

Sivert Skovereng

# Parametrization of a classical force field model (CT-PDI) for the calculation of molecular polarizabilities

Master's thesis in Chemistry  
Supervisor: Per-Olof Åstrand  
June 2023



Sivert Skovereng

# **Parametrization of a classical force field model (CT-PDI) for the calculation of molecular polarizabilities**

Master's thesis in Chemistry  
Supervisor: Per-Olof Åstrand  
June 2023

Norwegian University of Science and Technology  
Faculty of Natural Sciences  
Department of Chemistry







## **Acknowledgements**

This masters thesis was conducted under the supervision of Per-Olof Åstrand, whose help and guidance in this project has been greatly appreciated. Further i would like to thank my parents who have been a source of unwavering support and motivation through everything. Finally I would like to thank all my friends for their support.

## Abstract

Two parametrizations of a classical electrostatic model for the polarizability has been undertaken. The parametrizations were conducted on a large set of small organic molecules, containing many normal functional groups. For one parametrization, the set of molecules were supplemented with molecules where one of the bonds had been stretched or compressed. This was done such in order to include a larger number of less represented bond lengths to the parametrization. Care was taken not to stretch bonds between bond orders.

It was found that carbon-carbon bond lengths in the interval  $1.46 - 1.48 \text{ \AA}$  were under-represented. These consisted of a number of aromatic bonds and bonds near electronegative functional groups. A number of carbon bonds were added in this interval. It was found that this had little effect on the parametrization. A smaller number of carbon-fluorine bonds were also added. This was found to give better effect, but mainly due to these bonds initially representing a far smaller number of the total bonds. It shows that adding geometries like this may have some effect for under-represented bond distances, but that the focus should be on increasing the number of bonds that are under-represented for the whole data set, not just within each category of bonds.

The optimization algorithm used was found to be able to get stuck at certain parameter values if they were initially set to be zero. The model compensated for this by changing other parameter values in order to minimize the error. This showed a weakness in using molecular properties to optimize atomic contributions. Care should be taken with the initial parameter values to avoid this.

## Sammendrag

To parametriseringer av en klassisk elektrostatiske modell for polariserbarheten er blitt gjennomført. Parametriseringene er gjennomført på et stort sett av små organiske molekyler, med mange vanlige funksjonelle grupper. For den ene parametriseringen ble settet med molekyler utvidet med molekyler der geometrien hadde blitt endret for å komprimere eller strekke en av de kjemiske bindingene. Dette ble gjort med hensikten å forbedre beskrivelsen av disse underrepresenterte kjemiske bindingslengdene. Bindinger ble ikke strukket utover egen bindingsorden.

Det ble funnet at for karbon-karbon bindinger var bindingslenger mellom 1.46 – 1.48 Å var underrepresentert. Disse bindingene var sammensatt av aromatiske bindinger og bindinger nær elektronegative funksjonelle grupper. Flere karbon-karbon bindinger ble lagt til i dette intervallet. Det ble funnet at dette hadde liten effekt på parametriseringen. Et mindre antall karbon-fluor bindinger ble lagt til. Dette ble funnet å gi bedre effekt for parametriseringen, muligens fordi disse bindingene opprinnelig utgjorde et mye mindre antall av de totale bindingene. Dette viser at å legge til flere geometrier kan ha et positivt bidrag for underrepresenterte bindinger, men at fokuset av disse geometriene burde være på å legge til bindinger som er underrepresentert i hele datasettet, ikke bare innenfor hver kategori av bindinger.

Optimeringsalgoritmen som ble brukt ble funnet å kunne sette seg fast på visse parameterverdier dersom de opprinnelig ble satt til null. Det ble observert at modellen kunne kompensere for dette med å endre andre parameterverdier for å minimere feilen. Dette demonstrerer en svakhet ved å bruke molekylære egenskaper for å optimalisere parameterene for atomiske bidrag. For å unngå dette må aktsomhet vises for valg av de initiale parameterverdiene.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theory</b>	<b>2</b>
2.1	Quantum chemistry . . . . .	2
2.1.1	The Schrödinger equation . . . . .	2
2.1.2	Born-Oppenheimer Approximation . . . . .	2
2.1.3	Orbitals and Slater determinants . . . . .	3
2.1.4	Basis sets . . . . .	3
2.1.5	Variation theorem . . . . .	4
2.1.6	Hartree-Fock equations . . . . .	4
2.2	Density functional theory (DFT) . . . . .	5
2.2.1	Kohn-Sham DFT . . . . .	6
2.2.2	Exchange correlation functionals . . . . .	7
2.3	Perturbation theory . . . . .	8
2.3.1	Time independent perturbation theory . . . . .	8
2.3.2	Time dependent perturbation theory . . . . .	10
2.3.3	The Frequency dependent polarizability . . . . .	11
2.4	Force-field models for polarizability . . . . .	13
2.4.1	Electronegativity equilization model . . . . .	13
2.4.2	Point dipole model . . . . .	14
2.4.3	Drude oscillator . . . . .	15
2.5	CT-PDI model . . . . .	16
2.5.1	Representation of charges . . . . .	17
2.5.2	Charge-Charge interaction energy . . . . .	17
2.5.3	Charge-dipole interaction energy . . . . .	18
2.5.4	Dipole-dipole interaction energy . . . . .	19
2.5.5	Modification of the atomic polarizability . . . . .	19
2.5.6	Modification of the chemical hardness . . . . .	20
2.6	Calculation of the molecular polarizability in the CT-PDI model . . . . .	21
2.6.1	Evaluation of Euler-Lagrange terms . . . . .	22
2.7	Calculation of the dipole gradient . . . . .	24
2.7.1	Electric field shielding tensor . . . . .	25
2.7.2	Polarization . . . . .	26
2.7.3	Perturbation theory polarization gradient . . . . .	27
2.7.4	Expression of the shielding tensor . . . . .	28
2.8	The dipole gradient in the CT-PDI model . . . . .	29
<b>3</b>	<b>Method</b>	<b>29</b>
3.1	DFT-Calculations . . . . .	29

3.1.1	Choice of functional . . . . .	29
3.1.2	Choice of basis sets . . . . .	30
3.2	Modification of geometries . . . . .	30
3.3	Generation of parameter input files . . . . .	31
3.4	Optimization of parameters . . . . .	31
<b>4</b>	<b>Results and discussion</b>	<b>33</b>
4.1	Parametrization of the force field . . . . .	33
4.2	Description of bond distributions . . . . .	35
4.2.1	Carbon-Carbon bonds . . . . .	35
4.2.2	Carbon-Hydrogen bonds . . . . .	36
4.2.3	Carbon-Fluorine bonds . . . . .	37
4.2.4	Carbon-Nitrogen bonds . . . . .	38
4.2.5	Carbon-Oxygen bonds . . . . .	38
4.2.6	Carbon-Sulfur bonds . . . . .	39
4.2.7	Carbon-Chlorine bonds . . . . .	40
4.3	Optimized parameters . . . . .	41
4.3.1	Static polarizability parameters . . . . .	42
4.3.2	Frequency dependent polarizability parameters . . . . .	43
4.3.3	Smaller parametrization using different bonds . . . . .	43
4.3.4	Comparison of chlorinated molecules . . . . .	45
4.3.5	Parametrization of hydrocarbons . . . . .	49
4.3.6	Parametrization of oxygen . . . . .	54
4.4	The effect of stretching a bond . . . . .	55
4.5	Result of the addition of stretched geometries . . . . .	58
4.5.1	Addition of carbon bonds . . . . .	60
4.6	Parametrization of cyclic compounds . . . . .	65
<b>5</b>	<b>Conclusion</b>	<b>68</b>
	<b>References</b>	<b>78</b>
	<b>Appendix A Molecules used in parametrization</b>	<b>79</b>

# 1 Introduction

The use of computational models are a driving force for research on chemical systems [1]. While much research being conducted employs quantum chemical calculations, for large enough systems the computational load of purely quantum mechanical solutions are still very demanding [2, 3]. To simulate larger systems, such as biomolecules [4] or condensed phases [5] the use force field methods are often a requirement [6], still today it is not possible to simulate very ambitious systems in complete detail, even using force fields [7]. Still work is being done to improve and introduce new force fields [8–11].

The polarizability is a molecular property with great importance for the behaviour of many simulations [12]. The molecular polarizability is connected to many important properties like induction and dispersion, and is important for spectroscopic determination of molecules [13]. For simulations of ionic liquids, the inclusion of polarizability may be necessary for accurate results [14, 15], and many of the best performing force field do include polarization [16], for which the polarizability is an important property. Many applications of force field methods exists for use on fullerenes and graphene systems, and the polarizabilities have been shown to be of great importance in such systems [17, 18]. Still the proper incorporation of polarizability in classical force fields continues to be researched [19].

The dipole moment gradient is another molecular property, it is related to the intensities in the infrared spectre [20–23] and to Raman intensities [22]. The inclusion of the dipole moment gradient may be a better metric for a parametrization of the force field because it is a size localized to an atom, and may thus be better tied to the atomic parameters.

Quite some work has gone into the development of the charge transfer - point dipole interaction (CT-PDI) model [24–28]. The model consists of the combination of an electronegativity equalization model, where atomic charges are redistributed to form an equal chemical potential at each atom [29], and a point dipole model that places a dipole at polarizable points in the system [30]. The ct-pdi model has been used for the calculation of local fields, which can be used to find areas of a molecular system where rare events can occur [31]

A good parametrization is necessary to get useful data from a force field model [32, 33]. While many force fields exist, many are focused on certain groups of molecules [34] it would be ideal to generate a force field which has good transferability of parameters between larger and smaller molecules.

In this project, a large set of molecules have been generated, in bond length intervals where a smaller representation of bond lengths were found, the set was supplemented by modifying certain geometries to include these abnormal geometries, and the effect of this on the parametrization has been investigated.

## 2 Theory

### 2.1 Quantum chemistry

#### 2.1.1 The Schrödinger equation

To examine chemical systems in full detail, it is necessary to include the effects of quantum mechanics. The most fundamental equation in quantum chemistry is the Schrödinger-equation eq. (38) [35].

$$\hat{\mathcal{H}}\Psi(x, t) = i\hbar\frac{\partial\Psi(x, t)}{\partial t} = E\Psi(x, t) \quad (1)$$

Where  $i$  is the imaginary number,  $\hbar$  is the reduced Planck constant,  $E$  is the total energy, and  $\Psi$  is the wave function. The time dependent Schrödinger-equation eq. (38) describes the time evolution of the wavefunction [35]. For cases where the Hamiltonian  $\hat{\mathcal{H}}$  is independent of time [36], it can be shown that the Schrödinger equation can be separated into a time independent, and a time dependent part. This leads to a formulation of the time independent Schrödinger equation

$$\hat{\mathcal{H}}\Psi(x) = E\Psi(x) \quad (2)$$

This formulation has had great importance in the development of approximate solutions to the Schrödinger equation. Despite the ability to formulate the equation for the development of the system, the problem is not analytically solvable. This is due to the Schrödinger equations spatial dependence on the coordinates of the particles in the system [35].

#### 2.1.2 Born-Oppenheimer Approximation

The fundamental problem of solving the Schrödinger equation is the correlated movement of the quantum particles. This has the consequence of the movement of all the particles in the system being coupled, which is problematic for finding a solution. For nuclei, which have a much larger mass than electrons, it can be assumed that the coupling between the movement is negligible [35]. This would then allow for the separation of the total wavefunction for the system into a nuclear and a electronic wavefunction [36–38].

$$\Psi(x) = \Psi^{el}(x)\Psi^{nuc}(x) \quad (3)$$

This can be seen as the electronic wavefunction being able to instantaneously adjust to changes in changes in the nuclear configuration [39] and allows for solving the two

wavefunctions separately, greatly simplifying the solution.

### 2.1.3 Orbitals and Slater determinants

When treating a many electron wavefunction, the problem of the correlated movements of electrons returns. The crudest way to approximate the wavefunction would then be to neglect this correlation and simply define the wavefunction as a product of many one electron, non-interacting wavefunctions. This is the simple Hartree-product [35]. By exchanging labels for the electrons in the Hartree product, a Slater determinant can be constructed [36]

$$\Phi = \frac{1}{\sqrt{N!}} \det \begin{vmatrix} \chi_1(1) & \chi_2(1) & \cdots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \cdots & \chi_N(2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(N) & \cdots & \cdots & \chi_N(N) \end{vmatrix} \quad (4)$$

This Slater determinant accounts for the symmetry constraints on the wavefunction imposed by the Pauli exclusion principle.

### 2.1.4 Basis sets

For non-periodic systems, the orbitals may be represented by a linear combination of atom centered functions [36]. One of the most common ways to represent these functions is Gaussian-type orbitals. These are functions on the form [35, 36]:

$$\phi(r) = N x^i y^j z^k \exp\{-ar^2\} \quad \text{with } i, j, k \geq 0 \quad (5)$$

Where  $a$  is an exponent and  $N$  is a normalization constant. These have the central advantage that the product of such Gaussians is a Gaussian centered between the two. However, these functions do not have a cusp at  $r = 0$ , but a linear combination of Gaussians will approach such behaviour. In this way a contracted Gaussian can be constructed from a linear combination of primitive Gaussian orbitals eq. (5) [35]

$$\chi_o = \sum_i d_{oi} \phi_i(r) \quad (6)$$

Using a single contracted Gaussian eq. (6) to describe an orbital is often not sufficient to approximate the wavefunction with good precision. Instead one can use two contracted Gaussians for each orbital. This is what is known as a double-zeta basis set. Further one can use three Gaussians for this, in this case it is a tripple-zeta basis set. As will be



shown, using a larger basis set will improve the resulting approximated wavefunction, but increase the computational cost [35].

For the calculation of properties, it is common to also add diffuse functions, Gaussians with a smaller parameter value for  $a$  eq. (5) such that it is more diffuse and has a greater density near the tails of the distribution [36].

### 2.1.5 Variation theorem

The variation theorem states that if the energy of any approximated wavefunction  $\tilde{\Phi}$  is,

$$\tilde{E} = \frac{\langle \tilde{\Phi} | H | \tilde{\Phi} \rangle}{\langle \tilde{\Phi} | \tilde{\Phi} \rangle} \quad (7)$$

Then the energy of this wavefunction is greater or equal to the true ground state energy  $E_0$ .

$$\tilde{E} \geq E_0 \quad (8)$$

This has the strength that if any improvement is made to the estimated energy  $\tilde{E}$ , by varying the wavefunction  $\tilde{\Phi}$ , then the resulting wavefunction is a better approximation of the true wavefunction and it provides an upper bound to the ground state energy [35].

### 2.1.6 Hartree-Fock equations

Considering the Hamiltonian of one electron orbitals, moving in a potential of the average of interactions with all other particles in the system. The Hamiltonian for such a system can be expressed as [35]

$$\hat{H} = \sum_i h_i + \frac{1}{2} \sum_{i \neq j} \frac{1}{r_{ij}} \quad (9)$$

Where  $h_i$  is the one-electron hamiltonian for non-interacting electrons and  $\frac{1}{r_{ij}}$  is the coulomb interaction. Atomic units have been used to simplify constants. Applying this to a wavefunction approximated by a slater determinant eq. (4), the energy is the expectation value [35]

$$E = \left\langle \Phi \left| \sum_i h_i + \frac{1}{2} \sum_{i \neq j} \frac{1}{r_{ij}} \right| \Phi \right\rangle \quad (10)$$

Applying the Slater-Cordon rules [35], the energy becomes

$$E = \sum_i^N \langle \phi_i | h_1 | \phi_i \rangle + \frac{1}{2} \sum_{i \neq j}^N \left( \langle \phi_i \phi_j | \frac{1}{r_{ij}} | \phi_i \phi_j \rangle - \langle \phi_i \phi_j | \frac{1}{r_{ij}} | \phi_j \phi_i \rangle \right) \quad (11)$$

Minimizing this, with respect to the energy  $\delta E$ , using the orthonormality of the one electron orbitals of the slater determinant  $\delta \langle \psi_i | \psi_j \rangle$  one obtains [35]:

$$\sum_i^N \langle \delta \phi_i | h_i | \phi_i \rangle + \frac{1}{2} \sum_{i \neq j}^N \left( \langle \delta \phi_i \phi_j | \frac{1}{r_{ij}} | \phi_i \phi_j \rangle - \langle \delta \phi_i \phi_j | \frac{1}{r_{ij}} | \phi_j \phi_i \rangle \right) - \sum_{i,j}^N \epsilon_{i,j} \langle \delta \phi_i | \phi_j \rangle = 0 \quad (12)$$

Since  $\delta \phi_i$  is chosen arbitrarily, the sum of all terms must be zero [35].

$$\sum_i^N h_i | \phi_i \rangle + \frac{1}{2} \sum_{i \neq j}^N \left( \langle \phi_j | \frac{1}{r_{ij}} | \phi_i \rangle | \phi_j \rangle - \langle \phi_j | \frac{1}{r_{ij}} | \phi_j \rangle | \phi_i \rangle \right) \sum_{i,j}^N \epsilon_{i,j} | \phi_j \rangle = 0 \quad (13)$$

Introducing the Coulomb operator  $J$  and the exchange operator  $K$ , for each orbital [35]

$$h_i \phi_i + \frac{1}{2} \sum_j^N (J_j \phi_i - K_j \phi_i) = \sum_j^N \epsilon_{i,j} \phi_j \quad (14)$$

Then the Fock operator can be defined [35,36]

$$f_1 = h_1 + \sum_{j=1}^N J_j(1) - K_j(1) \quad (15)$$

This operator then can be used to find the orbital energies.

## 2.2 Density functional theory (DFT)

Density functional theory (DFT) is a widely used quantum mechanical method [35,40]. Very generally the concept of DFT is that the energy of an electronic system can be written in terms of the electron density  $\rho$  of the system [35]. This offers many advantages over the wavefunction methods described above.

Fundamental for the development of DFT as a method for quantum chemical calculations are the Hohenberg-Kohn theorems [41]. The first of these theorems show that the

electron density  $\rho(r)$  can be used to uniquely define the external potential caused by the nuclei, and that its normalization determines the number of electrons  $N$  [41]. This allows for the determination of all the ground state properties of the molecules. This can be expressed as [36]

$$E[\rho(r)] = \int V_{ext}\rho(r)dr + F[\rho(r)] \quad (16)$$

With  $V_{ext}$  being the external potential,  $F$  being the remaining energy terms, and  $\rho(r)$  is the electron density.

The second of the Hohenberg-Kohn theorems states that using variational theory, the energy of a system does not change with a variation of the electron density  $\rho(r)$ , given that  $\rho(r)$  is always normalized to  $N$  in integration [41]. This then implies [36]

$$\frac{\delta}{\delta\rho(r)} \left( E[\rho(r)] - \mu \int \rho(r)dr \right) = 0 \quad (17)$$

Where  $E[\rho]$  is the energy from eq. (16) and  $\mu$  is a Lagrangian multiplier that can be associated with the chemical potential [36].

### 2.2.1 Kohn-Sham DFT

In the Kohn-Sham approach, the density functional is formulated in a set of solvable equations [35]. This includes partitioning the energy in eq. (16) as

$$E[\rho(r)] = \int V_{ext}\rho(r)dr + E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)] \quad (18)$$

Where  $E_{KE}$  describes the kinetic energy of an ideal electron gas,  $E_H$  corresponds to the coulomb term from the hartree-fock approximation eq. (14), and  $E_{XC}$  is the exchange correlation energy. By introducing orbitals, like in the Hartree-Fock approach, an estimate of the electronic density can be found by a linear combination of orbitals [35,36,42].

$$\rho(r) = \sum_{i=1}^N |\phi(r_i)|^2 \quad (19)$$

Using eq. (17) on the terms in eq. (18) this leads to an expression for the orbital energies [36,43,44].

$$\hat{f}_I^{KS} \phi_i = (V_{ext} + V_{KE} + V_H + V_{XC}) \phi_i = \epsilon_i \phi_i \quad (20)$$

Where  $\hat{f}_I^{KS}$  can be seen as an analogy to eq. (15). This includes the term  $V_{XC}$  which is the exchange-correlation functional  $V_{XC}$ . This term is not immediately known, and many ways of approximating this functional has been proposed [35].

### 2.2.2 Exchange correlation functionals

The exchange correlation functional  $V_{XC}$  eq. (20), can be separated into two contributions [35, 42, 43]

$$V_{XC} = V_X + V_C \quad (21)$$

Due to the KS-DFT method sharing the orbitals with the Hartree-Fock method, the exact exchange can be calculated using this the Hartree-Fock equations [42, 45]. However as the Hartree term in eq. (20) contains self interactions, simply using this does not necessarily give a good approximation for the correct exchange correlation potential  $V_{XC}$  [42]. The full form of the exchange correlation potential  $V_{XC}$  is however unknown, and much work has been done to approximate it [35, 46].

One method of approximating the exchange correlation functional is the use of the local spin density approximation (LSDA). Using this the energy can be estimated from the exchange correlation energy of the particles in a uniform electron gas [47, 48].

$$E_{xc}^{LSDA} = \int \epsilon_{xc}^{unif}(n_{\uparrow}, n_{\downarrow}) d^3r \quad (22)$$

Where  $\epsilon_{xc}^{unif}$  is the energy of particles in a uniform electron gas [48], and the arrows signify spins. To improve on this, the generalized gradient approach can be used. Here  $\epsilon_{xc}$  is chosen to be a function, not only of the spin density, but also of the gradient of the spin density [48]. This gives

$$E_{xc}^{GGA} = \int \epsilon_{xc}^{GGA}(n_{\uparrow}, n_{\downarrow}, |\nabla n_{\uparrow}|, |\nabla n_{\downarrow}|) d^3r \quad (23)$$

The next development involved the addition of the kinetic energy density [35]

$$\tau(r, n_{\uparrow/\downarrow}) = \frac{\hbar^2}{2m_e} \sum_i \nabla \phi_i^*(r) \nabla \phi_i(r) \quad (24)$$

Where  $m_e$  is the electron mass, and  $\phi_i$  is an occupied orbital. Functionals including this kinetic energy density are known as meta-GGA functionals [35, 44].

$$E_{xc}^{m-GGA} = \int \epsilon_{xc}^{m-GGA}(n_\uparrow, n_\downarrow, |\nabla n_\uparrow|, |\nabla n_\downarrow|, \tau_\uparrow(r), \tau_\downarrow(r)) d^3r \quad (25)$$

A problem with these approaches (collectively) is that they fail at longer range due to them giving little knowledge about the system far from an electron [47]. To overcome this it is possible to include the exact Hartree-Fock exchange energy, which can be calculated due to the similar choice of orbitals between the methods. Then one or more parameters are chosen, along with approximations for the energy to make up the total exchange-correlation energy functional. Some of these hybrid functionals also are range separated, with the idea being to include mostly DFT-exchange at short range, and mostly Hartree-Fock exchange at long range [49]. These then have a smooth transition from short range to long, with a percentage making up the different part of the energy at different ranges [50].

## 2.3 Perturbation theory

An important method for the approximation of the eigenvalue problem for the hamiltonian is perturbation theory [35]. Here the change of the system caused by a small perturbation of the system is divided into the assumed known solution to an unperturbed system, and an additional contribution caused by the perturbation [36].

### 2.3.1 Time independent perturbation theory

In Rayleigh-Schrödinger perturbation theory the Schrödinger equation can be written [36]

$$\hat{\mathcal{H}} |\psi_k\rangle = (\hat{\mathcal{H}}_0 + \lambda \hat{\mathcal{H}}_1) |\psi_k\rangle = \epsilon_k |\psi_k\rangle \quad (26)$$

Where  $\hat{\mathcal{H}}_0$  is the unperturbed hamiltonian,  $\hat{\mathcal{H}}_1$  is the perturbation of the hamiltonian, and  $\lambda$  is an order parameter. Including more terms the perturbed Schrödinger equation eq. (26), giving [51]

$$(\hat{\mathcal{H}}_0 + \lambda \hat{\mathcal{H}}_1) \left( \sum \lambda^n |\psi_k^{(n)}\rangle \right) = \left( \sum \lambda^n \epsilon_k^{(n)} \right) \left( \sum \lambda^n |\psi_k^{(n)}\rangle \right) \quad (27)$$

Where the sum is over  $n$  and,  $\psi_k^{(n)}$  and  $\epsilon_k^{(n)}$  represents the  $n$ -th order correction to the wavefunction. This equation must hold for all orders  $n$  in  $\lambda^n$  [36]. The expanded sum can be rearranged for orders of  $\lambda$  to give equations for the lowest order corrections. It is assumed the perturbed wavefunction obeys intermediate normalization such that

$\langle \psi_k^{(n)} | \psi_k^{(m)} \rangle = \delta_{nm}$  [36, 51]. Expanding eq. (27) by orders the first and second order corrections to the energy can be written as

$$\epsilon_k^{(1)} = \langle \psi_k^{(0)} | \hat{\mathcal{H}}_1 | \psi_k^{(0)} \rangle \quad (28)$$

$$\epsilon_k^{(2)} = \langle \psi_k^{(0)} | \hat{\mathcal{H}}_1 | \psi_k^{(1)} \rangle \quad (29)$$

Because the unperturbed hamiltonian constitutes a complete basis set of functions [35], the first order correction to the wavefunction can be written as

$$|\psi_k^{(1)}\rangle = \sum_{j=1} c_{j,k}^{(1)} |\psi_j^{(0)}\rangle \quad (30)$$

Looking at the terms of eq. (27) for order  $n = 1$ , this is

$$(\epsilon_k^{(0)} - \hat{\mathcal{H}}_0) |\psi_k^{(1)}\rangle = (\hat{\mathcal{H}}_1 - \epsilon_k^{(1)}) |\psi_k^{(0)}\rangle \quad (31)$$

Multiplying with  $\langle \psi_j^{(0)} |$  gives

$$(\epsilon_k^{(0)} - \epsilon_j^{(0)}) \langle \psi_j^{(0)} | \psi_k^{(1)} \rangle = \langle \psi_j^{(0)} | \hat{\mathcal{H}}_1 | \psi_k^{(0)} \rangle - \epsilon_k^{(1)} \langle \psi_j^{(0)} | \psi_k^{(0)} \rangle \quad (32)$$

For all  $j \neq k$  this gives

$$\langle \psi_j^{(0)} | \psi_k^{(1)} \rangle = \frac{\langle \psi_j^{(0)} | \hat{\mathcal{H}}_1 | \psi_k^{(0)} \rangle}{\epsilon_k^{(0)} - \epsilon_j^{(0)}} \quad (33)$$

Which is equivalent to eq. (30) multiplied with  $\langle \psi_j^{(0)} |$ , giving an expression

$$|\psi_k^{(1)}\rangle = \sum_{j \neq k} \frac{\langle \psi_j^{(0)} | \hat{\mathcal{H}}_1 | \psi_k^{(0)} \rangle}{\epsilon_k^{(0)} - \epsilon_j^{(0)}} |\psi_j^{(0)}\rangle \quad (34)$$

This can also be inserted into eq. (29) to find the second order correction to the energy [36].

### 2.3.2 Time dependent perturbation theory

For a time dependent system, the hamiltonian can be expressed as [35, 36]

$$\hat{\mathcal{H}}(t) = \hat{\mathcal{H}}_0 + \lambda \hat{\mathcal{H}}_1(t) \quad (35)$$

Where  $\hat{\mathcal{H}}_0$  is a time independent hamiltonian of the system, and  $\hat{\mathcal{H}}_1(t)$  is the perturbation. In this time dependent case the time dependent Schrödinger equation eq. (38) can be used and the time dependent unperturbed wavefunction can be written as [35, 36]

$$\psi_n^{(0)}(t) = \psi_n^{(0)} \exp\{-i\omega_n t\} \quad (36)$$

Where  $\omega_n = \frac{E_n^{(0)}}{\hbar}$ . The wavefunction can then be written as [36]

$$|\psi^{(0)}(t)\rangle = \sum_n a_n(t) \exp\{-i\omega_n t\} |\psi_n^{(0)}\rangle \quad (37)$$

Where  $a_n(t)$  is a time dependent coefficient. Inserting this into the time dependent Schrödinger equation eq. (38) gives

$$\hat{\mathcal{H}} \sum_n a_n(t) \exp\{-i\omega_n t\} |\psi_n^{(0)}\rangle = i\hbar \frac{\partial}{\partial t} \sum_n a_n(t) \exp\{-i\omega_n t\} |\psi_n^{(0)}\rangle \quad (38)$$

Using the product rule on the right hand side and eq. (35) on the left, this becomes [36]

$$\begin{aligned} & \sum_n a_n(t) \exp\{-i\omega_n t\} \hat{\mathcal{H}}_0 |\psi_n^{(0)}\rangle + \lambda \hat{\mathcal{H}}_1(t) \sum_n a_n(t) \exp\{-i\omega_n t\} |\psi_n^{(0)}\rangle \\ & = i\hbar \sum_n \frac{\partial a_n(t)}{\partial t} \exp\{-i\omega_n t\} |\psi_n^{(0)}\rangle + \epsilon_n^{(0)} \sum_n a_n(t) \exp\{-i\omega_n t\} |\psi_n^{(0)}\rangle \end{aligned} \quad (39)$$

Where  $\omega_n = \frac{E_n^{(0)}}{\hbar}$  has been used. As  $\hat{\mathcal{H}}_0 |\psi_n^{(0)}\rangle = \epsilon_n^{(0)} |\psi_n^{(0)}\rangle$ , the first and last term cancel [35, 36]. Multiplying from the left with  $\langle \psi_m^{(0)} | \exp\{-i\omega_m t\}$

$$\sum_n a_n(t) \langle \psi_m^{(0)} | \lambda \hat{\mathcal{H}}_1(t) |\psi_n^{(0)}\rangle \exp\{-i(\omega_m - \omega_n)t\} = i\hbar \frac{\partial a_m(t)}{\partial t} \quad (40)$$

Where the sum disappeared due to the orthonormal states If  $a_n$  can be expanded in terms of the order parameter  $\lambda$  so  $a_n = \sum_i \lambda^i a_n^{(i)}$  if  $i = 1$  [36], the equation becomes:

$$i\hbar \frac{\partial a_m^{(1)}(t)}{\partial t} = \sum_n a_n^{(0)}(t) \langle \psi_m^{(0)} | \hat{\mathcal{H}}_1(t) | \psi_n^{(0)} \rangle \exp\{-i(\omega_m - \omega_n)t\} \quad (41)$$

Assuming the system was in state  $p$  at  $t = 0$  then the unperturbed coefficient  $a_n^{(0)}$  is  $\delta_{pn}$  in order to make sure this condition is filled. This makes the sum disappear. Integrating the surviving term yields [36]

$$a_m^{(1)}(t) - a_m^{(1)}(0) = \frac{1}{i\hbar} \int_0^t \langle \psi_m^{(0)} | \hat{\mathcal{H}}_1(t') | \psi_p^{(0)} \rangle \exp\{-i(\omega_m - \omega_p)t'\} dt' \quad (42)$$

Setting  $a_m^{(1)}(0) = 0$  by an initial condition gives [35]

$$a_m^{(1)}(t) = \frac{1}{i\hbar} \int_0^t \langle \psi_m^{(0)} | \hat{\mathcal{H}}_1(t') | \psi_p^{(0)} \rangle \exp\{-i(\omega_m - \omega_p)t'\} dt' \quad (43)$$

### 2.3.3 The Frequency dependent polarizability

Considering a system being perturbed by an oscillating electric field, the first order hamiltonian can be written [35, 36]

$$\hat{\mathcal{H}}_1(t) = 2\hat{V}e^{i\omega t} \cos \omega t = \hat{V}(e^{(\epsilon+i\omega)t} + e^{(\epsilon-i\omega)t}) \quad (44)$$

Where  $e^{\epsilon t}$  will approach zero when the field has been on for a long duration. Inserting this into eq. (43) gives

$$a_m^{(1)}(t) = \frac{1}{i\hbar} \int_0^t \langle \psi_m^{(0)} | \hat{V} | \psi_p^{(0)} \rangle (e^{(\epsilon+i\omega)t'} + e^{(\epsilon-i\omega)t'}) \exp\{-i(\omega_m - \omega_p)t'\} dt' \quad (45)$$

Solving the integral yields [36]

$$a_m^{(1)}(t) = \frac{1}{i\hbar} \langle \psi_m^{(0)} | \hat{V} | \psi_p^{(0)} \rangle \left( \frac{\exp\{(\epsilon + i\omega_{kp} + i\omega)t\}}{\omega_{kp} + \omega - i\epsilon} + \frac{\exp\{(\epsilon + i\omega_{kp} - i\omega)t\}}{\omega_{kp} - \omega - i\epsilon} \right) \quad (46)$$

Where it is used that  $\omega_{kp} = \omega_p - \omega_p$ . Allowing  $\epsilon$  to go to zero after a long time simplifies the result. As the polarizability is the field induced contribution to the dipole moment,



the time dependent dipole moment must be considered up to the first order [35]. This becomes [35,36]

$$\mu_\alpha^{(0)} = \langle \psi_0^{(0)} | \hat{\mu}_\alpha | \psi_0^{(0)} \rangle \quad (47)$$

And

$$\mu_\alpha^{(1)} = \sum_{k \neq 0} \left( \langle \psi_0^{(0)} | \hat{\mu}_\alpha | \psi_k^{(0)} \rangle a_k^{(1)} \exp\{-i\omega_{k0}t\} + \langle \psi_k^{(0)} | \hat{\mu}_\alpha | \psi_0^{(0)} \rangle a_k^{(1)*} \exp\{i\omega_{k0}t\} \right) \quad (48)$$

Where  $\alpha$  represents a Cartesian coordinate  $x, y$  or  $z$  and  $*$  indicating the complex conjugate. Inserting for  $a_k^{(1)}$  using eq. (46) with  $\hat{V} = -\hat{\mu}_\beta E_\beta$  This becomes [36]

$$\begin{aligned} \mu_\alpha^{(1)} = \frac{E_\beta}{\hbar} \sum_{k \neq 0} & \left( \langle \psi_0^{(0)} | \hat{\mu}_\alpha | \psi_k^{(0)} \rangle \langle \psi_k^{(0)} | \hat{\mu}_\beta | \psi_0^{(0)} \rangle \left( \frac{\exp\{i\omega t\}}{\omega_{k0} + \omega} + \frac{\exp\{-i\omega t\}}{\omega_{k0} - \omega} \right) \right. \\ & \left. + \langle \psi_0^{(0)} | \hat{\mu}_\beta | \psi_k^{(0)} \rangle \langle \psi_k^{(0)} | \hat{\mu}_\alpha | \psi_0^{(0)} \rangle \left( \frac{\exp\{-i\omega t\}}{\omega_{k0} + \omega} + \frac{\exp\{i\omega t\}}{\omega_{k0} - \omega} \right) \right) \end{aligned} \quad (49)$$

With the definition of the frequency dependent polarizability [36]

$$\mu_\alpha^{(1)} = \alpha_{\alpha\beta}(\omega) E_\beta (\exp\{i\omega t\} + \exp\{-i\omega t\}) \quad (50)$$

The polarizability term can be identified from eq. (49) by rearranging terms as [36]

$$\alpha_{\alpha\beta}(\omega) = \frac{1}{\hbar} \sum_{k \neq 0} \left( \frac{\langle \psi_0^{(0)} | \hat{\mu}_\alpha | \psi_k^{(0)} \rangle \langle \psi_k^{(0)} | \hat{\mu}_\beta | \psi_0^{(0)} \rangle}{\omega_{k0} - \omega} + \frac{\langle \psi_0^{(0)} | \hat{\mu}_\beta | \psi_k^{(0)} \rangle \langle \psi_k^{(0)} | \hat{\mu}_\alpha | \psi_0^{(0)} \rangle}{\omega_{k0} + \omega} \right) \quad (51)$$

Which is the frequency dependent polarizability. It is important to see that when the frequency  $\omega$  approaches  $\omega_{k0}$  the polarizability will diverge. This corresponds with excitation energies [52–54]. This is one of the applications of the frequency dependent polarizabilities to describe other molecular properties.

## 2.4 Force-field models for polarizability

The polarizability is a property that describes the change in charge distribution for a molecule, as a response to the environment [30]. The polarizability has important applications, and explicit treatment of it can give important insights not possible without it [55].

### 2.4.1 Electronegativity equilization model

The electronegativity equilization model (EEM) is a model that is based on the idea that when atoms form molecules, the electronegativity of the constituent atoms will equalize and the charge is redistributed in the molecule to achieve this [56]. By expanding the electrostatic energy to the second order with respect to atomic charge one obtains [57–59]

$$E(q_1, q_2, \dots, q_n) = \sum_{i=1}^n \left( E_i + \chi_i q_i + \frac{1}{2} \eta_i q_i^2 + \frac{1}{2} \sum_{j=1}^N \frac{q_i q_j}{R_{ij}} \right) \quad (52)$$

Where  $E_i$  is the charge independent energy,  $q_i$  is the atomic charge,  $\chi_i$  is the atomic electronegativity, used for calculating atomic charges,  $\eta_i$  is the atomic chemical hardness,  $R_{ij}$  and the final term may be modified by a screening factor [58, 59]. Taking the derivative of this eq. (52), with respect to the atomic charge  $q_i$ . A set of linear equations is then solved to equalize the electronegativity, with a constraint that the charge is conserved [56, 60]. This allows the model to calculate the minimum of the energy with atomic charges [58]. The EEM unfortunately does not scale correctly at long range [57].

One modification of the EEM model is the atom-atom-charge-transfer model (AACT) [58]. In this model the charges in the molecule is redistributed as a number of partial charges, describing charge transported between two atoms. [58]. By introducing a topology matrix, describing what atoms are bonded, the total charge is automatically accounted for. Describing the energy and minimizing this with respect to the chemical potential, the ideal charge distribution can be found [58].

By applying an electric field a linear response equation may be formulated [58]. Calculating the the atomic charges in the presence of this field the induced dipole moment can be calculated. Then from this the polarizability components can be found [58]. The AACT model has been shown to scale with the size of the system for both alkanes and alkenes [61]

### 2.4.2 Point dipole model

In the inducible point dipole model, a point dipole is associated with polarizable sites often at atomic centres, but implementations with dipoles on the bonds exists [30]. This polarizable site can then be expressed as [62–65].

$$\mu_i = \alpha_i \left[ E_i - \sum_{\substack{J=1 \\ j \neq i}}^N T_{ij}^{(2)} \mu_j \right] \quad (53)$$

Where  $E_i$  is the electric field at  $i$ ,  $\mu$  is a dipole, and  $T_{ij}^{(2)}$  is the dipole dipole interaction tensor eq. (80). Solving eq. (53) for the applied field yields [64, 66]:

$$E_i = \alpha_i^{-1} \mu_i + \sum_{\substack{J=1 \\ j \neq i}}^N T_{ij}^{(2)} \mu_j \quad (54)$$

Or in matrix form, the sum terms can be written as

$$\begin{pmatrix} \alpha_1^{-1} & T_{12}^{(2)} & \cdots & T_{1N}^{(2)} \\ T_{21}^{(2)} & \alpha_2^{-1} & \cdots & T_{2N}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ T_{N1}^{(2)} & T_{N2}^{(2)} & \cdots & \alpha_N^{-1} \end{pmatrix} \quad (55)$$

Defining a matrix  $B$  to be the inverse of the  $3N \times 3N$  matrix and assuming the applied field is uniform, the linear equations can be written [64, 67]

$$\mu_i = \sum_{j=1}^N B_{ij} E \quad (56)$$

So the molecular dipole moment is

$$\mu^{mol} = \sum_{i=1}^N \mu_i = \sum_{i=1}^N \sum_{j=1}^N B_{ij} E \quad (57)$$

and the molecular polarizability is recognized as [64, 65, 68]

$$\alpha^{mol} = \sum_{i=1}^N \sum_{j=1}^N B_{ij} \quad (58)$$

The induced dipoles depend on each other. To find values for the induced dipoles, these equations may be solved iteratively. Solving eq. (56) for a guess set of atomic dipoles, the dipoles are updated. This allows for a solution to eq. (55). Solving eq. (56) for the set of atomic dipoles, a new one is obtained. Allowing for an updated solution again. This may be done iteratively until the fields are self consistent. This is also known as the extended Lagrangian method [69]. The solving of this self consistent field, is usually the computational bottleneck in these types of calculations [70, 71].

### 2.4.3 Drude oscillator

The method of Drude oscillators uses two point charges to represent polarizable atoms [30, 72]. Each non-hydrogen atom can be described by the charges  $q$ , which is the core charge, and  $q_D$  which is a Drude particle, that are bound together by a harmonic spring [73, 74].

In an external field the drude particle will oscillate around a displaced position  $r + d$  where the displacement  $d$  is

$$d = \frac{q_D}{k_D} E \quad (59)$$

where  $E$  is the external electric field, and  $k_D$  is the force constant of the harmonic spring [70]. Then the induced dipole for the atomic charge is

$$\mu = q_D d = \frac{q_D^2}{k_D} E \quad (60)$$

So the isotropic atomic polarizability  $\alpha$  is [70, 75]

$$\alpha = \frac{q_D^2}{k_D} \quad (61)$$

The advantage to this model is the computational efficiency while still reproducing molecular polarizability, even without treating the hydrogens.

## 2.5 CT-PDI model

In the CT-PDI model, a combination of a charge-transfer and electronegativity equalization model is combined with a point-dipole interaction model [27].

In the model each atomic charge may be rewritten as a time dependent charge, where the charge is formulated in terms of charge transfer variables [24]. For an atom  $I$ , the atomic charge is [27]

$$q_I = \sum_K^N q_{IK}(t) \quad (62)$$

The Lagrangian for a molecular system in the model may be written as [26]

$$L = K^q + K^\mu - V \quad (63)$$

Where  $K^q$  is the charge kinetic energy,  $K^\mu$  is the atomic dipole moment kinetic energy, and  $V$  is the potential energy. Expanding the terms gives [24, 26]:

$$K^q = \frac{1}{2} \sum_{I,K>I}^N (c_I^{q*} + c_K^{q*}) R_{IK}^2 \dot{q}_{IK}^2 \quad (64)$$

Where  $c_I^{q*}$  is an atom type parameter. It is adopted that all subsequent atom-type parameters, being parameters specific to the atom at which they are located is denoted by a superscript \*. The parameters can be seen as the inverse number of oscillating charges [25, 28].  $R_{IK}$  is the internuclear distance between atoms  $I$  and  $K$ , and  $\dot{q}_{IK}$  is the time derivative of the charge passing as a current from atom  $I$ , to atom  $K$ .

$$K^\mu = \frac{1}{2} \sum_I^N c_I^{\mu*2} (\dot{\mu}_I)^2 \quad (65)$$

Here  $c_I^{\mu*}$  is an atom type parameter representing an inverse number of oscillating dipoles [25, 28], and  $\dot{\mu}_I$  is the time derivative of the atomic dipole moment for atom  $I$ .

$$V = V^{qq} + V^{q\mu} + V^{\mu\mu} \quad (66)$$

Where  $V^{qq}$  is the charge-charge interaction energy,  $V^{q\mu}$  is the charge dipole interaction energy, and  $V^{\mu\mu}$  is the dipole-dipole interaction energy.

### 2.5.1 Representation of charges

In the model charges are represented as Gaussian charge distributions on the form [24, 26, 27, 76].

$$\rho_I(r_i) = q_I \left( \frac{\Phi_I^*}{\pi} \right)^{\frac{3}{2}} \exp\{-\Phi_I^* r_i^2\} \quad (67)$$

Where  $\Phi_I^*$  is an atom type parameter and  $r_i$  represents the position in the distribution. The interaction between two Gaussian charge distributions with exponents  $\Phi_I^*$  and  $\Phi_J^*$  is [76]

$$V = \iint \frac{\rho_I(r_1)\rho_J(r_2)}{r_{12}} dr_1 dr_2 = q_I q_J \frac{\text{erf}(\sqrt{a}R_{IJ})}{R_{IJ}} \quad (68)$$

Where erf is the regular error function and  $a$  is [24, 26, 76]

$$a = \frac{\Phi_I^* \Phi_J^*}{\Phi_I^* + \Phi_J^*} \quad (69)$$

This result of eq. (68) may be used to introduce a modified distance  $\tilde{R}_{IJ}$  [24] such that

$$\tilde{R} = \frac{R_{IJ}}{\text{erf}(\sqrt{a}R_{IJ})} \rightarrow \tilde{V} = \frac{q_I q_J}{\tilde{R}} \quad (70)$$

An approximation of the modified distance eq. (70) is [24, 26, 76]

$$\tilde{R} = \sqrt{R_{IJ}^2 + \frac{\pi}{4a}} \quad (71)$$

Taking the limits of this approximation eq. (71) as  $R_{IJ}$  approached infinity causes  $\tilde{V}$  to approach 0. The same behaviour is found in the original result in eq. (68) where the error function is bound to be less than or equal one. If  $R_{IJ}$  approaches zero, both eq. (71) and eq. (68) will approach  $\sqrt{\frac{4a}{\pi}}$  as can be shown for eq. (68) using the definition of the error function and the fundamental theorem of calculus [77].

### 2.5.2 Charge-Charge interaction energy

The charge-charge interaction may be expressed as [24]

$$V^{qq} = \sum_I^N (\chi_I^* + \phi_I^{ext}) q_I + \frac{1}{2} \eta_I^* q_I^2 + \frac{1}{2} \sum_{I \neq J}^N q_I T_{IJ}^{(0)} q_J \quad (72)$$

$\chi_I^*$  is the atomic electronegativity and  $\eta_I^*$  is the atomic chemical hardness. Both are atom type parameters.  $\phi_I^{ext}$  is an external electrostatic potential at atom  $I$ .  $T_{IJ}^{(0)}$  is the inverse internuclear distance between atom  $I$  and  $J$ .  $T_{IJ}^{(0)} = \frac{1}{R_{IJ}}$ . Thus the first term can be interpreted as the energy of the charge caused by the attraction of the nuclei in the molecule and the applied external field. The chemical hardness term can be interpreted as the resistance to change in chemical potential from a change in number of electrons [78–80], this is used to calculate polarizabilities, and the final term can be seen as the Coulomb interaction. The electronegativity is included to calculate atomic charges, as it gives different elements an inherent potential difference [27].

Rewritten in terms of charge-transfer variables the charge-charge interaction energy becomes [24]:

$$V^{qq} = \sum_{I,K>I}^N (\chi_{IK} + \phi_{IK}^{ext})q_{IK} + \frac{1}{2} \sum_{\substack{I,K>I \\ J,M>J}}^N q_{IK}T_{IJ, JM}^{(0)}q_{JM} \quad (73)$$

Wherein it has been used that  $\chi_{IK} = \chi_I^* - \chi_K^*$ ,  $\phi_{IK}^{ext} = \phi_I^{ext} - \phi_K^{ext}$ , and  $T_{II}^{(0)} = \eta_I^*$ .  $T_{IK, JM}^{(0)}$  is defined as [24]:

$$T_{IK, JM}^{(0)} = T_{IJ}^{(0)} - T_{KJ}^{(0)} - T_{IM}^{(0)} + T_{KM}^{(0)} \quad (74)$$

For the case in eq. (74) where  $J = M$ , this becomes [24]

$$\begin{aligned} T_{IK, JM}^{(0)} &= T_{II}^{(0)} - T_{KI}^{(0)} - T_{IM}^{(0)} + T_{KM}^{(0)} \\ &= \eta_I^* - T_{KI}^{(0)} - T_{IM}^{(0)} + T_{KM}^{(0)} \end{aligned} \quad (75)$$

Where  $\eta_I^*$  is the atomic chemical hardness. If  $K = M$  and  $I = J$  [24]

$$\begin{aligned} T_{IK, JM}^{(0)} &= T_{II}^{(0)} - T_{KI}^{(0)} - T_{IK}^{(0)} + T_{KK}^{(0)} \\ &= \eta_I^* + \eta_K^* \end{aligned} \quad (76)$$

### 2.5.3 Charge-dipole interaction energy

In terms of charge transfer variables, the dipole-charge interaction from the potential energy eq. (66) may be written as [24]

$$V^{\mu q} = \sum_{I, J}^N q_I \tilde{T}_{IJ, \alpha}^{(1)} \mu_{J, \alpha} = \sum_{I, K>I, J}^N q_{IK} \tilde{T}_{IK, J, \alpha}^{(1)} \mu_{J, \alpha} \quad (77)$$

Where eq. (62) has been used. The Einstein summation convention for repeated subscripts has been used and  $\alpha$  and  $\beta$  represent the Cartesian coordinates  $x, y$  or  $z$ .  $\tilde{T}_{IJ, \alpha}^{(1)}$  is

the modified charge dipole interaction tensor, and  $\tilde{T}_{IK,J,\alpha}^{(1)} = \tilde{T}_{IJ,\alpha}^{(1)} - \tilde{T}_{KJ,\alpha}^{(1)}$ . The modified charge dipole interaction tensor is defined as [24]

$$\tilde{T}_{IJ,\alpha}^{(1)} = \frac{\partial \tilde{T}_{IJ}^{(0)}}{\partial \tilde{R}_{IJ,\alpha}} = -\frac{R_{IJ,\alpha}}{\tilde{R}_{IJ}^3} \quad (78)$$

Where  $\tilde{T}_{IJ}^{(0)} = \tilde{R}_{IJ}^{-1}$

#### 2.5.4 Dipole-dipole interaction energy

The final term of eq. (66) is  $V^{\mu\mu}$  [24, 26]

$$V^{\mu\mu} = \frac{1}{2} \sum_I^N \mu_{I,\alpha} (\alpha_{I,\alpha\beta})^{-1} \mu_{I,\beta} - \frac{1}{2} \sum_I^N \sum_{K \neq I}^N \mu_{I,\alpha} \tilde{T}_{IK,\alpha\beta}^{(2)} \mu_{K,\beta} - \sum_I^N E_{I,\alpha}^{ext} \mu_{I,\alpha} \quad (79)$$

Where  $\mu_{I,\alpha}$  is the dipole moment at I, and subscript  $\alpha$  is one of the Cartesian coordinates.  $\tilde{T}_{IK,\alpha\beta}^{(2)}$  is defined as [24]

$$\tilde{T}_{IK,\alpha\beta}^{(2)} = \frac{\partial \tilde{T}_{IK,\alpha}^{(1)}}{\partial R_{IK,\beta}} = \frac{3R_{IK,\alpha} R_{IK,\beta}}{\tilde{R}_{IK}^5} - \frac{\delta_{\alpha\beta}}{\tilde{R}_{IK}^3} \quad (80)$$

In eq. (79),  $\alpha_{I,\alpha\beta}$  is the atomic polarizability with a modification for the chemical surroundings [24, 26, 67], discussed below.

The atomic polarizability may be included in the sum by defining

$$\tilde{T}_{II,\alpha\beta}^{(2)} = \mu_{I,\alpha} (\alpha_{I,\alpha\beta})^{-1} \mu_{I,\beta} \quad (81)$$

#### 2.5.5 Modification of the atomic polarizability

For the anisotropy of the chemical environment one can define a symmetric matrix  $\Gamma$  such that [67]

$$\Gamma_{I,\alpha\beta} = \sum_{I \neq J}^N \bar{\alpha}_J (g_{IJ})^\kappa \frac{R_{IJ,\alpha} R_{IJ,\beta}}{R_{IJ}} \quad (82)$$

Where  $\bar{\alpha}_J$  is the average polarizability of atom  $J$ ,  $g_{IJ}$  is a electronic repulsion integral defined by the Ohno formula [67], and  $\kappa$  is an adjustable parameter; the repulsion exponent. Newer models have rewritten this in terms of an atom type parameter, and a



intermolecular charge transfer resistance  $S_{IJ}$  [26]

$$\Gamma_{I,\alpha\beta} = \sum_{I \neq J}^N \alpha_J^* S_{IJ} \frac{R_{IJ,\alpha} R_{IJ,\beta}}{R_{IJ}} \quad (83)$$

Where  $S_{IJ}$  is defined [24,27]

$$S_{IJ} = \exp\{(-a_{IJ}(R_{IJ} - R_I^* - R_J^*))\} \quad (84)$$

This formulation is constructed such that  $S_{IJ} \approx 1$  when the internuclear distance is within a regular bond length, such that there is little contribution to the polarizability from atoms far away.  $R_I^*$  is an atom type parameter.

Using eq. (83), the normalized atomic anisotropy tensor can be defined as [24,27,67]

$$G_{I,\beta\alpha} = \frac{3}{\text{Tr}(\Gamma_I)} \Gamma_{I,\beta\alpha} \quad (85)$$

Which is normalized such that the trace is three [67]. This then offers an isotropic and anisotropic modification to the atomic polarizability as

$$\alpha_{I,\alpha\beta} = \alpha_I^* (\delta_{\alpha\beta} + x_I^* (1 - G_{I,\beta\alpha})) \quad (86)$$

Where  $\delta_{\alpha\beta}$  is the Kronecker-delta,  $\alpha_I^*$  and  $x_I^*$  are atom type parameters. This term also gives correct properties with respect to rotations [24,27] It is imposed that the anisotropic polarizability parameter must be less than 0.5 [26], this is done such that if eq. (83), only has one contribution, eq. (85) becomes  $-3$ , and eq. (86) might become negative [26]. This is not a physical result, and it is therefore avoided.

### 2.5.6 Modification of the chemical hardness

To achieve the correct scaling of the polarizability for the systems, one can introduce a modification to the chemical hardness [24,26,27]. This becomes

$$\eta_I^* \rightarrow \eta_I^* S_{IK}^{-\frac{1}{2}} S_{IM}^{-\frac{1}{2}} g_{I,KM} \quad (87)$$

$S_{IK}$  is from eq. (84), and  $g_{I,KM}$  is the resistance to charge flow in the molecule. This resistance is defined to be [27]:

$$g_{I,KM} = \exp\{(C_{I,KM}(\Delta_{I,KM})^2)\} \quad (88)$$

With  $C_{I,KM}$  being  $(C_I^*)^2 C_K^* C_M^*$  where all  $C^*$  are atom type parameters and

$$\Delta_{I,KM} = R_{IK} - (R_{*I} + R_{*K}) + R_{IM} - (R_{*I} + R_{*M}) \quad (89)$$

With  $R^*$  being atom type parameters like in eq. (84)

It is possible to define different equations for this resistance [24, 26].

$$\begin{aligned} g_{I,KM} = & (g_{0,I}^*)^2 g_{0,K}^* g_{0,M}^* H_{I,KM}(\Delta_{I,KM}) \\ & + (g_{1,I}^*)^2 g_{1,K}^* g_{1,M}^* (1 - H_{I,KM}(\Delta_{I,KM})) \end{aligned} \quad (90)$$

$g_{0,I}^*$  and  $g_{1,I}^*$  are atom type parameters.  $H_{I,KM}$  is a smooth step function

$$H_{I,KM}(\Delta_{I,KM}) = \frac{1}{2}(1 + \tanh(C_{I,KM}\Delta_{I,KM})) \quad (91)$$

Previous work on the model found eq. (88) to give too sharp change with bond distance for certain molecules and therefore used eq. (90) [26], however it was eq. (88) which was used in this model.

## 2.6 Calculation of the molecular polarizability in the CT-PDI model

Assuming a molecule is in an external electric field and electrostatic potential oscillating with a frequency  $\omega$ , such that [24, 26]

$$E_{J,\alpha}^{ext} = \text{Re}(E_{J,\alpha}^\omega \exp(i\omega t)) \quad (92)$$

$$\phi_{IJ}^{ext} = \text{Re}(E_{IJ}^\omega \exp(i\omega t)) \quad (93)$$

This will cause an oscillation of the charge transfer and atomic dipole moments away from the static conditions [24, 26]

$$q_{IK} = \text{Re}(q_{IK}^{(0)} + q_{IK}^{(\omega)} \exp(i\omega t)) \quad (94)$$

$$\mu_{I,\alpha} = \text{Re}(\mu_{I,\alpha}^{(0)} + \mu_{I,\alpha}^{(\omega)} \exp(i\omega t)) \quad (95)$$

According to Hamilton's principle [81], the evolution of the system

$$J = \int \left\{ L(q(t), \dot{q}(t), \mu(t), \dot{\mu}(t)) \right\} - \lambda \left\{ \sum_I^N q_I - Q_{tot} \right\} \quad (96)$$

is minimized such that  $\delta J = 0$  [28]. Here  $L$  is the Lagrangian from eq. (63),  $q_I$  is the atomic charge and  $Q_{tot}$  is total charge of the molecule. For the condition that the molecule is neutral the final term in eq. (96) with the Laplace multiplier term  $\lambda$  may be neglected in favour of a topology matrix, controlling permitted charge transfers to be restricted to bonded nuclei [27]. Enforcing the condition the equations must be solved from eq. (96), results in the system being a stationary value, and the Euler-Lagrange equations apply [81]. For the system these equations become [26],

$$\frac{\partial}{\partial t} \left( \frac{\delta L}{\delta \dot{q}_{IK}} \right) - \frac{\delta L}{\delta q_{IK}} = 0 \quad (97)$$

$$\frac{\partial}{\partial t} \left( \frac{\delta L}{\delta \dot{\mu}_{I,\alpha}} \right) - \frac{\delta L}{\delta \mu_{I,\alpha}} = 0 \quad (98)$$

### 2.6.1 Evaluation of Euler-Lagrange terms

For the first term in eq. (97), the only term depending on  $\dot{q}_{IK}$  is the charge kinetic energy (eq. (64)) [26], thus

$$\begin{aligned} \frac{\partial}{\partial t} \left( \frac{\delta L}{\delta \dot{q}_{IK}} \right) &= \frac{\partial}{\partial t} \left( \frac{\delta K^q}{\delta \dot{q}_{IK}} \right) \\ &= (c_I^q + c_K^q) R_{IK}^2 \ddot{q}_{IK} \\ &= (c_I^q + c_K^q) R_{IK}^2 \omega^2 \operatorname{Re}(q_{IK}^{(\omega)} \exp(i\omega t)) \end{aligned} \quad (99)$$

Where eq. (94) has been used in the final step. Further, in the first term of eq. (98), only the dipole kinetic energy eq. (65) depends on the time derivative of the dipole moment  $\dot{\mu}_{I,\alpha}$  [26,27]

$$\begin{aligned} \frac{\partial}{\partial t} \left( \frac{\delta L}{\delta \dot{\mu}_{I,\alpha}} \right) &= \frac{\partial}{\partial t} \left( \frac{\delta K^\mu}{\delta \dot{\mu}_{I,\alpha}} \right) \\ &= c_I^\mu \ddot{\mu}_I \\ &= c_I^\mu \omega^2 \operatorname{Re}(\mu_{I,\alpha}^{(\omega)} \exp(i\omega t)) \end{aligned} \quad (100)$$

Where eq. (95) is used in the final step.

The final term in eq. (97) becomes [26, 27]

$$\begin{aligned}
\frac{\delta L}{\delta q_{IK}} &= -\frac{\delta V}{\delta q_{IK}} = -\frac{\delta(V^{qq} + V^{q\mu})}{\delta q_{IK}} \\
&= -\sum_{J < M}^N T_{IK, JM}^{(0)} q_{JM} \\
&\quad - (\chi_{IK} + \phi_{IK}^{ext}) \\
&\quad - \sum_J^N T_{IK, J, \alpha}^{(1)} \mu_{J, \alpha}
\end{aligned} \tag{101}$$

Where eqs. (77) and (79) have been used. The last term from eq. (98) is [26, 27]

$$\begin{aligned}
\frac{\delta L}{\delta \mu_{I, \alpha}} &= -\frac{\delta V}{\delta \mu_{I, \alpha}} = -\frac{\delta(V^{q\mu} + V^{\mu\mu})}{\delta \mu_{I, \alpha}} \\
&= -\sum_{J < M}^N T_{I, JM, \alpha}^{(1)} q_{JM} \\
&\quad + \sum_J^N T_{I, J, \alpha\beta}^{(2)} \mu_{J, \beta} + E_{I, \alpha}^{ext}
\end{aligned} \tag{102}$$

Putting together the equations eqs. (99) to (102), into the conditions from eqs. (96) to (98) can be expressed in matrix form as [24, 26]

$$\begin{aligned}
&\left( \begin{pmatrix} T_{IK, JM}^{(0)} & T_{IK, J, \alpha}^{(1)} \\ T_{I, JM, \alpha}^{(1)} & T_{I, J, \alpha\beta}^{(2)} \end{pmatrix} - \omega^2 \begin{pmatrix} C^q & 0 \\ 0 & C^\mu \end{pmatrix} \right) \\
&\quad \times \begin{pmatrix} q_J^{(\omega)} \\ \mu_{J, \beta}^{(\omega)} \end{pmatrix} = \begin{pmatrix} -\chi_{IK} - \phi_{IK}^{(\omega)} \\ E_{I, \alpha}^{(\omega)} \end{pmatrix}
\end{aligned} \tag{103}$$

Where the diagonal matrix elements  $C^q$  and  $C^\mu$  are the frequency dependent parts of the equations eq. (99) and eq. (100) respectively [24]. Dissipation may be added to the model by replacing the squared frequency  $\omega^2$  in eq. (99) and eq. (100) with  $\omega^2 - \omega \frac{i}{2} (\gamma_I^{q*} + \gamma_K^{q*})$  and  $\omega^2 - i\gamma_I^{\mu*} \omega$  respectively [26]. Thus the elements  $C^q$  and  $C^\mu$  with included dissipation become [24, 26]

$$C^q = (c_I^q + c_K^q) R_{IK}^2 \left(1 - \frac{i}{2\omega} (\gamma_I^{q*} + \gamma_K^{q*})\right) \delta_{IK} \delta_{JM} \tag{104}$$

$$C^\mu = c_I^\mu \left(1 - \frac{i}{2\omega} \gamma_I^\mu\right) \delta_{IK} \delta_{\alpha\beta} \tag{105}$$

Where  $\gamma_I^{q*}$  is an atom type parameter. If the polarizability  $\alpha$  is the linear response to an electric field [82,83] and assuming the molecule is in an electric field that can be treated as homogeneous, the polarizability may be calculated as the gradient of the induced dipole moment with the external electric field [24,26]. This may be found by taking the derivative of eq. (103) with respect to an external electric field. This gives:

$$\begin{aligned} & \left( \begin{pmatrix} T_{IK,JM}^{(0)} & T_{IK,J,\alpha}^{(1)} \\ T_{I,JM,\alpha}^{(1)} & T_{I,J,\alpha\beta}^{(2)} \end{pmatrix} - \omega^2 \begin{pmatrix} C^q & 0 \\ 0 & C^\mu \end{pmatrix} \right) \\ & \times \begin{pmatrix} \frac{\partial q_J^{(\omega)}}{\partial E_\gamma^{(\omega)}} \\ \frac{\partial \mu_{K,\beta}^{(\omega)}}{\partial E_\gamma^{(\omega)}} \end{pmatrix} = \begin{pmatrix} -R_{SP} \\ \delta_{\alpha\gamma} \end{pmatrix} \end{aligned} \quad (106)$$

Where it has been used that the electrostatic potential for an homogeneous external field is  $\phi_I^{ext} = R_{I,\alpha} E_\alpha^{ext}$  [26]. We thus arrive at a solvable equation for the molecular polarizability [24,26]

$$\alpha_{\alpha\beta}^{mol}(\omega) = \frac{\partial \mu_\alpha^{ind}}{\partial E_\beta^{(\omega)}} = \sum_{I,M>I}^N R_{IM,\alpha} \frac{\partial q_J^{(\omega)}}{\partial E_\beta^{(\omega)}} + \sum_I^N \frac{\partial \mu_\alpha^{ind}}{\partial E_\beta^{(\omega)}} \quad (107)$$

This must then be solved three times for  $\alpha, \beta = x, y, z$  [24]. In this way the molecular polarizability tensor elements may be estimated by the model.

## 2.7 Calculation of the dipole gradient

Hunt demonstrated a method for the calculation of the dipole gradient directly using only electrostatics [84,85]. Using the Hellmann-Feynmann theorem the force on a nucleus is [84,86,87]

$$f_I = Q_I E_I \quad (108)$$

Where the force on nucleus  $I$  is  $f_I$ ,  $Q_I$  is the charge of nucleus  $I$  and  $E_I$  is the electric field, caused by all the other particles in the molecule. For no applied external field. This is equivalent to the classical definition of the force from an electric field [82,87]. Following Hunts method [84], the field at point  $r$  due to the charge of nucleus  $I$ ,  $Q_I$ , is [82]

$$E_I(r) = \frac{Q_I}{|r - R_I|^2} \quad (109)$$

Where  $R_I$  is the position of nucleus  $I$  and  $r$  is a given position. The gradient of the electric field with respect to the cartesian coordinate  $\alpha$  is [84]

$$E_{I,\alpha}(r) = \frac{Q_I(r - R_I)_\alpha}{|r - R_I|^3} \quad (110)$$

If the nucleus  $I$  is shifted from  $R_I$  to  $R_I + \delta R_I$ , eq. (110) becomes [84]

$$\begin{aligned} E_{I,\alpha}(r) &= E_{I^{(0)},\alpha}(r) + E_{I,\alpha}(r) + \dots \\ &= \frac{Q_I(r - R_I)_\alpha}{|r - R_I|^3} + Q_I T_{\alpha\beta}(r, R_I) \delta R_{I,\beta} + \dots \end{aligned} \quad (111)$$

Where  $I^{(0)}$  indicates the original, unshifted position of  $R_I$ .  $T_{\alpha\beta}(r, R_I)$  is the dipole propagator [84]

$$\begin{aligned} T_{\alpha\beta}(r, R_I) &= \nabla_\alpha \nabla_\beta (|r - R_I|^{-1}) \\ &= \frac{3(r_\alpha - R_{I,\alpha})(r_\beta - R_{I,\beta}) - \delta_{\alpha\beta} |r - R_I|^2}{|r - R_I|^5} - \frac{4\pi}{3} \delta_{\alpha\beta} \delta(r - R_I) \end{aligned} \quad (112)$$

This is equivalent to the unmodified expression of the dipole-dipole interaction tensor eq. (80) with an additional term.

### 2.7.1 Electric field shielding tensor

With a molecule in a static uniform electric field  $F^e$ , the net field on nucleus  $I$  can be expanded as the series [84]

$$F^I = F^{I(0)} + (1 - \gamma^I) F^e + \dots \quad (113)$$

The energy of the field on the molecule can be written

$$E(F^e) = E^{(0)} + \mu F^e + \dots \quad (114)$$

The force on nucleus  $I$  can be expressed as the gradient of the energy  $E(F^e)$  with respect to  $R_I$ . Then [84]

$$\frac{\partial E(F^e)}{\partial R_I} = -\frac{\partial E^{(0)}}{\partial R_I} + \frac{\partial(\mu F^e)}{\partial R_I} + \dots \quad (115)$$

Equating this with the force found in eq. (108), then equating the terms linear with the electric field yields [84]

$$\frac{\partial \mu_\beta}{\partial R_{I,\alpha}} = Q_I (\delta_{\alpha\beta} - \gamma_{\alpha\beta}^I) \quad (116)$$

Where  $\delta_{\alpha\beta}$  is the Kronecker delta and  $\gamma_{\alpha\beta}^I$  is the electric field shielding tensor at nucleus  $I$ . An explicit expression for the tensor  $\gamma_{\alpha\beta}^I$  must be found to be able to calculate the dipole gradients.

### 2.7.2 Polarization

Shifting the position of nucleus  $I$  by  $\delta R_I$  will effect the polarization of the molecule. This effect is [84]

$$\delta P_\alpha(r) = \int \alpha_{\alpha\beta}(r, r') \delta f_{I,\beta}(r') dr' \quad (117)$$

Where polarization is the dipole moment per volume [82]. To evaluate this term the relation between the polarization  $P(r)$  and charge density  $\rho(r)$  is [84]:

$$\nabla P(r) = -\rho(r) \quad (118)$$

Further the ground state polarizability density  $\alpha_{\alpha\beta}(r, r')$  is [84]

$$\alpha_{\alpha\beta}(r, r') = \xi_{\alpha\beta} \sum_k' \frac{\langle 0 | \hat{P}_\alpha(r) | k \rangle \langle k | \hat{P}_\beta(r') | 0 \rangle}{(E_k - E_0)} \quad (119)$$

Where  $\xi_{\alpha\beta}$  imposes symmetry with respect to the polarizability operators  $\hat{P}_\alpha$  and  $\hat{P}_\beta$ .

Inserting eq. (111) and eq. (119) into eq. (117) yields

$$\delta P_\alpha(r) = \int \left( \xi_{\alpha\beta} \sum_k' \frac{\langle 0 | \hat{P}_\alpha(r) | k \rangle \langle k | \hat{P}_\beta(r') | 0 \rangle}{(E_k - E_0)} \right) \left( Q_I T_{\alpha\beta}(r, R_I) \delta R_{I,\beta} \right) dr' \quad (120)$$

Distributing  $\xi_{\alpha\beta}$  and expanding the dipole propagator eq. (112) the expression becomes

$$\delta P_\alpha(r) = \int \left( \sum_k' \frac{\langle 0 | \hat{P}_\alpha(r) | k \rangle \langle k | \hat{P}_\beta(r') | 0 \rangle + \langle 0 | \hat{P}_\beta(r') | k \rangle \langle k | \hat{P}_\alpha(r) | 0 \rangle}{(E_k - E_0)} \right) \left( Q_I \nabla_\alpha \nabla_\beta (|r - R_I|^{-1}) \delta R_{I,\beta} \right) dr' \quad (121)$$

Using then the relation in eq. (118)

$$\delta\rho(r) = \int \left( \sum_k \frac{\langle 0 | \hat{\rho}(r) | k \rangle \langle k | \hat{P}_\beta(r') | 0 \rangle + \langle 0 | \hat{P}_\beta(r') | k \rangle \langle k | \hat{\rho}(r) | 0 \rangle}{(E_k - E_0)} \right) \left( Q_I \nabla_\alpha \nabla_\beta (|r - R_I|^{-1}) \delta R_{I,\beta} \right) dr' \quad (122)$$

Integrating this expression with respect to  $r'$  by parts and using that  $\nabla_\alpha (|r' - R_I|^{-1}) = -\nabla_\alpha^I |r' - R_I|^{-1}$  gives

$$\delta\rho(r) = \int Q_I \nabla_\alpha (|r - R_I|^{-1}) \delta R_{I,\beta} \sum_k \frac{\langle 0 | \hat{\rho}(r) | k \rangle \langle k | \hat{\rho}(r') | 0 \rangle + \langle 0 | \hat{\rho}(r') | k \rangle \langle k | \hat{\rho}(r) | 0 \rangle}{(E_k - E_0)} dr' \quad (123)$$

The molecular dipole moment  $\mu^{mol}$  gradient with respect to the shift in position of nucleus  $I$  has two contributions [84]

$$\frac{\partial \mu^{mol}}{\partial R_{I,\alpha}} = \frac{\partial \mu_\beta^n}{\partial R_{I,\alpha}} + \frac{\partial \mu_\beta^e}{\partial R_{I,\alpha}} \quad (124)$$

The dipole moment is the integral of the polarization density over all space [84]. The nuclear component becomes

$$\frac{\partial \mu_\beta^n}{\partial R_{I,\alpha}} = Q_I \delta_{\alpha\beta} \quad (125)$$

### 2.7.3 Perturbation theory polarization gradient

The shift of the position of nucleus  $I$  by  $\delta R_I$  can be expressed as a perturbation of the Hamiltonian [84]

$$\begin{aligned} H &= H_0 + H_1 \\ &= H_0(R_I) + \frac{\partial H_0}{\partial R_{I,\alpha}} \delta R_{I,\alpha} + \dots \end{aligned} \quad (126)$$

Perturbation theory has also been used to show that the change in electronic charge density from the shift of  $\delta R_I$  is to the lowest order [84]

$$\delta\rho(r) = \sum_k \frac{[\langle 0 | \hat{\rho}(r) | k \rangle \langle 0 | H_1 | k \rangle + \langle 0 | H_1 | k \rangle \langle 0 | \hat{\rho}(r) | k \rangle]}{(E_0 - E_k)} \quad (127)$$



Inserting the expression from eq. (126) into eq. (128) becomes

$$\delta\rho(r) = \sum_k' \frac{[\langle 0 | \frac{\partial H_0}{\partial R_{I,\alpha}} \delta R_{I,\alpha} | k \rangle \langle 0 | H_1 | k \rangle + \langle 0 | H_1 | k \rangle \langle 0 | \frac{\partial H_0}{\partial R_{I,\alpha}} \delta R_{I,\alpha} | k \rangle]}{(E_0 - E_k)} \quad (128)$$

It can be shown that the first order correction to the hamiltonian in eq. (126) is [84]

$$\frac{\partial H_0}{\partial R_{I,\alpha}} = \int Q_I \nabla_\alpha^I |r - R_I|^{-1} \hat{\rho}(r) dr \quad (129)$$

Inserting eq. (129) into eq. (128) and moving the terms that do not operate on the wavefunction out of the sum gives

$$\delta\rho(r) = \int Q_I \nabla_\alpha (|r - R_I|^{-1}) \delta R_{I,\beta} \sum_k' \frac{\langle 0 | \hat{\rho}(r) | k \rangle \langle k | \hat{\rho}(r') | 0 \rangle + \langle 0 | \hat{\rho}(r') | k \rangle \langle k | \hat{\rho}(r) | 0 \rangle}{(E_k - E_0)} dr' \quad (130)$$

Which is equal to the previously found expression in eq. (123). This strengthens the assumption that eq. (117) is an appropriate expression for the polarization.

As the polarization is the total dipole moment over the volume [82] the dipole can be found by integrating eq. (123) over all space. This gives

$$\frac{\partial \mu_\beta^e}{\partial R_{I,\alpha}} = Q_I \int \alpha_{\alpha\gamma}(r, r') T_{\beta\gamma}(r', R_I) dr dr' \quad (131)$$

#### 2.7.4 Expression of the shielding tensor

For a molecule in an external field  $F^e$  The field at nucleus  $I$  becomes [84]

$$\begin{aligned} F^I &= F^{(0)}(R_I) + F^e(R_I) + \int T(R_I, r) \cdot P^{ind}(r) dr \\ &= F^{(0)}(R_I) + F^e(R_I) + \int T(R_I, r) \cdot \alpha(r, r') \cdot F^e(r') dr + \dots \end{aligned} \quad (132)$$

Then using eq. (113), equating terms gives [84]

$$\gamma_{\alpha\beta}^I = - \int T(R_I, r) \cdot \alpha(r, r') dr \quad (133)$$

This then gives an expression for the dipole gradient

$$\frac{\partial \mu_\beta}{\partial R_{I,\alpha}} = Q_I (\delta_{\alpha\beta} + \int T(R_I, r) \cdot \alpha(r, r') dr) \quad (134)$$

## 2.8 The dipole gradient in the CT-PDI model

The expression for the dipole gradient in the CT-PDI can be found by evaluating the equation eq. (134). In the model the integral will simply become the sum over atomic contributions to the total gradient. The term then becomes

$$\frac{\partial \mu_\beta}{\partial R_{I,\alpha}} = Q_I (\delta_{\alpha\beta} + \int T(R_I, r) \cdot \alpha(r, r') dr) \quad (135)$$

Inserting the equation for the molecular polarizability eq. (107) into the expression for the dipole gradient eq. (134), gives:

$$\frac{\partial \mu_\beta}{\partial R_{I,\alpha}} = Q_I (\delta_{\alpha\beta} + \sum_{I,M>I}^N \left( T(R_I, R_M) (R_{IM,\alpha} \frac{\partial q_I^{(\omega)}}{\partial E_\beta^{(\omega)}} + \frac{\partial \mu_\alpha^{ind}}{\partial E_\beta^{(\omega)}}) \right)) \quad (136)$$

## 3 Method

### 3.1 DFT-Calculations

An initial geometry for the molecules were generated using the AVOGADRO software [88] version 1.2.0, and was optimized using its implementation of the UFF force-field [89], except if there was potential for hydrogen bonding within the molecule, in which case the MMFF94s force field [90] was employed for the initial geometry. The dalton molecular electronic structure program was employed to carry out quantum chemical DFT calculations [91]. For the functional the CAM-B3LYP functional [49] was chosen. The Jensen polarization consistent basis sets apc-1 [92] were used. The same basis sets and functional was used in the calculation of the geometry and property calculation.

Basis set	apc-1
Functional	CAM-B3LYP

Table 3.1: Basis set and functional used in DFT calculations.

#### 3.1.1 Choice of functional

The CAM-B3LYP functional [49] was used in the DFT calculations. This was chosen because the B3LYP [93] functional does not offer sufficiently accurate results for the polarizability of longer chain molecules and excitations [49]. As the polarizability is of

importance to the CT-PDI model it was seen as reasonable to use the functional for the calculations, this has also been done in other experiments [94].

The geometry was also optimized using the CAM-B3LYP functional. As the CAM-B3LYP offers a correction to the B3LYP functional at longer range [49], for smaller molecules the use of the CAM-B3LYP functional should not be greatly different than having used normal B3LYP. For longer conjugated  $\pi$  systems, the B3LYP functional does not give an accurate geometry [94]. As most molecules in the study were in these categories the CAM-B3LYP functional is useful in the parametrization of the model.

Also the dipole moment, quadrupole moment and Dipole gradients were calculated using the CAM-B3LYP functional.

### 3.1.2 Choice of basis sets

The Jensen polarization consistent basis sets [92] were used for the geometry optimization and calculation of molecular properties. Previous works have employed the Dunning correlation consistent basis sets [95, 96].

The polarization consistent sets have the advantage that higher angular momentum functions have geometrically decreasing importance, while the correlation consistent sets have arithmetically decreasing importance. As such they may have faster convergence than correlation consistent sets [92].

Polarization functions were also added to improve description of the molecular properties. The addition of polarization functions improves the descriptions of molecular properties [1] and for dipole and quadrupole moments, as well as for static polarizabilities the addition of diffuse d-functions were important, with higher angular momentum function being less important [97]. The use of the apc-1 basis set has also been used in parametrization of another force field [10]

## 3.2 Modification of geometries

A distribution of the bond lengths in the molecules were visualized. From this and the optimized geometries the final geometry was modified to stretch or compress select bonds such that the intervals with less common bond lengths are better represented. In this way a better parametrization could be achieved for these bond lengths. When modifying molecular bonds, only one bond was modified at a time.

With the modified geometries, the single point properties were recalculated, and a new input file was generated.

### 3.3 Generation of parameter input files

Using the results of the DFT calculations described in section 3.1. Final geometry, dipole moment, quadrupole moment, static- and frequency-dependent polarizabilities, and the dipole gradient of the molecule was extracted. For the frequency-dependent polarizability, if the molecule an excitation energy lower than the frequency of the polarizability, all higher frequencies were discarded as the model will not correctly model excited states.

Discarding the frequency dependent polarizability if it was found to be above the excitation energy was done regardless of whether the excitation was symmetry forbidden. This was because many molecules were to have their modified as discussed in section 3.2. As such when the geometry is changed, the symmetry of the molecule would also be effected, and may cause a change in the oscillator strength, which would be incorrect if used in the parametrization.

### 3.4 Optimization of parameters

After the generation of input files, the files are used with the simplex algorithm [98] for optimizing the atom type parameters discussed in section 2. The simplex algorithm has been used for the optimization of force field parameters in several studies [99–101]. The parameters are optimized with respect to a given molecular property. For the project the parameters were optimized with respect to minimizing the root mean square error of the polarizability [25]. For each molecule in the parametrization, each component was divided by the isotropic polarizability. This was done to keep every molecule equally important. To parametrize the frequency dependence, the difference between the reference and static polarizability was used for the parametrization [26].

The parameters of the model given in table 3.2

Parameter	Interpretation	Equation
$\alpha^*$	Atomic isotropic polarizability	eqs. (83) and (86)
$x^*$	Atomic anisotropy factor	eq. (86)
$\eta_I^*$	Atomic chemical hardness	eqs. (72) and (87)
$C^*$	Steepness of resistance to charge transfer in molecule.	eq. (88)
$R^*$	Atomic radius for nuclear overlap and bond order	eqs. (84) and (88)
$c_I^{q*}$	Inverse number of oscillating charges.	eqs. (64) and (104)
$c_I^{\mu*}$	Inverse number of oscillating dipoles.	eqs. (65) and (105)
$\Phi_I^*$	Width of Gaussian charge distribution.	eqs. (68) and (69)
Locked:		
$\chi_I^*$	Atomic electronegativity	eq. (72)
$\gamma_I^q$	Dissipation of charge	eq. (104)
$\gamma_I^d$	Dissipation of dipole	eq. (105)

Table 3.2: Atom type parameters in the model. All parameters are given in atomic units.

It was chosen to lock certain parameters at fixed values. This was done to the atomic electronegativity  $\chi_I^*$  due to the polarizability calculation being independent of the atomic electronegativity [27] as shown in eq. (106), It was therefore chosen that as this parameter did not effect the result of the calculation, it would slow the optimization and not effect the resulting values, however it would be useful for other calculations, such as the dipole moment [27]. Likewise the dissipation of the atomic dipoles and atomic charges was locked to zero. This will only effect the frequency dependent part of the polarizability, as seen in eqs. (104) and (105). Because the dataset consists mostly of smaller molecules, it is unlikely that this will introduce a large error.

## 4 Results and discussion

### 4.1 Parametrization of the force field

A selection of 374 molecular geometries were used in the parametrization of the force field. In total the molecules consisted of molecules of the type C, H, O, F, Cl, N and S. The total number of nuclei in the dataset were:

		<hr/>	
		Nucleus	Number
<hr/>		<hr/>	
Unique molecules	224	H	2375
Modified geometries	150	C	1536
<hr/>		F	348
Total	374	O	143
<hr/>		N	60
(a) Geometries in parametrization		Cl	17
		S	30

(b) Different nuclei in the parametrization, for both unmodified and modified geometries together.

Table 4.1: Overview of geometries and nuclei in the parametrizations

The molecules in the parametrization were selected to belong to one or several functional groups. The number of molecules belonging to each groups is shown in table 4.2.

Classifications	Number	Stretched
Alcohols	10	0
Aldehyde	11	26
Alkanes	12	4
Alkenes	11	5
Alkynes	8	19
Amines	14	0
Branched alkanes	5	0
Carboxylic acids	8	0
Chlorinated	10	0
Cyclic alkanes	3	0
Cyclic alkenes	6	0
Dichlorinated	2	0
Difluorated	4	12
Esters	9	6
Ethers	5	0
Fluorinated	8	18
Hetrocycles	11	0
Hydroxy groups	9	0
Ketone	5	0
Nitriles	8	35
Partially fluorinated polyenes	9	0
Perfluorated	5	13
Perfluoro polyenes	8	0
Phenyl	5	0
Polyenes	8	2
Thioaldehydes	4	0
Thioethers	7	0
Thioketone	4	0
Thiols	9	4
Trichlorinated	1	0
Trifluorinated	5	6

Table 4.2: Number of molecules and stretched geometries belonging to different categories of molecules in the dataset.

## 4.2 Description of bond distributions

The description of bond distributions in the dataset is important to visualize what bond distributions are under-represented in the dataset. It is believed that the low coverage of bonds in some areas in the optimization may cause the optimization algorithm to misestimate the correct parameters for these areas. The error molecules containing these bonds can be comparatively large by themselves, but achieving a lower error for all of the better represented bond interactions may cause the optimization algorithm to favour this as the total error decreases. It should therefore be noted what bond distributions are most represented in the data.

### 4.2.1 Carbon-Carbon bonds

The distribution of bond lengths in the parametrization were visualized, the distribution of bond length, for the unmodified bondlengths is shown below in Figure 4.1a. In Figure 4.1b the modified geometries are included. The primary focus of the stretched geometries were on the carbon carbon bonds. This is the most common bond in the dataset, can be very sensitive to chemical environment, and it was believed achieving a good parametrization for the atomic carbon parameters were of primary importance.

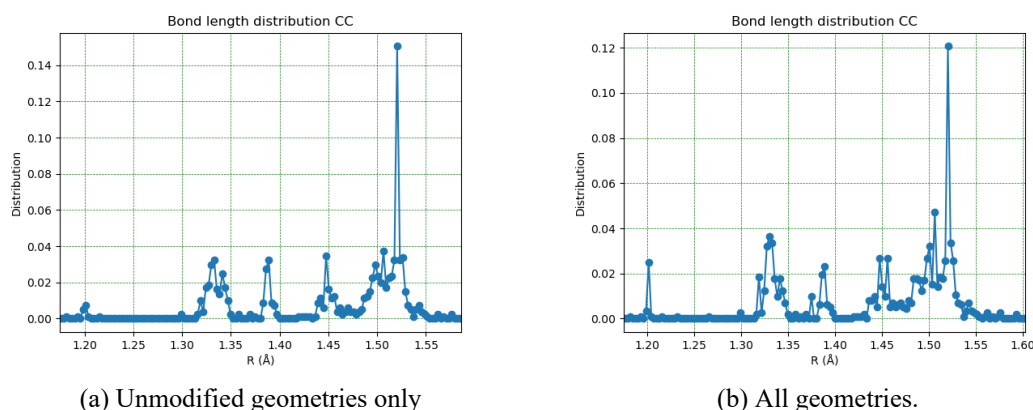


Figure 4.1: Carbon-carbon bond length distribution

It can be seen that some additional bonds have been added between fig. 4.1a and fig. 4.1b the area between 1.45 Å and 1.50 Å for example.

By supplying the force field optimization algorithm with more data in this region, a better estimate for the true parameters may be reached. Care must be taken not to stretch between bond orders. The area between 1.30 Å - 1.40 Å consists of double bonded carbon bonds, the area of larger bond lengths than this consists of the single bonded carbon



atoms. As the model contains explicit parameters for normal bond lengths, eq. (84), this can be problematic if a single-bond is distorted into the double bond range. The peak around 1.20 Å is the triple bonded carbons. Above 1.48 Å, the distribution consists of single bonded carbons. In the area 1.45 Å and 1.50 Å there exists a mixture of aromatic bonds and single bonds next to a particularly electronegative functional group, such as the bond between carbon  $\alpha$  to, and included in a nitrile group.

#### 4.2.2 Carbon-Hydrogen bonds

Bond lengths for hydrogen is visualized below fig. 4.2.

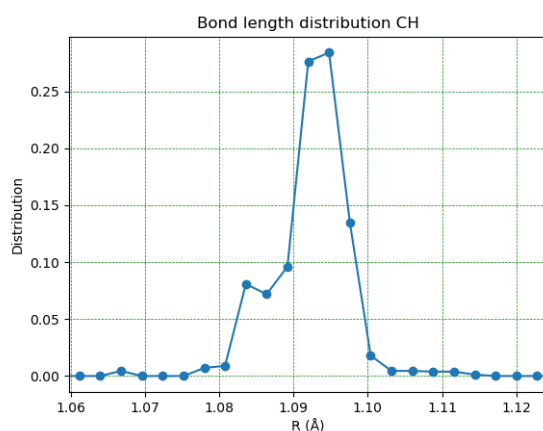


Figure 4.2: Carbon-hydrogen bond length distribution.

It can be seen that the carbon hydrogen bonds show a wide distribution of different bond lengths. There can not be seen any areas where the distributions is very low within the normal bond lengths. The shortest bonds in small peak at 1.07 Å are hydrogen bound to terminal alkynes, while the longest at 1.10 Å consists of hydrogen along a carbon chain, such as octane, these may be slightly longer due to steric hindrance from neighbouring carbon and hydrogen. As such the carbon-hydrogen bond lengths did not need to be supplemented.

### 4.2.3 Carbon-Fluorine bonds

The bond distributions for carbon-fluorine bonds is shown below in figs. 4.3a and 4.3b.

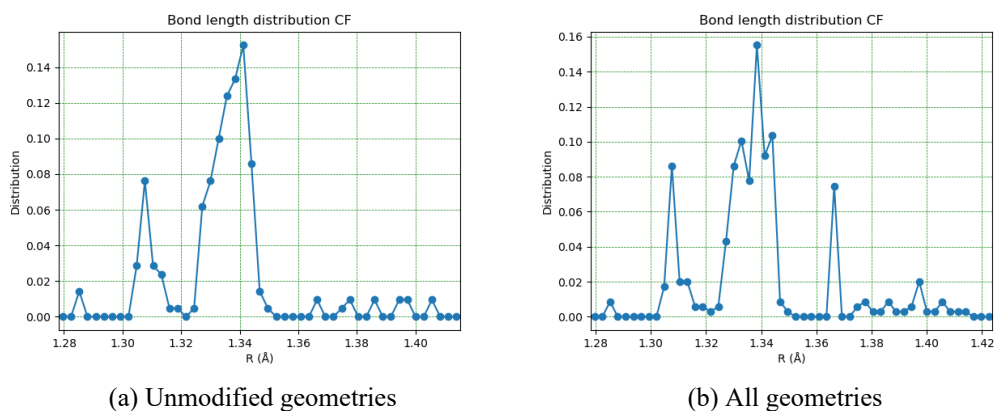


Figure 4.3: Carbon-fluorine bond length distribution

The peak at 1.285 Å consists of fluorine bonded to tripple bonded carbon, for example difluoroethyne. In the area between 1.30-1.32Å the bonds consist of the terminal bonds from the per-fluorinated polyenes. The bond between 1.32-1.35 consists of the C-F bonds from perfluorinated polyenes not in the terminal position, as well as trifluoromethyl groups substituted on the end of the partially fluorinated polyenes. Above this the distribution consists of Fluorine connected to only single bonded carbons.

It is expected that the non terminal bonds are longer in the perfluorinated polyenes, this can be because of steric effects from the carbon chain, while the terminal bonds can be more relaxed and draw closer to the carbon. The terminal trifluoromethyl groups also show lower bond lengths than other fluorines bonded to single bonded carbon. This could be explained by the double bonded  $\alpha$  carbon. the conjugated  $\pi$ -bonds may allow more electron mobility and thus allowing the fluorine bonds to relax.

The large peak appearing in the stretched parametrization at 1.37 arises from the stretching of the carbon carbon bond in 11-difluoroethane. Due to this the two carbon-fluorine bonds in the molecule is counted multiple times.

#### 4.2.4 Carbon-Nitrogen bonds

The bond distribution of carbon nitrogen bonds is shown as fig. 4.4

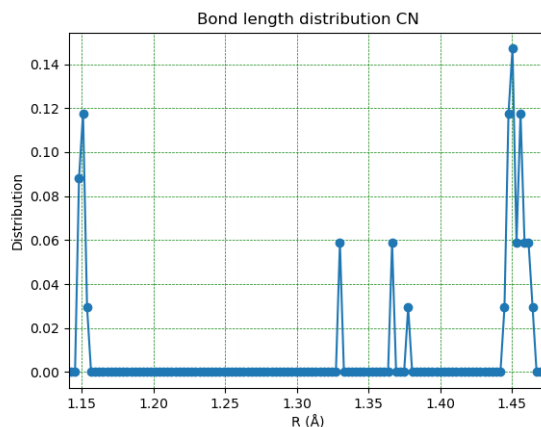


Figure 4.4: Carbon-nitrogen bond length distribution for unmodified geometries

The peak at 1.15Å is nitrile groups. The triple bond between carbon and nitrogen causes the short bond distance. In the area between 1.44-1.46Å the amines can be found, the single bond allows for longer bond lengths than the double and triple bonded molecules. At 1.33Å is pyridine, and at 1.37 is pyrrole. The bond in pyridine is aromatic and is therefore has an unusual bond length, while pyrrole is only singly bonded to the carbon in the ring, and not part of the aromaticity, it is bonded to the doubly bonded carbons, which has the effect of constraining the size of the ring, explaining the bond lengths of pyrrole while tetrahydropyrrole is found among the amine bonds. At 1.38 is prop-1-ene-2-amine, this too is bonded to the double bonded carbon, and thus a shorter bond length is achieved.

#### 4.2.5 Carbon-Oxygen bonds

Oxygen-Carbon bond distribution is shown in fig. 4.5

Between 1.18 Å to 1.22 Å is carbon oxygen double bonds, consisting of several functional groups such as aldehydes, ketones, esters, and carboxylic acids. Between 1.31-1.33 Å there are the single bonds to the hydroxy-group in the dicarboxylic acids ethanedioic-acid and propanedioic-acid. These are shorter than the C-O bond to the hydroxy-group in for example ethanoic acid (1.35 Å) compared to ethanedioic-acid (1.325 Å and 1.337 Å). This may be due to the possibility for intramolecular hydrogen bonds between the two carboxylic acid groups in the dicarboxylic acid giving a different electron density

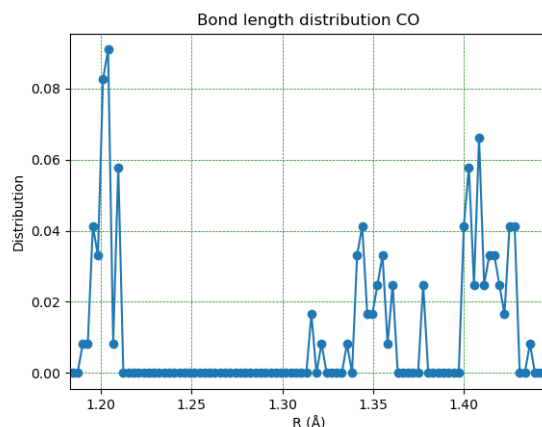


Figure 4.5: Carbon-oxygen bond length distribution for unmodified geometries

near the double bonded oxygen and allowing for a shorter bond distance. Between 1.34-1.36 Å are the normal length carbon-hydroxyl bonds in carboxylic acids. Also in this bond area is the bond to the ether bridge in esters, or esters where there is no double bond  $\alpha$  to the ester group.

From 1.39 Å and above are alcohols, ethers, and singly bonded carbon oxygen bonds. As these are singly bonded to  $sp^3$ -hybridized carbon, they are expected to have a longer bond length than bonds to doubly bonded carbon.

#### 4.2.6 Carbon-Sulfur bonds

Sulfur-Carbon bond distribution is shown in fig. 4.6.

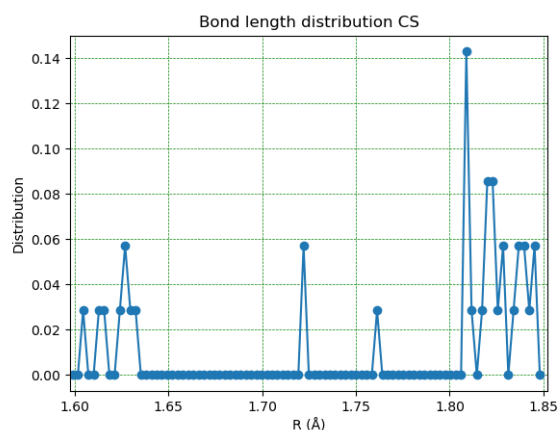


Figure 4.6: Carbon-sulfur bond length distribution for unmodified geometries

The bonds between 1.58 Å to 1.65 Å are thioaldehydes and thioketones. At 1.70 Å is thiophene, these two bonds are singly bonded to double bonded carbons. The peak at 1.76 Å is the bond to an ethenyl group in the thioether methylsulfanylene. This bond is somewhat longer than the bond in thiophene due to being less constrained by the cyclicity of the ring in thiophene. Between 1.80 Å and 1.85 Å are the carbon sulfur bonds in thiols and thioethers. These are singly bonded to  $sp^3$  carbon and are therefore found to be at the longest distance from the carbon.

#### 4.2.7 Carbon-Chlorine bonds

Chlorine-Carbon bond distribution is shown in fig. 4.7.

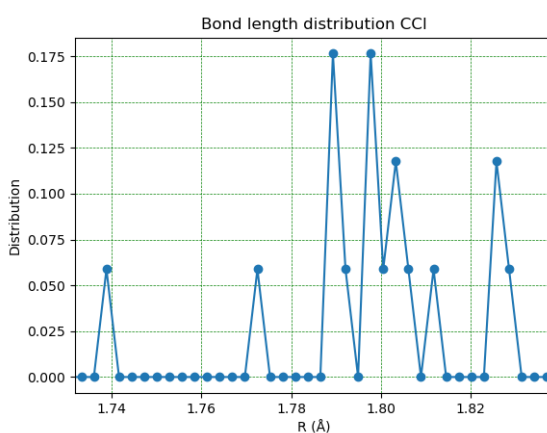


Figure 4.7: Carbon-Chlorine bond length distribution for unmodified geometries

At 1.74 Å there is chlorine bonded to a double bonded carbon in 1,3-dichloroprop-1-ene. This causes the bond length between carbon and chlorine to shorten as the electron density is more shared between the carbon-carbon double bond. At 1.775 Å is chlorocyclopropane. The constrained nature of the cyclopropane bonds may cause less steric hindrance to the chlorine as the carbon bonds are at a smaller angle than normal for an alkane and this allows for some smaller carbon chlorine bond distance. Between 1.79 Å and 1.80 Å is trichlorobutane, the single bonded chlorine of 1,3-dichloroprop-1-ene, and chloromethane. These chlorine atoms are bonded to  $sp^3$  hybridized carbon and therefore have similar bond lengths to each other.

### 4.3 Optimized parameters

The optimized parameters for the model based on unmodified geometries is given in table 4.3. The optimized parameters using all geometries is shown in table 4.4

Model parameters unstretched geometries [a.u.]							
Atom							
Parameter	C	Cl	F	H	N	O	S
Static polarizability parameters							
$\alpha^*$	6.556	12.909	2.042	1.663	6.867	3.119	15.331
$x^*$	0.015	0.064	0.134	0.416	0.008	0.056	0.081
$\eta^*$	1.446	2.982	3.640	3.588	2.432	2.586	0.750
$C^*$	1.072	22.437	12.140	3.310	27.866	3.429	1.939
$R^*$	1.265	3.122	1.093	0.656	1.216	1.289	1.968
$\Phi^*$	0.024	78.112	1.552	5.633	0.270	2.086	1.507
Frequency dependent polarizability parameters							
$c^q^*$	0.665	-0.000	0.000	4.103	0.000	0.000	0.000
$c^\mu^*$	0.763	0.000	0.000	2.925	0.000	-0.000	0.000

Table 4.3: Model parameters for unstretched geometries in atomic units. Only parameters used in the calculation is included

Model parameters stretched geometries [a.u.]							
Atom							
Parameter	C	Cl	F	H	N	O	S
Static polarizability parameters							
$\alpha^*$	6.953	12.782	1.981	1.767	5.471	2.896	15.594
$x^*$	0.013	0.082	0.125	0.303	0.006	0.048	0.070
$\eta^*$	1.760	3.381	3.640	4.684	1.948	2.796	0.861
$C^*$	0.989	5.442	12.254	5.025	29.147	7.506	4.011
$R^*$	1.263	3.388	1.067	0.651	1.226	1.315	2.026
$\Phi^*$	0.038	12.166	1.935	6.195	0.387	1.862	0.645
Frequency dependent polarizability parameters							
$c^q^*$	0.680	-0.000	0.000	8.997	-0.000	0.000	0.000
$c^\mu^*$	0.695	0.000	0.000	1.751	0.000	0.000	0.000

Table 4.4: Model parameters for stretched geometries in atomic units. Only parameters used in the calculation is included

### 4.3.1 Static polarizability parameters

The isotropic and anisotropic polarizability parameters  $a_I^*$  and  $x_I^*$  are in close agreement for all atoms. This may indicate these parameters are important for the parametrization and that the model may be sensitive to changes in this parameter. The isotropic and anisotropic parameters are included in the calculation of the calculation of the polarizability, through the dipole-dipole interaction tensor  $T_{II,\alpha\beta}^{(2)}$  in eq. (81). The calculation of the atomic polarizability is also related to the  $\Phi_I^*$  and  $R_I^*$  parameters through the calculation of the intermolecular charge transfer resistance  $S_I J$  in eq. (84). This resistance then scales the contribution of the  $\Gamma_{I,\alpha\beta}$  matrix eq. (83). This relation then describes the relation between the polarizability parameters and the parameters  $\Phi_I^*$  and  $R_I^*$ .

The atomic isotropic polarizability  $a_I^*$  is in close agreement between the two parametrizations except for nitrogen where the unstretched parametrization gives a value of about 1.4 a.u. higher than the stretched geometry parametrization. It is seen that while the  $a_I^*$  parameter of carbon is higher in the stretched geometry parametrization, all the other atoms have decreased. This may be an effect of the optimization target is the result of a molecular property, meaning the sum is of polarizability contributions is important but that the individual atomic contributions are not necessarily important. This allows for several solutions to the molecular polarizability, and therefore allows for several parameter values with similar results, given that they provide the same sum.

The anisotropic polarizability  $x_I^*$  is a correction to the polarizability eq. (86) such that the out-of-plane elements of the polarizability are given an additional contribution. These elements are expected to be low compared to the in-plane polarizability. It is seen that the anisotropy is on a smaller order of magnitude than the isotropic parameter, which is in line with this expectation. The anisotropic polarizabilities are low, but the value for fluorine and hydrogen can be seen to be comparatively high, indicating a larger degree of anisotropy. Due to the comparatively large values of  $\alpha_I^*$  however the atomic polarizability is still expected to be dominated by the in-plane elements.

The atomic chemical hardness is related to the polarizability through the calculation of the interatomic distance, where the sum over the same index is replaced with the chemical hardness eq. (75). As the atomic chemical hardness can be interpreted as the molecules resistance to change in chemical potential, it is expected that more electronegative atoms should have larger hardness. Looking at the parametrizations, it is seen that the halogens fluorine and chlorine have the greatest hardness. Oxygen, nitrogen, and carbon too follow this trend from the periodic table, but hydrogen can be seen to have a large hardness in both parametrization, and especially in the parametrization of all the geometries. This can be seen as there being a large energy cost from charge transfer away from hydrogen atoms. As the hydrogen only has one electron it therefore explains that charge transfer is especially energetically disadvantageous. Sulfur is a group 16 molecule but is here

seen as having the lowest atomic chemical hardness in both parametrizations. This can indicate the sulfur atoms are quite likely to transfer charge, and that this can indicate a larger charge contribution for sulfated compounds.

$\Phi_I^*$  and  $R_I^*$  are related to the atomic chemical hardness in the modification of the chemical hardness eq. (87). In this modification the intermolecular charge transfer resistance  $S_{IJ}$  eq. (84) is, like for the dipole-dipole interaction tensor, as well as the intramolecular charge transfer resistance  $g_{I,KM}$  eq. (88). Through  $g_{I,KM}$  the  $C^*$  parameter is also used. Through these calculations all the parameters of the model are employed, all remaining terms are based on scaled interatomic distances, which again rely on the parameter  $\Phi_I^*$ .

Looking at the parameters, first it is seen that the atomic radius  $R_I^*$  follows a trend of increasing for heavier atoms except for fluorine. As these radii determine the inter- and intramolecular charge transfer resistance, it is an important parameter. The atomic radius should be examined in detail for all atoms, as this is related to the parameters  $\Phi_I^*$  and  $C^*$ , this will be discussed in detail for the individual possibilities.

### 4.3.2 Frequency dependent polarizability parameters

It can be seen that for both the stretched and unstretched geometries, that the frequency dependent parameters  $c^{q^*}$  and  $c^{\mu^*}$  are only non-zero for carbon and hydrogen. This indicates that the entire frequency dependence can be described by oscillating charges and dipoles associated with hydrogen and carbon. It seems unintuitive that none of the heteroatoms should give a contribution to the frequency dependence, however it is noted that all the molecules in the parametrizations contain both at least one carbon and multiple hydrogens, as the model parametrization minimizes the molecular polarizability, a molecular property, it is possible that the carbon and hydrogen parameters simply compensate for the lacking contribution of the heteroatom.

### 4.3.3 Smaller parametrization using different bonds

It is possible that one of the contributing reasons for the parameters values of  $c^{\mu^*}$  and  $c^{q^*}$  being zero for heteroatoms is due to the lacking bonds between these type of atoms. For example there are no fluorine-oxygen bonds in the data set, nor were there nitrogen-nitrogen bonds like in an azo-group, or sulfur-sulfur bonds. If these functional groups were included in the set, it is possible the optimization would assign values to these parameters. To test this a smaller parametrization was undertaken. This set of molecules contained a number of interatomic bonds not represented in the original dataset. The nuclei in this set of molecules is described in table A.2 in appendix A. Allowing this parametrization to go to convergence showed that values for  $c^{\mu^*}$  had been assigned to chlorine (0.4) and sulfur (0.44) Still no nuclei except hydrogen and carbon were assigned a value for  $c^{q^*}$ . This shows that the model will predict values for these parameters for



hetroatoms, but it does not appear logical that oscillating dipoles associated with oxygen and nitrogen do not contribute to the frequency dependent polarizability.

To see if the model would accept a solution with this dipole contribution, the parameter value for nitrogen was manually set to 0.2 and the parametrization restarted. It was found that with this the parameter value increased to 1.26 after optimization, this would indicate that the oscillating dipoles associated with nitrogen are more important than carbon in describing the frequency dependence of the polarizability. This is more in line with what is expected from a compound like azomethane. The result also illustrates the important point that the optimization algorithm optimizes towards a molecular property, so that the sum of contributions is what determines the result. This allows some atomic parameters to potentially compensate for a lacking contribution from another atom, explaining the two different results.

To test the rest of the parameters, all values of  $c^{\mu*}$  and  $c^{q*}$ , that were zero were set to an initial value of 0.2. The optimization of parameters were then started. It was found that the optimization converged with none of the parameters becoming zero, the final parameters are shown in table 4.5. This shows that it is perhaps possible for the optimization algorithm to get stuck at parameter values if they are set to zero. The initial values were zero for all atoms except carbon and hydrogen for the full parametrization, explaining why this unphysical result was found.

Model parameters for the small parametrization [a.u.]							
Atom							
Parameter	C	Cl	F	H	N	O	S
Static polarizability parameters							
$\alpha^*$	7.912	12.640	2.165	2.130	5.705	3.076	15.175
$x^*$	0.109	0.024	0.019	0.250	0.018	0.106	0.014
$\eta^*$	2.888	2.284	3.558	9.446	2.457	3.963	3.566
$C^*$	1.687	20.381	51.069	54.398	69.964	1.935	4.268
$R^*$	1.288	16.271	1.306	0.709	1.338	1.423	2.064
$\Phi^*$	0.128	73.124	181.976	20.883	0.129	0.226	0.126
Frequency dependent polarizability parameters							
$c^{q*}$	0.471	1.239	0.897	18.745	2.824	1.103	0.882
$c^{\mu*}$	0.159	0.442	0.244	3.541	0.625	0.230	0.411

Table 4.5: Model parameters for the smaller parametrization in atomic units. Only parameters used in the calculation is included

Looking at the values for the frequency dependent parameters in table 4.5, it is seen that

compared to the full parametrization, the  $c^{\mu^*}$  and  $c^q^*$  parameter values for carbon has decreased. This may be in line with the assumption that in the full parametrization, the carbons compensate for the lacking contributions of the other atoms by increasing its values. Meanwhile hydrogen has increased in these values, perhaps to compensate for the smaller contribution of carbon. This indicates that the carbon parameters of  $c^{\mu^*}$  and  $c^q^*$  may be very important for the frequency dependent polarizability in the model. This is in line with the fact that carbon makes up a large part of the atoms in the dataset.

For the static polarizability parameters in table 4.5, the values for  $\Phi^*$  is very large for chlorine, fluorine, and hydrogen. All these can only bond singly to the other nuclei, and is therefore not expected to be sensitive to the slope in  $S_{IJ}$ , and a very large value will not effect the scaled interatomic distance if it is bonded to a nucleus where  $\Phi^*$  is low.  $C^*$  is seen to be high for chlorine, fluorine, and nitrogen. For the former two, this is likely due to the nuclei only being singly bonded to other nuclei, thus if  $R^*$  is sufficiently large  $C^*$  becomes less important, and may be tuned to be very large. For nitrogen, it is possible to observe single, double and tripple bonds, so this large  $C^*$  parameter may be an indication that the types of bonds are treated quite differently. This is also seen in the full parametrizations tables 4.3 and 4.4.

#### 4.3.4 Comparison of chlorinated molecules

It is noted that the size of parameters between the two parametrizations in tables 4.3 and 4.4 are largely comparable except for chlorine. Especially it can be seen that the  $C^*$  parameter, related to the steepness of the step function is very large for the model containing only unmodified geometries. This indicates that the model has a very steep change between a higher and lower order bond, such that the longer bonds have a strong resistance to charge transfer.

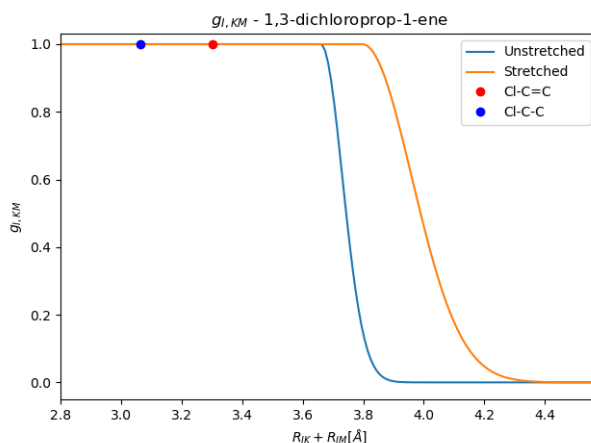


Figure 4.8: Intramolecular charge transfer resistance for 1,3-dichloroprop-1-ene. The two points indicate the function value for the sum of interatomic distances in the molecule. The blue point indicates the chlorine connected to the  $sp^3$  carbon and the red point is chlorine bonded to the  $sp^2$  carbon. The x axis indicates the sum of two interatomic distances.

By visualizing  $g_{I,KM}$  for the two chlorine atoms in 1,3-dichloroprop-1-ene, as shown in fig. 4.9, it can be seen that for both the distances the function value is one. As these two points are well within the area where the function remains constant at one, and the two points indicate the shortest carbon-chlorine bond and one of the longer, as discussed in section 4.2.7, it is likely that the function value is one for all carbon-chlorine bonds in the set and that the steepness of the step function, as defined by the atomic parameter  $C_{Cl}^*$  is completely irrelevant for the molecules in the set. As chlorine can only bond singly to carbon it is reasonable that there is only one value for the intramolecular charge transfer resistance, since there is only one type of bond. The comparatively large value of  $R_{Cl}^*$  indicates that the bond distances must be quite long before the resistance is effected so that it was perhaps not be necessary to optimize the  $C^*$  parameter for chlorine.

Turning to  $\Phi_{Cl}^*$  the parameter is also comparatively very large. The parameter represents the width of the Gaussian charge distribution and such a large value indicates a very localized charge near the chlorine atoms. Further the very small size of the  $\Phi_C^*$  parameter indicates that the intermolecular charge transfer resistance eq. (84) may be very dominated by the  $R_{Cl}^*$  parameter, and that the value will remain constant for all C-Cl bonds in the parametrization. Visualizing the function gives

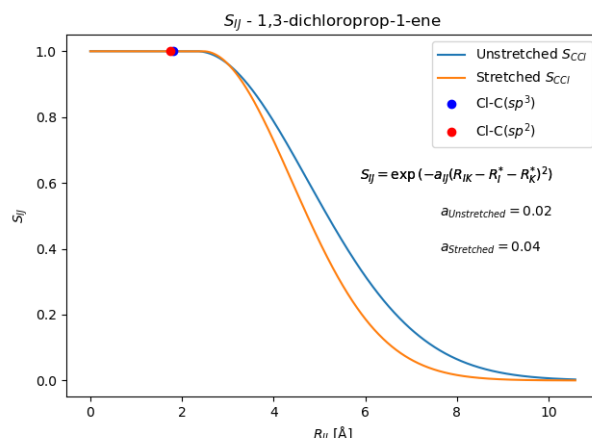


Figure 4.9: Intermolecular charge transfer resistance for 1,3-dichloroprop-1-ene. The two points indicate the function value for the sum of interatomic distances in the molecule. The blue point indicates the chlorine connected to the  $sp^3$  carbon and the red point is chlorine bonded to the  $sp^2$  carbon.

Here too it is observed that the function is one for both shorter and longer values of C-Cl bond length. It is noted  $\Phi_I^*$  is used to approximate the scaled distance using eq. (71). As the parameter value of  $\Phi_C^*$  is so small, the product in the numerator of eq. (69) while the denominator sum is still large due to  $\Phi_{Cl}^*$  that the approximated modified distance may have a large contribution from the modification to the distance. A plot of the polarizability for 2-chloropentane is shown in figs. 4.10 and 4.11

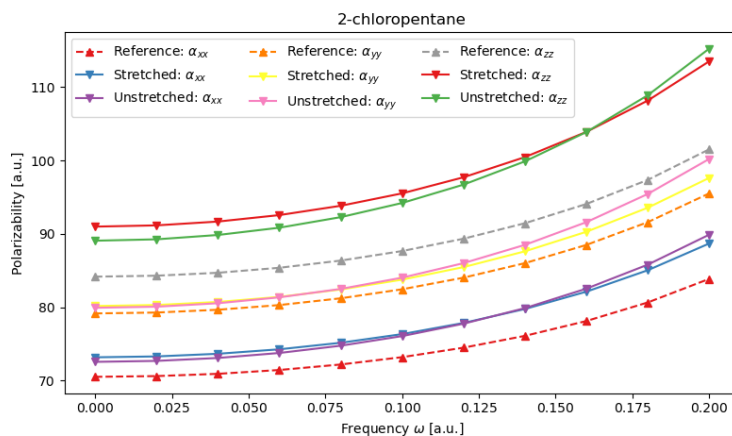


Figure 4.10: Polarizability of 2-chloropentane for the model including stretched and only including unstretched geometries.

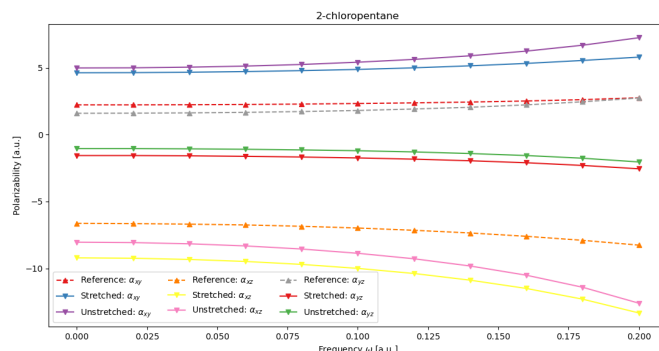


Figure 4.11: Polarizability of 2-chloropentane for the model including stretched and only including unstretched geometries. Showing the off diagonal elements

It can be seen from fig. 4.10 that there is a close agreement between the parametrizations despite tables 4.3 and 4.4 showing large difference in parameter values for chlorine. The frequency independent part of the parametrization appears to be slightly better for the unstretched parametrization. This might indicate that the parameters for carbon are better optimized for other functional groups. None of the molecules that were stretched included a carbon-chlorine bond. The amount of carbon-carbon bonds included from chlorinated compounds is therefore proportionally less than it is for the unmodified geometries and this gives the chlorinated compounds more importance than it gives for the parametrization including all geometries. It can be seen that the full parametrization achieves lower error for higher frequency  $\omega$ . This indicates that the frequency dependent part of the model is better parametrized for the inclusion of all geometries. This is maybe unexpected as both parametrizations had zero contribution from the parameters of any heteroatoms. It may show that the model parameters are optimized such that the parameters for carbon and hydrogen compensate for the lacking contribution of the heteroatom. This would explain why a larger set of molecular geometries may be more general than the smaller set of only including unmodified geometries.

Looking at the dipole and charge contributions can be seen in fig. 4.12. From fig. 4.10, the largest error is in the  $\alpha_{zz}$  component, and therefore only it has been visualized.

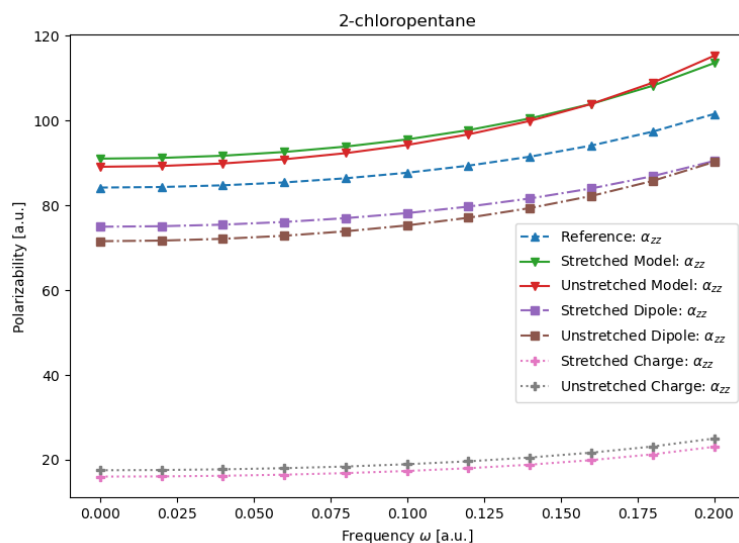


Figure 4.12: Polarizability of 2-chloropentane for the model only including the charge and dipole polarizability contributions.

It is seen that the unstretched parametrization predicts a slightly higher contribution of the polarizability from charges, while the stretched geometry predict higher dipole contribution. The frequency dependence to the dipole contribution has a slightly higher trend in the unstretched parametrization. This may indicate that the unmodified geometry parametrization predicts an excitation energy at a lower frequency than the full parametrization and the DFT reference data.

### 4.3.5 Parametrization of hydrocarbons

Alkanes contain only singly bonded carbon and hydrogens. This leaves a somewhat simple molecule with not many different types of interactions. For carbon, it can be seen that the parameters are in close agreement between the parametrizations tables 4.3 and 4.4, for hydrogen however there is some disagreement. It can be seen that the for the frequency dependent parameters,  $c^{q*}$  is much larger in the stretched parametrization (9.00) than the unstretched parametrization (4.10) and  $c^{\mu*}$  is nearly twice as large in the unstretched parametrization (2.93) than the stretched parametrization (1.75). This shows that the presence of oscillating charges is much more important in the parametrization containing stretched geometries, while oscillating dipoles are more important in the parametrization containing only unmodified geometries.

Further it can be seen that the parameters for the static polarizability there is close agreement between the two parametrizations in the case of carbon. For the  $C^{f*}$  parameter,

it is seen that carbon has the lowest value of all the atomic parameters. This shows that the steepness of the step function  $g_{I,KM}$  eq. (84) is much lower than for the other parameter. This is in line with the purpose of modifying the chemical hardness for different bond orders of carbon. Visualizing the step function for the two carbon-carbon interatomic distances in propane illustrates this.

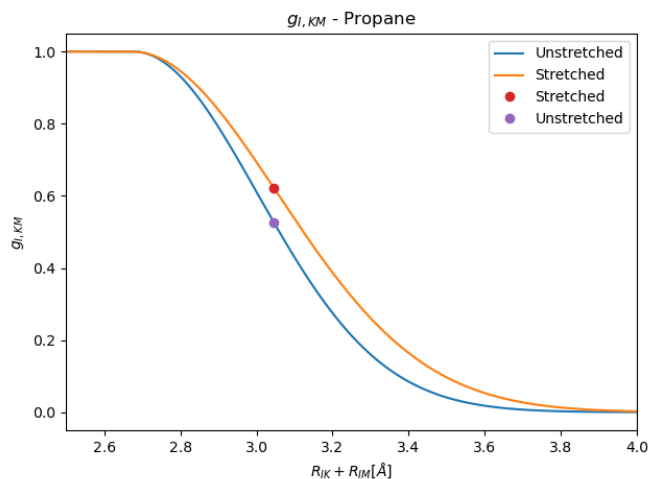


Figure 4.13: Step function for three carbon atoms, and example function value for propane carbon-carbon distances. The x-axis represents the sum of two interatomic distances.

It is seen that the step function does offer a modification the chemical hardness for bond lengths near those found in propane. If either bond is shortened, as it would be for an  $sp^2$ - or  $sp$  hybridized carbon, the function value would increase for this shorter distance. This shows that the step function does have an appropriate shape for the parameters of carbon atoms, since it is able to separate between  $\sigma$  bonded and  $\pi$  bonded carbon, however it is not able to separate between double and tripple bonded carbon.

It can also be seen that the  $\Phi^*$  parameter is very small for carbon, indicating the Gaussian distribution is very diffuse. This small Gaussian can be seen as the carbon charge being very spread out, this might make sense for conjugated  $\pi$  bond systems where the electrons may be seen as quite delocalized, it would indicate that the carbons charges generally have large overlap. This can be seen by visualizing the intermolecular charge transfer resistance  $S_{IJ}$  for carbon carbon bonds fig. 4.19.

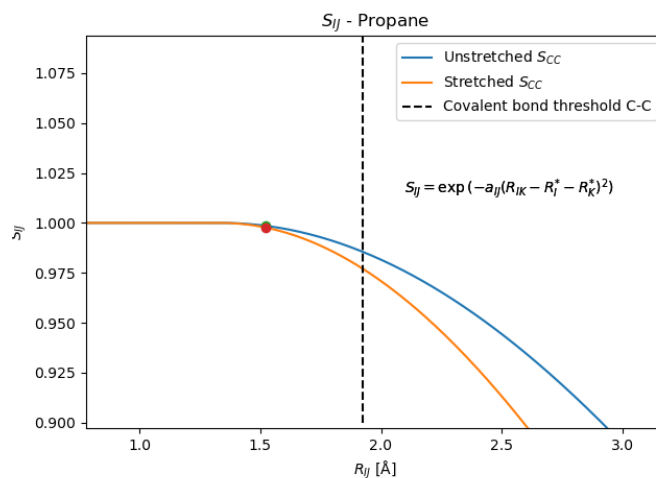


Figure 4.14: Carbon-carbon overlap function with a point for the function value for a C-C bond length in propane.

It is seen from fig. 4.19, that the overlap function  $S_{IJ}$  is very high for any reasonable carbon-carbon bond distance. This is connected to the low value of  $\Phi^*$  as the steepness of the curve is connected to  $a_{IJ}$  eq. (69). This shows that the overlap function functions well for bonded carbon as it remains close to one for any carbon-carbon bonds. It further shows that  $R_i^*$  is an appropriate size for carbon, in line with what was seen in fig. 4.13.

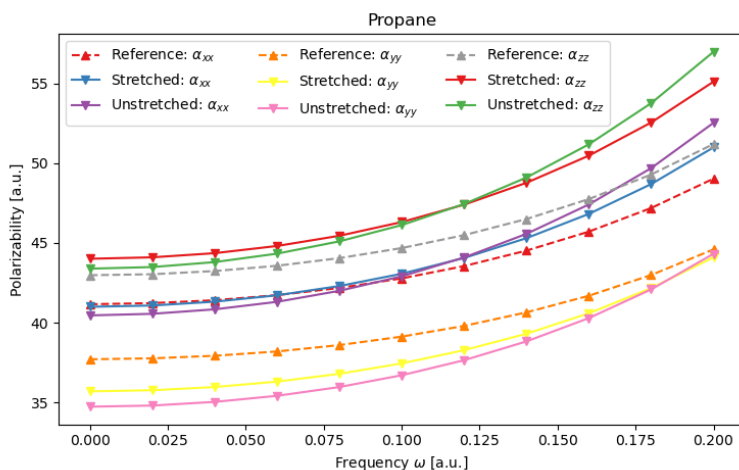


Figure 4.15: In-plane components of the polarizability of propane for the two parametrizations. The stretched parametrization is here the parametrization with added geometries, and the unstretched parametrization only includes unmodified geometries.



From fig. 4.15 it can be seen that there is better agreement between the unstretched parametrization and the reference data in the x-direction, while in the y, and z directions, the parametrization containing all geometries is in better agreement. It also seems the frequency dependence is overestimated for both models, but with better agreement for the full parametrization for all components. This implies the dipole contribution of hydrogen to the frequency dependent polarizability is overestimated for the model based on unmodified geometries. Looking at methane fig. 4.16:

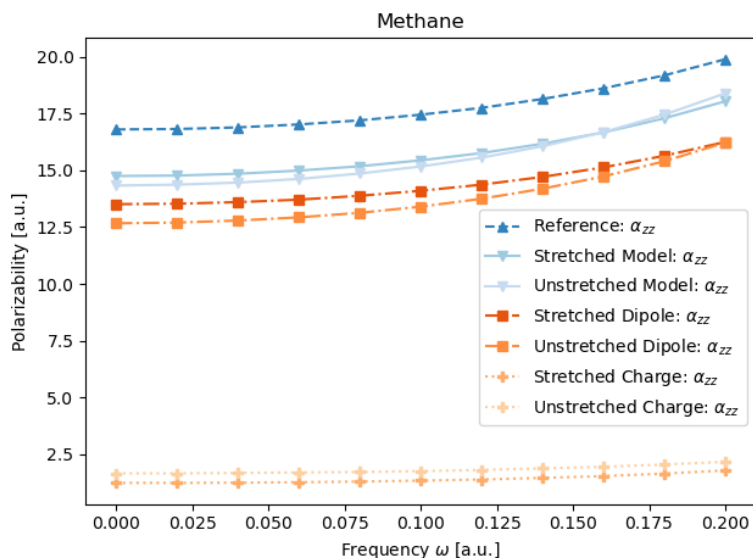


Figure 4.16: Polarizability along the z direction for two parametrizations of methane. Polarizability contributions are shown.

In methane there is only one carbon bonded to each of the hydrogen, as such the only interaction is between hydrogen the hydrogens and this central carbon. It can be seen that it is the dipole component of the polarizability that is giving too large a contribution for the unstretched parametrization, for higher frequency. As  $c^{\mu*}$  only has a parametric dependence on the inverse number of oscillating dipoles when neglecting dissipation eq. (105), it can be seen that the parametrization using only the unmodified geometries does not estimate this correctly. When solving eq. (106) for the polarizability contribution, it is the inverse matrix of the first term that is solved. Therefore the higher value obtained for  $c^{\mu*}$  in the unmodified geometry parametrization may lower any excitation frequency in the solution, giving a sharper gradient for the frequency dependence.

By tuning the parameters, the effect of this can be visualised by manually tuning the parameter values for the stretched parametrization fig. 4.17. The stretched parametrization

was chosen as it appeared from fig. 4.16 that these parameters were better for describing the polarizability values for methane.

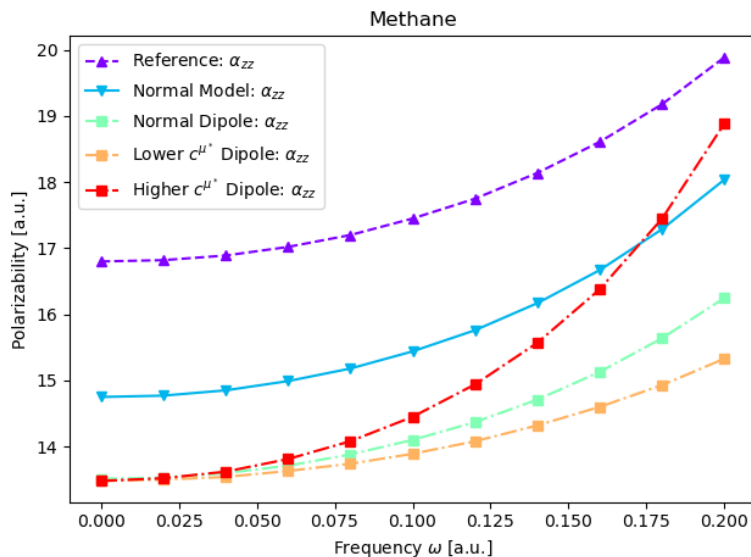


Figure 4.17: Dipole component of methane polarizability with tuning of the  $c^{\mu^*}$  parameter for carbon. Values are based on the parameter values of the stretched parametrization. Higher and lower values for  $c^{\mu^*}$  are produced simply by multiplying or dividing the parameter value by a factor of two.

The charge contribution to the polarizability remains very similar between the parametrizations fig. 4.16, despite  $c^{q^*}$  for hydrogen being higher by more than 1 for the unmodified geometries. The parametrization of  $c^{q^*}$  for carbon also gives a contribution to the dipole component for hydrogen through eq. (104).

Visualizing  $g_{I,KM}$  for the interatomic distance between two hydrogens and the central carbon shows fig. 4.18. It can be seen that the curve is quite steep, due to the high value of  $C^*$  for hydrogen. This means the polarizability contribution from the hardness will vary quite drastically with bond distance. However the carbon-hydrogen bond distances vary quite little in the data set fig. 4.2.

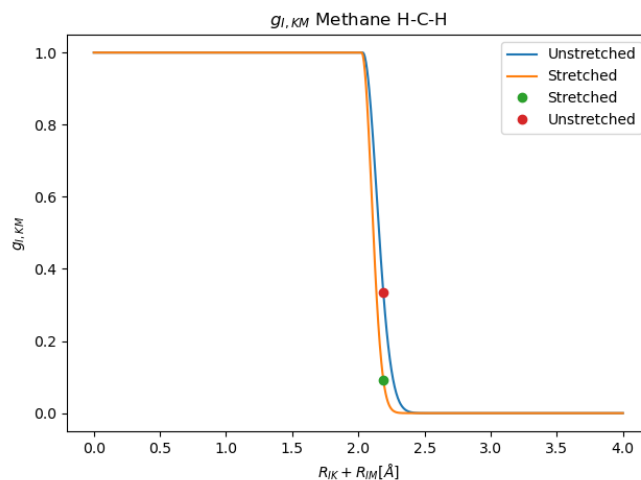


Figure 4.18: Step function for two hydrogen carbon bonds, and example function value for methane carbon-hydrogen distances. The x-axis represents the sum of two interatomic distances.

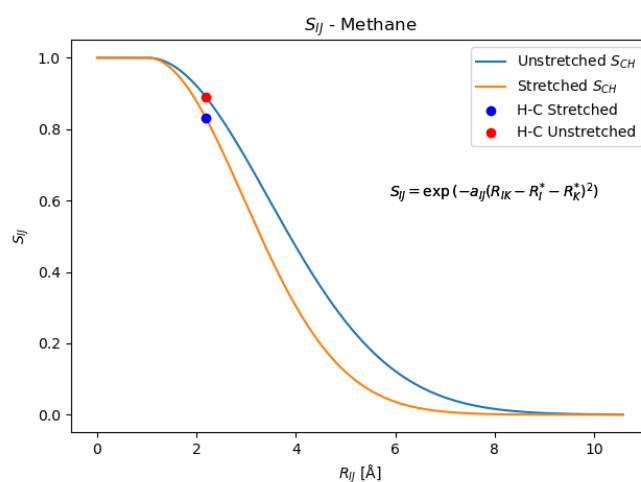


Figure 4.19: Carbon-hydrogen overlap function with a point for the function value for a C-H bond length in methane.

### 4.3.6 Parametrization of oxygen

Carbon can bond to oxygen singly or doubly. It is therefore worth visualizing the  $g_{I,KM}$  for an alcohol group. For ethanol it can be seen that the singly bonded C-O is located somewhat around the middle of the curve. As such it can be seen that if the oxygen was

double bonded, like in a ketone, the charge transfer resistance would be less important. This indicates that the parametrization is able to distinguish between different oxygen bond orders.

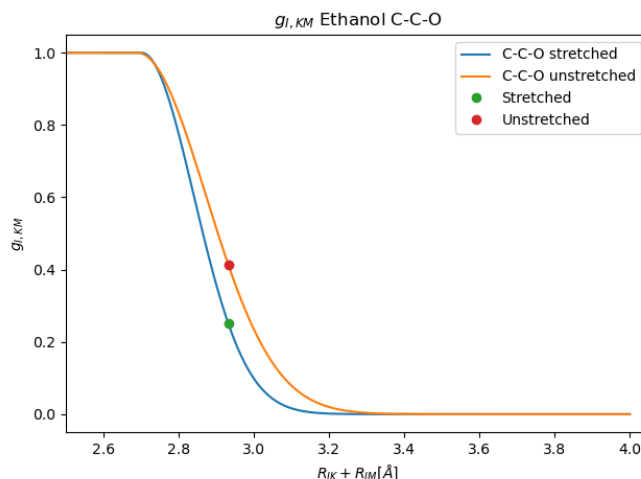


Figure 4.20: Step function for three carbon atoms, and example function value for ethanol carbon-carbon-oxygen distances. The x-axis represents the sum of two interatomic distances and the unstretched parametrization indicates the parametrization without the added geometry, and stretch indicates the parametrization with added geometries.

#### 4.4 The effect of stretching a bond

The effect of stretching a bond on the polarizability is of interest. When the bond is changed, this has an effect on the interatomic distance between two atoms. This is used to calculate the interaction tensors eqs. (78) and (80), as well as the potential overlap and charge transfer resistance between atoms in the molecule. By stretching the central carbon-carbon bond in 111-trifluoroethane it can be seen that fig. 4.21 the DFT polarizability increases with a longer bond length. This may be due to the molecule being a strong permanent dipole due to the large electronegativity of the fluorine. In the presence of an external field, this may stabilize the induced dipole, such that the dipole will increase with larger bond distance.

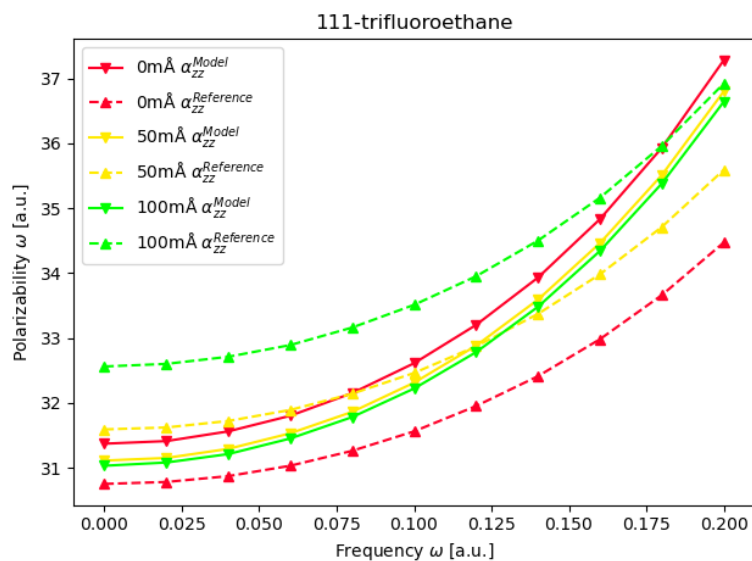


Figure 4.21:  $\alpha_{zz}$  for polarizabilities of trifluoroethane and some of the stretched variants

It is seen that as the bond is stretched, there is less effect on the polarizability. Stretching the bond from equilibrium with 50 mÅ has more effect than stretching from 50 mÅ to 100 mÅ. Further it can be seen that the model polarizability has the same shape for all model values. The size of the effect is also noteworthy. It can be seen that there is less than a half a.u. difference between the unmodified and stretched geometry. Then this shows that the stretching of the bond has little effect. This can be connected to the correction to the chemical hardness, as it is distance dependent. As both fluorine and hydrogen have quite high  $C^*$  parameters, it is likely that this curve could be quite steep. Thus showing why the effect dissipates at longer distances. Further, the polarizability depends on the internuclear distance through the interaction tensors  $T^{(0)}$ ,  $T^{(1)}$  and  $T^{(2)}$ . These scale inversely with the interatomic distances. Thus it should be expected that this would increase with increased bond length. The sum of these two effects may explain the difference in static polarizability between the molecular geometry. The correcting factor to the chemical hardness  $g_{I,KM}$  is shown in fig. 4.22.

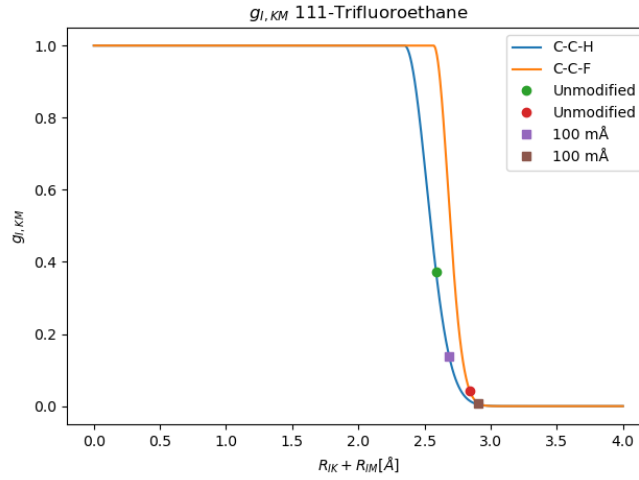


Figure 4.22:  $g_{I,KM}$  for 111-trifluoroethane and a variant stretched along the carbon-carbon bond. Curves are visualized for the carbon-carbon-fluorine and carbon-carbon-hydrogen two atomic distances.

Looking in detail at the polarizability contributions to the polarizability of the stretched molecule fig. 4.23, it can be seen that the unstretched geometry has a slightly higher frequency dependence for the charge contribution. This arises from the dependence on the interatomic distance in eq. (104). This term means that modifying the geometry to increase this bond length causes the pole to shift to a higher frequency as the pole must be inversely proportional to the bond length. This explains the slightly higher frequency dependent polarizability for the charge contribution.

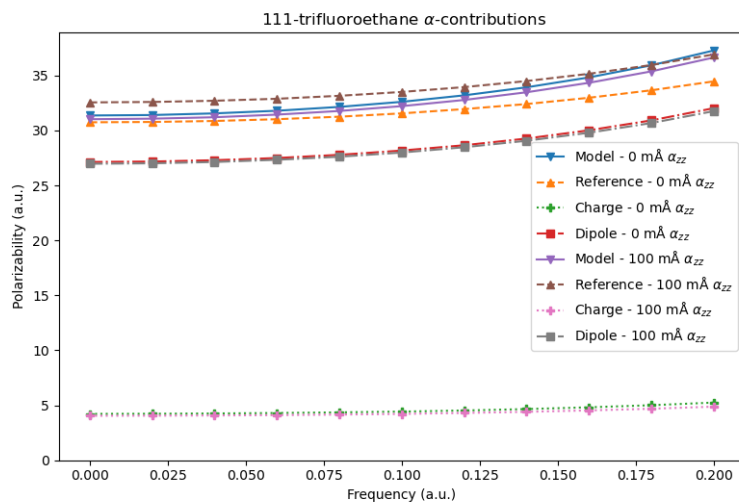


Figure 4.23:  $\alpha_{zz}$  for polarizabilities of trifluoroethane and a stretched variant, with polarizability contributions

## 4.5 Result of the addition of stretched geometries

The overall effect of the addition of adding geometries to the parametrization can be seen from figs. 4.24 and 4.25. It can be seen that both parametrizations are approximately equal in the static polarizabilities. From this it can be seen that the addition of modified geometries did not generally improve the description of static polarizability. From fig. 4.25, the full parametrization was found to be somewhat better. This shows that overall the full parametrization was in closer agreement with the reference data.

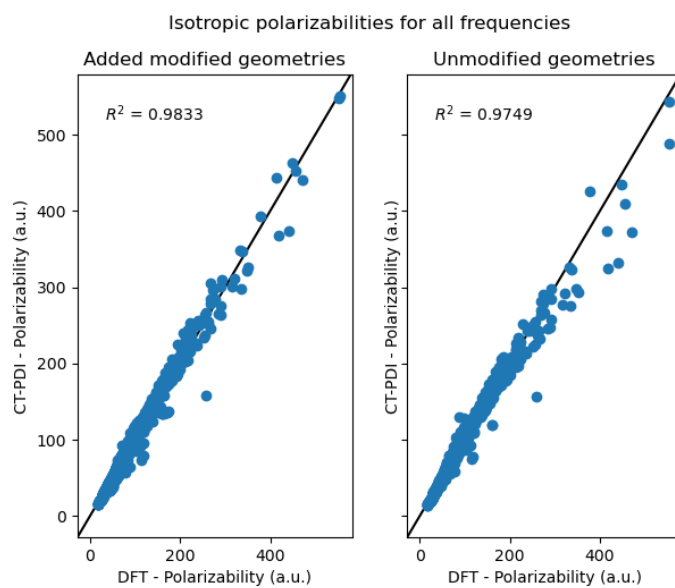


Figure 4.24: Isotropic polarizability (mean trace component) for each molecule in the parametrization including all frequencies. All molecules were given equal weight in the calculation of  $R^2$ .

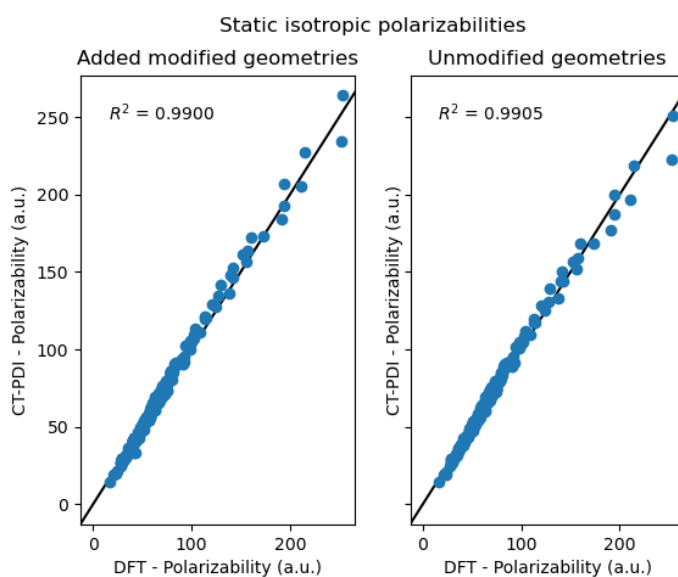


Figure 4.25: Static isotropic polarizability (mean trace component) for each molecule in the parametrization. All molecules were given equal weight in the calculation of  $R^2$ .



### 4.5.1 Addition of carbon bonds

The additional geometries were added in order to better represent less common bond lengths in the dataset. The major portion of this was done for carbon-carbon bonds in the bond length range between 1.46 – 1.48. In this interval the bond between the carbon bond to the nitrile carbon was found. 3-cyano-penta-13-diene shows some large deviation for the higher frequencies. The polarizability is therefore visualized figs. 4.26 and 4.27. It is seen that there is general agreement between the two polarizabilities, and there cannot be said generally that either result is more favourable.

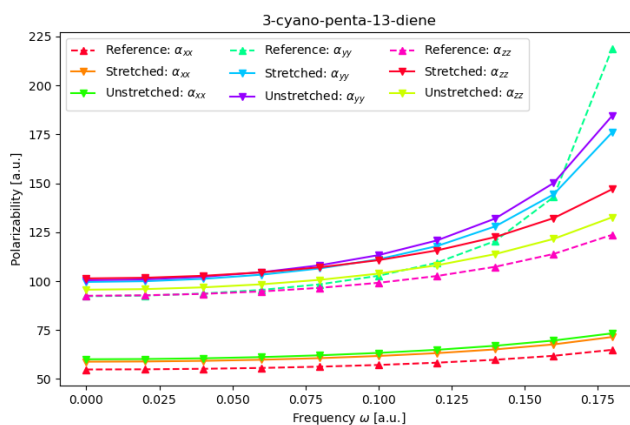


Figure 4.26: In-plane components of the polarizability of 3-cyano-penta-13-diene for the two parametrizations.

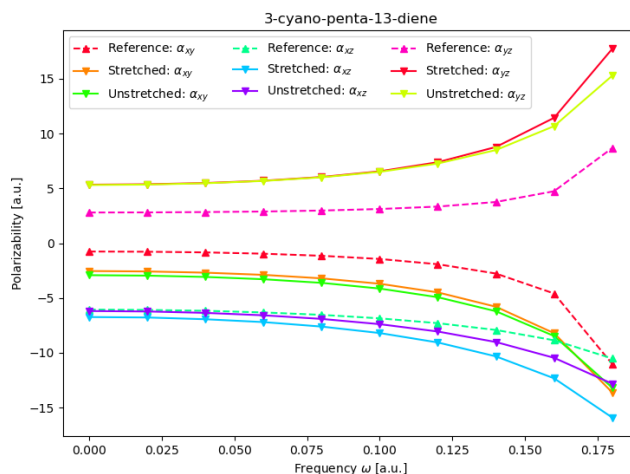


Figure 4.27: Out-of-plane components of the polarizability of 3-cyano-penta-13-diene for the two parametrizations.

The result shows that the description of nitriles were not significantly improved by the addition of more geometries. The addition of these geometries may not have had enough significance for the large number of molecules included, then it be expected not to see a clear improvement.

Looking at molecules with the overall largest error, the polyenes and phenyl molecules are represented. These molecules also contain bond lengths in the interval with added geometries. The errors for polyenes are shown in tables 4.6 and 4.7. It is seen that the RMSE is quite high. This is due to the very high values of the polarizabilities for these molecules. Then even if the error is only on a few percent, it may appear large as a root mean squared error, as this does not account for the magnitude of the reference polarizability, just the difference. It is noted that for the optimization algorithm, steps were taken to keep every molecule equally important as described in section 3.4.

Root mean square error for Unstetched polyenes												
Molecule	Frequency $\omega$ [a.u.]:											Mean
	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	
Penta-13e-diene	1.80	1.83	1.89	2.02	2.22	2.54	3.04	3.80	4.98	6.72	3.38	3.11
Hexa-13e5-triene	2.12	2.12	2.13	2.13	2.16	2.24	2.47	4.00	17.35			4.08
Hepta-13e5e-triene	2.18	2.21	2.31	2.51	2.88	3.53	4.69	7.04	13.83			4.58
Octa-13e5e7-tetraene	4.50	4.56	4.75	5.18	6.13	8.37	14.80	46.67				11.87
Nona-13e5e7e-tetraene	4.22	4.25	4.37	4.67	5.40	7.11	11.73	35.57				9.67
Deca-13e5e7e9-pentaene	12.40	12.72	13.78	16.05	20.82	32.53	75.91					26.32
Undeca-13e5e7e9e-pentaene	13.17	13.50	14.59	16.91	21.79	33.88	80.56					27.77
Dodeca-13e5e7e9e10-pentaene	26.38	27.23	30.09	36.32	50.21	89.44						43.28
Mean:	8.35	8.55	9.24	10.72	13.95	22.46	27.60	19.42	12.06	6.72	3.38	

Table 4.6: RMSE for between model and DFT polarizability at frequency (a.u.) for unstetched polyenes.

Root mean square error for Stetched polyenes												
Molecule	Frequency $\omega$ [a.u.]:											Mean
	0.00	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	
Penta-13e-diene	1.31	1.32	1.36	1.44	1.57	1.80	2.16	2.79	4.00	7.44	28.14	4.85
Hexa-13e5-triene	1.51	1.53	1.57	1.67	1.89	2.30	3.13	5.11	13.16			3.54
Hepta-13e5e-triene	2.40	2.45	2.60	2.90	3.43	4.35	6.15	10.70	33.32			7.59
Octa-13e5e7-tetraene	3.28	3.36	3.62	4.12	5.03	6.78	10.75	25.64				7.82
Nona-13e5e7e-tetraene	3.71	3.80	4.10	4.67	5.72	7.71	12.36	34.42				9.56
Deca-13e5e7e9-pentaene	8.12	8.33	9.04	10.51	13.42	20.07	42.82					16.04
Undeca-13e5e7e9e-pentaene	8.21	8.42	9.10	10.49	13.16	18.87	35.96					14.89
Dodeca-13e5e7e9e10-pentaene	19.19	19.79	21.81	26.16	35.59	61.55						30.68
Mean:	5.97	6.12	6.65	7.75	9.98	15.43	16.19	15.73	16.83	7.44	28.14	

Table 4.7: RMSE for between model and DFT polarizability at frequency (a.u.) for stetched polyenes.

The largest component of the polyene polarizability is the  $\alpha_{zz}$ , it is visualized below. There can not be seen a clear trend in what parametrization is better.

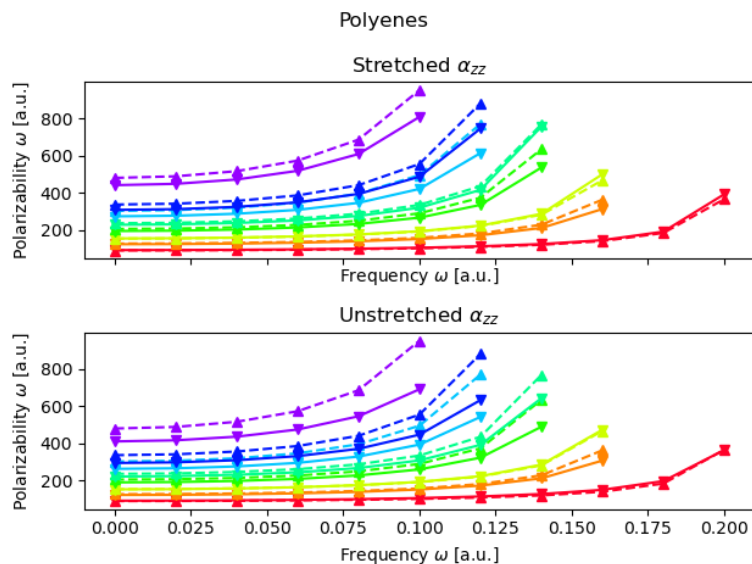


Figure 4.28:  $\alpha_{zz}$  component of the polarizability for polyenes from penta-1(E)3(E)-diene (lowest) to Dodeca-13(E)5(E)7(E)9(E)10-pentaene (highest). The two parametrizations are in close agreement, with the full parametrization using all geometries generally showing less error for the higher frequency.

Looking in detail at Octa-13E5E7-tetraene( fig. 4.29) it can be seen that there is good agreement between the two parametrizations and there is only a slight improvement for the parametrization of all geometries at higher frequency. This strengthens the suspicion that the addition of geometries in this interval has not significantly improved the parametrization.

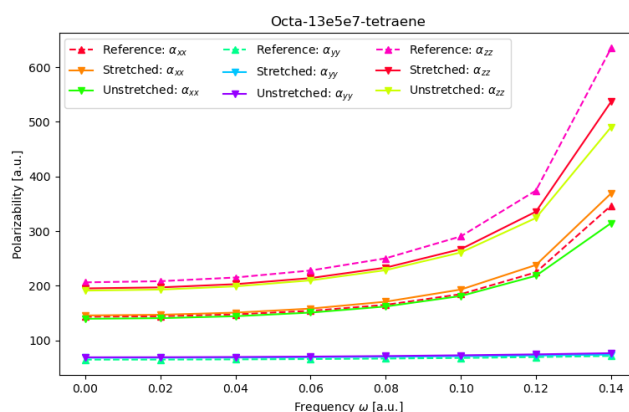


Figure 4.29: In plane polarizability of Octa-13E5E7-tetraene for the two parametrizations.

The same pattern can be seen for perfluorated-polyenes fig. 4.30

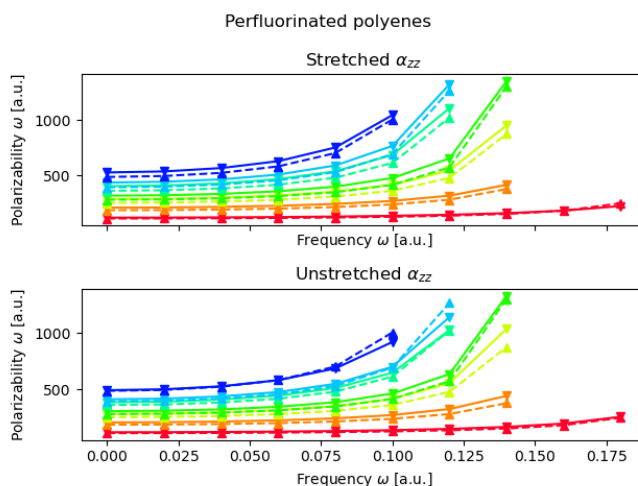


Figure 4.30:  $\alpha_{zz}$  component of the polarizability for polyenes from perfluoro-penta-1(E)3(E)-diene (lowest) to perfluoro-dodeca-13(E)5(E)7(E)9(E)10-pentaene (highest) except for perfluoro-hexa-13(E)5(E)-triene due to an error in the generation of the input file. The two parametrizations are in close agreement.

The single largest error molecule for both parametrizations was phenylethene (styrene). It can be seen from visualizing the polarizability for the two parametrizations fig. 4.32 that this error stems from the large growth in the polarizability for the frequency  $0.18a.u.$  in the reference data, which the model completely fails to capture.

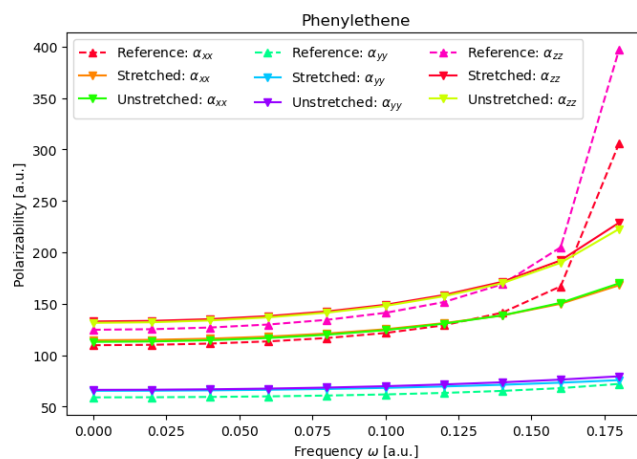


Figure 4.31: In plane polarizability of phenylethene for the two parametrizations.

Looking at the polarizability contributions to phenylethene, it is found that to the degree the frequency dependent polarizability is modelled, it is the charge contribution that depends most on the frequency in the area. This then shows that also the charge contribution may participate in an excitation, however the steepness is not well modelled. It is also noteworthy that both contributions are of almost similar size.

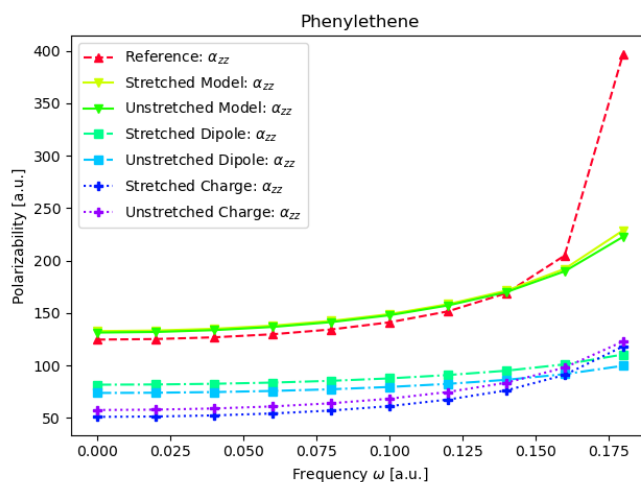


Figure 4.32: Polarizability of phenylethene for the model including the charge and dipole polarizability contributions for both parametrizations.

Looking at an example in 12-difluoroethane fig. 4.33 it can be seen that there is better agreement with the static polarizability for all three components. Further the parametriza-

tion containing the unmodified geometries appear to show a too steep frequency dependence. One reason for this may be the addition of stretched geometries in the bond length interval, giving more weight to these bond lengths. This shows that it can in principle be possible to improve the description of less represented bond distances by addition of stretched geometries, however it should not be seen as a strength of the parametrization that there was a comparatively low number of fluorine bonds to singly bound carbon, as this could have easily been represented with the addition of more perfluorated alkanes, for example.

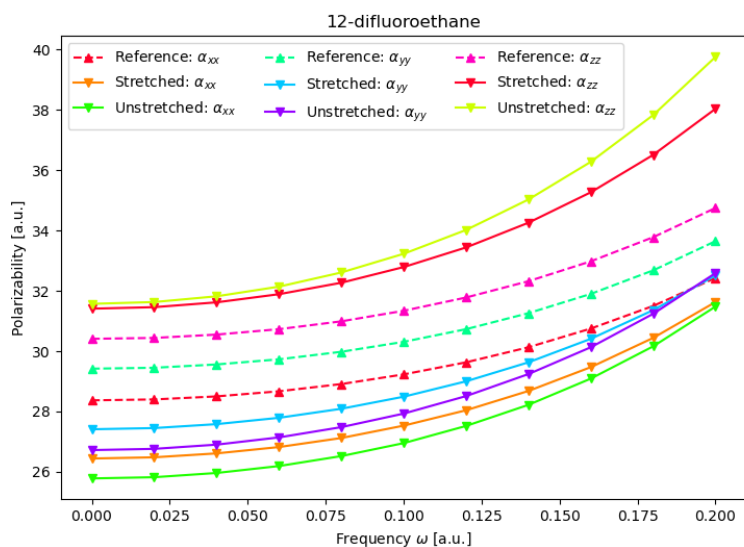


Figure 4.33: Polarizability of 12-difluoroethane for both parametrizations. The ‘Stretched’ parametrization refers to the parametrization including all geometries, the ‘Unstretched’ refers to parametrization including only unmodified geometries.

## 4.6 Parametrization of cyclic compounds

Some cyclic and heterocycles have been included in the parametrization. For cyclopentadiene, there can be seen some differences between the parametrizations fig. 4.36. It can be seen that the model predicts an excitation for the ‘stretched’ parametrization with poles in the x, and y direction. For the ‘Unstretched’ optimization, the pole is not reached but the sharp increase in the polarizability shows that the pole is very close. This then indicates that the model underestimates the excitation value. It is noted that as the pole is reached in the stretched optimization, the polarizability proceeds to fall to below the reference polarizability. As the optimization attempts to minimize the error in the polarizability the excitation may be considered more favourable for the algorithm, despite being a physically incorrect excitation.

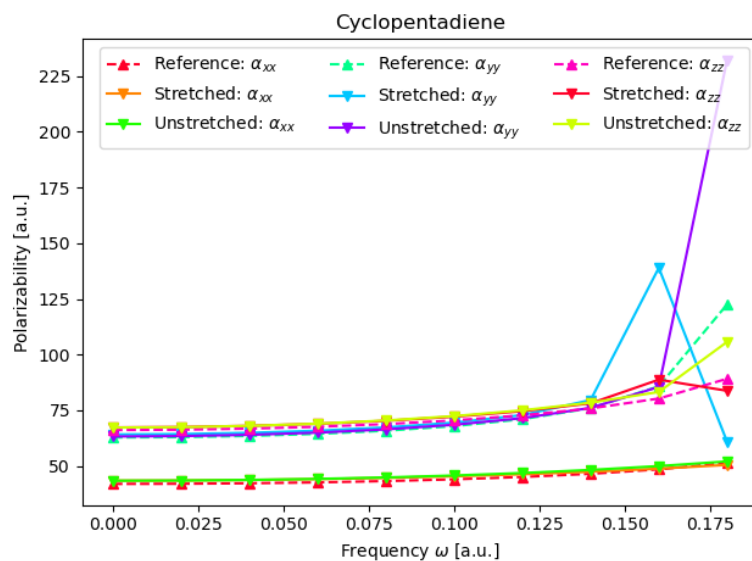


Figure 4.34: Polarizability of cyclopentadiene for both parametrizations. The ‘Stretched’ parametrization refers to the parametrization including all geometries, the ‘Unstretched’ refers to parametrization including only unmodified geometries.

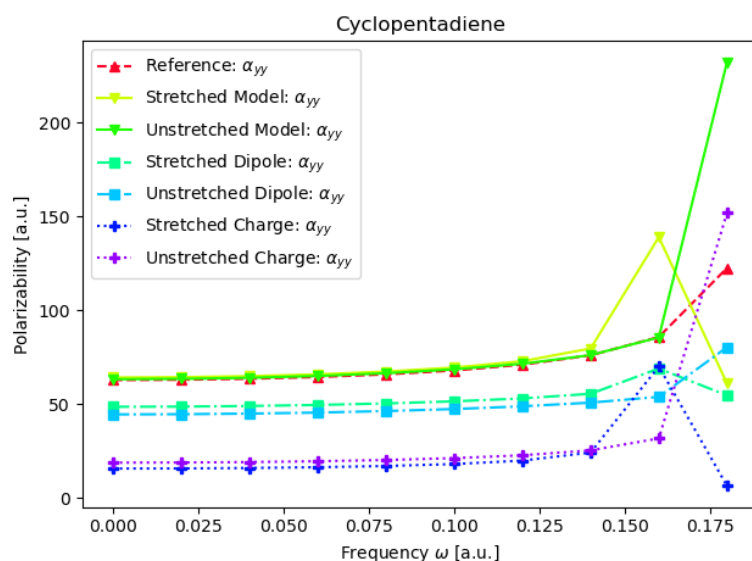


Figure 4.35: Polarizability  $y$  component of cyclopentadiene for both parametrizations with dipole and charge contributions. The ‘Stretched’ parametrization refers to the parametrization including all geometries, the ‘Unstretched’ refers to parametrization including only unmodified geometries.

A similar excitation can be seen for thiophene. At the frequencies 0.6 and 0.8 a.u., a weak excitation can be seen for the ‘stretched’ parametrization including all geometries. The same excitation is found between 0.8 and 0.10 a.u., for the parametrization of the unmodified ‘unstretched’ geometries. These excitations are very weak, but occur well below the DFT predicted excitation above 0.2 a.u., This shows the model is vulnerable to predicting unphysical excitations also well below the true excitation energy. A possible reason for this is that the sulfur does not correctly participate in the polarization, such that the parametrization would be different if there was a sulfur contribution to the frequency dependent polarizability, this pole would have occurred at a higher frequency. Further the discussion of the bond lengths in the parametrization showed that these aromatic rings were often less represented in the bond distribution than they perhaps could have been section 4.2.

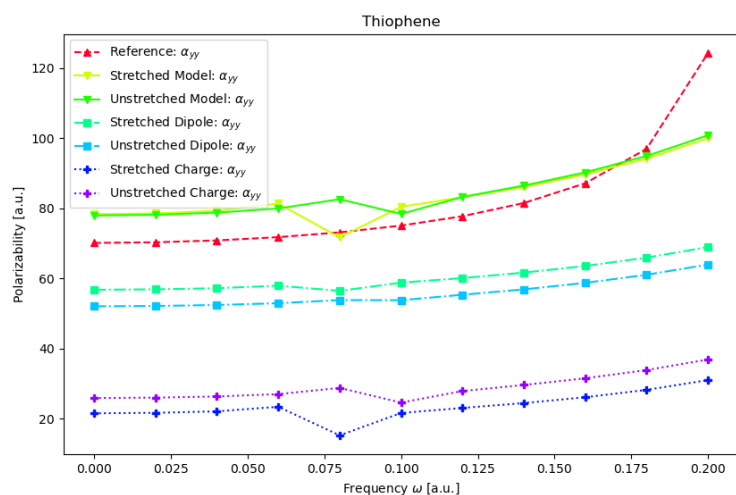


Figure 4.36: Polarizability  $\gamma$  component of thiophene for both parametrizations with dipole and charge contributions. The ‘Stretched’ parametrization refers to the parametrization including all geometries, the ‘Unstretched’ refers to parametrization including only unmodified geometries.



## 5 Conclusion

Two parametrizations of a classical electrostatic model for the polarizability was undertaken. A large set smaller organic molecules were used for the nuclei of types, carbon, hydrogen, nitrogen, oxygen, sulfur, fluorine and chlorine. For the bond lengths in the dataset for which there was less coverage, the molecules were supplemented with modified geometries that had one of their bonds stretched or compressed in order to better represent these bond lengths. The modified geometries were primarily modified by stretching or compressing carbon-carbon bonds, however some carbon-fluorine bonds were also modified.

While the two parametrizations were found to have differing numerical values. There was not a significant improvement to the parametrization that included the larger set of modified geometries. The largest errors in the model came from large highly polarizable molecules. Many of these were the polyenes.

The noticeable difference between the parameter values in the parametrizations were thought to arise from the fact that the model is trained to optimize the parameters with respect to a molecular property, as a sum of atomic contributions. This allows the polarizability to have many equally valid solutions provided the sum of atomic contributions returns the molecular polarizability.

The simplex optimization algorithm that was used was found to be able to get stuck on parameter values for the inverse number of oscillating charges  $c_I^{q*}$  and oscillating dipoles  $c_I^{\mu*}$  if they were initially set to zero, as they were for the main parametrizations. These parameters effect the frequency dependent polarizability. This complicates the results and makes it difficult to make a conclusion on the effect of supplementing the dataset with modified geometries, at least for the frequency dependent components. It is possible that having started the parametrization with better initial values for the parameter values either parametrization could have given a better result.

From the limited amount of modified geometries of carbon-fluorine bonds, this was found to give somewhat better results for fluorinated compounds. This might indicate that if the dataset is to be supplemented by modified geometries, a focus should be on carbon-heteroatom bonds.

Further work on the model should attempt to implement the equations for the dipole moment gradient, which is a local atomic property and may therefore be able to be solved less ambiguously with respect to atomic parameters than the molecular polarizability. This may aid the parametrization of the static polarizability, which has been found to be a source of error.

## References

- [1] A. R. Leach, “Molecular modelling : principles and applications,” 2001.
- [2] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, “Machine learning force fields,” *Chemical Reviews*, vol. 121, no. 16, pp. 10142–10186, 2021.
- [3] J. Morado, P. N. Mortenson, J. W. M. Nissink, J. W. Essex, and C.-K. Skylaris, “Does a machine-learned potential perform better than an optimally tuned traditional force field? a case study on fluorohydrins,” *Journal of Chemical Information and Modeling*, vol. 0, no. 0, p. null, 0. PMID: 37071825.
- [4] M. R. Tucker, S. Piana, D. Tan, M. V. LeVine, and D. E. Shaw, “Development of force field parameters for the simulation of single- and double-stranded dna molecules and dna–protein complexes,” *The Journal of Physical Chemistry B*, vol. 126, no. 24, pp. 4442–4457, 2022. PMID: 35694853.
- [5] D. Vasseti, M. Pagliai, and P. Procacci, “Assessment of gaff2 and opl-aa general force fields in combination with the water models tip3p, spce, and opc3 for the solvation free energy of druglike organic molecules,” *Journal of Chemical Theory and Computation*, vol. 15, no. 3, pp. 1983–1995, 2019.
- [6] L. D. Schuler, X. Daura, and W. F. Van Gunsteren, “An improved gromos96 force field for aliphatic hydrocarbons in the condensed phase,” *Journal of computational chemistry*, vol. 22, no. 11, pp. 1205–1218, 2001.
- [7] S. Kawamoto, H. Liu, Y. Miyazaki, S. Seo, M. Dixit, R. DeVane, C. MacDermaid, G. Fiorin, M. L. Klein, and W. Shinoda, “Spica force field for proteins and peptides,” *Journal of Chemical Theory and Computation*, vol. 18, no. 5, pp. 3204–3217, 2022. PMID: 35413197.
- [8] Y. Qiu, D. G. A. Smith, S. Boothroyd, H. Jang, D. F. Hahn, J. Wagner, C. C. Bannan, T. Gokey, V. T. Lim, C. D. Stern, A. Rizzi, B. Tjanaka, G. Tresadern, X. Lucas, M. R. Shirts, M. K. Gilson, J. D. Chodera, C. I. Bayly, D. L. Mobley, and L.-P. Wang, “Development and benchmarking of open force field v1.0.0—the parsley small-molecule force field,” *Journal of Chemical Theory and Computation*, vol. 17, no. 10, pp. 6262–6280, 2021. PMID: 34551262.
- [9] B. Han, C. M. Isborn, and L. Shi, “Incorporating polarization and charge transfer into a point-charge model for water using machine learning,” *The Journal of Physical Chemistry Letters*, vol. 0, no. 0, pp. 3869–3877, 0. PMID: 37067482.

- [10] S. Cardamone and P. L. A. Popelier, "Prediction of conformationally dependent atomic multipole moments in carbohydrates," *Journal of computational chemistry*, vol. 36, no. 32, pp. 2361–2373, 2015.
- [11] P. P. Poier and F. Jensen, "Describing molecular polarizability by a bond capacity model," *Journal of Chemical Theory and Computation*, vol. 15, no. 5, pp. 3093–3107, 2019. PMID: 30920212.
- [12] A. Mayer, A. L. González, C. M. Aikens, and G. C. Schatz, "A charge–dipole interaction model for the frequency-dependent polarizability of silver clusters," *Nanotechnology*, vol. 20, p. 195204, apr 2009.
- [13] D. M. Wilkins, A. Grisafi, Y. Yang, K. U. Lao, R. A. DiStasio, and M. Ceriotti, "Accurate molecular polarizabilities with coupled cluster theory and machine learning," *Proceedings of the National Academy of Sciences*, vol. 116, no. 9, pp. 3401–3406, 2019.
- [14] D. Bedrov, J.-P. Piquemal, O. Borodin, A. D. MacKerell Jr, B. Roux, and C. Schröder, "Molecular dynamics simulations of ionic liquids and electrolytes using polarizable force fields," *Chemical reviews*, vol. 119, no. 13, pp. 7940–7995, 2019.
- [15] F. Dommert, K. Wendler, R. Berger, L. Delle Site, and C. Holm, "Force fields for studying the structure and dynamics of ionic liquids: A critical review of recent developments," *ChemPhysChem*, vol. 13, no. 7, pp. 1625–1637, 2012.
- [16] B. C. B. Symons and P. L. A. Popelier, "Application of quantum chemical topology force field flux to condensed matter simulations: Liquid water," *Journal of Chemical Theory and Computation*, vol. 18, no. 9, pp. 5577–5588, 2022. PMID: 35939826.
- [17] D. S. Sabirov, "Polarizability as a landmark property for fullerene chemistry and materials science," *RSC Adv.*, vol. 4, pp. 44996–45028, 2014.
- [18] P. Calaminici, J. Carmona-Espindola, G. Geudtner, and A. M. Köster, "Static and dynamic polarizability of c540 fullerene," *International Journal of Quantum Chemistry*, vol. 112, no. 19, pp. 3252–3255, 2012.
- [19] A. Rahnamoun, K. A. O’Hearn, M. C. Kaymak, Z. Li, K. M. J. Merz, and H. M. Aktulga, "A polarizable cationic dummy metal ion model," *The Journal of Physical Chemistry Letters*, vol. 13, no. 23, pp. 5334–5340, 2022. PMID: 35675715.

- [20] J. Simons and P. Jørgensen, “Geometrical derivatives of dipole moments and polarizabilities,” *International Journal of Quantum Chemistry*, vol. 25, no. 6, pp. 1135–1150, 1984.
- [21] D. Jayatilaka, P. E. Maslen, R. D. Amos, and N. C. Handy, “Higher analytic derivatives,” *Molecular Physics*, vol. 75, no. 2, pp. 271–291, 1992.
- [22] Y. Cornaton, M. Ringholm, O. Louant, and K. Ruud, “Analytic calculations of anharmonic infrared and raman vibrational spectra,” *Physical Chemistry Chemical Physics*, vol. 18, no. 5, pp. 4201–4215, 2016.
- [23] M. R. Pederson, T. Baruah, P. B. Allen, and C. Schmidt, “Density-functional-based determination of vibrational polarizabilities in molecules within the double-harmonic approximation: derivation and application,” *Journal of Chemical Theory and Computation*, vol. 1, no. 4, pp. 590–596, 2005. PMID: 26641679.
- [24] S. Haghani, N. Davari, R. Sandnes, and P.-O. Åstrand, “Complex frequency-dependent polarizability through the  $\pi \rightarrow \pi$  excitation energy of azobenzene molecules by a combined charge-transfer and point-dipole interaction model,” *The journal of physical chemistry. A, Molecules, spectroscopy, kinetics, environment, & general theory*, vol. 118, no. 47, pp. 11282–11292, 2014.
- [25] N. Davari, S. Haghani, P.-O. Åstrand, and G. C. Schatz, “Local electric field factors by a combined charge-transfer and point-dipole interaction model,” *RSC Advances*, vol. 5, no. 40, pp. 31594–31605, 2015.
- [26] H. S. Smalø, P.-O. Åstrand, and A. Mayer, “Combined nonmetallic electronegativity equalisation and point-dipole interaction model for the frequency-dependent polarisability,” *Molecular physics*, vol. 111, no. 9-11, pp. 1470–1481, 2013.
- [27] H. S. Smalø, P.-O. Åstrand, and L. Jensen, “Nonmetallic electronegativity equalization and point-dipole interaction model including exchange interactions for molecular dipole moments and polarizabilities,” *The Journal of chemical physics*, vol. 131, no. 4, pp. 044101–044101–19, 2009.
- [28] A. Mayer, P. Lambin, and P.-O. Åstrand, “An electrostatic interaction model for frequency-dependent polarizability: methodology and applications to hydrocarbons and fullerenes,” *Nanotechnology*, vol. 19, no. 2, pp. 025203–025203 (12), 2008.
- [29] Z. Jing, C. Liu, S. Y. Cheng, R. Qi, B. D. Walker, J.-P. Piquemal, and P. Ren, “Polarizable force fields for biomolecular simulations: Recent advances and applications,” *Annual Review of Biophysics*, vol. 48, no. 1, pp. 371–394, 2019. PMID: 30916997.

- [30] C. M. Baker, “Polarizable force fields for molecular dynamics simulations of biomolecules,” *WIREs Computational Molecular Science*, vol. 5, no. 2, pp. 241–254, 2015.
- [31] N. Davari, C. D. Daub, P.-O. Åstrand, and M. Unge, “Local field factors and dielectric properties of liquid benzene,” *The Journal of Physical Chemistry B*, vol. 119, no. 35, pp. 11839–11845, 2015.
- [32] J. V. Vermaas, L. Petridis, J. Ralph, M. F. Crowley, and G. T. Beckham, “Systematic parameterization of lignin for the charmm force field,” *Green Chemistry*, vol. 21, no. 1, pp. 109–122, 2019.
- [33] T. Verstraelen, V. Van Speybroeck, and M. Waroquier, “The electronegativity equalization method and the split charge equilibration applied to organic systems: Parametrization, validation, and comparison,” *The Journal of chemical physics*, vol. 131, no. 4, p. 044127, 2009.
- [34] R. Galvelis, S. Doerr, J. M. Damas, M. J. Harvey, and G. De Fabritiis, “A scalable molecular force field parameterization method based on density functional theory and quantum-level machine learning,” *Journal of Chemical Information and Modeling*, vol. 59, no. 8, pp. 3485–3493, 2019. PMID: 31322877.
- [35] P. Atkins, *Molecular quantum mechanics*. Oxford: Oxford University Press, 5th ed. ed., 2011.
- [36] P.-O. Åstrand, *Atomistic Models: Concepts in Computational Chemistry*. bookboon.com, 2016.
- [37] R. Woolley and B. Sutcliffe, “Molecular structure and the born—oppenheimer approximation,” *Chemical Physics Letters*, vol. 45, no. 2, pp. 393–398, 1977.
- [38] P. Echenique and J. L. Alonso, “A mathematical and computational review of hartree–fock scf methods in quantum chemistry,” *Molecular Physics*, vol. 105, no. 23-24, pp. 3057–3098, 2007.
- [39] L. J. Butler, “Chemical reaction dynamics beyond the born-oppenheimer approximation,” *Annual Review of Physical Chemistry*, vol. 49, no. 1, pp. 125–171, 1998. PMID: 15012427.
- [40] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme, “Best-practice dft protocols for basic molecular computational chemistry,” *Angewandte Chemie International Edition*, vol. 61, no. 42, p. e202205735, 2022.
- [41] P. Geerlings, F. De Proft, and W. Langenaeker, “Conceptual density functional theory,” *Chemical Reviews*, vol. 103, no. 5, pp. 1793–1874, 2003. PMID: 12744694.

- [42] I. Y. Zhang, J. Wu, and X. Xu, “Extending the reliability and applicability of b3lyp,” *Chemical Communications*, vol. 46, no. 18, pp. 3057–3070, 2010.
- [43] S. F. Sousa, P. A. Fernandes, and M. J. Ramos, “General performance of density functionals,” *The Journal of Physical Chemistry A*, vol. 111, no. 42, pp. 10439–10452, 2007.
- [44] G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, and A. Fazzio, “From dft to machine learning: recent approaches to materials science—a review,” *Journal of Physics: Materials*, vol. 2, p. 032001, may 2019.
- [45] C. Fiolhais, F. Nogueira, and M. A. Marques, *A primer in density functional theory*, vol. 620. Springer Science & Business Media, 2003.
- [46] J. Liang, X. Feng, D. Hait, and M. Head-Gordon, “Revisiting the performance of time-dependent density functional theory for electronic excitations: Assessment of 43 popular and recently developed functionals from rungs one to four,” *Journal of Chemical Theory and Computation*, vol. 18, no. 6, pp. 3460–3473, 2022. PMID: 35533317.
- [47] J. P. Perdew and A. Ruzsinszky, “Fourteen easy lessons in density functional theory,” *International Journal of Quantum Chemistry*, vol. 110, no. 15, pp. 2801–2807, 2010.
- [48] K. Burke and L. O. Wagner, “Dft in a nutshell,” *International Journal of Quantum Chemistry*, vol. 113, no. 2, pp. 96–101, 2013.
- [49] T. Yanai, D. P. Tew, and N. C. Handy, “A new hybrid exchange–correlation functional using the coulomb-attenuating method (cam-b3lyp),” *Chemical physics letters*, vol. 393, no. 1, pp. 51–57, 2004.
- [50] G. Santra, R. Calinsky, and J. M. L. Martin, “Benefits of range-separated hybrid and double-hybrid functionals for a large and diverse data set of reaction energies and barrier heights,” *The Journal of Physical Chemistry A*, vol. 126, no. 32, pp. 5492–5505, 2022. PMID: 35930677.
- [51] J. Killingbeck, “Quantum-mechanical perturbation theory,” *Reports on Progress in Physics*, vol. 40, p. 963, sep 1977.
- [52] D. J. Tozer and N. C. Handy, “Improving virtual kohn–sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities,” *The Journal of chemical physics*, vol. 109, no. 23, pp. 10180–10189, 1998.
- [53] W. A. Parkinson and M. C. Zerner, “The calculation of dynamic molecular polarizability,” *The Journal of chemical physics*, vol. 90, no. 10, pp. 5606–5611, 1989.

- [54] P. T. Gee and K. T. Tang, "Error bounds on dynamic polarizability and oscillator strength," *Phys. Rev. A*, vol. 7, pp. 1863–1866, Jun 1973.
- [55] J. A. Lemkul, J. Huang, B. Roux, and A. D. J. MacKerell, "An empirical polarizable force field based on the classical drude oscillator model: Development history and recent applications," *Chemical Reviews*, vol. 116, no. 9, pp. 4983–5013, 2016. PMID: 26815602.
- [56] P. Bultinck, W. Langenaeker, P. Lahorte, F. De Proft, P. Geerlings, M. Waroquier, and J. P. Tollenaere, "The electronegativity equalization method i: Parametrization and validation for atomic charge calculations," *The Journal of Physical Chemistry A*, vol. 106, no. 34, pp. 7887–7894, 2002.
- [57] I. Leven, H. Hao, S. Tan, X. Guan, K. A. Penrod, D. Akbarian, B. Evangelisti, M. J. Hossain, M. M. Islam, J. P. Koski, S. Moore, H. M. Aktulga, A. C. T. van Duin, and T. Head-Gordon, "Recent advances for improving the accuracy, transferability, and efficiency of reactive force fields," *Journal of Chemical Theory and Computation*, vol. 17, no. 6, pp. 3237–3251, 2021. PMID: 33970642.
- [58] R. Chelli, P. Procacci, R. Righini, and S. Califano, "Electrical response in chemical potential equalization schemes," *The Journal of chemical physics*, vol. 111, no. 18, pp. 8569–8575, 1999.
- [59] S. Patel and C. L. B. III, "Fluctuating charge force fields: recent developments and applications from small molecules to macromolecular biological systems," *Molecular Simulation*, vol. 32, no. 3-4, pp. 231–249, 2006.
- [60] A. d. O. Cavalcante, M. C. Ribeiro, and M. S. Skaf, "Polarizability effects on the structure and dynamics of ionic liquids," *The Journal of Chemical Physics*, vol. 140, no. 14, p. 144108, 2014.
- [61] K. S. Smirnov, "Assessment of split-charge equilibration model for development of polarizable force fields," *Modelling and Simulation in Materials Science and Engineering*, vol. 23, p. 074006, sep 2015.
- [62] A. K. Das, O. N. Demerdash, and T. Head-Gordon, "Improvements to the amoeba force field by introducing anisotropic atomic polarizability of the water molecule," *Journal of Chemical Theory and Computation*, vol. 14, no. 12, pp. 6722–6733, 2018. PMID: 30428257.
- [63] T. A. Halgren and W. Damm, "Polarizable force fields," *Current Opinion in Structural Biology*, vol. 11, no. 2, pp. 236–242, 2001.
- [64] J. Applequist, J. R. Carl, and K.-K. Fung, "Atom dipole interaction model for molecular polarizability. application to polyatomic molecules and determination

- of atom polarizabilities,” *Journal of the American Chemical Society*, vol. 94, no. 9, pp. 2952–2960, 1972.
- [65] L. Jensen, M. Swart, P. T. Van Duijnen, and J. G. Snijders, “Medium perturbations on the molecular polarizability calculated within a localized dipole interaction model,” *The Journal of chemical physics*, vol. 117, no. 7, pp. 3316–3320, 2002.
- [66] L. Jensen, P.-O. Åstrand, K. O. Sylvester-Hvid, and K. V. Mikkelsen, “Frequency-dependent molecular polarizability calculated within an interaction model,” *The Journal of Physical Chemistry A*, vol. 104, no. 7, pp. 1563–1569, 2000.
- [67] R. R. Birge, “Calculation of molecular polarizabilities using an anisotropic atom point dipole interaction model which includes the effect of electron repulsion,” *The Journal of chemical physics*, vol. 72, no. 10, pp. 5312–5319, 1980.
- [68] H. Wang and W. Yang, “Determining polarizable force fields with electrostatic potentials from quantum mechanical linear response theory,” *The Journal of Chemical Physics*, vol. 144, no. 22, p. 224107, 2016.
- [69] G. A. Kaminski, R. A. Friesner, and R. Zhou, “A computationally inexpensive modification of the point dipole electrostatic polarization model for molecular simulations,” *Journal of Computational Chemistry*, vol. 24, no. 3, pp. 267–276, 2003.
- [70] J. Huang, P. E. M. Lopes, B. Roux, and A. D. J. MacKerell, “Recent advances in polarizable force fields for macromolecules: Microsecond simulations of proteins using the classical drude oscillator model,” *The Journal of Physical Chemistry Letters*, vol. 5, no. 18, pp. 3144–3150, 2014. PMID: 25247054.
- [71] G. Lamoureux and B. Roux, “Modeling induced polarization with classical drude oscillators: Theory and molecular dynamics simulation algorithm,” *The Journal of chemical physics*, vol. 119, no. 6, pp. 3025–3039, 2003.
- [72] M. M. Ghahremanpour, P. J. van Maaren, C. Caleman, G. R. Hutchison, and D. van der Spoel, “Polarizable drude model with s-type gaussian or slater charge density for general molecular mechanics force fields,” *Journal of Chemical Theory and Computation*, vol. 14, no. 11, pp. 5553–5566, 2018. PMID: 30281307.
- [73] A. Savelyev and A. D. MacKerell Jr., “All-atom polarizable force field for dna based on the classical drude oscillator model,” *Journal of Computational Chemistry*, vol. 35, no. 16, pp. 1219–1239, 2014.
- [74] V. M. Anisimov, I. V. Vorobyov, B. Roux, and A. D. MacKerell, “Polarizable empirical force field for the primary and secondary alcohol series based on the



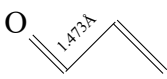
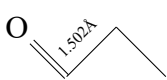

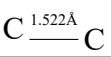
- classical drude model,” *Journal of Chemical Theory and Computation*, vol. 3, no. 6, pp. 1927–1946, 2007. PMID: 18802495.
- [75] D. Dubbeldam, K. S. Walton, T. J. H. Vlugt, and S. Calero, “Design, parameterization, and implementation of atomic force fields for adsorption in nanoporous materials,” *Advanced Theory and Simulations*, vol. 2, no. 11, p. 1900135, 2019.
- [76] L. Jensen, P. O. Astrand, A. Osted, J. Kongsted, and K. V. Mikkelsen, “Polarizability of molecular clusters as calculated by a dipole interaction model,” *The Journal of chemical physics*, vol. 116, no. 10, pp. 4001–4010, 2002.
- [77] R. A. Adams, “Calculus : a complete course,” 2018.
- [78] R. G. Parr and R. G. Pearson, “Absolute hardness: companion parameter to absolute electronegativity,” *Journal of the American chemical society*, vol. 105, no. 26, pp. 7512–7516, 1983.
- [79] J. L. Reed, “Electronegativity: chemical hardness i,” *The Journal of Physical Chemistry A*, vol. 101, no. 40, pp. 7396–7400, 1997.
- [80] C. Cárdenas, P. Ayers, F. De Proft, D. J. Tozer, and P. Geerlings, “Should negative electron affinities be used for evaluating the chemical hardness?,” *Physical Chemistry Chemical Physics*, vol. 13, no. 6, pp. 2285–2293, 2011.
- [81] G. B. Arfken, H. J. Weber, and F. E. Harris, *Mathematical Methods for Physicists: A Comprehensive Guide*. San Diego: Elsevier Science & Technology, 2011.
- [82] D. J. Griffiths, “Introduction to electrodynamics,” 2017.
- [83] P. Atkins, “Atkins’ physical chemistry,” 2018.
- [84] K. HUNT, “Relationships between electric-field shielding tensors and infrared or raman intensities - an explanation based on nonlocal polarizability densities,” *The Journal of chemical physics*, vol. 90, no. 9, pp. 4909–4915, 1989.
- [85] K. Hunt, Y. Liang, R. Nimalakirithi, and R. Harris, “Changes in electronic polarizability densities due to shifts in nuclear positions, and a new interpretation for integrated intensities of vibrational raman bands,” *The Journal of chemical physics*, vol. 91, no. 9, pp. 5251–5254, 1989.
- [86] P. Fowler and A. Buckingham, “Relation of the electric field at a nucleus to other molecular properties,” *Chemical physics*, vol. 98, no. 2, pp. 167–178, 1985.
- [87] R. P. Feynman, “Forces in molecules,” *Physical review*, vol. 56, no. 4, p. 340, 1939.

- [88] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, “Avogadro: an advanced semantic chemical editor, visualization, and analysis platform,” *Journal of cheminformatics*, vol. 4, no. 1, pp. 1–17, 2012.
- [89] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. I. Goddard, and W. M. Skiff, “Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations,” *Journal of the American Chemical Society*, vol. 114, no. 25, pp. 10024–10035, 1992.
- [90] T. A. Halgren, “Mmff vi. mmff94s option for energy minimization studies,” *Journal of Computational Chemistry*, vol. 20, no. 7, pp. 720–729, 1999.
- [91] 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. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. Aa. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, 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. Pawłowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Snegov, 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. D. Wilson, M. Ziolkowski, and H. Ågren, “The Dalton quantum chemistry program system,” *WIREs Comput. Mol. Sci.*, vol. 4, no. 3, pp. 269–284, 2014.
- [92] F. Jensen, “Polarization consistent basis sets: Principles (vol 115, pg 9113, 2001),” *The Journal of chemical physics*, vol. 116, no. 8, pp. 3502–3502, 2002.
- [93] 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,” *Journal of physical chemistry (1952)*, vol. 98, no. 45, pp. 11623–11627, 1994.
- [94] P. A. Limacher, K. V. Mikkelsen, and H. P. Lüthi, “On the accurate calculation of polarizabilities and second hyperpolarizabilities of polyacetylene oligomer chains using the cam-b3lyp density functional,” *The Journal of chemical physics*, vol. 130, no. 19, pp. 194114–194114–7, 2009.
- [95] T. H. Dunning, “Gaussian-basis sets for use in correlated molecular calculations .1. the atoms boron through neon and hydrogen,” *The Journal of chemical physics*,

- vol. 90, no. 2, pp. 1007–1023, 1989.
- [96] R. A. Kendall, T. H. Dunning Jr, and R. J. Harrison, “Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions,” *The Journal of chemical physics*, vol. 96, no. 9, pp. 6796–6806, 1992.
- [97] F. Jensen, “Polarization consistent basis sets. iii. the importance of diffuse functions,” *The Journal of chemical physics*, vol. 117, no. 20, pp. 9234–9240, 2002.
- [98] H. Nabli, “An overview on the simplex algorithm,” *Applied Mathematics and Computation*, vol. 210, no. 2, pp. 479–489, 2009.
- [99] R. Faller, H. Schmitz, O. Biermann, and F. Müller-Plathe, “Automatic parameterization of force fields for liquids by simplex optimization,” *Journal of Computational Chemistry*, vol. 20, no. 10, pp. 1009–1017, 1999.
- [100] R. M. Betz and R. C. Walker, “Paramfit: Automated optimization of force field parameters for molecular dynamics simulations,” *Journal of Computational Chemistry*, vol. 36, no. 2, pp. 79–87, 2015.
- [101] M. Wan, J. Song, W. Li, L. Gao, and W. Fang, “Development of coarse-grained force field by combining multilinear interpolation technique and simplex algorithm,” *Journal of Computational Chemistry*, vol. 41, no. 8, pp. 814–829, 2020.

## Appendix A Molecules used in parametrization

Table A.1: Molecules in the dataset. For stretched molecules the stretched bonds are shown by writing explicitly the unmodified bond length directly above the bond.

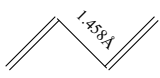
Molecule	Stretch [mÅ]
Alcohols	
Cyclopentanol	0
Benzenol	0
Methanol	0
Propanol	0
2-butanol	0
Propan-2-ol	0
Ethanol	0
Butanol	0
Hex-3-ol	0
Aldehyde	
1-butanal	0
Prop-2-en-1-al	0, 6, 12, 18, 24, 30, 36, 42, 48
	
Ethanal	0
1-hexanal	0
Propanal	-10, -5, 0
	
Pent-2e4-dien-1-al	0
Hex-2e4-en-1-al	0
But-2e-en-1-al	-18, -14, -11, -7, -4, 0, 4, 7, 11, 14, 18
	
1-pentanal	0
Alkanes	
Methane	0
Ethane	0, 8, 16, 25, 33
	

Continues on next page

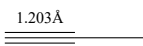

Continuation of table A.1

Propane	0
Butane	0
Pentane	0
Hexane	0
Heptane	0
Octane	0
Nonane	0
Decane	0
Dodecane	0
Undecane	0

Alkenes

E-but-2-ene	0
Butene	0
Buta-1,3-diene	-20, -16, -12, -8, -4, 0
	
Propa-1,2-diene	0
2,3-dimethylbut-2-ene	0
E-hex-3-ene	0
Z-but-2-ene	0
Ethene	0
Pent-2-ene	0
Z-hex-3-ene	0

Alkynes

Cyclopentyne	0
Pent-2-yne	0
Propyne	0, 33, 66, 99, 131, 164, 197, 230
	
Ethyne	0
Hex-3-yne	0
But-2-yne	-11, -6, 0, 3, 7, 10, 14, 17, 20, 24, 27, 31, 34
	

Amines

2-aminopentane	0
----------------	---

Continues on next page

---

Continuation of table A.1

---

1R-phenylethane-1-amine	0
NN-dimethylamino-ethane	0
Ethaneamine	0
2S-amino-3-phenylpropanoic-acid	0
2-N-methylamino-propan-1-ol	0
Propan-1-amine	0
N-methylamino-ethane	0
N-ethylethamine	0
Prop-1-en-2-amine	0
2-aminoethan-1-ol	0
3-aminopropanoic-acid	0
2-propanamine	0
2-n-methylamino-propane	0

---

Branched alkanes

---

2-methylpropane	0
2-methylbutane	0
2,2-dimethylpropane	0
2,2-dimethylbutane	0
1,2-dimethylcyclopentane	0

---

Carboxylic acids

---

Butanoic-acid	0
Methanoic-acid	0
Ethanoic-acid	0
Propanoic-acid	0
Butanedioic-acid	0
Ethanedioic-acid	0
Propanedioic-acid	0
But-2-enedioic-acid	0

---

Chlorinated

---

2-chloropentane	0
2-chloropropane	0
Chlorocyclopropane	0
3-chloro-prop-1-en	0
3-chloropropanol	0

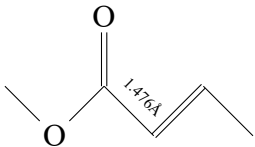
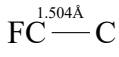
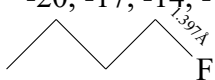
---

Continues on next page

---

Continuation of table A.1	
2-chlorobutane	0
Chloromethane	0
Chloroethane	0
Chlorobutane	0
Chloropropane	0
Cyclic alkanes	
Cyclopropane	0
Cyclobutane	0
Cyclopentane	0
Cyclic alkenes	
Cyclopentene	0
1,2-dimethylcyclopentene	0
1,2-dimethylcyclobutene	0
2,4-cyclohexadiene	0
Cyclobutene	0
Cyclohexene	0
Dichlorinated	
1,2-dichloroethane	0
1,3-dichloro-prop-1-en	0
Difluorated	
Difluoroethyne	0
1,2-difluoroethane	0
1,1-difluoroethane	-15, -12, -10, -8, -5, -2, 0, 5, 10, 15, 20, 25, 30
	$\begin{array}{c} \text{F}_2\text{C} \\ \swarrow \text{1.496\AA} \\ \text{C} \end{array}$
2,2-difluoropentane	0
Esters	
Ethyl-ethanoate	0
Methyl-prop-2e-enoate	0
Ethenyl-prop-2e-enoate	0
Continues on next page	

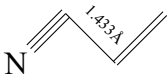
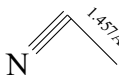

Continuation of table A.1

Methyl-but-2e-enoate		-20, -10, 0, 5, 10, 15, 20
		
Ethenyl-but-2e-enoate		0
Methyl-propanoate		0
Methyl-butanoate		0
Methyl-ethanoate		0
Ethenyl-ethanoate		0
Ethers		
Methoxymethane		0
Methoxyethane		0
Methoxyethene		0
Ethoxyethane		0
3-methoxyprop-1-ene		0
Fluorinated		
Fluorocyclopropane		0
1-fluoro-prop-1-yne		0
Fluoroethane		0, 15, 30, 45, 60
		
Fluorobutane		0
2-fluorobutane		0
2-fluoropropane		0
Fluoropropane		-20, -17, -14, -11, -9, -6, -3, 0, 3, 6, 9, 11, 14, 17, 20
		
Hetrocycles		
Benzene		0
Toluene		0
Pyridine		0
Cyclopentadiene		0
Cis-butadiene		0
Tetrahydrofuran		0
Furane		0

Continues on next page

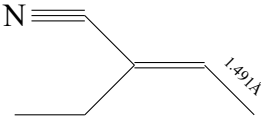
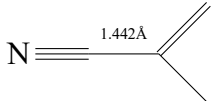


Continuation of table A.1

Tetrahydrothiophene		0
Thiophene		0
Tetrahydropyrrole		0
Pyrrole		0
Hydroxy groups		
11-dihydroxyprop-2-ene		0
22-dihydroxypropane		0
11-dihydroxypent-2e4e-diene		0
11-dihydroxybut-2e-ene		0
11-hydroxybut-2e-ene		0
22-dihydroxybutane		0
1144-tetrahydroxybut-2e-ene		0
33-dihydroxybut-1-ene		0
11-dihydroxypent-2e-ene		0
Ketone		
2-butanone		0
Cyclopentanone		0
Cyclopropanone		0
Propanone		0
2-pentanone		0
Nitriles		
3-cyano-penta-1,3-diene		0
Prop-2-ene-1-nitrile		0, 10, 20, 30, 40, 50, 60, 70
		
Ethanenitrile		-11, -6, 0, 6, 12, 19, 25, 31
		
Propane-1-nitrile		0
But-2-yne-1-nitrile		0, 8, 17, 25, 33, 42, 50, 58, 66, 75, 83
		

Continues on next page

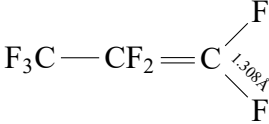
Continuation of table A.1

2-ethylbut-2-ene-1-nitrile		-20, -10, 0, 12, 24, 36, 48, 60
2-methylprop-2-ene-1-nitrile		0, 20, 40, 60, 80

Partially fluorinated polyenes


2-trifluoromethyl-1113-tetrafluoro-octa-2e4e6e-triene		0
2-trifluoromethyl-1113-tetrafluoro-hepta-2e4e-diene		0
2-trifluoromethyl-1113-tetrafluoro-pent-2e-ene		0
2-trifluoromethyl-1113-tetrafluoro-nona-2e4e6e8-tetraene		0
2-trifluoromethyl-1113-tetrafluoro-hepta-2e4e6-triene		0
2-trifluoromethyl-1113-tetrafluoro-deca-2e4e6e8e-tetraene		0
2-trifluoromethyl-1113-tetrafluoro-nona-2e4e6e-triene		0
2-trifluoromethyl-1113-tetrafluoro-hexa-2e4e-diene		0
2-trifluoromethyl-1113-tetrafluoro-but-2e-ene		0

Perfluorated

Tetrafluoroethene		0
Hexafluoroethane		0
Perfluoroprop-1-ene		0, 2, 5, 7, 9, 12, 14, 16, 18, 21, 23, 25, 28, 30
Perfluoro-but-2-ene		0

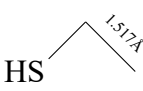
Continues on next page

Continuation of table A.1

Perfluoro polyenes		
Perfluoro-buta-1,3-diene		0
Perfluoro-penta-1,3e-diene		0
Perfluoro-hepta-1,3e5e-triene		0
Perfluoro-octa-1,3e5e7-tetraene		0
Perfluoro-nona-1,3e5e7e-tetraene		0
Perfluoro-deca-1,3e5e7e9-pentaene		0
Perfluoro-undeca-1,3e5e7e9e-pentaene		0
Perfluoro-dodeca-1,3e5e7e9e10-pentaene		0
Phenyl		
1s-phenylethan-1-ol		0
2-phenylpentane		0
Phenylethen-2-ol		0
Phenylmethanol		0
Phenylethene		0
Polyenes		
Penta-1,3e-diene		0, 15, 30
Hexa-1,3e5-triene		0
Hepta-1,3e5e-triene		0
Octa-1,3e5e7-tetraene		0
Nona-1,3e5e7e-tetraene		0
Deca-1,3e5e7e9-pentaene		0
Undeca-1,3e5e7e9e-pentaene		0
Dodeca-1,3e5e7e9e10-pentaene		0
Thioaldehydes		
Prop-1-en-3-thial		0
Propane-1-thial		0
Ethanethial		0
Methanethial		0

Continues on next page

Continuation of table A.1

Thioethers	
Ethyl-methyl-sulfide	0
Ethylsulfanylethane	0
1-methylsulfanylpropane	0
Methylsulfanylmethane	0
3-methylsulfanylprop-1-ene	0
Propanylsulfanylpropane	0
Methylsulfanylethene	0
Thioketone	
Propane-2-thione	0
But-1-yne-2-thione	0
But-1-ene-2-thione	0
Butan-2-thione	0
Thiols	
Methanethiol	0
But-1-en-2-thiol	0
Prop-1-en-2-thiol	0
Butan-2-thiol	0
Propan-1-thiol	0
Ethanethiol	0, 20, 40, 60, 80
	
Prop-1-yn-3-thiol	0
Propan-2-thiol	0
But-1-yn-3-thiol	0
Trichlorinated	
111-trichlorobutane	0

Continues on next page

---



---

Continuation of table A.1

---



---

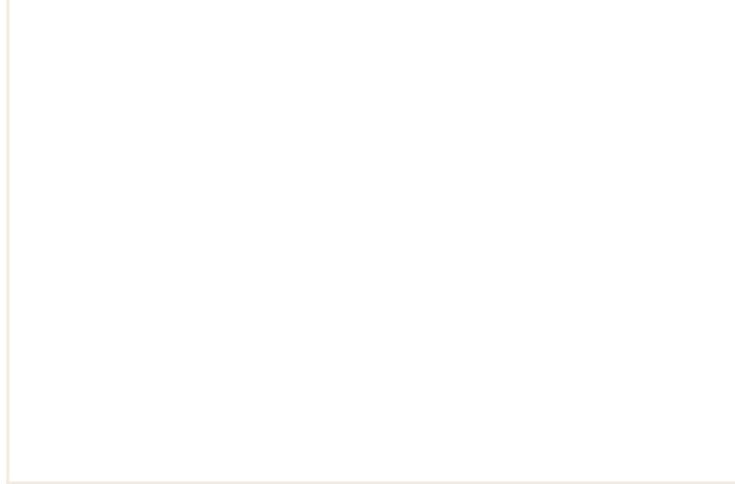
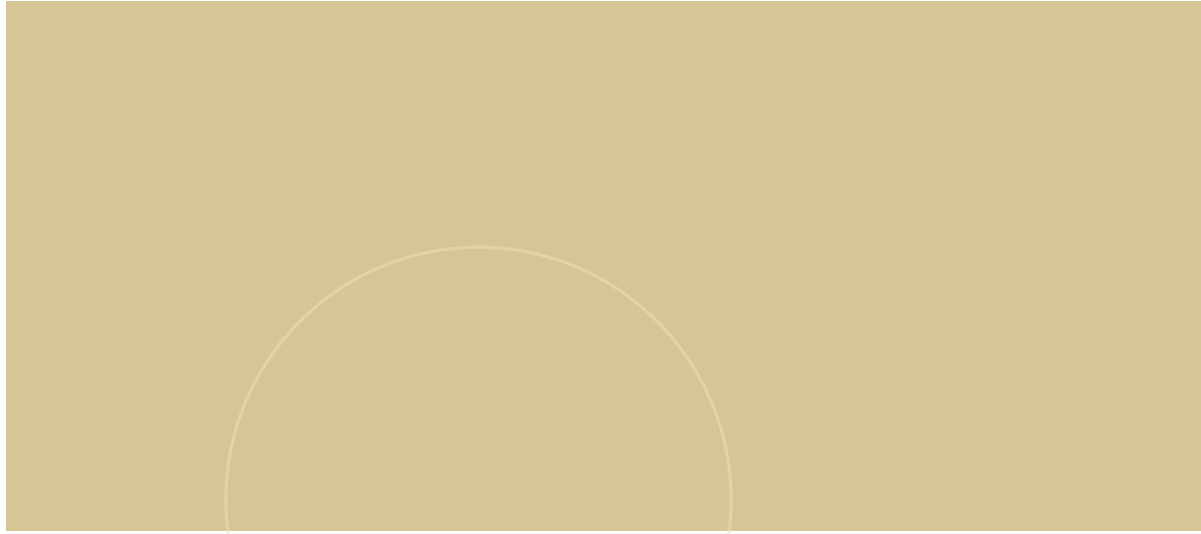
Trifluorinated		
111-trifluoroethane		0, 17, 33, 50, 67, 83, 100
	$\text{F}_3\text{C} \overset{1.495\text{\AA}}{\text{---}} \text{CH}_3$	
111-trifluorobutane		0
222-trifluoro-1-phenylethane		0
112-trifluoroprop-1-ene		0
Trifluoroethene		0

---

Table A.2: Molecules in the smaller parametrization.

Molecule	Stretch [mÅ]
Small parametrization	
N-hydroxy-propan-2-imine	0
Ethaneperoxoic-acid	0
Diazomethane	0
Dimethylsulfone	0
Methanesulfinic-acid	0
Nn-difluoroaminoethane	0
Chloroxy-ethane	0
Nitrosoethene	0
Difluoroxyperfluoromethane	0
Fluoroxymethane	0
Ethaneperoxol	0
Ethyl-methyl-disulfide	0
Ethanesulfinic-acid	0
Diazoethane	0

---



 **NTNU**

Norwegian University of  
Science and Technology