869	-	1.281
Combination				
5 1	4900	4713	-	0.440
5 2	4885	4703	-	0.427
6 1	4896	4712	-	0.478
6 2	4880	4685	-	0.386
7 1	4837	4670	-	0.380
7 2	4821	4652	-	0.342
8 7	2705	2643	-	0.301
9 7	2697	2634	-	0.364
10 4	3520	3483	-	1.947
11 8	1618	1596	-	1.851
11 9	1610	1585	-	0.869
12 8	1612	1591	-	0.762
12 9	1604	1580	-	0.814

[a] given in units of cm^{-1}; [b] given in units of km·mol^{-1}

Table 7: Harmonic and anharmonic wavenumbers and IR intensities acetonitrile in water; transitions from ground state. Only transitions with intensities of 0.3 km·mol^{-1} or higher are shown.

State	Wavenumbers[a]		IR[b]	
	Harmonic	Anharmonic	Harmonic	Anharmonic
Fundamental				
1	3307	3180	2.516	2.353
2	3292	3161	1.304	1.490
3	3213	3142	2.701	2.720
4	2572	2543	114.322	117.706
5	1596	1545	15.291	13.486
6	1587	1536	13.670	12.415
7	1531	1498	1.347	1.894
8	1167	1144	4.031	4.454
9	1158	1134	4.601	5.005
10	951	960	13.734	14.810
11	467	463	8.138	8.216
12	458	454	6.532	6.087
Overtone				
1	6614	6298	-	0.459
2	6583	6269	-	0.462
5	3192	3081	-	0.378
6	3174	3052	-	0.359
11	933	920	-	0.368
12	915	893	-	0.356
Combination				
5 1	4903	4711	-	0.450
5 2	4888	4696	-	0.428
6 1	4894	4703	-	0.481
6 2	4879	4680	-	0.428
7 1	4838	4669	-	0.359
7 4	4104	4041	-	0.319
8 7	2699	2638	-	0.343
9 7	2689	2627	-	0.449
10 4	3523	3482	-	2.851
10 7	2482	2427	-	0.612
11 8	1634	1610	-	1.869
11 9	1625	1601	-	2.221
12 8	1625	1602	-	2.529
12 9	1616	1592	-	1.828

[a] given in units of cm^{-1}; [b] given in units of km·mol^{-1}

Paper III

Dalton Project: A Python platform for molecular- and electronic-structure simulations of complex systems

J. M. H. Olsen, S. Reine, O. Vahtras, E. Kjellgren, P. Reinholdt, K. O. H. Dundas, X. Li, J. Cukras, M. Ringholm, E. D. Hedegård, R. Di Remigio, N. H. List, R. Faber, B. N. C. Tenorio, R. Bast, T. B. Pedersen, Z. Rinkevicius, S. P. A. Sauer, K. V. Mikkelsen, J. Kongsted, S. Coriani, K. Ruud, T. Helgaker, H. J. Aa. Jensen, and P. Norman

J. Chem. Phys. **152** (2020), 214115.

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

Dalton Project: A Python platform for molecular- and electronic-structure simulations of complex systems

Cite as: J. Chem. Phys. 152, 214115 (2020); doi: 10.1063/1.5144298
Submitted: 3 January 2020 • Accepted: 29 March 2020 •
Published Online: 5 June 2020

Jógvan Magnus Haugaard Olsen,[1,a] Simen Reine,[2,b] Olav Vahtras,[3] Erik Kjellgren,[4] Peter Reinholdt,[4] Karen Oda Hjorth Dundas,[1] Xin Li,[3] Janusz Cukras,[5] Magnus Ringholm,[1,3] Erik D. Hedegård,[6] Roberto Di Remigio,[1] Nanna H. List,[7] Rasmus Faber,[8] Bruno Nunes Cabral Tenorio,[8] Radovan Bast,[1] Thomas Bondo Pedersen,[2] Zilvinas Rinkevicius,[3,9] Stephan P. A. Sauer,[10] Kurt V. Mikkelsen,[10] Jacob Kongsted,[4] Sonia Coriani,[8] Kenneth Ruud,[1] Trygve Helgaker,[2] Hans Jørgen Aa. Jensen,[4] and Patrick Norman[3,c]

AFFILIATIONS

[1] Department of Chemistry, Hylleraas Centre for Quantum Molecular Sciences, UiT The Arctic University of Norway, N-9037 Tromsø, Norway
[2] Department of Chemistry, Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, N-0315 Oslo, Norway
[3] Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden
[4] Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark
[5] Department of Chemistry, University of Warsaw, 02-093 Warsaw, Poland
[6] Division of Theoretical Chemistry, Lund University, SE-223 62 Lund, Sweden
[7] Department of Chemistry and the PULSE Institute, Stanford University, Stanford, California 94305, USA and SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA
[8] DTU Chemistry, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
[9] Department of Physics, Faculty of Mathematics and Natural Sciences, Kaunas University of Technology, LT-51368 Kaunas, Lithuania
[10] Department of Chemistry, University of Copenhagen, DK-2100 Copenhagen Ø, Denmark

Note: This article is part of the JCP Special Topic on Electronic Structure Software.
[a] **Electronic mail:** jogvan.m.olsen@uit.no
[b] **Electronic mail:** simen.reine@kjemi.uio.no
[c] **Author to whom correspondence should be addressed:** panor@kth.se

ABSTRACT

The *Dalton Project* provides a uniform platform access to the underlying full-fledged quantum chemistry codes Dalton and LSDalton as well as the PyFraME package for automatized fragmentation and parameterization of complex molecular environments. The platform is written in Python and defines a means for library communication and interaction. Intermediate data such as integrals are exposed to the platform and made accessible to the user in the form of NumPy arrays, and the resulting data are extracted, analyzed, and visualized. Complex computational protocols that may, for instance, arise due to a need for environment fragmentation and configuration-space sampling of biochemical systems are readily assisted by the platform. The platform is designed to host additional software libraries and will serve as a hub for future modular software development efforts in the distributed Dalton community.

I. INTRODUCTION

More than 20 years have passed since the first version of the Dalton program[1] was released as a result of merging the separate HERMIT, SIRIUS, ABACUS, and RESPONS codes that implemented one- and two-electron integrals, wavefunctions, energy derivatives, and response theory, respectively. Later, the adopted monolithic code development structure turned out to be prohibitively difficult to sustain, and it was interrupted with the release of the atomic-orbital (AO) based linear-scaling initiative as a separate executable named LSDalton.[2] By the time of the Dalton paper in 2014,[3] the two codes represented a powerful general-purpose program system and provided users with access to the most relevant and standard electronic structure theory methods and, moreover, a vast amount of molecular properties. In 2017, all past and present authors of the Dalton and LSDalton codes unanimously voted in favor of open-sourcing the codes under the GNU Lesser General Public License version 2.1 (LGPLv2.1). In the present work, we will briefly recapitulate the functionalities of the codes and detail some of the developments provided in Dalton suite releases from 2015 until today, including the Dalton2020 release. With inspiration from the Molecular Sciences Software Institute (MolSSI) project,[4,5] we also take the opportunity to initiate a transition in the Dalton software engineering practices and we signal this paradigm shift by referring to the Dalton community effort as the *Dalton Project* (DP) initiative.[6] From the developer's perspective, we are taking steps to make it easier to develop, sustain, and maintain a large general-purpose software ecosystem for first-principles quantum molecular modeling of complex systems, and from the user's perspective, we are modifying the design of the user interface to enable new access and interaction patterns.

The general design strategy for the DP platform is that of software modularity[7–9] and based on a hybrid programming language approach, as illustrated in Fig. 1. We introduce an upper layer written in Python with support from specialized libraries, such as NumPy,[10] SciPy,[11] and MPI4Py.[12] This layer is hardware-aware and capable of managing computer resources, handling user interactivity, steering computation, and performing data processing of results. The lower layer contains libraries written in a language of choice based on the programmer's preference and the task to be addressed, but compute-intensive tasks will typically be performed by libraries written in Fortran, C, or C++. The two layers interact by any one of three means of communication, namely, conventional file input/output (I/O), Python bindings, e.g., through CFFI (C Foreign Function Interface)[13] or pybind11,[14] or pure Python module import. In this scheme, we view the Dalton and LSDalton executables as libraries serving the DP platform, and although further modular library decomposition would be desirable, it is hampered by code legacy and entanglement. More important than offering this new perspective, however, the DP platform encourages future development to be made in the form of modules with clear and specific tasks (or subtasks) that undergo strict unit testing. Modules, or coherent sets of modules, build up libraries that are developed, maintained, and released independently from one another such that the DP ecosystem will see more of a continuous evolution as compared to conventional monolithic program releases. Regarding communication, it is our ambition for the ecosystem to move toward libraries that provide clear application programming interfaces (APIs) and native bindings to Python. The latter allows importing such libraries directly into Python scripts or interactive sessions, enabling fast development, read–eval–print loop (REPL) style, without sacrificing performance. We believe that this software development model will serve us well as we constitute a distributed community of contributors belonging to network nodes with different scientific objectives and timelines.

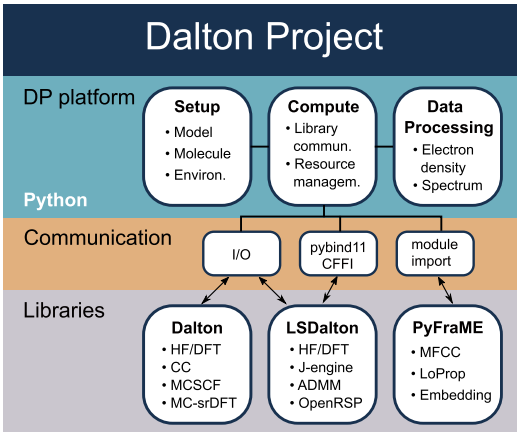


FIG. 1. Overview of the Dalton Project platform structure.

Within the field of quantum chemistry, the adoption of more modern software engineering strategies with APIs written in Python is in vogue at the moment, and we have been strongly influenced by (i) the Psi4NumPy project that exposes efficient computational kernels from the Psi4 program[15] to enable quick NumPy prototyping of novel science[16] and (ii) the PySCF program that, primarily in Python, implements self-consistent field (SCF) and post-Hartree–Fock (post-HF) electronic structure theory for finite and periodic systems.[17] Moreover, a source of inspiration as well as practical experience for the present work is provided by the VeloxChem project (and program)[18] that, with a hybrid Python/C++ programming model, implements real and complex response theory[19] at the SCF level of theory for execution in high-performance computing (HPC) cluster environments. In VeloxChem, Python is used for a split message passing interface (MPI) communicator management of large-scale distributed hardware resources with an anticipation of heterogeneous cluster nodes to become a future reality. Without noticeable sacrifice in computational efficiency or program execution stability, the higher-level quantum chemical methods and iterative linear response equation solvers are implemented in Python with the use of NumPy and underlying threaded math kernel libraries. With this as background, we have gained sufficient confidence to steer our project into a new direction as far as software engineering practices are concerned.

Our presentation is organized as follows: In Sec. II, we briefly mention some of the key features in Dalton and LSDalton that have already been presented[3] and provide a more detailed description

of novel functionalities that have been added thereafter. Moreover, we also present the features of the PyFraME package[20] for the handling of complex chromophore environments. In Sec. III, we give a more comprehensive view of the design of the DP platform as well as provide a concrete illustration of new library-access patterns and program execution practices for the user. In Sec. IV, we present six concrete examples of DP platform runs before closing with an outlook into the future for the *Dalton Project*.

II. CAPABILITIES OF DP PLATFORM LIBRARIES

A. Dalton and LSDalton up until 2014

In 2014, a presentation of the Dalton program system, including the Dalton and LSDalton codes, was published,[3] and functionalities listed in this presentation are, of course, still available and therefore only briefly mentioned here. The two codes are primarily written in Fortran, but parts involving density functional theory (DFT) kernels are mostly written in C. In Dalton, the routines for correlated wavefunction calculations are implemented only for serial execution—but can be linked to standard threaded linear algebra libraries—whereas the self-consistent field (SCF), i.e., HF and DFT, routines are implemented for parallel execution using MPI. LSDalton, on the other hand, comes with a native hybrid OpenMP/MPI parallelization scheme that enables shared memory data handling on central processing unit (CPU) sockets and/or compute nodes. None of the two codes come with support for hardware acceleration, such as general-purpose graphical processing units (GPUs).

The common foundation for the Dalton and LSDalton quantum chemistry programs is that of a nonrelativistic Hamiltonian, basis sets expanded in localized Gaussian AOs, and multi-electron reference states expanded in spin-restricted determinants or configuration-state functions. Relativistic corrections to the zeroth-order one-electron Hamiltonian are available in Dalton in terms of the spin-free second-order Douglas–Kroll–Hess (DKH2) Hamiltonian and effective-core potentials (ECPs). As a perturbative correction to the Hamiltonian, Dalton also offers an implementation of the full Breit–Pauli spin–orbit operator.

LSDalton provides efficient acceleration techniques for SCF-based property calculations and an implementation of the linear-scaling divide–expand–consolidate (DEC) scheme for second-order Møller–Plesset (MP2) and coupled cluster (CC) energy calculations. The code was initially developed to alleviate the restrictions of the Dalton code for calculations on large systems by introducing linear-scaling AO-based SCF and response capabilities based on an exponential *ansatz* of the AO density matrix.

Dalton provides implementations of most of the standard electronic-structure methods, including SCF, MP2, a hierarchy of CC methods [CC2, CCSD, CCSDR(3), CC3, and CCSD(T)], configuration interaction (CI), and multi-configurational SCF (MCSCF) based on the generalized active space (GAS) concept. MCSCF wavefunctions are optimized with a robust trust-region-based second-order approach.

Molecular gradients and Hessians are determined analytically for SCF and MCSCF reference states, and analytic gradients are also available at the levels of MP2, CC2, CCSD, and CCSD(T). In the absence of analytic gradients and Hessians, Dalton can determine these quantities by numerical differentiation and thereby offers an extensive functionality for exploring potential energy surfaces. The combination of geometric and electric-field perturbations allows for calculations of infrared (IR) and Raman intensities. Analytic linear and nonlinear response functions describing the interactions with external and internal (in general, time-dependent) electromagnetic fields are implemented for the entire selection of electronic-structure methods and enable simulations of a plethora of spectroscopies, too rich to be listed here. At this time, Dalton also included the means for structure-less and atomistic descriptions of chromophore environments through the polarizable continuum model (PCM) and polarizable embedding (PE) model, respectively.

B. Added features in Dalton

The functionalities of Dalton have been expanded in several directions. Here, we provide a summary of selected new features. To bring some structure and order into these developments, we have chosen to divide them into the three categories: (i) electronic-structure theory, (ii) spectroscopy simulations, and (iii) environment modeling. In the first one, we list general quantum-chemical method developments providing new means to describe the electronic structure of ground and excited states. In the second, we describe developments that are more specifically targeting and enabling simulations of certain spectroscopies. Such simulations are connected with certain electronic-structure theory methods and typically also environment models, but the primary objective of the development has been the spectroscopy at hand. In the third, we present approaches aimed at improving the effective description of the chromophore environment. These developments are, of course, made in combination with specific electronic-structure methods, but the environment is at focus.

1. Electronic-structure theory

Based on a range-separated Hamiltonian as proposed by Savin,[21,22] a rigorous combination can be made of wavefunction and density-functional theories for the treatment of the long- and short-range electron–electron Coulomb interactions, respectively. In Dalton, this approach has been implemented at the level of MP2,[23] CI,[24,25] MCSCF,[26–28] and NEVPT2[29] wavefunction theories, and it is now made available in the Dalton2020 release. In conjunction with MCSCF, the main idea is that static (or strong) electron correlation can be effectively accounted for by means of typically quite short determinant expansions of the wavefunction at the same time as dynamic electron correlation can be effectively accounted for by means of DFT with its low computational cost. The resulting method is referred to as multi-configurational short-range DFT (MC-srDFT), and it is available for closed-shell and open-shell systems.[28] Apart from calculations of energies, linear-response properties are available for both singlet and triplet perturbations.[30–32] More details are provided in Sec. IV B where an example is provided in terms of the calculation of the ultraviolet–visible (UV/Vis) absorption spectrum of a retinylidene Schiff base chromophore.

Using Löwdin's inner projection in conjunction with a one- and two-electron excitation operator manifold and an MP2 reference state, the second-order polarization propagator approximation (SOPPA) arises as a means to address the electronic

structure of excited states. Modifications of the original form of this approach have been implemented, including the SOPPA(SCS-MP2) and SOPPA(SOS-MP2) models[33] where spin-component-scaled[34,35] and scaled opposite-spin[36] versions of the Møller–Plesset correlation coefficients are employed. Going instead toward approximations of the SOPPA model, the random phase approximation with second-order non-iterative doubles corrections model [RPA(D)][37] has been extended to triplet excitations,[38] and a similarly derived higher RPA with non-iterative doubles corrections model [HRPA(D)] has been implemented.[38] The RPA(D) and HRPA(D) models are enabled for calculations of not only transition properties but also linear response functions.[39]

Furthermore, with regard to CC approaches, Dalton also offers a new and more efficient implementation of the CC3 model for ground- and excited-state energies,[40] although it does not support full Abelian point-group symmetry, it reduces the computational cost.

2. Spectroscopy simulations

Applying the Liouville equation to pure states in the density-matrix formalism of quantum mechanics has been shown to be equivalent to applying the Ehrenfest theorem to state-transfer operators in Hilbert space, thereby leading to a means to phenomenologically introduce relaxation mechanisms into wavefunction theories.[19,41,42] The resulting complex polarization propagator (CPP) theory defines frequency-dependent response functions for exact and approximate states that are physically sound in all regions of the spectrum, resonant as well as conventional nonresonant, and also x-ray as well as conventional UV/Vis. These response functions fulfill the Kramers–Kronig relations with real and imaginary parts that are associated with separate dispersive and absorptive spectroscopies, such as optical rotatory dispersion (ORD)[43] and electronic circular dichroism (ECD).[44]

Extensions of the CPP theory have been made to allow for the description of nonlinear external-field interactions,[41,45] and the latest release of the Dalton program also offers CPP/DFT simulations of resonant-enhanced hyper-Rayleigh scattering (HRS),[46,47] magnetic circular dichroism (MCD),[48,49] magneto-chiral dichroism (MChD) and birefringence (MChB) dispersion,[50] nuclear spin-induced optical rotation (NSOR) and dichroism (NSCD),[51] and two-photon absorption (TPA) cross sections.[45,52]

Core excitation processes are associated with large valence-electron relaxation and polarization effects that, in a polarization propagator or response theory approach, require multi-electron excited configurations to be properly accounted for.[53] Along this line, Dalton provides a hierarchy of CC methods to model a variety of x-ray spectroscopies including near-edge x-ray absorption fine structure (NEXAFS),[54–57] photo-electron spectroscopy (PES),[56–59] transient x-ray absorption spectroscopy (TRXAS),[60,61] and resonant inelastic x-ray scattering (RIXS).[62] The referred-to hierarchy of CC methods includes the CCS, CC2, and CCSD levels of theory, but core-excitation and core-ionization energies are also available for the CCSDR(3) and CC3 approximations. Both singlet and triplet excited states are encompassed, and the latter are obtained in a spin-adapted formalism.[61] The core–valence separation (CVS) approximation has been made available to decouple core and valence excited states. It can be applied either at the excited-state level only[56,57] or both during the determination of the ground state and excited states in a frozen-core variant (fc-CVS).[63] An example illustration of a DP platform XAS calculation using the CVS approximation is provided in Sec. IV F.

3. Modeling of chromophore environments

The capabilities in Dalton for including effects from a molecular environment have been extended in several directions. The PCM for efficient modeling of bulk solvent effects can now also be performed at the SOPPA level.[64] In this PCM-SOPPA/RPA model, the static solvent contributions are treated at the SOPPA level, while the dynamic solvent contributions are evaluated at the time-dependent HF level. The PCM model can also be used in combination with the MC-srDFT method.

The PE model[65,66] is a fragment-based (semi-)quantum–classical scheme designed for efficient and accurate inclusion of environment effects in calculations of spectroscopic properties of large and complex molecular systems. The environment is included effectively through an embedding potential whose parameters consist of distributed multipoles and polarizabilities, both of which are derived from quantum-mechanical calculations on the individual fragments that make up the environment. The PyFraME package, which is made available on the DP platform and is described in Sec. II D, can be used to automatize the generation of the embedding-potential parameters. The PE model is implemented in the Polarizable Embedding library (PElib)[67] based on an AO density-matrix-driven formulation, which facilitates a loose-coupling modular implementation in host programs. The PElib was included in the Dalton2013 release, but at that time, it was limited to PE-HF and PE-DFT.[65,66] Since then the implementation has been extended to PE-CC [specifically, PE-CC2, PE-CCSD, and PE-CCSDR(3)],[68] PE-MCSCF,[69] PE-MC-srDFT,[70] and PE-SOPPA.[71] The Dalton2020 release supports linear-, quadratic-, and cubic-response properties for PE-HF/DFT,[66] while PE-CC is limited to linear- and quadratic-response properties, and only linear-response properties are available for PE-MCSCF and PE-MC-srDFT. For PE-HF/DFT, it is also possible to compute properties based on resonant-convergent response theory.[72] London AOs (LAOs) are supported for magnetic linear-response properties that involve a single derivative with respect to a magnetic field.[73] The capabilities have also been extended to enable analytic quantum-mechanical molecular gradients at the PE-HF/DFT level, thus enabling geometry optimization of the core quantum region embedded in a fixed polarizable environment.[74] Local-field effects may also be included in PE-HF/DFT calculations where they are termed effective external field (EEF) effects.[75,76] Electronic energy transfer (EET) couplings can be calculated based on the PE model, including both direct and environment-induced contributions, and using QFITLIB[77] to derive transition-density-fitted multipoles.[78] Bulk solvation effects can be included through the FixSol conductor-like solvation model using the FIXPVA2 cavity tessellation scheme.[79,80] An overview of the developments related to the PE model can be found in Ref. 81, while a tutorial review is available in Ref. 82.

The PE model, and classical models in general, does not include Pauli repulsion between the chromophore and its environment. Such models can therefore suffer from so-called electron spill-out, where the electron density of the chromophore leaks out into the environment, thus causing an over-stabilization of the ground state

and, in particular, the excited states of the embedded chromophore. Negatively charged chromophores or excited states of even partial Rydberg-like character are especially susceptible.[83,84] The polarizable density embedding (PDE) model has been formulated to improve the electrostatic interactions between the chromophore and its environment and to address the electron spill-out issue.[85,86] In this model, the permanent charge distribution of the fragments in the environment is described by their full electronic densities, thus avoiding divergences of the multipole expansions, while still keeping the distributed polarizabilities to efficiently account for polarization effects. In addition, the PDE model contains a Huzinaga–Cantu-like projection operator[87] that models Pauli repulsion effects and thereby effectively prevents electron spill-out. The PDE model has been implemented in PElib using the same AO density-matrix-driven formulation as for the PE model. It can therefore straightforwardly be combined with the same DFT and wavefunction methods as the PE model both in terms of ground-state and response calculations, with the exception of LAOs and analytic molecular gradients.

The latest Dalton release has also received basic frozen density embedding (FDE)[88,89] capability. The FDE implementation enables import of a static embedding potential that has been pre-calculated on a numerical integration grid by another code that implements FDE.[90,91] A matrix representation of the embedding potential is constructed based on the grid and added to the one-electron Fock matrix. The implementation can thus be used with all available DFT and wavefunction methods in Dalton.[92,93]

C. Added features in LSDalton

1. Integral evaluation

Integrals sit at the heart of any quantum chemistry program, both when it comes to computational performance and available methods and properties. The development of an efficient and flexible integral-evaluation code has therefore been essential to the development of LSDalton. Since 2014,[3] four main integral developments have been added: high-order derivative integrals (HODI), integrals and differentiated integrals for embedding techniques involving interaction with point charges and higher-order multipoles, acceleration of the exchange contribution through developments of the auxiliary-density-matrix method (ADMM),[94] and interface with the XCFun library of DFT exchange–correlation (XC) functionals.[95,96] The XCFun library is based on forward-mode automatic differentiation[97] and can therefore generate arbitrary-order derivatives of these functionals.

The one- and two-electron HODI implementation employs the solid-harmonic Hermite scheme of Ref. 98, allowing for a unified scheme for undifferentiated and differentiated integrals by expanding the solid-harmonics in Hermite rather than Cartesian Gaussians; differentiation merely increments one of the quantum numbers of the Hermite Gaussians, whereas differentiation of Cartesian Gaussians gives linear combinations of Cartesian Gaussians. The HODI integrals have been extended to allow interactions with general-order point multipoles (charges, dipoles, and so on) needed for classical embedding techniques.

The exchange contribution is the main computational bottleneck in hybrid DFT calculations. The development of efficient and accurate acceleration techniques for the exchange contribution will thus greatly improve overall DFT timings and increase the scope and applicability of DFT in general. One such approach is the ADMM,[94,99,100] where the time-critical exchange contribution is instead evaluated in a smaller basis, and corrected with the difference between the local generalized-gradient approximation (GGA) exchange in the full and the small basis. The ADMM has been implemented in LSDalton with different variants for the projection to the smaller basis and GGA correction functional options,[99] and with tailored auxiliary basis sets (admm-*n*) for the pcseg-*n* and aug-pcseg-*n* basis sets.[100]

2. Exploiting the locality of electron correlation

The DEC[101–103] strategy employs highly local orbitals[104,105] to recover the inherent locality of dynamical correlation for large molecules in a linear-scaling fashion. Over the last few years, the DEC framework has been extensively developed and now includes resolution-of-the-identity (RI) accelerated MP2 (DEC-RI-MP2[106–108]), Laplace-transformed RI-MP2 (DEC-LT-RI-MP2[109]), CC theory through DEC-CCSD and DEC-CCSD(T),[110] and through the multilayer DEC framework ML-DEC,[111] which allows for efficient calculations by systematic treatment of the *pair-fragment* at different levels of theory. In addition to energies, densities, and electrostatic potentials, gradients are available at the DEC-MP2 and DEC-RI-MP2 levels,[102,107] and excitation energies are available through the local framework for calculating excitation energies (LoFEx)[112–114] and the correlated natural transition orbital framework for a low-scaling excitation energy (CorNFLEx) approach.[115] Due to the embarrassingly parallel nature of the DEC scheme, excellent scalability to a large number of CPU cores is possible. As an example, a DEC-RI-MP2/cc-pVDZ gradient calculation of the insulin molecule (787 atoms and 7604 basis functions) finished in less than 10 h using 32 000 cores (2000 nodes, each with 16 cores, on the Titan supercomputer).[107]

3. Molecular properties

Several property developments have been undertaken since 2014, including quasi-Newton transition-state optimization, the high-order path-expansion (HOPE)[116] method for improved geometry-optimization steps, automated counterpoise correction, and the same-number-of-optimized-parameters (SNOOP)[117,118] scheme as an improved alternative to the counterpoise correction, and nuclear-selected NMR shielding,[119] to mention a few.

On a longer-term development line, LSDalton has been interfaced with OpenRSP,[120] to allow, in principle, arbitrary-order molecular properties. OpenRSP is an open-ended response-theory library that manages the generation and solution of the response equations needed for the evaluation of arbitrary-order response properties. The current implementation in LSDalton enables the calculation of a sizable selection of (mixed) electrical and geometrical properties for HF and DFT, for the latter also involving an interface to the XCFun and XCint[121] libraries. This includes properties related to IR, Raman, and hyper-Raman spectral intensities, molecular gradients, Hessians, and cubic force constants. The capabilities of the LSDalton/OpenRSP/XCint/XCFun combination are illustrated for the calculation of IR and Raman spectra of benzene through the DP platform in Sec. IV D.

For modeling of solvent effects, the PCMSolver[122] library for continuum electrostatic solvation has been interfaced to LSDalton. This implementation is available for SCF and electric-dipole response properties up to fourth order.

D. PyFraME: Python framework for Fragment-based Multiscale Embedding

PyFraME[20] is a Python package providing a framework for managing fragment-based multiscale embedding calculations. The basic principle of embedding models in quantum chemistry is the division of a molecular system into two domains: a central core region and its environment. The core region is described at the highest level of theory using DFT or a wavefunction method, while the effects from the environment are included effectively through an embedding potential. To manually set up embedding calculations of large and complex molecular systems can be highly complicated, tedious, and error-prone. This is especially true when considering that configuration-space sampling, e.g., through molecular dynamics (MD) simulations, is usually required, which, in turn, means that the procedure has to be repeated many times.

The highly flexible PyFraME package automatizes workflows, starting from the initial molecular structure to the final embedding potential. It enables the user to easily set up a multilayered description of the environment. Each layer can be described either by a standard embedding potential, i.e., using a predefined set of parameters, or by deriving the embedding-potential parameters based on first-principles calculations. For the latter, a fragmentation method is used to subdivide large molecular structures into smaller computationally manageable fragments. The number of layers, as well as the composition and level of theory used for each layer, can be fully customized.

The basic workflow consists of three main steps. First, a molecular structure is given as an input. Currently, PyFraME supports input files in the PDB format. The input file reader extracts information about the structure and composition of the system, and it also defines the basic units of the system, i.e., fragments. Small molecules would typically constitute a fragment on their own, but larger molecules are usually broken down into small computationally manageable fragments. For example, for proteins, a fragment would usually consist of an amino-acid residue, while for nucleic acids, it could be a nucleotide. The molecular system to be used for the embedding calculation is then built by extracting subsets from the full list of fragments according to specified criteria, such as name, chain ID, distance, or a combination thereof, and placed into separate regions. As mentioned above, any number of regions may be added, and each can be fully customized. Once the system has been built, the final step is the derivation of the embedding potential. Depending on the specifics, it may involve a large number of separate calculations on the individual fragments in order to compute the embedding-potential parameters. For large molecules, where the parameters cannot be computed directly, PyFraME uses a fragmentation method based on the molecular fractionation with conjugate caps (MFCC) approach[123] to derive the parameters. The individual fragment calculations are typically performed by Dalton and the LoProp Python package,[124,125] but this can be customized. The fragmentation of the system, fragment calculations, and subsequent joining of parameters to build the embedding potential are fully automatized and can make full use of large-scale HPC resources.

III. DP PLATFORM DESIGN AND FEATURES

The ultimate goal of the DP platform is to establish a flexible, robust, and uniformly accessible environment that can be used for both large-scale applications and to facilitate the development of novel methodology. The challenges for quantum chemistry today are far more complex than earlier, both concerning the complexity of the chemical systems, adaptation to HPC facilities, and the number of tools and approaches needed for applications. As a result, it becomes essential to be able to easily combine the tools and approaches in meaningful ways. The main motivation of the Dalton Project is to provide a platform that can be used to combine the functionality of the tools and methods that are developed by the individual research groups in the Dalton community.

For a long time, we have relied on a monolithic codebase (first Dalton and later also LSDalton) for the development of new computational methodology. These programs have served us well in the past, and we expect this to continue into the foreseeable future. It is clear, however, that the codebase has accumulated substantial technical debt. The tight coupling between the software modules is particularly problematic because it complicates optimization and modernization of even small pieces of code. Moreover, implementation of new methodology often requires unnecessarily high efforts and easily leads to additional technical debt. The risks of relying on a monolithic codebase are especially high when the codebase is maintained by a scientific community such as ours whose primary goal is to perform research. In recognition of this, and the fact that individual groups have different research aims and preferences in terms of software development, we have in later years moved toward a more distributed codebase. This has resulted in the development of a series of software libraries, such as Gen1Int,[126,127] OpenRSP,[120] PCMSolver,[122] PElib,[67] QcMatrix,[128] XCFun,[95,96] and XCint.[121] This has, to some degree, alleviated the problem of the monolithic codebase for some developments, but the main issues remained.

We have now taken the next step and moved completely to a distributed codebase with the Dalton Project, whose main task is to integrate and provide interoperability between the individual software libraries that are developed and maintained by the different nodes in our community. At the same time, however, we acknowledge that there is vast functionality developed in our community during the last few decades that we do not wish to abandon, which is primarily implemented in Dalton and LSDalton. The DP platform thus has to accommodate a wide variety of software from large monolithic programs with a wide range of features to small libraries that provide very specific functionality. The design of the platform has to take this into account in a sustainable manner, so that it can act as a platform for present-day use cases and, importantly, for future developments based on modern software engineering practices. Moreover, the DP platform must be able to exploit current HPC facilities and be prepared for the upcoming exascale supercomputers.

To meet our goals and requirements, we devised a platform structure that is illustrated in Fig. 1. At the top, we have the DP platform itself, written in pure Python (3.6+), that interfaces to external libraries through different communication mechanisms. Python was chosen as the platform language because of its extensibility, emphasis on code readability, and comprehensive standard library, as well

as a large number of specialized libraries. We note here that the term library is used liberally to signify any type of software that can be interfaced to the platform, including libraries in the traditional sense, program executables, and Python modules and packages. The libraries thus require very different means of communication, and to accommodate this, we provide three different mechanisms: file I/O, Python bindings, and pure Python imports. The file I/O communication mechanism is provided to make the vast functionality implemented in Dalton and LSDalton more accessible. In fact, most of the functionality provided on the DP platform currently involves Dalton and LSDalton, but we will gradually move toward using loosely coupled libraries written in pure Python or hybrid Python and Fortran, C, or C++, with the hybrid approach being used for the more compute-intensive numerical tasks. Initially, the DP platform will interface Dalton and LSDalton as well as PyFraME, all of which have been described in Sec. II. In the immediate future, we expect that many of the aforementioned libraries that are currently interfaced to Dalton and LSDalton will be interfaced directly to the DP platform.

The DP platform is a Python package with an API that consists of a set of classes and functions used to set up molecular systems, perform numerical calculations, and process data. Usage of the platform would typically consist of three stages used in succession, namely, *setup*, *compute*, and *data processing*. By separating the compute and data processing stage, the DP platform may be easily employed in large-scale application workflows in which a number of calculations are typically run first, by submitting them to a queuing system on a supercomputer, and subsequently data manipulation, analysis, and visualization are performed.

The setup stage consists of instantiation of one or more of the five base classes: `Molecule`, `Basis`, `QCMethod`, `Property`, and `Environment`, which are used in the compute and data processing stages. The classes have been designed to be library-agnostic so that they can be used and reused for all libraries. However, not all classes are necessarily needed. It depends on the specific type of calculation that is performed. For example, the first four (or all five if an environment model is used) are needed to run, e.g., a TPA calculation employing Dalton or LSDalton, whereas other functionality may only need some of them (e.g., more fine-grained functionality can be obtained from LSDalton, as illustrated in Sec. IV A).

The `Molecule` class contains information about the molecular structure, which can be a single atom, a molecule, a fragment of a molecule, or a set of molecules. It requires, as a minimum, that atomic elements and coordinates are defined. Reasonable defaults are used for other attributes, such as the total charge, spin multiplicity, atomic masses, atomic labels, and, if enabled, information related to point-group symmetry. The atomic elements, coordinates, and, optionally, labels, can be given either as a file, e.g., in XYZ format (optionally with atom labels in the fifth column), or as a string. The DP platform also provides a function that can read the standard molecule file format used by Dalton and LSDalton and return instances of the `Molecule` and `Basis` classes. The atomic labels are used in the `Basis` class as described below, but also to specify, e.g., ghost atoms and point charges.

The `Basis` class holds all basis-set information, which includes the main basis set and, optionally, auxiliary basis sets used for, e.g., RI and ADMM approximations. The basis set can be given either as a string, in which case the specified basis set is used for all atoms, or as a dictionary using atom labels as the keys and basis sets as the corresponding values. The basis sets are obtained from the basis set exchange (BSE) Python package.[129]

The `QCMethod` class specifies the method, for example, HF, DFT, MCSCF, and CC, together with any associated attributes, such as XC functional or definition of active space. Approximations used to compute Coulomb and exchange contributions are also given here (and also require that the corresponding basis sets are defined in a `Basis` instance). In addition, this class is used to specify additional settings, such as convergence thresholds, maximum number of iterations, and so on.

The `Property` class is used to define which properties to compute. This includes single-point energy, geometry optimization, excitation energies, and so on, as well as additional specifications related to the property, which could be, e.g., the algorithm to use in the geometry optimization or the number of states to include in the calculation of excitation energies. Currently, the DP platform supports only a limited set of properties out of the great number that are available in Dalton and LSDalton, but this will be continuously extended. A selection of some of the current capabilities is demonstrated in the illustrations presented in Sec. IV.

The `Environment` class defines the environment, if present, surrounding a molecule or fragment, which is defined in a `Molecule` instance. It contains information about the type of environment model, e.g., PCM, PE, or PDE, as well as all the parameters and settings belonging to the specific model. For example, for the PE model, this class contains the coordinates of the classical sites and the associated multipoles and polarizabilities.

The libraries are used in the compute and/or data processing stages. For each library, there is an interface provided as a subpackage of the main DP package. The interface API consists of functions that allow the user to interact directly with the libraries. The exact implementation of the interface varies depending on the nature of the library, e.g., what functionality it provides and how the API of the library itself is defined. However, the interface API functions that are exposed to the user must conform to the standards laid out by the DP platform to ensure that there is uniform access to all libraries as well as interoperability between them. This means, for example, that libraries with similar functionality must also provide API functions with identical names and signatures. Moreover, apart from the classes defined by the DP platform, the types and data structures must be either Python built-ins, e.g., integers, floats, lists, and dictionaries, or NumPy arrays.

The default ordering of AO integrals on the DP platform is the Dalton ordering: atoms are ordered according to the user input, and for each atom it is angular-momentum components first and contracted functions second. Other AO orderings may be used on the platform, but the interface API must provide transformation functions to and from the default Dalton ordering. Transformations can then be made on the platform in the cheapest way possible, e.g., on the MO coefficients rather than on the four-center integrals for MP2.

Numerical compute-intensive tasks are performed at the compute stage. This can include anything from the calculation of integral components, all the way to a complete calculation of a molecular property, as well as more complex workflow protocols. In the first development release of the DP platform, users will be able to directly

calculate one- and two-electron integrals using LSDalton through CFFI-based Python bindings and have seamless access to a selection of wavefunctions and properties computed by Dalton and LSDalton using the file I/O communication mechanism. To deal with complex molecular systems that are too large for a full quantum-chemical treatment, we provide an interface to PyFraME that enables workflows involving fragmentation and quantum–classical embedding.

The results obtained at the compute stage can be used directly at the data processing stage, which includes data extraction, manipulation, analysis, and visualization. The feature set that will be available in the first development release includes the ability to perform vibrational analysis to obtain vibrational frequencies and normal modes, natural orbital occupation analysis to assist in the selection of active spaces, and partitioning of properties into atomic and interatomic contributions using the LoProp approach. In addition, the DP platform has capabilities for plotting spectra and visualizing orbitals, electrostatic potentials, and electronic densities.

The DP platform can automatically detect and manage available hardware resources. Manual specification is also possible and allows fine-grained control. In auto-detection mode, the DP platform will first check for resources reserved through a standard HPC job scheduler and, if no scheduler is found, fall back to use resources on the local computer. Resource usage is optimized according to the capabilities of the used libraries, exploiting OpenMP, MPI, or hybrid OpenMP/MPI parallelism when available. Moreover, a task-farming functionality is provided for use cases where many separate calculations are to be performed.

We conclude this section with a brief walk-through example that illustrates how the DP platform can be used to streamline a workflow going from initial molecular structures to TPA spectra, employing both Dalton and LSDalton. In Sec. IV, additional illustrations are presented to demonstrate the fine-grained access to integrals and show some of the new features available on the DP platform.

Assuming that the DP platform has been imported as `import daltonproject as dp` and a list of XYZ filenames is available in `xyz_filenames`, we start by creating a list of `Molecule` instances

```
molecules = []
for filename in xyz_filenames:
    molecule = dp.Molecule(input_file=filename)
    molecules.append(molecule)
```

Then, we create the `Basis`, `QCMethod`, and `Property` instances that will be used for a geometry optimization of the molecules using LSDalton. The order in which the classes are instantiated is unimportant. We here start with `QCMethod`

```
b3lyp = dp.QCMethod(qc_method='DFT',
                    xc_functional='B3LYP',
                    coulomb='DF',
                    exchange='ADMM')
```

specifying that we want to use the B3LYP XC functional together with density-fitting (DF) and ADMM to accelerate the calculation of Coulomb and exchange contributions, respectively. Both acceleration techniques require an auxiliary basis set, which is specified when instantiating the `Basis` class

```
small_basis = dp.Basis(basis='pcseg-1',
                       ri='def2-universal-JKFIT',
                       admm='admm-1')
large_basis = dp.Basis(basis='pcseg-2',
                       ri='def2-universal-JKFIT',
                       admm='admm-2')
```

Finally, we create a `Property` instance

```
geo_opt = dp.Property(geometry_optimization=True)
```

using the default optimization algorithm. With these instances we can proceed to the compute stage.

For Dalton and LSDalton, when used through the file I/O mechanism, we use the `compute()` function. It returns an `OutputParser` instance that contains methods for transparently fetching relevant results from the output files, as shown further below. The `compute()` function creates a unique filename based on the specific input, which is stored in the `OutputParser` instance. The filename can also be specified through an optional argument in which case it is up to the user to ensure its uniqueness.

We can iterate through the list of molecules, one by one, and perform both the pre-optimization and final optimization in the same loop (updating the molecule coordinates after each step)

```
for molecule in molecules:
    pre = dp.lsdalton.compute(molecule=molecule,
                              basis=small_basis,
                              qc_method=b3lyp,
                              properties=geo_opt)
    molecule.coordinates = pre.final_geometry
    final = dp.lsdalton.compute(molecule=molecule,
                                basis=large_basis,
                                qc_method=b3lyp,
                                properties=geo_opt)
    molecule.coordinates = final.final_geometry
```

LSDalton can exploit the available CPU resources using a hybrid OpenMP/MPI scheme. The optimal use for LSDalton is typically one MPI process per CPU and one OpenMP thread per CPU core, up to a maximum of 20 threads. For example, on a supercomputer with two CPUs per node and 12 cores per CPU, LSDalton will use two MPI processes per node and 12 OpenMP threads per MPI process.

The optimized molecular structures are passed to Dalton for the calculation of TPA cross sections employing the CAM-B3LYP functional

```
camb3lyp = dp.QCMethod(qc_method='DFT',
                       xc_functional='CAM-B3LYP')
```

and a mixed basis set which is defined through a dictionary

```
basis_dict = {'H': 'pcseg-2',
              'C': 'aug-pcseg-2',
              'O': 'aug-pcseg-2'}
mixed_basis = dp.Basis(basis=basis_dict)
```

It is here assumed that the molecules only contain hydrogens, carbons, and oxygens, and that the XYZ files do not contain atomic labels in the fifth column in which case the atomic labels default to

the element symbols. Finally, we create a new `Property` instance for the TPA cross section

```
tpa = dp.Property(two_photon_absorption=True)
tpa.two_photon_absorption(states=4)
```

where we have additionally specified that we want to include four states. We then proceed with the calculation of the TPA cross sections using Dalton, collecting the results for each molecule in the `tpa_results` list

```
tpa_results = []
for molecule in molecules:
    result = dp.dalton.compute(molecule=molecule,
                               basis=mixed_basis,
                               qc_method=camb3lyp,
                               properties=tpa)
    tpa_results.append(result)
```

By default, Dalton will here adopt a purely MPI-parallel execution, corresponding to one MPI process per CPU core. Alternatively, the task farming functionality can be used by supplying the list of molecules to the `compute()` function

```
tpa_results = dp.dalton.compute(molecule=molecules,
                                basis=mixed_basis,
                                qc_method=camb3lyp,
                                properties=tpa)
```

In this scenario, a separate calculation will run for each molecule concurrently, dividing the available CPU resources among them.

Finally, we can plot the TPA spectra using the spectrum module of the DP platform

```
axs = []
for result in tpa_results
    ax = dp.spectrum.plot_two_photon_spectrum(result)
    axs.append(ax)
```

where `ax` is an instance of the Matplotlib[130] `Axes` class. Further customization can then be performed to produce publication-ready figures, which can be plotted as normally done with Matplotlib.

IV. DP PLATFORM ILLUSTRATIONS

In this section, we provide six use cases of the DP platform with the intent to demonstrate novel platform functionalities in terms of data exposure, processing, and visualization as well as to illustrate some of the added features of the platform libraries. The Python scripts used for these DP platform illustrations, along with the corresponding input/output files, are deposited at https://doi.org/10.5281/zenodo.3710462.

A. NumPy-exposure of one- and two-electron integrals

We here demonstrate how to access primitive and contracted integrals on the DP platform. This functionality is made available by LSDalton through CFFI-based Python bindings and enables access to a variety of one- and two-electron integrals including Coulomb and exchange integral matrices. We will soon add exposure of geometrically differentiated integrals (in principle, to arbitrary order), integrals for embedding techniques involving interaction with classical charges and higher-order multipoles, Gaussian-geminal type integrals needed for F12-type theories, attenuated two-electron integrals, multipole-moment integrals, and magnetically differentiated London integrals (to first order for the two-electron integrals).

Integrals and integral components are available on the DP platform in the form of NumPy arrays and currently require instances of `Molecule`, `Basis`, and `QCMethod` classes, as outlined in Sec. III. For example,

```
h2o = dp.Molecule('O  0.000000  0.000000  0.000000;'
                  'H  0.758602  0.000000  0.504284;'
                  'H  0.758602  0.000000 -0.504284')
pcs1 = dp.Basis('pcseg-1')
dft = dp.QCMethod('DFT', 'B3LYP')
```

where `import daltonproject as dp` is assumed. In the following, we also import the LSDalton module: `import daltonproject.lsdalton as lsd`. We are then ready to compute basic DFT integral components, such as one-electron matrices

```
S = lsd.overlap_matrix(h2o, pcs1, dft)
h = lsd.kinetic_matrix(h2o, pcs1, dft) \
    + lsd.nuclear_electron_attraction_matrix(h2o, pcs1, dft)
D = lsd.diagonal_density(h, S, h2o, pcs1, dft)
```

two-electron matrices

```
J = lsd.coulomb_matrix(D, h2o, pcs1, dft)
K = lsd.exchange_matrix(D, h2o, pcs1, dft)
E_xc, V_xc = lsd.exchange_correlation(D, h2o, pcs1, dft)
g = 2.0 * J - K + V_xc
F_ks = h + g
```

Molecular-orbital energies and coefficients can then be obtained, e.g, using SciPy

```
E_mo, C_mo = scipy.linalg.eigh(a=F_ks, b=S)
```

Two-electron four-center integrals can be obtained through

```
eri = lsd.eri4(h2o, pcs1, dft)
```

and, for example, two- and three-center RI integrals, by specifying an auxiliary RI basis, through

```
basis = dp.Basis('pcseg-1', 'def2-universal-JKFIT')
ab_alpha = lsd.ri3(h2o, basis, dft)
alpha_beta = lsd.ri2(h2o, basis, dft)
```

With these integrals, we can easily construct the resolution-of-the-identity (RI) Coulomb matrix

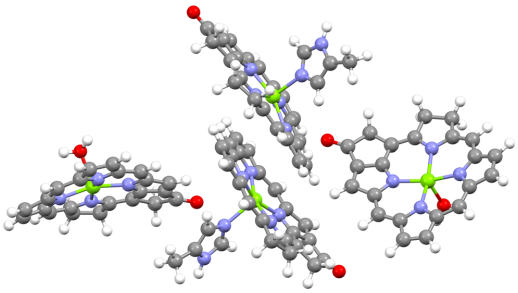


FIG. 2. The structure of the tetrameric model consisting of units of the P700 pigment of photosystem I.

$$\tilde{J}_{ab} = \sum_{cd}\sum_{\alpha,\beta}(ab|\alpha)(\alpha|\beta)^{-1}(\beta|cd)D_{cd}, \tag{1}$$

through the sequence

$$\begin{aligned} g_\beta &= \sum_{cd}(\beta|cd)D_{cd}, \\ c_\alpha &= \sum_{\beta}(\alpha|\beta)^{-1}g_\beta, \\ \tilde{J}_{ab} &= \sum_{\alpha}(ab|\alpha)c_\alpha \end{aligned} \tag{2}$$

according to

```
g_beta = np.einsum('cdB,cd', ab_alpha, density)
AB_inv = np.linalg.inv(alpha_beta)
c_alpha = np.einsum('AB,B', AB_inv, g_beta)
J_ab = np.einsum('A,abA', c_alpha, ab_alpha)
```

We note that this example is only meant to illustrate the ease with which algorithms can be implemented on the platform and that there are better ways to construct the density-fitted Coulomb matrix in Eq. (1) (see, for instance, Ref. 131 and references therein).

We end this illustration with a concrete example of a CAM-B3LYP/pcseg-2 single-point energy calculation of a tetrameric model of the P700 pigment of photosystem I,[132] depicted in Fig. 2, to illustrate the computational performance of the LSDalton library. This triple-zeta quality P700 tetramer model consists of 198 atoms and 4744 contracted basis functions. The wall time, with 32 Intel E5-2683v4 dual socket nodes (1024 cores), was 15 min. The calculation used J-engine accelerated density-fitting of the Coulomb contribution and the ADMM2 variant of ADMM for the exchange contribution, and employed one MPI process per CPU and 16 OpenMP threads per process. For more details about the acceleration techniques, consult Ref. 100, where these techniques were studied more extensively.

B. Combined treatment of static and dynamic electron correlation: The multi-configurational short-range density functional theory method

The calculation of UV/Vis spectra with the MC-srDFT method is here illustrated with the retinylidene Schiff base chromophore, as originally addressed in previous works.[31,70] Spectra for this system are shown in Fig. 3 together with the π-orbitals that constitute the active space, denoted by π_1–π_6. We first note that range-separated calculations introduce a range-separation parameter μ that affects the results as long as one remains short of the full-CI limit. From a practical point of view, the optimal μ-value is that which delivers reliable results with the least amount of computational effort. Benchmark studies have shown that a value around $\mu = 0.4\ a_0^{-1}$ is close to optimal for ground and excited states.[26,133,134] We have adopted this value for the present illustration.

The following workflow was employed in the spectrum calculations: First, the natural orbitals were calculated from the MP2-srPBE ground-state wavefunction and the occupation numbers were used to select a suitable active space. The DP platform provides seamless processing of the MP2-srPBE results, and, based on a user-defined selection criterion, an automatic selection of strongly and weakly occupied orbitals is made. Our adopted CAS(6,6) active space came as a result of including orbitals with occupation numbers

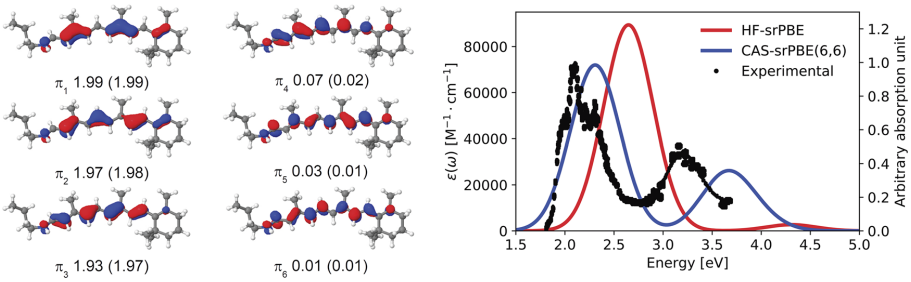


FIG. 3. UV/Vis spectrum of the retinylidene Schiff base chromophore. Results are obtained at the level of CAS(6,6)-srPBE/6-31+G* with $\mu = 0.4\ a_0^{-1}$. Both CAS and MP2 (in parentheses) occupation numbers are given. The experimental spectrum presented in arbitrary units is taken from Ref. 135.

below 1.99 (occupied space) and above 0.01 (virtual space). Second, the CAS(6,6)-srPBE calculation is performed. Third, the DP platform processes data by extracting excitation energies and oscillator strengths and generates the requested absorption spectrum. The following definition of the frequency-dependent molar absorption coefficient $\varepsilon(\omega)$ was used:

$$\varepsilon(\omega) = \frac{e^2\pi^2 N_A}{\ln(10)2\pi\epsilon_0 n m_e c}\sum_i \frac{\omega f_i}{\omega_i} g(\omega, \omega_i, \gamma_i), \quad (3)$$

where N_A is Avogadro's constant, e is the elementary charge, m_e is the electron mass, c is the speed of light, ϵ_0 is the vacuum permittivity, n is the refractive index (here set to unity), f_i is the calculated oscillator strength of the ith transition, and ω_i is the corresponding transition angular frequency. Equation (3) also introduces the line-broadening function g with the phenomenological parameter γ_i corresponding to the half-width at half-maximum (HWHM). The DP platform offers spectral broadenings based on normalized Gaussian or Lorentzian line profiles.

With the basis of the results in Fig. 3, we briefly discuss some key features of the MC-srDFT method. We note that the CAS-srPBE spectrum shows two distinct peaks in good agreement with the experimentally observed S_1 and S_2 bands[135,136]—transition energies and oscillator strengths are provided in Table I. It is here essential to employ a multi-configurational wavefunction as can be seen by the comparison to the single-determinant (HF-srPBE) spectrum, which essentially corresponds to the range-separated LC-PBE model. Without account of static correlation, peak positions are strongly blue-shifted and the intensity of the second band is severely underestimated. It is noted that occupation numbers from a regular MP2 calculation suggest significantly larger active spaces.[137] However, the use of CAS(6,6)-srPBE provides the same accuracy as literature CASPT2 results based on a much larger CAS(12,12) active space.[138] This demonstrated opportunity to employ relatively small active spaces with MC-srDFT is not limited to the retinylidene Schiff base chromophore, but has also been shown in other contexts, e.g., to describe the mechanism of [NiFe]-hydrogenase.[139]

The reason behind the failure of single-reference DFT approaches to properly describe the electronic transition underlying the S_2 band in the retinylidene Schiff base chromophore becomes apparent from an analysis of the MC-srDFT response vectors. For S_1, the dominant element of the response vector refers to the configuration associated with a single-electron excitation from π_3 to π_4 (CI coefficient of 0.78 in Table I). For S_2, on the other hand, there are three almost equally important configurations appearing in the response vector, one of which is associated with a double-electron excitation from π_3 to π_4 (CI coefficient of 0.42 in Table I) that is unaccounted for in standard time-dependent DFT.

TABLE I. Vertical excitation energies (eV) and oscillator strengths (in parentheses) for the two lowest singlet states in the retinylidene Schiff base chromophore.

Method	S_1	S_2
HF-srPBE	2.62 (1.980)	4.29 (0.060)
CAS(6,6)-srPBE	2.28 (1.592)	3.62 (0.525)
Expt.[135,136]	2.03	3.22
Assignment[a]	CI coeff.	CI coeff.
$\pi_2(\uparrow) \rightarrow \pi_4(\uparrow)$	0.30	−0.53
$\pi_3(\uparrow) \rightarrow \pi_4(\uparrow)$	0.78	0.46
$\pi_3(\uparrow\downarrow) \rightarrow \pi_4(\uparrow\downarrow)$	−0.25	0.42

[a]Largest CI coefficients in the CAS(6,6)-srPBE response vectors. Iso-density orbital plots are shown in Fig. 3.

C. Modeling complex systems through fragment-based quantum–classical approaches

We here provide an illustration that demonstrates not only the ability of the DP libraries to perform spectrum simulations of *complex systems* but also the ability and potential of the hosting DP platform to manage the workflow of *complex computational protocols* associated with environment fragmentation and configuration-space sampling. It is an indisputable fact that first-principles methods in quantum chemistry come with a computational cost that hampers applications to relevant chemical and biochemical systems such as solutions and protein-embedded chromophores. Methods that allow for low-cost approximate yet accurate modeling of environments are therefore scientifically enabling, and one such approach is the PE model briefly described in Sec. II B 3. In the PE scheme, the environment is represented by atom-centered multipoles (typically up to and including quadrupoles) and atom-centered anisotropic dipole–dipole polarizabilities that allow for a mutual polarization between the quantum part and the classical environment.

As further described in Sec. II B 3, the adopted multipole expansion in the standard PE formulation is in the PDE model replaced with an exact Coulomb interaction as well as a description of non-electrostatic exchange repulsion. The total embedding operator in PDE consists of terms that describe the electrostatic, induction, and exchange-repulsion effects from the environment onto the core quantum region. The electrostatic operator contains the Coulomb interaction with the electrons and nuclei of the environment. For the construction of the electrostatic operator, density-matrix elements of the environment and intermolecular (core–fragment) two-electron integrals are required. The repulsion operator models the effects of exchange repulsion between the quantum core and environment fragment wavefunctions through a projection operator that scales as the square of the intermolecular overlap. The PDE model shows clear advantages over the standard PE model. It effectively avoids artificial stabilization of diffuse excited states and electron spill-out.

Both the PE and PDE models are applicable to large and complex (bio-)molecular systems. Since the parameters describing the environment are derived based on a fully *ab initio* description, the setup of the spectrum calculation involves a large number of preliminary calculations on the separate fragments of the environment. Figure 4 illustrates this situation in terms of TPA spectrum calculations of the Nile red chromophore embedded in the β-lactoglobulin protein. Including the solvent, this system contains 32 582 classical sites. The entire protein and water molecules within 15 Å of the Nile red chromophore were treated with either PE or PDE, while water molecules beyond this distance were

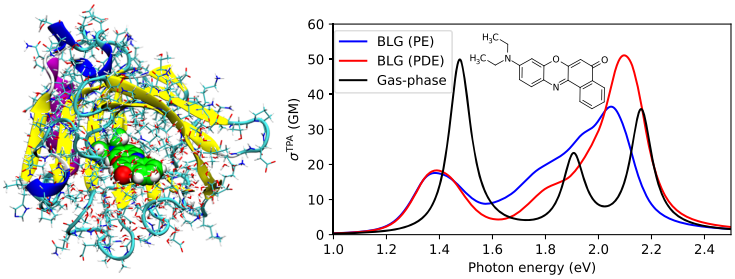


FIG. 4. The left panel shows the Nile red chromophore embedded in the β-lactoglobulin (BLG) protein. The right panel shows the simulated TPA spectra of the Nile red in vacuum and embedded in the protein (described using the PE or PDE models). Results are obtained at the CAM-B3LYP/6-31+G* level of theory.

described using the TIP3P water model (in total 10 000 water molecules).

Configuration sampling of the environment was performed by averaging over 150 snapshots sampled from a quantum mechanics/molecular mechanics (QM/MM) MD simulation.[140] The calculation of embedding-potential parameters for each snapshot took *circa* 1 h and 4 h for PE and PDE, respectively, on 20 nodes (dual socket E5-2680v3, 12 cores). The TPA spectrum calculations (five lowest singlet states) included EEF effects and took *circa* 4.5 h and 5.0 h per snapshot for PE and PDE, respectively, on eight nodes.

Assuming a monochromatic laser source, the TPA cross section (in units of GM) is computed as[141]

$$\sigma^{\mathrm{TPA}}(\omega) = \frac{8\pi^3 \alpha a_0^5}{c} \sum_i \omega^2 \delta_i^{\mathrm{TP}} g(2\omega, \omega_i, \gamma_i), \tag{4}$$

where α is the fine-structure constant, a_0 is the Bohr radius, and g is here chosen to be a Lorentzian with a HWHM of $\gamma = 0.1$ eV [see also Eq. (3)]. The two-photon (TP) transition strength of the ith transition is computed assuming linearly polarized light as

$$\delta_i^{\mathrm{TP}} = \frac{1}{15} \sum_{a,b} \left(2S_{ab}S_{ab}^* + S_{aa}S_{bb}^*\right), \tag{5}$$

where S_{ab} are the associated TP transition matrix elements.

The gas-phase TPA spectrum of Nile red shows three distinct peaks at 1.5 eV, 1.9 eV, and 2.2 eV. Embedded in the protein, the lowest band becomes red-shifted by 0.1 eV, and it also becomes both broader and lower in intensity. At higher energies, the PDE model preserves the structure from the gas phase partially, with an intense peak at 2.1 eV, and a shoulder at 1.7 eV. In contrast, the PE model predicts a large peak at 2.1 eV with a broad shoulder toward lower energies. This occurs due to the lack of non-electrostatic repulsion, which leads to a high density of states in this region due to artificial stabilization of higher-lying excited states.

D. Open-ended response theory for electric and geometric perturbations: Infrared and Raman spectroscopy with the OpenRSP and SpectroscPy modules

Spectroscopic techniques involving molecular vibrations are useful for characterizing molecular systems, and with the DP platform, an option for the calculation of vibrational spectra is available through the combined use of LSDalton, OpenRSP,[120] and SpectroscPy.[142] We here illustrate this functionality through a calculation of IR and Raman spectra of benzene, including a brief outline of the key aspects of the LSDalton/OpenRSP/SpectroscPy combination, and its use on the platform.

From the DP platform, energy-derivative tensors generated by LSDalton/OpenRSP are passed on to SpectroscPy that processes them to generate spectroscopic properties. OpenRSP manages the calculation of response properties by an open-ended quasienergy-based formulation of response theory,[143] employing a recursive formalism.[144] OpenRSP has an API through which it can be connected to different host programs. OpenRSP is currently called through LSDalton, which provides the necessary integral derivatives (briefly outlined in Sec. II C 1), XC contributions through modules XCFun[95] and XCint,[121] and response equation solver capability.[145] The use of QcMatrix[128] allows OpenRSP to be agnostic to the details of the matrix operations, i.e., independent of their specific implementation when passing matrices from/to a host program and when carrying out matrix operations inside the OpenRSP core functionality. SpectroscPy is used to produce spectroscopic properties involving molecular vibrations by performing vibrational analysis generating vibrational frequencies and absorption properties. Presently, SpectroscPy offers functionality for IR and Raman spectroscopy in the harmonic approximation and can also combine data from calculations on a series of molecular configurations, which is useful, for example, in applications dealing with flexible molecular systems that require configuration-space sampling. Future extensions will include anharmonic corrections, hyper-Raman spectroscopy, and other spectroscopic processes.

The IR molar decadic absorption coefficient for normal mode i can, in the double harmonic approximation, be expressed as[19]

$$\varepsilon_i(\tilde{\nu}) = \frac{N_A}{12\ln(10)\varepsilon_0 c^2} \sum_{\alpha}^{x,y,z} \left(\frac{\partial \mu_\alpha}{\partial Q_i}\right)^2 g(\tilde{\nu}; \tilde{\nu}_i, \gamma_i), \tag{6}$$

where μ is the molecular dipole moment, Q_i is the normal mode coordinate i, and g is here chosen as a Lorentzian [see also Eq. (3)]. The molar absorption coefficient is a function of the wavenumber $\tilde{\nu}$ and depends parametrically on the vibrational wavenumbers $\tilde{\nu}_i$ and the HWHM broadening γ_i of mode i. Derivatives are evaluated at the equilibrium geometry.

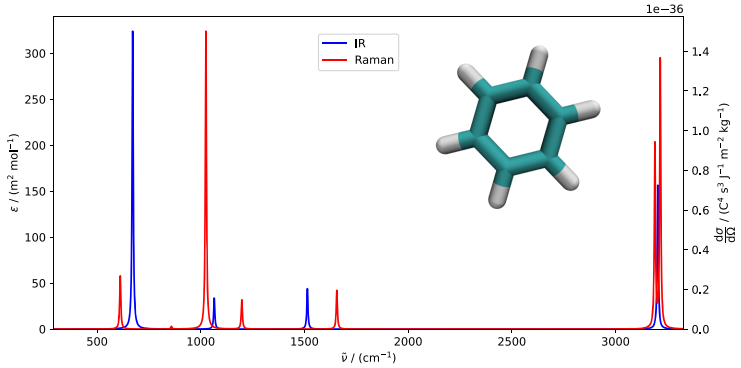


FIG. 5. IR molar decadic absorption coefficient and Raman scattering cross section of benzene using a common HWHM of 3.2 cm^{-1}. The spectra were calculated at the PBE0/pcseg-2 level of theory. The Raman scattering cross sections were calculated for $\tilde{\nu}_0 = 2.195 \times 10^4$ cm^{-1} and $T = 298$ K.

Similarly, the harmonic differential Raman scattering cross section σ_i' is given by[146]

$$\sigma_i'(\tilde{\nu}) = \frac{d\sigma_i(\tilde{\nu})}{d\Omega} = \frac{h(\tilde{\nu}_0 - \tilde{\nu}_i)^4}{16\pi^3 c^2 \tilde{\nu}_i \left(1 - \exp\left[-\frac{hc\tilde{\nu}_i}{kT}\right]\right)} \times \left(45a_i'^2 + 7b_i'^2\right) g(\tilde{\nu}; \tilde{\nu}_i, \gamma_i), \tag{7}$$

where h is Planck's constant, $\tilde{\nu}_0$ is the wavenumber of the incident light, k is the Boltzmann constant, and T is the temperature. We have here introduced

$$a_i' = \frac{1}{3}\sum_{\alpha}^{x,y,z} \frac{\partial \alpha_{\alpha\alpha}}{\partial Q_i}; \quad b_i'^2 = \sum_{\alpha}^{x,y,z}\sum_{\beta\neq\alpha}^{x,y,z}\left[\frac{1}{2}\left(\frac{\partial \alpha_{\alpha\alpha}}{\partial Q_i} - \frac{\partial \alpha_{\beta\beta}}{\partial Q_i}\right)^2 + 3\left(\frac{\partial \alpha_{\alpha\beta}}{\partial Q_i}\right)^2\right], \tag{8}$$

where α is the molecular dipole–dipole polarizability.

The spectra shown in Fig. 5 were generated by SpectroscPy using the molecular Hessian and the first-order geometrical derivatives of the molecular dipole moment and polarizability calculated by LSDalton/OpenRSP. SpectroscPy first calculated harmonic vibrational frequencies by carrying out an eigenanalysis of the molecular Hessian[147] and projecting out translational and rotational degrees of freedom.[148] The requested intensity-related quantities were then calculated and plotted as a function of the frequency. The DP platform allows for the whole pipeline to be run in an automated fashion. The libraries are seamlessly connected through the platform, which thus allows for running the calculations, processing the data, and visualizing the results through a simple Python script.

E. Open-shell properties free from spin-contamination: The restricted–unrestricted response theory formalism

One of the unique capabilities of Dalton is the ability to compute various linear and nonlinear response properties of open-shell systems at the spin-restricted open-shell Kohn–Sham (ROKS) level of theory,[149,150] and the DP platform provides the means for seamless and immediate visualization of spin densities to facilitate the interpretation of the results. The spin-restricted formalism ensures that the ground-state electron density is free from spin-contamination, and, as such, it provides a better starting point for molecular property calculations compared to the more widely used unrestricted Kohn–Sham (UKS) approach. The advantages of ROKS over UKS are most apparent in calculations of electron paramagnetic resonance (EPR) spin Hamiltonian parameters, which are explicitly dependent on an effective expectation value of the electronic spin operator, and benchmark studies on organic radicals show that the ROKS approach is able to better predict electronic g-tensors and hyperfine coupling constants.[151–154]

The main strength of the ROKS approach is the ability to produce a spin-contamination free electron density for the high-spin ground state. This is achieved by imposing spin-symmetry restrictions during the SCF optimization.[149] The side effect of these restrictions is that the hereby obtained KS orbitals have a non-vanishing gradient with respect to orbital rotations of triplet spin-symmetry,[152,155] and this side effect manifests itself in the computation of electronic spin-dependent properties, such as hyperfine coupling constants. For example, the expectation value of the one-electron spin-dependent operator $\hat{A}$ in the ROKS approach is given by the direct spin-density and spin-polarization contributions,[152]

$$\langle \hat{A} \rangle = \mathrm{Tr}(\mathbf{A}\,\mathbf{D}_{\mathrm{spin}}) + \mathrm{Tr}(\mathbf{A}\,\mathbf{D}_{\mathrm{pol}}); \quad A_{ij} = \langle \phi_i | \hat{A} | \phi_j \rangle, \tag{9}$$

where the first contribution is computed in the same way as in the UKS approach by contracting the electron spin-density $\mathbf{D}_{\mathrm{spin}}$ with operator matrix $\mathbf{A}$ in the AO basis and the second contribution is computed similarly to the first contribution by replacing $\mathbf{D}_{\mathrm{spin}}$ with the spin-polarization density $\mathbf{D}_{\mathrm{pol}}$.

The spin-polarization density, $\mathbf{D}_{\mathrm{pol}}$, is determined by solving restricted–unrestricted response equations that account for the relaxation of KS orbitals in the presence of the spin-dependent perturbation. The relative importance of the spin-polarization density contribution varies greatly between different molecular properties: from being prominent for isotropic hyperfine coupling constants[152,155] to being negligible for electronic g-tensors.[154]

The use of the spin and spin-polarization densities as obtained from the restricted–unrestricted response formalism is not limited to the calculation of spin-dependent molecular properties, but

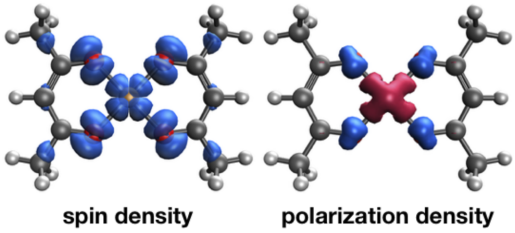


FIG. 6. Spin-density and spin-polarization densities for the $^2B_{1g}$ ground-state of the copper acetylacetonate complex, $Cu(AcAc)_2$. Isodensity surfaces are based on volumetric data files generated on the DP platform by performing restricted–unrestricted response theory calculations of the isotropic hyperfine coupling constant. At the ROKS/B3LYP/Wachtersa+f/Huz-III level of theory, the isotropic hyperfine coupling constant of the copper atom, $A(^{63}Cu)$, amounts to 539 MHz and only the spin-polarization contributes to its value.

can also be employed for topological analyses of open-shell system responses to spin-dependent perturbations. To illustrate this point, we depict the isodensity surfaces of the spin and spin-polarization in Fig. 6 for the $^2B_{1g}$ ground state of the copper acetylacetonate complex, $Cu(AcAc)_2$. The spin-density contribution for Cu(II) complexes is expected to primarily stem from the $3d_{xy}$-orbital of the copper atom,[154] but it is here seen to also acquire significant contributions from the coordinating oxygen atoms. The spin-polarization density is also delocalized across the copper and oxygen atoms, and based on the localization of these two densities, one can predict the behavior of spin-dependent properties and qualitatively estimate the importance of the spin-density and the spin-polarization contributions. We emphasize that the presented disentangled visualization of spin-density and spin-polarization contributions to molecular properties is uniquely accessible in the restricted–unrestricted response formalism and not available in the UKS approach.

F. Coupled cluster methods for inner-shell spectroscopy

As an illustration of the use of the DP platform for the simulation of x-ray spectroscopies, we present in Fig. 7 the NEXAFS spectrum of acrolein. We here adopt the CCSD level of theory in conjunction with the CVS approximation, as implemented in Dalton (see Sec. II B 2). Based on transition energies and oscillator strengths from CVS-CCSD response theory, the absorption spectrum is determined from Eq. (3) with the use of a Lorentzian line shape function and an HWHM broadening of 0.27 eV for all transitions.

The three lowest excitations at the near carbon edge correspond to the first two bands in the experimental spectrum, These excitations are assigned to $1s \rightarrow \pi^*$ transitions from the three carbon atoms by means of a natural transition orbital (NTO) analysis. The transition associated with the carbonyl group is chemically blue shifted by some 1.5 eV from the other two (see Fig. 7). There is an overall excellent spectrum agreement between theory and experiment, and, although simple, this example illustrates well the value of spectrum simulations for the characterization and interpretation of experiments. The most prominent added value of the DP platform for ground-state x-ray spectroscopy simulations using Dalton comes at this stage of analysis and visualization.

In the example above, the CVS approximation is employed only during the determination of the core-excited states. It is implemented as a projector that, during the iterative solution of the CC eigenvalue problem,[56,57] $\mathbf{A}R^f = \omega_f R^f$, only retains elements R^f_{aI}, $R^f_{aI,bj}$, $R^f_{ai,bJ}$, and $R^f_{aI,bJ}$ (CCSD case), where I and J refer to core orbitals and a and b refer to virtual orbitals. A schematic representation of the CVS approximation is shown in Fig. 8, where only the green matrix

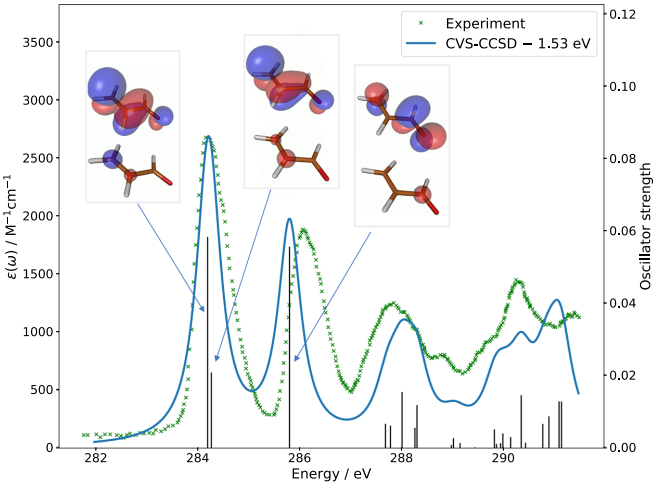


FIG. 7. Near carbon *K*-edge x-ray absorption spectrum of acrolein, C_3H_4O, obtained at the CVS-CCSD/6-311++G(d,p) level of theory. The theoretical spectrum is red shifted by 1.53 eV, and the experimental spectrum is taken from Ref. 156 and presented in arbitrary units. Natural transition orbitals for the first three excitations, all of $1s \rightarrow \pi^*$ character, are shown.

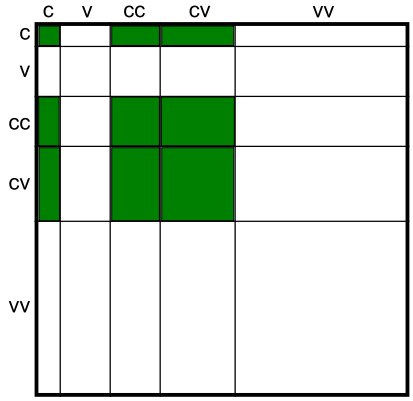


FIG. 8. Schematic representation of the CC Jacobian in the CVS approximation where only the green matrix sub-blocks are considered in solving the eigenvalue equation. The illustration refers to the case where single and double excitations are present. Labels c and v refer to occupied core and valence orbital indices, respectively.

sub-blocks are effectively kept after applying the projector during the iterative procedure of the response solver. Core and valence excitations become decoupled and x-ray spectra are obtained at basically the same computational cost as UV/Vis spectra with the use of identical bottom-up iterative techniques. The ability of a specific CC method to reproduce specific core-excitation spectral signatures depends on the amount of relaxation as well as of single, double, or higher excitation character of the transition. CCSD generally yields core spectra in rather good (semi-quantitative) agreement with experiment, with rigid blue-shifts ranging in between 0.5 eV and 3 eV, depending on the *K*-edge considered and the basis set adopted.

In addition to facilitating data analysis and visualization, the DP platform greatly simplifies the more complicated setup involved with simulations of transient (excited-state) x-ray absorption and emission spectra. Due to design legacies in Dalton, such simulations without the use of the DP platform will require a substantial amount of manual file handling as the necessary valence and core excitation vectors must be obtained from separate code executions and then later recombined. On the platform, all of these steps are handled in an automated way.

V. OUTLOOK

We have given a presentation of the *Dalton Project* that marks a paradigm shift in the software engineering practices for the Dalton community. At the heart of the Dalton Project, we find a hardware-aware platform written in Python with support from NumPy, SciPy, and MPI4Py, which provides a modern user interface and defines a communication standard for seamless library access and interoperability. At present, the DP platform supports three libraries, namely, Dalton, LSDalton, and PyFraME, but further modular extensions are underway. Libraries are written in any of the programming languages Python, Fortran, C, or C++, and they communicate with the platform by means of file I/O, Python bindings through CFFI or pybind11, or pure Python module import. Modular programming with enforced unit testing will become the standard for newly started developments, and it is to be anticipated that the existing few monolithic codes will gradually be phased out and replaced by a larger number of libraries with more specific tasks. The DP platform is open source and distributed under the GNU General Public License version 3.0 or later (GPLv3+).

The DP platform is developed for execution on personal computers as well as more powerful supercomputers in HPC environments, and, although without guarantees, the user should expect it to run under Windows, MacOS, and Linux operating systems equipped with Python 3.6 (or later) installations (see the DP website https://daltonproject.org for instructions). Installation of the DP libraries is a separate issue, and the platform will gracefully signal to the user when it encounters called for but missing libraries. It is, of course, fully possible to use the DP platform alone on a personal computer to analyze the results available in output files produced on a remote HPC system. As our model to reach a sustainable software ecosystem adopts the notion of separate and quite independent release and distribution policies for the libraries, the main burden of work and the most dependency issues in the installation process are expected to be found in the phase of library installation. Driven by the stimulus of becoming recognized and used, it is anticipated and expected that newly started library developments will care to offer a smooth installation process on a widespread selection of platforms and operating systems.

For developers, the DP platform can lead to rapid prototyping of novel scientific ideas, as illustrated in Sec. IV A with an examination of the RI approximation. However, this example also points out something else of great importance, namely, the educational aspect of the DP platform. Our experience tells us that the process of implementing methods to solve fundamental equations is supremely efficient to reach a deeper understanding of the topic at hand, but only few students are granted this opportunity as core program modules of scientific software were written a long time ago and often made obscure by code optimization and entanglement. What is here illustrated is the access to the needed building blocks to explore quantum chemistry in very much the same manner that we can use the Python NumPy package to explore linear algebra. At the early stage of a Ph.D. education, we believe that this can be of high value and help overcome some of the initial hurdles faced during a career in quantum chemical theory and program development.

ACKNOWLEDGMENT

This work was financially supported by the Norwegian Research Council through Grant Nos. 250743, 261873, and 274918 and also through the CoE Hylleraas Centre for Quantum Molecular Sciences (Grant Nos. 262695 and 231571), the Danish Council for Independent Research (Grant Nos. 7014-00050B and 7014-00258B), the Swedish Research Council (Grant No. 2018-4343), the European Commission through Project No. 745967 and also through the ITN Computational Spectroscopy in Natural Sciences and Engineering (COSINE) (Project No. 765739), and the European Research

Council (Grant No. 279619). Computational resources are provided by the DeiC National HPC Centre, the Norwegian Supercomputing Program (NOTUR, Grant No. NN4654K), and the Swedish National Infrastructure for Computing (SNIC).

REFERENCES

[1] Dalton, a molecular electronic structure program, Release v2020.0 (2020), see https://daltonprogram.org/.
[2] LSDalton, a linear scaling molecular electronic structure program, Release v2020.0 (2020), see https://daltonprogram.org/.
[3] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenaes, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. Aa. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjaergaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnaes, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. Wilson, M. Ziolkowski, and H. Ågren, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 269 (2014).
[4] The Molecular Sciences Software Institute (MolSSI) (2020), see https://molssi.org/.
[5] A. Krylov, T. L. Windus, T. Barnes, E. Marin-Rimoldi, J. A. Nash, B. Pritchard, D. G. A. Smith, D. Altarawy, P. Saxe, C. Clementi, T. D. Crawford, R. J. Harrison, S. Jha, V. S. Pande, and T. Head-Gordon, J. Chem. Phys. **149**, 180901 (2018).
[6] Dalton Project: A Python platform for molecular- and electronic-structure simulations of complex systems (2020), see https://daltonproject.org/.
[7] K. Roberts, Comput. Phys. Commun. **1**, 1 (1969).
[8] E. W. Dijkstra, Commun. ACM **11**, 341 (1968).
[9] D. L. Parnas, Commun. ACM **15**, 1053 (1972).
[10] S. van der Walt, S. C. Colbert, and G. Varoquaux, Comput. Sci. Eng. **13**, 22 (2011).
[11] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, and P. van Mulbregt, Nat. Methods **17**, 261 (2020).
[12] L. Dalcín, R. Paz, M. Storti, and J. D'Elía, J. Parallel Distrib. Comput. **68**, 655 (2008).
[13] A. Rigo and M. Fijalkowski, CFFI: C Foreign Function Interface for Python, 2018, see https://cffi.readthedocs.io/.
[14] W. Jakob, J. Rhinelander, and D. Moldovan, pybind11—Seamless operability between C++11 and Python, 2017, see https://github.com/pybind/pybind11.
[15] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill, J. Chem. Theory Comput. **13**, 3185 (2017).
[16] D. G. A. Smith, L. A. Burns, D. A. Sirianni, D. R. Nascimento, A. Kumar, A. M. James, J. B. Schriber, T. Zhang, B. Zhang, A. S. Abbott, E. J. Berquist, M. H. Lechner, L. A. Cunha, A. G. Heide, J. M. Waldrop, T. Y. Takeshita, A. Alenaizan, D. Neuhauser, R. A. King, A. C. Simmonett, J. M. Turney, H. F. Schaefer, F. A. Evangelista, A. E. DePrince, T. D. Crawford, K. Patkowski, and C. D. Sherrill, J. Chem. Theory Comput. **14**, 3504 (2018).
[17] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K. L. Chan, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **8**, e1340 (2018).
[18] Z. Rinkevicius, X. Li, O. Vahtras, K. Ahmadzadeh, M. Brand, M. Ringholm, N. H. List, M. Scheurer, M. Scott, A. Dreuw, and P. Norman, Wiley Interdiscip. Rev.: Comput. Mol. Sci. e1457 (published online 2019).
[19] P. Norman, K. Ruud, and T. Saue, *Principles and Practices of Molecular Properties* (John Wiley & Sons, Ltd., Chichester, UK, 2018).
[20] J. M. H. Olsen, PyFraME: Python framework for Fragment-based Multiscale Embedding, 2020, see https://gitlab.com/FraME-projects/PyFraME.
[21] A. Savin, in *Recent Developments and Applications of Modern Density Functional Theory*, edited by J. Seminario (Elsevier, 1996), Vol. 4, pp. 327–357.
[22] A. Savin and H.-J. Flad, Int. J. Quantum Chem. **56**, 327 (1995).
[23] E. Fromager and H. J. Aa. Jensen, Phys. Rev. A **78**, 022504 (2008).
[24] J. K. Pedersen, "Description of correlation and relativistic effects in calculations of molecular properties," Ph.D. thesis, University of Southern Denmark, 2004.
[25] E. D. Hedegård, S. Knecht, J. S. Kielberg, H. J. Aa. Jensen, and M. Reiher, J. Chem. Phys. **142**, 224108 (2015).
[26] E. Fromager, J. Toulouse, and H. J. Aa. Jensen, J. Chem. Phys. **126**, 074111 (2007).
[27] E. Fromager, F. Réal, P. Wåhlin, U. Wahlgren, and H. J. Aa. Jensen, J. Chem. Phys. **131**, 054107 (2009).
[28] E. D. Hedegård, J. Toulouse, and H. J. Aa. Jensen, J. Chem. Phys. **148**, 214103 (2018).
[29] E. Fromager, R. Cimiraglia, and H. J. Aa. Jensen, Phys. Rev. A **81**, 024502 (2010).
[30] E. Fromager, S. Knecht, and H. J. Aa. Jensen, J. Chem. Phys. **138**, 084101 (2013).
[31] E. D. Hedegård, F. Heiden, S. Knecht, E. Fromager, and H. J. Aa. Jensen, J. Chem. Phys. **139**, 184308 (2013).
[32] E. R. Kjellgren, E. D. Hedegård, and H. J. Aa. Jensen, J. Chem. Phys. **151**, 124113 (2019).
[33] S. P. A. Sauer, H. F. Pitzner-Frydendahl, M. Buse, H. J. Aa. Jensen, and W. Thiel, Mol. Phys. **113**, 2026 (2015).
[34] S. Grimme, J. Chem. Phys. **118**, 9095 (2003).
[35] S. Grimme, L. Goerigk, and R. F. Fink, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **2**, 886 (2012).
[36] Y. S. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, J. Chem. Phys. **121**, 9793 (2004).
[37] O. Christiansen, K. L. Bak, H. Koch, and S. P. A. Sauer, Chem. Phys. Lett. **284**, 47 (1998).
[38] P. A. B. Haase, R. Faber, P. F. Provasi, and S. P. A. Sauer, J. Comput. Chem. **41**, 43 (2020).
[39] A. K. Schnack-Petersen, P. A. B. Haase, R. Faber, P. F. Provasi, and S. P. A. Sauer, J. Comput. Chem. **39**, 2647 (2018).
[40] R. H. Myhre and H. Koch, J. Chem. Phys. **145**, 044111 (2016).
[41] P. Norman, D. M. Bishop, H. J. Aa. Jensen, and J. Oddershede, J. Chem. Phys. **123**, 194103 (2005).
[42] P. Norman, Phys. Chem. Chem. Phys. **13**, 20519 (2011).
[43] P. Norman, K. Ruud, and T. Helgaker, J. Chem. Phys. **120**, 5027 (2004).
[44] A. Jiemchooroj and P. Norman, J. Chem. Phys. **126**, 134102 (2007).
[45] T. Fahleson and P. Norman, J. Chem. Phys. **147**, 144109 (2017).
[46] B. F. Milne, P. Norman, F. Nogueira, and C. Cardoso, Phys. Chem. Chem. Phys. **15**, 14814 (2013).
[47] B. F. Milne and P. Norman, J. Phys. Chem. A **119**, 5368 (2015).
[48] H. Solheim, K. Ruud, S. Coriani, and P. Norman, J. Phys. Chem. A **112**, 9615 (2008).
[49] H. Solheim, K. Ruud, S. Coriani, and P. Norman, J. Chem. Phys. **128**, 094103 (2008).
[50] J. Cukras, J. Kauczor, P. Norman, A. Rizzo, G. L. J. A. Rikken, and S. Coriani, Phys. Chem. Chem. Phys. **18**, 13267 (2016).
[51] J. Vaara, A. Rizzo, J. Kauczor, P. Norman, and S. Coriani, J. Chem. Phys. **140**, 134103 (2014).
[52] T. Fahleson, H. Ågren, and P. Norman, J. Phys. Chem. Lett. **7**, 1991 (2016).
[53] P. Norman and A. Dreuw, Chem. Rev. **118**, 7208 (2018).
[54] S. Coriani, O. Christiansen, T. Fransson, and P. Norman, Phys. Rev. A **85**, 022507 (2012).

[55]S. Coriani, T. Fransson, O. Christiansen, and P. Norman, J. Chem. Theory Comput. **8**, 1616 (2012).
[56]S. Coriani and H. Koch, J. Chem. Phys. **143**, 181103 (2015).
[57]S. Coriani and H. Koch, J. Chem. Phys. **145**, 149901 (2016).
[58]J. Cukras, S. Coriani, P. Decleva, O. Christiansen, and P. Norman, J. Chem. Phys. **139**, 094103 (2013).
[59]B. N. C. Tenorio, T. Moitra, M. A. C. Nascimento, A. B. Rocha, and S. Coriani, J. Chem. Phys. **150**, 224104 (2019).
[60]T. Wolf, R. Myhre, J. Cryan, S. Coriani, R. Squibb, A. Battistoni, N. Berrah, C. Bostedt, P. Bucksbaum, G. Coslovich, R. Feifel, K. Gaffney, J. Grilj, T. Martinez, S. Miyabe, S. Moeller, M. Mucke, A. Natan, R. Obaid, T. Osipov, O. Plekan, S. Wang, H. Koch, and M. Gühr, Nat. Commun. **8**, 29 (2017).
[61]R. Faber, E. F. Kjønstad, H. Koch, and S. Coriani, J. Chem. Phys. **151**, 144107 (2019).
[62]R. Faber and S. Coriani, J. Chem. Theory Comput. **15**, 520 (2019).
[63]M. L. Vidal, X. Feng, E. Epifanovsky, A. I. Krylov, and S. Coriani, J. Chem. Theory Comput. **15**, 3117 (2019).
[64]J. J. Eriksen, L. M. Solanko, L. J. Nåbo, D. Wüstner, S. P. A. Sauer, and J. Kongsted, Comput. Theory Chem. **1040-1041**, 54 (2014).
[65]J. M. Olsen, K. Aidas, and J. Kongsted, J. Chem. Theory Comput. **6**, 3721 (2010).
[66]J. M. H. Olsen and J. Kongsted, in *Advances in Quantum Chemistry* (Elsevier, 2011), Vol. 61, pp. 107–143.
[67]J. M. H. Olsen, N. H. List, C. Steinmann, A. H. Steindal, M. S. Nørby, and P. Reinholdt, PElib: The Polarizable Embedding library, 2020, see https://gitlab.com/pe-software/pelib-public.
[68]K. Sneskov, T. Schwabe, J. Kongsted, and O. Christiansen, J. Chem. Phys. **134**, 104108 (2011).
[69]E. D. Hedegård, N. H. List, H. J. Aa. Jensen, and J. Kongsted, J. Chem. Phys. **139**, 044101 (2013).
[70]E. D. Hedegård, J. M. H. Olsen, S. Knecht, J. Kongsted, and H. J. Aa. Jensen, J. Chem. Phys. **142**, 114113 (2015).
[71]J. J. Eriksen, S. P. A. Sauer, K. V. Mikkelsen, H. J. Aa. Jensen, and J. Kongsted, J. Comput. Chem. **33**, 2012 (2012).
[72]M. N. Pedersen, E. D. Hedegård, J. M. H. Olsen, J. Kauczor, P. Norman, and J. Kongsted, J. Chem. Theory Comput. **10**, 1164 (2014).
[73]C. Steinmann, J. M. H. Olsen, and J. Kongsted, J. Chem. Theory Comput. **10**, 981 (2014).
[74]N. H. List, M. T. P. Beerepoot, J. M. H. Olsen, B. Gao, K. Ruud, H. J. Aa. Jensen, and J. Kongsted, J. Chem. Phys. **142**, 034119 (2015).
[75]N. H. List, H. J. Aa. Jensen, and J. Kongsted, Phys. Chem. Chem. Phys. **18**, 10070 (2016).
[76]N. H. List, P. Norman, J. Kongsted, and H. J. Aa. Jensen, J. Chem. Phys. **146**, 234101 (2017).
[77]C. Steinmann, QFITLIB: A library to do multipole fitting in quantum chemistry codes, 2020, see https://github.com/cstein/qfitlib.
[78]C. Steinmann and J. Kongsted, J. Chem. Theory Comput. **11**, 4283 (2015).
[79]N. M. Thellamurege and H. Li, J. Chem. Phys. **137**, 246101 (2012).
[80]M. S. Nørby, C. Steinmann, J. M. H. Olsen, H. Li, and J. Kongsted, J. Chem. Theory Comput. **12**, 5050 (2016).
[81]N. H. List, J. M. H. Olsen, and J. Kongsted, Phys. Chem. Chem. Phys. **18**, 20234 (2016).
[82]C. Steinmann, P. Reinholdt, M. S. Nørby, J. Kongsted, and J. M. H. Olsen, Int. J. Quantum Chem. **119**, e25717 (2019).
[83]L. J. Nåbo, J. M. H. Olsen, N. H. List, L. M. Solanko, D. Wüstner, and J. Kongsted, J. Chem. Phys. **145**, 104102 (2016).
[84]D. Hršak, J. M. H. Olsen, and J. Kongsted, J. Chem. Theory Comput. **14**, 1351 (2018).
[85]J. M. H. Olsen, C. Steinmann, K. Ruud, and J. Kongsted, J. Phys. Chem. A **119**, 5344 (2015).
[86]P. Reinholdt, J. Kongsted, and J. M. H. Olsen, J. Phys. Chem. Lett. **8**, 5949 (2017).
[87]S. Huzinaga and A. A. Cantu, J. Chem. Phys. **55**, 5543 (1971).
[88]A. S. P. Gomes and C. R. Jacob, Annu. Rep. Prog. Chem., Sect. C: Phys. Chem. **108**, 222 (2012).
[89]C. R. Jacob and J. Neugebauer, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 325 (2014).
[90]A. S. P. Gomes, C. R. Jacob, and L. Visscher, Phys. Chem. Chem. Phys. **10**, 5353 (2008).
[91]C. R. Jacob, S. M. Beyhan, R. E. Bulo, A. S. P. Gomes, A. W. Götz, K. Kiewisch, J. Sikkema, and L. Visscher, J. Comput. Chem. **32**, 2328 (2011).
[92]S. Höfener, A. S. P. Gomes, and L. Visscher, J. Chem. Phys. **136**, 044104 (2012).
[93]S. Höfener, A. S. P. Gomes, and L. Visscher, J. Chem. Phys. **139**, 104106 (2013).
[94]M. Guidon, J. Hutter, and J. VandeVondele, J. Chem. Theory Comput. **6**, 2348 (2010).
[95]U. Ekström and contributors, XCFun: A library of exchange-correlation functionals with arbitrary-order derivatives, 2020, see https://xcfun.readthedocs.io/.
[96]U. Ekström, L. Visscher, R. Bast, A. J. Thorvaldsen, and K. Ruud, J. Chem. Theory Comput. **6**, 1971 (2010), available at http://jmlr.org/papers/v18/16-107.html.
[97]A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei, J. Mach. Learn. Res. **18**, 1 (2017), available at http://jmlr.org/papers/v18/16-107.html.
[98]S. Reine, E. Tellgren, and T. Helgaker, Phys. Chem. Chem. Phys. **9**, 4771 (2007).
[99]P. Merlot, R. Izsák, A. Borgoo, T. Kjærgaard, T. Helgaker, and S. Reine, J. Chem. Phys. **141**, 094104 (2014).
[100]C. Kumar, H. Fliegl, F. J. Jensen, A. M. Teale, S. Reine, and T. Kjaergaard, Int. J. Quantum Chem. **118**, e25639 (2018).
[101]M. Ziółkowski, B. Jansík, T. Kjærgaard, and P. Jørgensen, J. Chem. Phys. **133**, 014107 (2010).
[102]K. Kristensen, P. Jørgensen, B. Jansík, T. Kjærgaard, and S. Reine, J. Chem. Phys. **137**, 114102 (2012).
[103]K. Kristensen, I.-M. Høyvik, B. Jansik, P. Jørgensen, T. Kjærgaard, S. Reine, and J. Jakowski, Phys. Chem. Chem. Phys. **14**, 15706 (2012).
[104]B. Jansík, S. Høst, K. Kristensen, and P. Jørgensen, J. Chem. Phys. **134**, 194104 (2011).
[105]I.-M. Høyvik, B. Jansík, and P. Jørgensen, J. Chem. Theory Comput. **8**, 3137 (2012).
[106]P. Baudin, P. Ettenhuber, S. Reine, K. Kristensen, and T. Kjærgaard, J. Chem. Phys. **144**, 054102 (2016).
[107]D. Bykov, K. Kristensen, and T. Kjærgaard, J. Chem. Phys. **145**, 024106 (2016).
[108]D. Bykov and T. Kjaergaard, J. Comput. Chem. **38**, 228 (2017).
[109]T. Kjærgaard, J. Chem. Phys. **146**, 044103 (2017).
[110]J. J. Eriksen, P. Baudin, P. Ettenhuber, K. Kristensen, T. Kjærgaard, and P. Jørgensen, J. Chem. Theory Comput. **11**, 2984 (2015).
[111]A. L. Barnes, D. Bykov, D. I. Lyakh, and T. P. Straatsma, J. Phys. Chem. A **123**, 8734 (2019).
[112]P. Baudin and K. Kristensen, J. Chem. Phys. **144**, 224106 (2016).
[113]P. Baudin, D. Bykov, D. Liakh, P. Ettenhuber, and K. Kristensen, Mol. Phys. **115**, 2135 (2017).
[114]P. Baudin, T. Kjærgaard, and K. Kristensen, J. Chem. Phys. **146**, 144107 (2017).
[115]P. Baudin and K. Kristensen, J. Chem. Phys. **146**, 214114 (2017).
[116]V. V. Rybkin, U. Ekström, and T. Helgaker, J. Comput. Chem. **34**, 1842 (2013).
[117]K. Kristensen, P. Ettenhuber, J. J. Eriksen, F. Jensen, and P. Jørgensen, J. Chem. Phys. **142**, 114116 (2015).
[118]T. H. Rasmussen, Y. M. Wang, T. Kjærgaard, and K. Kristensen, J. Chem. Phys. **144**, 204102 (2016).
[119]C. Kumar, T. Kjærgaard, T. Helgaker, and H. Fliegl, J. Chem. Phys. **145**, 234108 (2016).
[120]R. Bast, D. H. Friese, B. Gao, D. J. Jonsson, S. S. Reine, M. Ringholm, and K. Ruud, OpenRSP: An open-ended response property library, 2020, see https://openrsp.org/.
[121]R. Bast and contributors, XCint: Exchange–correlation integrator, 2020, see https://xcint.readthedocs.io/.
[122]R. Di Remigio, A. H. Steindal, K. Mozgawa, V. Weijo, H. Cao, and L. Frediani, Int. J. Quantum Chem. **119**, e25685 (2019).

[123]D. W. Zhang and J. Z. H. Zhang, J. Chem. Phys. **119**, 3599 (2003).
[124]O. Vahtras, LoProp for Dalton, 2014, see https://github.com/vahtras/loprop.
[125]L. Gagliardi, R. Lindh, and G. Karlström, J. Chem. Phys. **121**, 4494 (2004).
[126]B. Gao, Gen1Int: An open-ended integral library, 2012, see https://gitlab.com/bingao/gen1int.
[127]B. Gao, A. J. Thorvaldsen, and K. Ruud, Int. J. Quantum Chem. **111**, 858 (2010).
[128]B. Gao, QcMatrix: An abstract matrix library, 2015, see https://gitlab.com/bingao/qcmatrix.
[129]B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, and T. L. Windus, J. Chem. Inf. Model. **59**, 4814 (2019).
[130]J. D. Hunter, Comput. Sci. Eng. **9**, 90 (2007).
[131]S. Reine, A. Krapp, M. F. Iozzi, V. Bakken, T. Helgaker, F. Pawłowski, and P. Sałek, J. Chem. Phys. **133**, 044102 (2010).
[132]C.-M. Suomivuori, N. O. C. Winter, C. Hättig, D. Sundholm, and V. R. I. Kaila, J. Chem. Theory Comput. **12**, 2644 (2016).
[133]M. Hubert, E. D. Hedegård, and H. J. Aa. Jensen, J. Chem. Theory Comput. **12**, 2203 (2016).
[134]M. Hubert, H. J. Aa. Jensen, and E. D. Hedegård, J. Phys. Chem. A **120**, 36 (2016).
[135]I. B. Nielsen, L. Lammich, and L. H. Andersen, Phys. Rev. Lett. **96**, 018304 (2006).
[136]L. H. Andersen, I. B. Nielsen, M. B. Kristensen, M. O. A. El Ghazaly, S. Haacke, M. B. Nielsen, and M. Å. Petersen, J. Am. Chem. Soc. **127**, 12347 (2005).
[137]H. J. Aa. Jensen, P. Jørgensen, H. Ågren, and J. Olsen, J. Chem. Phys. **88**, 3834 (1988).
[138]A. Cembran, R. González-Luque, P. Altoè, M. Merchán, F. Bernardi, M. Olivucci, and M. Garavelli, J. Phys. Chem. A **109**, 6597 (2005).
[139]G. Dong, U. Ryde, H. J. Aa. Jensen, and E. D. Hedegård, Phys. Chem. Chem. Phys. **20**, 794 (2018).
[140]M. Scheurer, P. Reinholdt, E. R. Kjellgren, J. M. H. Olsen, A. Dreuw, and J. Kongsted, J. Chem. Theory Comput. **15**, 6154 (2019).
[141]M. T. P. Beerepoot, D. H. Friese, N. H. List, J. Kongsted, and K. Ruud, Phys. Chem. Chem. Phys. **17**, 19306 (2015).
[142]K. O. H. Dundas, M. Ringholm, Y. Cornaton, and B. Ofstad, SpectroscPy: Python tools for spectroscopy, 2020, see https://gitlab.com/kdu002/SpectroscPy.
[143]A. Thorvaldsen, K. Ruud, K. Kristensen, P. Jørgensen, and S. Coriani, J. Chem. Phys. **129**, 214108 (2008).
[144]M. Ringholm, D. Jonsson, and K. Ruud, J. Comput. Chem. **35**, 622 (2014).
[145]S. Coriani, S. Høst, B. Jansík, L. Thøgersen, J. Olsen, P. Jørgensen, S. Reine, F. Pawłowski, T. Helgaker, and P. Sałek, J. Chem. Phys. **126**, 154108 (2007).
[146]J. Bloino and V. Barone, J. Chem. Phys. **136**, 124108 (2012).
[147]E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra* (Courier Corporation, 1980).
[148]T. Helgaker, Acta Chem. Scand. **42a**, 515 (1988).
[149]Z. Rinkevicius, I. Tunell, P. Sałek, O. Vahtras, and H. Ågren, J. Chem. Phys. **119**, 34 (2003).
[150]Z. Rinkevicius, P. C. Jha, C. I. Oprea, O. Vahtras, and H. Ågren, J. Chem. Phys. **127**, 114101 (2007).
[151]Z. Rinkevicius, L. Telyatnyk, P. Sałek, O. Vahtras, and H. Ågren, J. Chem. Phys. **119**, 10489 (2003).
[152]Z. Rinkevicius, L. Telyatnyk, O. Vahtras, and H. Ågren, J. Chem. Phys. **121**, 7614 (2004).
[153]X. Chen, Z. Rinkevicius, Z. Cao, K. Ruud, and H. Ågren, Phys. Chem. Chem. Phys. **13**, 696 (2011).
[154]Z. Rinkevicius, K. J. de Almeida, and O. Vahtras, J. Chem. Phys. **129**, 064109 (2008).
[155]C. I. Oprea, L. Telyatnyk, Z. Rinkevicius, O. Vahtras, and H. Ågren, J. Chem. Phys. **124**, 174103 (2006).
[156]D. Duflot, J.-P. Flament, I. C. Walker, J. Heinesch, and M.-J. Hubin-Franskin, J. Chem. Phys. **118**, 1137 (2003).