# DISSERTATION

Defence held on 06/11/2023 in Luxembourg

to obtain the degree of

## DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

## EN PHYSIQUE

by

## Szabolcs GÓGER
Born on 18 November 1994 in Sopron (Hungary)

## DEVELOPMENT OF PRACTICAL NON-LOCAL MANY-BODY POLARIZATION FUNCTIONALS

## Dissertation defence committee

Dr Alexandre Tkatchenko, dissertation supervisor
*Professor, Université du Luxembourg*

Dr Massimiliano Esposito, Chair
*Professor, Université du Luxembourg*

Dr Jorge Íñiguez, Vice-Chair
*Professor, Luxembourg Institute of Science and Technology*

Dr Toon Verstraelen
*Professor, Ghent University*

Dr Katharine L. C. Hunt
*Professor, Michigan State University*

# Abstract

Electronic structure calculations can now achieve the highly coveted chemical accuracy (less than 1 kcal/mol average error in energy differences) for molecules with a few dozens of atoms; however, extending these approaches to larger systems is an area of active research. A major challenge is devising methods that are widely applicable to both molecules and materials. To achieve this goal, computational advances must be coupled to a deep understanding of the related physical principles. Response functions and, in particular, polarizability, play a central role in our conceptual understanding of both electron correlation and polarization/dispersion interactions – quantum mechanical effects that are notably hard to properly capture due to the underlying non-local nature of the quantities needed to compute these interactions.

To develop a practical formalism for non-local polarizability, one first needs to deeply elaborate the corresponding local and semi-local approaches. This can be achieved by studying model systems, atoms, and molecules, as the numerical results for molecular systems can be complemented by the physical understanding from the model results. By analyzing quantum systems ranging from model Hamiltonians to real molecules, in this work it is shown that polarizability can be factored into a spectrum-dependent and geometry-dependent part. Notably, the geometry-dependent part influences polarizability by a four-dimensional scaling law, enabling the proper description of response properties of individual atoms within molecules. A novel parametrization for representing the response of atoms by an effective harmonic oscillator model is also introduced, showing that spatially resolved polarization potentials can be predicted using just integrated dipolar properties of atoms.

Moving from model systems and atoms to molecules, it is found that the corresponding polarizability and HOMO–LUMO (highest occupied molecular orbital – lowest unoccupied molecular orbital) gap are independent. In parallel, the theoretical foundations of non-local polarizability are examined, presenting expressions for a range of model systems *via* the polarization field correlation function. By following a rigorous derivation of this response function, it is shown that not only existing methods can be obtained from it as limiting cases, but the design of a general non-expanded many-body dispersion energy functional is also feasible.

Overall, this thesis aims to show that combining the fundamental physics of model systems, atoms, and molecules with a theory of non-local polarizability can lead to practical functionals for electronic structure calculations based on the advanced non-local description of response functions.

# Preface

This thesis is partly based on the following publications

– Szabó, P.; **Góger, S.**; Charry, J.; Karimpour, M. R.; Fedorov, D. V.; Tkatchenko, A. Four-Dimensional Scaling of Dipole Polarizability in Quantum Systems. *Physical Review Letters* **2022**, *128*, 070602.

– **Góger, S.**; Khabibrakhmanov, A.; Vaccarelli, O.; Fedorov, D. V.; Tkatchenko, A. Optimized Quantum Drude Oscillators for Atomic and Molecular Response Properties. *The Journal of Physical Chemistry Letters* **2023**, *14*, 6217–6223.

– **Góger, S.**, Medrano Sandonas, L., Müller, C., Tkatchenko, A. Data-Driven Tailoring of Molecular Dipole Polarizability and Frontier Orbital Energies in Chemical Compound Space. *Physical Chemistry Chemical Physics* **2023**, *25,* 22211–22222.

In addition to the novel results published in the aforementioned papers, a significant portion of my work was involved with method development in larger *ab initio* codebases. While this aspect is not extensively discussed in the thesis, the following publication contains the technical details of the library in the center of these efforts, libMBD

– Hermann, J., Stöhr, M., **Góger, S.**, Chaudhuri, S., Aradi, B., Maurer, R. J., Tkatchenko, A. libMBD: A general-purpose package for scalable quantum many-body dispersion calculations. *The Journal of Chemical Physics* **2023**, *159,* 174802.

libMBD was also used in obtaining most computational results presented in the thesis.

# Notations

$r, r', r_a, \ldots$     coordinates defined in real space

$\int_V \ldots \mathrm{d}^3 r$     integration in 3D coordinates

$r_i, r_{i,a}, \ldots$     coordinate of the $i$th particle in a many-body wavefunction

$\int \ldots \mathrm{d} r_N$     integration in wavefunction many-electron coordinates

$\chi(r, r', \omega)$     frequency-dependent non-local electric susceptibility

$\alpha(r, r', \omega)$     frequency-dependent non-local electric polarizability tensor

$\alpha_\ell$     mean static $\ell$th multipole polarizability

$\mathbf{T}(r, r')$     dipole-dipole coupling tensor

$\phi(r)$     electrostatic potential

$E(r)$     electric field

$P(r)$     polarization density field

$\rho(r)$     electron density

$\hat{d}$     dipole moment operator

$\hat{\Theta}$     quadrupole moment operator

$\xi^{(n)}_{\alpha\beta\ldots\omega}$     $n$th Cartesian multipole moment

$q_{\ell,m}$     spherical multipole moment

$Y_{\ell,m}(\theta, \phi)$     spherical harmonics

$R_{vdw}$     van der Waals radius

$\Gamma_{\substack{\mu\omega \\ \sigma\gamma}}$     reduced two-electron density matrix

$e_\alpha$     unit vector of the $\alpha$ component of a coordinate system

$\{q, \omega, \mu\}$     charge, frequency and reduced mass of the harmonic oscillator model

# Contents

# 1. Introduction

The fields of theoretical chemistry and molecular physics prompt two approaches to understanding existing and predicting new phenomena: connecting observable properties to fundamental laws of physics by theoretical derivations or, alternatively, calculating (and interpolating) a large number of quantum calculations of molecular properties. Historically, equations were laid out first and were implemented as efficient algorithms only later. Nowadays, these two approaches work hand in hand, and it is often numerical simulations that show areas where theoretical developments are needed.

One area where the synergy between theory and computation is driving the frontiers is the description of intermolecular interactions. Despite the significant impact of these interactions on the macroscopic properties of molecules and matter, theories suitable for tractable algorithmic implementations have only begun to be developed within the last two decades [1]. More recently, response functions, especially dipole polarizability, have been attracting attention because of a clear connection to not only non-covalent interactions but also many-electron correlation energy in general. To date, however, no unified polarizability functional has been found that would generally be applicable to atoms, molecules, and materials with a manageable computational cost [2].

Motivated by classical electrodynamics, the macroscopic model for polarizability smears out the quantum mechanical structure of the response, leading to only a local "point" polarizability. However, even a homogeneous electric field can induce a response in quantum systems that is non-local, anisotropic, and contains contributions from higher order multipole moments. Therefore, to devise a general functional based on correct physical principles, knowledge of the non-local polarizability tensor field is required. This thesis presents advances towards such a method by examining the properties of the polarizability of model systems, atoms, and molecules, and accounting for the general non-local nature of this response function.

**Polarizability and Interactions of Molecules**

The dipole polarizability refers to the tendency of a system to form dipoles to an external perturbing homogeneous electric field. This concept first arose in macroscopic electrodynamics; therefore, the polarizability of atoms and molecules was discussed even before the development of quantum

1

mechanics. The first polarizability measurements were conducted in the 19th century, based on the observation that capacitors made from different materials have different capacitances. The fundamental quantity describing this behavior, the relative permittivity $\varepsilon_r$, is connected to the molar polarization $P_M$ of the matter, i.e. the density of dipole moments. This dipole moment density is not only due to the permanent dipole moments $d$ of the molecules, but also induced by the external electronic field, dictated by the polarizability $\alpha$. This process – at a given temperature $T$ – is described by the Debye equation [3] ($N_A$ is Avogadro's constant)

$$\frac{\varepsilon_r - 1}{\varepsilon_r + 2} = \frac{\rho P_M}{M} \; ;$$

$$P_M = \frac{N_A}{3\varepsilon_0} \left( \alpha + \frac{d^2}{3kT} \right) \; .$$

(1.1)

Since relative permittivity and refractive index $n_r$ at a given wavelength are connected by $n_r^2 = \varepsilon_r$, polarizability also determines the refractive index, therefore polarizability can be determined not only *via* capacitance, but also by measuring the refraction of light.

Polarizability drives not only the interaction of molecules with external electromagnetic fields but also intermolecular interactions, since these can be understood by considering the molecules in the field of each other. Dispersion interaction between two atoms, for example, can be expressed within the dipole limit at a distance $R$ by integrating the polarizability over the imaginary frequency $\nu$ [4]

$$E_{\text{disp}} = -\frac{3\hbar}{(4\pi\varepsilon_0)^2 \pi R^6} \int_0^\infty \alpha^A(i\nu)\alpha^B(i\nu) \; \mathrm{d}\nu \; .$$

(1.2)

Textbooks introducing intermolecular interactions tend to emphasize their weak nature by comparing the binding energy of a pair of atoms that interact through dispersion interaction or covalent bonding [5]. This comparison – contrasting a few hundred kJ/mol covalent binding energy with just a few kJ/mol from the dispersion interaction – is not helpful, since it is now known that the total contribution of long-range interactions within extended systems is often comparable, or even exceeds, the energy of covalent bonds [6]. In fact, many observable properties of matter stem from long-range interactions; the existence of molecular solids or liquids is a proof of the strength of these forces. Therefore, it is essential for accurate simulation methods to account for intermolecular interactions, and polarizability is a key quantity in achieving this.

## Non-Additive and Multipolar Effects

The equations for intermolecular interactions (for example, the dispersion energy of Eq. (1.2) and the well-known Lennard-Jones potential) are commonly written for two interacting particles. However, it is known that long-range forces are not pairwise additive, since the presence of the environment modifies the interactions.
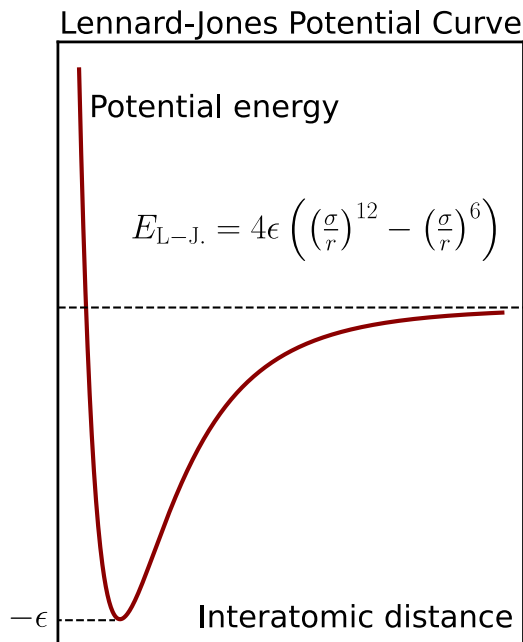


Lennard-Jones Potential Curve

Potential energy

$$E_{\mathrm{L-J.}} = 4\epsilon \left( \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right)$$

Interatomic distance

$-\epsilon$

Figure 1.1: The Lennard-Jones potential; arguably the best known long-range interatomic potential. The two parameters describing this potential are the particle size $\sigma$ (where the potential energy is zero) and the binding energy $\epsilon$. The equilibrium distance is given as $r_{\mathrm{eq.}} = 2^{1/6}\sigma$.

The physical effects behind the non-additive contributions to dispersion energy were categorized by Dobson: "type A" non-additivity refers to the effect of covalent bonding on the dispersion parameters of atoms; "type B" describes the screening of the interaction between the two partners by the presence of other particles; and "type C" non-additivity describes the effect of electron hopping between the pre-defined interaction centers. Most dispersion models can account for "type A" non-additivity, but "type B" and "type C" effects require explicit many-body models [7].

Another major challenge in electronic structure calculations is a seamless connection between inter- and intramolecular interactions. The reason for this difficulty is that most intermolecular models are based on the multipole expansion of the Coulomb interaction, assuming that the contributions due to higher order multipole moments and polarizabilities are diminishing [8]. The use of the dipole limit is justified for large intermolecular distances, but higher orders need to be taken into account in mid to short distances. However, accurate data for multipolar polarizabilities and dispersion

coefficients are scarce, limiting the practical applicability of the multipole expansion. Moreover, the multipole expansion suffers from convergence issues when the interparticle separation is not too large. A method not based on this expansion would be able to seamlessly connect inter- and intramolecular interactions, leading to more general and less empirical approaches for describing molecular interactions.

**Polarizability and Correlation Energy**

The mean-field approximation, where each electron is subject to an effective field from all other electrons, plays a central role in modern electronic-structure methods. In this approximation, we solve the $N$ one-electron Schrödinger equations instead of the full $N$-electron Schrödinger equation, taking the effect of the environment of all other electrons into account in a mean-field way. The Hartree-Fock theory is the most influential approach formalizing this idea, approximating the full $N$-electron solution by a Slater determinant of one-electron orbitals, thus ensuring that the Pauli principle is fulfilled. This method introduces key concepts such as exchange energy, defined as the energy difference between exchanging two electrons of different spin-orbitals having the same spin in the Slater determinant; and correlation energy, defined as the difference between the Hartree-Fock and the full $N$-electron energy. The hierarchy of quantum chemical methods is based on how much of the correlation energy is captured by each approach. For example, a density functional capable of treating the correlation energy exactly is often referred to as the holy grail of quantum chemistry.

Expressing the correlation by systematically increasing the sophistication in the wavefunction is a successful path in theoretical chemistry, but a different approach is taken in solid state physics. This alternative path uses response functions to express the electron correlation, similar to the use of response functions for the interaction between separate molecules. The theoretical formulation of this approach is the adiabatic connection fluctuation-dissipation theorem (ACFDT), which, in its most general form, expressed statistical correlations *within* a system through the response function of the same system to an *external* perturbation [9]. However, these approaches suffer from either an extremely large computational cost or approximations limiting their transferability. Although most ACFDT-based methods use susceptibility as the central response function, more recently, polarizability has been proposed as a viable alternative due to the fact that non-local polarizability is an extensive quantity – opposed to the non-local susceptibility –, prompting the use of coarse-grained approaches [10] and allowing the circumvention of excessive computational costs.
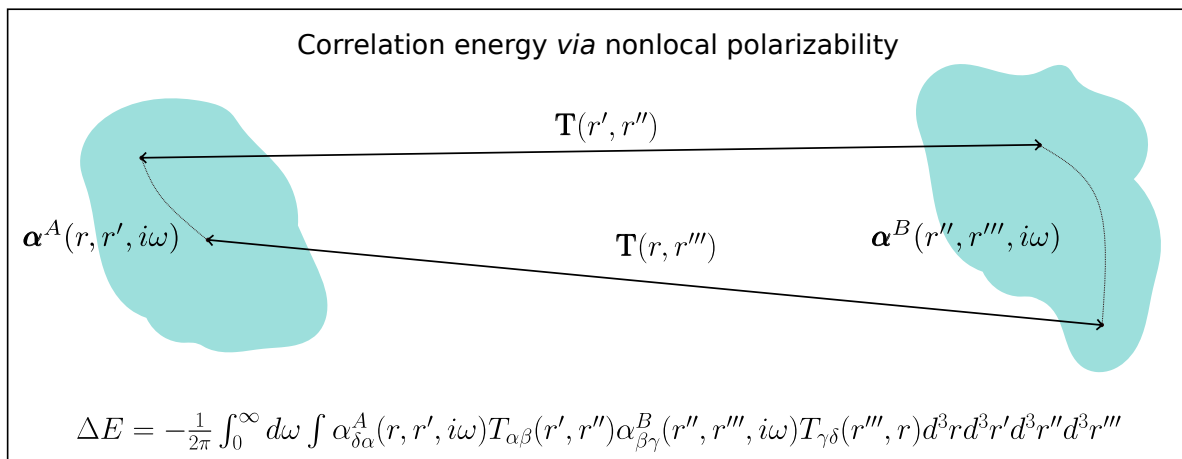
Figure 1.2: Interaction energy $\Delta E$ between *A* and *B* - the correlations within each structure are captured by the non-local polarizability $\boldsymbol{\alpha}$, the interaction between structures is expressed by the dipole-dipole coupling $\mathbf{T}$ of these polarizabilities. Note that in the most general case, the $\mathbf{T}$ tensor can also depend on frequency.

**The Scope of this Thesis**

Because of the central importance of polarizability in describing electron correlation and inter-molecular interactions, a multitude of models have been constructed in the past 150 years. Building on simple quantum mechanical model systems and tabulated atomic contribution to the molecular polarizability, the sophistication of the description can be systematically improved by accounting for the chemical environment of the atoms, interactions between them, as well as the semi-local and, eventually, the non-local nature of the response function. Coupled with modern electronic structure methods, many of these approaches are of practical importance in modeling molecules and materials. Consequently, the trade-off between physical sophistication and computational efficiency is a key question attracting contemporary research interest. While, in *theory,* the full non-local polarizability would enable us to perform accurate simulations accounting for effects such as non-locality or many-body interactions, there is currently no *practical* way to perform such calculations.

The outlook of this work is the eventual construction of a general many-body non-local nonexpanded polarization density functional, that can efficiently describe electron correlation and dispersion interactions in atoms, molecules, and materials. Since a complete description of polarizability is equivalent to obtaining the full solution of the Schrödinger equation, any feasible method must rely on approximations. Model systems and free atoms, for which polarizability can both be exactly

solved and scaling laws can be established, provide a natural framework for such approximations. Owing to this importance, Chapter 3 explores the polarizability of model systems and free atoms, with special attention to scaling laws with respect to their size. Chapter 4 presents several approaches for approximating polarizabilities of molecules, using correlations with different properties, and using a coupled atoms-in-a-molecule approach. Finally, Chapter 5 moves towards a more complete theory, by presenting the framework for the full non-local polarizability based on the polarization operator, and evaluating the derived expressions for select model systems.

# 2. Background

In Chapter 1, the importance of polarizability in describing electron correlation and long-range interactions was highlighted. This chapter delves into the physical basis of this quantity, as well as advanced computational models. First, Sect. 2.1 introduces some basic principles for understanding long-range interactions between atoms or molecules, the context in which polarizability is commonly used. Currently available methods with varying levels of sophistication to capture the physics of polarizability are presented in Sect. 2.2. Finally, some conceptually important model systems are introduced in Sect. 2.3.

## 2.1 Quantum Mechanical Description of Long-Range Interactions

The well-known concepts of multipole moments and multipolar polarizabilities can be introduced by examining the Taylor expansion of the Coulomb potential, enabling us to consider the interactions of molecules with external fields and other molecules in the same context. Therefore, the discussion will be built up as follows: first, the multipole expansion will be described, with the specific example of an atom or molecule interacting with an external electric field (with the notation loosely based on [11]). After the concept of multipole moments is introduced in such a way, the interaction between two separate atoms or molecules will be discussed, naturally leading to the definition of multipole polarizabilities. This arrangement of ideas follows the presentation as seen in [4].

The multipole expansion can be obtained from the Taylor series representation of the Coulomb potential [11]. We start from an arbitrary charge distribution $\rho(\mathbf{r}')$ and are interested in the potential due to this charge density at the point $\mathbf{r}$. We assume $\mathbf{r} \gg \mathbf{r}'$, that is, the charge density is only non-zero in the vicinity of $\mathbf{r}'$. The Coulomb potential due to the charge density is given by

$$V(\mathbf{r}) = \iiint_{V'} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}^3\mathbf{r}' \, . \tag{2.1}$$

The distance between the vectors can be written as

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \frac{1}{r\sqrt{\left|1 - 2\frac{\hat{\mathbf{r}}\cdot\mathbf{r}'}{r} + \left(\frac{r'}{r}\right)^2\right|}} \quad . \tag{2.2}$$

With the assumption $\mathbf{r} \gg \mathbf{r}'$, the term under the square root can be expressed with a small $\epsilon$ as

$$\frac{1}{|\mathbf{r} - \mathbf{r}'|} = \frac{1}{r\sqrt{1 + \epsilon}} \quad . \tag{2.3}$$

Using Taylor series expansion around $\epsilon = 0$, we obtain

$$(1 + \epsilon)^{-1/2} = 1 - \frac{\epsilon}{2} + \frac{3\epsilon^2}{8} - \frac{5\epsilon^3}{16} + \ldots \tag{2.4}$$

Note that if we allow for the extension of the binomial function using the Gamma function instead of the factorial, the Taylor expansion can be given in a closed form

$$(1 + \epsilon)^{-1/2} = \sum_{n=0}^{\infty} \binom{-1/2}{n} \epsilon^n \quad . \tag{2.5}$$

Finally, the general multipole operator $\xi$ of rank $n$ in Cartesian form (written for a one-center expansion for simplicity, with the indices $\alpha, \beta..., \omega$ representing the basis set components) is given as

$$\xi_{\alpha\beta...\omega}^{(n)} = \frac{(-1)^n}{n!} r^{2n+1} \partial_{r_\omega}...\partial_{r_\beta}\partial_{r_\alpha}\left(\frac{1}{r}\right) \quad . \tag{2.6}$$

Some examples of multipole operators are the dipole $\hat{\boldsymbol{d}}$, quadrupole $\hat{\boldsymbol{\Theta}}$ and octupole $\hat{\boldsymbol{\Omega}}$ moment operators, which have select components such as

$$d_x = -r^3 \partial_x \left(\frac{1}{r}\right) = x \ ,$$

$$\Theta_{xx} = \frac{r^5}{2} \partial_x \partial_x \left(\frac{1}{r}\right) = \frac{1}{2}\left(3x^2 - r^2\right) \ , \tag{2.7}$$

$$\Omega_{zzz} = \frac{-r^7}{6} \partial_z \partial_z \partial_z \left(\frac{1}{r}\right) = z\left(\frac{5}{2}z^2 - \frac{3}{2}r^2\right) \ .$$

The multipole contributions of a charge density $\rho(\boldsymbol{r})$ can then be identified, for example, by considering the potential generated at a faraway point $\boldsymbol{r}'$

$$\phi(\boldsymbol{r}) = \int_{V'} \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \, \mathrm{d}^3\boldsymbol{r}' = \sum_{n=0}^{\infty} \boldsymbol{\nabla}^n \left(\frac{1}{r}\right) \int_{V'} \rho(\boldsymbol{r}') \xi_{\alpha\beta\dots\omega}^{(n)}(\boldsymbol{r}') \, \mathrm{d}^3\boldsymbol{r}' \ . \tag{2.8}$$

Equations (2.6)-(2.8) are somewhat tedious. Moreover, some Cartesian indices in Eq. (2.7) are redundant due to Schwarz's theorem $\partial_i \partial_j f = \partial_j \partial_i f$. An alternative approach that results in mathematically more convenient equations is to use the spherical harmonics for the multipole expansion

$$\phi(\boldsymbol{r}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \frac{4\pi}{2\ell + 1} q_{\ell m} \frac{Y_{\ell m}(\theta, \phi)}{r^{\ell+1}} \ , \tag{2.9}$$

with the multipole moments defined as

$$q_{\ell m} = \int_{V'} Y_{\ell m}^*(\theta', \phi') r'^{\ell} \rho(\boldsymbol{r}') \, \mathrm{d}^3\boldsymbol{r}' \ . \tag{2.10}$$

As per Eq. (2.8), expansion of the Coulomb potential leads to the multipole moments of the charge density. The energy of a molecule in an external electric potential is also expressed using the same moments. To demonstrate this fact, let us consider the first-order contribution to the energy within Rayleigh-Schrödinger perturbation theory. If the perturbation is given by an arbitrary potential $H' = V(\boldsymbol{r} - \boldsymbol{r}')$, then the first-order correction to the energy reads

$$E^{(1)}(\boldsymbol{r}) = \langle 0|\hat{H}'|0\rangle = \int_{V'} \rho(\boldsymbol{r}')V(\boldsymbol{r}-\boldsymbol{r}')\,\mathrm{d}^3\boldsymbol{r}' = \sum_{n=0}^{\infty} \boldsymbol{\nabla}^n V(\boldsymbol{r}) \int_{V'} \rho(\boldsymbol{r}')\xi_{\alpha\beta...\omega}^{(n)}(\boldsymbol{r}')\,\mathrm{d}^3\boldsymbol{r}' \ . \quad (2.11)$$

If the potential is Coulombic (e.g. due to an external point charge), Eq. (2.11) is nominally equivalent to Eq. (2.8). In general, Eq. (2.11) shows that the $n$th multipole moment interacts with the $n$th gradient of the external potential; e.g., the total charge interacts with the electric potential, the dipole moment interacts with the gradient of the potential (the homogeneous component of the electric field), and so on.

The second-order contribution to the energy in the external field is given by

$$
\begin{aligned}
&-\sum_{n\neq0} \frac{\langle 0|V(\boldsymbol{r}-\boldsymbol{r}')|n\rangle\,\langle n|V(\boldsymbol{r}-\boldsymbol{r}')|0\rangle}{E_n - E_0} = \\
&-\sum_{n\neq0} \frac{\langle 0|\sum_m^{\infty}\boldsymbol{\nabla}^M V(\boldsymbol{r})\xi_{\alpha\beta...\omega}^{(m)}(\boldsymbol{r}')|n\rangle\,\langle n|\sum_m^{\infty}\boldsymbol{\nabla}^M V(\boldsymbol{r})\xi_{\alpha\beta...\omega}^{(m)}(\boldsymbol{r}')|0\rangle}{E_n - E_0}\ .
\end{aligned}
\quad (2.12)
$$

Similarly to the previous considerations, the matrix elements in Eq. (2.12) are integrated over $\boldsymbol{r}'$, factoring out the potential gradients. Multipole polarizabilities can be identified [4] as

$$E^{(2)}(\boldsymbol{r}) = -\frac{1}{2}\boldsymbol{\nabla}V(\boldsymbol{r})\boldsymbol{\alpha}\boldsymbol{\nabla}V(\boldsymbol{r}) - \frac{1}{3}\boldsymbol{\nabla}^2 V(\boldsymbol{r})\boldsymbol{A}\boldsymbol{\nabla}V(\boldsymbol{r}) - \frac{1}{6}\boldsymbol{\nabla}^2 V(\boldsymbol{r})\boldsymbol{C}\boldsymbol{\nabla}^2 V(\boldsymbol{r}) + ... \quad (2.13)$$

The most common example is the dipole-dipole polarizability, calculated using the matrix elements of the operators of the corresponding Cartesian coordinate components (with $q$ being the total charge)

$$\alpha_{xx} = 2q^2 \sum_{n\neq0} \frac{\langle 0|\hat{x}|n\rangle\,\langle n|\hat{x}|0\rangle}{E_n - E_0}\ . \quad (2.14)$$

Comparing the energy expressions of Eqs. (2.11)-(2.13), one can gain a physical interpretation of the multipole polarizability: it connects induced multipole moments to gradients of the external potential. This definition is also imminent in the spherical tensor representation of polarizability,

which defines $\alpha_{\ell m,\ell'm'}$ as the induced $(\ell m)$th multipole moment by the $(\ell'm')$th moment of the potential [4], given as

$$\alpha_{\ell m,\ell'm'} = \sum_{n\neq 0} \frac{\langle 0|q_{\ell m}|n\rangle \langle n|q_{\ell'm'}|0\rangle + \langle 0|q_{\ell'm'}|n\rangle \langle n|q_{\ell m}|0\rangle}{E_n - E_0} .$$ (2.15)

Multipole moments and multipole polarizabilities are used to describe the energy of atoms and molecules in external potentials, but intermolecular interactions can also be understood using these quantities. In this context, the Hamiltonian of the full system that contains nucleus–nucleus, nucleus–electron and electron–electron interactions (in atomic units) reads

$$\mathcal{H}_{\text{tot}} = -\sum_I \frac{\boldsymbol{\nabla}_I^2}{2M_I} - \sum_i \frac{\boldsymbol{\nabla}_i^2}{2} - \sum_{i,I} \frac{Z_I}{r_{iI}} + \sum_{I>J} \frac{Z_I Z_J}{r_{IJ}} + \sum_{i>j} \frac{1}{r_{ij}} .$$ (2.16)

The nuclear degrees of freedom in Eq. (2.16) are represented by capital letters $I$ and $J$, while the electronic degrees of freedom are represented with $i$ and $j$. The Born-Oppenheimer approximation will be employed throughout the thesis, focusing on the electronic subsystem only, meaning that the kinetic and potential energy operators for nuclei will be omitted (by shifting the reference energy), and the electrostatic repulsion between nuclei will be treated as a constant background charge field in the electronic Hamiltonian.

Equation (2.16) treats all electrons and nuclei on equal footing within the same Cartesian coordinate system. Long-range interactions, in general, are discussed in the context of atoms and molecules at a large distance. A natural way is to assume that some electrons and nuclei belong to one part $(A)$ of the entire system, whereas the others belong to another part $(B)$. The total Hamiltonian is then the sum of the non-interacting Hamiltonians of $A$ and $B$ as well as an interaction term

$$\mathcal{H}_{\text{tot}} = \mathcal{H}_A + \mathcal{H}_B + \sum_{a\in A}\sum_{b\in B} \frac{q_a q_B}{r_{ab}} ,$$ (2.17)

where $a$ and $b$ can represent a nucleus or an electron, with $q_a = Z_a$ for nuclei and $q_a = -1$ for electrons.

Intermolecular forces were historically discussed within quantum mechanics by treating the interaction term in Eq. (2.41) using Rayleigh-Schrödinger perturbation theory. To use perturbative

formulas, the wavefunctions of the unperturbed Hamiltonians are needed. Since the unperturbed Hamiltonian is a sum of two independent terms when the overlap is neglected, the total wavefunction can be written as a product of the individual eigenfunctions. The states of the unperturbed Hamiltonian are then written as $|mn\rangle$, where $m$ and $n$ are the quantum numbers belonging to $A$ and $B$, respectively, with $|00\rangle$ representing the ground state of the total (non-interacting) system. The first-order correction to the energy from the Rayleigh-Schrödinger theory is given as (using the notation $\mathcal{H}'$ for the perturbation)

$$E^{(1)} = \langle 00|\hat{\mathcal{H}}'|00\rangle \ , \tag{2.18}$$

whereas the second-order correction reads

$$E^{(2)} = - \sum_{mn \neq 0} \frac{\langle 00|\hat{\mathcal{H}}'|mn\rangle \langle mn|\hat{\mathcal{H}}'|00\rangle}{E_m^A + E_n^B - E_0^A - E_0^B} \ , \tag{2.19}$$

where it is required that $|mn\rangle$ is an excited-state wavefunction, so either $n$ or $m$ is non-zero: if only one subsystem is excited, we refer to the energy contribution as induction; if both subsystems are excited, then it is called as dispersion.

The first-order correction in Eq. (2.18) is the electrostatic interaction energy. Explicitly writing out the matrix element, one can see that the formula obtained can be interpreted as arising from the energy of species $A$ in the potential of $B$ (or vice versa)

$$E_{\text{electrostatic}} = \langle 00|\hat{\mathcal{H}}'|00\rangle = \int_{V'} \frac{\rho^A(\boldsymbol{r}')\rho^B(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \ \mathrm{d}^3\boldsymbol{r}' = \int_{V'} \rho^A(\boldsymbol{r}')V^B(\boldsymbol{r}') \ \mathrm{d}^3\boldsymbol{r}' \ , \tag{2.20}$$

which can also be interpreted using the gradients of the potential that act on the multipole moments of the charge distribution, as was already discussed for Eq. (2.11). As will be mentioned later, the exchange energy also arises from the first-order correction if antisymmetrized wavefunctions are considered.

Induction can be treated in a similar manner. Using the ground state $|0\rangle$ for atom $A$ and the excited state $|n\rangle$ for $B$, the induction energy, up to dipole contributions to the potential generated by $A$, is [4]

$$E_{\text{induction}} = -\sum_{n\neq0} \frac{\langle 00|\hat{\mathcal{H}}'|0n\rangle \langle 0n|\hat{\mathcal{H}}'|00\rangle}{E_n^B - E_0^B} =$$
$$-\left(q^A\boldsymbol{\nabla}\frac{1}{r} - \boldsymbol{d}^A\boldsymbol{\nabla}^2\frac{1}{r}\right)\sum_{n\neq0} \frac{\langle 0|\hat{\mathcal{H}}'|n\rangle \langle n|\hat{\mathcal{H}}'|0\rangle}{E_n^B - E_0^B}\left(q^A\boldsymbol{\nabla}\frac{1}{r} - \boldsymbol{d}^A\boldsymbol{\nabla}^2\frac{1}{r}\right) ,$$

(2.21)

with rewriting the part corresponding to atom $A$ similarly as in Eq. (2.20) and noticing that the sum-over-states expression gives (half) the dipole polarizability for atom $B$. The dipole contribution to the induction energy is then, using the electric field of atom $A$

$$E_{\text{induction}} = -\frac{1}{2}\boldsymbol{\nabla}V^A\boldsymbol{\alpha}^B\boldsymbol{\nabla}V^A ,$$

(2.22)

which shows that the induction energy can be interpreted as the interaction of atom $A$ with the *induced* multipole moments of atom $B$. The total induction energy is twice that of Eq. (2.22), as the interaction between $A$ and $B$ should have two symmetric contributions.

Similar considerations for the dispersion interaction lead to the formula (still in dipole approximation)

$$E_{\text{disp}} = -\sum_{mn\neq0} \frac{\langle 00|\hat{\mathcal{H}}'|mn\rangle \langle mn|\hat{\mathcal{H}}'|00\rangle}{E_n^B + E_m^A - E_0^B - E_0^A} =$$
$$-\left(\boldsymbol{\nabla}^2\frac{1}{r}\right)\sum_{mn\neq0} \frac{\langle 0|\hat{\boldsymbol{d}}^A|m\rangle \langle m|\hat{\boldsymbol{d}}^A|0\rangle \langle 0|\hat{\boldsymbol{d}}^B|n\rangle \langle n|\hat{\boldsymbol{d}}^B|0\rangle}{E_n^B + E_m^A - E_0^B - E_0^A}\left(\boldsymbol{\nabla}^2\frac{1}{r}\right) .$$

(2.23)

The expression in Eq. (2.23) cannot be easily factored into a product of the properties of atom $A$ and $B$ due to the mixing of terms in the denominator. An approach is to use the identity proposed by McLachlan [12]

$$\frac{1}{A + B} = \frac{2}{\pi}\int_0^\infty \frac{AB}{(A^2 + \omega^2)(B^2 + \omega^2)} \, \mathrm{d}\omega .$$

(2.24)

The energy denominator in Eq. (2.23) is written as $\omega_m^A + \omega_n^B$ (in atomic units, with $\hbar = 1$), interpreting the energy differences as transition frequencies, delivering the following formula

$$E_{\text{disp}} = -\left(\nabla^2 \frac{1}{r}\right) \int_0^\infty \sum_{m \neq 0} \frac{\langle 0|\hat{\boldsymbol{d}}^A|m\rangle \langle m|\hat{\boldsymbol{d}}^A|0\rangle}{(\omega_m^A)^2 + \omega^2} \sum_{n \neq 0} \frac{\langle 0|\hat{\boldsymbol{d}}^B|n\rangle \langle n|\hat{\boldsymbol{d}}^B|0\rangle}{(\omega_n^B)^2 + \omega^2} \; d\omega \left(\nabla^2 \frac{1}{r}\right) . \tag{2.25}$$

This formulation allows us to identify the polarizabilities at imaginary frequencies, when written (using the components of the dipole-dipole coupling tensor, $\nabla^2 1/r = \mathbf{T}$) as

$$E_{\text{disp}} = -\frac{1}{2\pi} T_{\alpha\beta} T_{\gamma\delta} \int_0^\infty \alpha_{\alpha\delta}^A(iv)\alpha_{\beta\gamma}^B(iv) \; dv , \tag{2.26}$$

commonly expressed in Cartesian coordinate basis (using the Kronecker delta notation) as

$$T_{\alpha\beta} = \frac{(3R_\alpha R_\beta - R^2 \delta_{\alpha\beta})}{R^5} , \tag{2.27}$$

and the integral over the imaginary frequencies referred to as the Casimir-Polder [13, 14] formula. For spherical atoms, we have

$$T_{\alpha\beta} T_{\gamma\delta} = \frac{(3R_\alpha R_\beta - R^2 \delta_{\alpha\beta})(3R_\gamma R_\delta - R^2 \delta_{\gamma\delta})}{R^{10}} = \frac{6}{R^6} , \tag{2.28}$$

leading to the well-known "R$^{-6}$" formula describing the decay of the dipole contribution to dispersion energy as a function of the interatomic distance $R$

$$E_{\text{disp}} = -\frac{3}{\pi R^6} \int_0^\infty \alpha^A(iv)\alpha^B(iv) \; dv . \tag{2.29}$$

This energy is often written as

$$E_{\text{disp}} = -\frac{C_6}{R^6} , \tag{2.30}$$

using $C_6$ coefficients that contain the full complexity of the two-body dipole-dipole dispersion interaction. It can be shown [4] that higher-order contributions to the dispersion energy can also be approximately recast into a relatively simple decay formula (having a Casimir-Polder formula for each $C_n$ coefficient written using the respective interacting higher-order polarizabilities), leading

14

to the expression [15] representing the dispersion energy as a sum of $n$th-order contributions

$$E_{\text{disp}} = \sum_n E_{\text{disp}}^{(n)} = - \sum_{n=6,8,10,\dots} \frac{C_n}{R^n} \ . \tag{2.31}$$

Finally, similarly to the electrostatic forces that were interpreted *via* the interaction between multipole moments and induction as an interaction between permanent and induced multipole moments, dispersion can also be seen as the interaction between fluctuating multipole moments.

Although the exchange interaction lies beyond the scope of this thesis, the effects caused by the Pauli repulsion can be still mentioned for the sake of completeness. A commonly used method to calculate intermolecular interactions is symmetry-adapted perturbation theory (SAPT), which can also account for exchange effects. This approach is based on the fact that the product wavefunction $|00\rangle$ cannot be used in intermediate to short ranges, but a properly antisymmetrized wavefunction requires more elaborate treatments than the Rayleigh-Schrödinger perturbation theory [4, 16]. The fundamental idea of SAPT-based methods is to consider a wavefunction that fulfills the Pauli repulsion

$$\Psi = \hat{A} |00\rangle \ , \tag{2.32}$$

where the unperturbed wavefunction is antisymmetrized by applying the operator $\hat{A}$, which is a proper sum of intermonomer perturbation operators. In practice, monomer wavefunctions are calculated using an approximate computational approach, leading to SAPT(HF), SAPT(DFT) and similarly named methods. The advantage of all these approaches is that no dimer calculation has to be performed; therefore, the computational scaling is favorable, as well as basis set superposition error does not play a role. The disadvantage is that the interaction between monomers is taken into account only in a perturbative way, so relaxation effects resulting from the interaction are very limited.

The interaction energy between monomers in SAPT is expanded in the series of a perturbing operator, the polarization part $E_{\text{pol}}^{(n)}$ obtained from the standard Rayleigh-Schrödinger theory and the exchange $E_{\text{exc}}^{(n)}$ obtained from the antisymmetrized theory

$$E_{\text{SAPT}} = \sum_{n=1}^{\infty} E_{\text{pol}}^{(n)} + E_{\text{exc}}^{(n)} . \tag{2.33}$$

Up to second order, SAPT accounts for electrostatic, exchange, induction, dispersion, and exchange-dispersion interactions.

In summary, this Section shows how the long-range interactions between molecules are described with the help of perturbation theory and the multipole expansion of the Coulomb potential. Electrostatic, induction, and dispersion energy is expressed using the interaction of multipole moments and multipole polarizabilities *via* the corresponding gradients of the potential; these quantities play a central role in the rest of the thesis.

## 2.2   Stairway of Polarizability Methods

Investigating the polarizability of atoms and molecules predates the theory of quantum mechanics [17, 18]. Measurements of relative permittivity and index of refraction were among the first experiments in establishing our modern understanding of the structure of matter, with the Lorentz-Lorenz equation linking microscopic quantities (polarizability $\alpha$ and number of molecules $N$) with the macroscopic refractive index $n_r$

$$\frac{n_r^2 - 1}{n_r^2 + 2} = \frac{N\alpha}{3} . \tag{2.34}$$

Interestingly, Eq. (2.34) discovered by Lorentz and Lorenz independently is equivalent to the Clausius-Mosotti and Debye formulas of Eq. (1.1) – a rare case of multiple discovery in the history of science, which also underlines the central role of polarizability in the development of molecular physics. It was also known at the turn of the 19th–20th century that the refractive index of molecules is approximately the sum of partial contributions, and it was Lorentz who first proposed that, due to Eq. (2.34), this additive property carries over to the polarizability of molecules, possibly providing the first intuitive theory of molecular polarizability [18].

Today, owing to the key role of polarizability for intermolecular interactions and electron correlation, there is a wide variety of predictive methods for this quantity, with varying levels of physical

sophistication behind them. This Section sets out to categorize some of these methods by organizing them into a "stairway" pattern, each step building on the physical picture of the ones behind it, eventually reaching a full theory accounting for all non-local effects. The basic physical model of each step can be summarized as follows
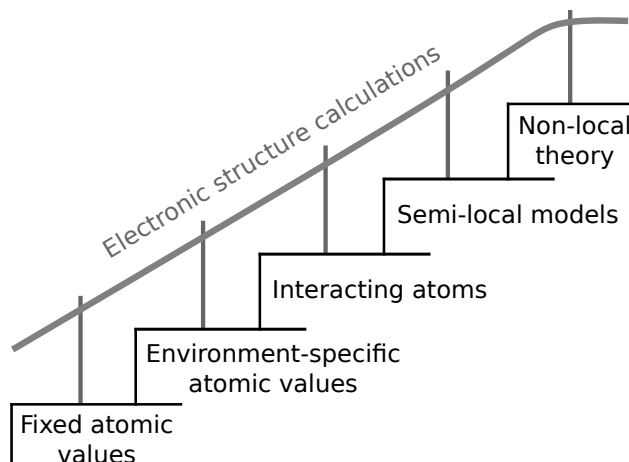


Figure 2.1: A stairway of different polarizability models with increasingly sophisticated physical description.

- **Step I: Fixed atomic values.** It has been known since before the development of quantum mechanics that molecular polarizability, up to a relatively good accuracy, can be predicted by using tabulated values per atom types. Such methods can be parametrized from experimental data of a handful of molecules or by computational analysis of a substantial subset of the chemical compound space.

- **Step II: Environment-specific atomic values.** The obvious flaw of using static, tabulated atomic contributions is that it neglects the actual chemical environment of these atoms within molecules. The models of Step II attempt to include the effect of the local chemical environment by using valence-dependent correction factors or by using a parametrization based on electron density.

- **Step III: Interacting atoms.** While the largest modifications to the "base" polarizability of atoms are due to the local chemical environment, a long-range electrostatic screening between atoms is necessary to properly capture the molecular polarizability, especially if high accuracy in terms of its anisotropy is desired.

17

- **Step IV: Semi-local models.** By properly parametrizing the atomic response accounting for both short- and long-range effects, it should be possible in principle to capture the polarizability of atoms-in-molecules. However, the spatial distribution of polarizability can lead to important insights into which regions in a molecule contribute the most to the response of a given atom, which is especially useful for coarse-graining principles.

- **Step V: Non-local theory.** Semi-local models can lead to an intuitive understanding of the spatial distribution of electronic response, but a generally proper theory can only be built on non-local (spatially two-point) response functions. Knowledge of the exact non-local polarizability would also deliver the exact correlation energy in a non-expanded way, enabling the description of intermolecular interactions and intramolecular electron correlation with the same functional.

In addition to the methods presented on the stairway, correlations between polarizability and various molecular properties can also lead to useful insights, such as the recent observation that polarizability scales with the fourth power of the characteristic size of the system [19].

The correlation between polarizability and orbital energies is relevant from a theoretical point of view, as it forms the basis of Pearson's hard-soft acid-base (HSAB) theory [20, 21]. Based on recent theoretical works [19, 22–28], we can postulate that polarizability can be expressed as a function of two factors that account for *i)* ground state geometry (e.g., van der Waals radius or molecular volume) and *ii)* electronic structure (e.g., ionization energy or hardness) [29]. Although these correlations provide useful conceptual insights, they have not been used to build accurate numerical methods [30].

## 2.2.1   Fixed Atomic Values

The tabulation of atomic contributions to the total molecular polarizability was the first attempt to understand the physical origin of the polarizability, and this method was proven to be largely successful. As already mentioned, this approach was suggested by Lorentz, based on the observation that refractive indices can be approximated with an additive approach and, because of their inherent connection, polarizability must be additive, too [18].

The statement that the polarizability of a molecule is seemingly the sum of its parts was common

knowledge in the last century [17], and although the interaction between atoms modifies this picture (see below), there were multiple attempts to improve the tabulated parameters [31]. Most recently, this approach was revisited by us [29], showing that the atomic additive approximation provides a relatively good performance even when thousands of diverse organic molecules are considered. We have also highlighted that this extensivity of polarizability inherently means that no correlation can be observed with size-independent intensive quantities, such as the HOMO–LUMO gap.

## 2.2.2 Environment-Specific Atomic Contributions

Instead of using tabulated atomic polarizabilities, a better prediction can be achieved by letting the contribution of each atom-type depend on the hybridization of the atom in the molecule [32]. This observation is also corroborated by methods that express the polarizability of chemical bonds rather than atoms [33].

Environment-specific polarizability models can be used in dispersion correction schemes for density functional theory because the geometry and electronic structure of the molecule in question are readily available from electronic structure calculations. A commonly used dispersion correction is Grimme's DFT–D3 [15, 34] scheme, where the dispersion energy as a function of the distance $r_{ij}$ between atoms $i$ and $j$ is given as

$$E_{\text{disp}} = -\sum_{i=1}^{N_{at}} \sum_{j<i} \left( f_{d,6}(r_{ij}) \frac{C_{6,ij}}{r_{ij}^6} + f_{d,8}(r_{ij}) \frac{C_{8,ij}}{r_{ij}^8} \right), \tag{2.35}$$

with $C_6$ and $C_8$ dispersion coefficients explicitly depending on the coordination of the atoms (and the damping functions $f$ of the method).

Grimme's DFT–D3 is used to calculate the dispersion energy. An alternative approach, the Tkatchenko-Scheffler (TS) method, was also originally proposed as a dispersion correction scheme, but it was shown (and it is also revisited later in this thesis) that it can be used to approximate the polarizability of molecules as a sum of atomic contributions [35]. In the TS method, one rescales free atomic reference polarizabilities and dispersion coefficients using the volume $V$ of the atoms in the molecule as

$$\alpha_n^{\text{eff}} = \alpha_n^{\text{free}} \left( \frac{V_n^{\text{eff}}}{V_n^{\text{free}}} \right)^p \quad ; \quad C_{6,n}^{\text{eff}} = C_{6,n}^{\text{free}} \left( \frac{V_n^{\text{eff}}}{V_n^{\text{free}}} \right)^q , \tag{2.36}$$

originally with setting $p = 1$ and $q = 2$, although a value of $q = p - 0.615$ was suggested by Gould [36], whereas the scaling law for polarizability is revisited in Chapter 3 of this thesis.

### 2.2.3 Interacting Atoms

An approach developed in parallel with increasingly sophisticated atomic parametrization is to account for the electrostatic screening between atoms. This idea was first suggested by Silberstein [37] by means of a dipole coupling between the tabulated atomic values, revisited by Applequist not long after [38]. A popular method based on interacting atoms is the Thole model [39], which was recently extended by an additional step of environment-specific screening [40] , written for the dipole moment $\boldsymbol{d}_i$ of an atom $i$ in an external field $\boldsymbol{E}$ employing the dipole polarizability and the dipole-dipole tensor as

$$\boldsymbol{d}_i = \boldsymbol{\alpha}_i (\boldsymbol{E} - \sum_j \mathbf{T}_{ij} \boldsymbol{d}_j) , \tag{2.37}$$

obtaining the corresponding polarizability tensors from a set of linear equations.

It is not hard to see that the dipole-dipole interaction is sufficient to describe the electrostatic screening between localized dipole polarizabilities. It has been shown by Mayer [41] that dipole-dipole screening of a set of oscillators can be self-consistently performed without resorting to pairwise calculations. This was later proven to be equivalent to diagonalizing a Hamiltonian of dipole-coupled quantum Drude oscillators, forming the basis of the many-body dispersion (MBD) method [42], with the Hamiltonian depending on the individual oscillator displacements $\xi$, their frequencies $\omega$ and polarizabilities $\alpha$, with a dipole-dipole coupling $\mathbf{T}$ between them in the interaction term

$$\mathcal{H} = -\frac{1}{2} \sum_i^N \nabla_{\boldsymbol{\xi_i}}^2 + \frac{1}{2} \sum_i^N \omega_i^2 \xi_i^2 + \sum_{i<j}^N \omega_i \omega_j \sqrt{\alpha_i \alpha_j} \boldsymbol{\xi_i} \mathbf{T}_{ij} \boldsymbol{\xi_j} . \tag{2.38}$$

This approach has recently been extended to higher-order multipole contributions [8, 43]. The coupling between higher multipole moments can be easily obtained by taking higher-order tensorial derivatives of the Coulomb potential, but obtaining accurate higher-order atomic polarizabilities remains a challenge [44].

The conceptual advances to go from the first model of Silberstein and Applequist [37, 38] to more recent self-consistent screening approaches should be clearly emphasized. The first models were based on interacting point dipoles centered on atoms, without trying to account for their spatial distribution. Thole [39] has introduced a damping function $\varphi$ that modifies the dipole-dipole interaction tensor in order to properly describe very short interatomic distances, that otherwise could lead to even infinite polarizabilities

$$\mathbf{T'} = -\mathbf{T}\frac{\partial^2 \varphi(|\mathbf{r}_i - \mathbf{r}_j|)}{\partial \mathbf{r}_i \partial \mathbf{r}_j} \; , \tag{2.39}$$

with a spherical potential $\varphi$ fit to satisfy some physical conditions [39, 45]. While Thole proposed that the potential should be expressed in terms of a charge distribution, it was Mayer [41] who popularized the error-function potential obtainable from the interaction of Gaussian charge distributions, which serves as a useful damping function.

Building on the observation that the *erf* damping function can be obtained from the electron density of the quantum Drude oscillator (see Sect. 2.3), Gobre [46] has extensively studied how potentials and scaling laws obtained from QDOs could be used to model coupled atomic response. This approach (named "coupled atomic response in matter", CARMA) was used to predict non-trivial scaling laws of different nanostructured materials with the advantage of using a parametrization that explicitly depends on the local electron density of molecules [47]. Within the CARMA approach, the dipole width $\sigma$ of an oscillator is frequency-dependent, being obtained from the free atomic parameters: polarizability $\alpha^0$, atomic QDO dipole density width $\sigma^{\text{free}}$ and the ratio of the free atomic volume and atom-in-molecule volume $V^{\text{rel}}$ as

$$\sigma(\mathbf{r}, \omega) = \left(\frac{\alpha^0(\omega)}{\alpha^0(0)}\right)^{1/4} \left(V^{\text{rel}}[n(\mathbf{r})]\right)^{1/3} \sigma^{\text{free}} \; . \tag{2.40}$$

## 2.2.4 Semi-Local Models

The general definition of a non-local response function $\chi(\boldsymbol{r}, \boldsymbol{r}')$ contains the derivative of a measured quantity $x(\boldsymbol{r})$ with respect to a perturbation $h(\boldsymbol{r}')$ as

$$\chi(\boldsymbol{r}, \boldsymbol{r}') = \frac{\partial x(\boldsymbol{r})}{\partial h(\boldsymbol{r}')} \ . \tag{2.41}$$

Equation (2.41) expresses the non-local nature of the response function, assuming that both the perturbation and the change in the observed property can be spatially resolved. For polarizability (as will be discussed in Chapter 5), the two connected quantities are the polarization field $\mathbf{P}(\mathbf{r})$ and the perturbing electric field $\mathbf{E}(\mathbf{r})$. Therefore, recasting Eq. (2.41) in integral form delivers

$$\delta\mathbf{P}(\mathbf{r}) = \int_V \boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}') \delta\mathbf{E}(\mathbf{r}') \, \mathrm{d}^3\mathbf{r}' \ . \tag{2.42}$$

Neglecting the fully non-local nature of the response and using only its semi-local form $\boldsymbol{\alpha}(\boldsymbol{r})$ was first suggested by Theimer [48], later elaborated by Oxtoby [49], assuming that the non-local nature of the polarizability can be represented within the integrand as $\boldsymbol{\alpha}(\boldsymbol{r}, \boldsymbol{r}') = \boldsymbol{\alpha}(\boldsymbol{r})\delta(\boldsymbol{r} - \boldsymbol{r}')$, which gives

$$\delta\mathbf{P}(\mathbf{r}) = \boldsymbol{\alpha}(\boldsymbol{r})\delta\mathbf{E}(\boldsymbol{r}) \ . \tag{2.43}$$

Semi-local response functions are also used in van der Waals (vdW) density functionals. Similarly to response functions in general, vdW density functionals also aim for a fully non-local description, with the correlation energy being expressed in terms of the electron density $\rho$ using a non-local kernel $\phi$ as [1]

$$E_{\mathrm{corr}} = \frac{1}{2} \iint \rho(\boldsymbol{r})\phi(\boldsymbol{r}, \boldsymbol{r}')\rho(\boldsymbol{r}') \, \mathrm{d}^3\boldsymbol{r} \, \mathrm{d}^3\boldsymbol{r}' \ . \tag{2.44}$$

One of the most influential functionals was vDW-DF (and later vDW-DF2) [50, 51], which is based on the assumption that the kernel $\phi$ is a function of the distance $|\boldsymbol{r} - \boldsymbol{r}'|$, which essentially makes it semi-local, the electron density and its gradient only, being proportional to the exchange-correlation

density of a gradient-corrected LDA functional.

Simpler functional forms were proposed by Vydrov and Van Voorhis (VV09 and VV10) by neglecting some physical constraints on the correlation kernel [52, 53]. In particular, VV10 was constructed to have a relatively simple form, given as

$$\phi(\mathbf{r}, \mathbf{r}') = \frac{-3}{2g(\mathbf{r})g(\mathbf{r}')(g(\mathbf{r}) + g(\mathbf{r}'))} \ , \tag{2.45}$$

where $g(\mathbf{r})$ is a function of the local density only. A semi-local polarizability expression can be obtained directly from the VV10 functional [54], as it is a functional of the local electron density

$$\alpha(\mathbf{r}, iu) = \frac{\rho(\mathbf{r})}{\frac{4\pi}{3}\rho(\mathbf{r}) + C \left|\frac{\nabla\rho(\mathbf{r})}{\rho(\mathbf{r})}\right|^4 + u^2} \ . \tag{2.46}$$

This semi-local polarizability expression from the VV10 functional was used by Hermann *et al.* [55] to extend the many-body dispersion approach to the MBD–NL model (NL stands for non-local). It is interesting to note that even though the VV10 functional provides a non-local expression for the correlation energy, the response function of MBD–NL constructed from it is still only semi-local.

The semi-local polarizability of a few model systems was obtained using the Dalgarno-Lewis perturbation theory [56] by Orttung [57, 58], which was also recently revisited [59]. While the underlying theory of the Dalgarno-Lewis approach is quite complex, for our purpose, it is essentially enough to understand that it allows one to express the perturbed wavefunction in the form [60]

$$\Psi^{(1)}(\mathbf{r}) = R(r)\Psi^{(0)}(\mathbf{r}) \ , \tag{2.47}$$

and, consequently, the polarizability becomes semi-local

$$\alpha(\mathbf{r}) = R(r)|\Psi^{(0)}(\mathbf{r})|^2 \ . \tag{2.48}$$

While this flavor of perturbation theory can deliver semi-local functions that give the exact polarizability when integrated, they still suffer from consistency issues, as explained below.

## 2.2.5 Non-Local Theory

As discussed above, using a semi-local theory for polarizability is a commonly used approach, despite the fact that it was shown in the 1970s by Sipe [61] that a semi-local polarizability description cannot be completely correct from both a quantum-mechanical and an electrodynamical points of view. A fully consistent theory should be built on a two-point response function, defined by Eq. (2.42) together with the connection between the induced charge density and the induced polarization density

$$\delta\rho(\boldsymbol{r}) = -\boldsymbol{\nabla} \cdot \delta\boldsymbol{P}(\boldsymbol{r}) \ . \tag{2.49}$$

The theory of non-local polarizability was later elaborated by Hunt, proving a number of important properties of this quantity: i) it contains all static multipole polarizabilities; ii) the dispersion energy between two interacting species can be expressed *via* this quantity; iii) in general, the non-local polarizability provides a convenient framework for describing molecular properties such as screening factors or Raman spectra [62–64]. Despite the theoretical usefulness of this quantity, it has only been explicitly analyzed for the homogeneous electron gas [65], and no practical calculations have been done within this framework until now.

Although non-local polarizability was not yet used in electronic structure calculations, non-local susceptibility is constructed as an intermediate quantity in methods such as RPA or SAPT. An interesting approach to obtaining the polarizability of atoms-in-molecules is to first construct the non-local susceptibility and then condense this quantity to atoms. While the polarizability obtained in such a way is not non-local on the response function level, this method offers accurate prediction due to the non-locality of the underlying quantities. Two examples of this approach are the ACKS$\omega$2 [66] and ISA–POL [67] methods, both starting with a sum-over-states expression of the non-local susceptibility $\chi$, then projecting this quantity to atomic values using a numerical basis set.

## 2.2.6 Electronic Structure Methods

In the stairway of the polarizability methods shown in Fig. 2.1, electronic structure calculations are represented as a handlebar, providing a stable reference for each step along the path. In principle,

almost any electronic structure method can be used to calculate polarizability, as the calculation only requires an additional one-electron integral due to a perturbing operator acting only on a single electron coordinate[see Eqs. (2.50)-(2.51)]. Moreover, since the computational cost of two-electron integrals dominates the cost of DFT and wavefunction-based methods, adding an external field to a molecule has only a small effect on the runtime of such calculations. Notable exceptions to this rule are zero differential overlap or near-linear scaling quantum chemical methods, since most of the two-electron integrals are neglected in these approaches, meaning that an additional one-electron integral can significantly increase costs.

Polarizability, by definition, can be expressed as the first derivative of a multipole moment or the second derivative of the energy with respect to an external perturbing electric field (the energy is denoted $U$ here).

$$\alpha_{ij} = \frac{\partial \xi_i}{\partial E_j} = -2 \frac{\partial^2 U}{\partial E_i \partial E_j} \ . \tag{2.50}$$

Equation (2.50) can lend itself to a simple implementation for most electronic structure codes, since the Hamiltonian perturbed by a homogeneous field of strength $|E|$ is just

$$\mathcal{H} = \mathcal{H}^0 + |E| \sum_i \hat{r}_i \ , \tag{2.51}$$

with a sum over the index of the electrons $i$, meaning that only the position operator, which is a one-electron operator, needs to be evaluated in addition [68]. Bearing in mind that the wavefunction in quantum chemistry calculation is normalized to the total number of electrons, the expectation value of $\sum_i \hat{r}_i$ directly gives the total dipole moment.

The results of this finite-field approach depend on a proper choice of the field strength $|E|$: It needs to be small enough so that perturbation theory is valid, but large enough so that differentiation can be carried out with low numerical error. A way to achieve controlled numerical error is the Romberg procedure, which iteratively refines the numerical grid applied with a trapezoidal scheme [69]. In principle, multipolar polarizabilities can also be evaluated in such a manner by calculating the derivative of the induced multipole moment with the corresponding gradient of the field (or taking the corresponding energy derivative), in which case the constant prefactors are found from the so-called Buckingham expansion [70].

Apart from setting an appropriate field strength, there are two practical difficulties with finite field calculations. For systems having nonzero total angular momentum, the polarizabilities of the electronic microstates can be different (see Fig. 2.2) since an applied external field can pick "preferred" electron configurations, giving distinct results for different Cartesian directions. In particular, the polarizability tensors of free atoms contained in the NIST database suffer from non-physical symmetry breaking due to this effect [71].

Electron configurations of the oxygen atom

$|X\rangle$    $|Y\rangle$    $|Z\rangle$

No electric field: Atom is spherical, $|X\rangle, |Y\rangle, |Z\rangle$ are degenerate

In electric field:
- Spherical symmetry is broken
- $|X\rangle, |Y\rangle, |Z\rangle$ become non-degenerate
- Each of $|X\rangle, |Y\rangle, |Z\rangle$ have different polarizability
- The actual orbital occupancies within *ab initio* calculation may oscillate between $|X\rangle, |Y\rangle, |Z\rangle$
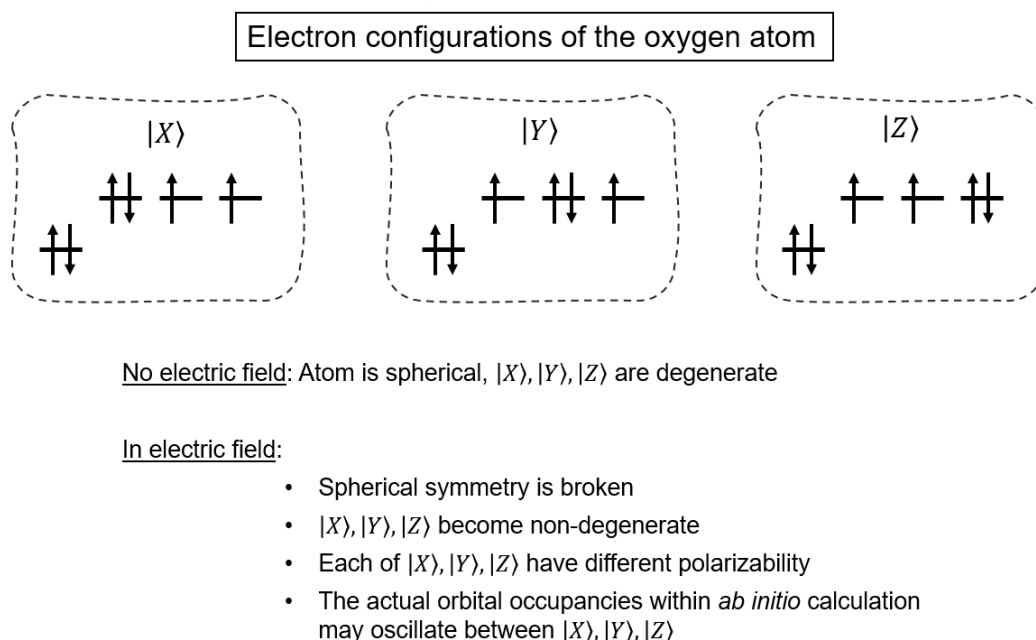
Figure 2.2: Illustration of the arbitrariness of the oxygen microstate energies in an oriented external electric field [44].

A second issue arises when considering the convergence of polarizability with respect to the basis set size. The observation that polarizability converges very slowly with respect to the size of the basis set was one of the main reasons behind the development of modern basis sets, resulting in the widely used Dunning and Sadlej basis functions [72, 73]. The slow convergence is illustrated for the case of the nitrogen atom in Fig. 2.3. In making this graph, perturbations by external electric fields were added to the core Hamiltonian (an option in the PySCF code [74]) as $\mathcal{H}' = -E_\alpha r_\alpha$ and $\mathcal{H}' = 0.5E_{\alpha\beta}(3z^2 - r^2)$, respectively, for the dipole and quadrupole polarizabilities. Polarizabilities were extracted from the numerical derivatives of energy with respect to the strength of applied electric fields (chosen to be 0.001 a.u.). Dunning's aug-cc-pVXZ basis sets were used, singly or doubly augmented (referred to as "single d" or "double d" in the figure), together with a coupled
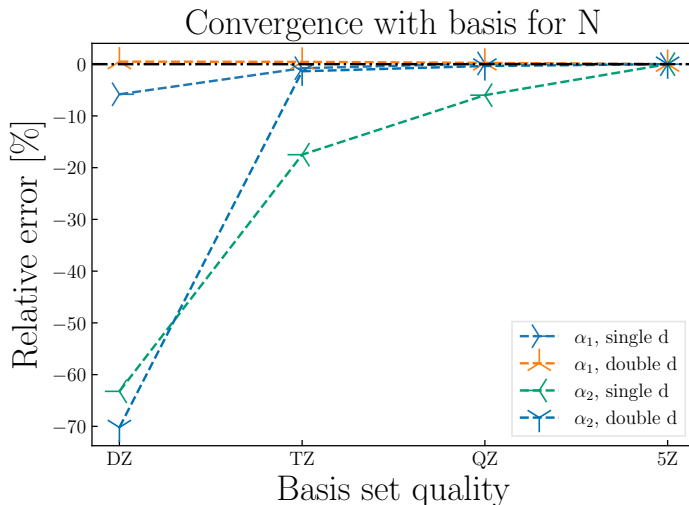
cluster UCCSD(T) calculation.



Figure 2.3: Prediction error in the dipole and quadrupole polarizabilities of the quadruplet nitrogen atom with respect to the basis set, calculated using UCCSD(T).

Due to these theoretical and practical difficulties, finite-field calculations are rarely used for large systems. The alternative approach, nowadays implemented in many *ab initio* codes, is to evaluate the perturbative expressions on the ansatz for the electronic wavefunction. The two most common methods are the coupled-perturbed Hartree-Fock (CPHF) and the density functional perturbation theory (DFPT), where the perturbation is taken on the Hartree-Fock or Kohn-Sham equations, respectively. In both methods, the perturbation is included via a perturbation parameter on the molecular orbitals themselves. A complete account for the CPHF equations would require a tedious treatment of both the unperturbed and perturbed theory [75, 76], so it is omitted from this thesis. Recently, such perturbative methods have been extended to multideterminant and multireference methods, enabling accurate calculations for both static and dynamic correlations.

Modern machine learning (ML) methods are also becoming more popular for evaluating response properties. Wilkins *et al.* [77] have used an ML algorithm - named ALPHAML - to learn the connection between dipole polarizability and molecular geometry. Their model was trained on 7000 geometries from the QM7b dataset [78], with polarizability calculated using the linear response coupled cluster (LR–CCSD) method. The resulting ALPHAML model can predict the polarizability tensor at a nominal cost, with an accuracy comparable to a DFT calculation. Due to the difficulty in extrapolating with ML models, ALPHAML is only accurate within the range of the training dataset.

## 2.3 Polarizability of Model Systems

A general perturbative expression for the multipolar polarizability was already introduced in Eq. (2.15), showing that the exact evaluation of the polarizability requires knowledge of the elements of the transient multipole moment matrix (between the ground and any excited state, for the polarizability of the ground state). This is infeasible for most real atoms and molecules and still challenging even for quantum-mechanical models. Consequently, the few exactly solvable systems play an important role both in our intuitive understanding of polarizability and in providing a basis for practical simulation methods.

The description of atomic spectra was one of the main driving forces in the development of quantum theory. The Stark effect, as the shift of atomic energy levels in an external electric field, was described in 1914 [79]. Curiously, even after the development of quantum mechanics, the description of the Stark effect even for hydrogen atom posed significant challenges, first solved by Waller [80], using ideas later formalized by Dalgarno [56] (called Dalgarno-Lewis perturbation theory in modern articles).

The source of the mathematical difficulty is the infinite nature of the summation in Eq. (2.15), which not only contains (an infinite number of) bound states, but must also account for continuum contributions. Ground-to-continuum-state transitions are not negligible, contributing 18.6% to the dipole, 64.4% to the quadrupole, and more than 91% to the octupole polarizability to hydrogen-like atoms [81]. Due to this difficulty, a two-level model of the hydrogen atom is still used for analytical derivations [82], even if the first-order correction to the hydrogen atom can be obtained analytically.

A way of formalizing the two-level approximation is the so-called Unsøld approximation [19, 83, 84] often used to understand qualitative trends of molecular polarizability. Within this approach, polarizability is written using an average excitation $\Delta E$, most often treated as a fitting parameter [85]

$$\boldsymbol{\alpha} = \frac{2}{\Delta E} \sum_{n \neq 0}^{\infty} \langle \Psi_0 | \hat{\mu} | \Psi_n \rangle \otimes \langle \Psi_n | \hat{\mu} | \Psi_0 \rangle \; . \tag{2.52}$$

The modeling of polarization of materials using an effective charged harmonic model also predated quantum mechanics [86]. Moreover, this classical Drude model is still used in polarizable molecular force field simulations [87, 88], with the quantum-mechanical analogue, the quantum Drude

oscillator (QDO), which is both an attractive analytical model for polarizability and a convenient Hamiltonian for Monte Carlo simulations [89, 90]. The quantum mechanics of the QDO model will be presented in Sect. 3.2.1. However, two conceptual observations can already be noted here. First, because of selection rules, the polarizabilities of the QDO model contain just a finite number of contributions (only the first excited state for dipole polarizability), so energies in an external field are analytically obtainable. This also means that setting the average excitation of Eq. (2.52) to $\Delta E = E_1 - E_0$ is exact for this model. Second, not only the energies, but also the wavefunction can be analytically written for external electric fields, making the QDO model a convenient choice for analytical derivations for interactions of atoms/molecules in a field [91].

Yet another model for which polarizability can be analytically obtained is the particle in a box, since the system has only bound states and the wavefunctions are mathematically convenient to work with. The dipole polarizability of the $n$th state (the ground state is $n = 1$) of the 1D particle in a box of length $A$ is [19, 92]

$$\alpha_n = \frac{\mu q^2}{\hbar^2} \frac{(15 - n^2\pi^2)}{12n^4\pi^4} A^4 \ . \tag{2.53}$$

The particle in a box is a model system that is rapidly gaining popularity, due to the practical importance of quantum dots and advances in the theory of confined systems in general [93–95], but it has not yet been widely used as a polarizability model, especially in excited states.

Polarizability, the quantity playing a central role in this Thesis, was introduced in this Chapter. It was shown that multipole moments and multipolar polarizabilities provide a natural backbone for the discussion of intermolecular interactions. The "traditional" theory of long-range interactions considers these quantities without any spatial extent, which is a consequence of the approximations behind the multipole expansion. It was also seen, however, that different methods approximate the non-local nature of polarizability to different levels, and fully non-local response functions can be used to approximate correlation energies without needing to resort either to range separation or multipole expansion. The following chapters will show how understanding the properties of molecular and atomic polarizabilities, together with the study of model systems, can lead to a fully non-local theory, eventually paving the way to a general polarizability density functional.

# 3. Polarizability of Model Systems and Atoms

This chapter is devoted to a consideration of systems in which qualitative as well as quantitative models for polarizability can be well tested. Due to the experimental and theoretical difficulties associated with evaluating the polarizability, such systems must be simple. As far as exact analytical results are concerned, only select model Hamiltonians are solvable, including the harmonic oscillator, the hydrogen atom, the two-state system, and the confined particle. Free atoms provide useful tests for polarizability models due to the availability of accurate numerical and experimental data and the convenient fact that the charge density is symmetrically centered around a single origin, making the discussions concerning gauge freedom, that is, a degree of arbitrariness regarding the choice of the coordinate system simpler.

A detailed discussion of analytical results for model systems is not fully presented here as they can be found in standard textbooks [96] and the Supplemental Material in [19]. Instead, this Chapter focuses on two findings: a correlation between polarizability and an intrinsic length scale of quantum mechanical systems and an efficient recipe for representing atomic response functions with a harmonic oscillator model.

Sect. 3.1 builds heavily on [19] and Sect. 3.2 contains an adapted version of the publication [44] – To this latter work, I have contributed by performing *ab initio* calculations, visualizing the results, and developing the theory of the polarization potential to explain the findings. The derivation of the correlated descriptor for atomic size was done with the help of Dr. Matteo Gori.

## 3.1 Polarizability and System Size in Quantum Mechanics

The dipole polarizability determines the strength of the electric response of a system of charged particles to applied electric fields as well as dispersion/polarization interactions between atoms or molecules [4, 96, 97], playing an important role in the interpretation of some experiments [98–102]. Efficient models for polarizability are useful to predictively describe various phenomena in physics, chemistry, and biology. Moreover, a detailed understanding of quantum-mechanical (QM) polarization mechanisms is desirable for developing a microscopic picture of intrinsic vacuum response properties [103–105].

In general, the dipole polarizability is a second-rank tensor determining the dipole moment induced by an applied electric field: $\mathbf{d} = \boldsymbol{\alpha}\mathbf{E}$. For anisotropic systems, the polarizability tensor can be diagonalized using the principal axes, whereas in the case of isotropic systems it effectively reduces to a scalar: $\alpha_{ii} = \alpha = \frac{1}{3}\operatorname{Tr}\boldsymbol{\alpha}$. For a QM system in its ground state, the dipole polarizability can be evaluated via the Rayleigh-Schrödinger perturbation theory [96]

$$\boldsymbol{\alpha} = 2\sum_{n\neq 0}^{\infty} \frac{\langle\Psi_0|\hat{\mathbf{d}}|\Psi_n\rangle \otimes \langle\Psi_n|\hat{\mathbf{d}}|\Psi_0\rangle}{E_n - E_0} \, , \tag{3.1}$$

where $\otimes$ indicates the dyadic vector product and the sum goes over all excited states. This formula describes transient or fluctuating electric dipoles as the matrix elements of the dipole operator $\hat{\mathbf{d}} = \sum_j \hat{\mathbf{d}}_j = \sum_j q_j\hat{\mathbf{r}}_j$, where $q_j$ and $\hat{\mathbf{r}}_j$ are the charge and position operator of the $j$th particle, respectively. For an accurate calculation of $\boldsymbol{\alpha}$, all bound and continuum states must be taken into account. Thus, Eq. (3.1), while being exact, is difficult to evaluate in practice. Therefore, various approximations [83, 106–108] have been developed for a more efficient evaluation of Eq. (3.1). In addition to their computational advantage, approximate models often provide a deeper insight into the polarizability and its relation to other physical observables.

According to Eq. (3.1), the polarizability should be related to a certain characteristic length for a given QM system. This has led to the proposition of a number of scaling laws with respect to different effective system sizes:

$$\alpha \propto R_{\text{cl}}^3 \, , \quad \alpha \propto R_{\text{conf}}^4 \, , \quad \alpha \propto R_{\text{vdW}}^7 \, . \tag{3.2}$$

The first relation stems from the classical formula, $\alpha = (4\pi\epsilon_0)R_{cl}^3$, where $\epsilon_0$ is the vacuum permittivity and $R_{cl}$ is the radius of a conducting spherical shell [109] or a hard sphere with uniform electron density and a positive point charge at its center [110]. This formula delivers the most commonly accepted scaling law, which is used in practice to describe the polarizability of atoms in molecules and materials [35, 40, 111–113]. The second relation in Eq. (3.2) holds for confined quantum systems of length $R_{conf}$, as was derived by Fowler [92]. This relation was observed for semiconductor nanocrystals by using terahertz time-domain spectroscopy [98, 99]. The third scaling law, $\alpha \propto R_{vdW}^7$, connecting the atomic polarizability and van der Waals (vdW) radius, was found [28, 114] by studying the balance between exchange and correlation forces for two interacting quantum Drude oscillators (QDO) [115–117]. Subsequently, the approach of [28] has been employed to improve effective models for vdW interactions [118, 119].

All three distinct scaling laws can be represented as

$$\alpha = (4\pi\epsilon_0)R_p^3\,(R_p/R_p^r)^p\,, \tag{3.3}$$

where $(R_p/R_p^r)^p$ is a correction to the classical formula. The renormalization length $R_p^r$ depends on the choice of the effective system size $R_p \in \{R_{cl}, R_{conf}, R_{vdW}\}$, which corresponds to $p = \{0, 1, 4\}$. Whereas $R_1^r$ depends on the system parameters [92], $R_4^r$ was found [28, 114] to be the same for all atoms. However, the vdW radius is an interacting radius rather than an effective system size, and its accurate evaluation independent of polarizability is difficult [28]. Therefore, it is desirable to establish a general relation of $\alpha$ to a concrete effective size of any given QM system, such as the scaling law for confined systems with a defined confinement radius [92]. Since Eq. (3.3) gives the right units of $\alpha$ for any value of $p$, the form of this general relation is not obvious *a priori*.

In Ref. [19], we have shown that for many distinct QM systems, the principal-axis components of the polarizability tensor in Eq. (3.1) are given by a unified expression

$$\alpha_{ii} = C_i(4\mu q^2/\hbar^2)L_i^4\,, \tag{3.4}$$

where the constant $C_i$ depends on the properties of the quantum particle with mass $\mu$ and charge $q$. The characteristic length $L_i$ measures the spatial spread of the ground-state wave function $\Psi_0$ with respect to its center of mass corresponding to the nuclear position in the case of atoms. The

Euclidean $\mathcal{L}^2$-norm of the position vector, $(\mathbf{r} - \mathbf{R})$, is defined for a QM system described by its ground-state wavefunction as

$$L_i = \sqrt{\int_V (r_i - R_i)^2 \, |\Psi_0(\mathbf{r})|^2 \, \mathrm{d}\mathbf{r}^N} \; , \tag{3.5}$$

where $N$ is the spatial dimension of the system.

Equation (3.4), connecting $\alpha_{ii}$ with the characteristic length $L_i$ along the $i$th principal axis, makes our approach applicable to QM systems of any spatial dimensionality. Moreover, for atom-like systems with a well-defined positively charged center of mass, the dimensionless constant $C_i$ turns out to be close to unity. Equation (3.4) scales as the relation obtained by Fowler [92] for confined systems using an exact derivation of the polarizability or its Unsøld [83] estimates. However, the size of such confined systems was imposed as a classical parameter, which cannot be defined for QM systems in free space. Using our choice of the characteristic length – a QM generalization of the conventional Euclidean $\mathcal{L}^2$-norm – it is shown here that one can properly describe the polarizability of any atom-like QM system.

### 3.1.1 Single Electron Models

Many objects in classical mechanics have well-defined boundaries, making the definition of size descriptors such as volume or area trivial. This is not true in quantum mechanics: many model systems have wavefunctions that decrease asymptotically only at infinity. Moreover, the uncertainty principle also complicates the discussion. The natural way to proceed is to work with the expectation values of the position operator. In particular, the second moment, e.g. $\hat{r}^2$ is a convenient choice, for two reasons. First, this operator becomes a sum of components in Cartesian coordinates, making it possible to treat the $i$th component of a $N$-dimensional system independently as

$$L_i = \left( \int_V r_i^2 \rho(\mathbf{r}) \, \mathrm{d}^N \mathbf{r} \right)^{0.5} \; , \tag{3.6}$$

with the assumption that there is a single center of charge at the origin. The second reason for this choice is that the perturbative expression for dipole polarizability (Eq. (2.14)) contains the square

of the matrix elements of the dipole operator between two states (called transition dipole matrix elements), hinting that using either dipole matrix elements or their squares might simplify later derivations.

The derivation of Eq. (3.4) can be found in [19], with a detailed consideration of some model systems in the Supplementary Material. The central idea is to use the Unsøld approximation [83, 106], that is, to treat the energy denominator in the perturbative expression for polarizability as an effective single value $\Delta E$. Written for the $ii$th component of the (dipole) polarizability tensor, this means

$$\alpha_{ii} = 2q^2 \sum_{n>0}^{\infty} \frac{\langle \Psi_0 \, |\hat{r}_i| \, \Psi_n \rangle \langle \Psi_n \, |\hat{r}_i| \, \Psi_0 \rangle}{E_n - E_0} = (2q^2/\Delta E_i) \sum_{n>0}^{\infty} \langle \Psi_0 \, |\hat{r}_i| \, \Psi_n \rangle \langle \Psi_n \, |\hat{r}_i| \, \Psi_0 \rangle \ . \tag{3.7}$$

The summation in Eq. (3.7) can be rewritten using the closure relation $\sum_{n=0}^{\infty} |\Psi_n\rangle\langle\Psi_n| = 1$:

$$\langle \Psi_0 | \hat{r}_i \sum_{n>0}^{\infty} \Big( |\Psi_n\rangle\langle\Psi_n| \Big) \hat{r}_i | \Psi_0 \rangle = \langle \Psi_0 | \hat{r}_i \Big( 1 - |\Psi_0\rangle\langle\Psi_0| \Big) \hat{r}_i | \Psi_0 \rangle = \langle 0 | r_i^2 | 0 \rangle - \langle 0 | r_i | 0 \rangle^2 = (\Delta r_i)^2, \tag{3.8}$$

delivering an intermediate expression

$$\alpha_{ii} = \frac{2q^2}{\Delta E} (\Delta r_i)^2 \ . \tag{3.9}$$

The issue of scaling the average excitation energy $\Delta E$ with geometric expectation values can be solved similarly, starting from the Thomas-Reiche-Kuhn (TRK) sum rule [120]

$$\frac{2\mu}{\hbar^2} \sum_{n>0}^{\infty} (E_n - E_0) \, \langle \Psi_0 | \hat{r}_i | \Psi_n \rangle^2 = 1 \ , \tag{3.10}$$

leading to a different average excitation energy $(\Delta\widetilde{E}_i)^{-1} = (2\mu/\hbar^2) \, (\Delta r_i)^2$. Assuming a transferable proportionality between the TRK and the Unsøld average excitation energy, the final expression for polarizability is

$$\alpha_{ii} = \frac{4q^2}{\hbar^2} (\Delta r_i)^4 \; . \tag{3.11}$$

Equation (3.11) implies that the polarizability is proportional to the second power of the variance in the position. Others have already obtained a similar scaling result under different approximations [92, 106, 121], resulting in slightly different proportionality constants. However, it should be kept in mind that the average energy approximation was used in two different sum rules, with no clear connection between the two average energies; therefore, instead of postulating the value of the proportionality constant, we have used an undetermined factor $C$ in [19]. Finally, if we only consider model systems with symmetric ground-state wavefunctions (which is true for the most common model Hamiltonians), we can appreciate that the variance $(\Delta r_i)^2$ is equal to the square of the characteristic length in Eq. (3.6) due to the vanishing expectation value of $\hat{r}_i$, giving the final concise result

$$\alpha_{ii} = \frac{4q^2}{\hbar^2} C L_i^4 \; . \tag{3.12}$$

The proportionality constant $C$ was empirically determined for select model systems by evaluating both $L_i$ and $\alpha$ and taking their ratio, with the detailed derivations found in the SI of [19]. The model systems include Hamiltonians with bound and unbound spectra, as well as ones having different dipole matrix selection rules. So far, it is unclear what the magnitude of the $C$ constant implies about the system. An interesting observation is that $C = 1$ for the ground state of the quantum Drude oscillator. The relationship for the $n$th excited state – using the variance of the ground state wavefunction $\sigma = \hbar/2\mu\omega$ – is

$$[\alpha_n]_{ii} = \frac{1}{(2n_i + 1)^2} \frac{4\mu q^2}{\hbar^2} L_{n,i}^4 \; , \qquad L_{n,i} = \sigma_i \sqrt{2n_i + 1} \; , \qquad (n = 0 \; \Leftrightarrow \; \Psi_0) \; . \tag{3.13}$$

Another important model system studied was the particle in a box of size $A_i$ along the $i$th Cartesian axis, for which the results are

$$[\alpha_n]_{ii} = \frac{3(15 - n_i^2 \pi^2)}{(n_i^2 \pi^2 - 6)^2} \frac{4\mu q^2}{\hbar^2} L_{n,i}^4 \; , \qquad L_{n,i} = A_i \frac{\sqrt{n_i^2 \pi^2 - 6}}{2\pi n_i \sqrt{3}} \; , \qquad (n = 1 \; \Leftrightarrow \; \Psi_0) \; . \tag{3.14}$$

The particle in a box Hamiltonian is useful for modeling the properties of quantum dots [94], so relationships such as Eq. (3.14) connecting geometrical properties to observable response functions may be useful in parametrizing quantitative models for such systems.
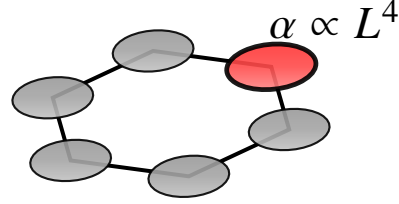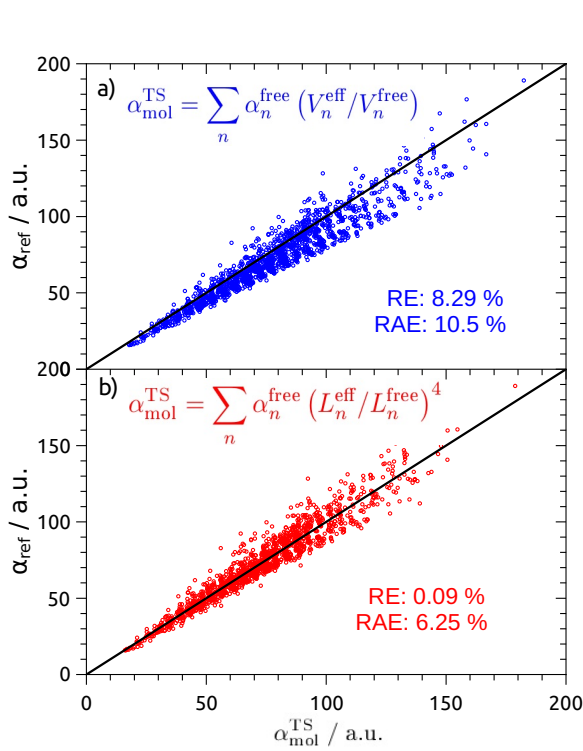


Figure 3.1: Comparing two implementations of the Tkatchenko-Scheffler (TS) method [35]: (a) the conventional TS method based on Eq. (3.15) and (b) the modified TS method based on Eq. (3.16). The corresponding molecular polarizabilities $\alpha_{\text{mol}}^{\text{TS}}$ are shown versus the reference values $\alpha_{\text{ref}}$ calculated by using density functional perturbation theory (DFPT) with the hybrid PBE0 functional [19]. RE and RAE stand for "relative error" and "relative absolute error", respectively, as defined in the text below. The illustration (right) shows the scaling law of the polarizability of atoms within the molecule.

Another manifestation of the scaling law can be found in the definition of the polarizability of atoms in molecules. As elaborated later in Sect. 4.2, polarizability of molecules can be generally constructed as a sum of atomic polarizabilities, as also done in the Tkatchenko-Scheffler (TS) method [35]

$$\alpha_{\text{mol}}^{\text{TS}} = \sum_n \alpha_n^{\text{eff}} = \sum_n \alpha_n^{\text{free}} \left( V_n^{\text{eff}} / V_n^{\text{free}} \right) \ , \tag{3.15}$$

with the scaling law implying that the re-scaling of the free atomic polarizability should be done instead as

$$\alpha_{\text{mol}}^{\text{TS}} = \sum_n \alpha_n^{\text{eff}} = \sum_n \alpha_n^{\text{free}} (L_n^{\text{eff}} / L_n^{\text{free}})^4 \ . \tag{3.16}$$

A calculation based on Eq. (3.16) was performed using the TABS dataset [122] containing optimized geometries of 1641 small organic molecules. Reference molecular polarizability calculations were performed with density functional perturbation theory (DFPT) (as implemented in the `FHI-aims` code [123]), employing the hybrid PBE0 functional and tight numerical basis set for all systems. Molecular volumes and characteristic length-scaled were calculated from the total electronic density based on Hirshfeld partitioning [124], that is, the volume is obtained as

$$V = \int d^3 \boldsymbol{r} w_j(\boldsymbol{r}) \rho(\boldsymbol{r}) r^3 \ , \tag{3.17}$$

and $L$ by taking the second power of $r$ instead of the third. In Eq. (3.17), the Hirshfeld weights $w_j(\boldsymbol{r})$ are obtained from Hirshfeld ratios

$$w_j(\boldsymbol{r}) = \frac{\rho_j(\boldsymbol{r})}{\sum_k \rho_k(\boldsymbol{r})} \ , \tag{3.18}$$

where $\sum_k \rho_k(\boldsymbol{r})$ is the density of a non-interacting system having the same atoms in the same positions, also called "promolecular" density.

The Hirshfeld volumes [124] for the same geometries as found in the TABS dataset were calculated using the PBE functional. (Fig. 3.1) shows that using Eq. (3.16) instead of Eq. (3.15) the average signed relative error $\langle \text{RE} \rangle = \frac{1}{n} \sum_i (\alpha_{\text{ref}}^i - \alpha_{\text{TS}}^i)/\alpha_{\text{ref}}^i$ drops from 8.29% to 0.09%. The practically vanishing systematic deviation and the decrease in the average absolute relative error $\langle \text{RAE} \rangle = \frac{1}{n} \sum_i (\left| \alpha_{\text{ref}}^i - \alpha_{\text{TS}}^i \right|)/\alpha_{\text{ref}}^i$ from 10.5% to 6.25% confirm the applicability of the scaling law used.

Finally, it should be noted that the authors of Ref. [25] also found that the molecular dipole polarizabilities based on the TABS database [122] seem to scale as $\alpha \propto V^{4/3}$. However, their empirical finding was reported without giving any fundamental reason for this non-linear relation. The finding that the four-dimensional scaling relation can be used to efficiently predict molecular polarizabilities from atomic data underlines the fundamental nature of this unraveled scaling law.

## 3.1.2 Effect of Electron Correlation in Atomic Systems

The four-dimensional scaling law between the characteristic length and the polarizability of quantum mechanical models discussed above was derived starting from considering one-electron systems.

All presented formulas rely on single-electron expectation values, whereas a fully physically correct description of quantum mechanical response should be based on response functions depending on two spatial coordinates [2]. Generally, electron correlation effects cannot be recast into one-electron operators, requiring the discussion to be formulated using two-electron expectation values [125]. Moreover, the derivation relied on multiple approximations of the sum rules, introducing average values that are not known even for single-electron systems.

In Ref. [19], we have treated many-electron atoms by generalizing the proportionality constant $C$ of Eq. (3.12) to be shell-dependent which, in part, is due to the shell-dependence of the electron correlation. The formula proposed for the (isotropic) atomic polarizability is

$$\alpha = \frac{4m_e e^2}{\hbar^2} \sum_k^{\text{occ}} \frac{C_k}{\eta_k} \frac{L_k^4}{N_k} \approx \tilde{C} \left( \frac{4m_e e^2}{\hbar^2} \right) \sum_k^{\text{occ}} \frac{L_k^4}{\eta_k N_k} , \tag{3.19}$$

where the sum runs over occupied orbitals with degenerate states treated together, $L_k$ is obtained by Eq. (3.5) for the $k$th orbital, and $N_k$ is its occupation number stemming from the many-electron version of the TRK sum-rule. Then, $\eta_k$ are orbital-dependent factors required for all atoms starting from Li ($\eta_k^{\text{He}} = 1$), empirically found to be $\eta_k = n_k^\ell N_k^{[1+(-1)^\ell]/2}$, where $\ell$ and $n_k$ are, respectively, the orbital and principal quantum numbers of the $k$th orbital.

In this section, a derivation directly accounting for electron correlation effects for many-electron atoms *by construction* is presented following the ideas of [125]. For reasons that should become clear after the derivation, only the case of $m = 0$ was elaborated in [125], with the general formulas appearing here for the first time.

This derivation relies on the fact that $(\ell m)$th spherical polarizability is written as the shift in the $(\ell m)$th multipole of the charge density due to an $\ell' m'$ perturbing field. The first-order change in the electron density can be written by minimizing an auxiliary Lagrangian, giving the connection between the perturbation and the corresponding response of the system under consideration.

Calculating the $2^l$-pole polarizability within perturbation theory is equivalent in obtaining the expectation value

$$\alpha_{l,m,l',m'} = -2 \langle \Psi_0 | \hat{Q}_{l,m} | \chi_{l',m'} \rangle , \tag{3.20}$$

where $\hat{Q}_{l,m}$ is the $l$th multipole operator and $\chi_l$ is the first order change of the ground-state wavefunction $\Psi_0$, obtained from Rayleigh-Schrödinger perturbation theory as

$$(\hat{\mathcal{H}}_0 - E_0)\chi_{l',m'} + \hat{V}_{l',m'}\Psi_0 = 0 \ . \tag{3.21}$$

This approach is equivalent to minimizing the Lagrangian

$$L[|\chi_{l'm'}\rangle] = \langle\chi_{l'm'}|\hat{\mathcal{H}}_0 - E_0|\chi_{l'm'}\rangle + \langle\chi_{l'm'}|\hat{V}_{l'm'}|\Psi_0\rangle + \langle\Psi_0|\hat{V}_{l'm'}|\chi_{l'm'}\rangle \ . \tag{3.22}$$

For the first order correction in the wavefunction, a variational multiplicative ansatz is proposed of the form

$$\chi_{l'm'} = \lambda_{l',m'}V_{l',m'}\Psi_0 \ . \tag{3.23}$$

Then, the multiplicative constant is found from Euler's equation for the functional to be

$$\frac{\partial L(\lambda_{l'm'})}{\partial \lambda_{l'm'}} = 0 \quad \Rightarrow \quad \lambda_{l'm'} = -\frac{\langle\Psi_0|\|\hat{Q}_{l'm'}\|^2|\Psi_0\rangle}{\langle\Psi_0|\hat{Q}_{l'm'}(\hat{H}_0 - E_0)\hat{Q}_{l'm'}|\Psi_0\rangle} \ , \tag{3.24}$$

where

$$\|\hat{Q}_{l'm'}\|^2 = \hat{Q}_{l'm'}^{\dagger}\hat{Q}_{l'm'} = \sum_{ij=1}^{N}(-1)^{m'}(r_i r_j)^{l'}Y_{l'(-m')}(\theta_i,\phi_i)Y_{l'm'}(\theta_j,\phi_j)|\boldsymbol{r}_1,...,\boldsymbol{r}_N\rangle\langle\boldsymbol{r}_1,...,\boldsymbol{r}_N| =$$

$$= \left[\sum_{i=1}^{N}(r_i)^{2l'}\sum_{L=1}^{2l'}(-1)^{m'}C_{l'0l'0}^{L0}C_{l'm'l'(-m')}^{L0}Y_{L0}(\theta_i,\phi_i)\right] \ . \tag{3.25}$$

Finally, the polarizability is given as

$$\alpha_{lm,l'm'} = \frac{16\pi}{2l'+1}\frac{\langle\Psi_0|(\hat{Q}_{l'm'})^2|\Psi_0\rangle\,\langle\Psi_0|\hat{Q}_{l'm'}^2|\Psi_0\rangle}{\langle\Psi_0|\hat{Q}_{l'm'}(\hat{H}_0 - E_0)\hat{Q}_{l'm'}|\Psi_0\rangle} \ . \tag{3.26}$$

In the next step, the denominator is converted to

$$\langle \hat{Q}^{\dagger}_{l'm'}(\hat{\mathcal{H}}_0 - E_0)\hat{Q}_{l'm'}\rangle_{\Psi_0} = \frac{1}{2}\langle \hat{C}\rangle_{\Psi_0} \,, \tag{3.27}$$

with

$$\hat{C} = \sum_{i=1}^{N} \frac{\hbar^2}{2M_i}(\boldsymbol{\nabla}_{\mathbf{r}_i}\, \hat{Q}^{\dagger}_{l'm'}) \cdot (\boldsymbol{\nabla}_{\mathbf{r}_i}\, \hat{Q}_{l'm'}) \,. \tag{3.28}$$

This gradient can be evaluated for an arbitrary $m$ as

$$\begin{aligned}
\boldsymbol{\nabla}_{\boldsymbol{r}_i}\hat{Q}_{l'm'} = Z_i e r_i^{l-1}\Bigg[ &lY_{lm}(\theta_i, \phi_i)\boldsymbol{e}_r + \Bigg(m\,\mathrm{ctg}(\theta_i)Y_{lm}(\theta_i, \phi_i) + \\
&+ \exp(-i\phi_i)\frac{\sqrt{\Gamma(1+l-m)\Gamma(2+l+m)}}{\sqrt{\Gamma(l-m)\Gamma(1+l+m)}}Y_{l(m+1)}(\theta_i, \phi_i)\Bigg)\boldsymbol{e}_\theta + \\
&+ im\,\csc(\theta_i)Y_{lm}(\theta_i, \phi_i)\boldsymbol{e}_\phi \Bigg]|r_1, \theta_1, \phi_1, ..., r_1, \theta_N, \phi_N\rangle\langle r_1, \theta_1, \phi_1, ..., r_N, \theta_N, \phi_N|
\end{aligned} \tag{3.29}$$

and

$$\begin{aligned}
\boldsymbol{\nabla}_{\boldsymbol{r}_i}\hat{Q}^{\dagger}_{l'm'} = Z_i e r_i^{l-1}\Bigg[ &lY^*_{lm}(\theta_i, \phi_i)\boldsymbol{e}_r - \Bigg(m\,\mathrm{ctg}(\theta_i)Y_{lm}(\theta_i, \phi_i) + \\
&+ \exp(-i\phi_i)\frac{\sqrt{\Gamma(2+l-m)\Gamma(1+l+m)}}{\sqrt{\Gamma(1+l-m)\Gamma(l+m)}}Y^*_{l(m-1)}(\theta_i, \phi_i)\Bigg)\boldsymbol{e}_\theta + \\
&- im\,\csc(\theta_i)Y^*_{lm}(\theta_i, \phi_i)\boldsymbol{e}_\phi \Bigg]|r_1, \theta_1, \phi_1, ..., r_1, \theta_N, \phi_N\rangle\langle r_1, \theta_1, \phi_1, ..., r_N, \theta_N, \phi_N| \,.
\end{aligned} \tag{3.30}$$

The previous equations yield to

$$\hat{C}_{m'} = \sum_{i=1}^{N} \frac{Z_i^2 e^2 \hbar^2 r_i^{2l-2}}{M_i}\Bigg[(l'^2 + m'^2\csc^2(\theta_i))\sum_{L=0}^{2l'}C^{L0}_{l'0l'0}C^{L0}_{l'm'l'(-m')}Y_{L0}(\theta_i, \phi_i)\Bigg] \,. \tag{3.31}$$

For a closed atom, the value of $\alpha$ is independent of $m$, therefore only the case of $m = 0$ was presented

in [125]. Since open-shell atoms generally do have non-zero magnetic quantum numbers, the full equations above are not practically feasible to be evaluated in quantum chemical codes. The $m = 0$ case, however, gives a workable formula

$$\hat{C}_0 = C = \left(\frac{4\pi}{2l+1}\right)^2 \sum_{i=1}^{N} r_i^{2l-2} \left(l^2 Y_{l,0}^2(\theta_i) + l(l+1)|Y_{l,1}(\theta_i, \phi_i)|^2\right) . \tag{3.32}$$

Combining Eqs. (3.26) and (3.32) results in the following expression for the multipole polarizability of the $l$th rank

$$\alpha_l = \frac{4\langle R^{2l}\rangle^2}{(2l+1)^2 l\langle r^{2l-2}\rangle} , \tag{3.33}$$

with

$$\langle r^n\rangle = \langle 0|\hat{r}^n|0\rangle \quad \text{and} \quad \langle R^{2n}\rangle = 4\pi \left\langle 0 \left| \sum_{i,j=1}^{N_e} r_i^n Y_{l,0}(\theta_i) r_j^n Y_{l,0}(\theta_j) \right| 0 \right\rangle . \tag{3.34}$$

For the specific case of dipolar response ($\ell=1$) this gives

$$\alpha = \frac{4\langle R^2\rangle^2}{9N_e} , \tag{3.35}$$

which, compared to Eq. (3.12), differs in the definition of the size operator $L := R$, which now is a two-electron operator able to describe electron correlation contributions to polarizability. A second difference between the two formulas is the pre-factor, as the constant in [19] was chosen to be unity for the harmonic oscillator, while Eq. (3.35) is exact for the hydrogen atom. Nevertheless, the four-dimensional scaling of the dipole polarizability is also confirmed using the correlated size descriptor outlined here.

The ground state wavefunction of free atoms is spherically symmetric, belonging to the SO(3) point group, which means that if the orbitals are properly symmetrized in *ab initio* calculations, only integrals of the form $\langle 0|r_i r_j|0\rangle$ must be evaluated. Having access to the two-electron reduced density matrix $\Gamma_{\mu\nu}^{\sigma\gamma}$ defined from the two-electron reduced density function (the trace of all but two electron coordinates of the form $\Psi^*\Psi$) in the basis of molecular orbitals $\phi$ as

$$\Gamma(r_1, r_1', r_2, r_2') = \frac{N!}{(N-2)!} \int dr_3 \dots dr_N \Psi^*(r_1, r_2, \dots, r_N) \Psi(r_1', r_2', \dots, r_N) \; ; \tag{3.36}$$

$$\Gamma_{\substack{\mu\nu \\ \sigma\gamma}} = \int dr_1 \, dr_1' \, dr_2 \, dr_2' \phi_\mu(r_1) \phi_\sigma(r_2) \Gamma(r_1, r_1', r_2, r_2') \phi_\nu(r_1') \phi_\gamma(r_2') \; , \tag{3.37}$$

such matrix elements are written (for $i \neq j$) as

$$\langle 0 | r_i r_j | 0 \rangle = \sum_{\substack{\mu\nu \\ \sigma\gamma}} \Gamma_{\substack{\mu\nu \\ \sigma\gamma}} r_{\mu\nu} r_{\sigma\gamma} \; , \tag{3.38}$$

with $r_{\mu\nu}$ being the matrix representation of the single-electron operator $\hat{r}$ in the basis of molecular orbitals $\phi$

$$r_{\mu\nu} = \int \phi_\mu^*(r_1) \mathbf{r}_1 \phi_\nu(r_1) \, dr_1 \; , \tag{3.39}$$

Equations (3.38)–(3.39) were implemented in PYSCF [74], where a native option of symmetrizing all integrals by SO(3) group is possible. The values of the uncorrelated $(\Delta r)^2$ as well as $\langle R^2 \rangle$ calculated with UCISD using the basis set aug-cc-pVTZ for noble gas atoms are shown in Table 3.1

| Atom | $(\Delta r^2)$ | $\langle R^2 \rangle$ | $(\Delta r^2)^2 / \langle R^2 \rangle^2$ | $n$ |
|------|------|------|------|---|
| He | 2.40 | 2.27 | 1.12 | 1 |
| Ne | 9.61 | 5.76 | 2.78 | 2 |
| Ar | 26.1 | 13.9 | 3.53 | 3 |
| Kr | 39.7 | 20.1 | 3.90 | 4 |

Table 3.1: Correlated and uncorrelated atomic "sizes" of noble gas atoms (in atomic units). The ratio of the square of the sizes is also shown, which is close to the principal quantum number $n$ of the valence shell in all cases.

One can see in Table 3.1 that the difference between the correlated and the uncorrelated size descriptor increases with the principal quantum number of the valence shells of the atoms. In Ref. [19], the principal quantum number was taken into account by introducing the empirical factor $\eta$, which is directly proportional to it. Remarkably, the ratio of the fourth powers of the two size descriptors closely follows the principal quantum number of the valence shell of each atom. This not only shows that electron correlation is a major factor in determining the $\eta$ parameter, but also

hints that electron correlation effects scale proportionally with dipole polarizability. Moreover, it can be concluded that expressing the polarizability using a correlated descriptor of the size enables a rigorous derivation of the connection between the $\eta$ parameter and the primary quantum number, previously introduced empirically in [19].

Finally, it should be noted (as also noticed in [125]) that the variational derivation presented here can compliment the perturbative Unsøld approach presented in Sect. 3.1.1 by giving an expression for the average excitation energy. On the other hand, the constant $C$ in Eq. (3.12) is calculated by the ratio of the full polarizability and the approximate expression, so it accounts for all other sources of errors, such as the choice of the variational ansatz and electrostatic screening effects. However, since the correlated descriptor presented in this section is based on the 2-electron reduced density matrix, it could, in principle, also be generalized for molecules, enabling us to define the quantum mechanical sizes of molecules consistently with the four-dimensional scaling law of the dipole polarizability.

In summary, we have established a formula for the dipole polarizability, $\alpha = C(4\mu q^2/\hbar^2)L^4$, valid for QM systems of varying spatial dimension, symmetry, excitation state, and number of particles. The universality of the $L^4$ scaling for $\alpha$ is connected to the unified QM metric $L$ measuring fluctuations in the position of the particles in terms of system parameters. On the contrary, the dimensionless coefficient $C$ reflects only the qualitative properties of the eigenvalue spectrum of each system. The geometric scaling of the polarizability for a system in its ground state is solely determined by the ground-state wavefunction, whereas the effect of excited states is encoded in $C$ only. Here, it was also shown using variational theory that the scaling law can be extended to many-electron systems simply by expressing the QM size of the system using correlated descriptors.

Another interesting finding is that the polarizability expression is directly proportional to the particle mass, which is opposite to the classical picture where the polarizability vanishes for infinite particle mass. The derived four-dimensional formula has many practical applications. It can be directly used to improve DFT-based methods for van der Waals interactions [35, 97], parametrize polarizable force fields [115–117], or efficiently calculate dynamic spectroscopic observables based on the polarizability (*i.e.*, Raman spectra and sum-frequency generation) [98–102]. All of these applications are based on an efficient and accurate evaluation of the polarizability from ground-state electron density, as enabled by our unified formula.

## 3.2 Quantum Drude Oscillators for Atomic Responses

The development of predictive model Hamiltonians that can describe various properties of realistic molecules and materials is a cornerstone of modern physics [126] and chemistry [127]. Of these model systems, the quantum Drude oscillator (QDO) is arguably the most powerful Hamiltonian for accurate and efficient modeling of the atomic and molecular response [2, 87, 115–117, 126, 128–130].

Despite the wide applicability of the coupled QDO model, its success in describing real atoms remains fundamentally unexplained, and the optimal mapping between atoms and oscillators has not been established. In [44], an optimized parametrization (OQDO) was developed where the parameters are fixed by using only the well-known atomic dipolar response properties. For direct comparison with the published results, this Chapter will differ from the rest of the thesis by not using atomic units but explicitly writing out the ISQ format for electrostatic quantities.

### 3.2.1 Description of the QDO Formalism

Within the quantum Drude oscillator (QDO) model, the response of all valence electrons is represented by a single quasiparticle (*drudon*) with a negative charge $-q$ harmonically bound to a positively charged pseudo-nucleus of charge $q$ with a characteristic frequency $\omega$, having a reduced mass $\mu$. The Hamiltonian of a single QDO is given by the well-known harmonic oscillator [117]

$$\hat{\mathcal{H}}_0 = -\frac{\hbar^2}{2\mu} \, \boldsymbol{\nabla}_{\mathbf{r}}^2 \; + \; \frac{1}{2} \, \mu\omega^2 \mathbf{r}^2 \; . \tag{3.40}$$

The corresponding eigenvalues and eigenfunctions, respectively, are given by

$$E_{\{n_x n_y n_z\}} = \hbar\omega \left( n_x + n_y + n_z + \frac{3}{2} \right) \, , \quad E_{n_x} = \hbar\omega \left( n_x + \frac{1}{2} \right) \, , \quad n_x = 0, 1, 2, ... \tag{3.41}$$

and

$$\Psi_{\{n_x n_y n_z\}}(\mathbf{r}) = \Psi_{n_x}(x)\Psi_{n_y}(y)\Psi_{n_z}(z) \, , \quad \Psi_{n_x}(x) = \frac{1}{\sqrt{2^{n_x} \, n_x!}} \left( \frac{\mu\omega}{\pi\hbar} \right)^{1/4} e^{-\frac{\mu\omega x^2}{2\hbar}} \, H_{n_x}\left( \sqrt{\frac{\mu\omega}{\hbar}}x \right) \, , \tag{3.42}$$

where $H_n(z) = (-1)^n e^{z^2} \frac{d^n}{dz^n} e^{-z^2}$ are the Hermite polynomials.

In order to describe electromagnetic interactions of a QDO with other species or external fields, the drudon acquires a negative charge $(-q)$ opposite to the charge $q$ of the related pseudo-nucleus possessing an infinite mass. The introduction of electric charges completes the QDO parameter set to $\{q, \mu, \omega\}$, making it different from the quantum harmonic oscillator (possessing just $\mu$ and $\omega$).

Within the QDO model [117] one obtains the following multipole polarizabilities

$$\alpha_l = \left(\frac{q^2}{\mu\omega^2}\right)\left[\frac{(2l-1)!!}{l}\right]\left(\frac{\hbar}{2\mu\omega}\right)^{l-1} . \tag{3.43}$$

For example, the first few polarizabilities as

$$\alpha_1 = \frac{q^2}{\mu\omega^2} , \quad \alpha_2 = \frac{3\hbar}{4\mu\omega}\alpha_1 , \quad \alpha_3 = \frac{5\hbar^2}{4(\mu\omega)^2}\alpha_1 \tag{3.44}$$

and the first three dispersion coefficients (solely expressed in terms of the QDO parameters)

$$C_6 = \frac{3}{4}\frac{\hbar q^4}{\mu^2\omega^3} \quad C_8 = \frac{5\hbar}{\mu\omega}C_6 , \quad C_{10} = \frac{245\hbar^2}{8(\mu\omega)^2}C_6 , \tag{3.45}$$

delivering the leading-order contributions to the dispersion energy, $E_{\text{disp}} = -\sum\limits_{n=3,4,\dots} C_{2n}/R^{2n}$ .

In the presence of an external (uniform) electric field $\mathbf{E}$, the QDO Hamiltonian modifies to

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2\mu}\nabla_{\mathbf{r}}^2 + \frac{1}{2}\mu\omega^2\mathbf{r}^2 - q\,(\mathbf{r}\cdot\mathbf{E}) , \tag{3.46}$$

which can be straightforwardly diagonalized by means of the coordinate transformation (see also [131]): $\mathbf{r} = \tilde{\mathbf{r}} + q\mathbf{E}/\mu\omega^2$. Consequently, the ground-state density acquires shifted coordinate

$$\rho(\mathbf{r}) = \rho_0(\mathbf{r}) = \frac{1}{(\sqrt{2\pi}\sigma)^3}e^{-\frac{r^2}{4\sigma^2}} , \quad \rho_E(\mathbf{r}) = \rho_0(\tilde{\mathbf{r}}) = \frac{1}{(\sqrt{2\pi}\sigma)^3}e^{-\frac{(\mathbf{r}-\alpha_1\mathbf{E}/q)^2}{4\sigma^2}} , \quad \sigma = \sqrt{\frac{\hbar}{2\mu\omega}} . \tag{3.47}$$

Substituting the above densities into Eq. (8) of the manuscript [44] leads to Eq. (8) there (Eq. (3.56) in this thesis).

## 3.2.2 Optimized QDO Parametrization

The quantum Drude oscillator (QDO) is an efficient yet accurate coarse-grained approach that has been widely used to model electronic and optical response properties of atoms and molecules, as well as polarization and dispersion interactions between them. Three effective parameters (frequency, mass, charge) fully characterize the QDO Hamiltonian and are adjusted to reproduce response properties. However, the soaring success of *coupled* QDOs for many-atom systems remains fundamentally unexplained and the optimal mapping between atoms/molecules and oscillators has not been established. Here, we present an optimized parametrization (OQDO) where the parameters are fixed by using only dipolar properties. For the periodic table of elements as well as small molecules, our OQDO model accurately reproduces atomic (spatial) polarization potentials and multipolar dispersion coefficients, elucidating the high promise of the model presented in the development of next-generation quantum-mechanical force fields for (bio)molecular simulations.

As mentioned before, the development of predictive model Hamiltonians that can describe various properties of realistic molecules and materials is a cornerstone of modern physics [126] and chemistry [127]. The quantum Drude oscillator (QDO) is arguably the most powerful Hamiltonian (see Sec. 3.2.1) for accurate and efficient modeling of the atomic and molecular response [2, 87, 115–117, 126, 128–130].

Within the coarse-grained QDO model, the response of valence electrons is described *via* a quasi-particle *drudon* with a negative charge $-q$ and mass $\mu$, harmonically bound to a positively charged pseudonucleus of charge $q$ with a characteristic frequency $\omega$. The many-body extension of the QDO model (the *coupled* QDO model) has been widely employed to study both molecules and materials, including their electronic [39, 97] and optical [132] properties, polarization [133, 134], dispersion [42, 128, 131, 133, 135–141], and exchange [28, 114, 142] interactions, as well as a wealth of non-additive field effects in quantum mechanics [141, 143] and quantum electrodynamics [131, 144].

Coupled QDOs are also extensively used in the development of van der Waals (vdW) density functionals [35, 42, 145], quantum mechanical [117, 126] and polarizable force fields [146–150] as well as recent machine learning force fields [151, 152]. Despite such a wide applicability of the coupled QDO model, its success in describing real atoms remains fundamentally unexplained and the optimal mapping between atoms and oscillators has not been established. In this Section, which contains the publication [44], an optimized parametrization (OQDO) is developed where the

parameters are fixed using only the well-known atomic dipolar properties. Remarkably, OQDO reproduces spatial atomic polarization potentials and atomic multipolar dispersion coefficients. Our OQDO model for atoms and small molecules also paves the way for the development of next-generation quantum-mechanical force fields for (bio)molecular simulations.

The three parameters $\{q, \mu, \omega\}$ fully define the QDO, and three atomic response properties could be chosen to fix them, which means that the choice of QDO parameters is not unique. In addition, all QDO response properties – multipolar polarizabilities and dispersion coefficients – are uniquely fixed by the three parameters *via* closed-form relations [117]. The static dipole polarizability of a QDO, $\alpha_1 = q^2/\mu\omega^2$, conveniently combines all three parameters, and it is natural to set this expression to the reference atomic $\alpha_1$. The QDO expression for the dipole-dipole dispersion coefficient $C_6 = \frac{3}{4}\hbar\omega\alpha_1^2$ is identical to the London formula and allows fixing $\omega$ if the reference atomic values of $C_6$ and $\alpha_1$ are given. Since $\alpha_1$ and $C_6$ are accurately known for all elements in the periodic table [46, 153, 154], they form a baseline for the QDO parametrization. However, one more condition is required to obtain $\{q, \mu, \omega\}$, for which different constraints can be imposed. A reasonable idea is to fix $q = 1$ a.u., since a QDO should reproduce the response of electrons. This results in the fixed-charge QDO (FQDO)

$$q = 1 \ , \quad \omega = 4C_6/3\hbar\alpha_1^2 \ , \quad \mu = 9\hbar^2\alpha_1^3/16C_6^2 \quad . \tag{3.48}$$

However, fixing $q$ and using QDO recursion relations for high-order response usually yields large errors in the multipolar response properties (see Fig. 2 and Refs. 111, 155). A more rigorous approach was suggested by Jones *et al.* [117] by employing the dipole-quadrupole dispersion coefficient $C_8$. The mapping $\{\alpha_1, C_6, C_8\} \rightarrow \{q, \mu, \omega\}$ yields the Jones QDO (JQDO) parametrization scheme

$$q = \sqrt{\mu\omega^2\alpha_1} \ , \quad \omega = 4C_6/3\hbar\alpha_1^2 \ , \quad \mu = 5\hbar C_6/\omega C_8 \ . \tag{3.49}$$

The JQDO approach improves the multipolar response over the FQDO model, while simulations using the coupled JQDO model captured many remarkable properties of the bulk water and its surface [149, 156]. However, the $C_8$ dispersion coefficient is not directly measurable, and accurate *ab initio* calculations of quadrupole ($\alpha_2$) and octupole ($\alpha_3$) polarizabilities and $C_8 - C_{10}$ dispersion coefficients are technically feasible currently only for closed-shell species (noble gas atoms and

small molecules) or alkali and alkaline earth atoms with $s$ valence shells [157–160]. For other open-shell atoms (containing $p$, $d$, or $f$ valence shells), convergence of quantum-chemical response calculations becomes a technical hurdle [161] (see also Sect. 2.2.6). Thus, using higher-order atomic response properties does not lead to a parametrization that would be universally applicable across the periodic table as well as for small molecules.

Here, we introduce an optimized QDO parametrization (OQDO), where we effectively map the dipolar atomic quantities $\{\alpha_1, C_6\}$ to the oscillator parameters. The third parameter is fixed by using the force balance equation for vdW-bonded dimers recently derived [28, 114, 142]. Two equations for $q$ and $\omega$ follow the JQDO scheme, while the third one is replaced with a transcendental equation for a product $\mu\omega$ to be solved numerically (*vide infra*)

$$
\mu = \frac{5\,\hbar\,C_6}{\omega\,C_8} \rightarrow \exp\left(\frac{2\mu\omega R_{\text{vdW}}^2}{\hbar}\right) = \frac{2^7 \cdot (\alpha_{\text{fsc}}^{-1/3} a_0)^4}{(3\hbar/\mu\omega)^2} \,, \tag{3.50}
$$

where $\alpha_{\text{fsc}} = e^2/4\pi\varepsilon_0\hbar c$ was suggested to be related to the fine-structure constant and $a_0$ is the Bohr radius. The vdW radius ($R_{\text{vdW}}$) is calculated *via* the universal formula connecting it with the dipole polarizability

$$
\alpha_1(R_{\text{vdW}}) = (4\pi\varepsilon_0)\,R_{\text{vdW}}^7/(\alpha_{\text{fsc}}^{-1/3} a_0)^4 \,, \tag{3.51}
$$

as established in Ref. 114 for atoms in the periodic table. Comparing it with its counterpart

$$
\alpha_1(\mu\omega, R_{\text{vdW}}) = (4\pi\varepsilon_0)\,R_{\text{vdW}}^7 \frac{2^7\,(\mu\omega/3\hbar)^2}{\exp(2\mu\omega R_{\text{vdW}}^2/\hbar)} \,, \tag{3.52}
$$

which was obtained within the QDO model [28, 142], delivers Eq. (3.50) to determine $\mu$ from $\mu\omega$.

The QDO Hamiltonian effectively captures the integrated atomic response. However, when modeling molecules or solids, coupled QDOs must properly describe noncovalent interactions between atoms. Considering two fragments $i$ and $j$ and using interatomic perturbation theory [4, 162], the interaction energy can be written as the integrated product of the electron density of the moiety $i$ with the electric potential generated by the moiety $j$ [4, 163]
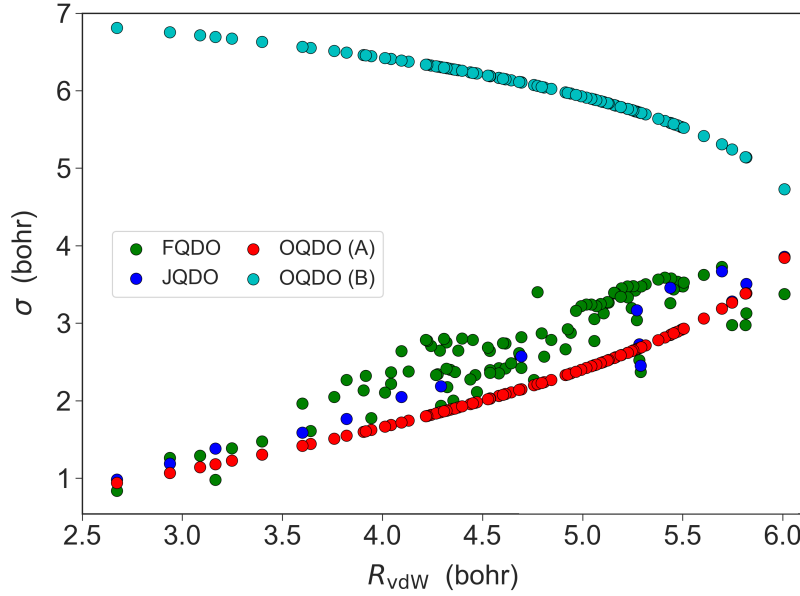
Figure 3.2: Correlation between $R_{vdW}$ and the QDO size $\sigma$ (within the three QDO parametrizations) for 102 elements in the periodic table. The schemes OQDO(A) and OQDO(B) correspond to two solutions of Eq. (3.58).

$$E_{\text{int}} = \int \rho_i(\mathbf{r}) V_j(\mathbf{r}) \, d^3\mathbf{r} \, . \tag{3.53}$$

This formula is valid for all noncovalent interactions: electrostatics, induction, exchange-repulsion, and dispersion. Its validity is evident for the former two cases [4, 163], and it was shown that exchange [164] and dispersion [64, 165, 166] interactions can be represented using the form of Eq. (3.53) with $\rho$ and $V$ being *effective* quantities different from free-atom counterparts.

The argument that dispersion interactions can be written using Eq. (3.53) goes back to Feynman's consideration of molecular forces [165], which was further elaborated by Hunt [64] with a focus on dispersion forces and finally extended to dispersion energies with a demonstration of its validity for real molecules and materials [167–170].

Response properties are given by variations of $E_{\text{int}}$ as

$$\delta E_{\text{int}} = \int \left( \delta\rho_i(\mathbf{r}) V_j(\mathbf{r}) + \rho_i(\mathbf{r}) \delta V_j(\mathbf{r}) \right) d^3\mathbf{r} \, . \tag{3.54}$$

For an external electric field $\mathbf{E}$, which can also model the effect of the environment, $\delta\rho(\mathbf{r}) = \rho_E(\mathbf{r}) - \rho(\mathbf{r})$, where $\rho_E(\mathbf{r})$ is the electron density under the external field. Then, the dominant contribution to $\delta V_j(\mathbf{r})$ is generated by the corresponding $\delta\rho_j(\mathbf{r}')$ via the polarization potential [171]

$$V_{\text{pol}}(\mathbf{r}) = -\frac{1}{4\pi\varepsilon_0} \int \frac{\rho_E(\mathbf{r}') - \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}^3\mathbf{r}' \,, \tag{3.55}$$

which describes the change in the electrostatic potential of the system as a result of the polarization of its charge density by the presence of another moiety (an electric field in this case). For the QDO in a uniform electric field, the integral in Eq. (3.55) can be evaluated analytically as

$$V_{\text{pol}}^{\text{QDO}}(\mathbf{r}) = \frac{-q}{4\pi\varepsilon_0} \left( \frac{\text{erf}(\tilde{r}/\sigma\sqrt{2})}{\tilde{r}} - \frac{\text{erf}(r/\sigma\sqrt{2})}{r} \right) \,, \tag{3.56}$$

where $\tilde{\mathbf{r}} = \mathbf{r} - \alpha_1 \mathbf{E}/q$ is the field-induced oscillator coordinate and $\sigma = \sqrt{\hbar/2\mu\omega}$ is the QDO spread [19].

In the Supporting Information of [44], we present $V_{\text{pol}}^{\text{QDO}}(\mathbf{r})$ compared to $V_{\text{pol}}(\mathbf{r})$ calculated for 21 atoms (between H – Ca and Kr) within hybrid density-functional theory DFT-PBE0 [172–176] shown to yield a highly accurate description of electronic response [177] comparable to coupled-cluster calculations. The strength of the electric field was chosen individually for each element depending on its reference static dipole polarizability [46, 154] so that the field-induced dipole moment is set as $\mathbf{d} = \alpha_1 \mathbf{E} = 0.01$ a.u., for all atoms.

To demonstrate that the agreement between the OQDO and the DFT polarization potential is not due to our choice of the PBE0 functional, we have also calculated this quantity using MP2 and CCSD(T). *Ab initio* curves shown in Fig. 3.3 were calculated using an orbital-optimization step from converged wave functions, as implemented in ORCA [178–180] for both MP2 and CCSD(T) projected densities. The basis set used for the calculation was aug-cc-pVQZ (uncontracted for the CCSD(T) curves), except for Si in MP2, where aug-cc-pVTZ was used. As shown in Fig. 3.3, our PBE0 results are in good agreement with the performed MP2 and CCSD(T) calculations.

Before comparing $V_{\text{pol}}(\mathbf{r})$ for real atoms with different QDO flavors, it is instructive to consider which atomic properties can be faithfully captured by a QDO. First, the QDO does not aim to describe static properties of the atomic electron density, but rather its response under applied static and fluctuating fields, as also demonstrated by the insets in Fig. 3.4a,b. The electrostatic potential (ESP) of a QDO is given by $V_{\text{el}}^{\text{QDO}} = -q \cdot \text{erf}(r/\sigma\sqrt{2})/r$, so the charge $q$ determines its magnitude. This explains why $V_{\text{el}}^{\text{FQDO}}$ is in good agreement with $V_{\text{el}}^{\text{DFT}}$ for hydrogen. However, the
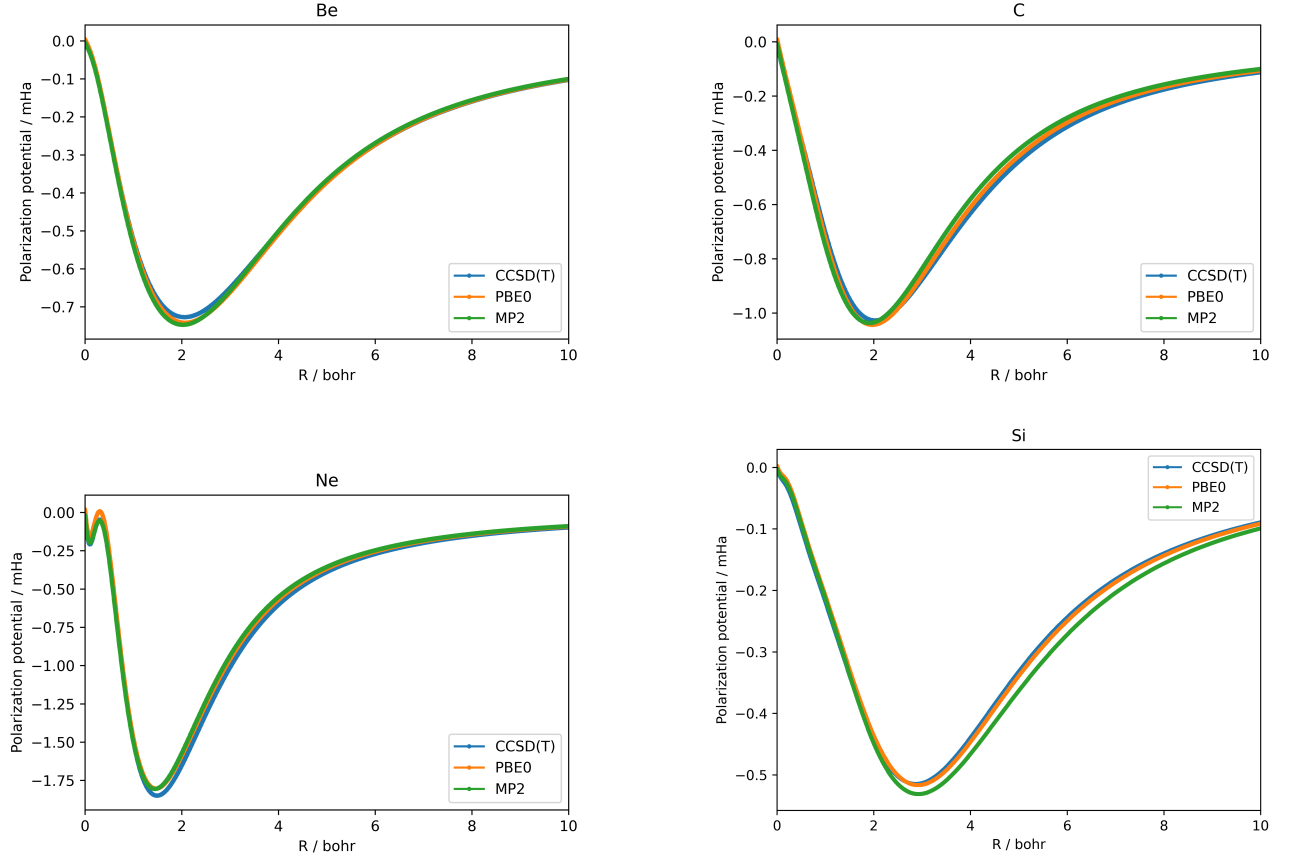
Figure 3.3: Comparison between PBE0 and orbital-optimized MP2/CCSD(T) calculations of the polarization potential, given by Eq. (3.56), performed for Be, C, Ne and Si atoms [44].

QDO model does not describe $V_{el}$ for many-electron atoms because $q \sim 1$ a.u., while the ESP of atoms scales non-linearly with $Z$ (see the example of carbon in the inset of Fig. 3.4b). Second, the harmonic response captured by a QDO model should be sufficient to accurately describe integrated electronic displacements induced by weak fields. However, it is much less clear how well different QDO parametrizations perform for distributed polarization potentials described by Eq. (3.55) for many-electron systems, given the analytical form of $V_{\text{pol}}^{\text{QDO}}(\mathbf{r})$ in Eq. (3.56). To answer this question, in Fig. 3.4 we compare $V_{\text{pol}}$ of real atoms and $V_{\text{pol}}^{\text{QDO}}(\mathbf{r})$ employing the three QDO models discussed above. We used the accurate *ab initio* reference data on $\alpha_1$ and $C_6$ [46, 154, 181] to parametrize FQDO and OQDO. When available, we also used the analogous data on $C_8$ [157–159] to parametrize the JQDO model. We observe that the OQDO model is capable of reproducing the full range of the polarization potential of real atoms with reasonable accuracy, showing significantly
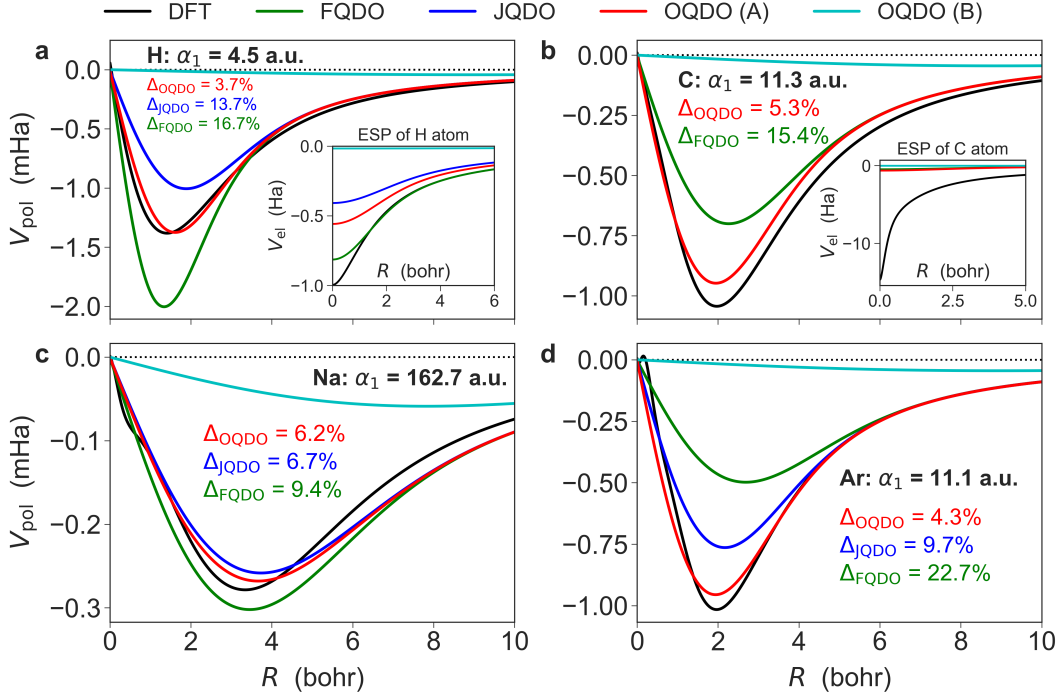
Figure 3.4: Polarization potential curves $V_{\mathrm{pol}}(\mathbf{r})$ calculated with DFT–PBE0 and various QDO parametrizations for (a) hydrogen, (b) carbon (no JQDO values are available), (c) sodium, and (d) argon atoms. The FQDO and JQDO parametrization schemes are described by Eqs. (3.48) and (3.49), respectively. OQDO(A) and OQDO(B) correspond to the two solutions of the transcendental equation given by Eq. (3.50). In all cases, the direction along the applied field was chosen for the plots. Reference values for the dipole polarizability $\alpha_1$ are shown for each element. The numerical values of the normalized root mean squared error ($\Delta$) are displayed for the three QDO flavors For hydrogen and carbon atoms, the unperturbed electrostatic potentials (ESP) $V_{\mathrm{el}}(\mathbf{r})$ are shown as insets, indicating that a QDO captures the response of the atomic electron density, but not the static potential itself [44].

better agreement with the DFT-PBE0 results than FQDO and JQDO. To quantify this, for each atom we calculated the root mean square error (RMSE) of the three QDO curves with respect to the PBE0 reference curves and normalized the RMSE using the equilibrium depth of the PBE0 curve. The OQDO flavor has an error of 8.9% when averaged over 21 atoms, while JQDO and FQDO produce average errors of 13.2% and 15.4%, respectively. We also emphasize that the predictions of the OQDO model remain accurate for many-electron atoms such as noble gases and alkali metals. It is especially reassuring that the OQDO model reproduces the non-linear $V_{\mathrm{pol}}(\mathbf{r})$ curves obtained from DFT calculations without any adjustments. In fact, the OQDO performance is sensitive to variations in the QDO parameters (solutions A or B in Fig. 3.4), so the satisfactory
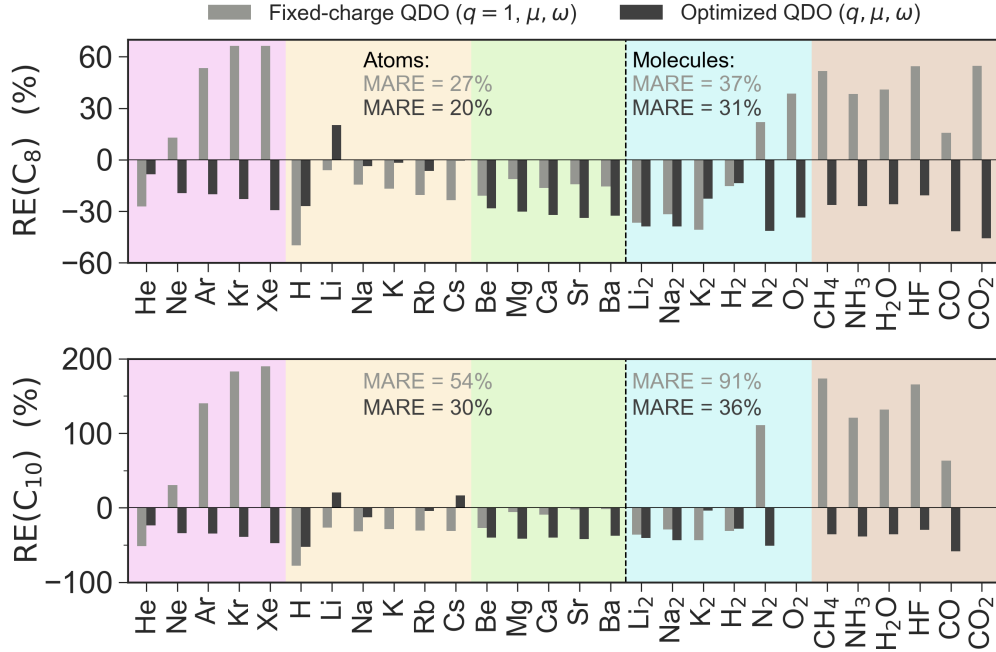
Figure 3.5: Multipolar dispersion coefficients $C_8$ and $C_{10}$ as predicted by FQDO ($q = 1$) and OQDO models. Relative error RE = $(C_j - C_j^{\text{ref}})/C_j^{\text{ref}}$ with respect to *ab initio* reference data [157–160] is plotted. For the two models, numerical values of mean absolute relative errors (MARE) are evaluated separately for atoms and molecules. In the case of $O_2$ and $CO_2$, no reliable *ab initio* reference data for $C_{10}$ could be found.

agreement shows that the chosen OQDO(A) model accurately describes real atoms. The significant differences between the predictions of various parametrizations for $V_{\text{pol}}(\mathbf{r})$ underline the importance of the optimal mapping between the atomic response properties and QDO parameters.

We now discuss the technical aspects of deriving the two solutions of the OQDO model (see OQDO(A) and OQDO(B) in Fig. 3.4), and their connection to real atoms. The starting point is Eq. (3.51) which connects the atomic vdW radius and its dipole polarizability. Within the QDO model, Eq. (3.51) can be written as follows [28]:

$$\alpha_1(\mu\omega, R_{\text{vdW}}) = \frac{2^7 \cdot (4\pi\varepsilon_0) \, R_{\text{vdW}}^7}{(3\hbar/\mu\omega)^2 \, \exp(2\mu\omega R_{\text{vdW}}^2/\hbar)} \; . \tag{3.57}$$

The OQDO parametrization imposes that the product $\mu\omega$ in Eq. (3.57) delivers the same $R_{\text{vdW}}$ as in Eq. (3.51), for $\alpha_1(\mu\omega, R_{\text{vdW}}) = \alpha_1(R_{\text{vdW}})$.

For simplicity, we rewrite Eq. (3.57) in terms of the dimensionless variable $x$ as

$$x = a\, e^{b\,x} \ , \quad x = \mu\,\omega\,a_0^2/\hbar = a_0^2/2\sigma^2 \ , \tag{3.58}$$

with the dimensionless coefficients $a$ and $b$ given by

$$a = \frac{3\,\alpha_{\text{fsc}}^{\frac{2}{3}}}{8\sqrt{2}} \ , \quad b = \frac{R_{\text{vdW}}^2}{a_0^2} = \frac{(\alpha_1/4\pi\varepsilon_0)^{\frac{2}{7}}}{\alpha_{\text{fsc}}^{\frac{8}{21}}a_0^{\frac{6}{7}}} \ , \tag{3.59}$$

where we used Eq. (3.51) to express $R_{\text{vdW}}$ in terms of $\alpha_1$. For all elements in the periodic table, we find that Eq. (3.58) has two solutions, A and B. This is illustrated by the inset of Fig. 3.2 for the case of Ar. It is instructive to consider that Eq. (3.58) has one solution when the polarizability of an atom is equal to the critical value

$$\alpha_1^{(c)}/4\pi\varepsilon_0 = \left(8\sqrt{2}/3e\right)^{\frac{7}{2}}\alpha_{\text{fsc}}^{-1}a_0^3 \approx 431 \text{ a.u.}, \tag{3.60}$$

which is greater than the largest atomic dipole polarizability ($\alpha_1 \approx 400$ a.u.) of Cs [154]. The existence of two solutions extends beyond the employed QDO model. We obtained an analogous result by using the Tang-Toennies potential [182] with the repulsive interaction treated by the Born-Mayer form (see the Supporting Information in [44]).

Since the OQDO frequency is fixed by the second condition of Eqs. (3.48) and (3.49), the solutions A and B for the product $\mu\omega$ differ in both mass and charge, giving quite different results. First, the $V_{\text{pol}}^{\text{QDO}}(\mathbf{r})$ constructed from solution B do not resemble the DFT potentials, while A is in good agreement with them (Fig. 3.4). Second, the overlap integral $S = \exp\left(-\frac{\mu\omega}{2\hbar}R_{\text{eq}}^2\right)$ between two QDOs at their equilibrium distance $R_{\text{eq}} = 2R_{\text{vdW}}$ is significantly larger for solution B, which violates the initial assumption used to derive Eq. (3.57) that $S$ is small at $R_{\text{eq}}$ [28]. Third, the QDO length $\sigma$ constructed from solution A follows the same periodic trend as the atomic vdW radii, whereas solution B does not seem to correlate well (see Fig. 3.6). Therefore, throughout this work, we refer to solution A as the optimized parametrization. For 102 atoms, the full set of QDO parameters corresponding to both solutions A and B is presented in the Supporting Information of [44] together with the reference values of $\{\alpha_1, C_6\}$. Another noteworthy property of the OQDO
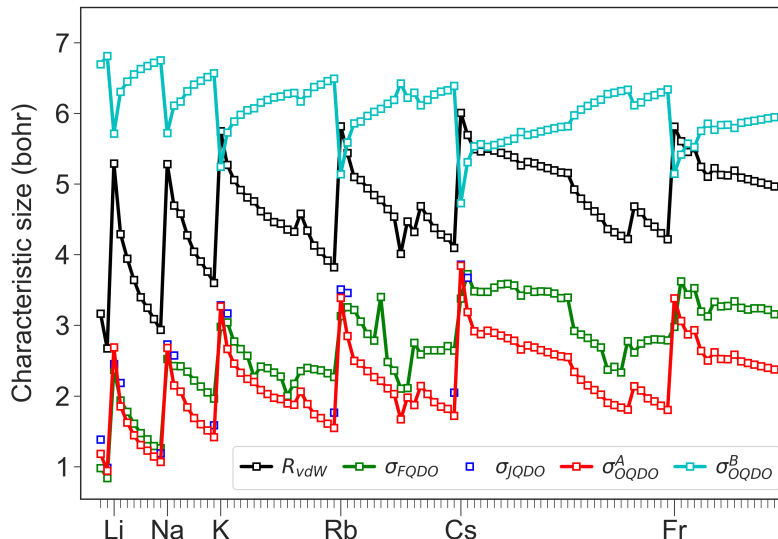
Figure 3.6: Periodic variations of the QDO size $\sigma = \sqrt{\hbar/2\mu\omega}$ with the atomic number for the three different parametrizations, as compared to the atomic van der Waals radii ($R_{\text{vdW}}$) which are evaluated *via* Eq. (3.2) using reference atomic polarizabilities [44].

model (see Fig. 3.2) is a quasi-linear correlation between the QDO length (model quantity) and the atomic vdW radius (physical observable). In fact, these quantities should be connected *via* the dipole polarizability [19, 28]. This property is not captured well by the FQDO or the JQDO models.

For practical calculations of the vdW energy and the construction of predictive force fields, the multipolar contributions associated with the $C_8$ and $C_{10}$ coefficients can become relevant [111, 155, 160]. The available reference data for higher-order molecular dispersion coefficients have significant uncertainties. Our careful examination of the literature reporting the reference values of $C_8$ and $C_{10}$ (see 157–160 and references therein) identifies uncertainties of up to 20% for the reference $C_8$ and $C_{10}$ values. Within the QDO formalism, it is straightforward to evaluate these coefficients using closed-form expressions derived by Jones *et al.* [117]. In Fig. 3.5, we present the predictions of $C_8$ and $C_{10}$ by FQDO and OQDO models compared to accurate reference values compiled from the literature [157–160] for a set of 16 atoms (including alkali and alkaline earth metals and noble gases) and 12 small molecules. In general, our results show that the OQDO parametrization improves the dispersion coefficients compared to the FQDO one, reducing the MARE from 31% to 25% for $C_8$ and from 68% to 33% for $C_{10}$ when averaged over all 28 (26 for $C_{10}$) systems. The OQDO model consistently surpasses FQDO in accuracy for all systems, except for alkaline earth metals, where FQDO gives more accurate results. Furthermore, Fig. 3.5 shows that the deviations of OQDO dispersion coefficients from the reference values are consistent in

terms of their sign and magnitude. That is, for most systems FQDO underestimates $C_8$ and $C_{10}$, but roughly for one third of them the dispersion coefficients are overestimated. The largest errors of FQDO are observed for Xe in the cases of $C_8$ (66%) and $C_{10}$ (190%). On the contrary, OQDO consistently underestimates both dispersion coefficients for all systems, except for $C_8$ of Li and $C_{10}$ of Li and Cs. The maximum errors of OQDO are observed for $CO_2$ (46%) in case of $C_8$ and CO (58%) in case of $C_{10}$, which are significantly smaller than the maximum errors of FQDO. The consistency of OQDO errors allows for a straightforward rescaling of dispersion coefficients: with our best rescaling factors, 1.3 for $C_8$ and 1.5 for $C_{10}$, one can decrease the MARE of OQDO to 15% and 22%, respectively, which is consistent with the uncertainty of the reference molecular $C_8$ and $C_{10}$ values.

A more detailed analysis of the dispersion coefficients (including JQDO and OQDO models as well as their scaled versions) can be found in the Supporting Information of [44], where we also discuss the static polarizabilities $\alpha_2$ and $\alpha_3$. The latter becomes less important in the QDO approach where the dispersion coefficients, determining the dispersion energy, are directly expressed in terms of the QDO parameters. It is important to mention that the effects of three-body interactions are captured by the OQDO scheme on equal footing with the JQDO scheme. The accuracy in determination of $C_6$ and $C_9$ coefficients is known to be comparable [166]. In Ref. 117 it was shown that within the QDO model the leading three-body dispersion coefficient is given by $C_9 = \alpha_1 C_6/4$. Thus, with the same reference $\alpha_1$ and $C_6$, there is no difference between the JQDO and OQDO parametrizations.

We presented the OQDO model based on a robust parametrization that only employs dipolar $\alpha_1$ and $C_6$, accurately known for all atoms in the periodic table. The new parametrization scheme eliminates the need for reference higher-order dispersion coefficients and delivers accurate polarization potentials, which improves the description of non-covalent interactions at short distances, keeping the accuracy of the JQDO model for large distances due to the proper dipolar response.

The key point of the proposed parametrization is employing the relation between the dipole polarizability and the vdW radius, both of which are integrated quantities with many-electron effects included. Thus, the OQDO scheme serves as an optimized and efficient mapping between atoms/molecules and oscillators, which substantially advances our ability to model a wide range of response properties of molecules and materials, also paving the way for the development of next generation quantum-mechanical force fields for (bio)molecular simulations.

## 3.3 Conclusions

The properties of polarizability of quantum mechanical models and (real) atoms were studied in this Chapter. Such model systems and free atoms allow one to deeply investigate the fundamental aspects of polarizability. The advantage of studying these systems is twofold: not only it is possible to obtain highly accurate – often exact – reference values, but such models also serve as the building blocks for coarse-grained approaches applicable to realistic complex systems.

It was shown that, contrary to arguments rooted in classical electrostatics, polarizability scales with the characteristic system size to the fourth power for a wide variety of models and real atoms. This scaling law seems to hold for a wide variety of model systems having qualitatively different spectra, as well as for many-electron atoms, provided that a definition accounting for electron correlation effects is applied to such many-body quantum systems. As will be shown in Section 4.2, the scaling law elaborated here is one of the reasons for the success of the model of interacting oscillators.

Recognizing the central role of the quantum Drude oscillator model in atoms-in-molecules approaches, a revised parametrization for atomic response functions was also presented here, showing that knowledge of just dipolar properties is sufficient to predict multipolar as well as spatially distributed responses. It is expected that this new parametrization will provide a more accurate approach for simulations in which quantum Drude oscillators are used as coarse-grained fragments.

# 4. Models for Molecular Polarizability

*Section 4.1 is based on the paper*
*Góger, S. et al., PCCP **2023**, 25, 22211–22222;*
*containing parts reproduced as permitted by the Creative Commons license.*

A goal of high practical importance is to develop the understanding and to create predictive methods for the polarizabilities of materials and large molecules, where long-range interactions play a key role in determining their properties. Smaller organic molecules provide a convenient stepping stone to this goal: the availability of large datasets of reference electronic-structure calculations enables testing of approximate efficient models.

In this Chapter, it is shown how the statistical independence of dipole polarizability and the energy difference between the highest occupied and lowest unoccupied molecular orbital (HOMO–LUMO gap) leads to a design principle that can be utilized when both quantities must be controlled. Notably, correlation between these two properties is expected based on textbook knowledge (see Sect. 4.1.1 for a detailed discussion), so the lack of correlation also unravels an unexpected behavior of polarizability.

Sect. 4.1 contains an adapted version of the publication [29] – To this work, I have contributed by conceptualizing the project, generating and analyzing the dataset as well as being involved in the visual presentation of the results and the writing of the text.

## 4.1   Empirical Correlations of Molecular Polarizability

Understanding the correlations – or lack thereof – between molecular properties is crucial to enable fast and accurate molecular design strategies. In this Chapter, based on [29], we explore the relation between two key quantities describing the electronic structure and chemical properties of molecular systems: the energy gap between the frontier orbitals and the dipole polarizability.

Based on the recently introduced QM7–X dataset [183], augmented with accurate molecular polarizability calculations, as well as analysis of functional group compositions, we show that polarizability and HOMO–LUMO gap are uncorrelated when considering sufficiently extended subsets of the chemical compound space. The relation between these two properties is further analyzed on specific examples of molecules with similar composition as well as homooligomers. Remarkably, the freedom brought by the lack of correlation between the molecular polarizability and the HOMO–LUMO gap enables the design of novel materials, as we demonstrate on the example of organic photodetector candidates.

## 4.1.1   Polarizability and HOMO-LUMO Gap

Data-driven molecular design is an increasingly pursued strategy in chemical physics and computational chemistry. The search for novel molecules with tailored physicochemical properties for a given functionality is continuously motivating the development of a wide variety of computer-aided molecular design approaches [184–186]. The ultimate goal is to establish a feasible protocol that can be used for exploring the chemical compound space (CCS) through systematic targeting of physical properties. Physicochemical quantities, such as color, conductivity, excited state lifetime, electron affinity, ionization potential, and solubility, are commonly used in the design of molecular photosensitizers or optoelectronic devices, for example [187–190]. Given the complexity of a multi-property design task, it is essential to first have a solid grasp of the physical relationships between the various target properties [191].

In this context, two fundamental quantum-mechanical (QM) electronic properties are the optical gap and the molecular dipole polarizability ($\alpha$). The optical gap is an experimental property that measures the energy corresponding to the lowest observed optical transition. Many computational studies use the HOMO–LUMO gap $\Delta E_{\mathrm{HL}}$ (the difference between the energies of frontier molecular orbitals in the ground state) as a starting point in approximating experimental optical gaps. This approximation is widely favored due to the computational challenges associated with employing highly accurate quantum mechanical methods incorporating orbital relaxation effects (e.g. time-dependent density functional theory or multi-configurational self-consistent field methods), especially when investigating vast areas of the CCS, macromolecules, molecular aggregates, or molecular junctions [192–194]. Thus, the HOMO–LUMO gap plays a crucial role in understanding various aspects of chemical reactivity, excitation energies, and several key optical properties in

these organic systems. For example, its calculation is essential to gain insight into optical absorption spectra, refractive indices, and conductivity [195–198]. For correctness of terminology, the HOMO–LUMO gap obtained from density functional calculations should be referred to as Kohn-Sham (KS) gap. Although the relations between different gaps (Kohn-Sham, fundamental, and optical) are subtle and have been discussed in detail in the literature [199, 200], in this manuscript we will use the KS gap as a proxy for observable experimental properties.

The molecular dipole polarizability $\alpha$ (referred to simply as polarizability in this Section as well as in the manuscript), on the other hand, describes the dipolar response of a molecule to an external electric field, becoming a key quantity for understanding intra- and intermolecular interactions (e.g. dispersion interactions, substituent and solvent effects as well as supramolecular structure formation) and for determining spectroscopic properties of molecules (Raman, Raman optical activity, and sum frequency spectroscopy) [4, 201–207]. These features make both polarizability and HOMO–LUMO gap essential in the derivation of structure-property/property-property relationships and, consequently, in the development of design strategies for molecules with a targeted array of QM properties for applications such as molecular dyes [208], optoelectronic devices [209], molecular junctions [210, 211], heterogeneous catalysts [212] and materials for non-linear optics [22, 213].

Various computational methods and predictive models have been developed to estimate HOMO–LUMO gaps and polarizabilities for organic molecules with different levels of trade-off between precision and computational cost [31, 77, 214, 215]. Lately, it has become feasible to access a plethora of highly accurate QM properties – including $\Delta E_{\mathrm{HL}}$ and $\alpha$ – for large swaths of the chemical compound space (CCS) [183, 216–219]. Comprehensive analyses of these extensive datasets may help to understand the deeper physical picture behind the inherent property-property relationships. With this motivation, we herein conduct an exhaustive investigation of the two-dimensional space defined by HOMO–LUMO gap and polarizability (i.e. $(\Delta E_{\mathrm{HL}}, \alpha)$-space) for small organic molecules with the aim of getting insights into the intrinsic relationship between these two properties. We find that while correlation might appear in homologous molecules (that is, molecules differing by a constant increment, meaning that their physicochemical properties follow a general trend), if a large enough subspace of CCS is considered, HOMO–LUMO gap and polarizability are essentially uncorrelated and their 2D space is represented as a structureless "blob". Through the analysis of diverse molecular sets, it is shown that this lack of correlation can be related to the fact that polarizability is primarily determined by the atomic composition, while

the HOMO–LUMO gap heavily depends on the arrangement of the atoms into chemical functional groups. Hence, we expect that our findings will assist the development of novel design principles in which the control of multiple electronic properties is relevant, as we finally demonstrate on the case of molecular photodetectors.

The outline of the discussion is as follows: in Sect. 4.1.2, we review accurate and approximate models for polarizability and HOMO–LUMO gap. In Sect. 4.1.3, we exhaustively examine the polarizabilities ($\alpha$) and HOMO–LUMO gaps ($\Delta E_{HL}$) of diverse molecular sets. In doing this, we have extended the QM7–X dataset [183] with functional group information and polarizabilities calculated with the hybrid PBE0 functional. In assessing our computational setting, we tested the predictive power of this functional against coupled cluster CCSD(T) calculations, and found an overall accuracy of 1.9%. As a first order approximation to predicting polarizabilities of small organic systems, we consider a linear combination of atomic contributions in Sec. 4.1.3. In Sect. 4.1.3, we then perform PBE0 calculations of polarizability (see Sect. 4.1.6 for computational details) for homologous molecules and explore the relationship with their HOMO–LUMO gaps. A statistical analysis of the $(\Delta E_{HL}, \alpha)$-space using a subset of molecules contained in QM7–X dataset is carried out in Sect. 4.1.3. Our proposed design principle is discussed and demonstrated in more detail on the case of organic photodetectors; see Sect. 4.1.4. The computational methods as well as the dataset used are presented in Sect. 4.1.6, following the main conclusions in Sect. 4.1.5.

## 4.1.2   Models for Polarizability and Frontier Orbital Energy Gap

Since our main focus is on having a better understanding of the relationship between polarizability and HOMO–LUMO gap in organic molecules, we first revisit the different qualitative and quantitative models used to compute them. In general, a variety of electronic structure methods can be employed to calculate both of these quantities. The choice of a computational level depends on the specific target property and the necessary trade-off between computational cost and accuracy. Although the calculation of the HOMO–LUMO gap is feasible using various mean-field electronic structure methods, orbital relaxation effects play a significant role in determining optical properties [220]. However, in computationally expensive studies such as the analysis of macro-molecules or extensive datasets (*e.g.* QM7–X considered in this manuscript), HOMO–LUMO gap is often used as a first approximation to experimental quantities. Polarizability ($\alpha$) is typically obtained from finite field, coupled perturbed Hartree-Fock or density functional perturbation

theory (DFPT) calculations [69, 214, 215]. However, these electronic structure methods require considerable computational resources when dealing with larger molecules or significant swaths of the CCS. Accordingly, we will next discuss alternative physical models, empirical correlations, and approximate methods that can be used to obtain these QM properties as well as to broaden the comprehension of property-property relationships in CCS. We will start with examining the polarizability, for which analytical models (such as the quantum Drude oscillator, or QDO) as well as empirical correlations and predictive semiempirical methods are available. After this, the models for the HOMO–LUMO gap will be mentioned, before concluding the section by analyzing what is known about the correlation between these two quantities.

A connection between HOMO–LUMO gap and polarizability can be anticipated starting from the perturbative expression for polarizability using the dipole moment operator $\hat{\mu}$ within second order perturbation theory as [4, 19]

$$\boldsymbol{\alpha} = 2 \sum_{n \neq 0}^{\infty} \frac{\langle \Psi_0 | \hat{\mu} | \Psi_n \rangle \otimes \langle \Psi_n | \hat{\mu} | \Psi_0 \rangle}{E_n - E_0} \ , \tag{4.1}$$

where $\Psi_0$ and $E_0$ are the ground state wavefunction and energy, respectively, and $n$ is the index of the excited states. Indeed, since $\Delta E_{\mathrm{HL}} = E_1 - E_0$ is commonly much smaller than the energy gap of higher excited states, the first term of the sum in Eq. 4.1 provides a first-order approximation to the infinite series, and hence there could exist an inversely proportional relationship between $\Delta E_{\mathrm{HL}}$ and $\alpha$, i.e. $\alpha \propto (\Delta E_{\mathrm{HL}})^{-1}$.

Equation 4.1 can be evaluated analytically only for simple model systems (such as the hydrogen atom or a quantum Drude oscillator). For many-electron systems, the sum can only be evaluated numerically and requires including bound-bound and bound-continuum transition dipoles [221]. Modeling atoms or larger coarse-grained fragments with QDOs and solving the dipole-dipole screening equations is known to be an effective method for predicting polarizability, and it is also the basis for the Many-Body Dispersion (MBD) method [41, 42]. Since the response properties of all atoms and molecules can be represented by QDOs by carefully setting the three parameters {charge $q$, frequency $\omega$, mass $\mu$} of the model, the analysis of the polarizability of the QDO Hamiltonian should generally be transferable to any system. Therefore, some attention is given to this model.

Due to selection rules of the dipole operator, only the first excited state contributes to the dipole

polarizability of a QDO [19, 117], making it effectively a two-state system

$$\alpha_{\mathrm{QDO}} = 2q^2 \frac{\langle \Psi_0 | \hat{\mu} | \Psi_1 \rangle \langle \Psi_1 | \hat{\mu} | \Psi_0 \rangle}{E_1 - E_0} = \frac{q^2}{\mu \omega^2} \, , \tag{4.2}$$

where $q$ is the magnitude of the charge bound by a harmonic potential with frequency $\omega$, having a mass $\mu$. The HOMO–LUMO gap of a QDO is $\Delta E_{\mathrm{HL}} = \hbar \omega$, which indeed appears in the denominator. However, $\alpha$ can be controlled separately through the other two individual QDO parameters $\{q, \mu\}$, independently from $\Delta E_{\mathrm{HL}}$. This means that for the QDO model, the polarizability and the HOMO–LUMO gap are mutually related, yet they could be tuned separately from each other.

The idea of approximating the polarizability using an effective two-state system (so-called Unsøld approximation) [19, 84] is also useful for understanding qualitative trends. Within this approximation, polarizability is written using an average excitation $\Delta E$ as a fitting parameter

$$\boldsymbol{\alpha} = \frac{2}{\Delta E} \sum_{n \neq 0}^{\infty} \langle \Psi_0 | \hat{\mu} | \Psi_n \rangle \otimes \langle \Psi_n | \hat{\mu} | \Psi_0 \rangle. \tag{4.3}$$

Setting the average excitation to $\Delta E_{\mathrm{HL}}$ is therefore exact for the QDO model, but the connection between these quantities for many-electron systems is not known in general [85].

Investigating correlations between polarizability and various molecular properties can lead to useful relationships, such as the recent observation that polarizability scales with the fourth power of the characteristic size of the system [19]. The correlation between polarizability and orbital energies is relevant from a theoretical point of view, as it forms the foundation of Pearson's hard-soft acid-base (HSAB) theory [20, 21]. Based on recent theoretical works, we can postulate that polarizability can be expressed as a function of two factors accounting for *i)* ground state geometry (e.g., van der Waals radius or molecular volume) and *ii)* electronic structure (e.g., ionization energy or hardness) [19, 22–28]. Although these correlations provide useful conceptual insights, they have not been used for constructing accurate numerical predictions.

There are two types of predictive models for polarizability with a lower computational burden than electronic structure calculations. First, approximations for polarizability can be constructed based on the group contribution principle, which divides polarizability into atomic or bond contributions [212, 222]. These models can offer somewhat accurate predictions with minimal molecular information and computational effort, and we will assess such models in this work. As a sec-

ond approach, machine learning (ML) models have been proposed as a cost-effective solution with improved accuracy [77]. However, the training process and accuracy of the ML models are strongly dependent on the features of the dataset (e.g. chemical diversity, molecular size, number of samples) and on the ML method itself.

For the case of HOMO–LUMO gap, there is a well-established underlying physical principle to determine this property: it is known that the HOMO–LUMO gap of individual functional groups (called chromophores in this context) is transferable, with values documented in standard reference texts [223]. These chromophores also form the foundation for accurate ML models and earlier empirical rules for the prediction of HOMO–LUMO gaps [224, 225]. The HOMO–LUMO gap of a single functional group can be understood based on molecular orbital theory, the most common version of which is the Hückel theory for conjugated systems. For instance, the inverse proportionality between the number of monomers and the HOMO–LUMO gap of polyenes is well explained within this theory [226]. In the case of noninteracting functional groups, their optical spectra are effectively independent and, consequently, the frontier energy gap of a molecule is determined by the lowest value for the constituent functional groups, making it an inherently size-independent (intensive) property.

In agreement with the analysis of the QDO model, recent studies relying on large datasets (7 k structures from the GDB-13 dataset as well as the TMQM dataset of 86 k transitional metal complexes) suggest that there is no overall correlation between HOMO–LUMO gap and polarizability [78, 227]. However, correlation has been observed both experimentally and computationally for different classes of structures (e.g. organic dyes and inorganic clusters [208, 228–230]), with the notable exception of smaller systems where the HOMO–LUMO transition is symmetry forbidden [206]. In the following section, we explore the source of such seemingly contradictory results by showing that investigating a reduced subset of the chemical compound space can lead to correlations between quantities that are generally uncorrelated.

### 4.1.3 Results and Discussion

**First Order Linear Atomic Additive Model for Polarizability**

To understand the correlation between polarizability and the HOMO–LUMO gap, we first show that polarizability can be determined, up to a large degree, by knowing only the atomic composition of

Table 4.1: Revised linear regression parameters for the atomic additive polarizability model of Bosque et al. Note that the values in the original paper are presented in Å$^3$, whereas the values here are in bohr$^3$. The relatively low influence of the intercept can be seen by comparing the last two rows; the rest of our manuscript uses the parameters presented under "This work"

|  | Intercept | C | Cl | H | N | O | S |
|---|---|---|---|---|---|---|---|
| Bosque | 2.14 | 10.20 | 14.60 | 1.17 | 6.95 | 3.85 | 20.20 |
| This work | 1.71 | 10.10 | 12.70 | 0.87 | 7.88 | 4.00 | 19.10 |
| No intercept | 0.00 | 10.37 | 13.00 | 0.88 | 8.11 | 4.24 | 19.37 |

a molecule. This analysis is done using the PBE0 polarizability values calculated for the QM7–X molecules, contrasting these quantum chemical calculations with a linear atomic additive method. The simplest atomic additive method (motivated by Bosque [31]) approximates the polarizability of a molecule through a linear combination of the number of each atom-types $n$ weighted with a type-specific factor $C_i$, together with an intercept $m$

$$\alpha = m + \sum_i C_i n_i \,.$$

(4.4)

Bosque's model was fitted directly using experimental data of 426 compounds. The fitted $m$ and $C_i$ values for C, Cl, H, N, O, and S are listed in Table 4.1. Consequently, we have used here the QM7–X dataset [183] (see Sect. 4.1.6) to validate the accuracy and reassess the model parameters on a significantly larger swath of the CCS. In doing so, we have considered the first conformer for each entry in the QM7–X dataset; a total of $\approx 13$ k structures. The linear regression parameters optimized on QM7–X yield the results listed in Table 4.1. Bosque's parameters hold relatively well for QM7–X molecules, accounting for a correlation coefficient ($R^2$ value) of 0.65 with a mean absolute percentage error (MAPE) of 6.11 %. However, the re-fitted parameters improve the correlation coefficient to 0.72 and reduce the MAPE value to 3.94 %, i.e. the prediction accuracy is increased by a factor of 1.6. Note that the presence of the intercept $m$ in Eq. 4.4 is just an artifact of the model, since the prediction should be zero when no atoms are present. Inclusion or omission of $m$, however, changes neither the goodness of the regression nor the numerical value of the atomic contributions to a meaningful degree (the mean absolute error of the linear model goes from 3.078 a.u. to 3.079 a.u.; similarly to what had also been observed by the authors). Therefore, we decided to include the intercept in our further analysis, to be consistent with Bosque's approach.

A shortcoming of atomic additive methods is that the same polarizability is predicted for all

structural isomers, since only the total number of each atom-types is used in the prediction. This is manifested in Fig. 4.1 as having a systematic error within each possible $\alpha_{\mathrm{pred}}$ value and further demonstrated in the inset on the case of molecules with chemical formula $C_6H_8O$. Indeed, the reference polarizabilities for this given chemical formula span a range of 30 a.u., but the predicted value is 73.4 a.u. for all molecules, irrespective of the chemical arrangement of the atoms. From Fig. 4.1, it can also be inferred that such a simple additive model will only become worse for molecules of increasing size. Indeed, a trend appears where larger molecules exhibit stronger deviations towards higher polarizabilities – a trend that an additive model is unable to describe. This can especially be the case for polymeric molecules which form long chains, whose polarizability is highly anisotropic and it behaves increasingly non-additively with size.
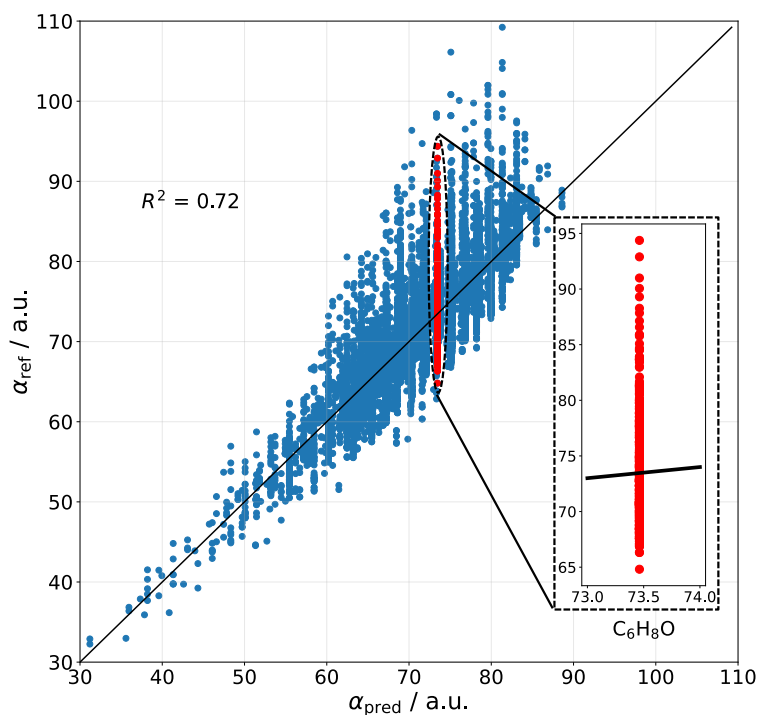


Figure 4.1: Performance of the atomic additive method of Eq. (4.4), using the parameters "This work" in Table 4.1 The inset shows the inherent shortcoming of the model, predicting the same polarizability for all $C_6H_8O$ isomers [29].

To differentiate between structural isomers, a descriptor that accounts for different geometric properties (for example, radius of gyration) might be constructed, since polarizability is an extensive property [25, 28]. This extensivity is only partially captured by atomic additive methods insofar as increasing the number of atoms in a molecule inherently increases the size as well. More accurate models should also differentiate between similar atoms based on their surrounding chemical environments, as is done for example in Ref. [19], where the short-range environment is taken into account by Hirshfeld partitioning as well as in the self-consistent screening approach used in

the Many-Body Dispersion (MBD) method [41, 42]. Therefore, while the shown first-order linear model is limited by its accuracy, it can serve as a baseline for more accurate methods involving coupling between atoms in a molecule.

In summary, a first-order approximation to polarizability can be constructed just by using an atomic additive model without explicit knowledge of the molecular spatial arrangement or the local chemical environments. While the predictive power of such a model is rather restricted (i.e. the chemical environment of each atom is not described), its rough correlation with reference electronic-structure calculations (see Fig. 4.1) gives a clear evidence that a significant fraction of the polarizability is determined by just the atomic composition.

**Case Studies for the Relation between the HOMO-LUMO Gap and Polarizability**

Having shown that polarizability depends mainly on the atomic structure of molecules, we now turn into exploring the correlations between polarizability and HOMO–LUMO gap. In doing so, we here discuss a set of case studies of select molecules, with all calculations being done using the PBE0 functional, as described in Sect. 4.1.6.

Experimental studies often focus on examining molecules with similar electronic structures, leading to hidden correlations between the optical gap and the polarizability. In our path toward the general understanding of the relationship between these QM properties, we now examine two different cases: (i) molecules having the same atom-type composition but slightly different chemical compositions, and (ii) molecules with the chemical properties fixed while increasing the system size (e.g. oligomers).

**Constitution isomers.** In general, the functional groups in a molecule govern the nature and order of the molecular orbitals, determining the HOMO–LUMO gap and the orbitals involved in the electronic transitions. To explore the relationship between HOMO–LUMO gap and polarizability as a function of chemical functionality, we present select examples of constitutional isomers, i.e. molecules with the same atomic composition that belong to different substance classes due to the presence of different functional groups.

As a first example, two constitutional isomers with the formula $C_5H_8O$, namely an $\alpha, \beta$- (3-penten-2-one) and a $\beta, \gamma$-unsaturated enone (4-penten-2-one) is considered (see Fig. 4.2(a)). It is noticeable that 3-penten-2-one has a smaller gap compared to 4-penten-2-one (by 1.12 eV)
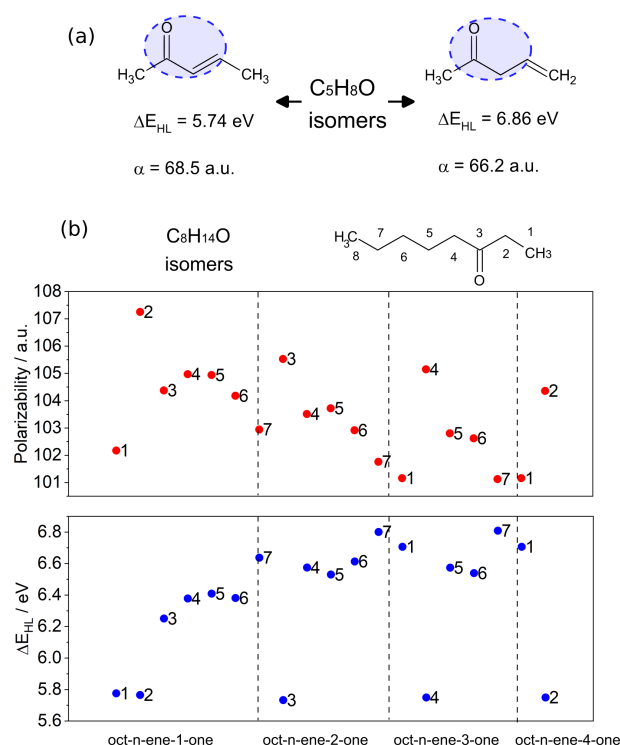
Figure 4.2: (a) Two constitution isomers, i.e. molecules with the same atomic composition but different chemical properties, showing similar polarizabilities but different HOMO–LUMO gaps. (b) HOMO–LUMO gap and polarizability of all possible linear structures having eight carbon atoms, an oxo group, and a double bond between two of the carbons. The numbering of carbon atoms is shown in the case of octane-3-one, with $n$ representing the numbering of the carbon atom at the start of the double bond (see Fig. S 5 of ESI [29] for the explicit structures). The polarizability and HOMO-LUMO gap values are results of PBE0 calculations as described in Sec. 4.1.6.

because delocalization results in greater mobility of $\pi$-electrons throughout the molecular structure. However, both molecules have a similar polarizability due to the identical atomic composition as well as a similar total size.

A second set of constitutional isomers with the formula $C_8H_{14}O$ was constructed for molecules bearing a C=O (oxo-group) and C=C (alkene-group) on an octane backbone. Therefore, these isomers are formed by the following classes of substances: one ketene, one conjugated aldehyde, four conjugated ketones, five non-conjugated aldehydes, and eight non-conjugated ketones (see Fig. 4.2(b) as well as Fig. S 5 of ESI [29]). Although these structures are chemically quite different, their orbital symmetries are largely similar, leading to a correlation between their polarizability and HOMO–LUMO gap. Note, however, that the polarizabilities of the structures are all within 4% of each other, while the variation of the HOMO–LUMO gap is about five times larger. As such, the statement that polarizability is mainly determined by the atomic composition and HOMO–LUMO gap by the chemical composition seems to hold, even though some correlation between these two quantities is observed due to the similarity of the structures.

**Homologous series of molecules.** As previously elaborated, HOMO–LUMO gap and polarizability can seemingly correlate for molecules that belong to a homologous series. This can be explained by

the fact that the electronic nature and order of the frontier orbitals are often identical for structurally and electronically similar molecules. Consequently, the decrease in the HOMO–LUMO gap can correlate with the increase in polarizability when considering molecules of a homologous series with an increasing number of repeating units. To support this assumption, we consider in the following a series of oligomers, namely alkanes ($C_mH_{2m+2}$) and alkenes ($C_mH_{2m}$; see Fig. 4.3). The example is taken from Afzal et al. [231], with polarizability and HOMO–LUMO gap recalculated within our computational setup (see Sect. 4.1.6).
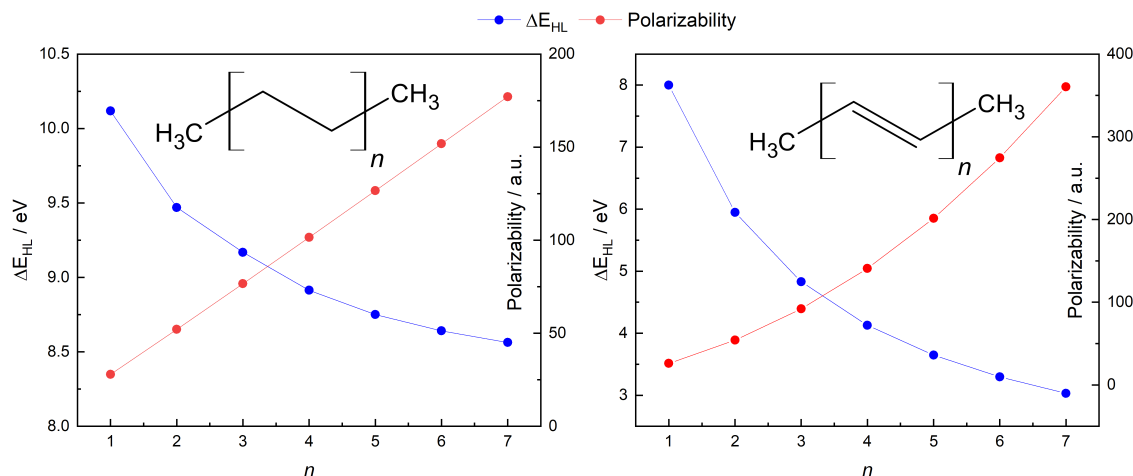


Figure 4.3: HOMO–LUMO gap (blue dots) and polarizability (red dots) of the oligomers of (a) ethylene and (b) acetylene. The calculations of both properties were carried out as described in Sect. 4.1.6.

Fig. 4.3 shows a decreasing behavior of HOMO–LUMO gap for oligoethylene and oligoacetylene as a function of the number of monomers $n$, in agreement with previous works, as well as qualitative predictions from the Hückel model [226, 232]. In fact, we have found that the absence of a qualitative change in the electronic structure within the ethylene oligomers leads to a relatively small HOMO–LUMO gap change going from $n = 1 \rightarrow 7$ ($\approx 1.5$ eV) compared to the acetylene oligomers, where every monomer modifies the conjugation, producing a more significant change of $\approx 5.0$ eV. Unlike the HOMO–LUMO gap, the behavior of polarizability in molecular chains cannot be simply explained. The observation that polarizability increases monotonously with $n$ is in line with both the principles of atomic additive models and the correlation with molecular size. However, the absolute magnitude of the polarizability values is significantly different for the two sets of oligomers, and this difference increases with increasing number of monomers. This quantitative difference cannot be explained by additive atomic models or correlations using

molecular size, but it correlates with the reductions in HOMO–LUMO gap. In particular, not even ALPHAML [77] can predict this behavior: the model predicts 181 a.u. for oligoethylene and 227 a.u. for oligoacethylene in the $n = 7$ case, with the DFPT results being 177 a.u. and 360 a.u., respectively. Therefore, the difference due to conjugation is underestimated by a factor of four even when using ML methods, and this error is expected to increase with increasing chain length. Until now, we are not aware of any other simple polarizability estimation method that can accurately predict the values in Fig. 4.3. These findings provide clear evidence that further work is necessary to enhance our understanding and improve the accuracy of computational methods used for calculating polarizability, even for relatively simple molecules such as hydrocarbon oligomers.

### Clustering of Structures in the $(\Delta E_{HL}, \alpha)$-Space

To draw more general conclusions about the relationship between the properties in question, we analyze the two-dimensional (2D) property space defined by HOMO–LUMO gap and polarizability for a selected subset of QM7–X molecules [191] (see Sec. 4.1.6). HOMO–LUMO gap values were taken from the dataset, while polarizability values were recalculated using the computational setup explained in Sect. 4.1.6. All examples presented previously might suggest that there is a correlation between HOMO–LUMO gap and polarizability. However, these examples considered similar molecules with respect to their functionality or chemical composition — factors that essentially determine both the HOMO–LUMO gap and polarizability. From optical spectroscopy it is known that the optical gap is primarily determined by the functional groups in a molecule. This is reflected in characteristic optical gaps (vertical excitation energies of the lowest electronic transitions) per functional group, for example, the $\pi\pi^*$ absorption of an isolated alkene-group as chromophore is between 7.51 and 6.70 eV. Since we are assuming that HOMO–LUMO gap is a good starting point for determining the optical gap of a molecule, it would be expected to find that $\Delta E_{HL}$ values are also clustered by certain functional groups. On the contrary, our analysis has shown that polarizability ($\alpha$) is primarily determined by the atomic composition of a molecule. The QM7–X dataset enables us to study the $(\Delta E_{HL}, \alpha)$ relationship more broadly because it covers a considerable number and variety of chemical compounds.

Fig. 4.4(a) shows the $(\Delta E_{HL}, \alpha)$-space for the QM7–X molecules – indicating no direct relationship between the two quantities across the chemical compound space spanned by this dataset ($R^2 = 0.13$). Furthermore, the role of the two main factors that determine $\Delta E_{HL}$ (functionality) and $\alpha$ (atomic

composition) are highlighted in Fig. 4.4. Panels (b) and (c) exemplarily display the distributions of $\Delta E_{HL}$ and $\alpha$ for aldehydes and primary alcohols, *i.e.*, molecules that bear one of the respective functional groups. Subplots (d) and (e) show the respective distributions for molecules with equal atomic compositions, namely with the molecular formulas $C_4H_8O$ and $C_4H_9N$, respectively.

**Functional groups & HOMO-LUMO gap.**

In Figs. 4.4(b,c), we highlight the frequency plots of $\Delta E_{HL}$ and $\alpha$ values for all non-conjugated aldehydes (blue) and primary alcohols (pink) of our select dataset. Fig. 4.4(c) clearly reflects the common notion of chromophores, namely that HOMO–LUMO gap is mainly determined by the type of chromophore (e.g., aldehyde or primary alcohol group) and the character of the lowest energy electronic transition (e.g. $n\pi^*$- or $n\sigma^*$-transition). Thus, $\Delta E_{HL}$ values for aldehydes only show a value of *circa* 6.5 eV while, for primary alcohol group, they extend from 7.2 eV to 7.8 eV.

To fully explore the role of functional groups in the ($\Delta E_{HL}$, $\alpha$) relationship, the QM7–X molecules were categorized into twelve major classes according to the functional groups they are bearing (see Sec. 4.1.6). Fig. S 4 of the Supporting Information in [29] shows the distribution of all functional groups detected in the dataset, confirming that $\Delta E_{HL}$ is clustered along the chemical properties of the molecules.

Unlike $\Delta E_{HL}$, molecules containing aldehydes (blue) and primary alcohols (pink) exhibit polarizabilities that extend throughout the entire range of the dataset (see Fig. 4.4(b)). This finding is further reflected in the average Kolmogorov-Smirnov-metric (measuring the statistical distance between two general distributions) of the individual molecular classes in the ($\Delta E_{HL}$, $\alpha$)-space, which is 0.81 and 0.40 for $\Delta E_{HL}$ and $\alpha$, respectively. To quantify the difference between the distributions of the eleven molecular classes studied in the $\Delta E_{HL}$-$\alpha$ space, the pairwise Kolmogorov-Smirnov-distances (*i.e.*, 55 unique pairs) were calculated between the normalized quantities using the SciPy implementation[233]. An (unnormalized) example for a unique pair is shown in the main article in Figure 1b ($\alpha$) and 1d ($\Delta E_{HL}$) for non-conjugated aldehydes and primary alcohols.

The Kolmogorov-Smirnov distance for two probability distributions $i$ and $j$ is defined using their individual empirical distribution functions $F(X)$ as

$$D_{ij} = \sup \left| F_i(x) - F_j(x) \right| , \tag{4.5}$$

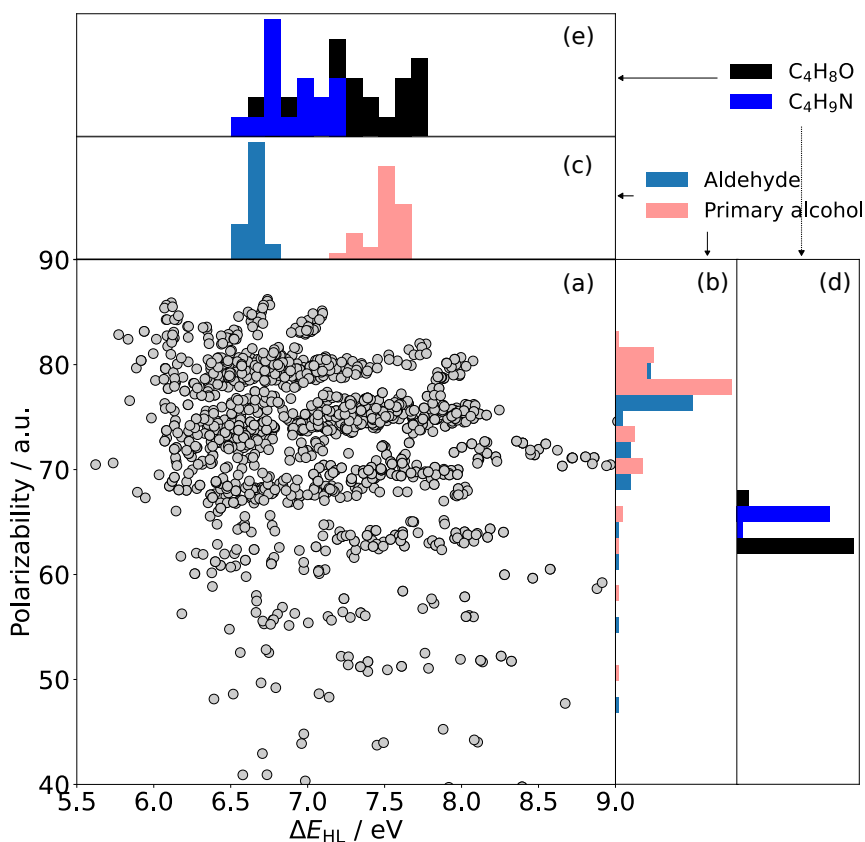Most commonly, the Kolmogorov-Smirnov distance is used in testing whether the probability

Figure 4.4: (a) Polarizability ($\alpha$) *vs.* HOMO–LUMO gap ($\Delta E_{HL}$) for molecules of the subset of QM7–X under study (see text). Histograms of the HOMO–LUMO gaps (b) for all non-conjugated aldehydes (blue) and primary alcohols (pink) and (d) for structures having the atomic composition $C_4H_8O$ (black) and $C_4H_9N$ (blue). Histograms of polarizabilities (c) for all non-conjugated aldehydes (blue) and primary alcohols (pink) and (e) for structures having the atomic composition $C_4H_8O$ (black) and $C_4H_9N$ (blue). The difference in the clustering in the two quantities is reflected in the degree of separation between the histograms.

distributions $i$ and $j$ have the same underlying distribution [234]. In our case, we calculate this metric for distributions that are known to be different, not for the purpose of a statistical test but to quantify the distances of the distributions. In this context, the absolute value of the Kolmogorov-Smirnov distance has little practical information; however, comparison of the distances confirm that the HOMO–LUMO gap depends on the functional groups present, whereas polarizability does not. Our analysis then demonstrates that functional groups primarily affect the HOMO–LUMO gap rather than polarizability, resulting in well-defined molecular clusters on the $\Delta E_{HL}$-axis.

**Atomic composition & polarizability.**

According to the Kolmogorov-Smirnov analysis, the functional groups only indirectly influence the magnitude of the polarizability in a given molecule, whereas the atomic composition is a crucial factor for the determination of the polarizability. This finding is also in line with the fact that a good correlation is achieved between the first-order atomic additive model and the reference DFT data shown in Fig. 4.1. The $\Delta E_{HL}$ and $\alpha$ values for a set of two constitution isomers, namely with the chemical formula $C_4H_8O$ (including aldehydes, dialkyl ethers, enol ethers, as well as primary and secondary alcohols) and $C_4H_9N$ (including carbonitriles and primary/secondary aliphatic amines) is also presented in Figs. 4.4(d), showing a narrow polarizability distribution. These results are another clear evidence that $\alpha$, to a reasonable approximation, is independent of the actual chemical arrangement of the atoms in the molecule, but it mainly depends on the total number of atom-types.

In summary, we can conclude that the lack of general correlation observed in $(\Delta E_{HL}, \alpha)$-space is a consequence of two main facts: (i) the HOMO–LUMO gap is determined by the nature of the chemical composition (see Fig. 4.4c *vs.* 4.4e) and (ii) the polarizability is largely determined by the atomic composition (see Fig. 4.4b *vs.* 4.4d).

### 4.1.4   Case Study: Design of Photodetectors

We have demonstrated in previous sections that polarizability and HOMO–LUMO gap do not correlate across large swaths of the CCS. Now, we present how this lack of correlation can be exploited for molecular design purposes. The property data used for this analysis is from the donor-acceptor (DA) dataset [235], which was designed to enumerate promising organic photodetector candidate molecules. The DA dataset contains only molecular structures and HOMO–LUMO gap values, while the estimation of polarizability was performed using the revised Bosque model, as elucidated in Sect. 4.1.3.

A common challenge in materials science is the effective design of photodetectors. These optoelectronic devices capture light and convert it into electric signal, therefore, playing an important role in sensing, monitoring, and optical communication. The wide range of physicochemical properties spanned by organic molecules allows various design strategies, which ultimately led to the emerging field of organic photodetectors [236, 237]. HOMO–LUMO gap is one of the key quantities that can be used to approximate the coupling strength of molecules with light; therefore, any design strategy

motivated by optics will be initially based on this property [235, 238]. Since the fundamental function of photodetectors is to convert light into electrical current, controlling the electrochemical behavior is also crucial. Specifically, the electrochemical work function plays a critical role in the description of organic photodetectors [239, 240], as opposed to organic semiconductors, where the focus is usually on charge carrier mobility [241]. The work function $\phi$ of an electrode is known to change with the polarizability of the absorbed molecules as well as the surface coverage, as described by the Topping equation [242] (written for a square lattice)

$$e\Delta\phi = \pm\frac{e\mu\theta}{\varepsilon_0 d^2}\left(1 + 9\alpha'\left(\frac{\theta}{d^2}\right)^{3/2}\right)^{-1}. \tag{4.6}$$

This expression highlights that the work function $\phi$ also depends on the the dipole moment $\mu$ and polarizability $\alpha$ of the molecules, besides the surface coverage $\theta$ and the lattice constant of the absorbate $d$. Notice that an effective polarizability $\alpha'$ is used to represent the properties of the absorbed molecules in Eq. 4.6, which is usually an order of magnitude larger than the free molecular polarizability [243, 244]. Although it is known to fluctuate with the coverage rate, this equation can serve as a useful initial reference to screen potential molecules for photosensitizers according to the intended work function [239, 240]. Indeed, this relationship between both properties makes it important to regulate the polarizability of molecules to achieve the desired electrochemical behavior. Through this connection, it can be seen that molecules with higher polarizability tend to facilitate electron injection while those with lower polarizability tend to facilitate hole injection [240].

In the preceding sections, we have postulated that polarizability and HOMO–LUMO gap are uncorrelated if a large enough subset of the CCS is considered. This law can now be translated to the domain of organic photodetectors: since HOMO–LUMO gap and polarizability are generally independent, it should be possible to design a photodetector with a given detection peak that has an arbitrary work function. Alternatively, if matching of electrochemical properties of different systems is the goal, it should be possible to design organic photodetectors with each having an arbitrary optical detection window, yet having the same effect on the work function of electrodes. To demonstrate this statement, we use a dataset generated by Xu et al. [235], who used a self-improving Bayesian search to predict possible photodetector molecules in a large subset of CCS. The selection criterion for possible photodetectors was based on both HOMO–LUMO gap and singlet-triplet energy gap, which were evaluated from ground state DFT and TD-DFT calculations, respectively. Of all predicted molecules that have a donor and an acceptor site (DA structures),
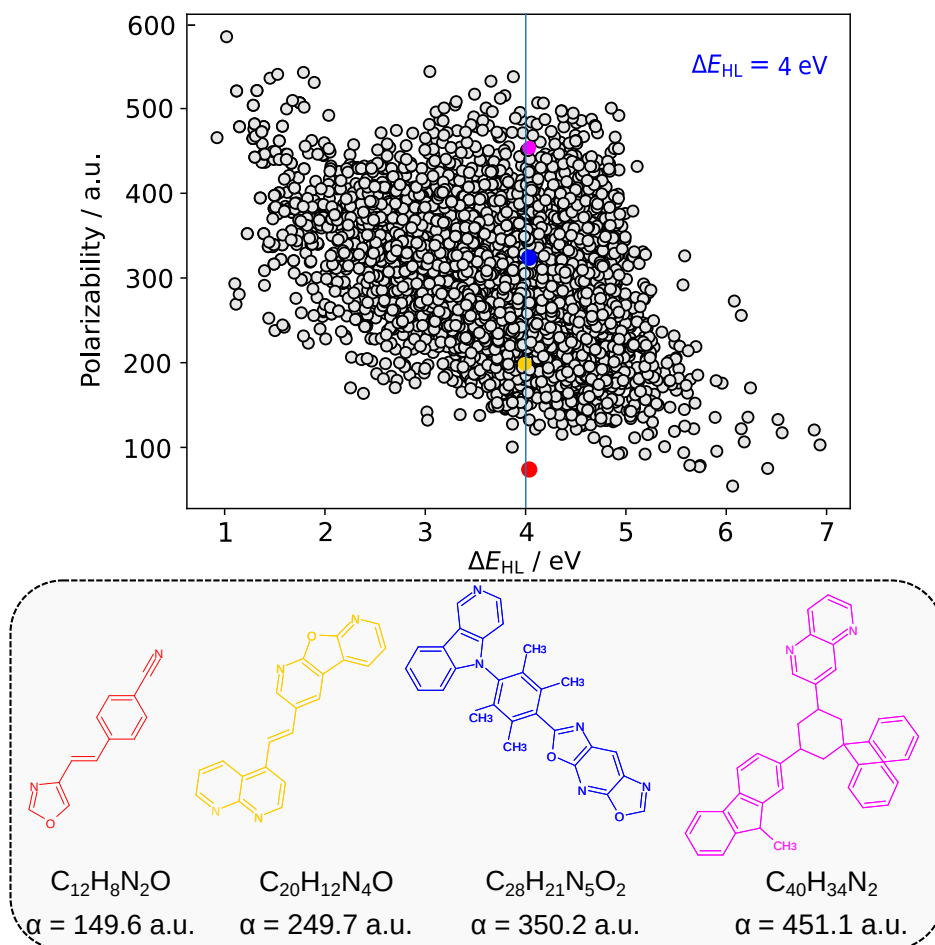
Figure 4.5: HOMO–LUMO gap and polarizability of the structures in the donor-acceptor (DA) dataset of Xu et al.[235]. The maximum of the HOMO–LUMO gap density (4 eV) is marked with a blue line. The four structures corresponding to the four quartiles (within ± 0.1 eV) in the predicted polarizability are also shown, together with the predicted values (in a.u.).

we have only selected those cases that have the same atom types as QM7–X molecules (see also Table 4.1), leading to a total of 5, 311 structures. Using the atomic additive model described in Sec. 4.1.3, we have estimated the polarizabilities of the selected structures; the graph of the polarizability versus HOMO–LUMO gap is shown in the top panel of Fig. 4.5. Here, one can see that most structures are found in a relatively extended region having $\Delta E_{\mathrm{HL}}$ between 3 and 5 eV and $\alpha$ between 200 and 400 a.u., with the possibility to find outliers in all directions around this cluster. In particular, if a high shift in the work function is desired, there appear to be several good candidates with varying optical absorption ranges (see bottom of the graph).

Moreover, our calculations show that the polarizability of these structures can vary by a factor of up to six, depending on the specific values of the HOMO–LUMO gap. To demonstrate this flexibility in $\alpha$, Fig. 4.5 also shows the four molecules corresponding to the four quartiles having a HOMO–LUMO gap of $4 \pm 0.1$ eV, selected to correspond to the maximum density of data. For this specific HOMO–LUMO gap, polarizability changes between 149.9 a.u. and 451.1 a.u. Taking into account the tenfold enhancement between the polarizability of the free molecule and the absorbed $\alpha'$ and using approximate values of $\mu = 4$ D and $d = 1.5$ nm with a full surface coverage, this would mean that changes in work function could range from 0.9 eV to 1.5 eV. These variations are larger than are usually achievable by modification of a semiconductor structure or control of surface coverage [244, 245]. Therefore, our analysis shows that the work function can be, for practical purposes, freely tailored, even with a very specific design requirement on the HOMO–LUMO gap.

This flexibility is also relevant in the task of designing wavelength-selective detectors, which would imply a hard constraint on the HOMO–LUMO gap. If $\Delta E_{\mathrm{HL}}$ and $\alpha$ could not be controlled independently, then optical design restrictions would directly influence electrochemical behavior. The decoupling of $\Delta E_{\mathrm{HL}}$ and $\alpha$ means that the wavelength of detection and the work function can be independently controlled. Fine-tuning the work functions to achieve matching on the metal-organic interface at the electrode is crucial for efficiency. Thus, with the existent "freedom of design" in $(\Delta E_{\mathrm{HL}}, \alpha)$-space, we have demonstrated that efficient detection can theoretically be achieved for any detection wavelength. Alternatively, since the work function can be tailored to match any detection wavelength, it is also possible to design detectors for different detection ranges having equivalent electrochemical properties such as sensitivity, dark current, and adhesion behavior as well as any other properties determined by the work function.

### 4.1.5 Conclusions

Predictive molecular design is an emerging tool in modern molecular physics and chemistry which heavily relies on the understanding of relationships between key structural and electronic properties. Identifying and explaining the correlations between properties requires either a deep physical understanding or exhaustive data analysis. Herein, we present a comprehensive investigation of the intricate interplay between the HOMO–LUMO gap and dipole polarizability – two central properties in the design of molecules with tailored optical properties and intermolecular interactions.

Despite the fact that both quantities have a root in the molecular electronic spectrum, understanding

their correlation is quite complex. On the one hand, the properties are essentially uncorrelated when a vast chemical space is taken into account. On the other hand, when examining a small subset of the chemical compound space with similar functionalities, such as homologous series of molecules like oligomeric hydrocarbons, we show that the properties can be observed as being correlated.

To perform a data-driven analysis, we extended the QM7–X database with functional group labels and accurate polarizabilities to explain the physical cause of this phenomenon. Our results demonstrate that the atomic composition plays an important role in determining the polarizability, while the arrangement of these atoms into chemical functional groups dictates the HOMO–LUMO gap. The physical origin of molecular polarizability was elaborated by studying conceptual models as well as interpreted with the help of a first order linear atomic additive model. Finally, the "freedom of design" arising from the interaction of HOMO–LUMO gap and polarizability was used in the example of organic photodetectors, demonstrating that the electrochemical properties of such molecules can be freely tailored even with specific requirements on the optical properties. The theoretical insights gained from this work can provide the basis for expanding the understanding of the relationship between HOMO–LUMO gap and polarizability by incorporating additional descriptors such as molecular size and electronic mobility. Additionally, the proven "freedom of design" could be applied to the development of new compounds with tailored optical and electronic properties for use in applications such as organic electronics, sensing, or energy harvesting.

### 4.1.6 Computational Methods

Generally, molecular design is a multi-property optimization problem and requires an exhaustive analysis of diverse structure-property and property-property relationships [184, 189]. Here, we have opted to focus on the two-dimensional property space defined by $\Delta E_{HL}$ and $\alpha$ (i.e. $(\Delta E_{HL}, \alpha)$-space), as motivated in the introduction (see Sect. 4.1.1). To calculate polarizability, two approaches were used: *(i)* the revised linear additive atomic model of Bosque (introduced in Sec. 4.1.3) was utilized for the prediction of polarizability of the organic photodetector candidates in Sec. 4.1.4 and *(ii)* density functional perturbation theory (for case studies in Sect. 4.1.3 as well as to analyze the QM7–X molecules in Sect. 4.1.3). The HOMO–LUMO gap was always obtained from DFT calculations, either by calculating it ourselves or utilizing the values provided in the QM7–X dataset.

**Target Molecular Property Space**

To perform a purely data-driven study, we utilize the QM7–X dataset [183] containing 42 physic-ochemical properties of $\approx 4.2$ M (equilibrium and non-equilibrium) organic molecules with up to seven heavy (non-hydrogen) atoms (including C, O, N, S and Cl), spanning a practically important subset of CCS. Accordingly, a subset of QM7–X considering only one equilibrium constitutional isomer and stereoisomer per unique molecular graph is selected for further analysis ($\approx 13$ k molecules). In QM7–X, the molecular structures were optimized using the third-order self-consistent charge density-functional tight binding method (DFTB3) [246] supplemented with a treatment of many-body dispersion/van der Waals interactions *via* the MBD approach [42, 139]. However, for our studies on polarizability, $\alpha$ was computed directly, using density functional perturbation theory (DFPT) [214] by means of the PBE0[247] functional as implemented in the `FHI-aims` code [123] (version 190205). To ensure the transferability of the values, we store the molecular (mean) polarizability (denoted as $\alpha$ and simply referred to as *polarizability* in other parts of the manuscript)

$$\alpha = \frac{1}{3} \left( \alpha_{xx} + \alpha_{yy} + \alpha_{zz} \right) , \tag{4.7}$$

which is independent of the molecular orientation. A second orientation-independent observable, the polarizability anisotropy ($\Delta\alpha$) is also often reported, defined as

$$(\Delta\boldsymbol{\alpha})^2 = 3(\alpha_{xy}^2 + \alpha_{xz}^2 + \alpha_{yz}^2) + \frac{1}{2}((\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{xx} - \alpha_{zz})^2 + (\alpha_{yy} - \alpha_{zz})^2).$$

This quantity is mainly used in the description of macromolecules and supramolecular systems, and since our focus is small organic molecules, we don't analyze the anisotropy in this manuscript.

Polarizability is known to be sensitive both to the choice of functional and the basis set size [72, 177, 207]. To converge our computational setup, the respective mean polarizabilities were compared with the highly accurate values of the QM7B database [248, 249]. This comparison ensures an accurate assessment of the prediction error for the following two reasons: *(i)* there is a large overlap between the structures in the QM7B and QM7–X databases, and *(ii)* QM7B provides highly accurate $\alpha$ values obtained at the linear-response coupled cluster singles and doubles (LR-CCSD) level of theory [27, 77]. Then, we computed the polarizabilities of 300 randomly selected structures of QM7B employing the same DFPT computational setup described above. We have found that the PBE0 hybrid functional using the default *light* basis set for all elements amended with three

additional functions from the *tight* level predicts $\alpha$ with a mean average error of 1.9% and a standard deviation of 1.1% (see Fig. S 1 of the ESI in [29]). The accuracy of our chosen computational setup is higher than the common DFT methods, and slightly better than the 2.84% found by Hait and Head-Gordon [177] for the PBE0 functional, which can be attributed to the fact that our study is only concerned with organic molecules. Polarizability anisotropy is predicted with a mean average error of 10.2% with a standard deviation of 5.1%, which is in line with previously reported values [247, 250]. In general, the mean polarizability is slightly underestimated, whereas the anisotropy is almost always overestimated by PBE0.

**Molecular Classification: Functional Groups**

A workflow has been implemented to identify chemical functional groups from the molecular structure in two steps: firstly, we save the Cartesian coordinates of molecules in a MDL Molfiles format using the standard implementation in `Open Babel` [251]. Second, `Checkmol` [252] is used to detect functional groups (204 tags) based on the connectivity tree. In total, 61 unique functional groups were detected for the subset of the $\approx 13$ k QM7–X-molecules [183], demonstrating that the dataset covers a considerable sector of CCS (see Fig. S 2 of the ESI in [29]). Since `Open Babel` predicts valencies only based on the distance between pairs of atoms, the functional group detection scheme is prone to errors for molecules with rare functional groups. Furthermore, the functional group definitions of `Checkmol` have significant overlaps, e.g., the molecules detected as alkylamines are also detected as primary amines. To ensure that these shortcomings do not influence our conclusions, we base our analyses only on the subset of the 14 k molecules that have certain functional groups. These groups are chosen to be chemically important, non-overlapping, and each of these categories contains at least 500 entries. The number of structures that contain one of these functional groups is 9604. For the analysis in Section 4.1.3, only molecules containing a single functional group are considered, that is, 1626 entries of our dataset (see Fig. 4.6).
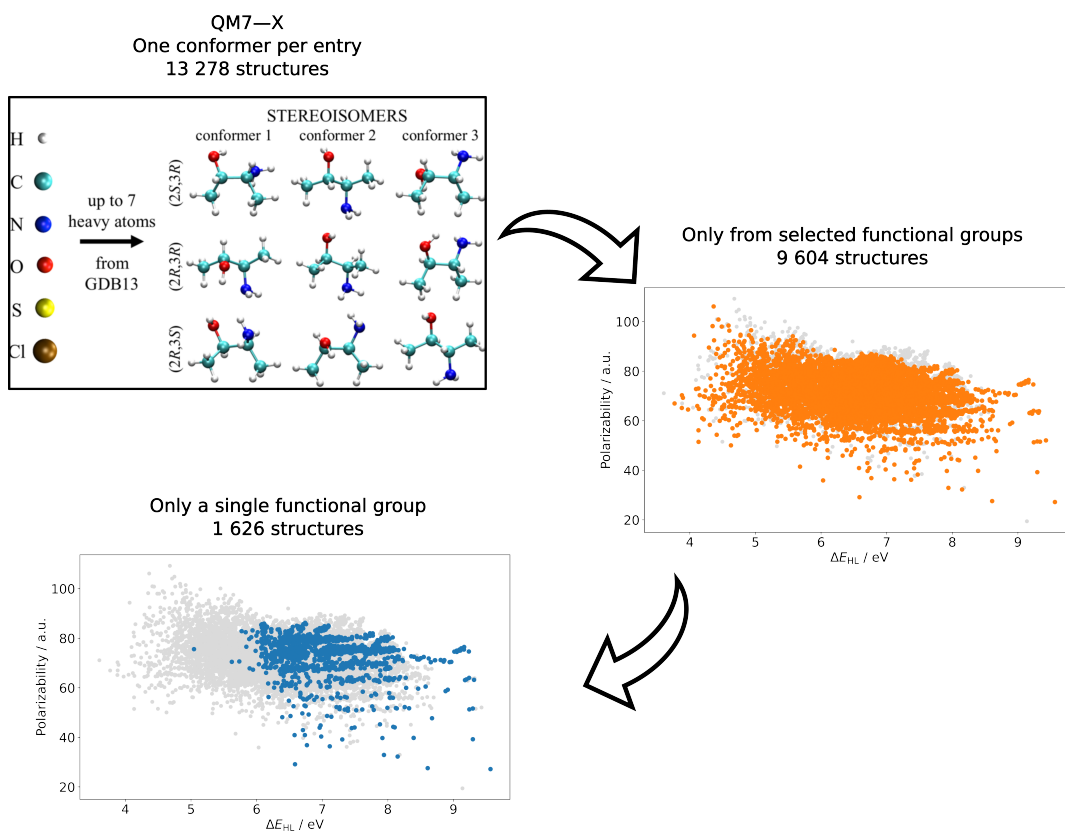
Figure 4.6: Schematic representation of the different data selection we used. Starting from the full QM7–X containing non-equilibrium structures and different conformers, we only select a single equilibrium conformer per entry. Then we select only those molecules that have functional groups exclusively from our selected list. For analyses where functional group labeling is needed, the structures that only have a single functional group are used.

Based on these constraints, the following eleven classes of molecules are identified: aldehydes, carbonitriles, dialkyl ether, enol ether, hydrazones, ketones, oximes, primary alcohols and amines, as well as secondary alcohols and amines (see labels in Fig. 4.4(a)).

## 4.2 Molecular Polarizability from Interacting Atoms in a Molecule

In Sect. 4.1 it was shown that the polarizability of small organic molecules can be well approximated by considering contributions from their constituent atoms. A simple linear combination of atomic values as a first-order approximation was also presented in Sect. 4.1.3. This model can have good

performance for small molecules, but it is limited by not accounting for the chemical environment. This Section will introduce a model based on self-consistently coupled atomic quantum Drude oscillator (QDO) capable of capturing both short-range environment as well as long-range geometrical effects. Building on the four-dimensional scaling law and the spatial properties of QDOs, this model efficiently captures both the magnitude and the anisotropy of the dipole polarizability for small organic molecules.

The effect of the chemical environment can be captured by using a self-consistent screening for response of atoms represented as dipoles. The name of this approach refers to the fact that *screened* atomic polarizabilities are obtained by using a Dyson-like self-consistent screening equation [112, 253]

$$\boldsymbol{\alpha}_j^{\text{SCS}}(\omega) = \boldsymbol{\alpha}_j^0(\omega) - \boldsymbol{\alpha}_j^0(\omega) \sum_{k \neq j} \mathbf{T} \boldsymbol{\alpha}_k^{\text{SCS}}(\omega) \ . \tag{4.8}$$

Here, $\boldsymbol{\alpha}_j^0(\omega)$ is the non-interacting polarizability, whereas the dipole propagator is calculated from an interaction potential as $\mathbf{T} = \boldsymbol{\nabla}_{\boldsymbol{R}_j} \otimes \boldsymbol{\nabla}_{\boldsymbol{R}_k} v(r_{jk})$. The molecular polarizability can then be expressed as a sum of atomic polarizabilities (including the screening effect)

$$\boldsymbol{\alpha}(\omega) = \sum_{j \in \text{atoms}} \boldsymbol{\alpha}_j^{\text{SCS}}(\omega) \ . \tag{4.9}$$

Eqs. (4.8)-(4.9) do not specify the source of the atomic polarizability. In principle, one could obtain these values from the fitting procedure of Sect. 4.1.3. Alternatively, Mayer *et al.* considered them as input parameters of the interacting model itself [41]. The approach I am using in this Section, which is also used in the calculation of dispersion energies in the MBD@RsSCS method, relies on partitioning molecular densities to the constituent atoms (although a range separation step is introduced in that method for the self-consistent screening procedure). The advantage of this way is that the chemical environment is already taken into account up to a certain degree. The most common partitioning method is to rescale free atomic polarizabilities with Hirshfeld ratios [35]

$$\boldsymbol{\alpha}_j^0(\omega) = \boldsymbol{\alpha}_j^{0,\text{free}}(\omega) \frac{\int \mathrm{d}^3\boldsymbol{r} w_j(\boldsymbol{r}) \rho(\boldsymbol{r}) r^3}{\int \mathrm{d}^3\boldsymbol{r} \rho_j^{\text{free}}(\boldsymbol{r}) r^3} \ , \tag{4.10}$$

where $\rho(\boldsymbol{r})$ is the charge density from the quantum mechanical calculation and $\rho_j^{\text{free}}(\boldsymbol{r})$ is the

density of a free (neutral) atom $j$. Hirshfeld weights $w_j(\mathbf{r})$ are obtained from Hirshfeld ratios

$$w_j(\mathbf{r}) = \frac{\rho_j(\mathbf{r})}{\sum_k \rho_k(\mathbf{r})} \ .$$

(4.11)

With the individual atomic parameters determined, the interaction between atoms is obtained by assuming that the interaction potential corresponds to the Coulomb coupling between two Gaussian charge densities (as in the case of QDOs)

$$T_{ab}^{\mathrm{GG}}(\mathbf{r}, \sigma) = \frac{\partial^2}{\partial r_a \partial r_b} \frac{\mathrm{erf}(\zeta)}{r} = \big( \mathrm{erf}(\zeta) - \Theta(\zeta) \big) T_{ab}(\mathbf{r}) + 2\zeta^2 \Theta(\zeta) \frac{r_a r_b}{r^5} \ ,$$

(4.12)

$$\text{where} \qquad \Theta(\zeta) = \frac{2\zeta}{\sqrt{\pi}} e^{-\zeta^2} \ , \qquad \zeta = \frac{r}{\sqrt{\sigma_j^2 + \sigma_k^2}} \ .$$

Finally, the connection between the individual atomic parameters and the interaction length $\sigma$ can be obtained from the limit of the interaction tensor with $j \to k$ [41]

$$\frac{1}{\alpha_k^0} = -\mathbf{T}^{k,k} = \sqrt{\frac{2}{\pi}} \frac{\sigma_k^3}{3} \ .$$

(4.13)

Here, it should be highlighted that while the analysis I present is novel, the model is essentially the same as the zero-frequency limit of the CARMA model described by Gobre [46, 47].

In the following, all analysis will be performed on the equilibrium structures of the QM7–X dataset (see Ref. [183]), which contains around 42000 equilibrium conformers of small organic molecules. The Hirshfeld ratios, on which the SCS method relies, were taken from the published database[1]. Dipole polarizabilities were obtained using the density functional perturbation theory (DFPT) implementation in `FHI-aims` code using the PBE0 hybrid functional and an augmented version of the LIGHT basis set, as described in Sect. 4.1.6 and Ref. [29], where it was also reported that this computational setup provides sub-2% accuracy with respect to reference data from coupled cluster calculations. The SCS calculations were performed by using a modified version of the Python

---

[1]Due to limitations of the `FHI-aims` code, the Hirshfeld ratios in QM7–X [183] were obtained with PBE free atomic and PBE0 atom-in-molecule data. To ensure consistent PBE0 accuracy, these ratios were multiplied with a correction factor obtained from the PBE0 and PBE free atomic volumes for each element. This amounts to an increase of about 5%, depending on the atomic composition of each entry.

interface of LIBMBD [254], turning off the range separation that is otherwise employed for the calculation of interaction energies.

It was concluded in Ref. [19] as well as in Sect. 3.1.1 that the Hirshfeld ratios of Eq. (4.10) need to be re-scaled by raising them to the (4/3)rd power. Therefore, the polarizability within this approach, in what follows, called SCS43, is calculated as

$$
\boldsymbol{\alpha}_j^0(\omega) = \boldsymbol{\alpha}_j^{0,\text{free}}(\omega) \left( \frac{\int \mathrm{d}^3\boldsymbol{r}\, w_j(\boldsymbol{r})\rho(\boldsymbol{r})r^3}{\int \mathrm{d}^3\boldsymbol{r}\, \rho_j^{\text{free}}(\boldsymbol{r})r^3} \right)^{4/3} . \tag{4.14}
$$

I have also evaluated two other prediction schemes, called TS and TS43, where the re-scaled atomic polarizabilities are directly summed, skipping the SCS step of Eq. (4.8). To compare the different methods, the mean absolute percentage error (MAPE) and the skewness (g) of the data were calculated, which are defined as

$$
\begin{aligned}
\text{MAPE} &= \frac{1}{n} \sum_i \frac{\left| \alpha_{\text{pred,i}} - \alpha_{\text{DFPT}} \right|}{\alpha_{\text{DFPT}}} ; \\
g &= \frac{m_3}{m_2^{3/2}} , \quad m_i = \frac{1}{N} \sum_{n=1}^{N} (x - \bar{x})^i .
\end{aligned} \tag{4.15}
$$

The mean absolute percentage errors of the methods "TS", "TS43", and "SCS43" are 21.12%, 12.38%, and 17.92%, with the skewness values $-0.88$, $-1.13$ and $0.11$, respectively. The errors for the methods "TS" and "SCS43" can also be visually observed in Fig. 4.7, with the anisotropy of each entry represented by the color of the individual points. The anisotropy values in Figs. 4.7-4.9 are obtained from the DFPT results with Eq. 4.8 (repeated here for clarity)

$$
(\Delta\boldsymbol{\alpha})^2 = 3(\alpha_{xy}^2 + \alpha_{xz}^2 + \alpha_{yz}^2) + \frac{1}{2}((\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{xx} - \alpha_{zz})^2 + (\alpha_{yy} - \alpha_{zz})^2).
$$

For best results, a method should fulfill two criteria: the lowest percentage errors measured by MAPE as well as a low skewness. The accuracy of the values obtained can be compared with commonly used DFT methods, which is between 2.5% and 3.8% [177] MAPE. Compared to these values, even the best scheme is about three times less accurate. However, it should be kept in mind that predicting the polarizability with atoms-in-molecules methods requires just a ground-state single-point calculation, circumventing the difficulties associated with finite-field and
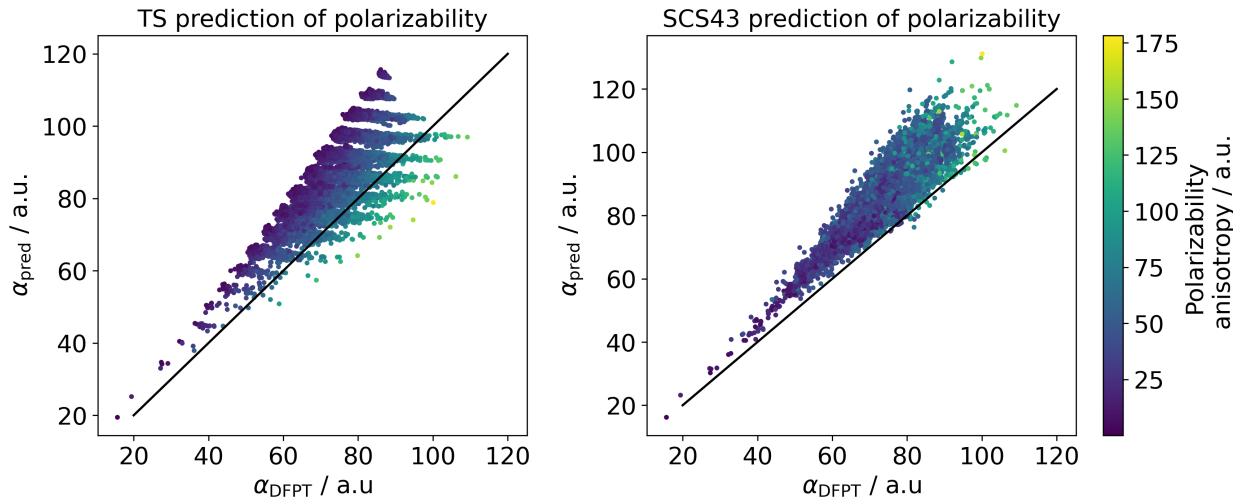
Figure 4.7: Comparing the error for the trace of dipole polarizabilities predicted by the "TS" and "SCS43" methods versus PBE0 reference values on the QM7–X dataset of small organic molecules [183]. The polarizability anisotropy of each molecule, obtained with the PBE0 functional, is also shown.

perturbative DFT calculations. Remarkably, while the errors of all the methods considered stem from an overestimation of polarizability, accounting for the four-dimensional scaling of the dipole polarizability seems to be essential in reducing this overestimation. In particular, the error of the "TS43" scheme is almost two times less than that of the "TS" scheme (21.12% vs. 12.38%), underlining the importance of relying on the correct scaling law.

Skewness measures the "symmetricity" of the distribution. Any method for prediction of molecular polarizabilities which possesses low skewness, even if it systematically over- or underestimates the corresponding quantity, provides one with the opportunity to obtain more accurate results by a simple rescaling the initially predicted values, as centrally distributed around the mean value. The main difference between "TS" methods and the "SCS43" approach is captured by their distinct skewness, showing that systematic error can be essentially eliminated using self-consistent screening and properly adjusting the mean value of the distribution. The difference in skewness is visualized in Fig. 4.8.

The source of the skewness can be understood by considering the interplay between the prediction error and the polarizability anisotropy, either by observing the color distribution in Fig. 4.7 or directly the error as a function of anisotropy in Fig. 4.9. From these graphs, it can be concluded that the "TS" methods cannot adequately account for the polarizability anisotropy, since obviously
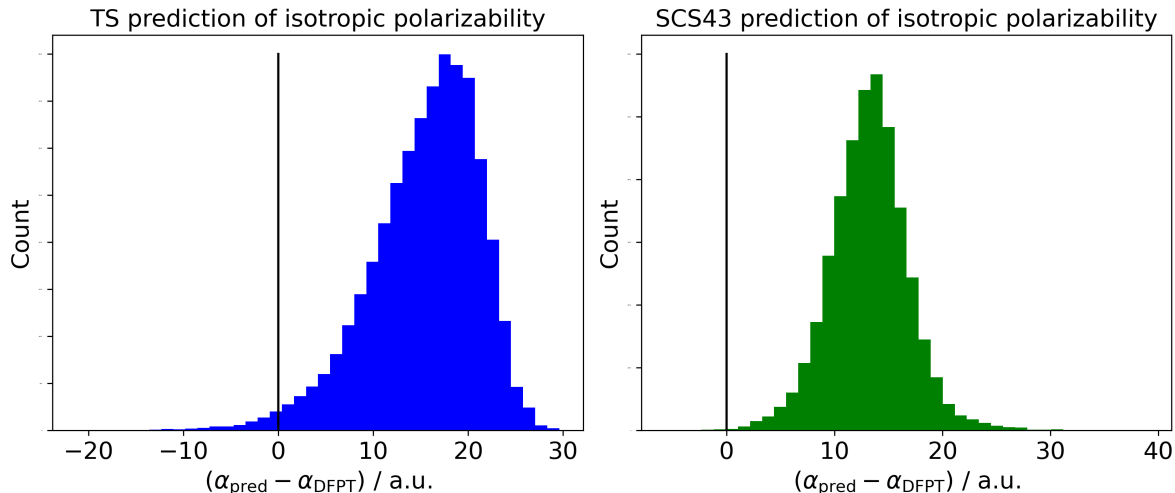
Figure 4.8: The distribution of the errors in the polarizability prediction of the "TS" and "SCS43" methods versus PBE0 reference values on the QM7–X dataset of small organic molecules [183].

the error and the anisotropy strength are strongly correlated. The systematic overestimation for the methods without the SCS step is explained by the relatively large number of points having lower anisotropy in polarizability stemming from the corresponding molecular geometry. This is due to the composition of the dataset, as it is limited by the total number of atoms in the molecules, meaning that branched and therefore more isotropic structures are overrepresented. This inadequacy of the "TS" schemes can be understood by taking into account that the molecular polarizability is built up just as a sum of atomic counterparts, with the individual contributions based on atomic volumes as $V = \int r^3 \rho(\boldsymbol{r}) \, \mathrm{d}^3\boldsymbol{r}$. Spatial contributions to this integral rapidly decrease when moving away from the atomic centers, meaning that long-range effects are not properly taken into account. Accounting for such long-range interactions *via* self-consistent screening leads to a method with an error independent of the polarizability anisotropy, as demonstrated by Fig. 4.9.

Two main shortcomings of this approach lie within the parametrization of the individual oscillators. In particular, localized oscillators could not adequately describe structures where the charge density is highly delocalized, as in the case of carbon nanostructures, where polarization effects must be taken into account by assigning a charge to each oscillator, in addition to the polarizability [112]. Moreover, the localization scheme based on Hirshfeld partitioning cannot adequately account for the contribution of electron correlation to polarizability, as outlined in Sect. 3.1.2. Devising a scheme that accounts for these effects would be highly impactful in the efficient modeling of the response properties of matter.
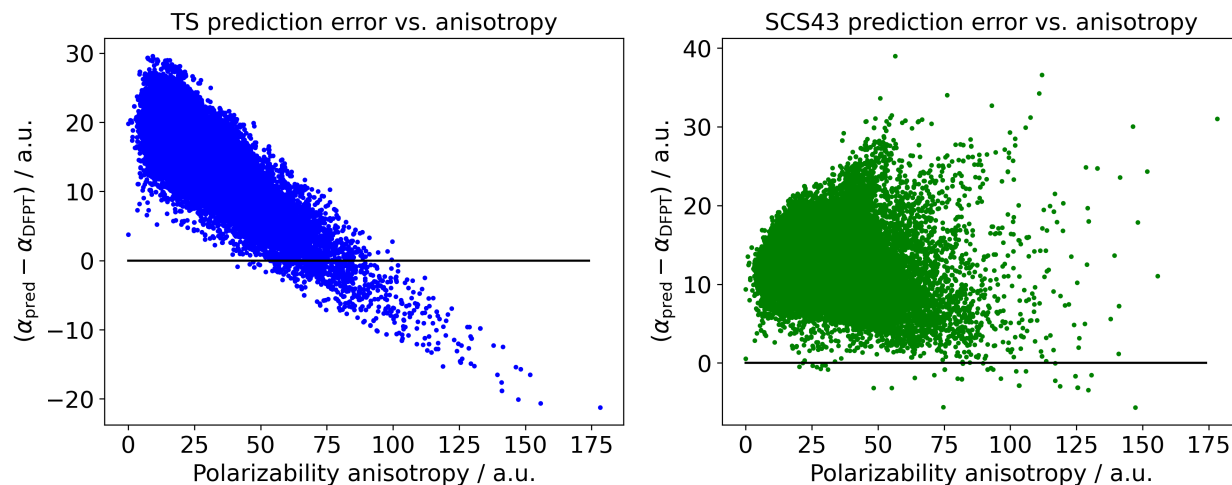
Figure 4.9: Error in the predicted polarizability values of the "TS" and "SCS43" methods versus PBE0 reference values shown as a function of polarizability anisotropy, obtained with the PBE0 functional, on the QM7–X dataset of small organic molecules [183].

As outlined in this section, polarizabilities of small organic molecules can be approximated with good accuracy and without artificial skewness by means of the self-consistent screening procedure, when accompanied with the four-dimensional scaling of the polarizability applied to atoms in a molecule. This approach, combining the screening obtained from the QDO model with the quantum mechanical scaling law of polarizability delivers a model, which is useful not only to calculate interaction energies within the MBD framework but also to deeply understand the response functions of molecules and materials starting from the corresponding atomistic data.

## 4.3 Conclusion and Outlook

Polarizability of small organic molecules was discussed in this chapter, starting with examining the empirical correlations of this quantity with other molecular descriptors. With a data-driven approach based on an extended QM7–X dataset, it was shown that there is no general correlation between the molecular polarizability and the HOMO–LUMO gap, in contrast to a common intuitive assumption (for instance, see Chapter 12 in Ref. [96]), and the most reasonable way to determine the polarizability of molecules should rely on the analysis of their atomic composition.

This finding suggests that the polarizability of organic molecules can be discussed on the basis of an

atom-in-molecule approach. Accordingly, a model where polarizability is approximated through electronic response of dipole-coupled atoms was examined. As a result, it is found that relatively accurate polarizabilities can be obtained using Hirshfeld partitioning to define individual atomic densities and rescaling free atomic polarizabilities using the quantum mechanical scaling law.

While this approach of interacting atomic oscillators seems to provide accurate polarizabilities, it still neglects some important physical effects. In particular, it is based on the coarse-grained atomic contributions, being limited by both the choice of the partitioning scheme as well as the fact that localized schemes cannot account for non-local effects, such as charge-flow contributions. Thus, the next chapter focuses on polarizability as a non-local quantity, enabling one to define spatially resolved contributions to the total polarizability without the need of a predefined level of resolution.

# 5. Polarizability as a Non-Local Quantity

*The new results contained in this Chapter are still unpublished.*

One of the main motivations of this thesis is to construct a generally applicable non-expanded many-body dispersion method, which - as was shown before - needs to be based on non-local response functions. In principle, two physical quantities can be used as a basis for such an approach: susceptibility, which measures the change of electron density to a perturbing potential, or polarizability, measuring the change of polarization density to a perturbing external electric field. While these two functions, in principle, contain the same physics, there is a practical difference between them: integrating the non-local polarizability over the full space will yield the dipole polarizability of the system, whereas non-local susceptibility has to integrate to zero if charge is conserved. The significance of this fundamental difference is that coarse-grained models are better suited for the non-local polarizability, as first pointed out by Dobson [10].

As will be elaborated in Sect. 5.1, the non-local polarizability can be expressed as a polarization-polarization correlation function [63]

$$\alpha_{ij}(\boldsymbol{r}, \boldsymbol{r}', \omega) = \sum_{n \neq 0} \frac{\omega_{n0} \left( \langle 0|\hat{P}_i(\mathbf{r})|n \rangle \, \langle n|\hat{P}_j(\mathbf{r}')|0 \rangle + \text{cc.} \right)}{\omega_{n0}^2 - \omega^2} . \tag{5.1}$$

Sect. 5.2 contains a discussion of this polarization operator, and Sect. 5.3 shows the application of the formalism to model systems. The correlation energy, connected to the dispersion interaction between $A$ and $B$, was already hinted at in the introduction of Fig. 1.2 to be

$$\Delta E = -\frac{1}{2\pi} \int_0^\infty \mathrm{d}\nu \int \alpha_{\delta\alpha}^A(\boldsymbol{r}, \boldsymbol{r}''', i\nu) T_{\alpha\beta}(\boldsymbol{r}''', \boldsymbol{r}'') \alpha_{\beta\gamma}^B(\boldsymbol{r}'', \boldsymbol{r}', i\nu) T_{\gamma\delta}(\boldsymbol{r}', \boldsymbol{r}) \, \mathrm{d}^3\boldsymbol{r} \, \mathrm{d}^3\boldsymbol{r}' \, \mathrm{d}^3\boldsymbol{r}'' \, \mathrm{d}^3\boldsymbol{r}''' . \tag{5.2}$$

Within second order perturbation theory, Eq. (5.2) is exact insofar as not relying on the multipole

expansion. This fact can be shown either by considering that any multipolar moment can be expressed from the non-local polarizability (as demonstrated in [62]), or by using the multipole expansion of the polarization operator (see details in Sect. 5.2, written here just for an one-centered single-electron system)

$$\mathbf{P}(\mathbf{r}) = q \left( \boldsymbol{r_i}\delta(\boldsymbol{r}) - \frac{(\boldsymbol{r_i} \cdot \boldsymbol{\nabla})^2}{2!}\delta(\boldsymbol{r}) + ... \right) \tag{5.3}$$

The first term of this expansion is the dipole term

$$\mathbf{P_d}(\mathbf{r}) = q\boldsymbol{r_i}\delta(\boldsymbol{r}) \ . \tag{5.4}$$

which, together with the connection between the polarization operator and polarizability leads to (with real wavefunctions)

$$\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}')_d = 2 \sum_{n\neq 0} \frac{\langle 0|\hat{\mathbf{P}}_d(\mathbf{r})|n\rangle \otimes \langle n|\hat{\mathbf{P}}_d(\mathbf{r}')|0\rangle}{E_n - E_0} = 2q^2 \sum_{n\neq 0} \frac{\langle 0|\boldsymbol{r_i}\delta(\boldsymbol{r})|n\rangle \otimes \langle n|\boldsymbol{r_i}\delta(\boldsymbol{r}')|0\rangle}{E_n - E_0} \ . \tag{5.5}$$

The individual integrals in the numerator are

$$\langle 0|\boldsymbol{r_i}\delta(\boldsymbol{r})|n\rangle = \int \Psi_0(\boldsymbol{r_i})\boldsymbol{r_i}\Psi_n(\boldsymbol{r_i})\delta(\boldsymbol{r}) \, \mathrm{d}\boldsymbol{r_i} = \delta(\boldsymbol{r}) \, \langle 0|\boldsymbol{r_i}|n\rangle \ , \tag{5.6}$$

meaning that the dipole contribution to the non-local polarizability is given by

$$\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}')_d = 2q^2\delta(\boldsymbol{r})\delta(\boldsymbol{r}') \sum_{n\neq 0} \frac{\langle 0|\boldsymbol{r_i}|n\rangle \otimes \langle n|\boldsymbol{r_i}|0\rangle}{E_n - E_0} = 2q^2\delta(\boldsymbol{r})2\delta(\boldsymbol{r}')\boldsymbol{\alpha} \ , \tag{5.7}$$

corresponding to the coarse-grained expression of the dipole polarizability, comparable with the MBD ansatz [253] in the one-center limit.

The aim of this Chapter is to evaluate the non-local polarizability for some model systems. On this way, non-local polarizability will first be contrasted with susceptibility in Sect. 5.1. After showing

that non-local polarizability should be expressed *via* the polarization operator, the properties of this object will be discussed in Sect. 5.2, which will be used to elaborate the non-local polarizabilities in Sect. 5.3.

While most of this Chapter contains results obtained by me, discussions with Dr. Péter Szabó, Dr. Matteo Gori, and Dr. Dmitry Fedorov were essential in developing them.

## 5.1 Connecting Non-local Susceptibility and Polarizability

To derive the connection between polarizability $\boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}')$ and susceptibility $\chi(\mathbf{r}, \mathbf{r}')$, one starts with the definition of these quantities as general linear response functions [255]. The dielectric susceptibility will relate the change of the charge density when an external potential is applied, while the polarizability will relate the change in polarization with an applied perturbing field [256] (note that non-local polarizability is a tensor field connecting two vector quantities)

$$\delta\rho(\mathbf{r}) = \int_V \chi(\mathbf{r}, \mathbf{r}')\delta\phi(\mathbf{r}')\, \mathrm{d}^3\mathbf{r}' \tag{5.8}$$

and

$$\delta\mathbf{P}(\mathbf{r}) = \int_V \boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}')\delta\mathbf{E}(\mathbf{r}')\, \mathrm{d}^3\mathbf{r}' \ . \tag{5.9}$$

In addition to these equations, Gauss's law can be used, connecting the electron density to the divergence of the polarization density [11] (omitting the delta $\delta$ symbol from now on)

$$-\rho(\mathbf{r}) = \boldsymbol{\nabla}_r \cdot \mathbf{P}(\mathbf{r}) - \rho^{\text{true}}(\boldsymbol{r}) \ . \tag{5.10}$$

In Eq. (5.10), $\rho^{\text{true}}(\boldsymbol{r})$ represents the electron density due to the *total* charge of a system, which is zero for neutral atoms and molecules – the polarization field is only defined for the part of the charge density that is offset by the nuclear charge.

Taking the divergence of Eq. (5.9) and using Eq. (5.10) and the fact that the divergence is taken

with respect to a variable different from the integration, one obtains

$$\rho(\mathbf{r}) = -\nabla_r \cdot \int_V \boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}') \mathbf{E}(\mathbf{r}') \, d^3\mathbf{r}' = -\int_V \nabla_r \cdot \big(\boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}') \mathbf{E}(\mathbf{r}')\big) \, d^3\mathbf{r}' \ . \tag{5.11}$$

Let us now evaluate the divergence of a product between a tensor field and a vector field. However, since the vector field is not defined in the same space as the divergence operator (the electric field is a function of $\mathbf{r}'$, whereas we take the divergence with respect to $\mathbf{r}$), it can be regarded as a constant and, therefore, factored out.

$$\int_V \nabla_r \cdot \big(\boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}') \mathbf{E}(\mathbf{r}')\big) \, d^3\mathbf{r}' = \int_V \big(\nabla_r \cdot \boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}')\big) \cdot \mathbf{E}(\mathbf{r}') \, d^3\mathbf{r}' \ . \tag{5.12}$$

The next step is to note that the electric field can always be written as the gradient of an electric potential: $\mathbf{E}(\mathbf{r}) = -\nabla\psi(\mathbf{r})$. Thus, we have

$$\rho(\mathbf{r}) = \int_V \big(\nabla_r \cdot \boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}')\big) \cdot \nabla_{r'}\phi(\mathbf{r}') \, d^3\mathbf{r}' \ . \tag{5.13}$$

To proceed further, we look at an identity of the divergence operator when applied to a product of a scalar and a vector field (note that the divergence of the non-local polarizability tensor is just a vector in $\mathbf{r}'$)

$$\nabla_{r'} \cdot (\phi\mathbf{A}) = \phi\nabla_{r'} \cdot \mathbf{A} + (\nabla_{r'}\phi) \cdot \mathbf{A} \ . \tag{5.14}$$

Integrating both sides of this equation and using the Gauss-Ostrogradsky theorem for the left side yields

$$\oiint_{\partial V} (\phi\mathbf{A}) \cdot \mathbf{n}_{r'} \, d\sigma = \int_V \phi\nabla_{r'} \cdot \mathbf{A} \, dV + \int_V (\nabla_{r'}\phi) \cdot \mathbf{A} \, dV \ . \tag{5.15}$$

In our case, the integration region $V$ is the full space, and integrating over the surface of the full space (by taking, for example, a sphere with radius going to infinity) gives zero for functions that go to zero at infinity. So, we can conclude that the right-hand side of Eq. (5.13) can be recast as

$$\int_V \left( \mathbf{\nabla}_r \cdot \boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}') \right) \cdot \mathbf{\nabla}_{r'} \phi(\mathbf{r}') \, \mathrm{d}^3 \mathbf{r}' = - \int_V \left( \mathbf{\nabla}_{r'} \cdot \mathbf{\nabla}_r \cdot \boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}') \right) \phi(\mathbf{r}') \, \mathrm{d}^3 \mathbf{r}' \; . \tag{5.16}$$

Comparing our result with Eq. (5.8), we obtain

$$\chi(\mathbf{r}, \mathbf{r}') = -\mathbf{\nabla}_{r'} \cdot \mathbf{\nabla}_r \cdot \boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}') \; . \tag{5.17}$$

This operation is expressed in component form as

$$\chi(\mathbf{r}, \mathbf{r}') = -\sum_{ij} \frac{\partial^2}{\partial r_i \partial r'_j} \alpha_{ij}(\boldsymbol{r}, \boldsymbol{r}') \; . \tag{5.18}$$

Finally, if we consider that the non-local susceptibility is understood as a density-density correlation function expressed *via* operators (see Sect. 5.2.1 for the transition between the macroscopic quantities and their corresponding operators)

$$\chi(\mathbf{r}, \mathbf{r}') = 2 \sum_{n \neq 0} \frac{\langle 0 | \hat{\rho}(\mathbf{r}) | n \rangle \, \langle n | \hat{\rho}(\mathbf{r}') | 0 \rangle}{E_n - E_0} \; , \tag{5.19}$$

then, utilizing Eq. (5.10) we find that non-local polarizability will act as a polarization-polarization correlation function

$$\alpha_{ij}(\boldsymbol{r}, \boldsymbol{r}') = 2 \sum_{n \neq 0} \frac{\langle 0 | \hat{P}_i(\mathbf{r}) | n \rangle \, \langle n | \hat{P}_j(\mathbf{r}') | 0 \rangle}{E_n - E_0} \; . \tag{5.20}$$

The expression for the dispersion energy using the non-local polarizability of Eq. (5.2) arises similarly, by starting from the non-local susceptibility (which is better known for expressing correlation energies [257]) similarly, by starting from a known expression utilizing the non-local susceptibility

$$\Delta E = -\frac{1}{2\pi} \int_0^\infty \mathrm{d}\nu \int \frac{\chi^A(\boldsymbol{r}, \boldsymbol{r}''', i\nu) \chi^B(\boldsymbol{r}'', \boldsymbol{r}', i\nu)}{|\mathbf{r} - \mathbf{r}''||\mathbf{r}' - \mathbf{r}'''|} \, \mathrm{d}^3 \boldsymbol{r} \, \mathrm{d}^3 \boldsymbol{r}' \, \mathrm{d}^3 \boldsymbol{r}'' \, \mathrm{d}^3 \boldsymbol{r}''' \; , \tag{5.21}$$

and utilizing the connection between the two response functions as above.

In conclusion, understanding the theory of non-local polarizability requires the discussion of polarization density as an operator. Before doing this in the next Section, however, it should be pointed out that the non-local polarizability, even though it was obtained by "smearing" the dipole polarizability, contains all higher order response properties [62]. Therefore, a model of intermolecular interactions based directly on the non-local polarizability will not suffer from the errors of the multipole expansion.

## 5.2 Polarization Operator for Isolated Systems

### 5.2.1 Electron Density and Electric Polarization as Operators

The non-local response quantities introduced in Sect. 5.1 come from electrodynamics and can be used equally in classical and quantum mechanical contexts. A simple quantum mechanical argument, based on second-order perturbation theory, was presented in [4], where it is shown that a perturbing potential $V(\boldsymbol{r})$ only at position $\mathbf{s}$ (that is, $V(\boldsymbol{r}) = V_0\delta(\boldsymbol{r} - \mathbf{s})$), for which the perturbing Hamiltonian is given as $\mathcal{H}' = \int_V V_0\delta(\boldsymbol{r} - \mathbf{s})\rho(\boldsymbol{r})\,\mathrm{d}\boldsymbol{r} = V_0\rho(\mathbf{s})$, results in a first-order shift in the charge density

$$\Delta\rho(\boldsymbol{r}) = -V_0\frac{\langle 0|\delta(\boldsymbol{r}_i - \mathbf{s})|n\rangle\,\langle n|\delta(\boldsymbol{r}_i - \boldsymbol{r})|0\rangle + cc.}{E_n - E_0} \ , \tag{5.22}$$

which is written as

$$\Delta\rho(\boldsymbol{r}) = -\int_V V(\boldsymbol{r}')\chi(\boldsymbol{r},\boldsymbol{r}')\,\mathrm{d}^3\boldsymbol{r}' = -\int_V V_0\delta(\boldsymbol{r}' - \mathbf{s})\chi(\boldsymbol{r},\boldsymbol{r}')\,\mathrm{d}^3\boldsymbol{r}' \ , \tag{5.23}$$

thereby identifying the non-local susceptibility (often referred to as charge density susceptibility) to be equal to the form in Eq. (5.21) for real wavefunctions.

The polarizability operator $\hat{\mathbf{P}}(\mathbf{r})$ was introduced in Eq. (5.20) based on the analogy of response functions in Eq. (5.21) and Eq. (5.20), while the connection between the two quantities was made on the level of expectation values in Eq. (5.10). To satisfy the constraint on the expectation values,

we postulate that the same equation holds between the operators

$$-\hat{\rho}(\mathbf{r}) = \boldsymbol{\nabla} \cdot \hat{\mathbf{P}}(\mathbf{r}) \ . \tag{5.24}$$

Formally, both $\hat{\rho}(\mathbf{r})$ and $\hat{\mathbf{P}}(\mathbf{r})$ could be written as belonging to a class of operators of the form

$$\hat{O}_{\mathbf{r}_k}(\mathbf{r}) = \hat{f}(\mathbf{r}) \otimes \sum_{n=1}^{N} \hat{I}_{(n \neq k)} \ , \tag{5.25}$$

where we act with the identity operator on all particles except the $k$th, for which we act on with an arbitrary one-electron operator $\hat{f}$. All the dependence of the external parameters is carried in $\hat{f}$. In the case of the charge density, we can write this operator as

$$\hat{\rho}_{\mathbf{r}_k}(\mathbf{r}) = |\mathbf{r}\rangle \langle\mathbf{r}| \otimes \sum_{n=1}^{N} \hat{I}_{(n \neq k)} \ . \tag{5.26}$$

An alternative way to write Eq. (5.26) is using the Dirac delta function.

$$\hat{\rho}_{\mathbf{r}_k}(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r_k}) \sum_{n=1}^{N} \hat{I}_{(n \neq k)} \ , \tag{5.27}$$

where it is clear that the operator $f$ acts as multiplication with the Dirac delta. Since all dependence on the real space coordinate $r$ is contained in a multiplicative term in Eq. (5.27), without loss of generality, we can also write the polarization operator for a time-independent system as

$$\hat{\mathbf{P}}_{\mathbf{r}_k}(\mathbf{r}) = \mathbf{P}_{\mathbf{r_k}}(\mathbf{r}) \sum_{n=1}^{N} \hat{I}_{(n \neq k)} \ , \tag{5.28}$$

with all the real space dependence is in a multiplicative function $P_{\mathbf{r_k}}(\mathbf{r})$. To satisfy Eq. (5.24), this function will need to satisfy the condition

$$\boldsymbol{\nabla} \cdot \mathbf{P}_{\mathbf{r_k}}(\mathbf{r}) = -\delta(\mathbf{r} - \mathbf{r_k}) \ . \tag{5.29}$$

The solution of Eq. (5.29), together with Eq. (5.28), provides us with a consistent definition of the polarization operator. A possible solution is obtained considering the multipole expansion of the Dirac delta function [258–260]. In the most general form, the multipole expansion along an arbitrary center $\mathbf{R}$ can proceed as

$$\rho(\mathbf{r}) = \sum_k q_k \sum_{n=0}^{\infty} \frac{1}{n!} (-(\mathbf{r_k} - \mathbf{R}) \cdot \nabla)^n \delta(\mathbf{r} - \mathbf{R}) . \tag{5.30}$$

If we consider a system with net zero charge, then the $n = 0$ term of Eq. (5.30) gives zero. The solution of Eq. (5.29) for a neutral system with this expansion in mind is given as

$$\mathbf{P}(\mathbf{r}) = \sum_k q_k (\mathbf{r_k} - \mathbf{R}) \left( \delta(\mathbf{r} - \mathbf{R}) - \frac{1}{2!}(\mathbf{r_k} - \mathbf{R}) \cdot \nabla \delta(\mathbf{r} - \mathbf{R}) + \frac{1}{3!}((\mathbf{r_k} - \mathbf{R}) \cdot \nabla)^2 \delta(\mathbf{r} - \mathbf{R}) - ... \right) . \tag{5.31}$$

This can be compared with the Taylor expansion of the function

$$\sum_k q_k (\mathbf{r_k} - \mathbf{R}) \delta(\mathbf{r} - \mathbf{r_k}) = \sum_k q_k (\mathbf{r_k} - \mathbf{R}) \sum_{n=0}^{\infty} \frac{1}{n!} (-(\mathbf{r_k} - \mathbf{R}) \cdot \nabla)^n \delta(\mathbf{r} - \mathbf{R}) , \tag{5.32}$$

where we can realize that each term is off by a factor of $1/n$. This can be included using the integral identity $\int_0^1 \lambda^{n-1} \, d\lambda = 1/n$, giving us the expression [258, 261]

$$\mathbf{P_{r_k}}(\mathbf{r}) = - \sum_k q_k (\mathbf{r_k} - \mathbf{R}) \int_0^1 \delta(\mathbf{r} - \mathbf{R} - \lambda(\mathbf{r_k} - \mathbf{R})) \, d\lambda . \tag{5.33}$$

Note once more that the polarization operator, if defined this way, does not include the part of the charge density that is due to the net charge of the system; this *true* density is taken into account in Eq. (5.10).

The simplest application of Eq. (5.33) is the case of the one center one electron system, where we take the center of charge to be the origin. Then, we have

$$\hat{\mathbf{P}}_{\mathbf{r_0}}(\mathbf{r}) = -\mathbf{r_0} \int_0^1 \delta(\mathbf{r} - \lambda\mathbf{r_0}) \, d\lambda , \tag{5.34}$$

and the polarization density reads

$$\langle\Psi|\hat{\mathbf{P}}_{\mathbf{r}_0}(\mathbf{r})|\Psi\rangle = \mathbf{P}(\mathbf{r}) = -q \int \mathbf{r_0} \int_0^1 \delta(\mathbf{r} - \lambda\mathbf{r_0})\, \mathrm{d}\lambda |\Psi(\mathbf{r_0})|^2 \, \mathrm{d}\mathbf{r_0} \ . \tag{5.35}$$

We can eliminate one of the integrals by considering the transformation $\mathbf{r}' = \lambda\mathbf{r_0}$ to give

$$\mathbf{P}(\mathbf{r}) = -q \int_V \frac{\mathbf{r}'}{\lambda} \int_0^1 \delta(\mathbf{r} - \mathbf{r}')\, \mathrm{d}\lambda \left|\Psi\left(\frac{\mathbf{r}'}{\lambda}\right)\right|^2 \lambda^{-3}\, \mathrm{d}^3\mathbf{r}' = -q\mathbf{r}\int_0^1 \left|\Psi\left(\frac{\mathbf{r}}{\lambda}\right)\right|^2 \lambda^{-4}\, \mathrm{d}\lambda \ . \tag{5.36}$$

Note that the factor $\lambda^{-3}$ appears due to having been integrated in the 3D space over $\mathrm{d}^3\mathbf{r_0}$. For different dimensions, this factor is different: for an $N$-dimensional space, one gets $\lambda^{-N}$ from $\mathrm{d}^N\mathbf{r_0}$. The scaling of $\mathbf{r}$ is also best treated in Cartesian coordinates where each component is scaled in the same way. In spherical coordinates only the radial component is scaled, since every other component is defined via ratios of the Cartesian ones.

## 5.2.2  Polarization Density of the 1D Quantum Drude Oscillator

If one considers a 1D quantum mechanical system, such as the 1D Drude oscillator, then Gauss's equation will have a straightforward solution

$$\frac{\partial P(x)}{\partial x} = -\rho(x) \quad\Longrightarrow\quad P(x) = -\int^x \rho(y)\, \mathrm{d}y \ . \tag{5.37}$$

This means that one can directly calculate the polarization densities, without the need to resort to the operator formalism developed for the general case. The ground-state polarization density for the 1D QDO is given by

$$P_{00}(x) = -\frac{q}{2}\, \mathrm{erf}(ax) \ , \tag{5.38}$$

with $a = \sqrt{\frac{m\omega}{\hbar}}$ scaling the coordinate of the QDO, and the error function defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)\, \mathrm{d}t \ . \tag{5.39}$$

96

Note that the polarization density, as expressed by Eq. (5.38), is an even function, so the integral of the function gives zero, in accordance with the matrix element $\langle 0|\hat{x}|0\rangle$. To obtain the dipole of the system, we can get the polarization density in a constant external electric field $E$, since the exact wavefunction is known. The polarization density in an external field is

$$P_{00}(x) = -\frac{q}{2}\,\mathrm{erf}(a(x - \alpha E/q)) , \qquad (5.40)$$

where $\alpha$ is the polarizability of the Drude oscillator, which is also the value of the integral $\int_0^\infty P(x)\,\mathrm{d}x$. This is shown schematically in Fig. 5.1.
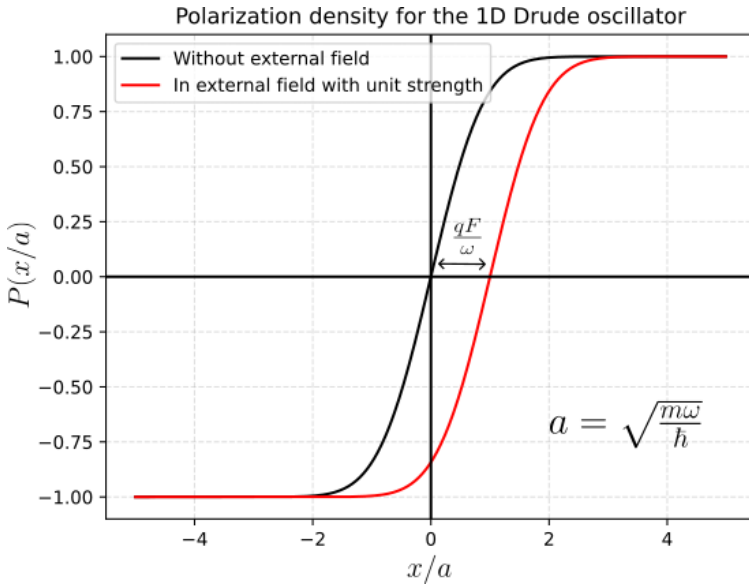


Figure 5.1: Polarization density of the ground state of QDO ($m = \hbar = \omega = q = 1$) with and without an external field of strength $|F| = 1$ a.u.

To evaluate the transition polarization densities, it is helpful to consider the following identity

$$\frac{\mathrm{d}\Psi_n(ax)}{\mathrm{d}x} = \frac{a}{\sqrt{2}}\left(\sqrt{n}\Psi_{n-1}(ax) - \sqrt{n+1}\Psi_{n+1}(ax)\right) . \qquad (5.41)$$

Using the definition of the transition density matrix, $\rho_{nm} = \langle \Psi_n|\Psi_m\rangle$, we can express the transition polarization density matrix as

$$P_{0,n}(ax) = \frac{\rho_{0,n-1}(ax)}{\sqrt{2n}a} . \qquad (5.42)$$

Equation (5.42) provides a direct formula for evaluating any transition polarization for the 1D Drude oscillator. Note that by deriving this formula, we relied on the fact that one of the states in question is the ground state, and such a simple formula will not be valid for exited-excited transitions.

### 5.2.3  Polarization Density of the Hydrogen Atom

While it is now well established that quantum Drude oscillators are good models for atomic response, it is nevertheless a natural development to evaluate the polarization density and the non-local polarizability on real atoms. In the realm of analytical derivations, this is only possible for the hydrogen atom, since the lack of electron correlation leads to analytical expressions for the wavefunctions of all states for the free hydrogen. To discuss the polarization density of the hydrogen atom, one needs to evaluate Eq. (5.36) directly with the ground state density to get (see also [258])

$$\mathbf{P}_{1s,1s}(\mathbf{r}) = -\frac{qr\mathbf{e_r}}{\pi a^3} \int_0^1 e^{-2r/\lambda a} \lambda^{-4} \, \mathrm{d}\lambda = \frac{-q\mathbf{e_r}}{4\pi r^2} e^{-2r/a} \left( 1 + 2(r/a) + 2(r/a)^2 \right) \ . \tag{5.43}$$

It is interesting to note that this vector field is completely radial, having zero curl. The divergence of the field is

$$\boldsymbol{\nabla} \cdot \mathbf{P}_{1s,1s}(\mathbf{r}) = \frac{1}{r^2} \frac{\partial}{\partial r} \frac{-q}{4\pi} e^{-2r/a} \left( 1 + 2(r/a) + 2(r/a)^2 \right) = \frac{q}{\pi a^3} e^{-2r/a} \ , \tag{5.44}$$

which gives us back the correct electron density, in agreement with Eq. (5.24).

Some additional interesting properties can be observed of $\mathbf{P}_{1s,1s}(\mathbf{r})$ when plotting the radial density of the contribution to the 3D integral (defined as $\mathbf{e_r} \cdot \mathbf{P}_{1s,1s}(\mathbf{r})4\pi r^2$, similar to the probability density often defined for the hydrogen atom when calculating the Bohr radius). First, the function has an inflection point at $r = a$, the Bohr radius, due to the connection between the electron density and the polarization density field. Second, the function decays relatively quickly within the van der Waals radius: the value of the function falls exactly to 0.05% at 3.148 a.u., meaning that almost 98%, of the function is found within the van der Waals radius of hydrogen. The ground-state polarization field and its density are shown in Fig. 5.2.

The polarization operator has the advantage that mixed-state, or transition polarization densities
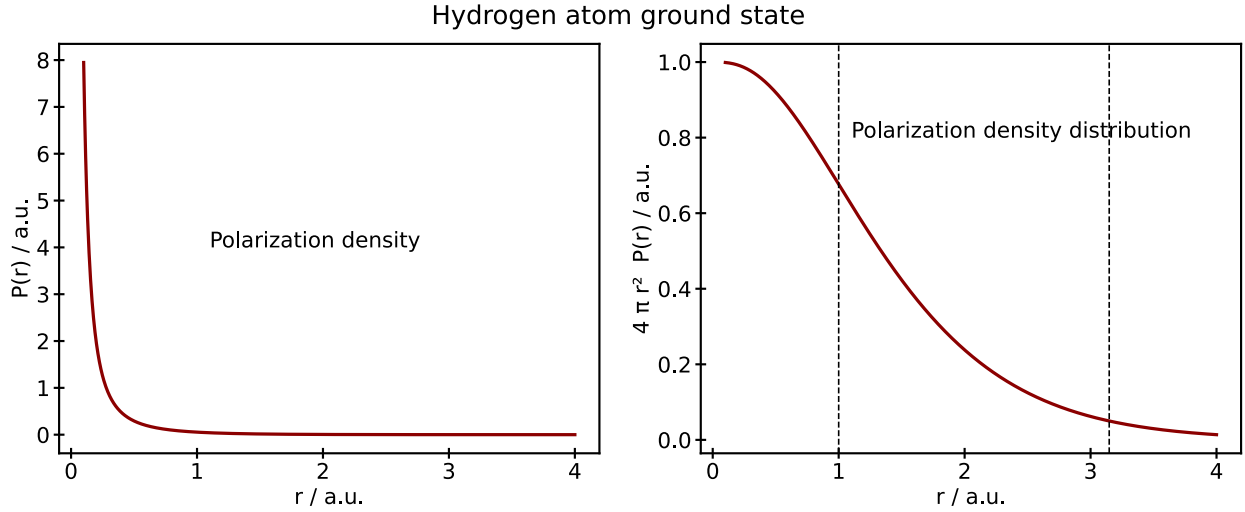
Hydrogen atom ground state



Figure 5.2: Polarization field and the density of the polarization field in the ground state of the hydrogen atom, with the Bohr and van der Waals radii marked.

can also be evaluated. For example, the transition polarization density between the $1s$ ground state and an excited $2s$ state is given by the field

$$\mathbf{P}_{1s,2s}(\mathbf{r}) = -\frac{re^{-\frac{3r}{2a_0}}}{6\sqrt{2}\pi a_0^3}\,\mathbf{e_r}\ . \tag{5.45}$$

For a more detailed analysis of the polarization density of the hydrogen atom, see [258]. Note that by Eq. (5.34) the polarization density, as a vector field, will only have a radial component when expressed in spherical coordinates. However, for most of the transition fields, this radial magnitude will have an angular dependence, resulting in the vector field not being irrotational.

In discussing the transition polarization densities, it is helpful to consider the wavefunction of the hydrogen atom as

$$\Psi_{n,l,m}(r, \theta, \phi) = N\frac{u_{n,l}(r)}{r}Y_{l,m}(\theta, \phi)\ , \tag{5.46}$$

where $N$ is the proper normalizing factor, $u(r)$ is the radial solution to the corresponding Sturm-Liouville problem of the central potential and all angular dependence is contained in the spherical harmonics $Y_{l,m}$. With this expression, the polarization field reduces to

$$\langle nlm|\mathbf{P}(\mathbf{r})|n'l'm'\rangle = \mathbf{e_r}N_{n,l,m}N_{n',l',m'}Y_{l,m}(\theta,\phi)Y_{l',m'}(\theta,\phi)\mathbf{r}\int_0^1 u_{n,l}(\mathbf{r}/\lambda)u_{n',l'}(\mathbf{r}/\lambda)\lambda^{-4}\,\mathrm{d}\lambda\;.$$

(5.47)

Equation (5.47) will be used in Sect. 5.2.4 in comparing the hydrogen atom with the isotropic harmonic oscillator, but since the wavefunction of the hydrogen atom is rather simple, full expressions for the ground state of the hydrogen atom between the $1s$ and the $2s$, $2p$, $3s$, $3p$ and $3d$ orbitals are presented here with all possible magnetic quantum numbers. These functions could be used to systematically approximate the non-local polarizability density by truncating the perturbative summation in Eq. (5.20).

$$\mathbf{P}_{1s,1s}(\mathbf{r}) = \frac{e^{-\frac{2r}{a_0}}\left(a_0^2 + 2a_0r + 2r^2\right)}{4\pi a_0^2 r^2}\,\mathbf{e_r}$$

(5.48)

$$\mathbf{P}_{1s,2s}(\mathbf{r}) = -\frac{re^{-\frac{3r}{2a_0}}}{6\sqrt{2}\pi a_0^3}\,\mathbf{e_r}$$

(5.49)

$$\mathbf{P}_{1s,2p_{-1}}(\mathbf{r}) = \frac{\left(16a_0^3 + 24a_0^2 r + 18a_0 r^2 + 9r^3\right)\sin(\theta)e^{-\frac{3r}{2a_0}-i\phi}}{108\pi a_0^3 r^2}\,\mathbf{e_r}$$

(5.50)

$$\mathbf{P}_{1s,2p_0}(\mathbf{r}) = \frac{e^{-\frac{3r}{2a_0}}\left(16a_0^3 + 24a_0^2 r + 18a_0 r^2 + 9r^3\right)\cos(\theta)}{54\sqrt{2}\pi a_0^3 r^2}\,\mathbf{e_r}$$

(5.51)

$$\mathbf{P}_{1s,2p_{+1}}(\mathbf{r}) = -\frac{\left(16a_0^3 + 24a_0^2 r + 18a_0 r^2 + 9r^3\right)\sin(\theta)e^{-\frac{3r}{2a_0}+i\phi}}{108\pi a_0^3 r^2}\,\mathbf{e_r}$$

(5.52)

$$\mathbf{P}_{1s,3s}(\mathbf{r}) = \frac{re^{-\frac{4r}{3a_0}}\left(r - 6a_0\right)}{54\sqrt{3}\pi a_0^4}\,\mathbf{e_r}$$

(5.53)

$$\mathbf{P}_{1s,3p_{-1}}(\mathbf{r}) = \frac{\left(243a_0^4 + 324a_0^3r + 216a_0^2r^2 + 96a_0r^3 - 32r^4\right)\sin(\theta)e^{-\frac{4r}{3a_0}-i\phi}}{3456\pi a_0^4 r^2}\,\mathbf{e_r} \tag{5.54}$$

$$\mathbf{P}_{1s,3p_0}(\mathbf{r}) = \frac{e^{-\frac{4r}{3a_0}}\left(243a_0^4 + 324a_0^3r + 216a_0^2r^2 + 96a_0r^3 - 32r^4\right)\cos(\theta)}{1728\sqrt{2}\pi a_0^4 r^2}\,\mathbf{e_r} \tag{5.55}$$

$$\mathbf{P}_{1s,3p_{+1}}(\mathbf{r}) = -\frac{\left(243a_0^4 + 324a_0^3r + 216a_0^2r^2 + 96a_0r^3 - 32r^4\right)\sin(\theta)e^{-\frac{4r}{3a_0}+i\phi}}{3456\pi a_0^4 r^2}\,\mathbf{e_r} \tag{5.56}$$

$$\mathbf{P}_{1s,3d_{-2}}(\mathbf{r}) = \frac{\left(243a_0^4 + 324a_0^3r + 216a_0^2r^2 + 96a_0r^3 + 32r^4\right)\sin^2(\theta)e^{-\frac{4r}{3a_0}-2i\phi}}{6912\pi a_0^4 r^2}\,\mathbf{e_r} \tag{5.57}$$

$$\mathbf{P}_{1s,3d_{-1}}(\mathbf{r}) = \frac{\left(243a_0^4 + 324a_0^3r + 216a_0^2r^2 + 96a_0r^3 + 32r^4\right)\sin(2\theta)e^{-\frac{4r}{3a_0}-i\phi}}{6912\pi a_0^4 r^2}\,\mathbf{e_r} \tag{5.58}$$

$$\mathbf{P}_{1s,3d_0}(\mathbf{r}) = \frac{e^{-\frac{4r}{3a_0}}\left(243a_0^4 + 324a_0^3r + 216a_0^2r^2 + 96a_0r^3 + 32r^4\right)(3\cos(2\theta)+1)}{6912\sqrt{6}\pi a_0^4 r^2}\,\mathbf{e_r} \tag{5.59}$$

$$\mathbf{P}_{1s,3d_{+1}}(\mathbf{r}) = -\frac{\left(243a_0^4 + 324a_0^3r + 216a_0^2r^2 + 96a_0r^3 + 32r^4\right)\sin(2\theta)e^{-\frac{4r}{3a_0}+i\phi}}{6912\pi a_0^4 r^2}\,\mathbf{e_r} \tag{5.60}$$

$$\mathbf{P}_{1s,3d_{+2}}(\mathbf{r}) = \frac{\left(243a_0^4 + 324a_0^3r + 216a_0^2r^2 + 96a_0r^3 + 32r^4\right)\sin^2(\theta)e^{-\frac{4r}{3a_0}+2i\phi}}{6912\pi a_0^4 r^2}\,\mathbf{e_r} \tag{5.61}$$

As per construction of the polarization density operator, the integrals of the field give the corresponding transition dipole matrix elements, *e.g.*
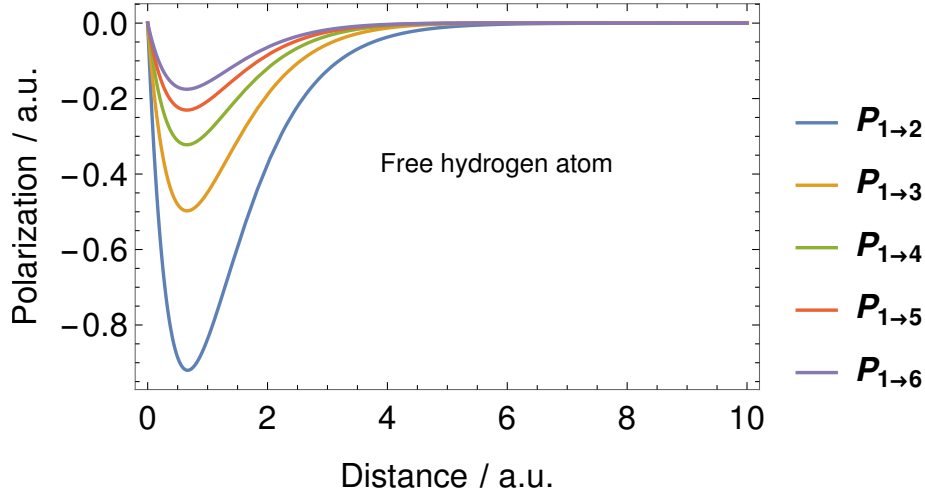
Figure 5.3: Transition polarization density fields between the ground state and the first few *s* states of the hydrogen atom.

$$\int_V \mathbf{P}_{nm,n'm'}(\mathbf{r})\, \mathrm{d}^3\boldsymbol{r} = \langle nm|\hat{\boldsymbol{r}}|n'm'\rangle \ . \tag{5.62}$$

It is known for the case of hydrogen atom that the transition dipole matrix elements between the ground state and higher excited states slowly converge to zero, leading to the fact that one needs to account for rather large excitations for calculating the polarizability, including 18.6% contribution from the continuum for dipole polarizability [81]. This convergence with respect to *n* integral can be implied from Fig. 5.3, showing that the transition polarization fields between the ground state and the higher excited *s* states qualitatively have the same shape, but decrease in magnitude.

## 5.2.4 Polarization Density of the 3D Isotropic Drude Oscillator

Apart from the hydrogen atom presented in Sect. 5.2.3 above, the 3-dimensional isotropic harmonic oscillator is also an example of a model system having a solvable central potential. In general, the N-dimensional quantum harmonic oscillator wavefunction can be expressed by the product of N 1-dimensional oscillators in Cartesian coordinates, with the total energy being

$$E = \sum_{i=1}^{N} (n_i + 0.5)\hbar\omega_i \ , \tag{5.63}$$

where $n_i$ and $\omega_i$ are the quantum number and frequency of the independent Cartesian oscillators. If all frequencies are equal in a 3D oscillator, it is possible to directly solve the radial equation instead of the Cartesian ones, giving a wavefunction (*cfi.* Eq. (5.46))

$$\Psi_{k,l,m}(r, \theta, \phi) = N\frac{u_{k,l}(r)}{r}Y_{l,m}(\theta, \phi) \ , \tag{5.64}$$

with the corresponding energies being

$$E = (k + l + 1.5)\hbar\omega \ . \tag{5.65}$$

Further expanding the similarity between the hydrogen atom and the harmonic oscillator, we can express the radial part of the solution for both systems (the *nl* subscript refers to the hydrogen atom, the *kl* to the isotropic harmonic oscillator) using the confluent hypergeometric function $F$ as [262]

$$
\begin{aligned}
u_{nl}(r) &= r^{l+1}F(-n + l + 1, 2l + 2, 2r/n)\exp{(-r/n)} \ , \\
u_{kl}(r) &= r^{l+1}F(-n, l + 3/2, r^2)\exp{(-r^2/2)} \ .
\end{aligned}
\tag{5.66}
$$

The allowed eigenstates are also different for the two systems, with the $n$ quantum number being a positive integer for the hydrogen atom and $k$ being a non-negative integer for the oscillator; and $l$ can take any positive integer value less than $n$ in the case of the hydrogen atom but $k$ and $l$ must also have the same parity for the isotropic oscillator; see Fig. 5.4 to compare the first few allowed states.

Combining the discussion of the allowed states above with the observation that, based on Eq. (5.47), only the radial part of the polarization field will be different between the hydrogen atom and the corresponding states of the isotropic 3D oscillator, it can be concluded that the two systems can be compared based on transitions between corresponding states. Such corresponding states are, for example, hydrogenic $1s \rightarrow 2p$ with the oscillator $00 \rightarrow 11$, or $1s \rightarrow 3s$ with $00 \rightarrow 20$.
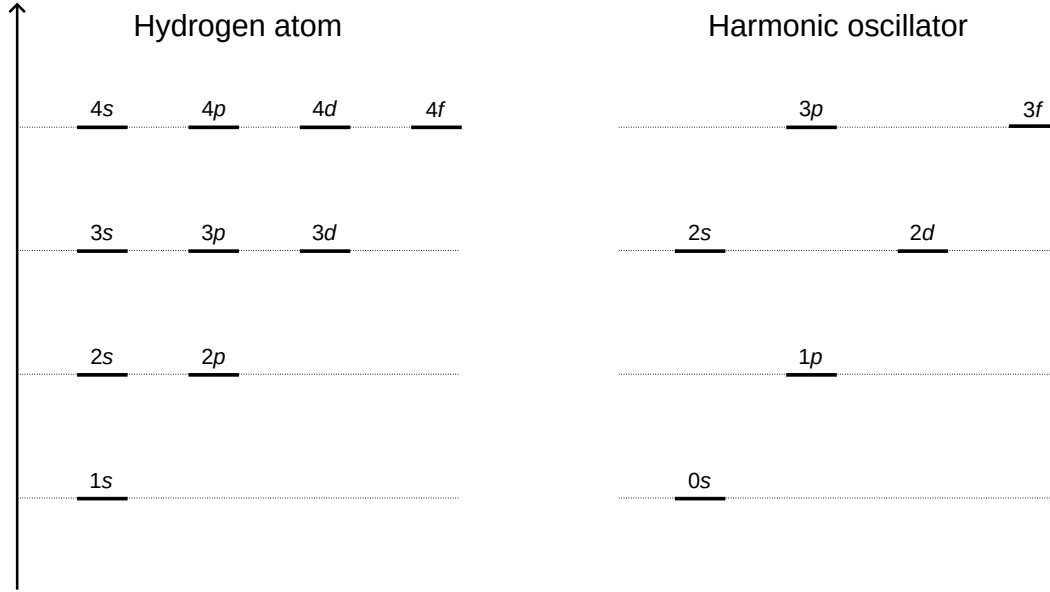
Figure 5.4:  Comparison of the allowed states of the hydrogen atom and the three-dimensional isotropic oscillator. Note that while the energy spectrum of the oscillator is equidistant in $n$, this is not the case for hydrogen atom, so the Figure is not representative of the energy difference between the states [263]

The transition polarization density curves of the normalized radial wavefunctions of some corresponding states are shown in Fig. 5.5, with the normalized radial wavefunctions defined as

$$U(r) = \frac{u(r)/r}{4\pi^2 \int_0^\infty (u(r)/r)^2 r^2 \, \mathrm{d}r} \quad . \tag{5.67}$$

Remarkably, the polarization fields in the ground state of the two systems are qualitatively similar, but the transition fields from the ground state to different excited states are different, even when the corresponding transitions are compared. This difference eventually leads to a qualitatively different non-local polarizability for the two systems.

## 5.2.5   Different Expressions for Polarization Operators

We have seen in Sect. 5.2.1 one definition of the polarization operator (*cfi.* [258]) to be
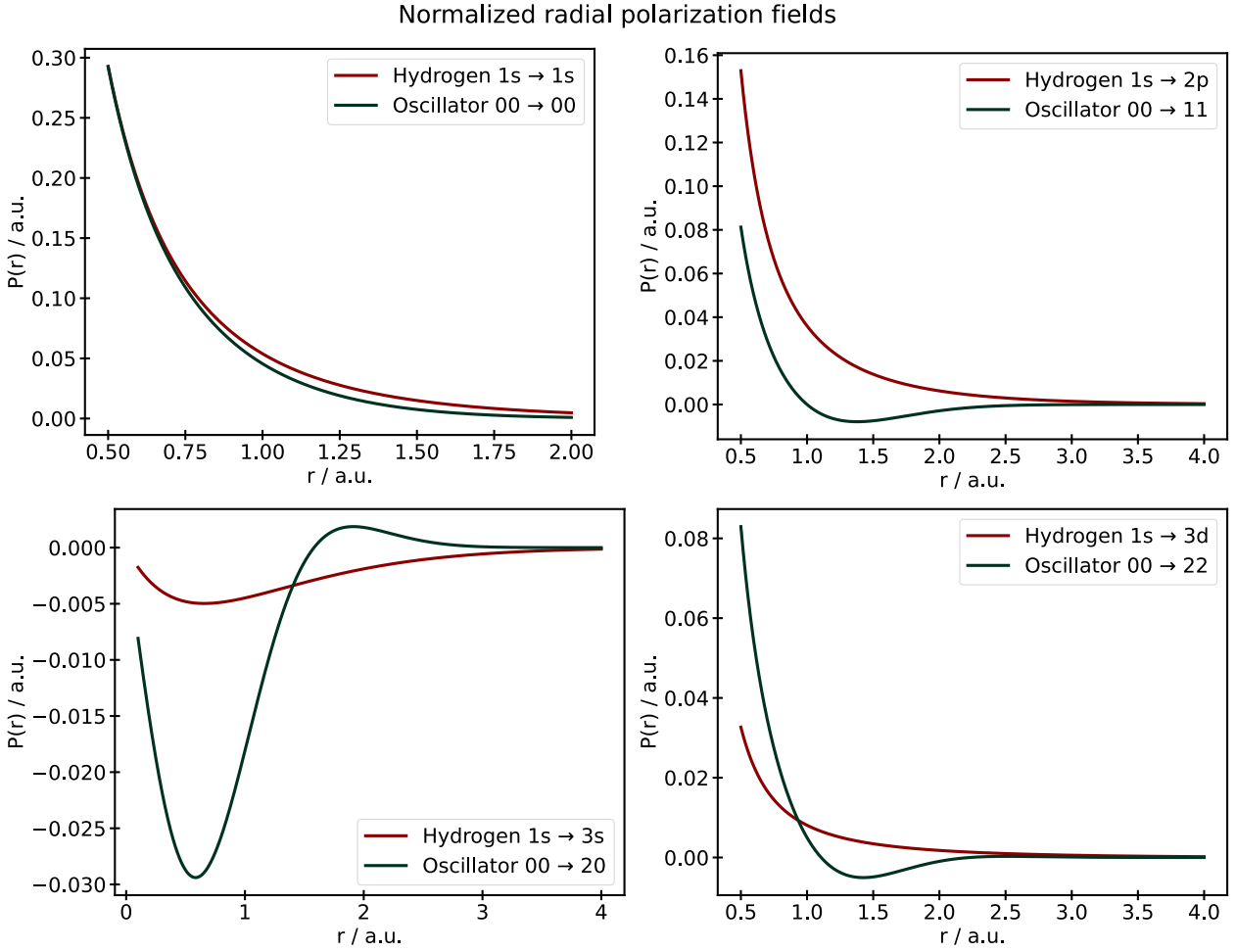
Figure 5.5: Comparison of the normalized radial transition polarization density fields between corresponding states of the hydrogen atom and the 3D isotropic quantum harmonic oscillator.

$$\mathbf{P}_{\mathbf{r_k}}(\mathbf{r}) = -\mathbf{r_k} \int_0^1 \delta(\mathbf{r} - \lambda\mathbf{r_k})\, d\lambda \,, \tag{5.68}$$

where the operator acts on the particle space $\mathbf{r_k}$ and contains the real space $\mathbf{r}$ as a parameter. Note that by repeating the formula here, I have omitted the dependence of the center of charge $\mathbf{R}$, and I am also taking the one-particle part of the operator into account.

In the recent paper of Woolley [264], a different formula for the polarization operator is shown:

$$\mathbf{P}(\mathbf{r}) = \int_C \delta(\mathbf{r} - \mathbf{z}) \, d\mathbf{z} \ . \tag{5.69}$$

This operator is expressed as a line integral along an arbitrary path $C$ that ends at the position of the particle $\mathbf{r_k}$.

In Sect. 5.2.2, deriving the polarization density for the Drude oscillator, I have also used a different definition. Namely, in that section, I have avoided expressing polarization as an operator, instead I have defined the polarization density to be the solution of Eq. (5.37).

It is not clear at first glance what is the difference between the three distinct ideas, what are their advantages and drawbacks, or do they even correspond to the same physical idea.

One caveat of the direct integration of Gauss's law is explained in [264] at Eqs. (27)–(30). In particular, we can consider the polarization operator to be expressed with the Green function of the divergence operator as

$$\mathbf{P}(\mathbf{r}) = \int_V \mathbf{g}(\mathbf{r}, \mathbf{r_0}) \rho(\mathbf{r_0}) \, d^3 \mathbf{r_0} \ , \tag{5.70}$$

with the kernel found from solving the usual defining equation for Green's functions

$$\nabla_{\mathbf{r}} \cdot \mathbf{g}(\mathbf{r}, \mathbf{r_0}) = -\delta(\mathbf{r} - \mathbf{r_0}) \ . \tag{5.71}$$

This, however, only defines the longitudinal part of the Green's function

$$\mathbf{g}^{\parallel}(\mathbf{r}, \mathbf{r_0}) = \nabla_r \frac{1}{4\pi \, |\mathbf{r} - \mathbf{r_0}|} \ . \tag{5.72}$$

One of the main conclusions in [264] is that the polarization field $\mathbf{P}(\mathbf{r})$ is not a longitudinal field, and the fact that the divergence is non-zero carries a physical interpretation. In this sense, the transversal part can not be treated as a nonphysical gauge freedom, so considering both parts of Green's function in Eq. (5.71) is essential. Therefore, not *all* solutions to Gauss's law can be treated as a polarization density in a physical sense.

The discussion of the polarization density and the non-local polarizability raises the question of

the gauge freedom. In the case of the polarization density, our defining equation is Gauss's law, Eq. (5.10). We can easily see from that equation that an arbitrary transversal field $\nabla \times \mathbf{A}$ can be added to $\mathbf{P}$ without changing the validity of the equation

$$\nabla \cdot \mathbf{P_0}(\mathbf{r}) = \nabla \cdot \left[ \mathbf{P_0}(\mathbf{r}) + \nabla \times \mathbf{A}(\mathbf{r}) \right] = -\rho(\mathbf{r}) \ . \tag{5.73}$$

This appears in contrast to the analysis in [258], where the longitudinal and transversal polarization components of the hydrogen atom are analyzed. It might also be interesting to note that the total integral of the $\mathbf{P}$-field, which is expected to be the dipole matrix element, also depends on the addition of an extra transversal term. Note that the analysis of the gauge freedom is more general, with essentially the same questions arising when studying the energy of a molecule in an external field [257].

The gauge freedom in the non-local polarizability $\boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}')$ will correspond to the direct product of the gauge freedoms of the polarization densities in $\mathbf{r}$ and $\mathbf{r}'$ separately. There are two arguments that can be made for this fact

1. Based on Eq. (5.20), $\boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}')$ is a direct product of two independent polarization densities; and

2. the integration by parts in $\mathbf{r}'$ can equivalently be done before or after applying the divergence in $\mathbf{r}$ in the derivation shown in Sect. 5.1.

Therefore, a detailed discussion of the gauge freedom in $\mathbf{P}$ carries over into the discussion in $\boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}')$.

One possible solution to the problem of gauge freedom and the unambiguous definition of $\mathbf{P}(\mathbf{r})$ is to follow the approach of [265] and refrain from detailing the physical significance of the polarization operator at this point.

## 5.3   Non-local Polarizability Density Models

### 5.3.1   Non-local Polarizability of the Drude Oscillator

We now have all the knowledge to work on the non-local polarizability of the 1D Drude oscillator: Eq. (5.42) makes it possible to calculate transition polarization matrix elements, while Eq. (5.20) gives the exact expression of $\boldsymbol{\alpha}(\mathbf{r}, \mathbf{r}')$ in terms of these matrix elements.

To proceed, we note that the full wavefunction is given as (using $a = \sqrt{m\omega/\hbar}$)

$$\Psi_n(x) = \left(\frac{a^2}{\pi}\right)^{1/4} (2^n n!)^{-1/2} H_n(ax) \exp\left(-\frac{a^2 x^2}{2}\right) , \tag{5.74}$$

while the energy reads

$$E_n = \hbar\omega(n + 1/2) . \tag{5.75}$$

We then write the non-local polarizability as

$$\alpha(x, y) = \sum_{n=1}^{\infty} \frac{P_{0,n}(x)P_{n,0}(y)}{n\hbar\omega} = \sum_{n=1}^{\infty} \frac{\rho_{0,n-1}(x)\rho_{n-1,0}(y)}{2n^2 a^2 \hbar\omega} = \sum_{n=0}^{\infty} \frac{\rho_{0,n}(x)\rho_{n,0}(y)}{2a^2(n+1)^2 \hbar\omega} . \tag{5.76}$$

Having transformed the sum to run over the transition density matrix, we can use the exact form of the wavefunction

$$\alpha(x, y) = \sum_{n=0}^{\infty} \frac{\rho_{0,n}(x)\rho_{n,0}(y)}{2a^2(n+1)^2 \hbar\omega} = \frac{\exp\left(-a^2(x^2+y^2)\right)}{2\pi\hbar\omega} \sum_{n=0}^{\infty} \frac{H_n(ax)H_n(ay)}{2^n n!(n+1)^2} . \tag{5.77}$$

The expression we get for the non-local polarizability resembles the identity kernel written in the basis of Hermite polynomials

$$\sum_{n=0}^{\infty} \frac{H_n(ax)H_n(ay)}{2^n n!} = \sqrt{\pi} \exp\left(\frac{1}{2}a^2(x^2 + y^2)\right)\delta(x - y) , \tag{5.78}$$

with a factor of $(n + 1)^{-2}$ missing from the summation. This resemblance hints that following a similar derivation to obtaining this identity kernel could be fruitful. The derivation uses the Fourier transform of the Gaussian function, which is another Gaussian with transformed width

$$\int_{-\infty}^{\infty} \exp\left(isx - \frac{s^2}{\rho^2}\right) ds = \rho\sqrt{\pi} \exp\left(\frac{-\rho^2 s^2}{4}\right) . \tag{5.79}$$

Using this idea on the exponential generating formula (Rodrigues' formula [266]) for the Hermite polynomial, we get

$$H_n(x) = (-1)^n \exp(x^2)\frac{d^n \exp(-x^2)}{dx^n} = (-1)^n \exp(x^2)\frac{d^n}{dx^n}\left(\frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left(isx - \frac{s^2}{4}\right) ds\right) . \tag{5.80}$$

By considering the effect of the $n$-th derivative on the Fourier transform, we get

$$H_n(x) = (-1)^n \exp(x^2)\left(\frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} (is)^n \exp\left(isx - \frac{s^2}{4}\right) ds\right) . \tag{5.81}$$

Using this formula in our expression from Eq. (5.77)

$$\alpha(x/a, y/a) = \sum_{n=0}^{\infty} \frac{H_n(x)H_n(y)}{2^n n!(n+1)^2}$$

$$= \frac{\exp(x^2+y^2)}{4\pi} \sum_{n=0}^{\infty} \iint_{-\infty}^{\infty} \frac{(-st)^n}{2^n n!(n+1)^2} \exp\left(isx + ity - \frac{s^2+t^2}{4}\right) ds\ dt$$

$$= \frac{\exp(x^2+y^2)}{4\pi} \iint_{-\infty}^{\infty} \left(\sum_{n=0}^{\infty} \frac{(-st)^n}{2^n n!(n+1)^2}\right) \exp\left(isx + ity - \frac{s^2+t^2}{4}\right) ds\ dt \ . \quad (5.82)$$

At this point, there are a set of different directions to take. In principle, the summation over $n$ can be evaluated, having the result

$$\sum_{n=0}^{\infty} \frac{(-st)^n}{2^n n!(n+1)^2} = \frac{2}{st}\left(\gamma + \Gamma\left(0, \frac{st}{2}\right) + \ln\left(\frac{st}{2}\right)\right) , \quad (5.83)$$

with $\gamma$ being the Euler gamma constant and $\Gamma(a, x)$ is the so-called incomplete gamma function. Further analytical work with this expression is so far inconclusive.

A different approach uses the simple 1D integral formula

$$\int_0^1 -u^n \log u\ du = \frac{1}{(n+1)^2} , \quad (5.84)$$

to be used in evaluating the sum as

$$\sum_{n=0}^{\infty} \frac{H_n(ax)H_n(ay)}{2^n n!(n+1)^2} = \int_0^1 \sum_{n=0}^{\infty} \frac{-u^n \log u H_n(ax)H_n(ay)}{2^n n!}\ du \ . \quad (5.85)$$

At this point, we can use the identity (obtained from Mehler's formula [267], valid for $u \in [0, 1[$)

$$\sum_{n=0}^{\infty} \frac{H_n(x)H_n(y)}{n!}\left(\frac{u}{2}\right)^n = \frac{1}{\sqrt{1-u^2}} e^{\frac{2u}{1+u}xy - \frac{u^2}{1-u^2}(x-y)^2} \ . \quad (5.86)$$

Equation (5.86), together with Eq. (5.85) leads us to the following representation

$$\sum_{n=0}^{\infty} \frac{H_n(x)H_n(y)}{2^n n!(n+1)^2} = \int_0^1 \frac{-\log u}{\sqrt{1-u^2}} \exp\left(\frac{2u}{1+u}xy - \frac{u^2}{1-u^2}(x-y)^2\right) du \ . \tag{5.87}$$
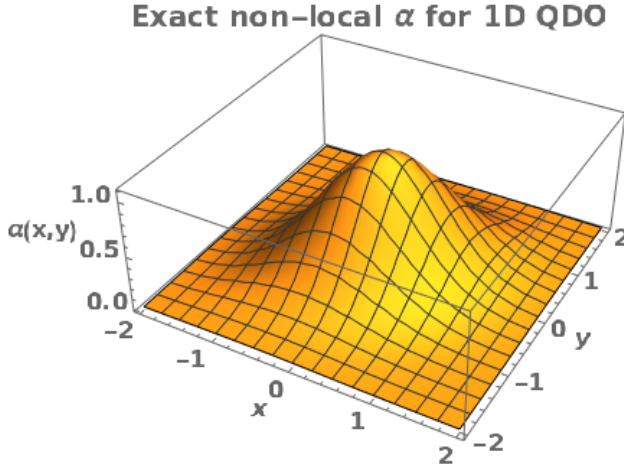
**Exact non−local α for 1D QDO**



Figure 5.6: Exact non-local polarizability for the 1D QDO model.

Equations (5.77)–(5.87), summarized, give

$$\alpha(x,y) = \frac{a^2}{2\pi\hbar\omega} \exp\left(-a^2(x^2+y^2)\right) \int_0^1 \frac{-\log u}{\sqrt{1-u^2}} \exp\left(\frac{2u}{1+u}xy - \frac{u^2}{1-u^2}(x-y)^2\right) du \ , \tag{5.88}$$

which can be evaluated numerically for every value of $x$ and $y$, so, in principle, the non-local polarizability can be plotted exactly, as is done in Fig. 5.6.

## 5.3.2 Non-local Polarizability of the Hydrogen Atom

While it would be desirable to obtain the non-local polarizability of the hydrogen atom, a tractable expression for this property is yet to be found. However, knowing some transition polarization field matrix elements (*cfi*. Sect. 5.2.3) allows us to discuss the approximate properties of the non-local polarizability by taking only the first $N$ terms into account

$$\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}') = \sum_{n\neq 0} \frac{\langle 0|\hat{\mathbf{P}}(\mathbf{r})|n\rangle \langle n|\hat{\mathbf{P}}(\mathbf{r}')|0\rangle}{E_n - E_0} \approx \sum_{n\neq 0}^{N} \frac{\langle 0|\hat{\mathbf{P}}(\mathbf{r})|n\rangle \langle n|\hat{\mathbf{P}}(\mathbf{r}')|0\rangle}{E_n - E_0} . \tag{5.89}$$

The resulting formulas are quite complicated, mainly due to the fact that one needs to account for the angular contributions in both $\mathbf{r}$ and $\mathbf{r}'$. Assuming that both the perturbation and the response are measured along the $x$ axis, the first ($N = 1$) term in the non-local polarizability is given as

$$\alpha(x,x') = \frac{e^{\frac{-3}{2}(|x|+|x'|)}\left(xx'\left(9|x|^3 + 24|x| + 18x^2 + 16\right)\left(9|x'|^3 + 24|x'| + 18x'^2 + 16\right) + 81x^4x'^4\right)}{2187\pi^2|xx'|^3} . \tag{5.90}$$
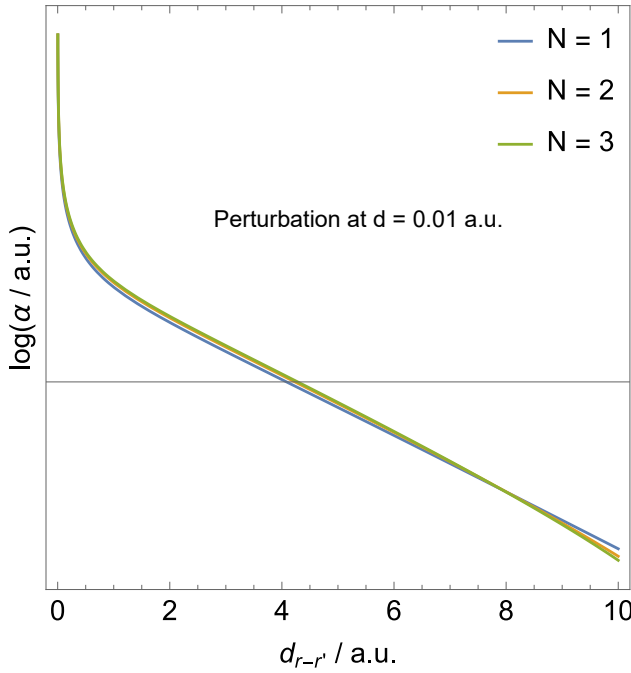


Figure 5.7: Decay of the non-local polarizability of a hydrogen atom; both the response and the polarization is taken along the same Cartesian axis. The number $N$ corresponds to the level of truncation in the perturbative expression, as written in Eq. (5.89).

Equation (5.90) shows that polarizability behaves in a similar way as the wavefunction, the curve being dominated by an exponential decay at medium to large distances. This localized behavior persists even when more contributions are taken into account, as can be concluded from Fig. 5.7.

## 5.4  Conclusions

The need for a non-local polarizability was already highlighted in Chapter 2, where different ways of expressing the dispersion energy through this quantity were discussed. This final Chapter of the thesis discussed a framework in which the non-local nature of the polarizability can be analyzed. The introduction of this object was done *via* treating it in parallel with the well-known non-local susceptibility, leading to the introduction of the polarization operator.

Since this operator was seldom discussed in the context of atomic and molecular systems, the properties of the polarization field operator were first revisited. Having analyzed this object on the case of the harmonic oscillator and the hydrogen atom, we had all the necessary tools to evaluate the non-local polarizability of these systems: the harmonic oscillator could be treated in an exact manner, whereas the only the first few contributing transition polarization operators could be analyzed for the hydrogen atom. Nevertheless, both of these systems seem to share a common property, with the decay of the non-local response function closely reflecting the decay of the electron density itself.

The formalism presented in this Chapter enables the evaluation of the non-local polarizability of model systems, atoms, and, eventually, molecules using an exact form of the polarization operator. Provided that this function can be obtained for a wide variety of realistic systems, its essential physical behavior could be studied, enabling us to capture the fundamental properties necessary to describe non-local quantum mechanical response and quantities derived from it with unprecedented accuracy.

# 6. Summary

Computational approaches that are desired to be applicable for a wide range of molecular systems and materials require transferable models based on universal physical principles. To this end, the polarizability, as a key electronic response function, which has been studied since long before the emergence of quantum mechanics, becomes a central quantity as being naturally used to describe both electron correlation and non-covalent interactions. Although it is well known that polarizability is fundamentally non-local, most existing computational approaches treat this quantity within local or semi-local approximations due to both conceptual and practical difficulties. By means of combining analytical results obtained for quantum mechanical models with those from computational treatment of realistic systems, the work presented in this thesis facilitates the development of a generalized (non-expanded) many-body van der Waals energy functional.

By studying the polarizability of exactly solvable model systems as well as free atoms, it was identified that the main physical factor in determining the magnitude of polarizability is the (effective) geometric extent of quantum-mechanical systems, which can be expressed through expectation values of one- or two-electron position operators. The novel four-dimensional scaling of the dipole polarizability with an effective system size is shown to be applicable for atoms in molecules by increasing the accuracy in evaluating molecular polarizabilities, achieved through an improvement of parametrizing non-interacting atomic polarizabilities from partitioned (atomic) electron densities.

The method of representing atoms by quantum Drude oscillators (QDOs) was also extensively discussed in this thesis. In accordance with the revealed quantum mechanical size-polarizability scaling law, it was found that molecular polarizabilities can be effectively captured by means of the model of dipole-coupled (atomic) oscillators as long as the proper scaling law is used to transform free-atomic response quantities into atoms-in-molecules ones. The optimized parametrization scheme to describe the electronic response by employing QDOs is also presented, demonstrating that dipolar response quantities (as the input data for the used coarse-grained approach) are sufficient to reproduce atomic polarization potentials in good agreement with *ab initio* calculations.

Based on the analysis of the interplay between the geometry of a system and its electronic response, it is reasonable to represent the polarizability as a product of two factors, one accounting for the spatial

extent of the system and the other encoding its electronic spectrum. Consequently, the relation between the polarizability and excitation properties of molecules was examined in a data-driven manner, using a large subset of the chemical space of organic compounds. Due to the diversity of such systems, it became possible to conclusively disprove the common intuitive assumption of a general correlation between the polarizability and HOMO–LUMO gap of molecules. This finding is a consequence of the fact that the polarizability, as an extensive quantity, depends on the number of atoms in a molecule, while the HOMO–LUMO gap, as an intensive quantity, is mostly defined by the arrangement of the atoms into chemical motifs within a molecule.

Understanding the relation between two physical quantities in the chemical compound space delivers guidelines for molecular engineering. In particular, this was demonstrated using the mutual independence of the HOMO–LUMO gap and the polarizability in the design of organic photodetectors, where it is essential to have control over both optical and electrochemical properties at the same time. Since the two kinds of properties are decoupled by the independence of the related quantum mechanical quantities, it is possible to devise organic photodetectors capable of detecting a given wavelength of light while still having freely tailored electrochemical properties.

Finally, the fundamental aspects of the most general approach to describe intermolecular interactions were elaborated on practical examples using the concept of non-local response functions. Special emphasis was placed on the polarization field operator, considering the absence of a well-developed description of this quantity for molecular systems. The knowledge gained on the polarization field operator paves the way for the development of a robust approach for the non-local polarizability. Since the formalism is quite elaborate, it has been applied so far only to a few model systems. In particular, the hydrogen atom and quantum Drude oscillator were considered in detail with a thorough discussion of their common and distinct properties.

The results presented in this thesis serve to systematically improve our general understanding of atomic and molecular response properties at the microscopic level. In particular, the insights gained into electronic polarizability, as the central quantity considered within the presented work, are important for the design of novel energy functionals capable of properly capturing intermolecular (including the ubiquitous dispersion) interactions. To this end, contributions were made in studying the connection between the system geometry related to ground-state properties and the response of electrons caused by their excitations, the switch from the expanded to non-expanded description of the dispersion forces through the use of the non-local formalism, and the intrinsically present many-body effects within the interaction energy being built based on the non-local polarizability.

Altogether, the presented results provide important knowledge required to develop new efficient and transferable methods applicable to complex (bio)molecular systems and materials.

## 6.1   Outlook and Perspectives

The novel results published in this thesis, with respect to the microscopic level of description of electronic response in molecules as well as the framework for the non-local polarizability, open up promising directions for future research. In what follows, some of them are briefly mentioned.

In this thesis, it was shown that the response properties of molecules (in particular, dipole polarizability) are well captured through the electronic response of adequately coupled atomic contributions. The predictive power of this approach derives from both the accurate determination of the individual atomic contributions and the proper description of the coupling between them. Within the presented work, the correct scaling law for the dipole polarizability of atoms in a molecule was established, which can now be used in similar approaches ensuring their improved accuracy. However, the proposed scheme relies on Hirshfeld partitioning of the molecular electron density, which (in its basic form) cannot properly account for polarization and charge transfer effects. Moreover, although a correlated size descriptor was already introduced in this thesis, the possibility of partitioning a two-point response function (as opposed to the electron density, which is a local quantity) has not been explored. Given the importance of the non-local nature of response properties, accounting for such effects seems to be crucial for devising a general and transferable approach. In this regard, a possible direction for further studies is to coarse-grain not the electron density but a non-local response function itself. By construction, the non-local response should account for polarization and charge transfer effects, and, provided an appropriate two-point coarse-graining scheme, electron correlation effects can also be incorporated into the individual atomic response within such an advanced approach.

In addition, the proposed optimized quantum Drude oscillator (OQDO) model dispenses with the need of employing higher order response properties, which are not so accurately known in comparison to the dipolar response quantities used in the new parametrization scheme presented in this thesis. The (atomic) polarization potential generated by QDOs was also examined, demonstrating that the OQDO model matches *ab initio* curves up to a high degree. Nevertheless, the electrostatic screening used to account for long-range interactions between atomic centers was still based on a

simple model stemming from the classical electrodynamical description of the interaction between Gaussian charge densities. It is yet to be explored whether the good prediction of the polarization potential can be eventually employed to replace the classical model with its quantum mechanical counterpart.

Another advantage of the model of coupled QDOs is that it is rooted in the analytically solvable harmonic Hamiltonian, so a variety of properties can also be derived relatively simply from it. Notably, since the (isolated) OQDO model performs accurately for higher order polarizabilities, it is expected – but not yet tested – that multipolar response could also be captured by coupled dipole oscillators. In principle, many other properties, including magnetic susceptibility or time-dependent behavior, could be solved for the coupled oscillator model. Showing that a relatively simple analytical model represents realistic response provides one with a powerful tool to study the properties of large and complex systems with techniques that are usually restricted to atoms and simpler molecules.

Finally, in this thesis, the evaluation of the non-local polarizability tensor through the exact form of the polarization field operator was elaborated. The spatial part of the polarization field was illustrated on the relatively simple models of the harmonic oscillator and the hydrogen atom, paving the way for further studies of molecules and material systems within the framework introduced here. Altogether, this delivers a tool to describe dispersion interactions and electron correlations in realistic systems based on truly non-local quantum physics.

# Bibliography

[1] Stöhr, M.; Van Voorhis, T.; Tkatchenko, A. Theory and Practice of Modeling van der Waals Interactions in Electronic-Structure Calculations. *Chemical Society Reviews* **2019**, *48*, 4118–4154.

[2] Hermann, J.; DiStasio, R. A. J.; Tkatchenko, A. First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chemical Reviews* **2017**, *117*, 4714–4758.

[3] Atkins, P.; Paula, J. *Atkins' Physical Chemistry*; Oxford University press, 2008.

[4] Stone, A. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, UK, 2013.

[5] Housecroft, C. E.; Constable, E. C. *Chemistry: An Introduction to Organic, Inorganic and Physical Chemistry*; Pearson Education: New York, NY, 2010.

[6] Herbert, J. M. Neat, Simple, and Wrong: Debunking Electrostatic Fallacies Regarding Noncovalent Interactions. *The Journal of Physical Chemistry A* **2021**, *125*, 7125–7137.

[7] Dobson, J. F. Beyond Pairwise Additivity in London Dispersion Interactions. *International Journal of Quantum Chemistry* **2014**, *114*, 1157–1161.

[8] Poier, P. P.; Adjoua, O.; Lagardère, L.; Piquemal, J.-P. Generalized Many-Body Dispersion Correction through Random-Phase Approximation for Chemically Accurate Density Functional Theory. *The Journal of Physical Chemistry Letters* **2023**, *14*, 1609–1617.

[9] Kubo, R. The Fluctuation-Dissipation Theorem. *Reports on Progress in Physics* **1966**, *29*, 255.

[10] Dobson, J. F.; Dinte, B. P. Constraint Satisfaction in Local and Gradient Susceptibility Approximations: Application to a van der Waals Density Functional. *Physical Review Letters* **1996**, *76*, 1780–1783.

[11] Jackson, J. D. *Classical Electrodynamics; 2nd Ed.*; Wiley: New York, NY, 1975.

[12] McLachlan, A. D.; Longuet-Higgins, H. C. Retarded Dispersion Forces in Dielectrics at Finite Temperatures. *Proceedings of the Royal Society of London. Series A.* **1963**, *274*, 80–90.

[13] Magnasco, V.; Ottonelli, M.; Figari, G.; Rui, M.; Costa, C. Long-Range Dispersion Coefficients for Like Centrosymmetric Linear Molecules and an Application to H2–H2. *Journal of Molecular Structure: THEOCHEM* **1998**, *430*, 231–239.

[14] Sato, T.; Nakai, H. Density Functional Method Including Weak Interactions: Dispersion Coefficients Based On the Local Response Approximation. *The Journal of Chemical Physics* **2009**, *131*, 224104.

[15] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chemical Reviews* **2016**, *116*, 5105–5154.

[16] Szalewicz, K. Symmetry-Adapted Perturbation Theory of Intermolecular Forces. *WIREs Computational Molecular Science* **2012**, *2*, 254–272.

[17] Kragh, H. The Lorenz-Lorentz Formula: Origin and Early History. *Substantia* **2018**, *2*, 7–18.

[18] Lorentz, H. A. *The Theory of Electrons and Its Applications to the Phenomena of Light and Radiant Heat*; GE Stechert & Company, 1916; Vol. 29.

[19] Szabó, P.; Góger, S.; Charry, J.; Karimpour, M. R.; Fedorov, D. V.; Tkatchenko, A. Four-Dimensional Scaling of Dipole Polarizability in Quantum Systems. *Physical Review Letters* **2022**, *128*, 070602.

[20] Pearson, R. G. Hard and Soft Acids and Bases, HSAB, Part 1: Fundamental Principles. *Journal of Chemical Education* **1968**, *45*, 581.

[21] Pearson, R. G. Hard and Soft Acids and Bases, HSAB, Part Ii: Underlying Theories. *Journal of Chemical Education* **1968**, *45*, 643.

[22] Meyers, F.; Marder, S. R.; Pierce, B. M.; Bredas, J. L. Electric Field Modulated Nonlinear Optical Properties of Donor-Acceptor Polyenes: Sum-Over-States Investigation of the Relationship between Molecular Polarizabilities ($\alpha$, $\beta$, and $\gamma$) and Bond Length Alternation. *Journal of the American Chemical Society* **1994**, *116*, 10703–10714.

[23] Brinck, T.; Murray, J. S.; Politzer, P. Polarizability and Volume. *Journal of Chemical Physics* **1993**, *98*, 4305–4306.

[24] Politzer, P.; Jin, P.; Murray, J. S. Atomic Polarizability, Volume and Ionization Energy. *Journal of Chemical Physics* **2002**, *117*, 8197–8202.

[25] Blair, S. A.; Thakkar, A. J. Relating Polarizability to Volume, Ionization Energy, Electronegativity, Hardness, Moments of Momentum, and Other Molecular Properties. *The Journal of Chemical Physics* **2014**, *141*, 074306.

[26] Choudhary, S.; Ranjan, P.; Chakraborty, T. Atomic Polarizability: A Periodic Descriptor. *Journal of Chemical Research* **2020**, *44*, 227–234.

[27] Yang, Y.; Lao, K. U.; DiStasio, R. A. Influence of Pore Size on the van der Waals Interaction in Two-Dimensional Molecules and Materials. *Physical Review Letters* **2019**, *122*, 026001.

[28] Fedorov, D. V.; Sadhukhan, M.; Stöhr, M.; Tkatchenko, A. Quantum-Mechanical Relation between Atomic Dipole Polarizability and the van der Waals Radius. *Physical Review Letters* **2018**, *121*, 183401.

[29] Góger, S.; Sandonas, L. M.; Müller, C.; Tkatchenko, A. Data-Driven Tailoring of Molecular Dipole Polarizability and Frontier Orbital Energies in Chemical Compound Space. *Physical Chemistry Chemical Physics* **2023**, *25*, 22211–22222.

[30] Liu, P.; Hunt, K. L. C. Molecular softness, hypersoftness, infrared absorption, and vibrational Raman scattering: New relations derived from nonlocal polarizability densities. *The Journal of Chemical Physics* **1995**, *103*, 10597–10604.

[31] Bosque, R.; Sales, J. Polarizabilities of Solvents from the Chemical Composition. *Journal of Chemical Information and Computer Sciences* **2002**, *42*, 1154–1163.

[32] Stout, J. M.; Dykstra, C. E. Static Dipole Polarizabilities of Organic Molecules. Ab Initio Calculations and a Predictive Model. *Journal of the American Chemical Society* **1995**, *117*, 5127–5132.

[33] Katritzky, A. R.; Kuanar, M.; Slavov, S.; Hall, C. D.; Karelson, M.; Kahn, I.; Dobchev, D. A. Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *Chemical Reviews* **2010**, *110*, 5714–5789.

[34] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

[35] Tkatchenko, A.; Scheffler, M. Accurate Molecular van der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Physical Review Letters* **2009**, *102*, 073005.

[36] Gould, T. How Polarizabilities and $C_6$ Coefficients Actually Vary With Atomic Volume. *The Journal of Chemical Physics* **2016**, *145*.

[37] Silberstein, L. VII. Molecular Refractivity and Atomic Interaction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1917**, *33*, 92–128.

[38] Applequist, J.; Carl, J. R.; Fung, K.-K. Atom Dipole Interaction Model for Molecular Polarizability. Application to Polyatomic Molecules and Determination of Atom Polarizabilities. *Journal of the American Chemical Society* **1972**, *94*, 2952–2960.

[39] Thole, B. Molecular Polarizabilities Calculated With a Modified Dipole Interaction. *Chemical Physics* **1981**, *59*, 341–350.

[40] DelloStritto, M.; Klein, M. L.; Borguet, E. Bond-Dependent Thole Model for Polarizability and Spectroscopy. *Journal of Physical Chemistry A* **2019**, *123*, 5378–5387.

[41] Mayer, A.; Åstrand, P.-O. A Charge-Dipole Model for the Static Polarizability of Nanostructures Including Aliphatic, Olephinic, and Aromatic Systems. *The Journal of Physical Chemistry A* **2008**, *112*, 1277–1285.

[42] Tkatchenko, A.; DiStasio Jr., R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Physical Review Letters* **2012**, *108*, 236402.

[43] Massa, D.; Ambrosetti, A.; Silvestrelli, P. L. Beyond-Dipole van der Waals Contributions Within the Many-Body Dispersion Framework. *Electronic Structure* **2021**, *3*, 044002.

[44] Góger, S.; Khabibrakhmanov, A.; Vaccarelli, O.; Fedorov, D. V.; Tkatchenko, A. Optimized Quantum Drude Oscillators for Atomic and Molecular Response Properties. *The Journal of Physical Chemistry Letters* **2023**, *14*, 6217–6223.

[45] van Duijnen, P. T.; Swart, M. Molecular and Atomic Polarizabilities: Thole's Model Revisited. *The Journal of Physical Chemistry A* **1998**, *102*, 2399–2407.

[46] Gobre, V. V. Efficient Modelling of Linear Electronic Polarization in Materials Using Atomic Response Functions. Ph.D. thesis, Fritz Haber Institute Berlin, 2016.

[47] Gobre, V. V.; Tkatchenko, A. Scaling Laws for Van Der Waals Interactions in Nanostructured Materials. *Nature Communications* **2013**, *4*, 2341.

[48] Theimer, O.; Paul, R. Anisotropic Light Scattering by Inner-Field Fluctuations in a Dense Monatomic Gas. *The Journal of Chemical Physics* **2004**, *42*, 2508–2517.

[49] Oxtoby, D. W. The Calculation of Pair Polarizabilities Through Continuum Electrostatic Theory. *The Journal of Chemical Physics* **2008**, *69*, 1184–1189.

[50] Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Physical Review Letters* **2004**, *92*, 246401.

[51] Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-Accuracy van der Waals Density Functional. *Physical Review B* **2010**, *82*, 081101.

[52] Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals Density Functional Made Simple. *Physical Review Letters* **2009**, *103*, 063004.

[53] Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals Density Functional: The Simpler the Better. *The Journal of Chemical Physics* **2010**, *133*, 244103.

[54] Vydrov, O. A.; Van Voorhis, T. Dispersion Interactions from a Local Polarizability Model. *Physical Reviews A* **2010**, *81*, 062708.

[55] Hermann, J.; Tkatchenko, A. Density Functional Model for van der Waals Interactions: Unifying Many-Body Atomic Approaches with Nonlocal Functionals. *Physical Review Letters* **2020**, *124*, 146401.

[56] Dalgarno, A. Atomic Polarizabilities and Shielding Factors. *Advances in Physics* **1962**, *11*, 281–315.

[57] Orttung, W. H.; Vosooghi, D. Polarizability Densities Within Atoms. 1. Simple One-Electron Systems. *The Journal of Physical Chemistry* **1983**, *87*, 1432–1437.

[58] Orttung, W. H.; St. Julien, D. Polarizability Densities Within Atoms. 2. Helium, Neon, and Argon. *The Journal of Physical Chemistry* **1983**, *87*, 1438–1444.

[59] Ricci, M.; Silvestrelli, P. L.; Dobson, J. F.; Ambrosetti, A. Exact Sum-Rule Approach to Polarizability and Asymptotic van der Waals Functionals-Derivation of Exact Single-Particle Benchmarks. *The Journal of Physical Chemistry Letters* **2022**, *13*, 8298–8304.

[60] Mavromatis, H. A. The Dalgarno–Lewis Summation Technique: Some Comments and Examples. *American Journal of Physics* **1991**, *59*, 738–744.

[61] Sipe, J.; Kranendonk, J. V. Limitations of the Concept of Polarizability Density As Applied to Atoms and Molecules. *Molecular Physics* **1978**, *35*, 1579–1584.

[62] Hunt, K. L. C. Nonlocal Polarizability Densities and van der Waals Interactions. *The Journal of Chemical Physics* **1983**, *78*, 6149–6155.

[63] Hunt, K. L. C. Nonlocal Polarizability Densities and the Effects of Short-Range Interactions on Molecular Dipoles, Quadrupoles, and Polarizabilities. *The Journal of Chemical Physics* **1984**, *80*, 393–407.

[64] Hunt, K. L. C. Dispersion Dipoles and Dispersion Forces: Proof of Feynman's "Conjecture" and Generalization to Interacting Molecules of Arbitrary Symmetry. *The Journal of Chemical Physics* **1990**, *92*, 1180.

[65] Nimalakirthi, R.; Hunt, K. L. C. Nonlocal Polarizability Density of a Model System: A Homogeneous Electron Gas at T=0. *The Journal of Chemical Physics* **1993**, *98*, 3066–3075.

[66] Cheng, Y.; Verstraelen, T. A New Framework for Frequency-Dependent Polarizable Force Fields. *The Journal of Chemical Physics* **2022**, *157*, 124106.

[67] Misquitta, A. J.; Stone, A. J. ISA-Pol: Distributed Polarizabilities and Dispersion Models from a Basis-Space Implementation of the Iterated Stockholder Atoms Procedure. *Theoretical Chemistry Accounts* **2018**, *137*, 1–20.

[68] Szabó, A.; Ostlund, N. S. *Modern quantum chemistry : introduction to advanced electronic structure theory*; Mineola (N.Y.) : Dover publications, 1996.

[69] Medveď, M.; Stachová, M.; Jacquemin, D.; André, J.-M.; Perpète, E. A. A Generalized Romberg Differentiation Procedure for Calculation of Hyperpolarizabilities. *Journal of Molecular Structure: THEOCHEM* **2007**, *847*, 39–46.

[70] Chen, H.; Liu, M.; Yan, T. Molecular Multipoles and (Hyper)polarizabilities from the Buckingham Expansion: Revisited. *Communications in Theoretical Physics* **2020**, *72*, 075503.

[71] NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 22, May 2022, Editor: Russell D. Johnson III, http://cccbdb.nist.gov/.

[72] Woon, D. E.; Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. Iv. Calculation of Static Electrical Response Properties. *The Journal of Chemical Physics* **1994**, *100*, 2975.

[73] Sadlej, A. J. Medium-Size Polarized Basis Sets for High-Level-Correlated Calculations of Molecular Electric Properties. *Theoretica Chimica Acta* **1992**, *81*, 339–354.

[74] Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based Simulations of Chemistry Framework. *WIREs Computational Molecular Science* **2018**, *8*, e1340.

[75] Balawender, R.; Geerlings, P. Nuclear Fukui Function from Coupled Perturbed Hartree–fock Equations. *The Journal of Chemical Physics* **2001**, *114*, 682–691.

[76] Osamura, Y.; Yamaguchi, Y.; Saxe, P.; Fox, D.; Vincent, M.; Schaefer, H. Analytic Second Derivative Techniques for Self-Consistent-Field Wave Functions. a New Approach to the Solution of the Coupled Perturbed Hartree-Fock Equations. *Journal of Molecular Structure: THEOCHEM* **1983**, *103*, 183–196.

[77] Wilkins, D. M.; Grisafi, A.; Yang, Y.; Lao, K. U.; DiStasio, R. A.; Ceriotti, M. Accurate Molecular Polarizabilities With Coupled Cluster Theory and Machine Learning. *Proceedings of the National Academy of Sciences* **2019**, *116*, 3401–3406.

[78] Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Müller, K.-R.; Von Lilienfeld, O. A. Machine Learning of Molecular Electronic Properties in Chemical Compound Space. *New Journal of Physics* **2013**, *15*, 095003.

[79] Stark, J. Beobachtungen über den Effekt des elektrischen Feldes auf Spektrallinien. I. Quereffekt. *Annalen der Physik* **1914**, *348*, 965–982.

[80] Waller, I. Der Starkeffekt zweiter Ordnung bei Wasserstoff und die Rydbergkorrektion der Spektra von He und Li+. *Zeitschrift für Physik* **1926**, *38*, 635–646.

[81] Tanner, A. C.; Thakkar, A. J. Discrete and Continuum Contributions to Multipole Polarizabilities and Shielding Factors of Hydrogen. *International Journal of Quantum Chemistry* **1983**, *24*, 345–352.

[82] Fiscelli, G.; Rizzuto, L.; Passante, R. Dispersion Interaction between Two Hydrogen Atoms in a Static Electric Field. *Physical Review Letters* **2020**, *124*, 013604.

[83] Unsøld, A. Quantentheorie des Wasserstoffmolekülions und der Born-Landéschen Abstoßungskräfte. *Zeitschrift für Physik* **1927**, *43*, 563–574.

[84] Sylvain, M. G.; Csizmadia, I. Average Dipole Polarizabilities from the Unsold Approximation and Ab Initio Data. *Chemical Physics Letters* **1987**, *136*, 575–582.

[85] Roberts, R. E. Scaled Unsöld Approximation for Atoms and Molecules. *Journal of Chemical Physics* **1967**, *47*, 1873–1873.

[86] Drude, P. *The Theory of Optics*; Longmans, Green: New York, 1902.

[87] Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. J. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chemical Reviews* **2016**, *116*, 4983–5013.

[88] Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. J. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *Journal of Chemical Theory and Computation* **2013**, *9*, 5430–5449.

[89] Jones, A.; Cipcigan, F.; Sokhan, V. P.; Crain, J.; Martyna, G. J. Electronically Coarse-Grained Model for Water. *Physical Review Letters* **2013**, *110*, 227801.

[90] Jones, A.; Thompson, A.; Crain, J.; Müser, M. H.; Martyna, G. J. Norm-Conserving Diffusion Monte Carlo Method and Diagrammatic Expansion of Interacting Drude Oscillators: Application to Solid Xenon. *Physical Reviews B* **2009**, *79*, 144119.

[91] Karimpour, M. R.; Fedorov, D. V.; Tkatchenko, A. Quantum Framework for Describing Retarded and Nonretarded Molecular Interactions in External Electric Fields. *Physical Review Research* **2022**, *4*, 013011.

[92] Fowler, P. Energy, Polarizability and Size of Confined One-Electron Systems. *Molecular Physics* **1984**, *53*, 865–889.

[93] Al-Hashimi, M.; Wiese, U.-J. From a Particle in a Box to the Uncertainty Relation in a Quantum Dot and to Reflecting Walls for Relativistic Fermions. *Annals of Physics* **2012**, *327*, 1–28, January 2012 Special Issue.

[94] Onyia, A.; Ikeri, H.; Nwobodo, A. Theoretical Study of the Quantum Confinement Effects On Quantum Dots Using Particle in a Box Model. *Journal of Ovonic Research* **2018**, *14*, 49–54.

[95] Katan, C.; Mercier, N.; Even, J. Quantum and Dielectric Confinement Effects in Lower-Dimensional Hybrid Perovskite Semiconductors. *Chemical Reviews* **2019**, *119*, 3140–3192.

[96] Atkins, P. W. P. W. *Molecular Quantum Mechanics*, 4th ed.; Oxford University Press: New York, 2005.

[97] Hermann, J.; Alfè, D.; Tkatchenko, A. Nanoscale $\pi$–$\pi$ Stacked Molecules Are Bound by Collective Charge Fluctuations. *Nature Communications* **2017**, *8*, 14052.

[98] Wang, J.; Yang, M.; Jellinek, J.; Wang, G. Dipole Polarizabilities of Medium-Sized Gold Clusters. *Physical Reviews A* **2006**, *74*, 023202.

[99] Dakovski, G. L.; Lan, S.; Xia, C.; Shan, J. Terahertz Electric Polarizability of Excitons in PbSe and CdSe Quantum Dots. *The Journal of Physical Chemistry C* **2007**, *111*, 5904–5908.

[100] Seufert, J.; Obert, M.; Scheibner, M.; Gippius, N. A.; Bacher, G.; Forchel, A.; Passow, T.; Leonardi, K.; Hommel, D. Stark effect and polarizability in a single CdSe/ZnSe quantum dot. *Applied Physics Letters* **2001**, *79*, 1033–1035.

[101] Empedocles, S. A.; Bawendi, M. G. Quantum-Confined Stark Effect in Single CdSe Nanocrystallite Quantum Dots. *Science* **1997**, *278*, 2114–2117.

[102] Kulakci, M.; Serincan, U.; Turan, R.; Finstad, T. G. The Quantum Confined Stark Effect in Silicon Nanocrystals. *Nanotechnology* **2008**, *19*, 455403.

[103] Milonni, P. W. *The Quantum Vacuum: An Introduction to Quantum Electrodynamics*; Academic press: Cambridge, Massachusetts, 2013.

[104] Leuchs, G.; Sánchez-Soto, L. L. A Sum Rule for Charged Elementary Particles. *The European Physical Journal D* **2013**, *67*, 1–5.

[105] Urban, M.; Couchot, F.; Sarazin, X.; Djannati-Atai, A. The Quantum Vacuum as the Origin of the Speed of Light. *The European Physical Journal D* **2013**, *67*, 1–6.

[106] Vinti, J. P. A Relation between the Electric and Diamagnetic Susceptibilities of Monatomic Gases. *Physical Review* **1932**, *41*, 813–817.

[107] Buckingham, A. D. Polarizability and Hyperpolarizability. *Philosophical Transactions of the Royal Society of London. Series A* **1979**, *293*, 239–248.

[108] Montgomery Jr., H.; Pupyshev, V. On Lower Bounds for Polarisability. *The European Physical Journal H* **2013**, *38*, 519–534.

[109] Hirschfelder, J.; Bird, R. B.; Curtiss, C. F. *Molecular Theory of Gases and Liquids*; Structure of matter series; Wiley: New York, NY, 1964; p 852.

[110] Griffiths, D. J. Introduction to Electrodynamics. 2017.

[111] Johnson, E. R.; Becke, A. D. A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *The Journal of Chemical Physics* **2006**, *124*, 174104.

[112] Mayer, A. Formulation in Terms of Normalized Propagators of a Charge-Dipole Model Enabling the Calculation of the Polarization Properties of Fullerenes and Carbon Nanotubes. *Physical Review B* **2007**, *75*, 045407.

[113] DeKock, R. L.; Strikwerda, J. R.; Eric, X. Y. Atomic Size, Ionization Energy, Polarizability, Asymptotic Behavior, and the Slater–Zener Model. *Chemical Physics Letters* **2012**, *547*, 120–126.

[114] Tkatchenko, A.; Fedorov, D. V.; Gori, M. Fine-Structure Constant Connects Electronic Polarizability and Geometric van-der-Waals Radius of Atoms. *The Journal of Physical Chemistry Letters* **2021**, *12*, 9488.

[115] Wang, F.; Jordan, K. D. A Drude-model Approach to Dispersion Interactions in Dipole-Bound Anions. *The Journal of Chemical Physics* **2001**, *114*, 10717.

[116] Sommerfeld, T.; Jordan, K. D. Quantum Drude Oscillator Model for Describing the Interaction of Excess Electrons with Water Clusters: An Application to $(H_2O)_{13}^-$. *The Journal of Physical Chemistry A* **2005**, *109*, 11531.

[117] Jones, A. P.; Crain, J.; Sokhan, V. P.; Whitfield, T. W.; Martyna, G. J. Quantum Drude Oscillator Model of Atoms and Molecules: Many-Body Polarization and Dispersion Interactions for Atomistic Simulation. *Physical Review B* **2013**, *87*, 144103.

[118] Silvestrelli, P. L.; Ambrosetti, A. van der Waals Interactions in DFT using Wannier Functions Without Empirical Parameters. *The Journal of Chemical Physics* **2019**, *150*.

[119] Rudden, L. S.; Degiacomi, M. T. Protein Docking Using a Single Representation for Protein Surface, Electrostatics, and Local Dynamics. *Journal of Chemical Theory and Computation* **2019**, *15*, 5135–5143.

[120] Wang, S. Generalization of the Thomas-Reiche-Kuhn and the Bethe Sum Rules. *Physical Review A* **1999**, *60*, 262–266.

[121] Buckingham, R. A.; Lennard-Jones, J. E. The Quantum Theory of Atomic Polarization I — Polarization by a Uniform Field. *Proceedings of the Royal Society of London. Series A* **1937**, *160*, 94–113.

[122] Blair, S. A.; Thakkar, A. J. TABS: A Database of Molecular Structures. *Computational and Theoretical Chemistry* **2014**, *1043*, 13–16.

[123] Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. Ab Initio Molecular Simulations With Numeric Atom-Centered Orbitals. *Computational Physics Communications* **2009**, *180*, 2175–2196.

[124] Heidar-Zadeh, F.; Ayers, P. W.; Verstraelen, T.; Vinogradov, I.; Vöhringer-Martinez, E.; Bultinck, P. Information-Theoretic Approaches to Atoms-in-Molecules: Hirshfeld Family of Partitioning Schemes. *The Journal of Physical Chemistry A* **2018**, *122*, 4219–4245.

[125] Lekkerkerker, H.; Coulon, P.; Luyckx, R. Dispersion Forces Between Closed Shell Atoms. *Physica A* **1977**, *88*, 375–384.

[126] Cipcigan, F. S.; Crain, J.; Sokhan, V. P.; Martyna, G. J. Electronic Coarse Graining: Predictive Atomistic Modeling of Condensed Matter. *Reviews of Modern Physics* **2019**, *91*, 025003.

[127] Akimov, A. V.; Prezhdo, O. V. Large-Scale Computations in Chemistry: A Bird's Eye View of a Vibrant Field. *Chemical Reviews* **2015**, *115*, 5797.

[128] Sadhukhan, M.; Manby, F. R. Quantum Mechanics of Drude Oscillators With Full Coulomb Interaction. *Physical Review B* **2016**, *94*, 115106.

[129] Schröder, C.; Steinhauser, O. Simulating Polarizable Molecular Ionic Liquids With Drude Oscillators. *The Journal of Chemical Physics* **2010**, *133*, 154511.

[130] Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. Determination of Electrostatic Parameters for a Polarizable Force Field Based on the Classical Drude Oscillator. *Journal of Chemical Theory and Computation* **2005**, *1*, 153–168.

[131] Karimpour, M. R.; Fedorov, D. V.; Tkatchenko, A. Quantum Framework for Describing Retarded and Nonretarded Molecular Interactions in External Electric Fields. *Physical Review Research* **2022**, *4*, 013011.

[132] Ambrosetti, A.; Umari, P.; Silvestrelli, P. L.; Elliott, J.; Tkatchenko, A. Optical van-der-Waals Forces in Molecules: From Electronic Bethe-Salpeter Calculations to the Many-Body Dispersion Model. *Nature Communications* **2022**, *13*, 813.

[133] Whitfield, T. W.; Martyna, G. J. A Unified Formalism for Many-Body Polarization and Dispersion: The Quantum Drude Model Applied to Fluid Xenon. *Chemical Physics Letters* **2006**, *424*, 409.

[134] Voora, V. K.; Ding, J.; Sommerfeld, T.; Jordan, K. D. A Self-Consistent Polarization Potential Model for Describing Excess Electrons Interacting with Water Clusters. *The Journal of Physical Chemistry B* **2013**, *117*, 4365–4370.

[135] Bade, W. L. Drude-Model Calculation of Dispersion Forces. I. General Theory. *The Journal of Chemical Physics* **1957**, *27*, 1280.

[136] Bade, W. L.; Kirkwood, J. G. Drude-Model Calculation of Dispersion Forces. II. The Linear Lattice. *The Journal of Chemical Physics* **1957**, *27*, 1284–1288.

[137] Reilly, A. M.; Tkatchenko, A. van der Waals Dispersion Interactions in Molecular Materials: Beyond Pairwise Additivity. *Chemical Science* **2015**, *6*, 3289.

[138] Sadhukhan, M.; Tkatchenko, A. Long-Range Repulsion Between Spatially Confined van der Waals Dimers. *Physical Review Letters* **2017**, *118*, 210402.

[139] Ambrosetti, A.; Alfè, D.; DiStasio Jr., R. A.; Tkatchenko, A. Hard Numbers for Large Molecules: Toward Exact Energetics for Supramolecular Systems. *The Journal of Physical Chemistry Letters* **2014**, *5*, 849.

[140] DiStasio Jr., R. A.; Gobre, V. V.; Tkatchenko, A. Many-body van der Waals Interactions in Molecules and Condensed Matter. *Journal of Physics: Condensed Matter* **2014**, *26*, 213202.

[141] Stöhr, M.; Sadhukhan, M.; Al-Hamdani, Y. S.; Hermann, J.; Tkatchenko, A. Coulomb Interactions Between Dipolar Quantum Fluctuations in van der Waals Bound Molecules and Materials. *Nature Communications* **2021**, *12*, 137.

[142] Vaccarelli, O.; Fedorov, D. V.; Stöhr, M.; Tkatchenko, A. Quantum-Mechanical Force Balance Between Multipolar Dispersion and Pauli Repulsion in Atomic van der Waals Dimers. *Physical Review Research* **2021**, *3*, 033181.

[143] Kleshchonok, A.; Tkatchenko, A. Tailoring van der Waals Dispersion Interactions With External Electric Charges. *Nature Communications* **2018**, *9*, 3017.

[144] Karimpour, M. R.; Fedorov, D. V.; Tkatchenko, A. Molecular Interactions Induced by a Static Electric Field in Quantum Mechanics and Quantum Electrodynamics. *The Journal of Physical Chemistry Letters* **2022**, *13*, 2197–2204.

[145] Ambrosetti, A.; Silvestrelli, P. L. van der Waals Interactions in Density Functional Theory Using Wannier Functions: Improved $C_6$ and $C_3$ Coefficients by a Different Approach. *Physical Review B* **2012**, *85*, 073101.

[146] Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B. Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field Based on the Classical Drude Oscillator. *Journal of Chemical Theory and Computation* **2006**, *2*, 1587–1597.

[147] Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *Journal of Chemical Theory and Computation* **2013**, *9*, 5430–5449.

[148] Adluri, A. N. S.; Murphy, J. N.; Tozer, T.; Rowley, C. N. Polarizable Force Field with a $\sigma$-Hole for Liquid and Aqueous Bromomethane. *The Journal of Physical Chemistry B* **2015**, *119*, 13422–13432.

[149] Sokhan, V. P.; Jones, A. P.; Cipcigan, F. S.; Crain, J.; Martyna, G. J. Signature Properties of Water: Their Molecular Electronic Origins. *Proceedings of the National Academy of Sciences* **2015**, *112*, 6341–6346.

[150] Poier, P. P.; Lagardère, L.; Piquemal, J.-P. O(N) Stochastic Evaluation of Many-Body van der Waals Energies in Large Complex Systems. *Journal of Chemical Theory and Computation* **2022**, *18*, 1633–1645.

[151] Muhli, H.; Chen, X.; Bartók, A. P.; Hernández-León, P.; Csányi, G.; Ala-Nissila, T.; Caro, M. A. Machine Learning Force Fields Based On Local Parametrization of Dispersion Interactions: Application to the Phase Diagram of $C_{60}$. *Physical Review B* **2021**, *104*, 054106.

[152] Poier, P. P.; Jaffrelot Inizan, T.; Adjoua, O.; Lagardère, L.; Piquemal, J.-P. Accurate Deep Learning-Aided Density-Free Strategy for Many-Body Dispersion-Corrected Density Functional Theory. *The Journal of Physical Chemistry Letters* **2022**, *13*, 4381–4388.

[153] Gould, T.; Bučko, T. $C_6$ Coefficients and Dipole Polarizabilities for All Atoms and Many Ions in Rows 1–6 of the Periodic Table. *Journal of Chemical Theory and Computation* **2016**, *12*, 3603.

[154] Schwerdtfeger, P.; Nagle, J. K. 2018 Table of Static Dipole Polarizabilities of the Neutral Elements in the Periodic Table. *Molecular Physics* **2019**, *117*, 1200–1225.

[155] Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A Generally Applicable Atomic-Charge Dependent London Dispersion Correction. *The Journal of Chemical Physics* **2019**, *150*, 154122.

[156] Jones, A.; Cipcigan, F.; Sokhan, V. P.; Crain, J.; Martyna, G. J. Electronically Coarse-Grained Model for Water. *Physical Review Letters* **2013**, *110*, 227801.

[157] Porsev, S. G.; Derevianko, A. Accurate Relativistic Many-Body Calculations of van der Waals Coefficients C8 and C10 for Alkali-Metal Dimers. *The Journal of Chemical Physics* **2003**, *119*, 844–850.

[158] Porsev, S. G.; Derevianko, A. High-Accuracy Calculations of Dipole, Quadrupole, and Octupole Electric Dynamic Polarizabilities and van der Waals Coefficients $C_6$, $C_8$, and $C_{10}$ for Alkaline-Earth Dimers. *Journal of Experimental and Theoretical Physics* **2006**, *102*, 195–205.

[159] Jiang, J.; Mitroy, J.; Cheng, Y.; Bromley, M. W. J. Effective Oscillator Strength Distributions of Spherically Symmetric Atoms for Calculating Polarizabilities and Long-Range Atom–atom Interactions. *Atomic Data and Nuclear Data Tables* **2015**, *101*, 158–186.

[160] Tao, J.; Rappe, A. M. Communication: Accurate Higher-order van der Waals Coefficients Between Molecules from a Model Dynamic Multipole Polarizability. *The Journal of Chemical Physics* **2016**, *144*, 031102.

[161] As stated in DiStasio *et al.* [268], *"From a theoretical point of view, an accurate and reliable description of $\alpha_l$ can be quite demanding, and often requires sophisticated treatment of electron correlation in conjunction with large (and diffuse) basis sets [14–19]. From an experimental point of view, $\alpha_l$ measurements are susceptible to (zero-point) vibrational contributions, thermal effects, as well as origin and orientational dependencies [20–22]."* In addition, Woon *et al.* [72] provide a detailed analysis of the slow basis-set convergence of beyond-dipole multipole polarizabilities.

[162] Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes. *Chemical Reviews* **1994**, *94*, 1887.

[163] Kaplan, I. G. *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*; John Wiley & Sons: Ontairo, Canada, 2006.

[164] Van Vleet, M. J.; Misquitta, A. J.; Stone, A. J.; Schmidt, J. R. Beyond Born-Mayer: Improved Models for Short-Range Repulsion in ab Initio Force Fields. *Journal of Chemical Theory and Computation* **2016**, *12*, 3851.

[165] Feynman, R. P. Forces in Molecules. *Physical Reviews* **1939**, *56*, 340.

[166] von Lilienfeld, O. A.; Tkatchenko, A. Two- and Three-Body Interatomic Dispersion Energy Contributions to Binding in Molecules and Solids. *Journal of Chemical Physics* **2010**, *132*, 234109.

[167] Ferri, N.; DiStasio Jr., R. A.; Ambrosetti, A.; Car, R.; Tkatchenko, A. Electronic Properties of Molecules and Surfaces With a Self-Consistent Interatomic van der Waals Density Functional. *Physical Review Letters* **2015**, *114*, 176802.

[168] Ferri, N.; Ambrosetti, A.; Tkatchenko, A. Electronic Charge Rearrangement at Metal/organic Interfaces Induced by Weak van der Waals Interactions. *Physical Review Materials* **2017**, *1*, 026003.

[169] Allen, M. J.; Tozer, D. J. Helium dimer dispersion forces and correlation potentials in density functional theory. *The Journal of Chemical Physics* **2002**, *117*, 11113–11120.

[170] Thonhauser, T.; Cooper, V. R.; Li, S.; Puzder, A.; Hyldgaard, P.; Langreth, D. C. Van der Waals density functional: Self-consistent potential and the nature of the van der Waals bond. *Phys. Rev. B* **2007**, *76*, 125112.

[171] Often the term *polarization potential* refers to a potential given by an expansion in terms of multipole polarizabilities [269] to describe the long-range interaction between an atom and a charged particle. Here, we use this term for the function given by Eq. (3.55) which is another kind of polarization potential caused by an external field.

[172] Adamo, C.; Barone, V. Toward Reliable Density Functional Methods Without Adjustable Parameters: The PBE0 Model. *The Journal of Chemical Physics* **1999**, *110*, 6158–6170.

[173] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865.

[174] Epifanovsky, E., et al. Software for the Frontiers of Quantum Chemistry: An Overview of Developments in the Q-Chem 5 Package. *The Journal of Chemical Physics* **2021**, *155*, 084801.

[175] Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *Journal of Computational Chemistry* **2012**, *33*, 580–592.

[176] Zhang, J.; Lu, T. Efficient Evaluation of Electrostatic Potential With Computerized Optimized Code. *Physical Chemistry Chemical Physics* **2021**, *23*, 20323–20328.

[177] Hait, D.; Head-Gordon, M. How Accurate Are Static Polarizability Predictions from Density Functional Theory? an Assessment Over 132 Species at Equilibrium Geometry. *Physical Chemistry Chemical Physics* **2018**, *20*, 19800.

[178] Neese, F. The ORCA Program System. *WIREs Computational Molecular Science* **2012**, *2*, 73–78.

[179] Neese, F. Software Update: The Orca Program System, Version 4.0. *WIREs Computational Molecular Science* **2018**, *8*, e1327.

[180] Veis, L.; Antalík, A.; Brabec, J.; Neese, F.; Legeza, O.; Pittner, J. Coupled Cluster Method with Single and Double Excitations Tailored by Matrix Product State Wave Functions. *The Journal of Physical Chemistry Letters* **2016**, *7*, 4072–4078.

[181] Derevianko, A.; Porsev, S. G.; Babb, J. F. Electric Dipole Polarizabilities at Imaginary Frequencies for Hydrogen, the Alkali–metal, Alkaline–Earth, and Noble Gas Atoms. *Atomic Data and Nuclear Data Tables* **2010**, *96*, 323–331.

[182] Tang, K. T.; Toennies, J. P. The van der Waals Potentials Between All the Rare Gas Atoms from He to Rn. *The Journal of Chemical Physics* **2003**, *118*, 4976–4983.

[183] Hoja, J.; Sandonas, L. M.; Ernst, B. G.; Vazquez-Mayagoitia, A.; DiStasio Jr, R. A.; Tkatchenko, A. QM7-X, a Comprehensive Dataset of Quantum-Mechanical Properties Spanning the Chemical Space of Small Organic Molecules. *Scientific data* **2021**, *8*, 1–11.

[184] Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361*, 360–365.

[185] von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Exploring Chemical Compound Space With Quantum-Based Machine Learning. *Nature Reviews Chemistry* **2020**, *4*, 347–358.

[186] Bilodeau, C.; Jin, W.; Jaakkola, T.; Barzilay, R.; Jensen, K. F. Generative Models for Molecular Discovery: Recent Advances and Challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2022**, *12*, e1608.

[187] Müller, C.; Bold, S.; Chavarot-Kerlidou, M.; Dietzek-Ivanšić, B. Photoinduced Electron Transfer in Triazole-Bridged Donor-Acceptor Dyads – A Critical Perspective. *Coordination Chemistry Reviews* **2022**, *472*, 214764.

[188] Kranz, C.; Wächtler, M. Characterizing Photocatalysts for Water Splitting: From Atoms to Bulk and from Slow to Ultrafast Processes. *Chemical Society Reviews* **2021**, *50*, 1407–1437.

[189] Zhao, Z.-J.; Liu, S.; Zha, S.; Cheng, D.; Studt, F.; Henkelman, G.; Gong, J. Theory-Guided Design of Catalytic Materials Using Scaling Relationships and Reactivity Descriptors. *Nature Reviews Materials* **2019**, *4*, 792–804.

[190] Wang, X.; Zhang, G.; Yang, L.; Sharman, E.; Jiang, J. Material Descriptors for Photocatalyst/Catalyst Design. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1369.

[191] Medrano Sandonas, L.; Hoja, J.; Ernst, B. G.; Vázquez-Mayagoitia, Á.; DiStasio, R. A.; Tkatchenko, A. "Freedom of Design" in Chemical Compound Space: Towards Rational in silico Design of Molecules with Targeted Quantum-Mechanical Properties. *Chemical Science* **2023**, Advance Article.

[192] Roncali, J. Molecular Engineering of the Band Gap of Pi-Conjugated Systems: Facing Technological Applications. *Macromolecular Rapid Communications* **2007**, *28*, 1761–1775.

[193] Beaujuge, P. M.; Reynolds, J. R. Color Control in Pi-Conjugated Organic Polymers for Use in Electrochromic Devices. *Chemical Reviews* **2010**, *110*, 268–320.

[194] Kubatkin, S.; Danilov, A.; Hjort, M.; Cornil, J.; Brédas, J.-L.; Stuhr-Hansen, N.; Hedegård, P.; Bjørnholm, T. Single-Electron Transistor of a Single Organic Molecule With Access to Several Redox States. *Nature* **2003**, *425*, 698–701.

[195] Katritzky, A. R.; Sild, S.; Karelson, M. Correlation and Prediction of the Refractive Indices of Polymers by QSPR. *Journal of Chemical Information and Computer Sciences* **1998**, *38*, 1171–1176.

[196] Takimiya, K.; Yamamoto, T.; Ebata, H.; Izawa, T. Design Strategy for Air-Stable Organic Semiconductors Applicable to High-Performance Field-Effect Transistors. *Science and Technology of Advanced Materials* **2007**, *8*, 273–276.

[197] Mazinani, S. K. S.; Meidanshahi, R. V.; Palma, J. L.; Tarakeshwar, P.; Hansen, T.; Ratner, M. A.; Mujica, V. Polarizability as a Molecular Descriptor for Conductance in Organic Molecular Circuits. *The Journal of Physical Chemistry C* **2016**, *120*, 26054–26060.

[198] Gryn'ova, G.; Lin, K.-H.; Corminboeuf, C. Read between the Molecules: Computational Insights into Organic Semiconductors. *Journal of the American Chemical Society* **2018**, *140*, 16370–16386.

[199] Bredas, J.-L. Mind the Gap! *Materials Horizons* **2014**, *1*, 17–19.

[200] Stein, T.; Eisenberg, H.; Kronik, L.; Baer, R. Fundamental Gaps in Finite Systems from Eigenvalues of a Generalized Kohn-Sham Method. *Physical Review Letters* **2010**, *105*, 266802.

[201] Karplus, M.; Kolker, H. J. Van der Waals Forces in Atoms and Molecules. *The Journal of Chemical Physics* **1964**, *41*, 3955–3961.

[202] Hohm, U. Is There a Minimum Polarizability Principle in Chemical Reactions? *The Journal of Physical Chemistry A* **2000**, *104*, 8418–8423.

[203] Xie, C.; Oganov, A. R.; Dong, D.; Liu, N.; Li, D.; Debela, T. T. Rational Design of Inorganic Dielectric Materials With Expected Permittivity. *Scientific Reports* **2015**, *5*, 16769.

[204] Sabirov, D. S. Polarizability as a Landmark Property for Fullerene Chemistry and Materials Science. *RSC Advances* **2014**, *4*, 44996–45028.

[205] Kamada, K.; Ueda, M.; Nagao, H.; Tawa, K.; Sugino, T.; Shmizu, Y.; Ohta, K. Molecular Design for Organic Nonlinear Optics: Polarizability and Hyperpolarizabilities of Furan Homologues Investigated by Ab Initio Molecular Orbital Method. *The Journal of Physical Chemistry A* **2000**, *104*, 4723–4734.

[206] Pouchan, C.; Bégué, D.; Zhang, D. Y. Between Geometry, Stability, and Polarizability: Density Functional Theory Studies of Silicon Clusters $Si_n$ (n=3–10). *The Journal of Chemical Physics* **2004**, *121*, 4628–4634.

[207] Cheeseman, J. R.; Frisch, M. J. Basis Set Dependence of Vibrational Raman and Raman Optical Activity Intensities. *Journal of Chemical Theory and Computation* **2011**, *7*, 3323–3334.

[208] Targema, M.; Obi-Egbedi, N. O.; Adeoye, M. D. Molecular Structure and Solvent Effects On the Dipole Moments and Polarizabilities of Some Aniline Derivatives. *Computational and Theoretica Chemistry* **2013**, *1012*, 47–53.

[209] Kim, T.-D.; Lee, K.-S. D-$\pi$-A Conjugated Molecules for Optoelectronic Applications. *Macromolecular Rapid Communications* **2015**, *36*, 943–958.

[210] Wang, D.; Fracasso, D.; Nurbawono, A.; Annadata, H. V.; Sangeeth, C. S. S.; Yuan, L.; Nijhuis, C. A. Tuning the Tunneling Rate and Dielectric Response of SAM-Based Junctions via a Single Polarizable Atom. *Advanced Materials* **2015**, *27*, 6689–6695.

[211] Kaasbjerg, K.; Flensberg, K. Strong Polarization-Induced Reduction of Addition Energies in Single-Molecule Nanojunctions. *Nano Letters* **2008**, *8*, 3809–3814.

[212] Jaque, P.; Toro-Labbé, A. Characterization of Copper Clusters Through the Use of Density Functional Theory Reactivity Descriptors. *The Journal of Chemical Physics* **2002**, *117*, 3208–3218.

[213] Otto, P. Recent Developments in the Theoretical Design of Low-Gap Polymers and Their Nonlinear Optical Properties. *International Journal of Quantum Chemistry* **1994**, *52*, 353–364.

[214] Shang, H.; Raimbault, N.; Rinke, P.; Scheffler, M.; Rossi, M.; Carbogno, C. All-Electron, Real-Space Perturbation Theory for Homogeneous Electric Fields: Theory, Implementation, and Application Within DFT. *New Journal of Physics* **2018**, *20*, 073040.

[215] Karna, S. P. A "Direct" Time-Dependent Coupled Perturbed Hartree–Fock–Roothaan Approach to Calculate Molecular (Hyper) Polarizabilities. *Chemical Physics Letters* **1993**, *214*, 186–192.

[216] Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-Throughput Highway to Computational Materials Design. *Nature Materials* **2013**, *12*, 191–201.

[217] Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Scientific Data* **2014**, *1*, 140022.

[218] Gromski, P. S.; Henson, A. B.; Granda, J. M.; Cronin, L. How to Explore Chemical Space Using Algorithms and Automation. *Nature Reviews Chemistry* **2019**, *3*, 119–128.

[219] Gadaleta, D.; Benfenati, E. A Descriptor-Based Analysis to Highlight the Mechanistic Rationale of Mutagenicity. *Journal of Environmental Science and Health, Part C* **2021**, *39*, 269–292.

[220] Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons: West Sussex, 2013.

[221] Buckingham, R. The Quantum Theory of Atomic Polarization I–Polarization by a Uniform Field. *Proceedings of the Royal Society of London. Series A* **1937**, *160*, 94–113.

[222] Thakkar, A. J. A Hierarchy for Additive Models of Polarizability. *AIP Conference Proceedings* **2012**, *1504*, 586–589.

[223] Gilbert, A.; Baggott, J. E. *Essentials of Molecular Photochemistry*; Wiley-Blackwell: Oxford, 1991.

[224] Joung, J. F.; Han, M.; Hwang, J.; Jeong, M.; Choi, D. H.; Park, S. Deep Learning Optical Spectroscopy Based on Experimental Database: Potential Applications to Molecular Design. *JACS Au* **2021**, *1*, 427–438.

[225] Woodward, R. B. Structure and the Absorption Spectra of a,b-Unsaturated Ketones. *Journal of the American Chemical Society* **1941**, *63*, 1123–1126.

[226] Bahnick, D. A. Use of Huckel Molecular Orbital Theory in Interpreting the Visible Spectra of Polymethine Dyes: An Undergraduate Physical Chemistry Experiment. *Journal of Chemical Education* **1994**, *71*, 171.

[227] Balcells, D.; Skjelstad, B. B. tmQM Dataset–Quantum Geometries and Properties of 86k Transition Metal Complexes. *Journal of Chemical Information and Modeling* **2020**, *60*, 6135–6146.

[228] Wang, J.; Yang, M.; Wang, G.; Zhao, J. Dipole Polarizabilities of Germanium Clusters. *Chemical Physics Letters* **2003**, *367*, 448–454.

[229] Alyar, H.; Kantarci, Z.; Bahat, M.; Kasap, E. Investigation of Torsional Barriers and Nonlinear Optical (NLO) Properties of Phenyltriazines. *Journal of Molecular Structure* **2007**, *834-836*, 516–520.

[230] De Proft, F.; Ayers, P. W.; Fias, S.; Geerlings, P. Woodward-Hoffmann Rules in Density Functional Theory: Initial Hardness Response. *The Journal of Chemical Physics* **2006**, *125*, 214101.

[231] Afzal, M. A. F.; Cheng, C.; Hachmann, J. Combining First-Principles and Data Modeling for the Accurate Prediction of the Refractive Index of Organic Polymers. *The Journal of Chemical Physics* **2018**, *148*, 241712.

[232] Zade, S. S.; Bendikov, M. From Oligomers to Polymer: Convergence in the HOMO-LUMO Gaps of Conjugated Oligomers. *Organic Letters* **2006**, *8*, 5243–5246.

[233] Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272.

[234] DeGroot, M. H. *Probability and Statistics*; Addison-Wesley Pub. Co., 1986.

[235] Xu, S.; Li, J.; Cai, P.; Liu, X.; Liu, B.; Wang, X. Self-Improving Photosensitizer Discovery System via Bayesian Search with First-Principle Simulations. *Journal of the American Chemical Society* **2021**, *143*, 19769–19777.

[236] Brédas, J. L.; Calbert, J. P.; da Silva Filho, D. A.; Cornil, J. Organic Semiconductors: A Theoretical Characterization of the Basic Parameters Governing Charge Transport. *Proceedings of the National Academy of Sciences* **2002**, *99*, 5804–5809.

[237] Yang, D.; Ma, D. Development of Organic Semiconductor Photodetectors: From Mechanism to Applications. *Advanced Optical Materials* **2019**, *7*, 1800522.

[238] Narsaria, A. K.; Poater, J.; Fonseca Guerra, C.; Ehlers, A. W.; Lammertsma, K.; Bickelhaupt, F. M. Rational Design of Near-Infrared Absorbing Organic Dyes: Controlling the Homo–lumo Gap Using Quantitative Molecular Orbital Theory. *Journal of Computational Chemistry* **2018**, *39*, 2690–2696.

[239] Castellani, M.; Winkler, S.; Bröker, B.; Baumgarten, M.; Müllen, K.; Koch, N. Work Function Increase of Transparent Conductive Electrodes by Solution Processed Electron Acceptor Molecular Monolayers. *Applied Physics A* **2014**, *114*, 291–295.

[240] Peng, X.; Hu, L.; Qin, F.; Zhou, Y.; Chu, P. K. Low Work Function Surface Modifiers for Solution-Processed Electronics: A Review. *Advanced Material Interfaces* **2018**, *5*, 1701404.

[241] Kunkel, C.; Schober, C.; Margraf, J. T.; Reuter, K.; Oberhofer, H. Finding the Right Bricks for Molecular Legos: A Data Mining Approach to Organic Semiconductor Design. *Chemistry of Materials* **2019**, *31*, 969–978.

[242] Topping, J. On the Mutual Potential Energy of a Plane Network of Doublets. *Proceedings of the Royal Society London. Series A* **1927**, *114*, 67–72.

[243] Maurer, R. J.; Ruiz, V. G.; Tkatchenko, A. Many-Body Dispersion Effects in the Binding of Adsorbates On Metal Surfaces. *The Journal of Chemical Physics* **2015**, *143*, 102808.

[244] Widdascheck, F.; Hauke, A. A.; Witte, G. A Solvent-Free Solution: Vacuum-Deposited Organic Monolayers Modify Work Functions of Noble Metal Electrodes. *Advanced Functional Materials* **2019**, *29*, 1808385.

[245] Naghdi, S.; Sanchez-Arriaga, G.; Rhee, K. Y. Tuning the Work Function of Graphene Toward Application As Anode and Cathode. *Journal of Alloys and Compounds* **2019**, *805*, 1117–1134.

[246] Seifert, G.; Porezag, D.; Frauenheim, T. Calculations of Molecules, Clusters, and Solids With a Simplified LCAO-DFT-LDA Scheme. *International Journal of Quantum Chemistry* **1996**, *58*, 185–192.

[247] Adamo, C.; Cossi, M.; Scalmani, G.; Barone, V. Accurate Static Polarizabilities by Density Functional Theory: Assessment of the PBE0 Model. *Chemical Physics Letters* **1999**, *307*, 265–271.

[248] Blum, L. C.; Reymond, J.-L. 970 Million Druglike Small Molecules for Virtual Screening in the Chemical Universe Database GDB-13. *Journal of the American Chemical Society* **2009**, *131*, 8732–8733.

[249] Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies With Machine Learning. *Physical Review Letters* **2012**, *108*, 058301.

[250] Hickey, A. L.; Rowley, C. N. Benchmarking Quantum Chemical Methods for the Calculation of Molecular Dipole Moments and Polarizabilities. *The Journal of Physical Chemistry A* **2014**, *118*, 3678–3687.

[251] O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: an Open Chemical Toolbox. *Journal of Cheminformatics* **2011**, *3*, 1–14.

[252] Haider, N. Functionality Pattern Matching as an Efficient Complementary Structure/Reaction Search Tool: an Open-Source Approach. *Molecules* **2010**, *15*, 5079–5092.

[253] Tkatchenko, A.; Ambrosetti, A.; DiStasio, J., Robert A. Interatomic Methods for the Dispersion Energy Derived from the Adiabatic Connection Fluctuation-Dissipation Theorem. *The Journal of Chemical Physics* **2013**, *138*, 074106.

[254] Hermann, J.; Stöhr, M.; Góger, S.; Chaudhuri, S.; Aradi, B.; Maurer, R. J.; Tkatchenko, A. libMBD: A general-purpose package for scalable quantum many-body dispersion calculations. *The Journal of Chemical Physics* **2023**, *159*, 174802.

[255] Hermann, J. Towards Unified Density-Functional Model of Van Der Waals Interactions. Ph.D. thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2018.

[256] Maaskant, W.; Oosterhoff, L. Theory of Optical Rotatory Power. *Molecular Physics* **1964**, *8*, 319–344.

[257] Mandal, A.; Hunt, K. L. C. Gauge-Invariant Expectation Values of the Energy of a Molecule in an Electromagnetic Field. *The Journal of Chemical Physics* **2016**, *144*, 044109.

[258] Babiker, M.; Power, E. A.; Thirunamachandran, T.; Massey, H. S. W. Atomic Field Equations for Maxwell Fields Interacting With Non-Relativistic Quantal Sources. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **1973**, *332*, 187–197.

[259] Kocher, C. A. Point-Multipole Expansions for Charge and Current Distributions. *American Journal of Physics* **1978**, *46*, 578–579.

[260] Woolley, R. G. Molecular Quantum Electrodynamics. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* **1971**, *321*, 557–572.

[261] Craig, D. P.; Thirunamachandran, T. *Molecular quantum electrodynamics: an introduction to radiation-molecule interactions*; Courier Corporation, 1998.

[262] Liu, Y.; Lei, Y.; Zeng, J. Factorization of the Radial Schrödinger Equation and Four Kinds of Raising and Lowering Operators of Hydrogen Atoms and Isotropic Harmonic Oscillators. *Physics Letters A* **1997**, *231*, 9–22.

[263] Cohen-Tannoudji, C.; Diu, B.; Laloë, F.; Hemley, S. R. *Quantum Mechanics*; A Wiley-Interscience publication; Wiley: New York, NY, 1977.

[264] Woolley, R. G. Power-Zienau-Woolley Representations of Nonrelativistic QED for Atoms and Molecules. *Physical Review Research* **2020**, *2*, 013206.

[265] Andrews, D. L.; Jones, G. A.; Salam, A.; Woolley, R. G. Perspective: Quantum Hamiltonians for Optical Interactions. *The Journal of Chemical Physics* **2018**, *148*, 040901.

[266] Szegö, G. *Orthogonal Polynomials*; American Mathematical Society: Providence, RI, 1975.

[267] Erdélyi, A. *Higher Transcendental Functions*; McGraw-Hill Book Company: New York, NY, 1953.

[268] K. U. Lao, Y. Yang, and R. A. DiStasio, Jr., On The Higher-Order Static Polarizabilities and Dispersion Coefficients of the Fullerenes: An Ab Initio Study, doi:10.26434/chemrxiv.12940652.v1.

[269] Zhang, Y.-H.; Tang, L.-Y.; Zhang, X.-Z.; Jiang, J.; Mitroy, J. Convergence of the Multipole Expansions of the Polarization and Dispersion Interactions for Atoms Under Confinement. *The Journal of Chemical Physics* **2012**, *136*, 174107.