UNIVERSITÉ DU LUXEMBOURG

PhD-No PhD-FSTM-2025-032
The Faculty of Science, Technology and Medicine

DISSERTATION

Presented on 10/03/2025 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN PHYSIQUE

by

Almaz KHABIBRAKHMANOV

Born on 17 March 1997 in Kukmor (Russian Federation)

BRIDGING QUANTUM DRUDE OSCILLATORS AND ELECTRONIC-STRUCTURE THEORY FOR VAN DER WAALS DISPERSION INTERACTIONS

Dissertation defense committee

Dr. Alexandre Tkatchenko, Supervisor
Professor, University of Luxembourg

Dr. Massimiliano Esposito, Chairman
Professor, University of Luxembourg

Dr. Adolfo Del Campo, Vice-Chairman
Professor, University of Luxembourg

Dr. Krzysztof Szalewicz
Professor, University of Delaware

Dr. Reinhard Maurer
Professor, University of Warwick

I dedicate this work to my beloved parents, whose nurturing attitude fostered my education and personal growth, making this milestone possible. Thank you for everything, Mom and Dad!

Abstract

Van der Waals (vdW) dispersion forces are fundamental to the structure and behavior of biomolecular, solid-state, and polymeric systems. These interactions, arising from Coulomb-correlated quantum fluctuations in charge density, in principle, demand sophisticated quantum chemistry methods, such as coupled cluster and quantum Monte Carlo. However, the high computational cost of these approaches limits their practical application to large and complex systems. Approximate methods, like classical force fields or semi-local density functional theory (DFT), fall short of capturing the intricacies of vdW dispersion forces. This thesis addresses these limitations by advancing the theoretical description of vdW dispersion interactions through the quantum Drude oscillators (QDOs) framework – a versatile, coarse-grained model for electronic response.

To this end, we first develop a universal, analytical vdW potential based on the QDO model, applicable across the periodic table. With minimal parametrization, this potential is designed for noble gases and generalized to atomic and molecular dimers, achieving high accuracy compared to experimental data and high-level *ab initio* calculations. Relying on just two atomic parameters – dipole polarizability and dipolar dispersion coefficient, our vdW-QDO potential is twice as accurate as the widely used Lennard-Jones potential. This marks a significant advance for biomolecular force fields, where accurate vdW modeling is critical yet remains challenging.

While vdW interaction energies are known to scale with system size, their broader influence on other properties remains less explored. Using the dipole-coupled QDO framework within the many-body dispersion (MBD) method, we examine how vdW dispersion interactions impact electron density. Our findings reveal that these interactions induce significant charge polarization even in systems as small as 100 atoms – a phenomenon often overlooked in semi-local DFT, where vdW forces are usually treated as a *post hoc* correction. To address this, we propose a fully coupled, optimally tuned variant of the MBD model based on vdW-QDO parameters, effectively capturing vdW-induced polarization in diverse systems from small molecules to proteins. Our results indicate potential improvements in density functional approximations by incorporating vdW polarization effects.

While the one-body density contains full information about the ground state of a system in DFT, the universal functional required to extract this information remains elusive. In contrast, the two-body density matrix represents electronic correlations more directly. Building on this idea, this thesis introduces a density-matrix reformulation of the MBD method. This approach facilitates real-space visualization of vdW dispersion interactions, while also linking the dipole-coupled QDO framework of MBD to nonlocal correlation functionals in DFT. The resulting nonlocal MBD correlation kernel is critically assessed against existing nonlocal functionals, offering a deeper insight into the theoretical underpinnings of vdW dispersion interactions.

In conclusion, this thesis highlights the QDO model as a robust framework for advancing vdW dispersion modeling across multiple levels of theory. From developing accurate interatomic potentials to uncovering vdW-induced polarization effects and linking to nonlocal correlation functionals, the QDO framework provides a unified platform to address key challenges in the field. This work enhances the understanding of vdW forces and offers a versatile toolbox for future studies in computational chemistry, biomolecular modeling, and beyond.

Keywords: vdW dispersion interactions, quantum Drude oscillator, many-body dispersion, density functional theory

Preface

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Prof. Alexandre Tkatchenko, for offering me the opportunity to pursue my doctoral studies and for his invaluable guidance, insightful discussions, and unwavering support throughout this journey. I am equally thankful to both past and present members of the Theoretical Chemical Physics group for their collaboration, stimulating conversations, and shared dedication to our collective efforts. In particular, I am deeply grateful to Dr. Dmitry Fedorov and Dr. Matteo Gori, whose scientific advice and personal encouragement have been instrumental in bringing this work to life.

My sincere thanks also extend to my friends and colleagues at the University of Luxembourg, whose presence has enriched not only my professional endeavors but also my personal life, making these past four years both fulfilling and memorable.

I gratefully acknowledge the financial support from the Fonds National de la Recherche Luxembourg (PRIDE19/14063202 grant), which enabled this research and facilitated my participation in inspiring international workshops and conferences. Furthermore, the results presented in this thesis would not have been possible without access to the HPC resources of the University of Luxembourg. I am also thankful to the Institute for Pure and Applied Mathematics at the University of California, Los Angeles, for hosting me in Spring 2022 and providing an intellectually flourishing and productive environment.

Lastly, I want to express my deepest gratitude to all the wonderful people who have supported me throughout my academic journey. To my family and friends – thank you for your unwavering belief in me, your endless patience, and your steadfast support over all these years.

Note on publications

This thesis builds upon and incorporates material from the following (published) papers:

1. **A. Khabibrakhmanov**, D. V. Fedorov, A. Tkatchenko, "Universal Pairwise Interatomic van der Waals Potentials Based on Quantum Drude Oscillators" *J. Chem. Theory Comput.* **19**, 7895-7907, 2023.

2. **A. Khabibrakhmanov**, M. Gori, C. Müller, A. Tkatchenko, "Noncovalent Interactions in Density Functional Theory: All the Charge Density We Do Not See" prepared for submission (2025). Preprint available at https://doi.org/10.26434/chemrxiv-2024-lf3nr.

3. **A. Khabibrakhmanov**, M. Gori, A. Tkatchenko, "Reduced Density Matrix Formalism for the Many-Body Dispersion Method", in preparation (2025).

4. **A. Khabibrakhmanov**, D. V. Fedorov, A. Tkatchenko, *et al.*, "Surprising Success of the Quantum Drude Oscillator Model to Describe Polarization Response and van der Waals Interactions in Atomistic Systems", review article for *J. Chem. Phys.* in preparation (2025)

Other co-authored publications:

5. S. Góger, **A. Khabibrakhmanov**, O. Vaccarelli, D. V. Fedorov, A.Tkatchenko, "Optimized Quantum Drude Oscillators for Atomic and Molecular Response Properties" *J. Phys. Chem. Lett.* **14**, 6217-6223, 2023.

6. A. Kabylda, J. T. Frank, S. S. Dou, **A. Khabibrakhmanov**, L. M. Sandonas, O. T. Unke, S. Chmiela, K. R. Müller, A. Tkatchenko, "Molecular Simulations with a Pretrained Neural Network and Universal Pairwise Force Fields" prepared for submission (2025). Preprint available at https://doi.org/10.26434/chemrxiv-2024-bdfr0-v2.

Articles published during the PhD studies but related to prior research:

7. **A. Khabibrakhmanov**, P. B. Sorokin, "Electronic Properties of Graphene Oxide: Nanoroads Towards Novel Applications" *Nanoscale* **14**, 4131-4144, 2022.

Table of Contents

List of Abbreviations

ACFD	Adiabatic-Connection Fluctuation-Dissipation
AIM	Atom-In-Molecule
CCSD	Coupled Cluster theory with Single and Double excitations
CCSD(T)	Coupled Cluster theory with Single, Double and perturbative Triple excitations
DFA	Density Functional Approximation
DFT	Density Functional Theory
DFTB	Density Functional Tight-Binding
HF	Hartree-Fock
HK	Hohenberg-Kohn
GGA	Generalized Gradient Approximation
KS	Kohn-Sham
LDA	Local Density Approximation
LJ	Lennard-Jones
MAE	Mean Absolute Error
MARE	Mean Absolute Relative Error
MBD	Many-Body Dispersion
ML	Machine Learning
NCI	Noncovalent Interactions
PBE	Perdew-Burke-Enzerhof
QDO	Quantum Drude Oscillator
QM	Quantum Mechanics
QMC	Quantum Monte Carlo
(RS)PT	(Rayleigh-Schrödinger) Perturbation Theory
RPA	Random Phase Approximation
(rs)SCS	(range-separated) Self-Consistent Screening
SAPT	Symmetry-Adapted Perturbation Theory
sc	self-consistent
SE	Schrödinger Equation
TS	Tkatchenko-Scheffler
TT	Tang-Toennies
vdW	van der Waals
VV	Vydrov-Van Voorhis
xc	exchange-correlation
XDM	Exchange-hole Dipole Moment

List of Symbols and Notation

$\nabla_{\mathbf{r}}$	gradient operator: $\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)^T$
Δ, ∇^2	Laplacian operator: $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$
$\int d\mathbf{r}$	integral over three-dimensional space: $\int\int\int dx\,dy\,dz$
$\mathbb{1}$	(three-dimensional) identity matrix
Tr	trace operator
$\boldsymbol{\alpha}$	dipole polarizability tensor
α_0	static dipole polarizability (scalar)
α_l	2^l-pole scalar polarizability
$\gamma^{(N)}$	N-body (reduced) density matrix
δ_{ij}	Kronecker delta
$\delta(\mathbf{r})$	(three-dimensional) Dirac delta-function
ε	(orbital) eigenvalue / xc energy per particle
λ	adiabatic coupling constant
v	Coulomb potential (in atomic units): $v(\mathbf{r},\mathbf{r}') = 1/\lvert\mathbf{r}-\mathbf{r}'\rvert$
ρ	electron density
χ	density-density response function
Ψ	(total) wave function
ψ	(spatial) orbital
ω	angular frequency
C_{2n}	(a sum of) pairwise (2^{l_1}-pole)-(2^{l_2}-pole) dispersion coefficients, $l_1 + l_2 + 1 = n \geq 3$
D_e, R_e	binding energy and equilibrium distance in vdW-bonded atomic dimers
E	(total, relative, or interaction) energy
e	energy density per volume / elementary charge / Euler's constant
f_{damp}	damping function
$\hat{H}$	Hamiltonian operator
k_e	Coulomb constant: $k_e = 1/4\pi\varepsilon_0$
m	(effective) mass
q	(effective) charge
$\mathbf{r}$	electronic or drudonic position
$\mathbf{R}$	nuclear (atomic) position
R_{vdW}	van der Waals radius
$\hat{T}$	kinetic energy operator
$\mathbf{T}$	dipole-dipole coupling tensor
$\hat{V}$	(perturbing) potential
Z	atomic number

List of Figures

Appendices

List of Tables

Chapter 1
Introduction

Why does matter stick together? Why do gases condense to liquids? These and related questions have accompanied humanity throughout history; however, for a very long time, no satisfactory answers were known. A significant step forward was made in 1873, when a Dutch secondary school teacher Johannes Diderick van der Waals defended his doctorate dissertation [1]. In his thesis, he presented the now-famous equation of state named after him:

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT \,, \tag{1.1}$$

where P, V, and T are the pressure, molar volume, and temperature of a gas (liquid), and a and b are empirical constants. The van der Waals (vdW) equation explains the condensation of gases to liquids and predicts the existence of a critical point when gas and liquid phases become indistinguishable. In 1880, van der Waals took the next step and formulated the *law of the corresponding states* [2]:

$$\left(\pi + \frac{3}{v^2}\right)\left(v - \frac{1}{3}\right) = \frac{8}{3}\tau \,, \tag{1.2}$$

where $\pi = P/P_c$, $v = V/V_c$, and $\tau = T/T_c$ are thermodynamic variables normalized by their values at the critical point. This equation states that the behavior of all fluids is governed by this universal law. Obviously, for real fluids this relation is not exact, reflecting the phenomenological nature of the vdW equation of state. Nevertheless, it still provides an essentially correct description, which makes it a very useful and powerful principle. For instance, the law of corresponding states was used by Kamerlingh Onnes to correctly predict the critical point of helium $T_c = 5.2$ K [3], followed by its successful liquefaction and the discovery of superconductivity in 1911. Thus, the law of corresponding states "participated" in two Nobel Prizes in Physics, one awarded to van der Waals in 1910 and another to Kamerlingh Onnes in 1913. For the time, when many leading scientists still did not believe in the existence of atoms [4], the work of van der Waals was extremely important, catalyzing further progress in understanding molecules and laying the foundation for the emergence of chemical physics.

While the meaning of the parameter b in Eq. (1.1) as a measure of molecular volume was clear already to van der Waals, the microscopic origin of the interactions described by a remained obscure until the advent of quantum mechanics (QM). Formally, the first explanation of that mysterious attractive force between molecules belongs to S.C. Wang, who derived the QM expression for the long-range force between two hydrogen atoms in 1927 [5]. However, Wang did not connect that to the attractive term in the vdW equation, which was done only by Fritz London in 1930, who also generalized Wang's derivation to arbitrary molecules [6]. London coined the term *dispersion force* for this QM interaction since it is determined by (virtual) electronic excitations similar to the dispersion of light, even within the same frequency range [7].

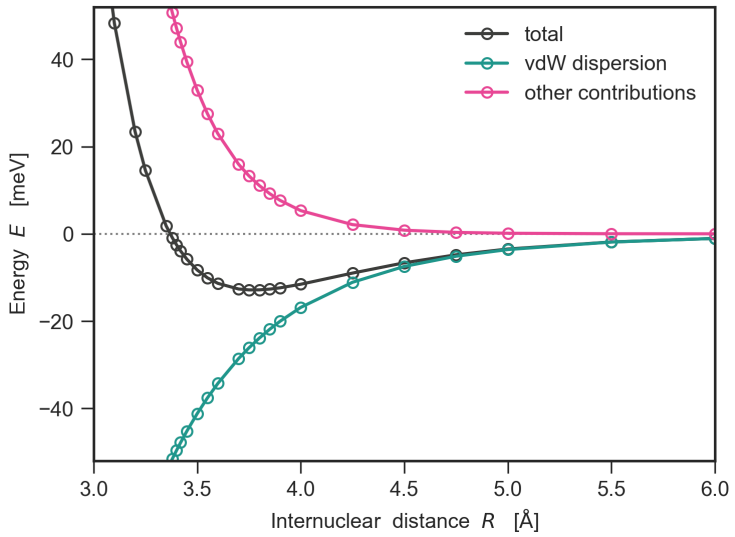

Figure 1.1: vdW dispersion is binding argon dimer. Plotted using the symmetry-adapted perturbation theory (SAPT) data from Ref. [8]. See Section 2.2.1 for further explanations.

More precisely, vdW dispersion arises due to the fluctuations (between ground and excited states) of electrons in interacting moieties, correlated due to the long-range Coulomb interaction. Those fluctuations can be understood e.g. as a consequence of Heisenberg's uncertainty principle, or as a manifestation of a fluctuating vacuum field in quantum electrodynamics. Hence, vdW dispersion is a purely quantum, inherently many-body, and omnipresent force. For instance, it is the primary cohesive force between non-overlapping, spherically symmetric charge distributions, as seen in noble gases (Figure 1.1).

Van der Waals (vdW) dispersion, together with electrostatics, polarization (or induction), and Pauli repulsion, forms the fundamental quartet of noncovalent interactions [7, 9, 10]. Although individually weaker than covalent bonds, noncovalent interactions are ubiquitous and play a decisive role in determining the structure, stability, and behavior of complex systems, especially under ambient conditions where chemical bonds remain largely intact. In biomolecular systems, for instance, protein folding involves intricate rearrangements of amino acid chains while preserving the integrity of the covalent backbone. Here, noncovalent forces – notably vdW dispersion – govern the spatial configuration and overall dynamics, with covalent bonds primarily acting as structural constraints. Beyond biomolecules, vdW interactions are indispensable in a wide range of systems, including layered materials, molecular crystals, supramolecular host-guest complexes, and polymers [11, 12]. In each of these cases, vdW dispersion not only stabilizes molecular assemblies but also underpins their emergent physical and chemical properties.

Atomistic Simulations As a Tool To Study Noncovalent Interactions

In 1929, Paul Dirac famously remarked that chemistry was essentially "solved" after the formulation of the Schrödinger equation (SE) [13], as its exact solution, in principle, provides a complete (non-relativistic) description of all interactions, including noncovalent ones. However, it immediately became apparent that solving the SE exactly for anything beyond the simplest systems was an insurmountable mathematical challenge. This realization gave rise to quantum chemistry, which later evolved into the broader field of computational chemistry. The primary goal of this discipline is to develop accurate approximations of the SE solution while keeping computational costs manageable. Decades of sustained effort have led to a diverse array of computational methods, enabling detailed atomistic simulations of structures, dynamics, and various physical-chemical properties. Today, atomistic simulations serve as an indispensable bridge between theoretical predictions and experimental observations, offering insights that

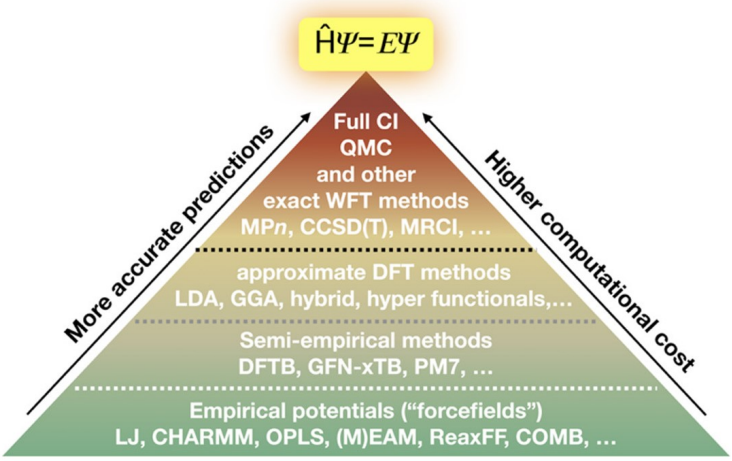

Figure 1.2: Hierarchy of atomistic simulation methods. Adapted from Ref. [18] under the CC-BY-NC-ND 4.0 license. © 2021 The Authors. Published by American Chemical Society.

are often inaccessible through traditional approaches.

This is especially true for noncovalent interactions, which are inherently difficult to measure experimentally and challenging to disentangle from other effects [7, 9, 10]. Modern computational models have become essential tools in advancing biology, chemistry, and materials science. In drug discovery, for example, computational pre-screening significantly reduces the time and cost associated with identifying promising candidates for synthesis. Such a paradigm shift has been driven by breakthroughs in molecular crystal structure prediction and polymorph ranking algorithms [14, 15]. Another milestone is the recent success of neural network models, such as AlphaFold2, in predicting protein structures with unprecedented accuracy [16]. The awarding of the 2024 Nobel Prize in Chemistry for *"for computational protein design"* and *"for protein structure prediction"* [17] (enabled by AlphaFold2) underscores the transformative role of atomistic simulations in contemporary science.

The inherent trade-off between accuracy and computational cost leads to a natural hierarchy of computational methods (Figure 1.2). At the top of this hierarchy lie wave function-based methods, which explicitly treat electronic correlations and are primarily focused on achieving highly accurate energies. The Hartree-Fock (HF) method, which accounts for mean-field effects but neglects electronic correlations, serves as the baseline for correlated methods. Positioned alongside HF are approaches based on density functional theory (DFT), which adopt a distinct theoretical framework but share similar limitations regarding long-range correlation effects. Below these, semi-empirical methods – often derived from tight-binding approximations – offer reduced computational cost at the price of increased parametrization. Finally, at the base of the hierarchy, empirical force fields employ fixed analytical expressions parameterized for specific classes of systems or problems, enabling large-scale simulations at minimal computational expense. A more detailed account of the theoretical methods relevant to the present thesis is provided in Chapter 2.

When considering van der Waals (vdW) dispersion, correlated *ab initio* methods inherently include it as part of the computed electronic correlation energy, offering highly accurate reference data. However, despite advancements in high-performance computing, such methods remain limited to relatively small systems due to their steep computational scaling. For instance, the widely regarded "gold standard" CCSD(T) method scales as $\mathcal{O}(N^7)$ with the number of electrons, making its application impractical for large-scale systems. In contrast, density functional approximations (DFAs) represent the workhorses of computational chemistry, balancing $\mathcal{O}(N^3)$ computational efficiency with a reasonable description of electronic correlation effects at near-equilibrium distances. Nevertheless, most DFAs fail to account for long-range

vdW dispersion, requiring *post hoc* corrections via auxiliary vdW models. These corrections, while effective, often introduce empiricism and coarse-graining, leading to trade-offs between accuracy and interpretability [11, 12, 19, 20].

Recent advances have improved these methods significantly [11, 12, 21], yet a truly universal and fully consistent DFA+vdW framework is yet to be developed (see Section 2.3 for a detailed discussion). A key limitation in practical DFA+vdW methods lies in the treatment of range separation, often relying on empirical damping functions. These functions, while balancing the overall accuracy of the DFA+vdW method, obscure the physical interpretation of the vdW energy contribution. Another research gap is that most DFA+vdW methods neglect the vdW-induced polarization of electron density – a phenomenon first predicted by Feynman [22]. Though often negligible in small molecules, vdW-induced polarization can scale with system size and significantly contribute to ground-state energy, as highlighted in recent studies [23–26], which underlines the importance of the systematic investigation of vdW density polarization.

Physics-Derived Models Are Still Valuable For Modeling

In classical force fields [27–29], vdW dispersion is typically described by the renowned Lennard-Jones (LJ) potential, written in its canonical form as:

$$E_{\mathrm{LJ}} = 4\varepsilon \left[\left(\frac{\sigma}{R}\right)^{12} - \left(\frac{\sigma}{R}\right)^{6} \right] , \tag{1.3}$$

where ε is the potential well depth, and σ is the distance at which the energy equals zero. Proposed in 1924 [30], before the quantum mechanics era, the LJ potential was originally parameterized based on experimental data from noble gases. The exponents were empirically chosen: the repulsive term (later understood as effectively modeling *Pauli repulsion*) was estimated to fall within the range of 11–14, while the attractive term (now recognized as *vdW dispersion*) decayed with a power between 5–7. With London's theoretical insights, it was established that the attractive force universally follows an R^{-6} decay, and the repulsive exponent was set to 12 for computational convenience – despite Pauli repulsion actually decaying exponentially.

This historical example underscores a critical principle: while data fitting can yield models applicable in a certain range of potential energy surfaces, correct asymptotic behavior cannot be reliably captured without an underlying physical model. This principle remains valid even in the era of big data and machine learning (ML). Although ML force fields (MLFFs) offer exceptional flexibility in fitting energies and forces, they must be augmented with physically motivated models to ensure both short-range stability and accurate long-range descriptions. Recognizing this, the MLFF community has increasingly embraced the integration of coarse-grained quantum mechanical (QM) models for noncovalent interactions [31–33]. This hybrid approach is essential for building robust and transferable force fields capable of describing complex interactions present in proteins, lipids, and other macromolecular systems, motivating the further development of coarse-grained QM models for noncovalent forces.

Aim and Outline

The primary objective of this thesis is to deepen our understanding of collective van der Waals (vdW) forces across a diverse spectrum of systems, ranging from simple diatomic molecules to complex biomolecular structures such as proteins. To achieve this, the thesis integrates analytical, density-functional, and *ab initio* approaches into practical methodologies for modeling and visualizing vdW interactions. A key milestone in this pursuit is the development of a universal, broadly applicable interatomic vdW potential that can be further embedded within classical and machine-learning force fields. Future applications of the novel potential might

help to advance the understanding of noncovalent interactions in (bio)molecular systems. Furthermore, the work seeks to uncover vdW dispersion effects that remain inaccessible to current density functional approximations (DFAs), such that vdW-induced density shifts. By exploring the reverse engineering of coarse-grained atomic models towards density functional theory, we propose a route to overcome those limitations, having a potential impact on the design of next-generation DFAs.

This thesis is organized as follows: Chapter 2 introduces the theoretical foundations and methodological tools essential for this work, with an emphasis on the vdW dispersion interactions and the density functional theory. Chapter 3 delves into the construction of universal interatomic vdW potentials, leveraging the quantum analog of the law of corresponding states, the charged quantum harmonic oscillator model, and rigorous analysis of highly accurate quantum chemistry reference data. The relevance of these findings for the development of both classical and machine-learning force fields is highlighted. In Chapter 4, vdW-induced charge density polarization in large and polarizable systems is examined using a dipolar many-body framework, benchmarked against state-of-the-art CCSD and SAPT calculations. The results are discussed in terms of scaling with system size and their implications for biomolecular systems and noncovalent interactions analysis. Chapter 5 introduces a reduced-density matrix reformulation of the many-body dispersion (MBD) method, enabling real-space visualization of vdW dispersion energy as a scalar energy density field. The potential impact on constructing improved nonlocal correlation kernels is envisaged and illustrated through proof-of-concept calculations for diatomic and molecular systems. Finally, Chapter 6 shortly summarizes the main findings, outlines future research directions, and reflects on the broader implications of the presented results. The thesis adopts a semi-cumulative structure, with each chapter prefaced by a concise introduction and discussion tailored to its specific content.

Chapter 2

Theoretical Framework

Section 2.1 provides a brief introduction to the basics of electronic structure theory in the context of quantum chemistry and density functional theory (DFT), with a special emphasis on the adiabatic-connection fluctuation-dissipation theorem (ACFDT). For a more complete overview of the theory and practical methods, see the books by Szabo & Ostlund [34] and Parr & Yang [35] as well as Ref. [36], which this section is based on.

Section 2.2 introduces the textbook classification of noncovalent interactions (NCI) based on the perturbational analysis. The main ideas and equations of symmetry-adapted perturbation theory (SAPT) [37, 38] used for this purpose are summarized. This is followed by a brief overview of two practical models for noncovalent interactions: the Tang-Toennies interatomic potentials and the quantum Drude oscillator model (QDO). The latter is especially important for this thesis as the coarse-grained model for the electronic response of atoms and molecules, laying the foundation for some of the popular vdW methods, such as Tkatchenko-Scheffler (TS) [39] or many-body dispersion (MBD) [40, 41], and providing a fully analytical framework to study noncovalent interactions.

Section 2.3 summarizes theoretical approaches to vdW dispersion interactions, mostly focusing on the practical methods within the framework of DFT. The section includes an exact formulation based on the ACFDT, followed by its successive approximations leading to pairwise-additive vdW models, nonlocal correlation functionals, and the many-body dispersion (MBD) framework [40, 41], the analysis and development of which is in the focus of this thesis.

All derivations provided in this chapter are given in atomic units unless otherwise noted.

2.1 Electronic Many-Body Problem

We consider an N-electron system (atom or molecule) in non-relativistic approximation. Since nuclei are (at least) 1836 times heavier than electrons, the electrons can be considered to move in the field of static nuclei, i.e. nuclear kinetic energy can be neglected. This forms the basis for the *Born-Oppenheimer approximation*, which is central to quantum chemistry. The potential energy of Coulomb repulsion between nuclei can therefore be considered fixed, and one can focus on the electronic Hamiltonian, which reads

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_{\mathbf{r}_i}^2 + \sum_{i=1}^{N} v_{\text{ext}}(\mathbf{r}_i) + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \equiv \hat{T} + \hat{V}_{\text{ext}} + \hat{V}_{\text{ee}}\,, \tag{2.1}$$

where $\mathbf{r}_i$ is the electron coordinate, and the external potential created by N_A nuclear charges $\{Z_A\}$ located at positions $\{\mathbf{R}_A\}$ is given by:

$$v_{\text{ext}}(\mathbf{r}_i) = -\sum_{A=1}^{N_A} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|} \,. \tag{2.2}$$

The stationary electronic states are determined by the time-independent Schrödinger equation (SE):

$$\hat{H}\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) = E\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) \,, \tag{2.3}$$

where $\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)$ is the total wave function in the coordinate representation depending on space-spin coordinates $\mathbf{x}_i = (\mathbf{r}_i, \sigma_i)$ ($\sigma_i \in \{\uparrow, \downarrow\}$), and E is the associated total energy. Due to indistinguishability of electrons and their fermionic spin statistics, the electronic wave function must be *antisymmetric* with respect to permutation of any pair of particles, i.e.

$$\Psi(\mathbf{x}_1, ..., \mathbf{x}_j, ..., \mathbf{x}_i, ..., \mathbf{x}_N) = -\Psi(\mathbf{x}_1, ..., \mathbf{x}_i, ..., \mathbf{x}_j, ..., \mathbf{x}_N) \,. \tag{2.4}$$

This immediately implies that if $\mathbf{x}_i = \mathbf{x}_j$, the wave function is zero meaning that there is zero probability of finding two (or more) same-spin electrons at the same point $\mathbf{r}$ in space. This statement is known as *Pauli exclusion principle* and represents the physical justification for the so-called Pauli- or exchange-repulsion between electrons. The solution of Eq. (2.3) under the constraint (2.4) is the central task of electronic-structure theory. However, the exact (analytical) solution is possible only for hydrogen or hydrogen-like atoms. Therefore, manifold methods have been developed to solve the electronic SE for general many-electron systems.

2.1.1 Hartree-Fock Method

In the Hartree-Fock (HF) theory, the following wave function ansatz called a *Slater determinant* is proposed:

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & ... & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & ... & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & ... & \psi_N(\mathbf{x}_N) \end{vmatrix} \,, \tag{2.5}$$

where $\psi_j(\mathbf{x}_i) = \phi_j(\mathbf{r}_i)\chi_j(\sigma_i)$ is the single-particle spin-orbital, usually written as the product of spatial part ϕ and spin part χ. The Slater determinant by construction satisfies the antisymmetry condition; moreover, any antisymmetric wave function can be represented as an (infinite) linear combination of Slater determinants [34].

The basic idea of the HF method is viewing electrons as independent particles, described by their spin orbitals, moving in the average potential due to other electrons (and nuclei). Therefore, the HF method is often called a *mean-field* theory. Formally, this can be seen from the HF equations for spin-orbitals:

$$\left[-\frac{1}{2}\nabla^2_{\mathbf{r}} + v_{\text{ext}}(\mathbf{r}) + v_{\text{HF}}(\mathbf{r})\right] \psi_i(\mathbf{x}) = \varepsilon_i \psi_i(\mathbf{x}) \,, \tag{2.6}$$

where $v_{\text{HF}}(\mathbf{r})$ is the mean-field HF potential:

$$v_{\text{HF}}(\mathbf{r}) = \sum_{j \neq i} J_j(\mathbf{r}) - K_j(\mathbf{r}) \,. \tag{2.7}$$

The Coulomb potential J_j and the exchange potential K_j operators are defined as

$$J_j(\mathbf{r})\psi_i(\mathbf{x}) = \left[\int d\mathbf{x}' \, \frac{\psi_j^*(\mathbf{x}')\psi_j(\mathbf{x}')}{|\mathbf{r}'-\mathbf{r}|}\right]\psi_i(\mathbf{x}) \ , \tag{2.8}$$

$$K_j(\mathbf{r})\psi_i(\mathbf{x}) = \left[\int d\mathbf{x}' \, \frac{\psi_j^*(\mathbf{x}')\psi_i(\mathbf{x}')}{|\mathbf{r}'-\mathbf{r}|}\right]\psi_j(\mathbf{x}) \ . \tag{2.9}$$

In contrast to the local Coulomb potential, the exchange operator is *nonlocal*, requiring the knowledge of ψ_i throughout all space to evaluate K_j at a given point $\mathbf{r}$. In addition, Eq. (2.6) for ψ_i depends on all other spin orbitals ψ_j. Thus, the HF equations form a set of coupled nonlinear equations that have to be solved iteratively. Starting from an initial guess for orbitals, the iterative cycle continues until the updated orbitals and the resulting mean-field potential become self-consistent. This explains why the HF method is often labeled a *self-consistent field* (SCF) approach. The resulting Hartree–Fock energy – if a self-consistent solution is reached numerically – is variationally bounded from above by the exact ground-state energy E_0. This is ensured by the *variational principle*, the fundamental theorem stating that the expectation value of Hamiltonian over any physically reasonable approximate wave function $\tilde{\Psi}$ provides an upper bound to the exact ground-state energy E_0. Using Dirac notation, this reads

$$\langle\tilde{\Psi}|\hat{H}|\tilde{\Psi}\rangle \geq E_0 \ , \tag{2.10}$$

where the normalization condition $\langle\tilde{\Psi}|\tilde{\Psi}\rangle = 1$ is assumed. This principle, which is behind the derivation of the HF equations, tells that the optimal orbitals can be obtained by minimizing the total energy with respect to them.

In practice, the HF equations are solved using a *basis set expansion* for orbitals:

$$\phi_i(\mathbf{r}) = \sum_j c_{ij}\varphi_j(\mathbf{r}) \ , \tag{2.11}$$

where c_{ij} are expansion coefficients and φ_j are basis functions. Using basis set functions transforms the integro-differential HF equations (2.6)–(2.9) to algebraic matrix equations, which can be efficiently solved for coefficients c_{ij} using linear algebra methods. The real molecular orbitals ϕ_i are of Slater type, i.e. they decay exponentially. However, the mathematical properties of Gaussians greatly simplify the computation of emerging two-electron integrals, which led to their widespread use as basis functions φ_j in quantum chemistry. The exponential decay of orbitals is mimicked by using several Gaussian functions centered on the same atom. To date, a large variety of high-quality Gaussian basis sets optimized for various tasks are readily available in the literature.

The HF approximation works quite well for simple organic molecules with nondegenerate ground states. However, a significant weakness is the neglect of electron correlations. In particular, a Slater determinant (2.5) incorporates *exchange* or *Pauli correlation*, which means that the motion of the same spin electrons is correlated due to wave function antisymmetry. However, the so-called *Coulomb correlation*, which arises from the $1/r$ repulsion between electrons and acts between pairs of parallel and opposite spins, is neglected in the Hartree-Fock theory. Although the contribution of the Coulomb correlation to *total energy* is much smaller than that of the Pauli correlation, for (long-range) *interaction energy* (rigorously defined in Section 2.2.1) the Coulomb correlation is often the most important effect. Since interaction energies are of actual interest for chemical physics, post-Hartree-Fock methods including the Coulomb correlation (further called simply *correlation*) were devised which use the HF approximation as a starting point. The prominent examples include *configuration interaction*, *Møller-Plesset perturbation theory*, and *coupled cluster theory* [34].

2.1.2 Electron Correlation Energy

In quantum chemistry, the electron correlation energy, E_c, is defined as the difference between the exact (non-relativistic) solution of the electronic SE and the HF mean-field approximation [34]:

$$E_c = E_{\text{exact}} - E_{\text{HF}} \,. \tag{2.12}$$

Furthermore, the electron (Coulomb) correlation can be divided into the so-called *static* and *dynamic* correlation. Static correlation refers to the cases when a single Slater determinant is a qualitatively wrong approximation. This typically occurs for (near-)degenerate electronic systems and requires using several reference Slater determinants, leading to the so-called multi-reference methods [34]. However, this thesis focuses only on non-degenerate systems, where static correlation is not relevant. Therefore, throughout this work electron correlation is synonymous to dynamic correlation, which denotes the energy missed due to a mean-field approximation.

In a broader context, the correlation energy can be defined also in other ways. For example, the Kohn-Sham DFT can be used as the mean-field reference instead of the HF method. Therefore, the definition of correlation energy is not unique and depends on the context. In quantum chemistry, the correlated methods seek to accurately approximate the difference in (2.12), while a different strategy is chosen within the framework of the density functional theory. Common approximations of DFT include much of the short-range correlation energy but miss the long-range correlation energy, which can be included, in principle, exactly using the ACFDT formalism (see Section 2.1.4).

2.1.3 Density Functional Theory

In 1964, Pierre Hohenberg and Walter Kohn proved two theorems [42], which state that the ground-state energy is a functional of electron density $\rho(\mathbf{r})$, thereby reducing the description of the system from a $3N$ dimensional wave function Ψ to a functional of three space variables ρ. However, the Hohenberg-Kohn (HK) theorems only prove the existence of a density functional, but do not provide any insight on how to find it. This was overcome in 1965 when Kohn and Sham [43] introduced an auxiliary system of effective non-interacting electrons whose density equals the density of the real system. The KS-DFT enabled the construction of useful and practical density functional approximations (DFA), and in what follows, we briefly summarize its main aspects.

Electron density, the central quantity in the DFT, is expressed in terms of the many-body wave function as:

$$\rho(\mathbf{r}) = N \int |\Psi(\mathbf{x}, \mathbf{x}_2, ..., \mathbf{x}_N)|^2 \, d\sigma \, d\mathbf{x}_2 \ldots d\mathbf{x}_N \,, \tag{2.13}$$

where coefficient N arises due to the indistinguishability of electrons. The physical meaning of $\rho(\mathbf{r})$ is therefore the probability to find *any* of the N electrons in elementary volume $d\mathbf{r}$. Trivially, integrated over all space $\rho(\mathbf{r})$ gives N. The Hohenberg-Kohn theorems tell us that the total energy is a functional of ρ:

$$E[\rho] = T[\rho] + V_{\text{ee}}[\rho] + V_{\text{ext}}[\rho] = F_{\text{HK}}[\rho] + \int \rho(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) \, d\mathbf{r} \,, \tag{2.14}$$

where $T[\rho]$ is the kinetic energy functional, $V_{\text{ee}}[\rho]$ is the Coulomb energy, and $V_{\text{ext}}[\rho]$ is the external potential (due to the nuclei). In the above, we also defined *the universal Hohenberg-Kohn functional*, $F_{\text{HK}}[\rho] = T[\rho] + V_{\text{ee}}[\rho]$. If it was known, the ground-state energy could be computed

exactly for any system. Kohn and Sham introduced a noninteracting reference system described by the KS Hamiltonian:

$$\hat{H}_{\mathrm{KS}} = \hat{T}_{\mathrm{KS}} + \hat{V}_{\mathrm{KS}} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_{\mathbf{r}_i}^2 + \sum_{i=1}^{N} v_{\mathrm{eff}}(\mathbf{r}_i)\,, \tag{2.15}$$

where there is no electron repulsion term, and v_{eff} is the effective one-electron potential. The ground-state wave function of this Hamiltonian is given by a Slater determinant (2.5), with the corresponding density:

$$\rho(\mathbf{r}) = \sum_{i=1}^{N}\sum_{\sigma}|\psi_i(\mathbf{r},\sigma)|^2\,. \tag{2.16}$$

The kinetic energy of the Kohn-Sham system is given by

$$T_s[\rho] = \sum_{i=1}^{N}\langle\psi_i|-\tfrac{1}{2}\nabla_{\mathbf{r}_i}^2|\psi_i\rangle\,, \tag{2.17}$$

which allows one to rewrite $F_{\mathrm{HK}}[\rho]$ as

$$F_{\mathrm{HK}}[\rho] = T_s[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho]\,, \tag{2.18}$$

where we defined the *exchange-correlation (xc) functional*

$$E_{\mathrm{xc}} \equiv T[\rho] - T_s[\rho] + V_{\mathrm{ee}}[\rho] - J[\rho]\,. \tag{2.19}$$

$J[\rho]$ is the classical part of the Coulomb potential energy (*Hartree energy*):

$$J[\rho] = \frac{1}{2}\int\frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}\,d\mathbf{r}\,d\mathbf{r}'\,, \tag{2.20}$$

which corresponds to the same Coulomb operator (2.8) as in the HF theory. The total energy functional thus takes the form

$$E[\rho] = T_s[\rho] + V_{\mathrm{ext}}[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho]\,. \tag{2.21}$$

Essentially, the Kohn-Sham (KS) approach to DFT provides a good approximation to the kinetic energy of electrons, which was the major problem in earlier attempts to construct a density functional, e.g., the Thomas-Fermi model [44, 45]. The remaining unknown parts of the HK functional are "hidden under the carpet" by defining E_{xc} and further approximating it by relatively simple functional forms. The success of the KS-DFT is based on the fact that almost all kinetic energy $T[\rho]$ is recovered by $T_s[\rho]$ and only a small correction enters E_{xc}. In addition, E_{xc} also describes the beyond-classical part of the Coulomb potential, which includes exchange and correlation effects (hence the name).

Using the variational principle (2.10), the Kohn-Sham equations for orbitals can be obtained in the form:

$$\left[-\frac{1}{2}\nabla_{\mathbf{r}}^2 + v_{\mathrm{eff}}(\mathbf{r})\right]\psi_i(\mathbf{x}) = \varepsilon_i\psi_i(\mathbf{x})\,, \tag{2.22}$$

$$v_{\mathrm{eff}}(\mathbf{r}) = v_{\mathrm{ext}}(\mathbf{r}) + \int\frac{\rho(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}\,d\mathbf{r}' + v_{\mathrm{xc}}(\mathbf{r})\,, \tag{2.23}$$

with the *exchange-correlation potential*

$$v_{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}}{\delta\rho(\mathbf{r})}\,. \tag{2.24}$$

Since the effective KS potential depends on ρ, these equations have to be solved together with Eq. (2.16) using a self-consistent iterative procedure, starting from an initial guess for orbitals and minimizing the total energy. Mathematically, the KS equations (2.22) are completely analogous to the HF equations (2.6), with the only important difference that instead of the nonlocal, orbital-dependent potential v_{HF}, in the KS scheme there is a local, orbital-free (i.e. the same for all KS electrons) xc potential v_{xc}. This locality makes the KS equations somewhat easier to solve, with a computational scaling of $\mathcal{O}(N^3)$ compared to $\mathcal{O}(N^4)$ for the canonical HF algorithm.

Another fundamentally important difference is that the KS-DFT provides, in principle, *exact* framework to solve Schrödinger equation, assuming the exact functional $E_{\mathrm{xc}}[\rho]$ were known. In contrast, the HF method is approximate by construction, as it neglects electron correlation altogether. While HF can be systematically improved through a hierarchy of post-HF correlation methods, KS-DFT relies on approximations to $E_{\mathrm{xc}}[\rho]$ for practical use. Although there is no universally systematic path for improving these functionals, significant progress has been made by incorporating exact constraints, physical insights, and empirical data.

To design practical approximations, the xc energy is split into two parts $E_{\mathrm{xc}}[\rho] = E_{\mathrm{x}}[\rho] + E_{\mathrm{c}}[\rho]$ called exchange and correlation functionals, respectively. In the same 1965 paper [43], Kohn and Sham proposed the first parametrization of $E_{\mathrm{xc}}[\rho]$, based on homogeneous electron gas model system and widely known as the *local density approximation* (LDA). In general, the xc energy term is approximated as an explicit functional of $\rho(\mathbf{r})$ and its gradients in conventional LDA and generalized gradient approximations (GGAs), and as a functional of the orbitals $\psi_i(\mathbf{r})$ in more advanced functionals (meta-GGA, hybrids, etc.). However, before diving into specific examples of *density functional approximations* (DFAs), it is instructive to start with the formally exact way to construct xc-functional using the adiabatic connection formalism.

2.1.4 Adiabatic-Connection Fluctuation-Dissipation Theorem

The idea of adiabatic connection is to construct a continuous path from the non-interacting KS system to the fully interacting physical system by slowly (adiabatically) switching on interelectronic Coulomb interaction in the Hamiltonian. This is usually done via introducing a coupling constant λ to the Hamiltonian:

$$\hat{H}_\lambda = \hat{T} + \lambda \hat{V}_{\mathrm{ee}} + \hat{V}_\lambda \,, \tag{2.25}$$

where $\hat{V}_\lambda$ is the external potential, coinciding with the KS potential $\hat{V}_{\mathrm{KS}}$ at $\lambda = 0$ and with the real external potential $\hat{V}_{\mathrm{ext}}$ at $\lambda = 1$. There are infinite paths connecting $\lambda = 0$ and $\lambda = 1$; in DFT, the path is chosen such that the ground-state electron density is kept fixed at its physical value for all $\lambda \in [0, 1]$, which is imposed by the $\hat{V}_\lambda$ operator, having a non-trivial (unknown explicitly) dependence on λ.

Just like for the physical system, it is possible to define a universal functional of density associated with the system of Eq. (2.25) for every λ:

$$F_\lambda[\rho] = T[\rho] + \lambda V_{\mathrm{ee}}[\rho] \,, \tag{2.26}$$

as well as the total energy functional:

$$E(\lambda)[\rho] = F_\lambda[\rho] + V_\lambda[\rho] \,. \tag{2.27}$$

The electronic energy of the real system E ($\lambda = 1$) can be obtained from the KS energy E_0 ($\lambda = 0$) by performing integration:

$$E[\rho] = E_0[\rho] + \int_0^1 d\lambda \frac{dE(\lambda)}{d\lambda} = E_0[\rho] + \int_0^1 d\lambda \frac{dF_\lambda[\rho]}{d\lambda} + \int_0^1 d\lambda \frac{dV_\lambda[\rho]}{d\lambda} \,. \tag{2.28}$$

According to the Hellmann-Feynman theorem, in the derivative

$$\frac{dF_\lambda[\rho]}{d\lambda} = \langle \frac{\partial \Psi_\lambda[\rho]}{\partial \lambda} \big| \hat{T} + \lambda \hat{V}_{\text{ee}} \big| \Psi_\lambda[\rho] \rangle + \langle \Psi_\lambda[\rho] \big| \hat{V}_{\text{ee}} \big| \Psi_\lambda[\rho] \rangle + \langle \Psi_\lambda[\rho] \big| \hat{T} + \lambda \hat{V}_{\text{ee}} \big| \frac{\partial \Psi_\lambda[\rho]}{\partial \lambda} \rangle \quad (2.29)$$

the first and third terms vanish, since the system is always in the ground state along the integration path, and any variation of $\Psi_\lambda[\rho]$ keeping the density constant gives a vanishing variation of $F_\lambda[\rho]$ ($\Psi_\lambda[\rho]$ denotes many-electron wave function yielding the fixed density ρ) [36]. Inserting the second term of the derivative into (2.28) and integrating the term with $\frac{\partial V_\lambda[\rho]}{\partial \lambda}$, we obtain:

$$E[\rho] = T_s[\rho] + V_{\lambda=0}[\rho] + \int_0^1 d\lambda \, \langle \Psi_\lambda[\rho] \big| \hat{V}_{\text{ee}} \big| \Psi_\lambda[\rho] \rangle + V_{\lambda=1}[\rho] - V_{\lambda=0}[\rho] \, . \quad (2.30)$$

By introducing the electron density operator

$$\hat{\rho}(\mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r} - \hat{\mathbf{r}}_i) \, , \quad (2.31)$$

the integrand in (2.30) can be rewritten as:

$$\langle \Psi_\lambda \big| \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \big| \Psi_\lambda \rangle = \frac{1}{2} \iint d\mathbf{r} d\mathbf{r}' \, \langle \Psi_\lambda \big| \frac{\hat{\rho}(\mathbf{r})[\hat{\rho}(\mathbf{r}') - \delta(\mathbf{r} - \mathbf{r}')]}{|\mathbf{r} - \mathbf{r}'|} \big| \Psi_\lambda \rangle \, . \quad (2.32)$$

Further, by adding and subtracting Hartree term $J[\rho]$ (2.20) one can get:

$$E[\rho] = T_s[\rho] + V_{\text{ext}}[\rho] + J[\rho] + \frac{1}{2} \int_0^1 d\lambda \iint d\mathbf{r} d\mathbf{r}' \, \langle \Psi_\lambda \big| \frac{\hat{\rho}(\mathbf{r})[\hat{\rho}(\mathbf{r}') - \delta(\mathbf{r} - \mathbf{r}')] - \rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \big| \Psi_\lambda \rangle \, . \quad (2.33)$$

Finally, comparing to Eq. (2.21) we obtain the formally exact expression for the xc energy:

$$E_{\text{xc}}[\rho] = \frac{1}{2} \int_0^1 d\lambda \iint d\mathbf{r} d\mathbf{r}' \, \frac{\rho^\lambda_{\text{xc}}(\mathbf{r}, \mathbf{r}')\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} \, , \quad (2.34)$$

where we introduced the so-called *exchange-correlation hole*:

$$\rho^\lambda_{\text{xc}}(\mathbf{r}, \mathbf{r}') = \frac{\langle \Psi_\lambda \big| \delta\hat{\rho}(\mathbf{r}) \delta\hat{\rho}(\mathbf{r}') \big| \Psi_\lambda \rangle}{\rho(\mathbf{r})} - \delta(\mathbf{r} - \mathbf{r}') \, , \quad (2.35)$$

with $\delta\hat{\rho}(\mathbf{r}) = \hat{\rho}(\mathbf{r}) - \rho(\mathbf{r})$ being the fluctuation of the density operator around its expectation value. Exchange-correlation hole describes how the presence of an electron at point $\mathbf{r}$ affects the probability density of all other electrons at another point $\mathbf{r}'$ [46], i.e. it is related to the density-density correlation function.

The next step involves connecting the density correlations to the system's response properties through the zero-temperature fluctuation-dissipation theorem (FDT). This theorem, a cornerstone of statistical physics, asserts that a system in thermodynamic equilibrium responds to weak external perturbations in the same way it responds to spontaneous internal fluctuations when no perturbation is applied [47]. The particular formulation of the FDT for density fluctuations leads to [48]

$$\langle \Psi_\lambda \big| \delta\hat{\rho}(\mathbf{r}) \delta\hat{\rho}(\mathbf{r}') \big| \Psi_\lambda \rangle = -\frac{1}{\pi} \int_0^\infty d\omega \, \text{Im}\chi_\lambda(\mathbf{r}, \mathbf{r}', \omega) \, , \quad (2.36)$$

where $\chi_\lambda(\mathbf{r},\mathbf{r}',\omega)$ is the linear density response of the λ-scaled system. The integral of its imaginary part over the real frequency can be substituted by the integration along the imaginary axis using the Wick rotation [49] as $\int_0^\infty d\omega\ \mathrm{Im}\,\chi_\lambda(\omega) = \int_0^\infty d\omega\,\chi_\lambda(i\omega)$. Ultimately, introducing a shorthand notation for Coulomb operator $v(\mathbf{r},\mathbf{r}') = 1/|\mathbf{r}-\mathbf{r}'|$, we arrive at the renowned ACFD expression for E_{xc}:

$$E_{\mathrm{xc}} = -\frac{1}{2\pi}\int_0^1 d\lambda \iint d\mathbf{r}d\mathbf{r}'\, v(\mathbf{r},\mathbf{r}')\left[\int_0^\infty d\omega\,\chi_\lambda(\mathbf{r},\mathbf{r}',i\omega) - \rho(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')\right]. \tag{2.37}$$

For the case of a non-interacting KS system, the response function can be expressed explicitly in terms of KS orbitals $\psi_i(\mathbf{r})$, orbital eigenvalues ε_i and occupation factors f_i as [50, 51]:

$$\chi_0(\mathbf{r},\mathbf{r}',i\omega) = \sum_{ij}(f_i - f_j)\frac{\psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')}{\varepsilon_i - \varepsilon_j + i\omega}. \tag{2.38}$$

This leads to a HF-like *exact exchange* (EX) term:

$$\begin{aligned} E_{\mathrm{x}}^{\mathrm{EX}} = -\frac{1}{2\pi}\iint d\mathbf{r}d\mathbf{r}'\, v(\mathbf{r},\mathbf{r}')\left[\int_0^\infty d\omega\,\chi_0(\mathbf{r},\mathbf{r}',i\omega) - \rho(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')\right] = \\ = -\sum_{ij} f_i f_j \iint d\mathbf{r}d\mathbf{r}'\psi_i^*(\mathbf{r})\psi_j(\mathbf{r})v(\mathbf{r},\mathbf{r}')\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}'). \end{aligned} \tag{2.39}$$

The remaining part of E_{xc} is the correlation energy:

$$E_{\mathrm{c}} = E_{\mathrm{xc}} - E_{\mathrm{x}}^{\mathrm{EX}} = -\frac{1}{2\pi}\int_0^\infty d\omega \iint d\mathbf{r}d\mathbf{r}'\, v(\mathbf{r},\mathbf{r}')\left[\int_0^1 d\lambda\,\chi_\lambda(\mathbf{r},\mathbf{r}',i\omega) - \chi_0(\mathbf{r},\mathbf{r}',i\omega)\right]. \tag{2.40}$$

Using the shorthand notation $\mathrm{Tr}[AB] = \int d\mathbf{r}\int d\mathbf{r}'\, A(\mathbf{r},\mathbf{r}')B(\mathbf{r}',\mathbf{r})$, this equation can be written in a more compact form:

$$E_{\mathrm{c}} = -\frac{1}{2\pi}\int_0^\infty d\omega \int_0^1 d\lambda\,\mathrm{Tr}[\chi_\lambda v - \chi_0 v]. \tag{2.41}$$

This formula would yield the exact correlation energy if the density-response function of the λ-scaled system was known. In practice, however, $\chi_\lambda(\mathbf{r},\mathbf{r}',i\omega)$ has to be approximated.

The interacting response function χ_λ and the KS response function χ_0 are related by the Dyson-like screening equation [52] of the time-dependent density functional theory (TDDFT) [53]:

$$\chi_\lambda(\mathbf{r},\mathbf{r}',i\omega) = \chi_0(\mathbf{r},\mathbf{r}',i\omega) + \iint d\mathbf{r}_1 d\mathbf{r}_2\,\chi_0(\mathbf{r},\mathbf{r}_1,i\omega)[\lambda v(\mathbf{r}_1,\mathbf{r}_2) + f_{\mathrm{xc},\lambda}(\mathbf{r}_1,\mathbf{r}_2,i\omega)]\chi_\lambda(\mathbf{r}_2,\mathbf{r}',i\omega), \tag{2.42}$$

where $f_{\mathrm{xc},\lambda}$ is the so-called exchange-correlation kernel [52], which is still unknown in general form. The simplest approximation is therefore to simply neglect $f_{\mathrm{xc},\lambda}$ in Eq. (2.42), which is known as *random phase approximation* (RPA) and captures accurately long-range vdW dispersion. However, it is quite inaccurate for short-range correlations, therefore practical calculations often rely on the range separation of correlation energy, when its short-range part is described by a semi-local or hybrid DFA and the long-range part is recovered using the ACFD

expression. To avoid double-counting at a short range, the Coulomb potential in the ACFD formula can be range-separated:

$$v(\mathbf{r},\mathbf{r}') = \big(1 - f(|\mathbf{r}-\mathbf{r}'|)\big)\, v(\mathbf{r},\mathbf{r}') + f(|\mathbf{r}-\mathbf{r}'|) v(\mathbf{r},\mathbf{r}') \,, \tag{2.43}$$

where $f(R)$ is a range-separation functions such that $f(0) = 0$ and $f(R) \to 1$ as $R \to \infty$. This leads to partitioning E_c into short- and long-range parts, $E_\mathrm{c} = E_\mathrm{c,sr} + E_\mathrm{c,lr}$. $E_\mathrm{c,sr}$ can then be evaluated using an approximate density functional, while $E_\mathrm{c,lr}$ can be obtained from the ACFD expression (2.41) using the long-range part of Coulomb potential $f(\mathbf{r}-\mathbf{r}')v(\mathbf{r},\mathbf{r}')$. For a more rigorous discussion of the range separation, see e.g. Refs. [54, 55].

However, even within RPA in modern implementations, computation of E_c using (2.40) scales as $\mathcal{O}(N^5)$, and further simplifying approximations are desirable to make accurate calculations of long-range correlation energy more scalable; some of them will be discussed in Section 2.3.

2.1.5 Exchange-Correlation Functionals

Hundreds of xc functionals have been proposed since 1965 when the modern DFT in its Kohn-Sham formulation was born. To navigate in this sea of functionals, the metaphoric *"Jacob's ladder"* concept was proposed by Perdew [56], which classifies DFAs on the rungs of this ladder depending on the level of rigor and accuracy. On the first rung, there is LDA functional [43]

$$E_\mathrm{xc}^\mathrm{LDA}[\rho] = \int \rho(\mathbf{r}) \varepsilon_\mathrm{xc}^\mathrm{HEG}(\rho(\mathbf{r}))\, d\mathbf{r} \,, \tag{2.44}$$

where $\varepsilon_\mathrm{xc}^\mathrm{HEG}(\rho)$ is the exchange-correlation energy per particle of the *homogeneous electron gas* (HEG) with the density ρ. The HEG consists of a uniformly distributed density of electrons in space and the positive background charge density, which is also constant and interacts with electrons and itself electrostatically. The density ρ of electrons therefore fully specifies the HEG system, which is the key physical model underlying the DFT.

The exchange energy per particle of the HEG was first derived by Dirac [57]:

$$\varepsilon_\mathrm{x}^\mathrm{HEG}(\rho) = -\frac{3}{4}\left(\frac{3\rho}{\pi}\right)^{1/3} = -\left(\frac{3^5}{4^4 \cdot \pi^2}\right)^{1/3} \frac{1}{r_s} \approx -\frac{0.458}{r_s} \,, \tag{2.45}$$

where we introduced the *Wigner-Seitz radius* $r_s = (3/4\pi\rho)^{1/3}$. The per-particle correlation energy ε_c cannot be derived analytically but was computed nearly exactly by Ceperley and Alder using Quantum Monte Carlo (QMC) [58], whose results were fitted by accurate analytical parametrizations [59]. Recently, the alternative many-body perturbation theory approach yielded a simple expression for ε_c [60], which agrees within 2 mHa with the near-exact QMC results [58]:

$$\varepsilon_\mathrm{c}^\mathrm{HEG}(\rho) \approx a \ln\left(1 + \frac{b}{r_s} + \frac{b}{r_s^2}\right) , \tag{2.46}$$

with $a = (\ln 2 - 1)/2\pi^2$ and $b = 20.4562557$.

LDA works fairly well for covalent solids and metals, whose density is relatively slowly varying, and it fails for molecules due to their highly inhomogeneous density. The main deficiency of the LDA is the so-called *self-interaction error*. The origin of this error can be understood if one considers a single electron, which should have zero interaction energy. This is correctly predicted by e.g. Hartree-Fock theory, where Hartree and exchange energies exactly cancel each other for the one-electron case [34]. In LDA, this cancellation is incomplete; moreover, the LDA

correlation functional does not vanish in one-electron systems [36], resulting in an unphysical contribution to the energy that can be interpreted as an electron interacting with itself.

This self-interaction error causes LDA and LDA-based functionals to artificially spread out (over-delocalize) the electron density. Another significant source of such an error, now broadly termed the *delocalization error*, stems from the failure of LDA (and many other DFAs) to enforce piece-wise-linear energy dependence for fractional electron numbers [61, 62]. The delocalization error leads to the underestimation of band gaps, spurious charge transfer, and poor reaction barriers [61, 62], and it has been described as *"the greatest outstanding challenge in density-functional theory"* [62].

The next milestone in the DFA development was the generalized gradient approximations (GGA), having the generic form

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho] = \int e_{\mathrm{xc}}^{\mathrm{GGA}}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}))\, d\mathbf{r}\,, \tag{2.47}$$

where $e_{\mathrm{xc}}^{\mathrm{GGA}}$ is some energy density function. To evaluate xc energy at a given point, GGAs use not only the local information via density at this point (like in LDA) but also include information from the vicinity of that point through density gradient. GGAs generally provide a good improvement over LDA on molecular systems and form the second rung of the functionals ladder; however, their accuracy is still limited due to the delocalization error inherited from LDA.

The Perdew-Burke-Ernzerhof (PBE) functional [63] is an outstanding example among GGAs. Unlike many other GGAs, the PBE functional does not have any fitted parameters; it is the most widely used functional for various DFT calculations across materials science and chemistry. In the spin-independent form, the PBE exchange energy density reads

$$e_{\mathrm{x}}^{\mathrm{PBE}}(\rho, \nabla\rho) = \rho\varepsilon_{\mathrm{x}}^{\mathrm{HEG}}(\rho) F_{\mathrm{x}}^{\mathrm{PBE}}(s)\,, \tag{2.48}$$

where we introduced the exchange enhancement factor F,

$$F_{\mathrm{x}}^{\mathrm{PBE}}(s) = 1 + \kappa - \frac{\kappa}{1 + \mu s^2/\kappa}\,, \tag{2.49}$$

with $s = |\nabla\rho|/(2k_F\rho)$ being the *reduced density gradient* (see Section 4.6 for more discussion about s), depending on the Fermi wave vector $k_F = (3\pi^2\rho)^{1/3}$, while μ and κ are numerical parameters. The PBE correlation energy density is given by

$$e_{\mathrm{c}}^{\mathrm{PBE}}(\rho_\uparrow, \rho_\downarrow, \nabla\rho_\uparrow, \nabla\rho_\downarrow) = \rho\left[\varepsilon_{\mathrm{c}}^{\mathrm{HEG}}(\rho_\uparrow, \rho_\downarrow) + H(\rho_\uparrow, \rho_\downarrow, t)\right]\,, \tag{2.50}$$

where $\rho = \rho_\uparrow + \rho_\downarrow$ is the sum of spin-up and spin-down densities, $H(\rho_\uparrow, \rho_\downarrow, t)$ is a rather elaborate mathematical function [63], and $t = |\nabla\rho|/(2\phi(\zeta)\rho\sqrt{4k_F/\pi})$ is another dimensionless density gradient (adapted to correlation), with spin-scaling function $\phi(\zeta) = [(1+\zeta)^{2/3} + (1-\zeta)^{2/3}]/2$ depending on the relative spin polarization $\zeta = (\rho_\uparrow - \rho_\downarrow)/\rho$.

The third rung of Jacob's ladder is represented by the *meta-GGA* functionals, which include the dependence on the Laplacian of the density and/or on the KS kinetic energy density $\tau(\mathbf{r})$:

$$\tau(\mathbf{r}) = \frac{1}{2}\sum_{i=1}^{N} |\nabla\psi_i(\mathbf{r})|^2\,. \tag{2.51}$$

Since GGA and meta-GGA functionals both rely on density gradient information around the given point, together they are often referred to as *semi-local* approximations. The motivation

to use $\tau(\mathbf{r})$ is because it contains information similar to the Laplacian, which can be seen from its second-order gradient expansion:

$$\tau(\mathbf{r}) \approx \tau^{\mathrm{HEG}}(\mathbf{r}) + \frac{1}{72}\frac{|\nabla\rho(\mathbf{r})|^2}{\rho(\mathbf{r})} + \frac{1}{6}\nabla^2\rho(\mathbf{r}) \,, \tag{2.52}$$

and $\tau(\mathbf{r})$ is also useful for identifying different spatial regions of electronic density [64]. This can be done by comparing it with the von Weizsäcker kinetic energy density,

$$\tau^{\mathrm{W}}(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|^2}{8\rho(\mathbf{r})} \,, \tag{2.53}$$

which is exact for one- and two-electron systems. For instance, the iso-orbital indicator [64]:

$$h(\mathbf{r}) = \frac{\tau(\mathbf{r}) - \tau^{\mathrm{W}}(\mathbf{r})}{\tau^{\mathrm{HEG}}(\mathbf{r})} \,, \tag{2.54}$$

where $\tau^{\mathrm{HEG}}(\mathbf{r}) = \frac{3}{10}(3\pi^2)^{2/3}\rho^{5/3}(\mathbf{r})$ is the non-interacting kinetic energy density of the HEG. The indicator $h(\mathbf{r})$ can distinguish one-orbital regions (e.g. covalent bonds) with $h = 0$, slowly varying density regions ($h \approx 1$), and regions of density tails overlap characteristic to noncovalent interactions ($h \gg 1$). The use of orbital-dependent $\tau(\mathbf{r})$ allows meta-GGA to essentially eliminate the self-interaction error in the correlation functional [36]. However, their exchange part is still contaminated with the self-interaction error; also, meta-GGAs often suffer from numerical instabilities. Among the prominent meta-GGA examples, TPSS [65] and SCAN [66] functionals are worth mentioning, while the development of better meta-GGAs remains the area of active research.

The self-interaction error in the exchange functional is substantially reduced in the *hybrid functionals* named so because of mixing a portion of HF-like *exact exchange* from Eq. (2.39) with the base GGA functional. This idea was first proposed by Becke [67, 68] in a three-parametric form, exemplified below with the B3LYP [69], the most popular functional of this form:

$$E_{\mathrm{xc}}^{\mathrm{B3LYP}}[\Phi] = aE_{\mathrm{x}}^{\mathrm{EX}}[\Phi] + bE_{\mathrm{x}}^{\mathrm{B88}}[\rho_\Phi] + (1-a-b)E_{\mathrm{x}}^{\mathrm{LDA}}[\rho_\Phi] + cE_{\mathrm{c}}^{\mathrm{LYP}}[\rho_\Phi] + (1-c)E_{\mathrm{c}}^{\mathrm{LDA}}[\rho_\Phi] \,, \tag{2.55}$$

with $a = 0.20$, $b = 0.72$, and $c = 0.81$ found by fitting the energies of small molecules. Note that B3LYP uses different GGA functionals for exchange (Becke, B88 [70]) and correlation (Lee-Yang-Parr, LYP [71]), the strategy very commonly adopted in the development of DFAs and leading to combinatorial increase in their number. Φ in Eq. (2.55) denotes the Slater determinant composed of the KS orbitals, and ρ_Φ is the corresponding density.

Later, a simpler one-parametric hybrid approximation was suggested, with PBE0 functional [72] as a prominent example:

$$E_{\mathrm{xc}}^{\mathrm{PBE0}}[\Phi] = aE_{\mathrm{x}}^{\mathrm{EX}}[\Phi] + (1-a)E_{\mathrm{x}}^{\mathrm{PBE}}[\rho_\Phi] + E_{\mathrm{c}}^{\mathrm{PBE}}[\rho_\Phi] \,. \tag{2.56}$$

The fraction of exact exchange is fixed at $a = 0.25$ following the rationale proposed in Ref. [73], which is used also in many other hybrid approximations. Hence, there is the '0' in the name, emphasizing the non-empirical nature of a (in the sense of no fitting to reference energies). In general, hybrid functionals (4th rung) offer a systematic improvement over semi-local approximations for molecules with large electronic gaps, while systems with stretched bonds or with transition metal elements remain challenging examples even for hybrids.

There are more advanced approximations, such as range-separated hybrids (the long-range exact exchange plus the short-range GGA exchange) and double hybrids (mixing a fraction of the

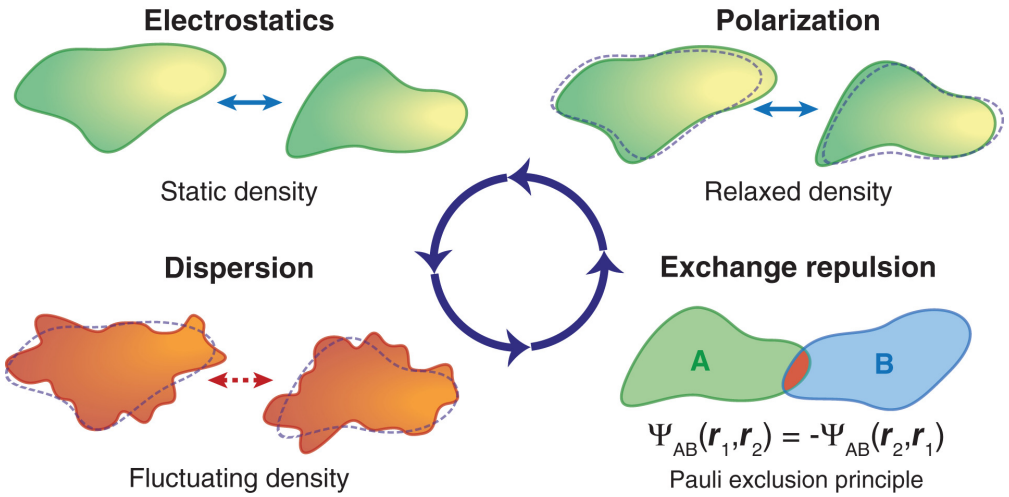

Figure 2.1: Pictorial representation of noncovalent interactions. Note that "polarization" is also often termed as "induction". Reproduced with permission from Ref. [74]. © 2017 John Wiley & Sons, Ltd.

MP2 correlation energy in addition to mixing the exact exchange). However, they are not applied in this thesis and hence not covered here, see e.g. Ref. [36] for further reading. The nonlocal vdW density functionals as well as semi-local DFT+vdW methods are discussed in detail in Section 2.3.

2.2 Noncovalent Interactions

Noncovalent interactions between closed-shell molecules can be broadly categorized into four types: **electrostatics**, **exchange repulsion**, **induction** (or **polarization**), and **dispersion**. In the framework of a textbook perturbation theory, electrostatics refers to the Coulomb interaction between static charge distributions, described by permanent multipoles at long range (e.g., dipole-dipole or dipole–quadrupole). Exchange repulsion force stems from the Pauli exclusion principle, which prevents same-spin electrons from occupying the same space. Induction describes the interaction between static density and the density polarization it induces on a neighboring molecule. Finally, dispersion results from interactions between instantaneous multipoles generated by zero-point electron fluctuations, which induce corresponding multipoles in adjacent molecules. Notably, the dispersion is universally present, as it does not rely on permanent multipoles. The physical nature of these four interactions is visually depicted in Figure 2.1.

We now provide a rigorous mathematical definition for each of the aforementioned interactions using the framework of the symmetry-adapted perturbation theory.

2.2.1 Symmetry-Adapted Perturbation Theory

Methods of quantum chemistry and DFT enable accurate computations of total energies E_{tot} for electronic systems. However, in most cases, we are interested in interaction energy rather than the total energy. For a complex AB at its equilibrium geometry, the interaction energy E_{int} is usually defined as

$$E_{\text{int}}(AB) = E_{\text{tot}}(AB) - E_{\text{tot}}(A) - E_{\text{tot}}(B)\,, \tag{2.57}$$

where total energies of fragments A and B are calculated at their geometries in complex (i.e., the deformation energy of fragments is ignored). This is the so-called *supermolecular* approach

to compute the interaction energy using any electronic structure method. Such a straightforward approach, however, implies obtaining a small number as a difference of two large numbers, which requires the method in use to be very accurate. This becomes especially relevant in the context of noncovalent interactions (NCI), whose energies are several orders of magnitude smaller than the total energies. In addition, the supermolecular approach is prone to the basis set superposition error (BSSE), resulting from a more saturated basis set in a dimer calculation compared to monomer counterparts.

The alternative approach, offered by the symmetry-adapted perturbation theory (SAPT), enables the direct calculation of the interaction energy *without* performing the actual correlated calculation for a dimer. Moreover, instead of a single number provided by the supermolecular approach, SAPT provides the total interaction energy as the sum of components corresponding to basic noncovalent interactions. As the name suggests, SAPT is based on the perturbational treatment of the molecular Hamiltonian. Within the context of this thesis, SAPT is mostly used as a benchmark method, providing reliable estimates for physical components of the interaction energy. Therefore, this section gives only a brief overview of the main concepts and equations of SAPT; more detailed discussions can be found e.g. in Refs. [37, 38].

Rayleigh-Schrödinger Perturbation Theory At Long Range

We start from the Hamiltonian of interacting monomers A and B

$$\hat{H} = \hat{H}_A + \hat{H}_B + \hat{V} \,, \tag{2.58}$$

where $\hat{H}_A$ and $\hat{H}_B$ are the Hamiltonians of isolated monomers, and the intermolecular interaction operator

$$\hat{V} = \sum_{i\in A}\sum_{j\in B}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{j\in B}\sum_{\alpha\in A}\frac{Z_\alpha}{|\mathbf{r}_j - \mathbf{R}_\alpha|} - \sum_{i\in A}\sum_{\beta\in B}\frac{Z_\beta}{|\mathbf{r}_i - \mathbf{R}_\beta|} + \sum_{\alpha\in A}\sum_{\beta\in B}\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \,, \tag{2.59}$$

will be treated as a perturbation. Notation is self-explanatory, with i, j marking electrons and α, β indexing nuclei. We assume that eigenstates and eigenfunctions of $\hat{H}_A$ and $\hat{H}_B$ are known:

$$\hat{H}_A\Psi_0^A = E_0^A\Psi_0^A \,, \quad \hat{H}_B\Psi_0^B = E_0^B\Psi_0^B \,, \tag{2.60}$$

leading to the zeroth-order wave function and energy of the dimer as

$$\Psi_0 = \Psi_0^A\Psi_0^B \,, \quad E_0 = E_0^A + E_0^B \,. \tag{2.61}$$

By labeling electrons such that the first N_A of them initially belong to monomer A, and the others with numbers $N_A+1, ..., N_A+N_B$ are initially assigned to monomer B, the zeroth-order wave function can be written as

$$\Psi_0(1, ..., N_A+N_B) = \Psi_0^A(1, ..., N_A) \times \Psi_0^B(N_A+1, ..., N_A+N_B) \,, \tag{2.62}$$

depending on the spatial and spin coordinates of electrons. Clearly, Ψ_0 respects the Pauli exclusion principle (2.4) within each monomer. However, the *intermolecular* Pauli exclusion principle is not obeyed by this wave function, which means that interaction energies obtained by perturbing Ψ_0 with operator (2.59) will miss the exchange contributions. This approach is known as *polarization approximation* [37, 75]. Despite neglecting the exchange effects, it provides the correct leading-order terms in the asymptotic expansion of the intermolecular interaction energy, where the exchange is negligibly small [9]. Moreover, the textbook Rayleigh-Schrödinger

perturbation theory (RSPT) [9] used in polarization approximation can be modified to incorporate exchange effects by a suitable permutational *symmetry adaptation* (*vide infra*).

The standard RSPT gives the first-order energy correction:

$$E^{(1)}_{\text{elst}} = \langle \Psi^A_0 \Psi^B_0 | \hat{V} | \Psi^A_0 \Psi^B_0 \rangle \,, \tag{2.63}$$

which can be rewritten in the form

$$E^{(1)}_{\text{elst}} = \iint \frac{\rho^{\text{tot}}_A(\mathbf{r}) \rho^{\text{tot}}_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \,, \tag{2.64}$$

in terms of total (electronic plus nuclear) charge densities of the fragments:

$$\rho^{\text{tot}}_X(\mathbf{r}) = -\rho_X(\mathbf{r}) + \sum_{\alpha \in X} Z_\alpha \delta(\mathbf{r} - \mathbf{R}_\alpha) \,, \quad X = A, B \,. \tag{2.65}$$

Eq. (2.64) provides a clear physical interpretation for $E^{(1)}_{\text{elst}}$ as the electrostatic between the unperturbed charge densities of the monomers.

The second-order RSPT energy correction consists of the two terms. The first one includes excitations only on one of the monomers and is called *induction energy*:

$$E^{(2)}_{\text{ind}} = E^{(2),A\leftarrow B}_{\text{ind}} + E^{(2),B\leftarrow A}_{\text{ind}} = -\sum_{k\neq 0} \frac{|\langle \Psi^A_0 \Psi^B_0 | \hat{V} | \Psi^A_k \Psi^B_0 \rangle|^2}{E^A_k - E^A_0} - \sum_{k\neq 0} \frac{|\langle \Psi^A_0 \Psi^B_0 | \hat{V} | \Psi^A_0 \Psi^B_k \rangle|^2}{E^B_k - E^B_0} \,, \tag{2.66}$$

where Ψ^X_k and E^X_k are the excited wave functions and energies of the operator $\hat{H}_X$, $X = A, B$. The notation $A \leftarrow B$ means that the monomer A is perturbed (polarized) by the presence of B. The corresponding first-order correction to the wave function is given by

$$\Psi^{(1)}_{\text{ind}} = -\sum_{k\neq 0} \frac{\langle \Psi^A_k \Psi^B_0 | \hat{V} | \Psi^A_0 \Psi^B_0 \rangle}{E^A_k - E^A_0} \Psi^A_k \Psi^B_0 - \sum_{k\neq 0} \frac{\langle \Psi^A_0 \Psi^B_k | \hat{V} | \Psi^A_0 \Psi^B_0 \rangle}{E^B_k - E^B_0} \Psi^A_0 \Psi^B_k \,. \tag{2.67}$$

Further insight into the meaning of the induction term can be obtained by expressing Eq. (2.66) through the densities as [38]

$$E^{(2),A\leftarrow B}_{\text{ind}} = -\frac{1}{2} \iint \frac{\rho^{(1)}_{\text{ind},A}(\mathbf{r}) \rho^{\text{tot}}_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \,, \tag{2.68}$$

where $\rho^{(1)}_{\text{ind},A}(\mathbf{r})$ is the first-order induced electron density on molecule A due to the molecule B [38]. Thus, the induction term can be interpreted as the density deformation of molecule A (by the electric field of B), that furthermore interacts with the unperturbed density of B. The $1/2$ coefficient can be understood as half of the Coulomb interaction energy between $\rho^{(1)}_{\text{ind},A}(\mathbf{r})$ and $\rho^{\text{tot}}_B(\mathbf{r}')$ is spent as the quantum-mechanical "work" to polarize the density of A [38], similar to the classical explanation for $1/2$ factor in polarization energy [9]. An analogous expression can be written for $E^{(2),B\leftarrow A}_{\text{ind}}$.

Another second-order term from the RSPT interaction energy, *dispersion energy*, involves excitations on both monomers:

$$E^{(2)}_{\text{disp}} = -\sum_{m\neq 0} \sum_{n\neq 0} \frac{|\langle \Psi^A_0 \Psi^B_0 | \hat{V} | \Psi^A_m \Psi^B_n \rangle|^2}{(E^A_m - E^A_0) + (E^B_n - E^B_0)} \,. \tag{2.69}$$

This expression was first derived by London [6], who applied it to the quantum harmonic oscillator as the model for atomic response and using the multipole expansion obtained the famous

$-C_6/R^6$ leading-order term for the asymptotic dispersion energy. Alternative theoretical and practical approaches to calculating dispersion energy are discussed separately in Section 2.3.

We note that the dispersion energy cannot be easily represented in terms of monomer densities, unlike the electrostatic and induction contributions, as the first-order dispersion correction to the wave function

$$\Psi^{(1)}_{\text{disp}} = -\sum_{m\neq 0}\sum_{n\neq 0}\frac{\langle \Psi^A_m \Psi^B_n|\hat{V}|\Psi^A_0\Psi^B_0\rangle}{(E^A_m - E^A_0)+(E^B_n - E^B_0)}\Psi^A_m\Psi^B_n \tag{2.70}$$

involves *intermonomer* correlation of electronic positions [38]. However, Eq. (2.69) can be reformulated in an alternative form known as the *generalized Casimir-Polder equation* [9, 76–79]:

$$E^{(2)}_{\text{disp}} = -\frac{1}{2\pi}\int_0^\infty d\omega \iint d\mathbf{r}d\mathbf{r}' \iint d\mathbf{r}''d\mathbf{r}''' \frac{\chi_A(\mathbf{r},\mathbf{r}',i\omega)}{|\mathbf{r}-\mathbf{r}''|}\frac{\chi_B(\mathbf{r}'',\mathbf{r}''',i\omega)}{|\mathbf{r}'-\mathbf{r}'''|}\,, \tag{2.71}$$

where $\chi_X(\mathbf{r},\mathbf{r}',i\omega)$ is the density-density response function of a system X, in this context sometimes called *frequency-dependent density susceptibility* (FDDS):

$$\chi_X(\mathbf{r},\mathbf{r}',\omega) = 2\sum_{m\neq 0}\frac{E^X_m - E^X_0}{(E^X_m - E^X_0)^2-\omega^2}\langle\Psi^X_0|\hat{\rho}(\mathbf{r})|\Psi^X_m\rangle\langle\Psi^X_m|\hat{\rho}(\mathbf{r}')|\Psi^X_0\rangle\,, \tag{2.72}$$

where $\hat{\rho}(\mathbf{r})$ is the density operator (2.31). While being formally exact, the expression (2.71) involves a 13-dimensional integration, making it a computationally expensive route to get the dispersion energy. If we substitute the density operator with the dipole operator in Eq. (2.72), we would obtain the expression for nonlocal dipole polarizability $\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}',i\omega)$, another linear response function (tensor). Using accurate (semi-)local approximations to $\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}',i\omega)$, the dispersion energy could be evaluated in much more practical ways. Please see Section 2.3.1 for the discussion about the connection between $\chi(\mathbf{r},\mathbf{r}',i\omega)$ and $\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}',i\omega)$.

Symmetry Adaptation

To make perturbation theory applicable throughout the full range of intermonomer distances, the antisymmetry of the dimer wave function relative to the intermonomer electron exchange should be enforced to account for exchange energy contributions. There are multiple ways to do this symmetry adaptation, see e.g. Ref. [37] for the review. However, all practical implementations of SAPT use the simplest *symmetrized Rayleigh-Schrödinger* (SRS) approach [80], sometimes also called *weak symmetry forcing* [37, 75]. The main idea is that the wave functions are still computed using the ordinary RSPT, but the energy expressions are antisymmetrized. Then the exchange energy contributions can be computed as the differences between the corresponding SRS and RS energy corrections [81], i.e.

$$E^{(n)}_{\text{exch}} = E^{(n)}_{\text{SRS}} - E^{(n)}_{\text{RS}}\,, \quad n = 1,2\,, \tag{2.73}$$

where the SRS energies are computed as

$$E^{(1)}_{\text{SRS}} = \frac{\langle\Psi^A_0\Psi^B_0|\hat{V}\mathscr{A}|\Psi^A_0\Psi^B_0\rangle}{\langle\Psi^A_0\Psi^B_0|\mathscr{A}|\Psi^A_0\Psi^B_0\rangle}\,, \tag{2.74}$$

$$E^{(2)}_{\text{SRS}} = \frac{\langle\Psi^A_0\Psi^B_0|(\hat{V}-E^{(1)}_{\text{SRS}})\mathscr{A}\Psi^{(1)}_{\text{RS}}\rangle}{\langle\Psi^A_0\Psi^B_0|\mathscr{A}\Psi^A_0\Psi^B_0\rangle}\,, \tag{2.75}$$

with $\Psi^{(1)}_{\mathrm{RS}} = \Psi^{(1)}_{\mathrm{ind}} + \Psi^{(1)}_{\mathrm{disp}}$. The operator $\mathscr{A}$ introduced in the equations is the all-electron *antisymmetrizer*, which can be formally represented as [75]

$$\mathscr{A} \approx \frac{N_A! N_B!}{(N_A + N_B)!} \mathscr{A}_A \mathscr{A}_B (1 + \mathscr{P}_1) \,, \tag{2.76}$$

where $\mathscr{A}_X$ is the antisymmetrizer for the monomer X and $\mathscr{P}_1$ is the *single-exchange operator*:

$$\mathscr{P}_1 = -\sum_{i \in A} \sum_{j \in B} P_{ij} \,, \tag{2.77}$$

with P_{ij} being the operator exchanging two electrons i and j between monomers A and B. In Eq. (2.76), we neglected the double and higher-order electron exchanges, which is also known as S^2 *approximation*, since the resulting exchange energy terms are quadratic in the overlap integral S [37, 38].

The first order exchange $E^{(1)}_{\mathrm{exch}} = E^{(1)}_{\mathrm{SRS}} - E^{(1)}_{\mathrm{elst}}$ is the dominant exchange contribution, accounting for about 90% of the total exchange energy at the van der Waals minima of dimers [37]. The second-order exchange naturally splits into the induction and dispersion parts:

$$E^{(2)}_{\mathrm{exch}} = E^{(2)}_{\mathrm{SRS}} - E^{(2)}_{\mathrm{RS}} = E^{(2),A\leftarrow B}_{\mathrm{exch-ind}} + E^{(2),B\leftarrow A}_{\mathrm{exch-ind}} + E^{(2)}_{\mathrm{exch-disp}} \equiv E^{(2)}_{\mathrm{exch-ind}} + E^{(2)}_{\mathrm{exch-disp}} \,, \tag{2.78}$$

which can be obtained by inserting Eqs. (2.67) and (2.70) into (2.75) [37, 38]. Both second-order exchange terms are repulsive in the ground state, reflecting the short-range quenching of the attractive induction and dispersion interactions imposed by the Pauli exclusion principle [75].

Thus, the total second-order SAPT interaction energy contains six terms:

$$E^{(2),\mathrm{SAPT}}_{\mathrm{int}} = E^{(1)}_{\mathrm{elst}} + E^{(1)}_{\mathrm{exch}} + E^{(2)}_{\mathrm{ind}} + E^{(2)}_{\mathrm{exch-ind}} + E^{(2)}_{\mathrm{disp}} + E^{(2)}_{\mathrm{exch-disp}} \,. \tag{2.79}$$

Often, it is convenient to pair the corresponding second-order contributions together,

$$E^{(2)}_{\mathrm{indx}} = E^{(2)}_{\mathrm{ind}} + E^{(2)}_{\mathrm{exch-ind}} \,, \quad E^{(2)}_{\mathrm{dispx}} = E^{(2)}_{\mathrm{disp}} + E^{(2)}_{\mathrm{exch-disp}} \,, \tag{2.80}$$

which reduces the number of terms to four, corresponding to four basic noncovalent interactions and enabling the comparison with other energy decomposition analysis (EDA) methods [38]. For instance, the induction terms $E^{(2)}_{\mathrm{ind}}$ and $E^{(2)}_{\mathrm{exch-ind}}$ almost exactly cancel out for neutral monomers with spherically symmetric densities, like in noble gases [8]. In Figure 1.1 for argon dimer, we plotted $E^{(2)}_{\mathrm{dispx}}$ in green and the sum of the remaining terms in Eq. (2.79) in pink, while the total SAPT interaction energy is displayed in black.

Practical SAPT Methods and Applications

All the discussion above, where we introduced the SAPT framework, assumed for simplicity that the monomer energies and wave functions are known *exactly* at the full CI level. However, such computations can be carried out in practice only for the smallest systems, like He–He or He–H dimers [75]. Therefore, in practical calculations, one has to opt for a simpler method to provide zeroth-order wave functions and energies. The natural choice is to use the Hartree-Fock method, which leads to the so-called SAPT0 method. However, the HF wave functions completely neglect electron correlations inside monomers, and the method can be improved by using the Møller-Plesset (MP) perturbation theory to the second or higher order to recover (at least partially) *intramolecular* correlation effects [37, 38, 75]. Additionally, the third-order SRS

perturbation terms for interaction energy can be included [37, 38, 75]. If very accurate interaction energies are needed, one might even start from CCSD-level monomer wave functions, leading to the SAPT(CCSD) [8, 82].

However, quantum chemistry methods like MPn or coupled cluster scale steeply with the number of orbitals, preventing their use within the SAPT framework beyond the small monomers. A more efficient alternative is using the KS-DFT for the monomer description, resulting in the SAPT(DFT) method [79, 83]. To achieve accurate first-order energies, it is important to perform the asymptotic correction of the xc potential to ensure the physically correct behavior [75]. For accurate induction[1] and dispersion energies, the FDDSs entering Eq. (2.71) are obtained by solving linear-response equations of the time-dependent DFT (TDDFT) [53], which is sometimes called the *coupled Kohn-Sham* (CKS) approach. Therefore, the SAPT(DFT) interaction energy can be written as

$$E_{\text{int}}^{\text{SAPT(DFT)}} = E_{\text{elst}}^{(1)}(\text{KS}) + E_{\text{exch}}^{(1)}(\text{KS}) + E_{\text{indx}}^{(2)}(\text{CKS}) + E_{\text{dispx}}^{(2)}(\text{CKS}) \,. \tag{2.81}$$

Overall, the SAPT(DFT) is a reliable practical method for computing interaction energies at an accuracy comparable to the "golden standard" CCSD(T) method [38, 75, 81]. Importantly, the SAPT(DFT) is much more accurate for interaction energies than the supermolecular DFT approach. The great advantage of SAPT(DFT) over the high-level quantum chemistry methods is its physical interpretability, as SAPT provides a decomposition of the total interaction energy onto the well-defined components instead of a single number. Last but not least, the $\mathcal{O}(N^5)$ scaling can be achieved for SAPT(DFT) by using density fitting techniques [79, 83], which is a significant speed-up over the $\mathcal{O}(N^7)$ scaling of CCSD(T); memory requirements are also much lower for SAPT(DFT) [37, 75].

In this thesis, SAPT will be mostly used as a benchmark method for the interaction energy components. In addition, the numerical analysis of the relationship between SAPT, the Tang-Toennies potentials (see below), and the universal vdW-QDO potential developed in this thesis will be discussed in Section 3.2.2.

2.2.2 Analytical Tang-Toennies Potentials

The Tang-Toennies (TT) potentials [85–87] provide a rigorous and analytically tractable framework for modeling vdW interactions. These potentials are particularly effective for noble-gas dimers, which are ideal benchmark systems due to their near-exclusive reliance on dispersion forces. The TT potentials describe interaction energies across arbitrary interatomic distances with CCSD(T)-level accuracy, seamlessly incorporating both short-range repulsion and long-range dispersion contributions.

The first widely adopted form of the TT potential was introduced in 2003 [85] and is expressed as:

$$V_{\text{TT}}(R) = Ae^{-bR} - \sum_{n=3}^{5} f_{2n}^{\text{TT}}(R)\frac{C_{2n}}{R^{2n}} \,. \tag{2.82}$$

Here, the exponential term models short-range repulsion, while the summation over n captures the long-range dispersion interactions by including the leading-order coefficients C_6, C_8, and C_{10}. These coefficients can be obtained from accurate many-body perturbation theory (MBPT) calculations [88, 89]. The parameters A and b are determined by the minimum conditions:

$$V_{\text{TT}}(R_e) = D_e \,, \quad \left.\frac{dV_{\text{TT}}}{dR}\right|_{R=R_e} = 0 \,, \tag{2.83}$$

[1] The induction energy can be expressed via FDDSs evaluated at zero frequency [9, 84].

where R_e and D_e are the equilibrium distance and binding energy of the dimer, typically derived from high-level reference methods like CCSD(T).

The Tang-Toennies damping function $f_{2n}^{\mathrm{TT}}(R)$ captures quenching of vdW dispersion interactions at short- and mid-ranges:

$$f_{2n}^{\mathrm{TT}}(z) = 1 - e^{-z} \sum_{k=0}^{n} \frac{z^k}{k!}\,, \quad z = bR\,. \tag{2.84}$$

This function is derived consistently within the TT model using asymptotic analysis [90]. In this thesis, a similar approach is adopted to derive a damping function within the quantum Drude oscillator (QDO) model (see Chapter 3 and Appendix A1).

Recently, the TT potential was extended to arbitrary interatomic distances, including the united-atom limit, and made conformal, meaning it has a fixed functional shape for all noble-gas dimers. This generalized form, known as the conformal Tang-Toennies-Sheng potential (TTS), is defined as [86]:

$$V_{\mathrm{TTS}}(R) = \frac{Z_A Z_B}{R} U_{\mathrm{short}}(x) + D_e U_{\mathrm{long}}(x)\,, \tag{2.85}$$

where Z_A, Z_B are nuclear charges, and $x = R/R_e$ is the reduced coordinate. The reduced dimensionless potentials $U_{\mathrm{short}}(x)$ and $U_{\mathrm{long}}(x)$ are given by:

$$\begin{aligned} U_{\mathrm{short}}(x) &= \frac{1}{x}\left(1 + a_1^* x + a_2^* x^2 + a_3^* x^3\right) e^{-\alpha^* x}\,; \\ U_{\mathrm{long}}(x) &= \left(1 - e^{-\alpha^* x}\right)\left[A^* x^{\gamma^*} e^{\beta^* x} - \sum_{n=3}^{5} f_{2n}^{\mathrm{TT}}(b^* x)\frac{C_{2n}^*}{x^{2n}}\right]\,. \end{aligned} \tag{2.86}$$

The short-range part, $U_{\mathrm{short}}(x)$, models approaching the united-atom limit and was first proposed by Buckingham [91]. The long-range component, $U_{\mathrm{long}}(x)$, is derived from the generalized Heitler-London theory [92], which provides a perturbative approach to interaction energies, which is an alternative to SAPT.

The Tang-Toennies potentials accurately describe binding curves for noble gases and group II dimers [85, 86, 93]. The TTS potential uses nine independent parameters (excluding nuclear charges), while the original TT potential in Eq. (2.82) requires five parameters. This complex parameterization, along with the need for highly accurate interaction coefficients and dimer parameters, limits the applicability of the TT potentials across the periodic table.

In this thesis, the TT potentials serve as an essential starting point and benchmark for the analytical description of vdW interactions. Inspired by their principles, including the balance between exchange repulsion and dispersion attraction, as well as the concept of conformality in the TTS variant, Chapter 3 introduces the universal vdW-QDO potentials. These potentials aim to achieve TT-level accuracy while maintaining a minimal two-parameter form applicable across the periodic table.

2.2.3 Quantum Drude Oscillator Model

The quantum Drude oscillator (QDO) model plays a pivotal role in this thesis as a framework for accurately describing the electronic response and noncovalent interactions in diverse systems. By representing coarse-grained charge density fluctuations of weakly-bound (valence) electrons, the QDO model provides an efficient and versatile approach to studying response

properties and vdW interactions. With suitable parametrization, it has been successfully applied to atoms, molecules of varying sizes, biomolecular systems, solids, nanostructures, and hybrid organic/inorganic interfaces [25, 40, 94–103].

This model delivers an accurate description (within a few percent) of polarization and dispersion interactions [25, 99, 100]. Moreover, it captures electron density redistribution induced by such interactions [25, 104] and provides a robust framework for describing vdW interactions influenced by external charges and spatial confinement [26, 105, 106]. Notably, the QDO model paves the way for obtaining full binding energy curves of vdW-bonded dimers [107] through an appropriate generalization of the fermionic Pauli exchange to this coarse-grained representation [102, 103, 108].

Drudons and QDO Parameters

In the QDO model, the electronic response is modeled using a quasiparticle called a *drudon*, which represents the collective behavior of valence electrons within an atom or molecule. The Hamiltonian for a single, unperturbed QDO is[2]

$$\hat{H}_0 = -\frac{\hbar^2}{2m}\nabla^2_{\mathbf{r}} + \frac{1}{2}m\omega^2(\mathbf{r}-\mathbf{R}_A)^2 \,, \tag{2.87}$$

where m is the drudon's mass and ω is the oscillator frequency. Each drudon is assigned an electric charge $(-q)$ and is harmonically bound to a positively charged pseudo-nucleus $(+q)$. Importantly, the drudon and its pseudo-nucleus do not interact via direct electromagnetic force. Their interaction is described solely by a harmonic bond representing time-dependent fluctuations of the charge density from its static equilibrium configuration. Thus, a QDO is fully characterized by its three parameters $\{q, m, \omega\}$.

Polarization Interactions

Multipole polarizabilities (α_l) are among the key quantities describing the electronic response of atoms and molecules. Within the QDO model, these can be derived by considering the perturbation caused by a test charge placed at a large distance from the QDO [99]. The resulting expressions are

$$\alpha_l = \left(\frac{q^2}{m\omega^2}\right)\left[\frac{(2l-1)!!}{l}\right]\left(\frac{\hbar}{2m\omega}\right)^{l-1} . \tag{2.88}$$

The dipole, quadrupole, and octupole polarizabilities can be recursively expressed via the QDO parameters as [99]

$$\alpha_1 = \frac{q^2}{m\omega^2} \,,\; \alpha_2 = \frac{3\hbar}{4m\omega}\alpha_1 \,,\; \alpha_3 = \frac{5\hbar^2}{4(m\omega)^2}\alpha_1 \,. \tag{2.89}$$

As it follows from Eq. (2.88), in the classical limit $\hbar \to 0$ only dipole polarizability survives, highlighting the quantum nature of higher-order polarizabilities [99, 109]. A useful invariant derived from these polarizabilities is given by [99]:

$$\gamma_{\mathrm{pol}} = \sqrt{\frac{20}{9}}\frac{\alpha_2}{\sqrt{\alpha_1\alpha_3}} = 1 \,. \tag{2.90}$$

Comparing this to the experimental ratio γ_{pol} for atoms and small molecules reveals a remarkable agreement [99], witnessing the versatility of the QDO model for computing higher-order polarizabilities. The agreement is especially good for noble gases and hydrogen, while for alkali elements the errors can reach 20% [99].

[2] The formulas in this section are written in atomic units, while explicitly keeping the Planck constant $\hbar$.

vdW Dispersion Interactions

The QDO model also facilitates closed-form calculations of dispersion coefficients, first demonstrated by London in his foundational work on dispersion forces [6]. For homonuclear dimers, the leading-order dispersion coefficients are [99]

$$C_6 = \frac{3}{4}\hbar\omega\alpha_1^2\,,\; C_8 = \frac{5\hbar}{m\omega}C_6\,,\; C_{10} = \frac{245\hbar^2}{8(m\omega)^2}C_6\,, \tag{2.91}$$

which can also be expressed in terms of polarizabilities:

$$C_8 = 5\alpha_1\alpha_2\hbar\omega\,,\; C_{10} = \left(\frac{21}{2}\alpha_1\alpha_3 + \frac{35}{2}\alpha_2\alpha_2\right)\hbar\omega\,. \tag{2.92}$$

These scaling laws for dispersion coefficients are well applicable to real atoms and molecules [25, 99, 110]. In parallel with Eq. (2.90), the following invariant can be derived for dispersion coefficients [99]:

$$\gamma_{\text{disp}} = \sqrt{\frac{49}{40}}\frac{C_8}{\sqrt{C_6 C_{10}}} = 1\,. \tag{2.93}$$

The model's predictive power extends also to mixed polarization-dispersion invariant [99]:

$$\gamma_{\text{3disp}} = \frac{\alpha_1 C_6}{4C_9} = 1\,, \tag{2.94}$$

where $C_9 = \alpha_1 C_6/4$ is the three-body dispersion coefficient (Axilrod-Teller-Muto term (2.119)) in the QDO model [99].

The London formula [6] $C_6 = \frac{3}{4}\hbar\omega\alpha_1^2$ combined with free-atom reference data is used to compute molecular C_6 dispersion coefficients from ground-state electron densities with 5.5% accuracy compared to reference vdW coefficients derived from experimental dipole-oscillator measurements in the Tkatchenko-Scheffler method [39] (Section 2.3.2). The dipole-coupled QDO framework further led to the development of the MBD method [40], which is discussed in Section 2.3.4. Hence, the QDO model is pivotal in developing accurate vdW methodologies.

Electrostatics

The first-order perturbation energy between two QDOs interacting via Coulomb interaction is [100]:

$$E_{\text{elst}} = q^2\left(\frac{1}{R} + \frac{\text{erf}(R/2\sigma)}{R} - \frac{2\text{erf}(R/\sigma\sqrt{2})}{R}\right), \tag{2.95}$$

where $\sigma = \sqrt{\hbar/2m\omega}$. However, directly applying Eq. (2.95) results in too strong interactions, since the QDO length-scale σ is fitted to response properties, rather than atomic size [100]. The appropriate QDO reparametrization is desired to achieve a better description of electrostatic interaction energy in the QDO model. Nevertheless, the existing parametrizations deliver accurate electrostatic *response*, such as the atomic polarization potential.

Moreover, the QDO model provides a useful framework to study e.g. field-induced dispersion effects [105] or Coulomb interactions between dipolar fluctuations [26], which provide an example of coupling between dispersion and polarization (induction) interactions. Such effects can significantly contribute to the interaction energy for large polarizable systems [26] and lead to a counterintuitive dispersion repulsion under confinement [26, 106].

Exchange-repulsion

The original QDO model [96, 97, 99] did not account for exchange-repulsion effects due to two main limitations. First, the coarse-grained wave function of a single Drude particle represents all valence electrons in an atom collectively, preventing a straightforward Pauli exchange treatment for individual electron pairs [92]. Second, Drude particles were initially modeled as distinguishable, each tied to a specific nucleus [99]. Recently, it has been shown [102, 103] that the exchange-repulsion between identical QDOs can be introduced via the Heitler-London approach [111] yielding the coarse-grained model for exchange-repulsion between real atoms. We refer to Chapter 3 for a detailed discussion of this matter.

QDO Parametrization

Closed-form expressions for the QDO response properties (2.88) and (2.91) can be inverted to set the three parameters $\{q, m, \omega\}$, exactly reproducing these properties. While there are six response properties in Eqs. (2.88) and (2.91), any three of them can be used to solve for $\{q, m, \omega\}$. For instance, Jones *et al.* [99] proposed using $\{\alpha_1, C_6, C_8\}$ to parameterize the QDO (JQDO scheme) as

$$q = \sqrt{m\omega^2\alpha_1}\ ,\ \omega = 4C_6/3\hbar\alpha_1^2\ ,\ m = 5\hbar C_6/\omega C_8\ . \tag{2.96}$$

Alternatively, when multipole polarizabilities are prioritized, α_1, α_2, C_6 can be used, leading to slightly different values for m but identical ω and q [109]:

$$q = \sqrt{m\omega^2\alpha_1}\ ,\ \omega = 4C_6/3\hbar\alpha_1^2\ ,\ m = 3\hbar\alpha_1/4\omega\alpha_2\ . \tag{2.97}$$

Another approach is the fixed-charge QDO (FQDO) parametrization, where always $q = 1$ a.u., assuming that drudons have the same charge as electrons. The FQDO parameters read

$$q = 1\ ,\ \ \omega = 4C_6/3\hbar\alpha_1^2\ ,\ \ m = 9\hbar^2\alpha_1^3/16C_6^2\quad . \tag{2.98}$$

The FQDO scheme is implicitly used in the many-body dispersion (MBD) model [40], where QDO masses are defined as $m = 1/\alpha_1\omega^2$ aligning with setting $q = 1$ a.u. in $\alpha_1 = q^2/m\omega^2$.

Recently, it has been shown that properly tuned QDO parametrization is crucial for achieving accurate results for non-parametrized observables, such as the atomic potential response to external electric field [110]. For molecular simulations, the QDO parametrization has to rely on atom-in-molecule (AIM) approaches, which extract AIM parameters (polarizabilities and dispersion coefficients) from electronic-structure theory. Please see Section 2.3.2 and Appendix A2 for a more detailed discussion.

2.3 van der Waals Dispersion Interactions

vdW dispersion interactions are crucial for a wide range of physical and chemical properties, including binding energies, structural and dynamical stability of proteins and molecular crystals, as well as mechanical, electronic, and optical characteristics. Originating from long-range electronic correlations, these interactions are inherently *collective*, *non-additive*, and *ubiquitous* in nature. This section provides an overview of theoretical and practical approaches to modeling vdW dispersion within the framework of the adiabatic-connection fluctuation-dissipation theorem (ACFD), following the notation and concepts established in Refs. [11, 12].

As discussed in Section 2.1.4, the ACFD theorem provides an exact expression for the electronic correlation energy, with vdW dispersion dominating the long-range (dynamic) correlation contribution. The ACFD formalism is particularly suitable for describing vdW interactions in the long-range limit, supported by practical approximations such as the random phase approximation (RPA). Many established computational approaches to vdW dispersion can be viewed as approximations of the ACFD-RPA energy expression, forming the conceptual basis of this section.

A remark has to be made about the terminology here. In the field of intermolecular forces, "van der Waals forces" historically encompass Keesom (orientation-averaged permanent multipoles), Debye (permanent–induced multipoles), and dispersion (induced–induced multipoles) contributions [7]. In chemistry, nonbonded Pauli repulsion is often included under vdW interactions to contrast them with covalent bonds, while in solid-state physics, "vdW interactions" are synonymous with dispersion interactions. Classical force fields usually represent vdW interactions through a Lennard-Jones (LJ) potential, combining repulsive and attractive contributions. In this thesis, we will use the term *"vdW potential"* to denote such interatomic potentials. This terminology is also consistent with the vdW equation of state, which accounts for both repulsion and attraction. The names *"vdW dispersion interactions"* and *"dispersion energy"* refer here specifically to the dispersion-only contribution to avoid ambiguity.

Relativistic retardation and finite temperature effects become relevant in mesoscopic systems (e.g., Casimir forces) but are beyond the scope of this thesis. For further insights, the reader is referred to Refs. [112–114].

2.3.1 ACFD Energy Expressed via Nonlocal Dipole Polarizability

The response of charge density to the frequency-dependent electric field $\mathscr{E}(\mathbf{r},\omega) = -\nabla\phi(\mathbf{r},\omega)$ can be described via the density response function:

$$\Delta\rho(\mathbf{r},\omega) = \int d\mathbf{r}'\,\chi(\mathbf{r},\mathbf{r}',\omega)\phi(\mathbf{r}',\omega)\,. \tag{2.99}$$

The discussion of the ACFDT in Section 2.1.4 was based on $\chi(\mathbf{r},\mathbf{r}',\omega)$ as the central quantity. For the description of approximate vdW dispersion models, it is more convenient to use an equivalent [115] reformulation of the ACFDT formalism in terms of nonlocal dipole polarizability [11]. The nonlocal dynamic polarizability $\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}',\omega)$ is related to the (anisotropic) point polarizability of a system $\boldsymbol{\alpha}(\omega)$ by a simple normalization condition $\boldsymbol{\alpha}(\omega) = \iint d\mathbf{r}d\mathbf{r}'\,\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}',\omega)$. The link between $\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}',i\omega)$ and density response function is given by [11, 12]:

$$\chi(\mathbf{r},\mathbf{r}',i\omega) = \nabla_{\mathbf{r}}\cdot\nabla_{\mathbf{r}}\cdot\boldsymbol{\alpha}(\mathbf{r},\mathbf{r}',i\omega) = \sum_{ij}\frac{\partial}{\partial r_i}\frac{\partial}{\partial r'_j}\alpha_{ij}(\mathbf{r},\mathbf{r}',i\omega)\,. \tag{2.100}$$

By introducing the dipole-coupling tensor

$$\mathbf{T}(\mathbf{r},\mathbf{r}') = -\nabla_{\mathbf{r}}\otimes\nabla_{\mathbf{r}'}v(\mathbf{r},\mathbf{r}')\,, \tag{2.101}$$

the ACFD expression (2.40) can be reformulated as:

$$E_{\mathrm{c}} = \frac{1}{2\pi}\int_0^\infty d\omega\int_0^1 d\lambda\iint d\mathbf{r}d\mathbf{r}'\,\mathrm{Tr}\Big\{\big[\boldsymbol{\alpha}_\lambda(\mathbf{r},\mathbf{r}',i\omega)-\boldsymbol{\alpha}_{\lambda=0}(\mathbf{r},\mathbf{r}',i\omega)\big]\,\mathbf{T}(\mathbf{r},\mathbf{r}')\Big\}\,, \tag{2.102}$$

where Tr{·} denotes the trace over the Cartesian components of the tensor [11, 12]. The nonlocal polarizability of the interacting system ($\boldsymbol{\alpha}_\lambda$) can be computed from the polarizability of

non-interacting KS system ($\boldsymbol{\alpha}_{\lambda=0}$) using the self-consistent Dyson equation in random phase approximation, analogous to Eq. (2.42):

$$\begin{aligned}\boldsymbol{\alpha}_\lambda(\mathbf{r},\mathbf{r}',i\omega) = \boldsymbol{\alpha}_0(\mathbf{r},\mathbf{r}',i\omega) - \iint d\mathbf{r}''d\mathbf{r}'''\ \boldsymbol{\alpha}_0(\mathbf{r},\mathbf{r}'',i\omega)\lambda\mathbf{T}(\mathbf{r}'',\mathbf{r}''')\boldsymbol{\alpha}_\lambda(\mathbf{r}''',\mathbf{r}',i\omega) = \\ = \boldsymbol{\alpha}_0 - \langle\lambda\boldsymbol{\alpha}_0\mathbf{T}\boldsymbol{\alpha}_\lambda\rangle = \sum_{n=0}^{\infty}\langle\boldsymbol{\alpha}_0(-\lambda\mathbf{T}\boldsymbol{\alpha}_0)^n\rangle\,,\end{aligned} \tag{2.103}$$

where the variable dependence of $\boldsymbol{\alpha}$ and $\mathbf{T}$ is omitted for brevity, and the shorthand notation $\langle\cdot\rangle$ denotes the implicit integrations over the spatial coordinates. For coupling with semi-local or hybrid density functionals in practical calculations, range separation of dipole tensor is often introduced, similar to Eq. (2.43):

$$\mathbf{T}(\mathbf{r},\mathbf{r}') = \big(1-f(|\mathbf{r}-\mathbf{r}'|)\big)\mathbf{T}(\mathbf{r},\mathbf{r}') + f(|\mathbf{r}-\mathbf{r}'|)\mathbf{T}(\mathbf{r},\mathbf{r}') = \mathbf{T}_{\mathrm{sr}}(\mathbf{r},\mathbf{r}') + \mathbf{T}_{\mathrm{lr}}(\mathbf{r},\mathbf{r}')\,. \tag{2.104}$$

Plugging this into Eq. (2.103) decomposes the n-th term in the Dyson equation into 2^n terms. Each of these terms is formed by a specific combination of short-range ($\mathbf{T}_{\mathrm{sr}}$) and long-range ($\mathbf{T}_{\mathrm{lr}}$) interaction components. To simplify the expression, we identify and contract all short-range interaction segments of the form $\ldots\boldsymbol{\alpha}_0\mathbf{T}_{\mathrm{sr}}\boldsymbol{\alpha}_0\ldots$ within these terms. This leads us to define an effective short-range screened polarizability, $\boldsymbol{\alpha}_{\mathrm{sr}}$, which encapsulates the effect of repeated short-range interactions. After this manipulation, the Dyson equation takes the following compact form:

$$\boldsymbol{\alpha}_\lambda = \sum_{n=0}^{\infty}\left\langle\boldsymbol{\alpha}_{\mathrm{sr}}(-\lambda\mathbf{T}_{\mathrm{lr}}\boldsymbol{\alpha}_{\mathrm{sr}})^n\right\rangle. \tag{2.105}$$

The equivalence of Eqs. (2.105) and (2.103) can be verified directly. Essentially, Eq. (2.105) is just a reorganization of the terms in the sum from (2.103).

Finally, combining Eq. (2.105) and the long-range part of the ACFD expression (2.102) gives:

$$E_{\mathrm{c}}^{(\mathrm{lr,RPA})} = \frac{1}{2\pi}\sum_{n=1}^{\infty}\int_0^\infty d\omega\int_0^1 d\lambda\iint d\mathbf{r}d\mathbf{r}'\ \mathrm{Tr}\left\{\left\langle\boldsymbol{\alpha}_{\mathrm{sr}}(-\lambda\mathbf{T}_{\mathrm{lr}}\boldsymbol{\alpha}_{\mathrm{sr}})^n\right\rangle\mathbf{T}_{\mathrm{lr}}(\mathbf{r},\mathbf{r}')\right\}, \tag{2.106}$$

where $n=0$ term in the sum cancels out.[3] Since within the applied random phase approximation the coupling tensor $\mathbf{T}_{\mathrm{lr}}$ does not depend on λ, we can integrate over the coupling constant analytically, which yields

$$E_{\mathrm{c}}^{(\mathrm{lr,RPA})} = -\frac{1}{2\pi}\sum_{n=2}^{\infty}\frac{(-1)^n}{n}\int_0^\infty d\omega\iint d\mathbf{r}d\mathbf{r}'\,\mathrm{Tr}\left\{\left\langle(\boldsymbol{\alpha}_{\mathrm{sr}}\mathbf{T}_{\mathrm{lr}})^n\right\rangle(\mathbf{r},\mathbf{r}',i\omega)\right\}. \tag{2.107}$$

Note that the index n has been shifted by 1 due to the integration, which also leads to the minus sign in front of the sum. To clarify the introduced shorthand notations, we write out the second-order term $E_{\mathrm{c,lr}}^{(2)}$ in a fully explicit form [11]:

$$E_{\mathrm{c,lr}}^{(2)} = -\frac{1}{2}\int_0^\infty\frac{d\omega}{2\pi}\iiiint d\mathbf{r}d\mathbf{r}'d\mathbf{r}''d\mathbf{r}'''\sum_{ijkl}\alpha_{ij}^{\mathrm{sr}}(\mathbf{r},\mathbf{r}',i\omega)\,T_{jk}^{\mathrm{lr}}(\mathbf{r}',\mathbf{r}'')\alpha_{kl}^{\mathrm{sr}}(\mathbf{r}'',\mathbf{r}''',i\omega)\,T_{li}^{\mathrm{lr}}(\mathbf{r}''',\mathbf{r})\,. \tag{2.108}$$

The expression (2.107), exact in the long-range limit, provides a starting point for all approximate vdW models, discussed in what follows and pictorially summarized in Figure 2.2. The

[3] In Eq. (2.106), we corrected a typo in Eq. (8) of Ref. [12]: there is an erroneous minus sign in front of the sum and a missing minus sign in front of λ.

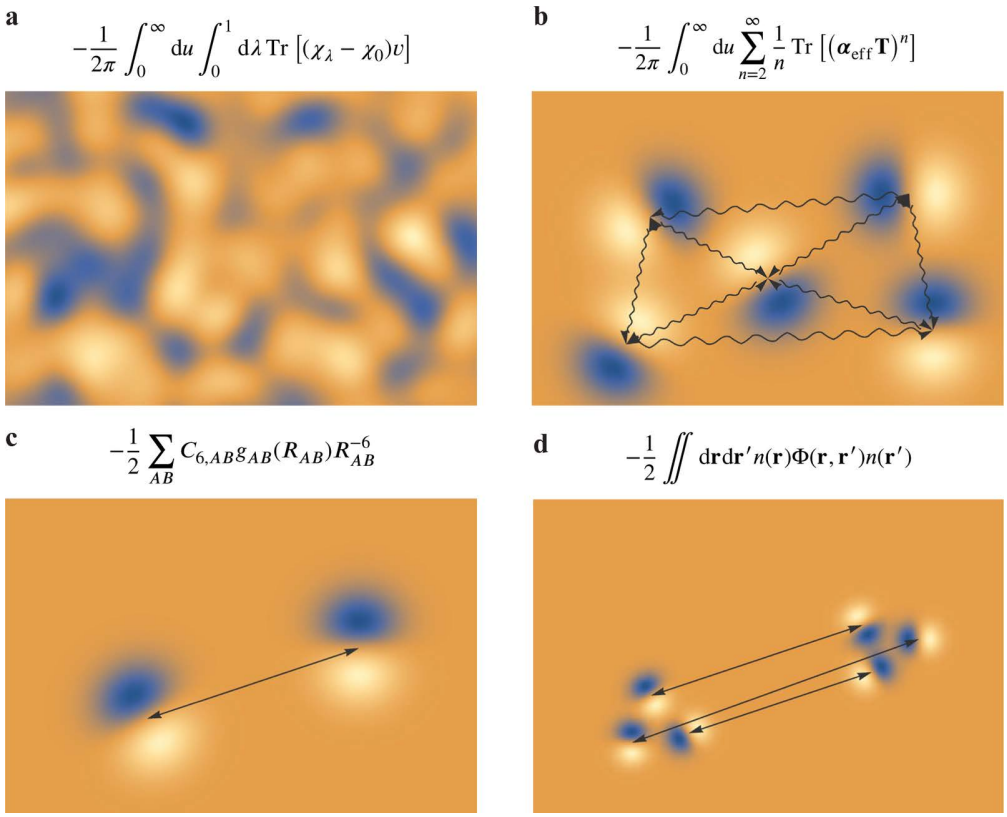

Figure 2.2: Theoretical approaches behind the first-principles vdW modeling. **(a)** Exact treatment of correlation energy within the ACFD framework, using the full nonlocal response function $\chi(\mathbf{r},\mathbf{r}',i\omega)$, which describes the propagation of fluctuations in the system. **(b)** Fragment-based many-body methods (e.g., MBD) correlate fluctuations to infinite order through model response functions in the ACFD formula. **(c)** Coarse-grained two-body methods (e.g., TS, XDM, DFT-D) compute the pairwise vdW dispersion energy by correlating only two dipoles (or multipoles) at a time. **(d)** Long-range nonlocal correlation functionals use a continuous representation of dipolar fluctuations, truncated at second-order in coupling. Reproduced with permission from Ref. [11]. © 2017 American Chemical Society.

primary rationale for dividing the problem into short- and long-range components lies in the tractability of the dipolar coupling $\mathbf{T}$ at long range, while the polarizability $\boldsymbol{\alpha}$ can be effectively modeled using the electron density at short range, as will be demonstrated in subsequent sections.

2.3.2 Pairwise-Additive vdW Approaches

One of the most common strategies for modeling vdW interactions relies on pairwise-additive potentials. Despite their widespread use, these models inherently neglect the many-body, non-additive nature of dispersion forces – a property that has been firmly established through extensive experimental and theoretical studies [11, 12, 77, 116–122]. This simplification persists largely due to the computational efficiency and practicality of pairwise approaches.

At their core, pairwise potentials for long-range correlation forces can be derived from the long-range ACFD formula (2.107). It's worth noting that this functional form can also emerge from alternative theoretical approaches, such as (many-body) perturbation theory (see Section 2.2.1) or other approximations to the ACFD expression. In this section, we outline the key approximations and derivation steps that lead to this fundamental form, providing insight into the origins and limitations of the pairwise perspective on vdW interactions.

The first step is to approximate the nonlocal polarizability as the sum of local atomic polarizabilities at positions $\{\mathbf{R}_A\}$:

$$\boldsymbol{\alpha}_{\mathrm{sr}}(\mathbf{r},\mathbf{r}',i\omega) \approx \sum_{A=1}^{N} \boldsymbol{\alpha}_A^{\mathrm{sr}}(i\omega)\delta(\mathbf{r}-\mathbf{R}_A)\delta(\mathbf{r}-\mathbf{r}') \equiv \sum_{A=1}^{N} \boldsymbol{\alpha}_A^{\mathrm{sr}} \,. \tag{2.109}$$

Inserting this into Eq. (2.107) yields:

$$E_{\mathrm{c}}^{(\mathrm{lr,RPA})} = -\frac{1}{2\pi}\sum_{n=2}^{\infty}\frac{(-1)^n}{n}\int_0^{\infty} d\omega \iint d\mathbf{r}d\mathbf{r}' \,\mathrm{Tr}\left\{\left\langle\left(\sum_{A=1}^{N}\boldsymbol{\alpha}_A^{\mathrm{sr}}\mathbf{T}_{\mathrm{lr}}\right)^n\right\rangle\right\} \tag{2.110}$$

For $n=2$, one obtains $\boldsymbol{\alpha}_A^{\mathrm{sr}}\mathbf{T}_{\mathrm{lr}}\boldsymbol{\alpha}_B^{\mathrm{sr}}\mathbf{T}_{\mathrm{lr}}$, for $n=3$, the term is $\boldsymbol{\alpha}_A^{\mathrm{sr}}\mathbf{T}_{\mathrm{lr}}\boldsymbol{\alpha}_B^{\mathrm{sr}}\mathbf{T}_{\mathrm{lr}}\boldsymbol{\alpha}_C^{\mathrm{sr}}\mathbf{T}_{\mathrm{lr}}$, etc. In this context, the order n does not strictly correspond to a pure n-body vdW interaction term as defined in a perturbative framework. For instance, when $n=3$, the terms include contributions where $C=A$, representing screened two-body interactions. Conversely, for $n=2$, the non-zero contributions involve only two distinct polarizability centers, A and B, making it a pure (but still incomplete) two-body vdW interaction. Focusing on the second-order term, $E_{\mathrm{c}}^{(2)}$, we can perform the spatial integration to derive:

$$E_{\mathrm{c}}^{(2)} = -\frac{1}{4\pi}\int_0^{\infty} d\omega \,\mathrm{Tr}\left\{\sum_{A,B}\boldsymbol{\alpha}_A^{\mathrm{sr}}\mathbf{T}_{AB}^{\mathrm{lr}}\boldsymbol{\alpha}_B^{\mathrm{sr}}\mathbf{T}_{BA}^{\mathrm{lr}}\right\}, \tag{2.111}$$

where $\mathbf{T}_{AB}^{\mathrm{lr}} = \mathbf{T}_{\mathrm{lr}}(\mathbf{R}_A,\mathbf{R}_B)$. Finally, if the polarizabilities are assumed to be isotropic, $\boldsymbol{\alpha}_A^{\mathrm{sr}} = \mathbb{1}\alpha_A^{\mathrm{sr}}$, they commute with dipole tensors and

$$E_{\mathrm{c}}^{(2)} = -\frac{1}{2}\sum_{A,B}\frac{3}{\pi}\int_0^{\infty}\alpha_A^{\mathrm{sr}}\alpha_B^{\mathrm{sr}}\,d\omega\,\frac{1}{6}\mathrm{Tr}\left\{\mathbf{T}_{AB}^{\mathrm{lr}}\mathbf{T}_{BA}^{\mathrm{lr}}\right\} = -\frac{1}{2}\sum_{A,B}C_6^{AB}\frac{1}{6}\mathrm{Tr}\left\{\mathbf{T}_{AB}^{\mathrm{lr}}\mathbf{T}_{BA}^{\mathrm{lr}}\right\}, \tag{2.112}$$

where we factorized the Casimir-Polder integral for the C_6 dispersion coefficient [123]. Recognizing that $\mathbf{T}_{AB}^{\mathrm{lr}} = f(|\mathbf{R}_A-\mathbf{R}_B|)\,\mathbf{T}_{AB}$ and taking the trace $\mathrm{Tr}\left\{\mathbf{T}_{AB}\mathbf{T}_{BA}\right\} = 6/|\mathbf{R}_A-\mathbf{R}_B|^6 = 6/R_{AB}^6$, one obtains:

$$E_{\mathrm{c}}^{(2)} = -\frac{1}{2}\sum_{A,B} f_{\mathrm{damp}}(R_{AB})\frac{C_6^{AB}}{R_{AB}^6}, \tag{2.113}$$

where we defined the *damping function* $f_{\mathrm{damp}}(R_{AB}) = f^2(|\mathbf{R}_A-\mathbf{R}_B|)$. In the asymptotic limit $R_{AB}\to\infty$, this formula reduces to the celebrated C_6/R^6 expression for vdW dispersion derived by London [6].

Note that here we considered only the leading dipole-dipole term emerging from a coarse-grained (atom-centered) multipole expansion of the nonlocal $\boldsymbol{\alpha}^{\mathrm{sr}}(\mathbf{r},\mathbf{r}')$ and $\mathbf{T}_{\mathrm{lr}}(\mathbf{r},\mathbf{r}')$ [11, 124]. In principle, the higher-order multipoles can be also considered, which would lead to the analogous C_8/R^8 and C_{10}/R^{10} pairwise terms [11, 124]. It is important to emphasize that in the *fine-grained*, *nonlocal* picture dipole polarizability $\boldsymbol{\alpha}^{\mathrm{sr}}(\mathbf{r},\mathbf{r}')$ and dipole coupling tensor $\mathbf{T}_{\mathrm{lr}}(\mathbf{r},\mathbf{r}')$ are sufficient to provide the *exact* description of the long-range correlation energy as demonstrated by Eq. (2.107), while the inclusion of higher-order multipole counterparts is required once *coarse-graining* is performed. Now we will briefly discuss the most common practical methods based on the energy expressions like (2.113)

Semiempirical DFT-D Methods

One of the earliest and most widely adopted pairwise van der Waals (vdW) approaches is the DFT-D family of methods developed by Grimme and collaborators [19]. The initial DFT-D1 model [125] employed tabulated C_6 coefficients for organic elements along with empirical scaling factors. However, as the atomic C_6 coefficients were averaged over possible hybridization states, this approach led to significant errors and limited transferability.

The subsequent DFT-D2 iteration [126] improved upon this by calculating atomic C_6 coefficients (in J nm^6 mol^{-1}) are calculated based on atomic ionization potentials I (in a.u.) and static polarizabilities α (in a.u.) using the empirical formula $C_6 = 0.05NI\alpha$, with $N = 2, 10, 18, 36, 54$ for atoms from rows 1-5 of the periodic table, correspondingly. This extension increased the periodic table coverage, but the dispersion coefficients remained environment-independent. Additionally, the model inherited incorrect asymptotic behavior from DFT-D1 due to the damping function, which renders both methods largely obsolete by modern standards.

The next-generation DFT-D3 method [127] introduced several significant improvements. First, the concept of an effective coordination number (CN) was introduced as

$$\mathrm{CN}^A = \sum_{B\neq A} \left\{ 1 + \exp\left[-k_1 \left(k_2 \frac{R_A^{\mathrm{cov}} + R_B^{\mathrm{cov}}}{R_{AB}} - 1 \right) \right] \right\}^{-1} , \tag{2.114}$$

where $k_1 = 16$ and $k_2 = 4/3$ are fitted to a set of organic molecules, and R_A^{cov} is a (scaled) covalent radius of atom A. CN enables incorporating the dependence of C_6 coefficients on the local chemical environment via

$$C_6^{AB}(\mathrm{CN}^A, \mathrm{CN}^B) = \frac{1}{W} \sum_i^{N_A} \sum_j^{N_B} C_{6,\mathrm{ref}}^{AB}(\mathrm{CN}_i^A, \mathrm{CN}_j^B) L_{ij} , \tag{2.115}$$

where $W = \sum_{i,j} L_{ij}$ and

$$L_{ij} = e^{-k_3 \left[\left(\mathrm{CN}^A - \mathrm{CN}_i^A\right)^2 + \left(\mathrm{CN}^B - \mathrm{CN}_j^B\right)^2 \right]} \tag{2.116}$$

is the Gaussian distance measure between the coordination numbers of A, B in the system of interest and in the reference systems i and j, for which the reference $C_{6,\mathrm{ref}}^{AB}(\mathrm{CN}_i^A, \mathrm{CN}_j^B)$ values have been precomputed. Reference systems were selected as hydrides of elements A and B, and $C_{6,\mathrm{ref}}^{AB}(\mathrm{CN}_i^A, \mathrm{CN}_j^B)$ were obtained by substracting the contributions of hydrogens. The global parameter was $k_3 = 4$ empirically determined to ensure smooth behavior across integer coordination numbers [127].

Second, the correct asymptotic behavior of the damping function was enforced as $R \to \infty$. Third, C_8 dispersion coefficients were computed recursively from C_6 and included in the energy expression, which ultimately has a form [127]:

$$E_{\mathrm{D3}}^{(2)} = -\frac{1}{2} \sum_{A\neq B} \left(f_{d,6}(R_{AB}) \frac{C_6^{AB}}{R_{AB}^6} + f_{d,8}(R_{AB}) \frac{C_8^{AB}}{R_{AB}^8} \right) . \tag{2.117}$$

Originally, the Chai–Head-Gordon damping functions [128] were applied,

$$f_{d,n}(R_{AB}) = \frac{s_n}{1 + 6\left(R_{AB}/s_{r,n}R_0^{AB}\right)^{-\alpha_n}} , \quad n = 6, 8 , \tag{2.118}$$

with $s_{r,8} = 1$, $\alpha_6 = 14$ and $\alpha_8 = 16$, while s_8 and $s_{r,6}$ are DFA-dependent scaling factors. Also, fixing $s_6 = 1$ ensures the correct asymptotic behavior of the damping function. R_0^{AB} is the "cut-off" radius, defined as a distance at which the so-called first-order KS-DFT interaction energy

of a given pair equals arbitrarily specified threshold [127]. Later, the DFT-D3 method was also coupled with the Becke-Johnson (BJ) damping function (*vide infra*), leading to the DFT-D3(BJ) variant [129].

Optionally, the DFT-D3 method can be extended to include the leading non-additive dispersion term due to the triple dipole interactions, known as the *Axilrod-Teller-Muto* (ATM) term [116, 117]:

$$E_{\mathrm{ATM}}^{(3)} = -\sum_{ABC} \frac{C_9^{ABC}}{(R_{AB}R_{BC}R_{CA})^3}(3\cos\theta_A\cos\theta_B\cos\theta_C + 1)\,, \tag{2.119}$$

where θ_A, θ_B, and θ_C are the internal angles of the triangle formed by atoms A, B, and C. In general, the three-body dipolar dispersion coefficient C_9^{ABC} is given by the generalization of the Casimir-Polder integral,

$$C_9^{ABC} = \frac{3}{\pi}\int_0^\infty d\omega\ \alpha_A(i\omega)\alpha_B(i\omega)\alpha_C(i\omega)\,, \tag{2.120}$$

but within the DFT-D3 framework it is approximated as the geometric mean [127]:

$$C_9^{ABC} \approx \sqrt{C_6^{AB}C_6^{BC}C_6^{AC}}\,. \tag{2.121}$$

The C_9 term is also damped using the damping function (2.118) with $\alpha = 16$, $s_r = 4/3$ and geometrically averaged cutoff radius $\bar{R}_0^{ABC}$ [127]. However, ambiguous definition of damping radius and intricate error cancellations between the DFA and ATM parts lead to uncertainties about whether the addition of the three-body term systematically improves vdW dispersion energy predictions [130–132].

Despite its reliance on empirical parameters and fine-tuning, DFT-D3 represented a significant step forward, particularly through its coordination-dependent C_6 coefficients. However, this approach lacks the physical rigor required for a robust description of local chemical environments. Nevertheless, it remains widely applied in computational chemistry, especially for solid-state calculations. The more advanced DFT-D4 method [133] was recently introduced, accounting for charge transfer and many-body effects. While it offers improved accuracy, its complex architecture and continued reliance on empirical parameters place its detailed theoretical discussion beyond the scope of this overview.

Exchange-Hole Dipole Moment Model

In the exchange-hole dipole moment (XDM) model [134, 135], vdW dispersion interactions are derived as the interaction of electronic multipoles between the moving electron and the exchange-hole $h_{X\sigma}$, accompanying it. Exchange-hole describes the instantaneous depletion in the probability of finding a parallel-spin electron at $\mathbf{r}'$, given a reference electron at $\mathbf{r}$:

$$h_{X\sigma}(\mathbf{r},\mathbf{r}') = -\frac{1}{\rho_\sigma(\mathbf{r})}\sum_{ij}\psi_{i\sigma}(\mathbf{r})\psi_{j\sigma}(\mathbf{r})\psi_{i\sigma}(\mathbf{r}')\psi_{j\sigma}(\mathbf{r}')\,, \tag{2.122}$$

where $\rho_\sigma(\mathbf{r})$ is the spin density, and $\psi_{i\sigma}$ are the spin orbitals (assumed to be real) of a single Slater determinant, which may be obtained from Hartree-Fock or Kohn-Sham theory. The exchange-hole integrates to -1 everywhere:

$$\int h_{X\sigma}(\mathbf{r},\mathbf{r}')d\mathbf{r}' = -1\,, \quad \forall\mathbf{r}\in\mathbb{R}^3\,, \tag{2.123}$$

which ensures the overall neutrality of the electron plus its exchange-hole. However, the exchange-hole is generally not spherically symmetric, creating a net dipole moment relative to the reference electron [134, 135]:

$$\mathbf{d}_{X\sigma}(\mathbf{r}) = \left[\frac{1}{\rho_\sigma(\mathbf{r})} \sum_{ij} \psi_{i\sigma}(\mathbf{r})\psi_{j\sigma}(\mathbf{r}) \int d\mathbf{r}'\, \mathbf{r}' \psi_{i\sigma}(\mathbf{r}')\psi_{j\sigma}(\mathbf{r}') \right] - \mathbf{r}\,. \tag{2.124}$$

Alternatively, the magnitude of the dipole $d_{X\sigma}(\mathbf{r})$ may be approximated using semi-local DFA densities and the Becke-Roussel exchange model [136], which provides similar results with lower computational cost.

By combining the exchange-hole dipole with the Unsöld approximation in second-order perturbation theory, the following expression for the C_6 dispersion coefficient is obtained [134, 135]:

$$C^{AB}_{6,\mathrm{XDM}} = \frac{\alpha_{0,A}\alpha_{0,B}\langle d_X^2\rangle_A \langle d_X^2\rangle_B}{\alpha_{0,A}\langle d_X^2\rangle_B + \alpha_{0,B}\langle d_X^2\rangle_A}\,, \tag{2.125}$$

where α_0 is the atomic static polarizability, and $\langle d_X^2\rangle$ represents the expectation value of the squared dipole magnitude over the electron density:

$$\langle d_X^2\rangle = \sum_\sigma \int \rho_\sigma(\mathbf{r}) d^2_{X\sigma}(\mathbf{r})\; d\mathbf{r}\,. \tag{2.126}$$

In a more general form, the formulae for dispersion coefficients between spherically symmetric atoms A and B are [20, 135]

$$\begin{aligned}
C^{AB}_{6,\mathrm{XDM}} &= \alpha_{0,A}\alpha_{0,B}\frac{\langle M_1^2\rangle_A\langle M_1^2\rangle_B}{\alpha_{0,A}\langle M_1^2\rangle_B + \alpha_{0,B}\langle M_1^2\rangle_A}\,,\\
C^{AB}_{8,\mathrm{XDM}} &= \frac{3}{2}\alpha_{0,A}\alpha_{0,B}\frac{\langle M_1^2\rangle_A\langle M_2^2\rangle_B + \langle M_2^2\rangle_A\langle M_1^2\rangle_B}{\alpha_{0,A}\langle M_1^2\rangle_B + \alpha_{0,B}\langle M_1^2\rangle_A}\,,\\
C^{AB}_{10,\mathrm{XDM}} &= \alpha_{0,A}\alpha_{0,B}\frac{2\langle M_1^2\rangle_A\langle M_3^2\rangle_B + 2\langle M_3^2\rangle_A\langle M_1^2\rangle_B + \frac{21}{5}\langle M_2^2\rangle_A\langle M_2^2\rangle_B}{\alpha_{0,A}\langle M_1^2\rangle_B + \alpha_{0,B}\langle M_1^2\rangle_A}\,.
\end{aligned} \tag{2.127}$$

In the above, the atomic l-pole moment M_l (its $m=0$ component) has been introduced as

$$M_l = \sqrt{4\pi}\sum_i r_i^l Y_l^0(\Omega_i) \tag{2.128}$$

where Y_l^0 is the corresponding spherical harmonic function of angular coordinates Ω_i, and the summation goes over all electrons in an atom. The squared multipole operator is defined as

$$M_l^2 = 4\pi \sum_i r_i^l Y_l^0(\Omega_i) \sum_j r_j^l Y_l^0(\Omega_j)\,, \tag{2.129}$$

and it is a two-electron function. While this does not pose any complications for isolated atoms, the two-electron nature of the operator prevents straightforward decomposition into atomic contributions, which becomes essential for molecular systems [137].

To address this issue, a simple model based on two point charges is introduced [135, 137]. This model consists of a negative unit charge (representing the reference electron) at a distance r from the atomic nucleus and a positive unit charge (representing the center of the exchange-hole) at a distance $(r - d_{X\sigma})$. This setup recovers exact dipole moment $\mathbf{d}_{X\sigma}$ of the exchange-hole

plus electron and additionally captures higher-order l-pole moments, with the magnitudes $r^l - (r - d_{X\sigma})^l$.

The *spherically averaged* integrals for the multipole moments of this two-point system are expressed as

$$\langle M_l^2 \rangle = \sum_\sigma \int \rho_\sigma(\mathbf{r}) \left[r^l - (r - d_{X\sigma})^l \right]^2 \, d\mathbf{r} \,. \tag{2.130}$$

Thus, the model allows efficient numerical evaluation of $\langle M_l^2 \rangle$ as one-electron integrals, relying only on the values of $d_{X\sigma}$, which are precomputed on a real-space grid using HF or DFT orbitals. The assumption that Eq. (2.130) provides a reasonable approximation for the expectation values of the squared multipole operator in Eq. (2.129) is a crucial step in extending the XDM method to practical calculations of higher-order dispersion coefficients [137].

For molecular systems, the Hirshfeld partitioning scheme is employed to assign atom-in-molecule (AIM) contributions:

$$\langle M_l^2 \rangle_A = \sum_\sigma \int w_A(\mathbf{r}) \rho_\sigma(\mathbf{r}) \left[r^l - (r - d_{X\sigma})^l \right]^2 \, d\mathbf{r} \,, \tag{2.131}$$

where $w_A(\mathbf{r})$ the Hirshfeld weighting factor:

$$w_A(\mathbf{r}) = \frac{\rho_A^{\text{free}}(\mathbf{r})}{\sum_B \rho_B^{\text{free}}(\mathbf{r})} \,. \tag{2.132}$$

Here, $\rho_A^{\text{free}}(\mathbf{r})$ denotes the spherically averaged charge density of the free atom A, while the denominator sums over the spherical free-atomic densities and is commonly named as a *promolecular density*.

The AIM polarizabilities are then derived by rescaling reference free-atomic values under the assumption of proportionality between atomic volumes and polarizabilities:

$$\alpha_{0,A} = \left(\frac{V_A^{\text{eff}}}{V_A^{\text{free}}} \right) \alpha_{0,A}^{\text{free}}, \tag{2.133}$$

where V_A^{eff} is the effective volume of the atom-in-molecule, and V_A^{free} represents the volume of the free atom. The ratio between these volumes is given by

$$\frac{V_A^{\text{eff}}}{V_A^{\text{free}}} = \frac{\int d\mathbf{r}\, r^3 w_A(\mathbf{r}) \rho(\mathbf{r})}{\int d\mathbf{r}\, r^3 \rho_A^{\text{free}}(\mathbf{r})} \,. \tag{2.134}$$

The XDM dispersion energy is computed as a sum over pairwise dispersion coefficients [20]:

$$E_{\text{XDM}}^{(2)} = -\frac{1}{2} \sum_{A \neq B} \sum_{n=3}^{5} f_{2n}^{\text{BJ}}(R_{AB}) \frac{C_{2n}^{AB}}{R_{AB}^{2n}} \,, \tag{2.135}$$

where $f_{2n}^{\text{BJ}}(R_{AB})$ is the Becke-Johnson (BJ) damping function:

$$f_{2n}^{\text{BJ}}(R_{AB}) = \frac{R_{AB}^{2n}}{R_{AB}^{2n} + (a_1 R_c^{AB} + a_2)^{2n}} \,. \tag{2.136}$$

In this equation, R_c^{AB} denotes a "critical" distance between atoms A and B, estimated as [138]

$$R_c^{ij} = \frac{1}{3}\left[\left(\frac{C_8^{AB}}{C_6^{AB}}\right)^{\frac{1}{2}} + \left(\frac{C_{10}^{AB}}{C_8^{AB}}\right)^{\frac{1}{2}} + \left(\frac{C_{10}^{AB}}{C_6^{AB}}\right)^{\frac{1}{4}}\right]. \tag{2.137}$$

The expression $a_1 R_c^{AB} + a_2$ essentially defines the vdW radius in the XDM method, with the two parameters a_1 and a_2 optimized for each density functional and perhaps the basis set [139].

A key advantage of the BJ damping function is its ability to extrapolate the dispersion energy to a finite value in the united-atom limit [129], aligning with the asymptotic result established by Koide [140]. This contrasts with other damping functions, such as Tang-Toennies (2.84), Chai–Head-Gordon (2.118), and Fermi-type (2.143), which drive the dispersion energy to zero at the united-atom limit, leading to unphysical repulsive forces at short distances [129]. Similar to the DFT-D3 method, XDM can be extended to account for the three-body Axilrod-Teller-Muto (ATM) term [130]. However, this extension has not been shown to consistently improve dispersion energies [130].

Since its introduction, significant theoretical advancements to the XDM method have been made by Ángyán [141, 142], Ayers [143] and Hesselmann [144], who have deepened the understanding of its theoretical underpinnings from first principles. Their analyses revealed that the expectation value $\langle \mathbf{r} \cdot \mathbf{d}_X \rangle$ holds a more fundamental role, entering the expressions for multipole moments, in contrast to the use of $\langle d_X^2 \rangle$, as adopted by XDM. Although XDM does not emerge from a strictly rigorous derivation, it remains a physically sound and computationally efficient framework for modeling dispersion interactions [20].

Tkatchenko-Scheffler Method

The Tkatchenko-Scheffler (TS) method [39] relies on the effective *atom-in-molecule* (AIM) C_6 dispersion coefficients obtained from accurate reference data for free-atom C_6 and semi-local electron density. Similar to the approach of London [6], the atomic response is modeled by a quantum harmonic oscillator (see more detailed discussion in 2.2.3). For polarizability, this leads to the frequency dependence as

$$\alpha_A(i\omega) = \alpha_{0,A} \big/ \left(1 + (i\omega/\omega_A)^2\right), \tag{2.138}$$

where $\alpha_{0,A} \equiv \alpha_A(0)$ is the static polarizability of atom A, and ω_A is the effective frequency of an oscillator. Inserting this into the Casimir-Polder integral, one obtains the London formula [6]:

$$C_6^{AB} = \frac{3}{2}\frac{\omega_A \omega_B}{\omega_A + \omega_B}\alpha_{0,A}\alpha_{0,B}. \tag{2.139}$$

Taking $A = B$ enables expressing the oscillator frequency via the atomic observables as $\omega_A = 4C_6^{AA}/3\alpha_{0,A}^2$, which combined with Eq. (2.139) gives the combination rule for heteronuclear coefficients [145]:

$$C_6^{AB} = \frac{2\alpha_{0,A}\alpha_{0,B}C_6^{AA}C_6^{BB}}{\alpha_{0,B}^2 C_6^{AA} + \alpha_{0,A}^2 C_6^{BB}}. \tag{2.140}$$

This combination rule gives a mean absolute relative error (MARE) of only 2.7% on a database of 70 heteronuclear C_6 coefficients for free atoms [39].

The AIM polarizabilities and dispersion coefficients are obtained by a simple rescaling of accurate free-atomic quantities (available from databases) based on the ratio of atomic volumes

from the Hirshfeld partitioning (see above):

$$\alpha_{0,A}^{\mathrm{TS}} = \left(\frac{V_A^{\mathrm{eff}}}{V_A^{\mathrm{free}}}\right)\alpha_{0,A}^{\mathrm{free}}\,, \quad C_{6,AA}^{\mathrm{TS}} = \left(\frac{V_A^{\mathrm{eff}}}{V_A^{\mathrm{free}}}\right)^2 C_{6,AA}^{\mathrm{free}}\,. \tag{2.141}$$

Note that the TS method uses the same scaling relation for polarizability as XDM (2.133), while the quadratic scaling of C_6 can be rationalized through the London formula (2.139). Recently, it was shown that slightly more accurate results can be obtained using a 4/3 scaling power for polarizability [146].

The key benchmark for any vdW model is a prediction of intermolecular C_6^{mol} coefficients. For molecules M_1 and M_2, it is defined as the sum over all atomic pair contributions:

$$C_6^{\mathrm{mol}} = \sum_{A\in M_1}\sum_{B\in M_2} C_{6,AB}^{\mathrm{TS}}\,. \tag{2.142}$$

The TS method shows a remarkable accuracy of 5.5% for C_6^{mol} on a database of 1225 molecular pairs [39], which is a factor of 2-3 more accurate than XDM [135, 137].

As in the DFT-D and XDM methods, the TS dispersion coefficients can be coupled with a damping function, avoiding double-counting of correlation energy at short distances, to compute the dispersion energy. The standard choice for the TS method is the Fermi-type damping

$$f_{\mathrm{damp}}(R_{AB}) = \frac{1}{1+\exp\left[-d\left(\frac{R_{AB}}{\beta R_{\mathrm{vdW}}^{AB}}-1\right)\right]}\,, \tag{2.143}$$

where $d = 20$, and β is the only fitting parameter, depending on the underlying density functional and optimized by minimizing the DFA+TS error on the benchmark S22 dataset, e.g. $\beta = 0.94$ for the PBE functional [39]. $R_{\mathrm{vdW}}^{AB} = R_{\mathrm{vdW}}^{A} + R_{\mathrm{vdW}}^{B}$ is the sum of effective vdW radii are obtained by rescaling the tabulated free-atomic vdW radii based on simple dimensionality arguments:

$$R_{\mathrm{vdW}}^{A} = \left(\frac{V_A^{\mathrm{eff}}}{V_A^{\mathrm{free}}}\right)^{1/3} R_{\mathrm{vdW}}^{A,\mathrm{free}}\,. \tag{2.144}$$

Later, it was shown that more accurate results can be obtained by using the quantum-mechanical scaling of vdW radius with polarizability [102]:

$$R_{\mathrm{vdW}}^{A} = 2.54\left(\alpha_{0,A}^{\mathrm{TS}}\right)^{1/7}\,. \tag{2.145}$$

The effects of vdW dispersion on the electronic structure can be studied using the self-consistent version of the TS method (sc-TS) [23]. This is achieved by incorporating an additional term to the KS potential, which is derived by considering all parameters of the method as functionals of electron density:

$$v_{\mathrm{TS}}[\rho(\mathbf{r})] = -\frac{1}{2}\sum_{AB}\left[\left(\frac{\delta f_{\mathrm{damp}}^{AB}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}\right)\frac{C_{6,AB}^{\mathrm{TS}}[\rho(\mathbf{r})]}{R_{AB}^6} + \frac{f_{\mathrm{damp}}^{AB}[\rho(\mathbf{r})]}{R_{AB}^6}\left(\frac{\delta C_{6,AB}^{\mathrm{TS}}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}\right)\right]\,. \tag{2.146}$$

Electrons "feel" additional potential due to vdW dispersion interactions, and the resulting KS orbitals and eigenvalues are changed accordingly. This enables investigating the vdW polarization of charge density in free and metal-adsorbed molecules [23, 24] or the effect of vdW interactions on the work function of metal surfaces with adsorbed molecules [23, 24]. These

vdW polarization effects on electron density are expected to become increasingly important for larger and more polarizable systems [25, 26], which is the key focus of Chapter 4.

Beyond the pairwise approximation and the lack of short-range electrodynamic screening of polarizability, a key limitation of the TS method lies in its reliance on Hirshfeld partitioning to capture local chemical environment effects. Standard Hirshfeld analysis tends to underestimate charge transfer [147], leading to a significant over- or underestimation of polarizabilities, depending on the sign of the transferred charge. This issue becomes particularly pronounced in ionic systems, as demonstrated in Refs. [148, 149].

Improvements can be achieved through iterative Hirshfeld partitioning [147, 149] or by introducing charge-dependent reference states for polarizability [150]. For hybrid organic-inorganic interfaces, the TS$^{\text{surf}}$ variant [151] incorporates metallic screening effects via Lifshitz-Zaremba-Kohn theory [77, 118], enabling accurate modeling of molecule-surface interactions. Moreover, the TS method can be coupled with the charge population analysis (CPA) approach [152] based on density-functional tight-binding (DFTB), enhancing its applicability to larger systems. In this case, the role of effective scaling parameter for polarizability is played by on-site contributions to Mulliken populations [152]. Finally, deep neural networks trained on DFT/DFTB data can be employed to predict Hirshfeld volume ratios in systems containing tens of thousands of atoms [153].

2.3.3 Nonlocal vdW Density Functionals

Nonlocal vdW density functionals have a general form [11, 12]

$$E_{\text{c,nl}} = \frac{1}{2}\iint d\mathbf{r}d\mathbf{r}'\,\rho(\mathbf{r})\Phi[\rho](\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')\,, \tag{2.147}$$

where $\Phi[\rho]$ is the *nonlocal kernel* correlating electron density at points $\mathbf{r}$ and $\mathbf{r}'$ in space. Note that nonlocal functionals define only correlation energy, as they are intended for use on top of the underlying semi-local or hybrid DFA. Nonlocal density functionals can be viewed as a truncated to the second order approximation to the ACFD formula (2.107) [11, 12]:

$$E^{(2)}_{\text{c,nl}} = -\frac{1}{4\pi}\int_0^\infty d\omega \iint d\mathbf{r}d\mathbf{r}'\,\mathrm{Tr}\left\{\left\langle \boldsymbol{\alpha}_{\text{sr}}\mathbf{T}_{\text{lr}}\boldsymbol{\alpha}_{\text{sr}}\mathbf{T}_{\text{lr}}\right\rangle(\mathbf{r},\mathbf{r}',i\omega)\right\}\,. \tag{2.148}$$

To bring it to the form (2.147), the nonlocal polarizability is usually approximated as a local isotropic function $\boldsymbol{\alpha}_{\text{sr}}(\mathbf{r},\mathbf{r}',i\omega) = \mathbb{1}\alpha_{\text{eff}}(\mathbf{r},i\omega)\delta(\mathbf{r}-\mathbf{r}')$. This allows one to get rid of the integrals over $\mathbf{r}'',\mathbf{r}'''$ and commute polarizability with dipole tensor, eventually moving α_{eff} outside the trace:

$$E^{(2)}_{\text{c,nl}} = -\frac{1}{4\pi}\int_0^\infty d\omega \iint d\mathbf{r}d\mathbf{r}'\,\alpha_{\text{eff}}(\mathbf{r},i\omega)\alpha_{\text{eff}}(\mathbf{r}',i\omega)\,\mathrm{Tr}\left\{\mathbf{T}_{\text{lr}}(\mathbf{r},\mathbf{r}')\mathbf{T}_{\text{lr}}(\mathbf{r},\mathbf{r}')\right\}\,. \tag{2.149}$$

Introducing the generalized "point-point" dispersion coefficient

$$C_6(\mathbf{r},\mathbf{r}') = \frac{3}{\pi}\int_0^\infty d\omega\,\alpha_{\text{eff}}(\mathbf{r},i\omega)\alpha_{\text{eff}}(\mathbf{r}',i\omega)\,, \tag{2.150}$$

and tracing the long-range coupling tensor as $\mathrm{Tr}\left\{\mathbf{T}_{\text{lr}}(\mathbf{r},\mathbf{r}')\mathbf{T}_{\text{lr}}(\mathbf{r},\mathbf{r}')\right\} = 6f^2(|\mathbf{r}-\mathbf{r}'|)/|\mathbf{r}-\mathbf{r}'|^6$, we obtain

$$E^{(2)}_{\text{c,nl}} = -\frac{1}{2}\iint d\mathbf{r}d\mathbf{r}'\,C_6(\mathbf{r},\mathbf{r}')\frac{f^2(|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|^6}\,. \tag{2.151}$$

Thus, the nonlocal density functionals are similar to pairwise-additive approaches in truncating the ACFD series at $n = 2$ but differ from them in not doing atomic coarse-graining (see Figure 2.2 for pictorial illustrations).

The typical choice for $\alpha_{\text{eff}}(\mathbf{r}, i\omega)$ is the local plasmon-type function

$$\alpha_{\text{eff}}(\mathbf{r}, i\omega) = \frac{\rho(\mathbf{r})}{\omega_p^2(\mathbf{r}) + \omega^2} , \tag{2.152}$$

where we introduced the *plasma frequency* of the uniform electron gas $\omega_p^2(\mathbf{r}) = 4\pi\rho(\mathbf{r})$ (in atomic units). Combining this with (2.150) and (2.151), we finally obtain $E_{\text{c,nl}}^{(2)}$ in the form (2.147) with the nonlocal kernel

$$\Phi[\rho](\mathbf{r}, \mathbf{r}') = \frac{3}{\pi} \int_0^\infty d\omega \, \frac{1}{\omega_0^2[\rho](\mathbf{r}) + \omega^2} \frac{1}{\omega_0^2[\rho](\mathbf{r}') + \omega^2} \frac{f^2(|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|^6} , \tag{2.153}$$

with the choice of local frequency functional $\omega_0^2[\rho]$ (which generally does not have to be $\omega_p^2(\mathbf{r})$) determining the differences between particular nonlocal functionals, together with the choice of the range-separation function f.

There are two primary categories of nonlocal correlation functionals: the "van der Waals density functionals" (vdW-DF) [154, 155] and the Vydrov–Van Voorhis (VV) functionals [156–159]. This discussion will focus exclusively on the most recent iteration of the VV family, namely the VV10 nonlocal correlation functional [159]. In this functional, the local excitation frequency is approximated as

$$\omega_0^2[\rho](\mathbf{r}) = \frac{\omega_p^2(\mathbf{r})}{3} + \omega_g^2(\mathbf{r}) = \frac{4\pi\rho(\mathbf{r})}{3} + C\frac{|\nabla\rho(\mathbf{r})|^4}{\rho^4(\mathbf{r})} , \tag{2.154}$$

where $\omega_g^2(\mathbf{r})$ is the local band gap descriptor based on the density gradient,[4] and C is the adjustable parameter controlling the accuracy of asymptotic C_6 coefficients. The VV10 correlation kernel is taken as

$$\Phi^{\text{VV10}}(\mathbf{r}, \mathbf{r}') = -\frac{3}{2g(\mathbf{r})g(\mathbf{r}')(g(\mathbf{r}) + g(\mathbf{r}'))} + \beta\delta(\mathbf{r} - \mathbf{r}') , \tag{2.155}$$

where

$$g(\mathbf{r}) = \omega_0(\mathbf{r})R^2 + \kappa(\mathbf{r}) , \quad R = |\mathbf{r} - \mathbf{r}'| , \tag{2.156}$$

and $\kappa(\mathbf{r}) = bk_F^2(\mathbf{r})/\omega_p(\mathbf{r})$, with the local Fermi wave vector $k_F(\mathbf{r}) = (3\pi^2\rho(\mathbf{r}))^{1/3}$. b is the second adjustable parameter responsible for the short-range damping of the kernel. The constant $\beta = (3/b^2)^{3/4}/16$ is tuned to ensure that $E_c^{\text{VV10}}[\rho]$ vanishes for the homogeneous electron gas limit [159]. This enables pairing the VV10 functional with existing DFAs without changing their description of the HEG.

In the short-range limit $|\mathbf{r} - \mathbf{r}'| \to 0$, the VV10 kernel behaves as [159]

$$\Phi^{\text{VV10}} = -A + B|\mathbf{r} - \mathbf{r}'|^2 + \dots , \tag{2.157}$$

in agreement with the asymptotic limit of Koide [140] for the dispersion energy. In the opposite infinite limit $|\mathbf{r} - \mathbf{r}'| \to \infty$,

$$\Phi^{\text{VV10}} \to -\frac{3}{2\omega_0(\mathbf{r})\omega_0(\mathbf{r}')(\omega_0(\mathbf{r}) + \omega_0(\mathbf{r}'))} \frac{1}{|\mathbf{r} - \mathbf{r}'|^6} , \tag{2.158}$$

in accordance with Eq. (2.153).

[4] Atomic and molecular density tails decay like $\propto e^{-2\sqrt{2I}r}$, where I is the ionization potential, meaning that the ratio $|\nabla\rho(\mathbf{r})|/\rho(\mathbf{r}) \propto \sqrt{I}$. This rationalizes why including density gradients is relevant for an accurate description of local excitation frequency at the long range.

2.3.4 Many-Body Dispersion Model

A very successful and efficient approach to beyond-pairwise vdW interactions is the many-body dispersion (MBD) formalism, which approximates atomic response functions with dipole-coupled (charged) harmonic oscillators. The MBD Hamiltonian for a system of N coupled oscillators centered at $\mathbf{R}_A$, with frequencies ω_A, charges q_A and masses m_A reads

$$\hat{H}_{\mathrm{MBD}} = \sum_A \left[-\frac{1}{2}\nabla^2_{\boldsymbol{\xi}_A} + \frac{1}{2}\omega_A^2 \xi_A^2 \right] + \frac{1}{2}\sum_{A\neq B} \omega_A \omega_B \sqrt{\alpha_{0,A}\alpha_{0,B}}\, \boldsymbol{\xi}_A \mathbf{T}_{AB}\big(|\mathbf{R}_A - \mathbf{R}_B|\big)\boldsymbol{\xi}_B\,, \tag{2.159}$$

where $\boldsymbol{\xi}_A = \sqrt{m_A}(\mathbf{r}_A - \mathbf{R}_A)$ are mass-weighted displacements of the oscillating charges, and $\mathbf{T}_{AB}$ is a 3×3 sub-block of the dipole interaction tensor $\mathbf{T}_{\mathrm{lr}}$. Since the MBD Hamiltonian represents a quadratic form, it can be exactly diagonalized, leading to the emergence of $3N$ collective oscillation modes with the frequencies $\tilde{\omega}_k$. The MBD energy is then expressed as the change in the zero-point energy of fluctuations due to the dipole interaction (plasmon-pole formula):

$$E_{\mathrm{MBD}} = \sum_{k=1}^{3N} \frac{\tilde{\omega}_k}{2} - \sum_{A=1}^{N} \frac{3\omega_A}{2}\,, \tag{2.160}$$

yielding result equivalent to Eq. (2.162) but Eq. (2.160) is much more efficient to evaluate, requiring only $3N\times 3N$ matrix diagonalization, which scales as $\mathcal{O}(N^3)$.

The static polarizabilities α_A and dispersion coefficients $C_{6,A}$ of atoms are connected to the oscillator parameters via simple relations:

$$C_{6,A} = \frac{3}{4}\hbar\omega_A\alpha_{0,A}^2\,, \quad \alpha_{0,A} = \frac{q_A^2}{m_A\omega_A^2}\,. \tag{2.161}$$

The particular choice of Hamiltonian parameters $\{\alpha_{0,A}, \omega_A\}$ and $\mathbf{T}_{AB}$ form defines different flavours of MBD [41, 160] (see below). Note that the MBD Hamiltonian, and hence the MBD modes and eigenvalues, depend only on two parameters per atom, and the usual convention is to fix $q_A = 1$ a.u. [25, 26, 160], making Eqs. (2.161) sufficient to fully parametrize the model (see also Section 2.2.3). However, some observables, such as the charge density of MBD modes, are not invariant to the choice of q_A. This question is discussed in detail in Chapter 4.

As other post-DFT vdW methods, MBD can be viewed as a certain approximation to the ACFD-RPA formula (2.107), which for the MBD model becomes [11, 41]:

$$E_{\mathrm{c}}^{\mathrm{MBD}} = -\frac{1}{2\pi}\int_0^\infty d\omega\, \mathrm{Tr}\big\{\ln\big(\mathbb{1}_{3N} - \boldsymbol{\alpha}^{\mathrm{sr}}(i\omega)\mathbf{T}_{\mathrm{lr}}\big)\big\} \tag{2.162}$$

where $\boldsymbol{\alpha}^{\mathrm{sr}}$ and $\mathbf{T}_{\mathrm{lr}}$ are now *coarse-grained* $3N\times 3N$ polarizability tensor and dipole coupling tensor, respectively. In practice, however, the equivalent yet more efficient formulation of MBD based on the model Hamiltonian (2.159) is used for calculating the MBD energy.

MBD@rsSCS Variant

The most widely used variant of MBD is the so-called MBD@rsSCS employing range-separated (rs) self-consistent screening (SCS). This flavor of MBD is based on the TS parametrization of QHOs through the Eqs. (2.141) and (2.161), followed by self-consistent screening to obtain a refined set of oscillator parameters before they enter the MBD Hamiltonian. Range separation is employed to avoid double-counting of short-range electrodynamic screening effects [40, 41].

The "bare" (isotropic) polarizability tensor of a system within the MBD framework is defined as $\mathbf{A}(i\omega) = \mathrm{diag}\{\boldsymbol{\alpha}_1^{\mathrm{TS}}(i\omega), ..., \boldsymbol{\alpha}_N^{\mathrm{TS}}(i\omega)\}$, where $\boldsymbol{\alpha}_A^{\mathrm{TS}} = \mathbb{1}\alpha_A^{\mathrm{TS}}$. The short-range screened atomic polarizabilities can then be obtained from the solution of the self-consistent screening equation (a coarse-grained analog of the Dyson equation) by matrix inversion and its partial contraction:

$$\alpha_A^{\mathrm{sr}}(i\omega) = \frac{1}{3}\mathrm{Tr}\Big\{\sum_C \mathbf{B}_{AC}\Big\}; \quad \mathbf{B} = \big(\mathbf{A}^{-1}(i\omega) + \mathbf{T}_{\mathrm{sr}}\big)^{-1} . \tag{2.163}$$

Summing over index C (over the rows of $\mathbf{B}$) in this equation recovers anisotropic screened atomic polarizabilities, which are further made isotropic through the trace operation. We note here that having isotropic tensors is not a strict requirement but this enables efficient analytical evaluation of MBD energy through Hamiltonian diagonalization, as described above. For the solution of the SCS equation, the short-range dipole coupling $\mathbf{T}_{\mathrm{sr}}$ needs to be defined.

The Coulomb interaction between two isotropic QHOs is screened due to the finite width of their Gaussian charge distributions:

$$v_{\mathrm{GG}}(R_{AB}) = \frac{\mathrm{erf}(R_{AB}/\sigma_{AB})}{R_{AB}}; \quad \sigma_{AB} = \sqrt{\sigma_A^2 + \sigma_B^2} . \tag{2.164}$$

The oscillator width is derived from the zero-distance limit of the classical dipole-dipole interaction [40, 161]:

$$\sigma_A(i\omega) = \left(\frac{1}{3}\sqrt{\frac{2}{\pi}}\alpha_A^{\mathrm{sr}}(i\omega)\right)^{1/3} \tag{2.165}$$

The full-coupling dipole-dipole tensor derived from v_{GG} is therefore:

$$\begin{aligned} \mathbf{T}_{AB}^{ij}(\mathbf{R}, \sigma) &= \frac{\partial^2}{\partial R_i \partial R_j}\frac{\mathrm{erf}(\zeta)}{R} = \big(\mathrm{erf}(\zeta) - \Theta(\zeta)\big)\mathbf{T}_{\mathrm{bare}}^{ij}(\mathbf{R}) + 2\zeta^2\Theta(\zeta)\frac{R_i R_j}{R^5}, \\ \Theta(\zeta) &= \frac{2\zeta}{\sqrt{\pi}}e^{-\zeta^2}, \; \zeta = \frac{R}{\sqrt{\sigma_A^2 + \sigma_B^2}} \quad (i, j = x, y, z) . \end{aligned} \tag{2.166}$$

Here, $\mathbf{T}_{\mathrm{bare}}^{ij}(\mathbf{R}) = (-3R_iR_j + \delta_{ij}R^2)/R^5$ is the "bare" dipole tensor. Within MBD@rsSCS, the short-range (frequency-dependent) coupling tensor is obtained as

$$\mathbf{T}_{AB}^{\mathrm{sr}} = \big(1 - f_{\mathrm{damp}}(R_{AB})\big)\mathbf{T}_{AB}(R_{AB}) , \tag{2.167}$$

where $f_{\mathrm{damp}}(R_{AB})$ is the same Fermi-type damping function as in the TS method:

$$f_{\mathrm{damp}}(R_{AB}) = \frac{1}{1 + \exp\left[-a\left(\frac{R_{AB}}{\beta(R_{\mathrm{vdW}}^A + R_{\mathrm{vdW}}^B)} - 1\right)\right]} , \tag{2.168}$$

where $a = 6$ and β is fitted to the chosen xc functional. With the so-defined $\mathbf{T}^{\mathrm{sr}}$, depending on frequency through σ, the SCS equation can be solved as described in eq. (2.163) on a quadrature grid of imaginary frequencies [94, 160]. The obtained screened polarizabilities are then used to renormalize oscillator frequencies and C_6 coefficients:

$$\bar{C}_{6,A} = \frac{3}{\pi}\int_0^\infty d\omega\,[\alpha_A^{\mathrm{sr}}(i\omega)]^2 , \quad \bar{\omega}_A = \frac{4\bar{C}_{6,A}}{3[\alpha_A^{\mathrm{sr}}(0)]^2} . \tag{2.169}$$

The screened polarizabilities and oscillator frequencies finally enter the Hamiltonian (2.159), with the dipole coupling tensor given by

$$\mathbf{T}_{AB}^{\mathrm{lr}} = f_{\mathrm{damp}}(R_{AB})\mathbf{T}_{AB}(R_{AB},\sigma) \approx f_{\mathrm{damp}}(R_{AB})\mathbf{T}_{\mathrm{bare}}(R_{AB}) \,. \tag{2.170}$$

Here, the "bare" dipole tensor is used to make $\mathbf{T}_{AB}^{\mathrm{lr}}$ frequency-independent, which ensures the equivalence between the RPA correlation energy (2.162) and the MBD interaction energy (2.160) and delivers excellent approximation to the full dipole tensor at long range (Figure A13).

The MBD@rsSCS method is designed to accurately describe vdW dispersion in finite-gap systems, where the localized atomic response approximation works well [41]. DFA+MBD@rsSCS approach provides accurate interaction energies for molecular dimers [41, 94, 162], molecular crystals [14, 95], supramolecular complexes [25, 162, 163], as well as anisotropic polarizabilities for molecular systems [11, 25, 94, 95].

MBD-NL Variant

Sometimes, the MBD@rsSCS method might fail due to the negative eigenvalues emerging in the diagonalization procedure, which is especially relevant for ionic and transition-metal compounds [21, 150]. However, this deficiency is due to the TS-based parametrization of QHOs and not due to the MBD framework itself. By choosing more appropriate initial parametrization for α and C_6, these issues might be avoided. The state-of-the-art nonlocal MBD method (MBD-NL) [21] uses the nonlocal VV10 polarizability functional for that purpose (see Section 2.3.3):

$$\alpha_{\mathrm{VV}}[\rho](\mathbf{r}, i\omega) = \frac{\rho(\mathbf{r})}{\frac{4\pi\rho(\mathbf{r})}{3} + C\frac{|\nabla\rho(\mathbf{r})|^4}{\rho^4(\mathbf{r})} + \omega^2} \,. \tag{2.171}$$

To avoid double counting of correlation energy due to the regions with slowly varying, jellium-like densities (which is already accounted for by the xc functional), the MBD-NL method smoothly cuts off the contributions of those regions to the polarizability. This is achieved using the iso-orbital indicator h (2.54) and the local ionization potential $I_{\mathrm{loc}} = |\nabla\rho(\mathbf{r})|^2/8\rho^2(\mathbf{r})$, which enter the cutoff function g [21]:

$$\alpha'_{\mathrm{VV}}[\rho](\mathbf{r}, i\omega) = g(I_{\mathrm{loc}}, h)\alpha_{\mathrm{VV}}[\rho](\mathbf{r}, i\omega) \,. \tag{2.172}$$

The function g is finely tuned not to influence the polarizabilities of simple molecules like benzene while cutting off the jellium-like parts [21]. Furthermore, the polarizability density is partitioned using Hirshfeld weights $w_A(\mathbf{r})$ (2.132), leading to atomic dynamic polarizabilities

$$\alpha_A^{\mathrm{VV}}(i\omega) = \int d\mathbf{r}\, w_A(\mathbf{r})\alpha'_{\mathrm{VV}}(\mathbf{r}, i\omega) \,. \tag{2.173}$$

This dispenses with the need for the SCS procedure, and the atomic C_6 coefficients can then be evaluated immediately using the Casimir-Polder integral:

$$C_{6,\mathrm{A}}^{\mathrm{VV}} = \frac{3}{\pi}\int_0^\infty d\omega\, \alpha_A^{\mathrm{VV}}(i\omega)\alpha_A^{\mathrm{VV}}(i\omega) \,. \tag{2.174}$$

To eliminate errors due to the uneven accuracy of VV10 polarizability functional across the periodic table, the obtained atomic parameters are renormalized with respect to the accurate reference free-atom values [21]:

$$\alpha_{0,\mathrm{A}}^{\mathrm{rVV}} = \alpha_{0,\mathrm{A}}^{\mathrm{VV}}\frac{\alpha_{0,\mathrm{A}}^{\mathrm{ref,free}}}{\alpha_{0,\mathrm{A}}^{\mathrm{VV,free}}} \,, \quad C_{6,\mathrm{A}}^{\mathrm{rVV}} = C_{6,\mathrm{A}}^{\mathrm{VV}}\frac{C_{6,\mathrm{A}}^{\mathrm{ref,free}}}{C_{6,\mathrm{A}}^{\mathrm{VV,free}}} \,. \tag{2.175}$$

Finally, the parameters $\left\{\alpha_{0,\mathrm{A}}^{\mathrm{rVV}},\, \omega_A\right\}$, with $\omega_A = 4C_{6,\mathrm{A}}^{\mathrm{rVV}}/3(\alpha_{0,\mathrm{A}}^{\mathrm{rVV}})^2$, enter the Hamiltonian (2.159), where the dipole tensor is damped at short-range via $\mathbf{T}_{AB}^{\mathrm{lr}} = f_{\mathrm{damp}}(R_{AB})\mathbf{T}_{\mathrm{bare}}(R_{AB})$ like in MBD@rsSCS. The only difference is in the definition of the vdW radii, which are obtained in MBD-NL using Eq. (2.145).

MBD-NL yields substantially improved binding energies for metals, transition-metal dichalcogenides, and ionic solids, while performing nearly identically to MBD@rsSCS for semiconductor, molecular crystals, and organic molecules [21]. In addition, MBD-NL can effectively treat organic-metal interfaces [21]. The only missing physics in MBD-NL are long-range coupled delocalized electronic fluctuations (type-C non-additivity in Dobson's classification [119]), taking place in conductors. This is, however, one of the biggest gaps in the theory of vdW interactions in general.

Alternative Developments

In the past decade, numerous developments have emerged within the MBD framework, addressing its physical limitations while enhancing its computational efficiency and broadening its applicability. These advancements reflect the growing interest in refining MBD to better capture complex dispersion interactions and improve its utility across diverse material systems.

For example, the so-called fractionally-ionic MBD (FI-MBD) method by Gould *et al.* [150] employs charge-dependent reference polarizabilities to address the deficiencies of Hirshfeld partitioning in ionic systems. In addition, eigenvalue remapping is performed in FI-MBD to avoid the issue with negative eigenvalues [150]. The follow-up "universal" MBD (uMBD) method [164] relies on the FI-MBD polarizabilities and introduces an additional smearing to the fluctuating dipoles in the MBD Hamiltonian (2.159) to attenuate the dispersion correction in a mid-range. This is accomplished by bringing in a scaling parameter a in the definition of σ in Eq. (2.165), which is subsequently fitted to minimize the error on the benchmark S22 dataset [164]. The optimal value $a = 0.77$ is below unity, indicating that uMBD effectively reduces the range of oscillators fluctuations.

Another direction is extending MBD beyond dipole approximation, as proposed in Refs. [165–167], where quadrupolar couplings within MBD were considered. This extension is significant because higher-order multipolar interactions, such as quadrupole couplings, are important for accurately modeling the vdW dispersion interactions [11, 12], and by including them the predictive accuracy of the MBD framework can be further enhanced. Recently, it was also demonstrated that the MBD method can provide a semi-quantitative description of optically excited states [168], serving as a much cheaper alternative to the solution of the Bethe-Salpeter equation.

The MBD+C method [169] represents yet another notable advancement, which tackles the challenge of capturing delocalized charge fluctuations in metallic systems by incorporating charge-hopping terms in the MBD model. From a computational standpoint, MBD+C introduces a larger prefactor due to the emergence of $4N \times 4N$ matrices, though it retains the cubic scaling of the standard MBD algorithm. This makes MBD+C significantly more efficient than ACFD-RPA, which is practically the only method capable of accurately capturing the type-C non-additivity [119].

Moreover, the computational scaling of MBD can be reduced to linear by using stochastic trace estimators, as shown by Piquemal and co-workers [170]. Using deep neural networks to predict Hirshfeld volume ratios [171] further alleviates the computational load related to MBD by dispensing with density-functional (tight-binding) calculations, normally used to set the oscillator

parameters. Last but not least, the second-quantized approach to MBD (SQ-MBD) [172] offers a promising path for exploring collective plasmon-like phenomena in large-scale biomolecular systems.

In Chapter 4 of this thesis, we introduce a novel variant of the MBD method, termed Fully-Coupled Optimally-Tuned MBD (MBD@FCO). This approach is designed to accurately and simultaneously capture both the dispersion energy and the impact of vdW dispersion on electronic density. The new method leverages the full dipole tensor (2.166) and incorporates an optimized parametrization of QHOs. Detailed explanations and derivations are provided in Section 4.1.

Chapter 3

Universal Pairwise Interatomic van der Waals Potentials

Parts of this chapter have been published in this or similar form in:
A. Khabibrakhmanov, D. V. Fedorov, A. Tkatchenko,
J. Chem. Theory Comput. **19**, 7895–7907, 2023.

Van der Waals (vdW) forces play an indisputably important role in determining the structure and dynamics of many biomolecular, solid-state, and polymeric systems [7, 9–12]. The accurate description of vdW interactions requires sophisticated quantum-mechanical treatment, using the adiabatic-connection fluctuation-dissipation theorem (ACFDT) in density-functional theory or high-level quantum chemistry methods, such as coupled cluster or quantum Monte-Carlo [11, 12]. However, the prohibitive computational cost of these methods precludes their applicability to extended (bio)molecular systems. Therefore, practical simulations of large and complex systems are often done using classical force fields such as AMBER [27], CHARMM [28], or GROMACS [29].

For the description of vdW forces, these popular force fields resort to the seminal Lennard-Jones (LJ) [30] (or an improved Buckingham [173]) potential as a practical shortcut. Two parameters, the well depth D_e and the equilibrium position R_e, fully specify the LJ potential. However, these parameters can be determined unambiguously only for relatively simple vdW-bonded systems, such as noble gas dimers or crystals. Moreover, the LJ potential is notorious for its lack of flexibility and very limited quantitative accuracy [174, 175]. On the other hand, the celebrated Tang-Toennies potentials [85–87, 176] are derived from first principles and yield high accuracy for dimers including noble gases and group II elements. To achieve such an accuracy, the Tang-Toennies potentials employ from 5 to 9 parameters, depending on the exact flavor [87]. Setting these parameters requires knowledge of R_e and D_e for each vdW bonded dimer [85, 86], which prevents a generalization of the Tang-Toennies models to the whole periodic table. Moreover, like the LJ potential, the most recent conformal Tang-Toennies (TTS) potential [86] is prone to large errors for dispersion coefficients (see Figure 3.1a) despite its high accuracy close to equilibrium distances. Hence, a vdW potential combining wide transferability across the periodic table, high accuracy, and minimal parametrization is not yet available.

Here, we develop a universal conformal pairwise vdW potential, which can be parametrized for all chemical elements based solely on two *non-bonded* atomic properties – static dipole polarizability α_1 and dipole-dipole dispersion coefficient C_6. Our potential is consistently derived within the framework of the quantum Drude oscillator (QDO) model [99] using the Heitler-

London perturbation theory [111, 177], and it is devoid of adjustable parameters. This is achieved by building connections between atomic scaling laws [102, 103, 178], the microscopic law of corresponding states [4, 179–181], and symmetry-adapted perturbation theory (SAPT) [37, 38] for intermolecular interactions. The derived exchange repulsion term in our potential obeys correct physical limits both at $R \to 0$ and $R \to \infty$, and the predicted C_6 dispersion coefficients are significantly more accurate compared to the other conformal Lennard-Jones and Tang-Toennies [86] potentials. The designed vdW-QDO potentials are twice as accurate as the LJ potentials when averaged over 15 noble-gas dimers. In addition, the vdW-QDO potential augmented by a damping function can accurately describe binding curves of dimers consisting of (closed-shell) group II atoms. Moreover, the vdW-QDO potential can be applied to molecular systems, when coupled with an atom-in-molecule (AIM) approach [39]. We demonstrate this by accurately reproducing the dispersion energy for dispersion-dominated molecular dimers from the S66×8 dataset [182].

We derive the vdW potential in the QDO framework, which is a coarse-grained model for the electronic response [96, 97, 109, 183–185] proved to be accurate and insightful in many applications across various fields [40, 41, 98, 99, 103, 109, 110, 168, 186–189]. Within the QDO model, the response of valence electrons is described *via* a quasi-particle (*drudon* or Drude particle) with a negative charge $-q$ and mass m, harmonically bound to a positively-charged pseudo-nucleus of charge q with a characteristic frequency ω. Coupled QDOs are also extensively used in the development of vdW density functionals [39, 40, 190], quantum mechanical [99, 109] and polarizable force fields [187, 191–194] as well as recent machine learning force fields [153, 195].

The QDO model has been already used to build interatomic vdW potentials for water or noble-gas dimers and crystals [99, 100, 109, 186, 187]. However, within the corresponding studies, the repulsive term was added in *ad hoc* manner, either by fitting Born-Mayer [173, 196] exponents to *ab initio* repulsive walls [99, 109, 186, 187] or by directly adding the Hartree-Fock exchange energy [100]. Therefore, such potentials cannot be generalized beyond the systems for which direct first-principles simulations are possible. In contrast, here we suggest a consistent treatment of both Pauli (exchange) repulsion and vdW dispersion within the QDO framework. To our knowledge, this is the first vdW potential of such a type, which does not directly utilize the reference binding energy of dimers or the Hartree-Fock exchange energy curve, but nevertheless provides relatively good accuracy.

3.1 Model Construction

The long-range vdW dispersion energy for two identical QDOs is given by the usual multipolar series [99, 185]

$$E_{\mathrm{disp}}(R) = -\frac{C_6}{R^6} - \frac{C_8}{R^8} - \frac{C_{10}}{R^{10}} - \dots , \tag{3.1}$$

where the dispersion coefficients are related to the oscillator parameters *via* the closed-form expressions (see Section 2.2.3):

$$C_6 = \frac{3}{4}\hbar\omega\alpha_1^2 k_e^2 , \quad C_8 = \frac{5\hbar}{m\omega} C_6 , \quad C_{10} = \frac{245\hbar^2}{8(m\omega)^2} C_6 , \tag{3.2}$$

where $\alpha_1 = q^2/m\omega^2$ is the QDO dipole polarizability and $k_e = 1/4\pi\varepsilon_0$.[1] Tang and Toennies showed [85, 176] that including the three leading dispersion terms is sufficient to obtain the accurate vdW potential. Therefore, we also truncate the series of Eq. (3.1) at the C_{10} term.

[1] All equations in this chapter are written in SI units.

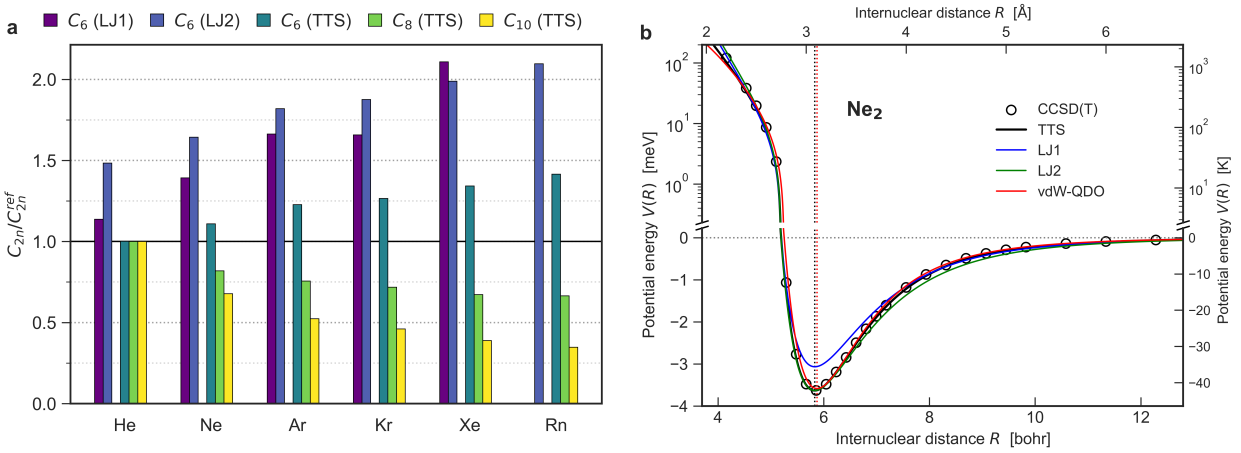

Figure 3.1: Dispersion coefficients of noble gases and the potential curve of a neon dimer. **(a)** Errors in dispersion coefficients arising from the LJ and TTS potentials. The TTS dispersion coefficients are obtained as $C_{2n} = C_{2n}^{*} \times D_e R_e^{2n}$ [86], and the reference dispersion coefficients C_{2n}^{ref} are given in Table 3.1. **(b)** The vdW-QDO potential for neon dimer benchmarked to the TTS potential and the reference CCSD(T) potential [197]. Vertical dotted lines indicate the equilibrium distance as predicted by CCSD(T) (black) and vdW-QDO potential (red). For comparison, the LJ potentials in two different parametrizations (see our discussion in the text) are also displayed in blue and green. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

The exchange repulsion is introduced into the model according to Refs. [102, 103], where multipole contributions to the exchange energy of a homonuclear dimer were derived by considering two identical drudons as bosons, assuming that they represent closed valence shells of atoms with a zero total spin. Consequently, the total wave function of a dimer is represented by a symmetrized product

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} \left(\psi_A(\mathbf{r}_1)\psi_B(\mathbf{r}_2) + \psi_A(\mathbf{r}_2)\psi_B(\mathbf{r}_1) \right) , \tag{3.3}$$

where $\psi_A(\mathbf{r}_1) = \left(\frac{m\omega}{\pi\hbar}\right)^{3/4} e^{-\frac{m\omega}{2\hbar}\mathbf{r}_1^2}$ and $\psi_B(\mathbf{r}_2) = \left(\frac{m\omega}{\pi\hbar}\right)^{3/4} e^{-\frac{m\omega}{2\hbar}(\mathbf{r}_2 - \mathbf{R})^2}$ are, respectively, the ground-state wave functions of drudons centered at nuclei A and B separated by $\mathbf{R}$. Within the Heitler-London perturbation theory [111, 177], the exchange energy of two identical vdW-bonded QDOs at near-equilibrium and larger distances is well approximated by the exchange integral [102, 103]:

$$E_{\mathrm{ex}} \approx J_{\mathrm{ex}} = \left\langle \psi_A(\mathbf{r}_1)\psi_B(\mathbf{r}_2) \right| \hat{V}_C \left| \psi_A(\mathbf{r}_2)\psi_B(\mathbf{r}_1) \right\rangle . \tag{3.4}$$

The evaluation of Eq. (3.4) with the multipole expansion of Coulomb coupling $\hat{V}_C$ between the two QDOs results in multipole contributions to the exchange energy [102, 103]. In dipole approximation, this yields

$$J_{\mathrm{ex}}^{(1)} = k_e q^2 S/2R , \quad S = \left| \langle \psi_A | \psi_B \rangle \right|^2 = e^{-\frac{m\omega}{2\hbar}R^2} , \tag{3.5}$$

where S is the overlap integral. Higher-order multipole contributions ($l > 1$) to the exchange repulsion energy $J_{\mathrm{ex}}^{(l)}$ have the same leading-term dependence on internuclear distance R, with the only difference in a proportionality coefficient, i.e. $J_{\mathrm{ex}}^{(l)} \propto k_e q^2 S/R$ [103]. Therefore, we introduce an effective exchange repulsion energy as

$$E_{\mathrm{ex}}^{\mathrm{eff}} = A k_e q^2 S/R , \tag{3.6}$$

with the proportionality coefficient A to be determined self-consistently, in what follows. In this way, we effectively include multipole contributions to all orders. Importantly, our $E_{\mathrm{ex}}^{\mathrm{eff}}$ has

Table 3.1: The reference parameters of noble-gas dimers: dipole polarizability α_1 (in a.u.), dispersion coefficients C_6, C_8, C_{10} (in a.u.), and dimer potential well parameters R_e (in bohr) and D_e (in meV) for noble-gas dimers. For R_e and D_e, the values in Å and Kelvin, respectively, are extra given in parentheses. Also, we compare their reference values (columns 5 and 7) to the predictions of Eqs. (3.10) and (3.17) (columns 6 and 8). The values for C_6, C_8, and C_{10} labeled with the star ($*$) are taken from Ref. [85] instead of Refs. [89, 101].

	α_1 [101]	C_6 [101]	C_8 [89]	C_{10} [89]	R_e^{ref} (in Å) [86]	R_e (in Å)	D_e^{ref} (in K) [86]	D_e (in K)
He_2	1.38	1.46	14.123	183.79	5.608 (2.97)	5.35 (2.83)	0.948 (10.99)	1.634 (19.0)
Ne_2	2.67	6.38	90.265	1532.8	5.83 (3.09)	5.87 (3.11)	3.632 (42.15)	4.049 (47.0)
Ar_2	11.1	64.3	1621.5	49033	7.11 (3.76)	7.20 (3.81)	12.319 (142.95)	12.00 (139.3)
Kr_2	16.8	129.6	4040	150130	7.589 (4.02)	7.64 (4.04)	17.310 (200.87)	16.94 (196.6)
Xe_2	27.3	285.9	12004	588210	8.273 (4.38)	8.19 (4.33)	24.126 (279.97)	24.64 (285.9)
Rn_2	33.54	420.6*	19263*	1067000*	8.37 (4.43)	8.43 (4.46)	34.885 (404.81)	30.38 (352.5)

$1/R$ dependence, which properly describes the infinite repulsive wall at short distances. Thus, in contrast to the Born-Mayer or Duman-Smirnov [92, 198, 199] functional forms for exchange repulsion possessing a finite value of E_{ex} at $R \to 0$, Eq. (3.6) agrees with the orbital overlap model for Pauli repulsion [200, 201]. Moreover, our $E_{\text{ex}}^{\text{eff}}$ does not rely on empiricism, as it explicitly depends only on the QDO parameters (*vide infra*), whereas the existing Pauli repulsion models require fitting to some *ab initio* data [92, 98, 99, 109, 201–203].

To determine the coefficient A in Eq. (3.6), we employ the force balance condition at the equilibrium distance, $\left(-\nabla_R E_{\text{ex}}^{\text{eff}} - \nabla_R E_{\text{disp}}\right)\big|_{R=R_e} = 0$, which yields

$$Ak_e q^2 \left[\frac{1}{R_e^2} + \frac{m\omega}{\hbar}\right] e^{-\frac{m\omega}{2\hbar}R_e^2} = \frac{6C_6}{R_e^7} + \frac{8C_8}{R_e^9} + \frac{10C_{10}}{R_e^{11}} \,. \tag{3.7}$$

To evaluate the equilibrium distance R_e in our model, we use the quantum-mechanical relation between the atomic (static) dipole polarizability and vdW radius [102]

$$\alpha_1 = \Phi \times R_{\text{vdW}}^7 \,, \tag{3.8}$$

where the proportionality coefficient Φ is given by [178]

$$\Phi = (4\pi\varepsilon_0 / a_0^4) \times \alpha_{\text{fsc}}^{4/3} \,, \tag{3.9}$$

with $\alpha_{\text{fsc}} = e^2/4\pi\varepsilon_0\hbar c \approx 1/137.036$ as the fine-structure constant. The relation given by Eqs. (3.8)–(3.9) turned out to be valid for real atoms. Especially, it is very accurate for noble gases, where the mean absolute relative error (MARE) $\langle |R_{\text{vdW}} - R_{\text{vdW}}^{\text{ref}}| / R_{\text{vdW}}^{\text{ref}} \rangle$ is about 1%. [102, 178] Since by definition R_{vdW} is a half of the equilibrium distance R_e in a homonuclear vdW bonded dimer [102, 204], accurate equilibrium distances can be obtained *via*

$$R_e = 2 \times R_{\text{vdW}} = 2 \times (\alpha_1/\Phi)^{1/7} \,. \tag{3.10}$$

With α_1 and C_6 being fixed, there are two unknown quantities in Eq. (3.7), A and $m\omega$, since C_8 and C_{10} are solely expressed in terms of C_6 and $m\omega$ via Eq. (3.2).

As shown in Ref. [110], the product $m\omega$ can be obtained from the force balance in the dipole approximation

$$\frac{k_e q^2}{2}\left[\frac{1}{R_e^2} + \frac{m\omega}{\hbar}\right] e^{-\frac{m\omega}{2\hbar}R_e^2} = \frac{6C_6}{R_e^7} \,, \tag{3.11}$$

with R_e substituted from Eq. (3.10). Solution of this transcendental equation allows to determine the three oscillator parameters $\{q, m, \omega\}$ given only $\{\alpha_1, C_6\}$. We denote this parametrization scheme as vdW-OQDO, similar to the recently suggested OQDO scheme [110]. The details of the procedure and the corresponding values of $\{q, m, \omega\}$ can be found in Appendix A2.

Solving Eqs. (3.11) and (3.7) together, one can obtain

$$A = \frac{1}{2} + \frac{2C_8}{3C_6R_e^2} + \frac{5C_{10}}{6C_6R_e^4} \tag{3.12}$$

and the total vdW potential

$$V_{\mathrm{QDO}} = A\frac{k_e q^2}{R} e^{-\frac{(\gamma R)^2}{2}} - \sum_{n=3}^{5} \frac{C_{2n}}{R^{2n}}\,, \quad \gamma = \sqrt{m\omega/\hbar}\,. \tag{3.13}$$

The vdW-QDO potential for neon is displayed by the red curve in Figure 3.1b, which shows excellent agreement with the TTS potential [86] as well as with the CCSD(T) calculations [197] across the whole range of distances from $0.7R_e$ (~4 Bohr) to infinity. Inclusion of C_8 and C_{10} dispersion coefficients together with the suggested approach to treat exchange repulsion energy allows us to predict the correct depth and shape of the potential without losing the accuracy in predicting the equilibrium distance, which is inherited from the dipole approximation. In addition, we compare our potential to the LJ potential, for which we use two different parametrizations: LJ1 derived from thermodynamical properties and LJ2 designed to reproduce reference R_e and D_e (see Appendix A3 for more details). We note that the present vdW-QDO potential (3.13) performs accurately in the whole range of distances, whereas the LJ1 potential (blue curve in Figure 3.1b) underestimates the energy in potential minimum region and the LJ2 potential overestimates the long-range energy (green curve), although both being reasonably accurate in the repulsive region. This imbalance and lack of flexibility of the LJ potential, which is observed for all noble gases, is one of the main issues limiting its quantitative predictive power [174, 175]. Moreover, the LJ potential severely overestimates C_6 coefficient (Figure 3.1a), which is responsible for the correct long-range energy. The proposed vdW-QDO potential overcomes these difficulties without increasing the number of parameters. Moreover, our potential recovers correct bonding behavior using only a free atom property α_1 and asymptotic interaction parameter C_6, which do not contain information about the interaction between atoms at short distances.

3.2 vdW-QDO Potential for Noble-Gas Dimers

With the accurate Ne_2 potential curve in hand, its counterparts for all other noble-gas dimers can be derived using the conformality of their potentials [86, 205, 206], which is a microscopic manifestation of the law of corresponding states [179–181]. Namely, for the vdW potential of other noble-gas dimers, we write

$$V_{\mathrm{QDO}}(R) = D_e U^{\mathrm{Ne}}_{\mathrm{QDO}}(x), \quad x = R/R_e\,, \tag{3.14}$$

where $U^{\mathrm{Ne}}_{\mathrm{QDO}}(x) = V^{\mathrm{Ne}}_{\mathrm{QDO}}(xR^{\mathrm{Ne}}_e)\big/V^{\mathrm{Ne}}_{\mathrm{QDO}}(R^{\mathrm{Ne}}_e)$ is the dimensionless potential (shape) of Ne_2 dimer

$$U^{\mathrm{Ne}}_{\mathrm{QDO}}(x) = \frac{A^*}{x} e^{-\frac{(\gamma^* x)^2}{2}} - \sum_{n=3}^{5} \frac{C^*_{2n}}{x^{2n}}\,, \tag{3.15}$$

with the numerical values of the starred (unitless) parameters and their definitions presented in Table 3.2.

Parameter	Definition	Numerical value
A^*	$Ak_e q^2/R_e D_e$	1508.917
γ^*	$R_e\sqrt{m\omega/\hbar}$	3.912
C_6^*	$C_6/D_e R_e^6$	1.1779
C_8^*	$5C_6/D_e R_e^6(\gamma^*)^2$	0.3848
C_{10}^*	$245C_6/8D_e R_e^6(\gamma^*)^4$	0.1540

Table 3.2: Dimensionless parameters of the vdW-QDO potential in Eq. (3.15). The Ne_2 dimer parameters used in the second column are D_e = 3.586 meV [Eq. (3.16)] and R_e = 5.875 bohr [Eq. (3.10)]. The QDO parameters for Ne_2 dimer are q = 1.18865, m = 0.37164, ω = 1.19326 (in a.u.).

Thus, only R_e and D_e for every dimer are required to obtain their vdW potential. For R_e, the accurate scaling law (3.10) is already established, whereas an analogous scaling law for D_e of noble-gas dimers is not yet known. Substituting $R = R_e$ to Eq. (3.13) and using Eq. (3.7) to eliminate $Ak_e q^2 e^{-\frac{(\gamma R)^2}{2}}/R$ yields

$$D_e = -V_{\mathrm{QDO}}^{\mathrm{Ne}}(R_e^{\mathrm{Ne}}) = \frac{C_6}{R_e^6}\left(1 - \frac{\beta-5}{\beta(1+\beta)} - \frac{40}{\beta(1+\beta)} + \frac{245}{8\beta^2} - \frac{2450}{8\beta^2(1+\beta)}\right), \tag{3.16}$$

with $\beta = \frac{m\omega}{\hbar}R_e^2 = (\gamma^*)^2$. Analyzing reference CCSD(T) data for D_e from Ref. [86], we found that Eq. (3.16) truncated at first two terms can accurately predict D_e for all noble-gas dimers

$$D_e \approx \frac{C_6}{R_e^6}\left(1 - \frac{\beta-5}{\beta(1+\beta)}\right). \tag{3.17}$$

In Figure 3.2, D_e by Eq. (3.17) are compared to the reference CCSD(T) data. The bar chart shows that Eq. (3.17) is accurate for homo- and heteronuclear dimers of He–Xe with all errors below 1 meV. For dimers with Rn, the errors are larger, with Rn_2 being underbound by 4.5 meV or 13%. The larger errors for Rn dimers likely stem from the fact that the reference coupled-cluster calculation [213] is less reliable than the corresponding calculations for the lighter dimers He_2–Xe_2 [197, 207–210]. For example, D_e of Xe_2 dimer reported in Ref. [213] is by 7.5% larger than the one of Ref. [210], which is state-of-the-art calculation. Thus, a similar or even larger overestimation of D_e should be expected for Rn_2 [213], where relativistic effects are more pronounced. Accounting for that, the estimated error of Eq. (3.17) for Rn_2 would not exceed 5.5%. We conclude that the suggested scaling law (3.17) allows one to accurately evaluate D_e for all noble-gas dimers given only $\{\alpha_1, C_6\}$ without involving any adjustable parameters.

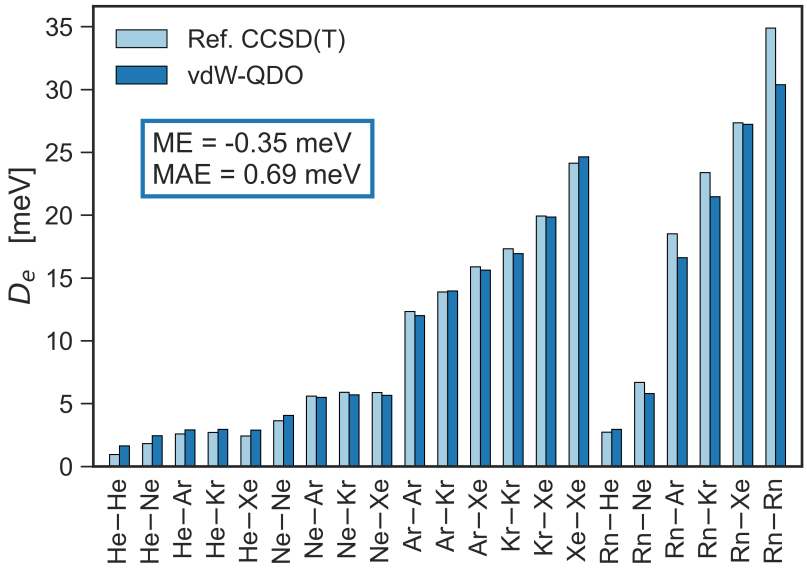

Figure 3.2: Binding energies of noble-gas dimers. D_e by Eq. (3.17) compared to the reference CCSD(T) values [86] for 21 noble-gas dimers. Mean error (ME) and mean absolute error (MAE) are displayed. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

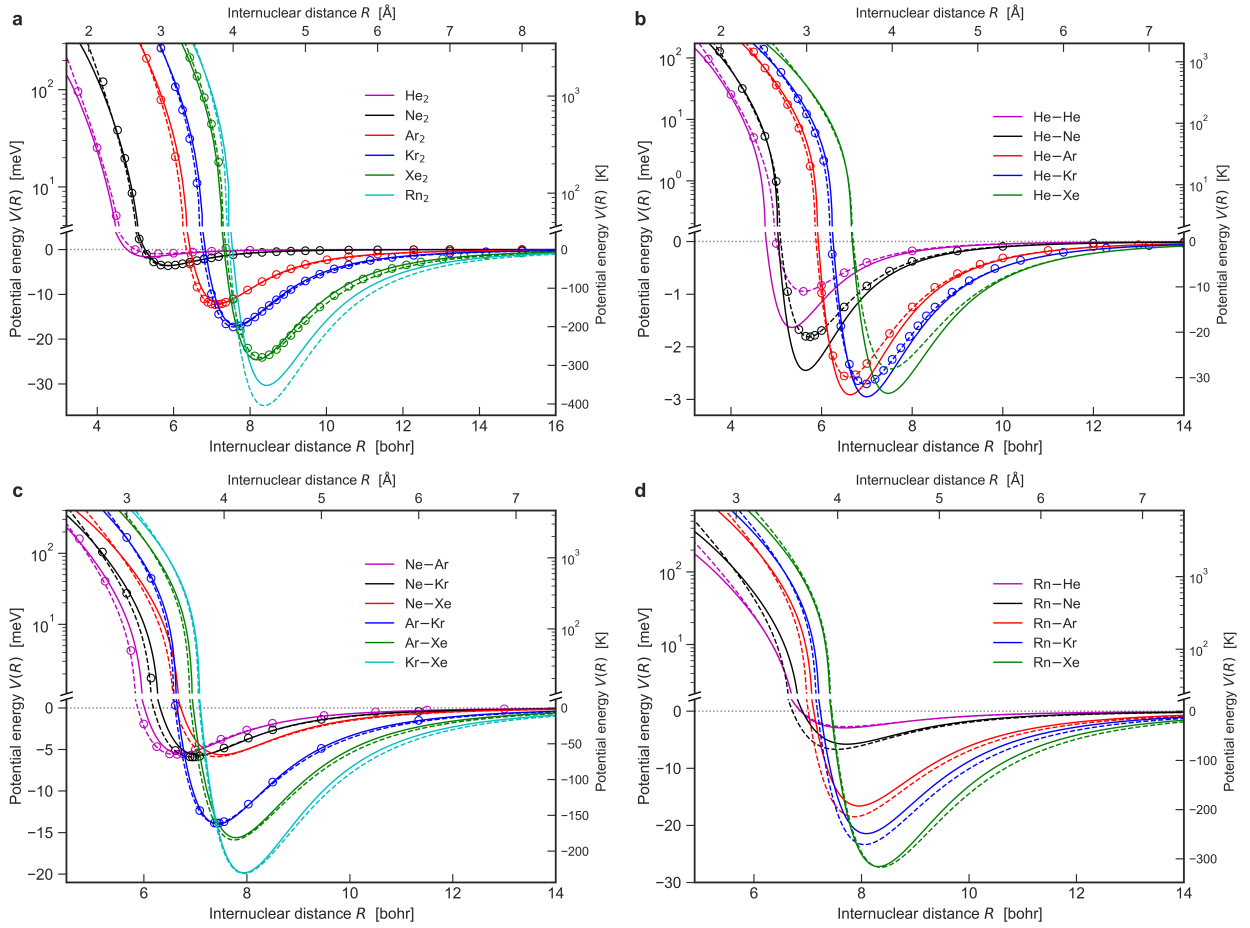

Figure 3.3: Potential curves of 21 noble-gas dimers. vdW-QDO potentials (solid lines) for **(a)** homonuclear and **(b-d)** heteronuclear noble-gas dimers benchmarked to the TTS potential [86] (dashed lines) and the reference CCSD(T) calculations (circles) [197, 207–212]. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

To extend the developed potential to heteronuclear dimers, combination rules for potential parameters can be used. The simplest ones are given by

$$R_e^{AB} = \left(R_e^A + R_e^B\right)/2\,, \quad D_e^{AB} = \sqrt{D_e^A D_e^B} \tag{3.18}$$

and known as the Lorentz-Berthelot rules, which are often used for the LJ potential and implemented in many molecular simulation packages [27–29]. However, the Lorentz-Berthelot rules are not accurate [102, 214–216]. Therefore, instead of mixing R_e and D_e, we use mixing rules for α_1 and C_6, since our potential is fully parametrized by these two quantities. With the effective mixed values $\{\alpha_1^{AB}, C_6^{AB}\}$, we can set three oscillator parameters $\{q, m, \omega\}$ through the same vdW-OQDO parametrization procedure as for homonuclear dimers (Appendix A2). By doing so, even the heteronuclear dimer AB is effectively represented by two *identical* oscillators, which still allows us to apply the formalism for exchange repulsion, Eqs. (3.3)–(3.4), developed for homonuclear dimers. For C_6 dispersion coefficient, the very accurate combination rule arising from the London formula is already well known [85, 145]

$$C_6^{AB} = \frac{2\alpha_1^A \alpha_1^B C_6^A C_6^B}{C_6^A (\alpha_1^B)^2 + C_6^B (\alpha_1^A)^2}\,. \tag{3.19}$$

To combine polarizabilities, we employ the robust mixing rule for vdW radii which was established in Ref. [102], where it was shown that accurate equilibrium distances in noble-gas dimers

(MARE = 1%) are delivered by

$$R_e^{AB} = 2 \times \Phi^{-1/7} \left[\left(\alpha_1^A + \alpha_1^B \right) / 2 \right]^{1/7} , \tag{3.20}$$

similar to the homonuclear case (3.8). Thus, the effective polarizability α_1^{AB} can be simply represented by

$$\alpha_1^{AB} = \left(\alpha_1^A + \alpha_1^B \right) / 2 . \tag{3.21}$$

Thereby, combining Eqs. (3.14), (3.15), (3.17), (3.19) and (3.21) with the vdW-OQDO parametrization scheme (see Appendix A2), we obtain vdW-QDO potentials for all 21 noble-gas dimers. They are shown in Figure 3.3 with the excellent agreement to both TTS potential and reference CCSD(T) calculations for homo- and heteronuclear dimers of Ne, Ar, Kr, and Xe on panels (a) and (c), as well as for He–Ar, He–Kr (b) and Rn–Xe (d). The other Rn dimers are challenging for our model due to the discrepancies in D_e, as was discussed above. In the case of He dimers, the error is caused to a large extent by the underestimated R_e of He_2, with 5.35 Bohr predicted by Eq. (3.10) against the reference value of 5.608 bohr [207]. In addition, the actual error in the potential for He dimers is small in magnitude (despite being seemingly large visually due to the scale of y-axis in Figure 3.3b).

To evaluate the accuracy of our potential quantitatively, we introduce the normalized area difference metric Δ_S between tested and reference potentials as

$$\Delta_S = \frac{1}{R_e D_e} \int_{0.8R_e}^{2.0R_e} |V_{\text{ref}}(R) - V_{\text{test}}(R)| \, dR . \tag{3.22}$$

The essential physical meaning of Δ_S is illustrated in Figure 3.4b which shows that this single unitless number represents a measure of difference between tested and reference potentials. The integration limits are set to $0.8R_e$ and $2.0R_e$ to evaluate accuracy close to the minimum region, whereas the long-range accuracy can be evaluated separately in terms of dispersion coefficients. Previously, Δ-gauge was used in benchmarks of various density-functional codes in calculations of equations of state for solids [217, 218]. The TTS potential [86] was chosen as the reference potential, and we benchmark vdW-QDO and LJ1 potentials relative to it (a similar benchmark for LJ2 can be found in Appendix A3). The computed Δ_S-matrices for 15 He – Xe dimers are displayed in Figure 3.4c-d (Rn dimers are omitted since there are no LJ parameters for Rn available in the literature). We note that the vdW-QDO potential has twice better accuracy than the LJ one with $\langle \Delta_S^{\text{QDO}} \rangle$ = 9.0% compared to $\langle \Delta_S^{\text{LJ1}} \rangle$ = 18.4% and $\langle \Delta_S^{\text{LJ2}} \rangle$ = 17.3% when averaged over 15 dimers. Helium is the obvious outlier for vdW-QDO with the highest Δ_S = 31.1%, whereas for all other dimers Δ_S is below 13.6%. In contrast, LJ potential shows much broader variations in Δ_S spanning from 3.7% for Ar–Kr to 43.2% for Xe_2. Although the LJ potential shows better Δ_S values than vdW-QDO for some of the dimers (He_2, Ne_2, He–Ne, Ne–Ar, Ar–Kr), overall the performance of vdW-QDO potential is more accurate and robust. Generally, Figure 3.4 supports the above conclusions about the accuracy of vdW-QDO potential based on Figure 3.3. We note that the predictions of LJ potential become worse for heteronuclear dimers composed of small and large atoms (e.g. He–Ar, He–Kr, Ne–Kr) than for atoms with a relatively close size (Ne–Ar, Ar–Kr) (Figure 3.4c). In contrast, the evenly accurate predictions of the vdW-QDO model (Figure 3.4d) suggest that the combination rules (3.19) and (3.21) employed in this work are more accurate and robust than the Lorentz-Berthelot rules (3.18).

To evaluate the quality of the potentials in the long-range limit, in Figure 3.4a we compare the dispersion coefficients predicted by vdW-QDO and TTS potentials to the reference *ab initio* values (Table 3.1). Such a comparison is fair since both potentials are built as conformal

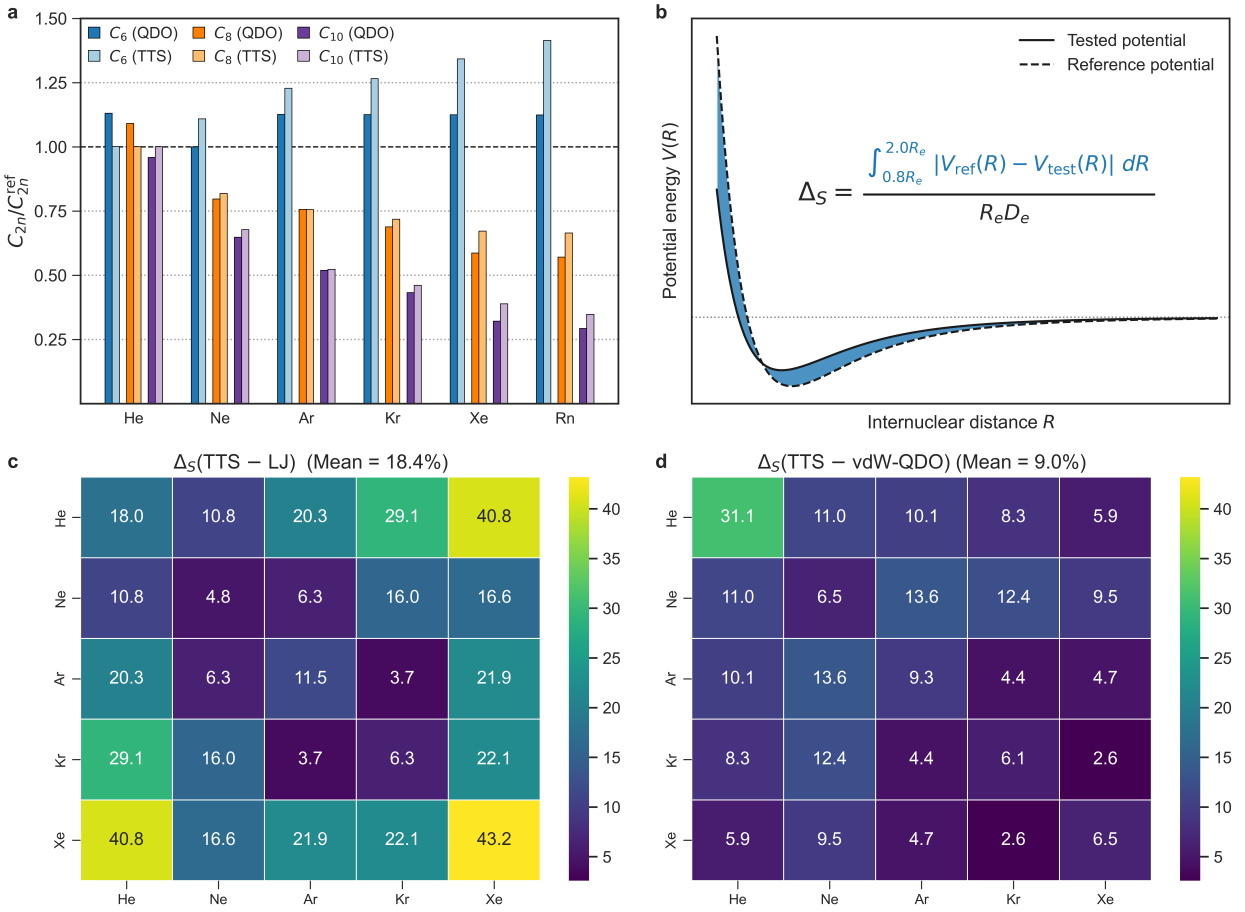

Figure 3.4: Assessment of the vdW-QDO potential accuracy. **(a)** Errors in dispersion coefficients predicted by vdW-QDO (dark colors) and TTS (light colors) potentials. **(b)** Schematic illustration of the Δ_S metric calculation. **(c,d)** Heatmaps showing Δ_S (in %) for LJ1 (left) and vdW-QDO (right) potentials relative to the reference TTS potential. The left and right colorbars have the same scale. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

ones, unlike the earlier TT-2003 potentials [85], which directly utilize reference C_{2n} coefficients for every element being therefore not strictly conformal. To recover the TTS dispersion coefficients, we used the reported scaling law $C_{2n} = C_{2n}^* \times D_e R_e^{2n}$ with $C_6^* = 1.3499$, $C_8^* = 0.4147$, $C_{10}^* = 0.1716$ [86]. For the vdW-QDO dispersion coefficients, from Eqs. (3.14), (3.15) and (3.17) we obtain

$$C_{2n} = C_{2n}^* \times \frac{C_6}{R_e^{2n-6}} \left[1 - \frac{\beta - 5}{\beta(\beta+1)} \right], \tag{3.23}$$

where $C_6^* = 1.1779$, $C_8^* = 0.3848$, $C_{10}^* = 0.1540$ (see Table 3.2). We found that both potentials severely underestimate C_8 and C_{10} and demonstrate the similar magnitude of these errors increasing with the atomic number. However, for C_6 vdW-QDO potential performs much better, showing a homogeneous overestimate of 12-13%, whereas the TTS potential again possesses an increasing magnitude of error, reaching its maximum of 41% for Rn. While C_8 and C_{10} are important to deliver accurate potential near the equilibrium, in the asymptotic limit the quality of the potential is fully determined by the leading C_6 coefficient. Therefore, we can conclude that our conformal vdW-QDO potential shows physically more reasonable long-range behavior than the conformal TTS potential.

3.2.1 Damped vdW-QDO Potential

In contrast to the Tang-Toennies potentials [85, 86], the above vdW-QDO model does not employ any damping of the dispersion energy. In fact, for noble-gas dimers damping the dispersion energy is not essential, and interatomic vdW potential can be effectively described even without a damping function, as was shown above. This provides additional reasoning why the scaling law for vdW radius (3.8), originally derived without damping [102], works so well. To uncover the effect of the damping function on the results, we derived the damped vdW-QDO potential (*vide infra*), where the QDO damping function reads

$$f_{2n}(z) = 1 - e^{-z} \sum_{k=0}^{n} \frac{z^k}{k!} \,, \quad z = \frac{(\gamma R)^2}{2} \,. \tag{3.24}$$

Interestingly, the damping function (3.24) differs from the TT damping function (2.84) only in the upper summation limit (for the TT it is $2n$) and in the physical meaning of the unitless variable ($z = bR$ for the TT damping function, with b stemming from the Born-Mayer repulsion term Ae^{-bR}). Note that, due to the distinct form of the Pauli repulsion in the vdW-QDO and TT models, the QDO damping function contains only even powers of R up to $2n$. The derivation of the QDO damping function and a more detailed discussion are presented in Appendix A1.

The derivation of the damped vdW-QDO potential is similar to the undamped one, with the only difference in the form of dispersion energy and hence dispersion force. The force-balance equation in the dipole approximation reads [102]

$$\frac{k_e q^2}{2} \left[\frac{1}{R_e^2} + \frac{m\omega}{\hbar} \right] e^{-\frac{m\omega}{2\hbar} R_e^2} = \frac{6 f_6(R_e) C_6}{R_e^7} - \frac{f_6'(R_e) C_6}{R_e^6} \,. \tag{3.25}$$

Incorporating C_8 and C_{10} terms and the constant A_d, we get the full force-balance equation

$$A_d k_e q^2 \left[\frac{1}{R_e^2} + \frac{m\omega}{\hbar} \right] e^{-\frac{m\omega}{2\hbar} R_e^2} = \sum_{n=3}^{5} \left(\frac{2n f_{2n}(R_e) C_{2n}}{R_e^{2n+1}} - \frac{f_{2n}'(R_e) C_{2n}}{R_e^{2n}} \right) \,. \tag{3.26}$$

Solving Eqs. (3.25) and (3.26) together, one can express A_d by

$$A_d = \left(\frac{1}{2} + \frac{C_8}{2C_6 R^2} \cdot \frac{8 f_8 - R f_8'}{6 f_6 - R f_6'} + \frac{C_{10}}{2C_6 R^4} \cdot \frac{10 f_{10} - R f_{10}'}{6 f_6 - R f_6'} \right) \Bigg|_{R=R_e} \tag{3.27}$$

and write the damped vdW-QDO potential as

$$V_{\mathrm{QDO}}^{\mathrm{damp}}(R) = \frac{A_d k_e q^2}{R} e^{-\gamma R^2/2} - \sum_{n=3}^{5} f_{2n}(R) \frac{C_{2n}}{R^{2n}} \,. \tag{3.28}$$

Similar to the undamped case, we use Ne_2 to obtain the dimensionless shape of the potential

$$U_{\mathrm{QDO}}^{\mathrm{Ne,damp}}(x) = \frac{A_d^*}{x} e^{-\frac{(\gamma^* x)^2}{2}} - \sum_{n=3}^{5} f_{2n}(\gamma^* x) \frac{C_{2n}^*}{x^{2n}} \,, \tag{3.29}$$

with the starred parameters for this equation defined in Table 3.3. For other dimers, the potentials are obtained via rescaling

$$V_{\mathrm{QDO}}^{\mathrm{damp}}(R) = D_e U_{\mathrm{QDO}}^{\mathrm{Ne,damp}}(x) \,, \quad x = R/R_e \,, \tag{3.30}$$

where for D_e we use Eq. (3.17) as for the undamped potential. The above derivation is not specific to noble gases, remaining valid also for group II dimers. The corresponding starred parameters are listed in Table 3.5.

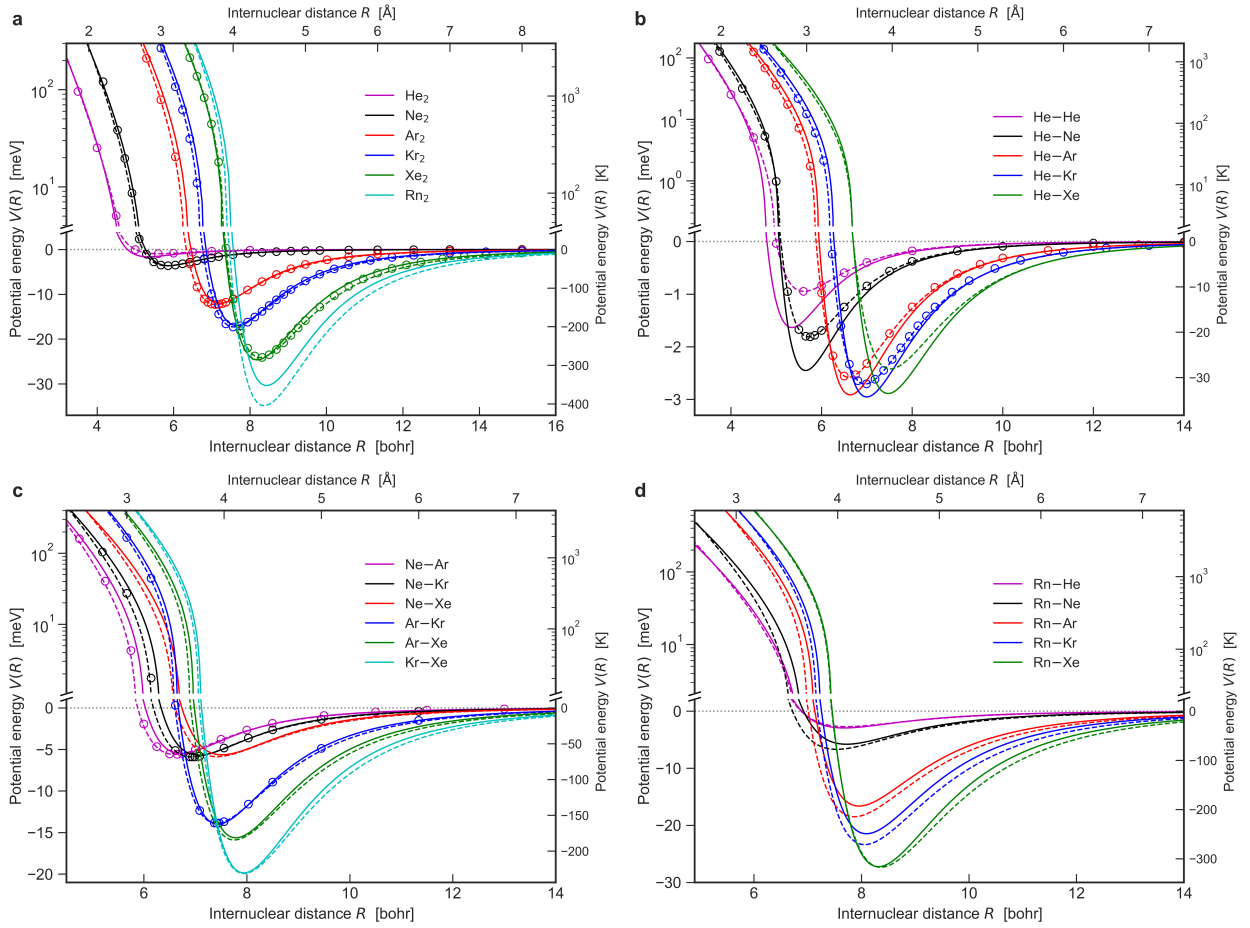

Figure 3.5: Damped vdW-QDO potentials of noble-gas dimers. The same as Figure 3.3 but for the damped vdW-QDO potential (3.28). Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

Figure 3.5 shows the obtained damped vdW-QDO potential curves for noble-gas dimers compared to the TTS potential [86] and reference CCSD(T) calculations. One can see almost the same curves as in the undamped case, having a perfect agreement with the references. The only noticeable differences to the undamped case can be observed in the short-range region.

Table 3.3: The damped vdW-QDO potential parameters for noble gases in Eq. (3.29). The parameters of Ne_2 dimer used in the second column are $R_e = 5.875$ bohr and $D_e = 3.620$ meV [Eq. (3.28)]. The damped QDO parameters for Ne_2 are $q_d = 1.20299$, $m_d = 0.38066$, $\omega = 1.19326$ (in a.u.).

Parameter	Definition	Numerical value
A_d^*	$A_d k_e q_d^2 / R_e D_e$	1415.607
γ^*	$R_e \sqrt{m_d \omega / \hbar}$	3.959
C_6^*	$C_6 / D_e R_e^6$	1.1667
C_8^*	$5C_6 / D_e R_e^6 (\gamma^*)^2$	0.3721
C_{10}^*	$245C_6 / 8D_e R_e^6 (\gamma^*)^4$	0.1454

3.2.2 Comparison of vdW-QDO and TT Models Against SAPT

We also observe that for both the TT and vdW-QDO models exchange repulsion and dispersion terms separately do not agree with their SAPT counterparts, whereas the total potentials show very close agreement with the sum of the SAPT terms. This finding is in line with the statement

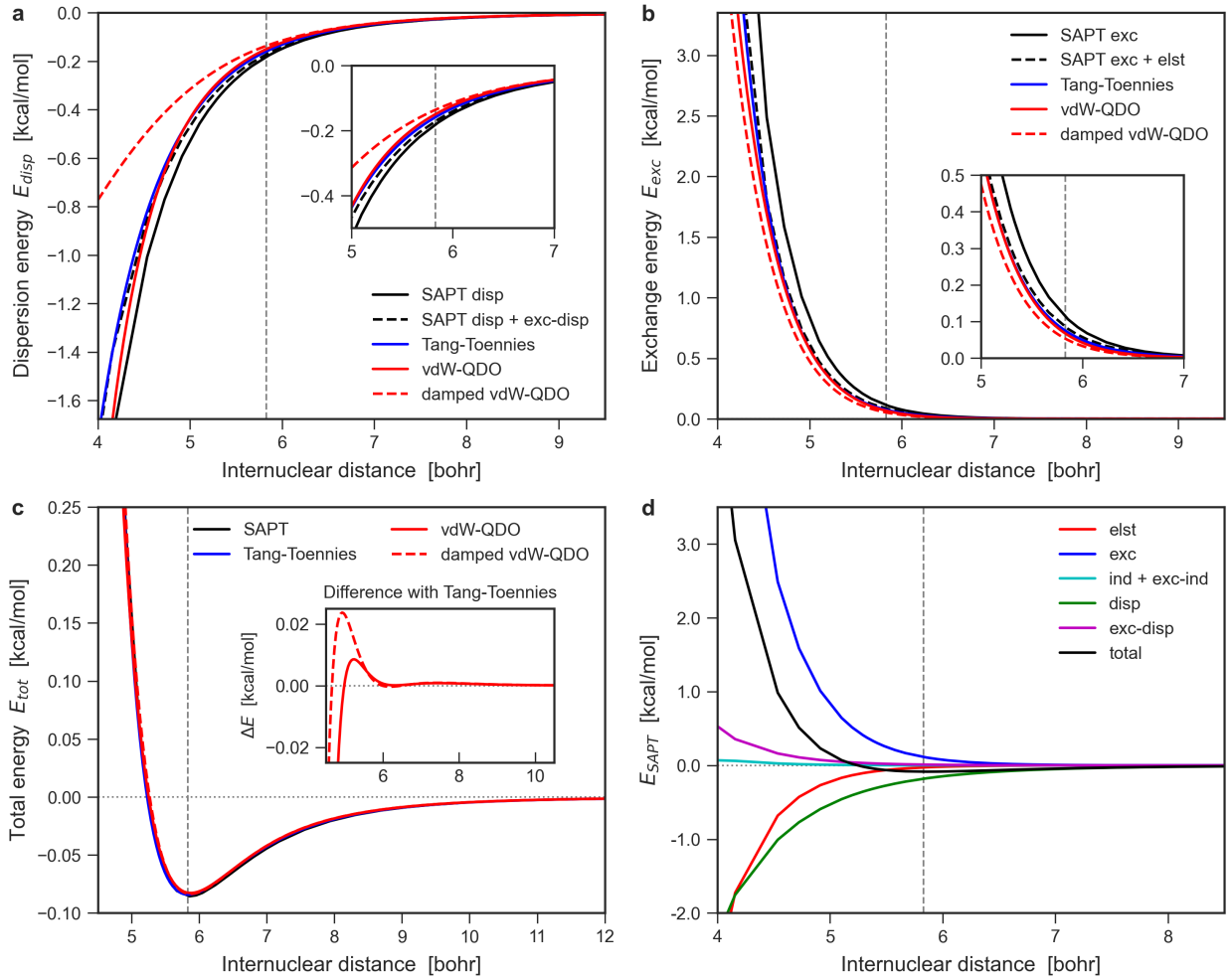

Figure 3.6: Comparison of SAPT, TT and vdW-QDO energy contributions. **(a)** Dispersion and **(b)** exchange repulsion components of the vdW-QDO (red) and Tang-Toennies [85] (blue) potentials compared to the SAPT energy contributions (black) for neon dimer. The comparison of total interaction curves is shown in **(c)**. The SAPT data was taken from Ref. [8]. Individual SAPT contributions for neon are displayed in **(d)**. The insets in **(a)** and **(b)** show zoom-in to near-equilibrium distances. The inset in **(c)** is the discrepancy between Tang-Toennies potential and undamped/damped vdW-QDO potentials. Everywhere the vertical dashed line denotes the equilibrium distance. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

of Ref. [92] that the generalized Heitler-London theory delivers a more compact expansion of interaction energy than the SAPT.

Within the SAPT framework [37, 38], the interaction energy of two noble-gas atoms can be represented as the sum of four contributions

$$E_{\text{int}}^{\text{SAPT}} = E_{\text{elst}}^{(1)} + E_{\text{exc}}^{(1)} + E_{\text{disp}}^{(2)} + E_{\text{exc-disp}}^{(2)}\,, \tag{3.31}$$

since induction and exchange-induction terms perfectly cancel out each other delivering $E_{\text{ind}}^{(2)} + E_{\text{exc-ind}}^{(2)} \approx 0$ (see Figure 3.6d). On the other hand, the Tang-Toennies model relies on a more compact decomposition of the interaction energy, consisting only of exchange repulsion and dispersion. Nevertheless, the total interaction energies obtained through the TT model and SAPT are practically the same for noble-gas dimers. This interesting fact has been already discussed in literature [92]. Similar to the TT model, our vdW-QDO potential also relies on the balance between exchange repulsion and dispersion. Therefore, we find it useful to report here

a simple analysis, which elucidates the reasons behind the high accuracy of the vdW-QDO and TT models.

In Figure 3.6a-b, we separately compare the dispersion and exchange repulsion terms of vdW-QDO and TT potentials for the neon dimer. A close agreement between the undamped vdW-QDO (red solid curve) and TT (blue curve) potentials is observed for both terms, with the discrepancy appearing more pronounced only at shorter distances. In contrast, the damped vdW-QDO dispersion and exchange repulsion (red dashed curves) are significantly different, although their sum shows a perfect agreement with undamped vdW-QDO and TT potentials (Figure 3.6c). In addition, we compare the energy terms of vdW-QDO and TT models to the corresponding energy contributions from the SAPT-CCSD calculations [8] (black solid curves). A noticeable difference for both dispersion and exchange repulsion energies can be spotted. In other words, the dispersion energy from the vdW-QDO and TT potentials does not *exactly* reproduce $E^{(2)}_{\text{disp}}$ curve from SAPT, and the same is true also for their exchange repulsion parts versus $E^{(1)}_{\text{exc}}$. However, if instead one compares to the sums $E^{(1)}_{\text{elst}} + E^{(1)}_{\text{exc}}$ and $E^{(2)}_{\text{disp}} + E^{(2)}_{\text{exc-disp}}$, then much closer agreement with the vdW-QDO and TT energy decomposition is observed, as illustrated by black dashed lines.

Thus, we conclude that both vdW-QDO and TT potentials rely on the effective description of exchange repulsion and electrostatics by the *exchange term* and on the effective description of dispersion and exchange-dispersion interactions by the *dispersion term* in the potential. We believe that this observation is important as it embeds the designed accurate interatomic potentials into the physical picture of noncovalent interactions.

3.3 Application to Group II Dimers

Another class of vdW systems where Tang-Toennies potentials work well consists of Me_2 dimers, with Me = Mg, Ca, Sr, Ba, Zn, Cd, and Hg. Although such systems are not purely vdW-bonded, it was demonstrated that their interatomic potentials can be also well described by the TT potential [93, 176, 219–223]. Moreover, potentials of the group II dimers also obey the principle of corresponding states, albeit the underlying potential shape is distinct from that of noble-gas dimers [93, 221, 222]. The only exception is the Be-Be dimer, which has been a long-standing puzzle for quantum chemistry. The potential curve of Be_2 has a remarkably different shape in the long-range region, compared to other alkaline-earth elements [93]. Since this dimer does not obey the principle of corresponding states, it is excluded from our consideration here. In what follows, we show that the vdW-QDO potential is also capable of describing the potential curves of the group II dimers upon several modifications.

Table 3.4: The reference parameters of group II dimers: dipole polarizability α_1 and C_6 dispersion coefficient (in a.u.), as well as R_e (in bohr) and D_e (in meV), for which the values in Å and Kelvin, respectively, are extra given in parentheses. The used data sources: [a] Ref. [224], [b] Ref. [225], [c] Ref. [226], [d] Ref. [222], [e] Ref. [227], [f] Ref. [228].

	α_1	C_6	R_e (in Å)	D_e (in K)
Mg_2	71.3[a]	627[a]	7.35 (3.89)[b]	53.81 (624.4)[b]
Ca_2	157.1[a]	2121[a]	8.13 (4.30)[b]	130.18 (1511)[b]
Sr_2	197.2[a]	3103[a]	8.88 (4.70)[b]	129.69 (1505)[b]
Ba_2	273.5[a]	5160[a]	9.43 (4.99)[b]	169.36 (1965)[b]
Zn_2	38.67[c]	359[d]	7.23 (3.83)[e]	28.64 (332.4)[e]
Cd_2	46[c]	686[d]	7.32 (3.87)[e]	40.91 (474.8)[e]
Hg_2	33.9[c]	392[d]	6.95 (3.68)[f]	48.60 (564.0)[f]

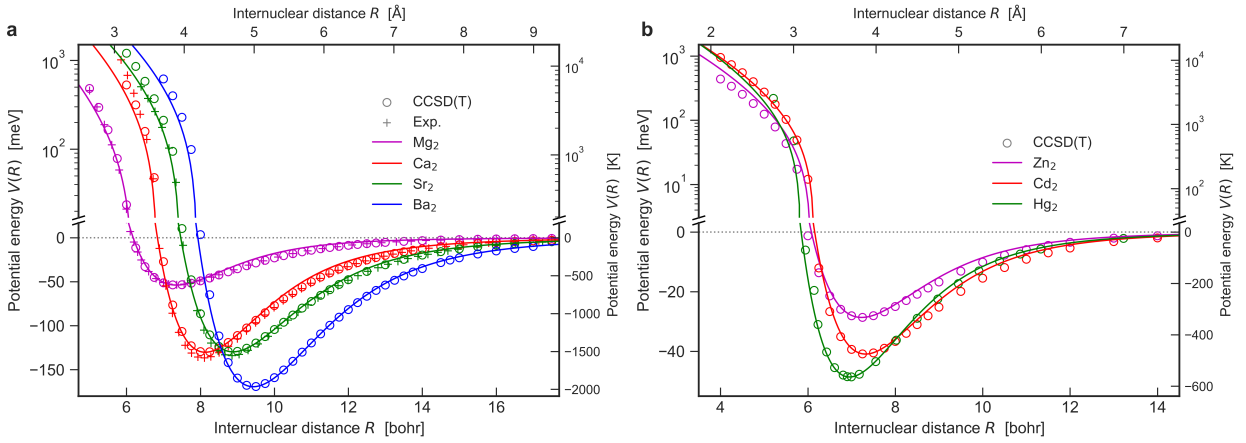

Figure 3.7: Potential curves of group II dimers: (a) Mg_2, Ca_2, Sr_2, Ba_2 and **(b)** Zn_2, Cd_2, Hg_2. The vdW-QDO potentials are shown by solid lines, circles mark coupled-cluster calculations [225, 227, 228], and crosses display experimental potential curves [229–231]. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

First, in contrast to noble gases, for Me_2 dimers the damping function (3.24) cannot be omitted due to much larger polarizabilities and hence dispersion coefficients than those of noble gases (see Table 3.4). As a result, without the damping function, a pronounced divergence of the vdW-QDO potential would already occur at near-equilibrium distances. Second, the scaling laws of Eqs. (3.10) and (3.17) are not so accurate for the group II dimers, since the bonding in Me_2 is not purely of vdW type [93, 221]. Therefore, the reference values of R_e and D_e (Table 3.4) were used for our vdW-QDO potential.

Following Tang *et al.* [93], we choose Sr_2 dimer as the reference system to get the shape of the potential curve and then rescale it onto other dimers. Similar to the case of noble gases, the damped vdW-QDO potential shape for Sr_2 dimer was derived as

$$U_{\mathrm{QDO}}^{\mathrm{Sr}}(x) = \frac{A_d^*}{x} e^{-\frac{(\gamma^* x)^2}{2}} - \sum_{n=3}^{5} f_{2n}(\gamma^* x) \frac{C_{2n}^*}{x^{2n}} , \tag{3.32}$$

with the numerical values of its parameters presented in Table 3.5. The QDO parameters $\{q, m, \omega\}$ for the Sr_2 dimer were set using α_1, C_6 and R_e following the damped vdW-OQDO procedure, as explained in Appendix A2. Altogether, the three physical quantities are employed to parametrize the vdW-QDO potential for Sr_2, compared to the five $\{R_e, D_e, C_6, C_8, C_{10}\}$ in case of the Tang-Toennies potential [93]. For other Me_2 dimers, we need only R_e and D_e from Table 3.4 to perform the rescaling

$$V_{\mathrm{QDO}}(R) = D_e U_{\mathrm{QDO}}^{\mathrm{Sr}}(x) , \quad x = R/R_e . \tag{3.33}$$

Parameter	Definition	Numerical value
A_d^*	$A_d k_e q^2 / R_e D_e$	58.051
γ^*	$R_e \sqrt{m\omega/\hbar}$	2.992
C_6^*	$C_6 / D_e R_e^6$	1.6209
C_8^*	$5C_6 / D_e R_e^6 (\gamma^*)^2$	0.9053
C_{10}^*	$245C_6 / 8D_e R_e^6 (\gamma^*)^4$	0.6194

Table 3.5: The damped vdW-QDO potential parameters for group II elements in Eq. (3.32). The Sr_2 dimer parameters used in the second column are D_e = 129.7 meV and R_e = 8.88 bohr [225]. The QDO parameters for Sr_2 dimer are q = 1.5433, m = 1.0671, ω = 0.1064 (in a.u.).

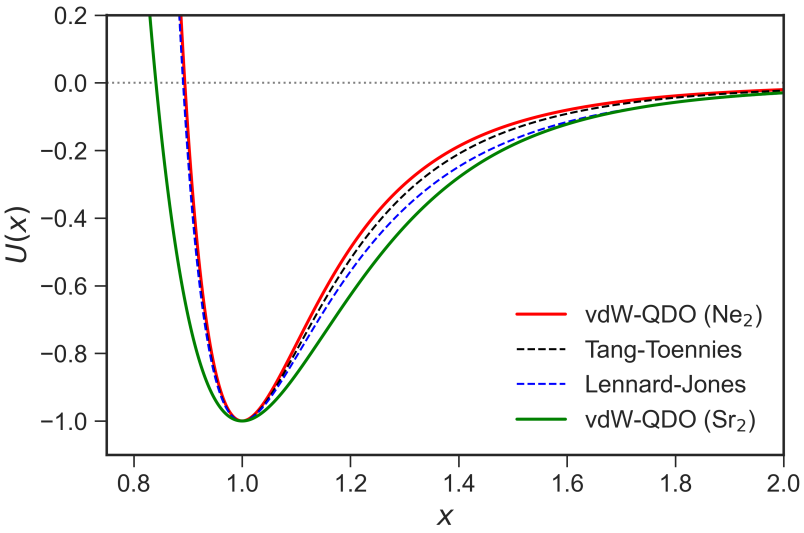

Figure 3.8: Dimensionless shape of vdW potential curves. vdW-QDO shapes for Ne_2 (red) and Sr_2 (green) compared to the shapes of the LJ (dashed blue) and TTS (dashed black) potentials. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

The results in Figure 3.7 show that vdW-QDO potentials are in excellent agreement with both *ab initio* and experimentally derived potentials (when they are available). This is a remarkable result since the binding energies of group II dimers are up to 5 times larger than those of the heaviest noble gases. Furthermore, the shape of their potentials is distinct from the one of noble gases (Figure 3.8). Thus, the vdW-QDO functional form is robust and well-suited to describe vdW potentials across various types of systems. In contrast, the LJ potential cannot be employed to describe group II dimers with any combination of parameters, since its energy well is too narrow for such binding curves (Figure 3.8).

3.4 vdW-QDO Potential Generalized to Molecules

We thank Dr. L. Medrano Sandonas for sharing his DFTB3+MBD results on molecular dimers.

Finally, we can show that the developed vdW-QDO potential is also applicable to molecular systems, with an example of eight dispersion-dominated molecular dimers from the S66×8 benchmark dataset [182], including benzene as well as aliphatic and cyclic molecules (see the full list in Appendix A4). Such systems were chosen to diminish the influence of electrostatic terms not included in the current vdW-QDO potential.

We compute the energy of vdW interaction between molecules A and B at a given intermolecular separation as

$$V_{\mathrm{QDO}}^{\mathrm{vdW}}(A,B) = V_{\mathrm{QDO}}^{\mathrm{exc}} + V_{\mathrm{QDO}}^{\mathrm{disp}} = \sum_{i\in A}\sum_{j\in B} V_{\mathrm{QDO}}^{ij}(R_{ij}) \,, \tag{3.34}$$

where summation goes over the atoms i and j of the molecules A and B, respectively, and R_{ij} is the interatomic distance. Interaction energy between a pair (i,j) is given by the damped vdW-QDO potential

$$V_{\mathrm{QDO}}^{ij}(R_{ij}) = A_d \frac{k_e q^2}{R_{ij}} e^{-\frac{(\gamma R_{ij})^2}{2}} - \sum_{n=3}^{5} f_{2n}(\gamma R_{ij}) \frac{C_{2n}}{R_{ij}^{2n}} \,. \tag{3.35}$$

To set the QDO parameters $\{q, m, \omega\}$, we apply the vdW-OQDO procedure coupled with the atom-in-molecule (AIM) approach to each pair of atoms (i,j). Following the TS method [39],

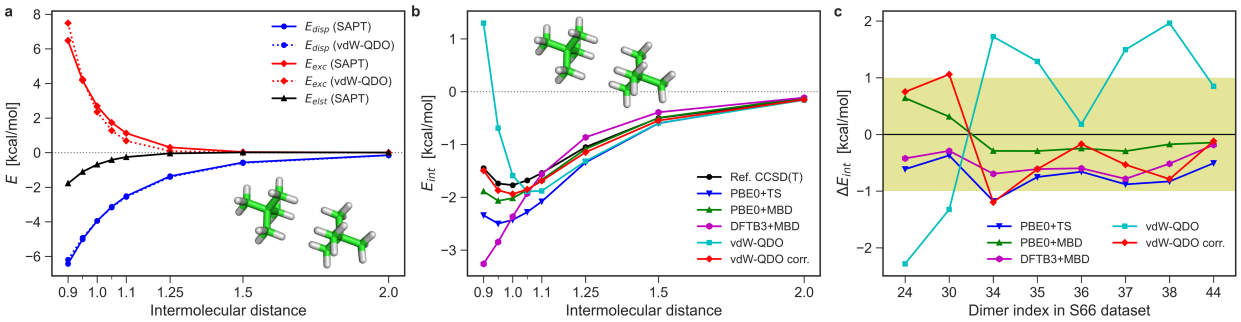

Figure 3.9: Interaction energy curves for a neopentane dimer and statistics for the selected S66×8 dimers. **(a)** Dispersion (blue) and exchange (red) contributions to the interaction energy of neopentane dimer (shown as inset) calculated by SAPT-DFT [234] (solid lines) and damped vdW-QDO potential of Eq. (3.34) (dotted lines). In addition, the electrostatic term from SAPT-DFT is displayed in black. **(b)** Binding energy curves of neopentane dimer as calculated by different methods: CCSD(T) [182] (black); PBE0+TS (blue) and PBE0+MBD (green); DFTB3+MBD (magenta); damped vdW-QDO potential (cyan); SAPT-corrected vdW-QDO potential (3.37) (red). **(c)** Errors in interaction energy of eight dispersion-dominated dimers for the five methods considered. Yellow filling depicts the "chemical accuracy" region of ±1 kcal/mol error. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

the AIM polarizabilities and dispersion coefficients are defined by

$$\alpha_{1,i}^{\mathrm{AIM}} = \alpha_{1,i}^{\mathrm{free}} \left(\frac{V_i^{\mathrm{AIM}}}{V_i^{\mathrm{free}}} \right), \; C_{6,i}^{\mathrm{AIM}} = C_{6,i}^{\mathrm{free}} \left(\frac{V_i^{\mathrm{AIM}}}{V_i^{\mathrm{free}}} \right)^2, \tag{3.36}$$

where V_i^{free} and V_i^{AIM} are the corresponding Hirshfeld volumes. To compute them, single-point DFT-PBE0 [72] calculations for every dimer were performed at their equilibrium geometry. Then, the effective polarizability α_1^{ij} and dispersion coefficient C_6^{ij} for a pair (i,j) were defined using the combination rules of Eqs. (3.19) and (3.21). Finally, the vdW-QDO parametrization procedure was applied to map $\{\alpha_1^{ij}, C_6^{ij}\}$ onto $\{q,m,\omega\}$. After performing pairwise summation in Eq. (3.34) and repeating the whole procedure for all 8 intermolecular separations, the vdW-QDO interaction curves for dimers are obtained and compared to the reference CCSD(T) interaction curves [182]. For comparison, interaction energies of dimers at PBE0+TS [39], PBE0+MBD [40, 41], and DFTB3+MBD [232] levels of theory were also calculated. All DFT calculations in this work were performed using FHI-aims code [233] with the 'tight' atomic basis sets.

We showcase the obtained results with an example of a neopentane dimer (Figure 3.9b). One can see that PBE0+TS overbinds neopentane dimer significantly and underestimates the equilibrium separation by 5%. The inclusion of many-body effects at the PBE0+MBD level improves the results of PBE0+TS for energy, but the predicted equilibrium separation is still 95% of the reference value. On the other hand, the DFTB3+MBD method clearly lacks repulsion and attraction at short-range and long-range distances, respectively, although at the equilibrium distance, the two errors largely cancel each other. As for the vdW-QDO potential, we note that the minimum of the intermolecular interaction curve is very close to the reference CCSD(T) energy, although our method overestimates the equilibrium separation by 10%. When considering the overall interaction curve, however, the vdW-QDO potential deviates significantly from the reference, being too repulsive at shorter distances. These conclusions remain valid also for the other seven dimers, as supported by the corresponding results in Appendix A4.

To get insights into such a behavior, we compared the exchange and dispersion terms in vdW-QDO potential (3.34) to their counterparts from SAPT-DFT calculations [234] (see Figure 3.9a). The dispersion part of vdW-QDO potential provides excellent agreement with $E^{(2)}_{\text{disp}}$ from SAPT-DFT, which illustrates the well-known fact that vdW dispersion interaction between small molecules can be effectively described by a pairwise potential, despite being essentially many-body in its nature [11, 25, 40]. In contrast, for the vdW-QDO exchange repulsion term, we observe a noticeable deviation from the $E^{(1)}_{\text{exc}}$ contribution of SAPT-DFT. This might indicate that exchange repulsion between molecules is not accurately described by the commonly used pairwise approach and hence exchange repulsion requires many-body treatment as well [235].

SAPT-DFT calculations [234] indicate also that electrostatic contributions are small but not negligible even for dispersion-dominated dimers (see Figure 3.9a and Figures A3–A5 in Appendix A4). To check a possible effect of the inclusion of accurate many-body exchange and electrostatic interactions on our results, we consider the corrected vdW-QDO potential, where the first-order SAPT-DFT energy was added to the dispersion energy from the vdW-QDO model

$$V^{\text{corr}}_{\text{QDO}}(A,B) = E^{(1)}_{\text{elst}} + E^{(1)}_{\text{exc}} + V^{\text{disp}}_{\text{QDO}} \,. \tag{3.37}$$

This approach is similar to the HFDc$^{(1)}$ scheme of Podeszwa *et al.* [236] with the important difference that here the dispersion energy is calculated from the QDO model, whereas in the HFDc$^{(1)}$ method this energy is computed from SAPT. The corrected vdW-QDO curve in Figure 3.9b delivers a much better description of the interaction energy than the original vdW-QDO method. Notably, at larger distances $V^{\text{corr}}_{\text{QDO}}$ also shows a good agreement with PBE0+MBD energy.

The overall statistics for the eight molecular dimers are shown in Figure 3.9c in terms of the error in the equilibrium energy obtained by five methods relative to the reference CCSD(T) values [182]. Purely analytical vdW-QDO predictions are scattered within the 2 kcal/mol range, which is remarkable considering the approximations made in the model. Including the first-order SAPT energy reduces errors roughly to 1 kcal/mol (chemical accuracy). This test demonstrates that the vdW-QDO approach is not limited to atomic dimers and can be generalized to molecular complexes, although in that case the accuracy is currently lower. Nevertheless, considering the approximations made, the fact that the vdW-QDO method (even without SAPT corrections) properly predicts binding of the considered weakly-bound molecular complexes from a set of AIM quantities (α_1 and C_6) is reassuring.

3.5 Summary and Outlook

We developed a universal pairwise vdW potential devoid of empiricism and parameterized by only two atomic non-bonded parameters. The developed vdW-QDO potential combines the strengths of the Lennard-Jones and Tang-Toennies models. Similarly to the former, the vdW-QDO potential is fully determined by only two parameters. At the same time, our potential is comparable in accuracy to the Tang-Toennies potential for noble-gas dimers, being twice as accurate as the Lennard-Jones one. Moreover, the vdW-QDO potential has advantages that are present neither in Lennard-Jones nor in Tang-Toennies models. First, the two parameters $\{\alpha_1, C_6\}$ are readily available for the whole periodic table [101, 226], being computed by highly accurate *ab initio* methods. This makes our potential widely applicable, as demonstrated for atomic dimers of group II elements as well as organic molecular dimers. Second, the conformal vdW-QDO potential has better long-range behavior than the Lennard-Jones and conformal

Tang-Toennies potentials. This is crucial for applications to extended systems, where errors in the long-range vdW energy accumulate over many atomic pairs.

The key idea behind the presented potential is the synergy between vdW scaling laws, coarse-grained QDO formalism for exchange repulsion, and the principle of corresponding states. In its current form, the vdW-QDO potential does not explicitly include the electrostatic contribution arising from charge penetration between atoms at short distances, which is non-negligible according to the SAPT. Although the short-range electrostatic interaction can be introduced into the QDO model in the form of Coulomb integral [100], this requires the introduction of an additional "electrostatic charge" parameter [109] into the vdW-QDO potential. Therefore, given the good accuracy of the current vdW-QDO potential, we decided to dispense with the explicit modeling of short-range electrostatics at this stage of the model development. In its current version, the vdW-QDO potential can be incorporated into classical force fields, as a non-empirical replacement for the LJ potential. For polar systems, an additional electrostatic/polarization term is needed in a force field (like it is done in the case of the LJ potential). Eventually, effort in this direction could deliver a general coarse-grained force field for all non-covalent interactions entirely based on the model system of coupled QDOs.

We consider the present vdW-QDO potential as an important step towards a new generation of universal vdW potentials to be used in force fields and for biomolecular applications. To extract atom-in-molecule parameters, we currently employ *ab initio* calculations, which is a certain limitation. One possible alternative is substituting Hirshfeld volumes with Mulliken charge populations extracted from DFTB calculations, which is shown to provide reliable approximation [152]. This is especially relevant for biomolecular systems like proteins, where DFT calculations become unaffordable. Moreover, active development of machine-learning models [153, 195, 237] paves the way to obtain atom-in-molecule partitioning parameters without any electronic-structure calculations. There is an increasing trend in creating extensive molecular datasets such as QM7-X [238] or Aquamarine [239], which include information about atom-in-molecule volumes. Such datasets can be used for training neural-network (NN) models predicting Hirshfeld volumes from atomic descriptors [153].

Very recently, the vdW-QDO potential was used to augment the SO3KRATES equivariant graph NN model to ensure physically correct behavior in the long-range limit [33]. In addition, the vdW-QDO potential was implemented by Thomas Plé within the FENNOL library for building, training, and running force-field-enhanced neural network potentials [240]. The code is open-source and available in [241], enabling the direct applicability of the vdW-QDO potential to arbitrary molecular systems together with other program modules as well as standalone. Finally, the vdW-QDO potential is also available in the modified version of the LAMMPS molecular dynamics simulator [242]. The code has not been made public yet but is available from the author of this thesis upon reasonable request.

Chapter 4

van der Waals Polarization of Electron Density

Parts of this chapter have been published in this or similar form in:
A. Khabibrakhmanov, M. Gori, C. Müller, A. Tkatchenko,
DOI: 10.26434/chemrxiv-2024-lf3nr, 2024.

The electronic charge density $\rho(\mathbf{r})$ is a fundamental observable that uniquely defines molecules and materials with non-degenerate electronic ground states [42, 43]. It plays a central role in density-functional approximations (DFAs), which are widely used in computational chemistry and materials science. Existing DFAs often produce reasonable estimates of the ground state energy, which makes them so useful in practice. However, the accuracy of DFAs in predicting energies has not necessarily translated into better predictions of $\rho(\mathbf{r})$. This trade-off was highlighted in the seminal work by Lyssenko *et al.* [243], revealing discrepancies between DFA-generated densities and near-exact *ab initio* references. Foundational works by Burke and co-authors on density-corrected DFAs found that evaluating DFAs on the Hartree-Fock (HF) density without self-consistency yields more accurate energies for important prototypical examples including e.g. stretched heteronuclear bonds or torsional barriers in flexible molecules [244–246]. Hence, accurate models for electron densities would benefit the developments of next-generation DFAs as well as machine-learned surrogate models for force fields [247–249] and other properties of atomistic systems [250–252] throughout physics, chemistry, and biology.

Large, polarizable systems require accurate treatment of non-local electronic correlations, such as van der Waals (vdW) dispersion, which affects both energies and $\rho(\mathbf{r})$ (and related properties). Nevertheless, most of the popular vdW methods, such as the DFT-D family by Grimme *et al.* [19, 127, 133] or exchange-hole dipole moment (XDM) model of Johnson and co-workers [20], treat vdW dispersion as *a posteriori* energy correction without accounting for the resulting vdW polarization of $\rho(\mathbf{r})$. While this may be sufficient for small molecules, it becomes inadequate for larger, more polarizable systems, where dispersion-driven stabilization is significant [25, 253]. Recent studies suggest that vdW interactions can induce substantial polarization in $\rho(\mathbf{r})$ for complex systems, e.g. π-stacked supramolecular dimers or molecules on a surface [23–26, 120], but a comprehensive assessment of the relative role of vdW-induced $\rho(\mathbf{r})$ polarization in comparison to semi-local (or hybrid) DFAs remains lacking.

Apart from being an experimental observable and a basic input to DFAs, $\rho(\mathbf{r})$ serves as a descriptor for various chemical analyses, including molecular electrostatic potential maps [254], electron localization functions [255, 256], and the Quantum Theory of Atoms in Molecules

(QTAIM) [257, 258]. The QTAIM formalism provides an inspiration for the widely used visualization of noncovalent interactions (NCI) [259, 260] based on the gradient and Hessian of $\rho(\mathbf{r})$. The NCI analysis [259] enables visualization of a broad variety of noncovalent interactions, such as hydrogen bonds and vdW dispersion. Typically, the NCI analysis relies on self-consistent DFA densities or even additive atomic densities [259, 261], neglecting vdW-induced polarization. Although regions contributing to dispersion interactions can be glanced from semi-local densities, the reliance on "vdW-lacking" densities has been recognized by the community as one of the key shortcomings [261, 262].

Here, we propose an efficient way to include vdW dispersion effects in $\rho(\mathbf{r})$ from a (semi-local) DFA. This can be accomplished by using the Many-Body Dispersion (MBD) [40] model, which has been successfully applied to describe anisotropic molecular polarizabilities [11, 94, 183, 184], polarization response of atoms and molecules to electric fields [110, 188, 189], optical excitonic spectra [168], and to accurately capture effects of polarization and vdW dispersion in (bio)molecules and materials [14, 25, 26, 99, 109, 120, 263]. We use the newly developed Fully-Coupled and Optimally-Tuned version of MBD (MBD@FCO) and validate this method by benchmarking against accurate *ab initio* calculations. We apply it to the S12L [264] and L7 [265] datasets of large organic dimers and show that the vdW interactions can displace charge by up to 80 % of the DFA-induced displacement or even exceed it, as in the case of alkane chains. We found that vdW polarization effects scale linearly with system size, highlighting their growing importance in extended molecules and materials. Accounting for vdW-induced polarization leads to striking differences in NCI isosurfaces as measured by their volume, which increases by several folds. The application of MBD@FCO to large (bio)molecules exemplified with the Fip35-WW protein allows us to identify smooth (as opposed to "patchy") interacting regions, facilitating chemical analysis of noncovalent interactions in these complex systems.

4.1 Methods

Computational Details

Unless otherwise noted, all DFT calculations in this chapter were performed using the Perdew-Burke-Ernzerhof (PBE) [63] functional as implemented in all-electron FHI-AIMS code [233], with 'tight' basis sets and integration grids. The Kohn-Sham self-consistent cycle was converged to 10^{-6} eV in energy and 10^{-5} electrons in density. The input α_0 and C_6 for the MBD Hamiltonian were obtained using the TS rescaling of the reference free-atom parameters [39], as shown in Eq. (2.141). The MBD@FCO calculations were carried out using the LIBMBD library [160]. For the PBE+XDM calculations, the Becke-Johnson damping parameters $a_1 = 0.6438$ and $a_2 = 1.8533$ Å corresponding to the 'tight' settings in FHI-AIMS [139] were utilized.

CCSD and HF calculations of electronic density were performed employing augmented correlation-consistent Dunning basis sets [266] with counterpoise correction and frozen-core approximation adopted for CCSD, and molecular symmetry was not utilized. Unless otherwise noted, PYSCF(v.2.6.2) [267] code was used for such calculations. In addition, auxiliary CCSD calculations using Q-CHEM(v.6.1) [268] and CCSDT calculations with MRCC program [269] were done. More technical details on coupled-cluster and HF calculations are given in Appendix A5.

Fully-Coupled and Optimally-Tuned Many-Body Dispersion Method

The MBD method efficiently captures many-body polarization and dispersion effects to infinite order in perturbation theory [40, 41] using coupled quantum Drude oscillators (QDOs) to model

collective electron density fluctuations. The MBD method maps atoms onto a model Hamiltonian of dipole-coupled QDOs centered at $\mathbf{R}_A$, with frequencies ω_A, charges q_A and masses m_A:

$$\hat{H}_{\mathrm{MBD}} = \sum_A \left[-\frac{1}{2}\nabla^2_{\xi_A} + \frac{1}{2}\omega_A^2 \xi_A^2 \right] + \frac{1}{2}\sum_{A\neq B} \omega_A \omega_B \sqrt{\alpha_{0,A}\alpha_{0,B}}\, \boldsymbol{\xi}_A \mathbf{T}_{AB}(|\mathbf{R}_A - \mathbf{R}_B|)\boldsymbol{\xi}_B\,, \tag{4.1}$$

where $\boldsymbol{\xi}_A = \sqrt{m_A}(\mathbf{r}_A - \mathbf{R}_A)$ are mass-weighted displacements of the oscillating charges, and $\mathbf{T}_{AB}$ is the dipole interaction tensor. To map atoms onto oscillators, we employ the optimized QDO parametrization (vdW-OQDO), designed to accurately describe the atomic response to electric fields[110] and vdW binding curves of atomic dimers [107]. This allows us to determine $\{m_A, \omega_A, q_A\}$ solely from two response properties of atoms: the static polarizability (α_0) and the dipolar dispersion coefficient (C_6) – both obtained from reliable free-atom reference data and semi-local electron density [39]. Please see Appendix A2 for more details on the vdW-OQDO parametrization.

Interoscillator coupling in MBD is introduced using the dipole-dipole interaction tensor $\mathbf{T}_{AB}$. The widely used MBD@rsSCS variant [41] employs empirical short-range damping in $\mathbf{T}_{AB}$, tuned for coupling with a specific DFA, see Eq. (2.170). The DFA+MBD@rsSCS approach is tailored to accurately reproduce total binding energies of dimers [41], however, the pure MBD@rsSCS dispersion energy is not comparable to *ab initio* reference (see Figure 4.1a). Instead, here we use the full dipole interaction:

$$\begin{aligned} \mathbf{T}^{ij}_{AB}(\mathbf{R},\sigma) &= \frac{\partial^2}{\partial R_i \partial R_j}\frac{\mathrm{erf}(\zeta)}{R} = (\mathrm{erf}(\zeta) - \Theta(\zeta))\,\mathbf{T}^{ij}_{\mathrm{bare}}(\mathbf{R}) + 2\zeta^2\Theta(\zeta)\frac{R_i R_j}{R^5}, \\ \Theta(\zeta) &= \frac{2\zeta}{\sqrt{\pi}}e^{-\zeta^2},\ \zeta = \frac{R}{\sqrt{\sigma_A^2+\sigma_B^2}}\ \ (i,j=x,y,z)\,, \end{aligned} \tag{4.2}$$

derived from the Gaussian-screened Coulomb interaction of two oscillator charge densities with widths σ_A^2 at distance $\mathbf{R}$. Here, $\mathbf{T}^{ij}_{\mathrm{bare}}(\mathbf{R}) = (-3R_iR_j + \delta_{ij}R^2)/R^5$ is the "bare" dipole tensor. The oscillator width is derived from the zero-distance limit of the classical dipole-dipole interaction [40, 161]:

$$\sigma_A = \left(\frac{1}{3}\sqrt{\frac{2}{\pi}}\alpha_{0,A}\right)^{1/3} \tag{4.3}$$

This dipole tensor is exempt from DFA-specific short-range damping. Together with the vdW-OQDO parametrization, this choice of dipole tensor defines the parameter-free MBD variant, which we abbreviate as MBD@FCO and apply throughout this chapter.

4.2 MBD@FCO Yields Accurate Dispersion Energies

In Figure 4.1a, we show the MBD@FCO, MBD@rsSCS, and XDM dispersion interaction energies (with respect to unrelaxed monomers) for the S12L dataset of supramolecular dimers [264], excluding the DFA contribution. These results are compared to the benchmark $E^{(2)}_{\mathrm{disp}}$ values [270] from the density-functional based symmetry-adapted perturbation theory (DFT-SAPT) [37, 79, 83]. The MBD@FCO variant yields dispersion energies clustered close to the perfect correlation line with the MARE = 18%, while MBD@rsSCS and XDM systematically underestimate dispersion by about 50 %. The discrepancy between MBD@FCO and SAPT is primarily due to the neglect of beyond-dipole couplings in the MBD model, which can contribute up to 30 % of the total dispersion energy at equilibrium [11, 12].

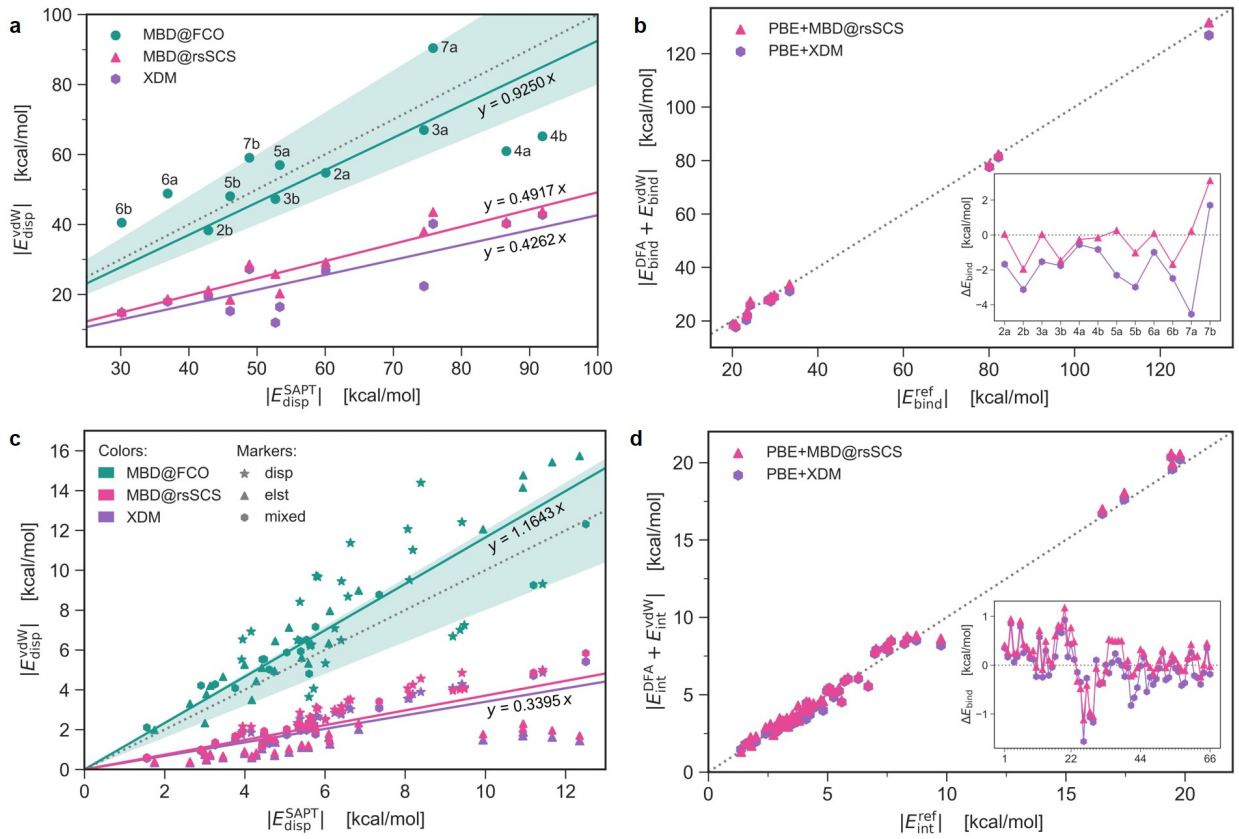

Figure 4.1: MBD@FCO method predicts accurate dispersion energies. **(a,c)** Parity plots of the dispersion energy magnitudes as computed by DFT-SAPT ($E^{(2)}_{\text{disp}}$) [234, 270] versus MBD@FCO, MBD@rsSCS and XDM methods for the **(a)** S12L [264] and **(c)** S66 [182] datasets. The solid lines display linear fits of data points, and the gray dotted line marks the perfect correlation, with the shaded area highlighting ±20% interval around. In **(c)**, different markers denote the S66 dimers dominated by dispersion (stars), electrostatics (triangles), or mixed interactions (hexagons), while colors refer to the vdW methods (see the legend). The linear regression equation for MBD@rsSCS (pink line) is $y = 0.3705x$ (not shown for clarity). **(b,d)** Parity plots of the total DFA+vdW binding/interaction energy magnitudes as computed by PBE+MBD@rsSCS and PBE+XDM methods versus the corresponding reference energies [182, 264] for the **(b)** S12L and **(d)** S66 datasets. The gray dotted line corresponds to the perfect correlation, while the insets show the error $\Delta E = E_{\text{DFA+vdW}} - E_{\text{ref}}$ resolved per system.

Similar trends are observed also for the S66 dataset composed of small molecular dimers [182] (Figure 4.1c). The MBD@FCO method performs much better than MBD@rsSCS and XDM, whose errors are as large as 60-70%, while MBD@FCO systematically overestimates the dispersion energy by roughly 16% compared to the DFT-SAPT reference [234] Interestingly, the strongest overestimation occurs for the dispersion-dominated dimers, according to the SAPT-based classification in Ref. [182]. Another curious observation is that the S66 dataset turns out to be more challenging for accurate predictions of the pure dispersion energy, than the seemingly more complex S12L dataset. This is possibly due to some cancellation of errors and system-size dependent balance between semi-local and nonlocal correlation energies, suggesting that more robust strategies than a simple energy-based fitting should be explored for damping functions design.

However, for the total binding/interaction energy including the DFA contribution, $E_{\text{DFT}} + E_{\text{vdW}}$, the PBE+MBD@rsSCS and PBE+XDM methods demonstrate excellent accuracy both for S12L and S66 (Figure 4.1b,d). In the case of S12L, the MAE of PBE+MBD@rsSCS and PBE+XDM rel-

ative to the experimental binding energies [264] are 0.86 and 2.04 kcal/mol, respectively, while for S66 their MAE in interaction energy are, correspondingly, 0.35 and 0.33 kcal/mol, as compared to the CCSD(T) reference values [182].[1]

Overall, our findings highlight that the empirical damping used in DFT+vdW methods obfuscates the physical meaning of the dispersion energy, making it an adaptive correction to the underlying DFA to compensate for its energetic errors rather than a physically meaningful contribution. This point has also been widely debated in the recent literature [271–273], underscoring its vital importance for the field. In this context, the MBD@FCO method delivers dispersion energies that are closer to the SAPT reference. The remaining challenge is how to couple the parameter-free MBD@FCO method with a semi-local DFA to avoid double-counting of the correlation energy. This question is beyond the scope of this thesis; nevertheless, we briefly discuss that in Section 4.7.

In the following, we show that besides the accurate dispersion energies, the MBD@FCO method yields physically relevant vdW-induced density polarization.

4.3 MBD Density Polarization

Collective charge density fluctuations, which mediate vdW dispersion interaction, induce density polarization [23–25]. Within the MBD framework, vdW-induced density polarization is naturally defined as the difference between the charge densities of interacting and non-interacting QDOs, constituting the model [25, 160]:

$$\rho_{\mathrm{pol}}(\mathbf{r}) = \langle\Psi|\hat{\rho}|\Psi\rangle - \langle\Psi_0|\hat{\rho}|\Psi_0\rangle = \rho_{\mathrm{MBD}}(\mathbf{r}) - \rho_0(\mathbf{r}), \quad \hat{\rho} = \sum_{A=1}^{N} q_A \delta(\mathbf{r} - \hat{\mathbf{r}}_A), \tag{4.4}$$

where Ψ and Ψ_0 are the ground states of interacting (MBD) and non-interacting oscillators, respectively. The ground state wave function for a system of N non-interacting quantum Drude oscillators is a Gaussian state of the form

$$\Psi_0\big(\{\mathbf{r}_A\}\big) = \big\langle\{\mathbf{r}_A\}\big|\Psi_0\big\rangle = \prod_{A=1}^{N}\left(\frac{m_A\omega_A}{\pi\hbar}\right)^{3/4} \exp\left[-\frac{m_A\omega_A}{2\hbar}(\mathbf{r}_A - \mathbf{R}_A)^2\right]. \tag{4.5}$$

The ground state of the MBD Hamiltonian is fully described by the $3N\times 3N$ symmetric positive-definite correlation matrix $\mathbf{C}$ between QDOs dipole fluctuations, which is defined as:

$$\mathbf{C} = \hbar^{-1}\mathbf{M}\mathbf{O}^T\tilde{\mathbf{\Omega}}\mathbf{O}\mathbf{M}, \tag{4.6}$$

where $\mathbf{M} = \mathrm{diag}\big\{\sqrt{m_1}, \sqrt{m_1}, \sqrt{m_1}, ..., \sqrt{m_N}, \sqrt{m_N}, \sqrt{m_N}\big\}$ is the matrix of the square root of QDO masses, $\mathbf{O}$ is the $3N\times 3N$ orthogonal matrix diagonalizing the quadratic MBD potential, i.e.

$$\tilde{\mathbf{\Omega}}^2 = \mathrm{diag}\big\{\tilde{\omega}_1^2, ..., \tilde{\omega}_{3N}^2\big\} = \mathbf{O}(\mathbf{\Omega}^2 + \mathbf{V}_{\mathrm{dip}})\mathbf{O}^T, \tag{4.7}$$

with $[\mathbf{V}_{\mathrm{dip}}]_{AB} = \omega_A\omega_B\sqrt{\alpha_{0,A}\alpha_{0,B}}\,\mathbf{T}_{AB}$ being the 3×3 dimensional block of dipole interaction matrix. The correlation matrix can also be viewed as composed of 3×3 blocks $\mathbf{C}_{AB}$, and with

[1] For S12L, the relaxed monomer geometries were used, while in the case of S66 the monomer geometries were the same as in dimers, which is consistent with the methodology of the original references [182, 264]. This explains the difference in the terminology used: "binding energy" for S12L and "interaction energy" for S66. The dispersion energies in Figure 4.1a,c were computed using the unrelaxed monomers, in accordance with the DFT-SAPT approach [234, 270].

this notation, the MBD ground state can be written as

$$\Psi_{\mathrm{MBD}}\left(\{\mathbf{r}_A\}\right) = \left\langle \{\mathbf{r}_A\} \middle| \Psi \right\rangle = \left[\left(\prod_{A=1}^{N} m_A^3 \right) \left(\prod_{\alpha=1}^{3N} \frac{\tilde{\omega}_\alpha}{\pi\hbar} \right) \right]^{1/4} \exp\left[-\frac{1}{2} \sum_{A,B=1}^{N} (\mathbf{r}_A - \mathbf{R}_A)^T \mathbf{C}_{AB} (\mathbf{r}_B - \mathbf{R}_B) \right] . \quad (4.8)$$

The Eq. (4.4) for the vdW-induced density polarization reads

$$\rho_{\mathrm{pol}}(\mathbf{r}) = \sum_{A=1}^{N} q_A \int_{\mathbb{R}^{3N}} \left(\left|\Psi\left(\{\mathbf{r}_A\}\right)\right|^2 - \left|\Psi_0(\{\mathbf{r}_A\})\right|^2 \right) \delta(\mathbf{r} - \mathbf{r}_A) \; \mathrm{d}^3\mathbf{r}_1 ... \mathrm{d}^3\mathbf{r}_N \, . \quad (4.9)$$

Substituting Eqs. (4.5) and (4.8) and performing the integration as in [124], one can obtain:

$$\rho_{\mathrm{pol}}(\mathbf{r}) = \sum_{A=1}^{N} \frac{q_A}{\pi^{3/2}} \left(\frac{1}{\Lambda_A^3} \exp\left[-(\mathbf{r} - \mathbf{R}_A)^T \mathbf{K}_{AA} (\mathbf{r} - \mathbf{R}_A) \right] - \frac{1}{\lambda_A^3} \exp\left[-\lambda_A^{-2} (\mathbf{r} - \mathbf{R}_A)^2 \right] \right) , \quad (4.10)$$

where $\lambda_A = \sqrt{\hbar / m_A \omega_A}$ and $\Lambda_A = \sqrt[6]{\det \mathbf{C}^{/A} / \det \mathbf{C}}$ are characteristic length scales, and $\mathbf{K}_{AA}$ is a 3×3 matrix obtained from the correlation matrix as

$$\mathbf{K}_{AA} = \mathbf{C}_{AA} - \mathbf{C}_A^T \left[\mathbf{C}^{/A} \right]^{-1} \mathbf{C}_A \, . \quad (4.11)$$

Here, $\mathbf{C}_{AA}$ is the A-th diagonal block of $\mathbf{C}$. The $(3N-3) \times (3N-3)$ square matrix $\mathbf{C}^{/A}$ is obtained from $\mathbf{C}$ by removing from it a row vector $\mathbf{C}_A^T$ sized $3 \times (3N-3)$ and a column vector $\mathbf{C}_A$ sized $(3N-3) \times 3$, which correspond to the A-th QDO. Note that $\mathbf{C}_A^T$ and $\mathbf{C}_A$ do not include the diagonal elements of $\mathbf{C}$, which enter the $\mathbf{C}_{AA}$ matrix. Finally, $[\mathbf{C}^{/A}]^{-1}$ denotes the inverse of $\mathbf{C}^{/A}$ matrix. Fig. 4.2 schematically explains the introduced notation. A more complete derivation of Eq. (4.10) based on the density-matrix formalism can be found in Section 5.1 or in a slightly different form in Refs. [124, 274].

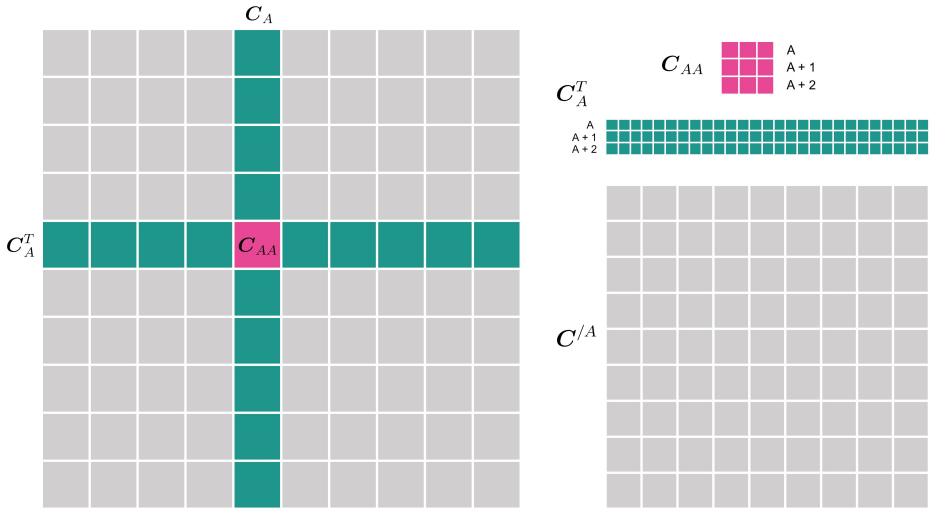

Figure 4.2: Schematic illustration of the notation for the MBD correlation matrix.

Connection to Electron Density

The derived polarization density can be used to refine the underlying KS electron density $\rho(\mathbf{r})$. We introduce the total wave function and the total density operator for the KS+MBD system as

$$\Phi = \Psi_{\mathrm{KS}}\left(\{\mathbf{r}_i\}\right) \cdot \Psi_{\mathrm{MBD}}\left(\{\mathbf{r}_A\}\right) , \quad \hat{\rho}_{\mathrm{tot}} = \hat{\rho}_{\mathrm{elec}} + \hat{\rho}_{\mathrm{MBD}} = \sum_{i=1}^{N_{\mathrm{elec}}} e\delta(\mathbf{r} - \hat{\mathbf{r}}_i) + \sum_{A=1}^{N} q_A \delta(\mathbf{r} - \hat{\mathbf{r}}_A) , \quad (4.12)$$

where $\Psi_{\mathrm{KS}}(\{\mathbf{r}_i\})$ is the Slater determinant of KS orbitals, depending on coordinates of *real* electrons, and $\Psi_{\mathrm{MBD}}(\{\mathbf{r}_A\})$ is the MBD wave function, depending on coordinates of *drudons* – quasiparticles, mimicking collective correlated response of valence electrons in an atom. It is important to emphasize that the MBD Hamiltonian does not directly depend on real electronic coordinates; in fact, the dipole-coupled oscillators represent an auxiliary model quantum system assisting in capturing the intricate electron correlations in a coarse-grained way. Moreover, for a given system there are N_{elec} electrons and N (equal to the number of atoms) drudons, with all electrons carrying elementary charge e and each oscillator having, in general, a different charge q_A (see Figure 4.3). Hence, $\hat{H}_{\mathrm{MBD}}$ and $\hat{H}_{\mathrm{KS}}$ exist in distinct Hilbert spaces, and the total wave function can be represented as a product state.

The total charge density of the KS+MBD system is straightforward to evaluate as

$$\rho_{\mathrm{tot}}(\mathbf{r}) = \langle\Phi|\hat{\rho}_{\mathrm{tot}}|\Phi\rangle = \langle\Psi_{\mathrm{KS}}|\hat{\rho}_{\mathrm{elec}}|\Psi_{\mathrm{KS}}\rangle + \langle\Psi_{\mathrm{MBD}}|\hat{\rho}_{\mathrm{MBD}}|\Psi_{\mathrm{MBD}}\rangle = \rho(\mathbf{r}) + \rho_{\mathrm{MBD}}(\mathbf{r})\,. \tag{4.13}$$

Since the MBD Hamiltonian uses a harmonic bond between drudon and its nuclei, it cannot model the full charge density reasonably. Nevertheless, the oscillators can model *changes* in density due to the long-range correlations (*vide infra*). Therefore, for the purpose of total density analysis in Section 4.6 we will use $\rho_{\mathrm{tot}}(\mathbf{r}) = \rho(\mathbf{r}) + \rho_{\mathrm{pol}}(\mathbf{r})$, with $\rho_{\mathrm{pol}}(\mathbf{r})$ defined in Eq. (4.10).

The optimal choice of charges q_A within the MBD model is essential for accurate density predictions. Using the vdW-OQDO parametrization, we assign species-specific charges based on α_0 and C_6 coefficients, in contrast to the earlier MBD approaches using arbitrary $q_A = 1$ a.u. charge values [25, 26]. Figure 4.3 shows the histogram of these charges for the S12L dataset.

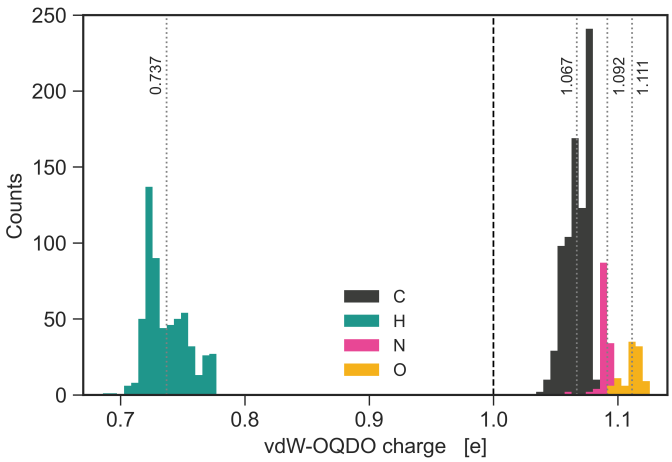

Figure 4.3: Distribution of vdW-OQDO charges across the S12L dataset. For every chemical species, the dotted gray line denotes the position of the mean value (displayed in the plot). In the conventional MBD parametrization, all charges are set to 1 a.u., which is highlighted by the black dashed line.

4.4 Density Polarization Induced by Long-Range Correlations

To focus on the dispersion interaction-induced polarization, we examine the MBD@FCO density *deformation* $\Delta\rho_{\mathrm{pol}}(\mathbf{r})$ for a dimer (D) composed of two monomers (M1, M2):

$$\Delta\rho_{\mathrm{pol}}(\mathbf{r}) = \rho^{\mathrm{D}}_{\mathrm{pol}}(\mathbf{r}) - \rho^{\mathrm{M1}}_{\mathrm{pol}}(\mathbf{r}) - \rho^{\mathrm{M2}}_{\mathrm{pol}}(\mathbf{r})\,. \tag{4.14}$$

This quantity describes the density shifts due to the long-range electronic correlations and thus can be directly compared to its *ab initio* counterparts (*vide infra*). We also calculate the vdW polarization of density using the self-consistent Tkatchenko-Scheffler method (sc-TS) [23, 39] on top of the Perdew-Burke-Ernzerhof (PBE) functional [63],

$$\Delta\rho_{\text{sc-TS}}(\mathbf{r}) = \left[\rho^{\mathrm{D}}_{\mathrm{PBE+sc\text{-}TS}}(\mathbf{r}) - \rho^{\mathrm{M1}}_{\mathrm{PBE+sc\text{-}TS}}(\mathbf{r}) - \rho^{\mathrm{M2}}_{\mathrm{PBE+sc\text{-}TS}}(\mathbf{r})\right] - \left[\rho^{\mathrm{D}}_{\mathrm{PBE}}(\mathbf{r}) - \rho^{\mathrm{M1}}_{\mathrm{PBE}}(\mathbf{r}) - \rho^{\mathrm{M2}}_{\mathrm{PBE}}(\mathbf{r})\right]\,, \tag{4.15}$$

to analyze the effects of self-consistency at the DFT level.

4.4.1 Choice of the Reference Method

As a reference for density deformation $\Delta\rho$ in Eq. (4.14), we use the double density difference between coupled-cluster singles and doubles (CCSD) and Hartree-Fock (HF) levels of theory, $\Delta\rho_{\mathrm{CCSD-HF}}(\mathbf{r})$,

$$\Delta\rho_{\mathrm{CCSD-HF}}(\mathbf{r}) = \left[\rho^{\mathrm{D}}_{\mathrm{CCSD}}(\mathbf{r}) - \rho^{\mathrm{M1}}_{\mathrm{CCSD}}(\mathbf{r}) - \rho^{\mathrm{M2}}_{\mathrm{CCSD}}(\mathbf{r})\right] - \left[\rho^{\mathrm{D}}_{\mathrm{HF}}(\mathbf{r}) - \rho^{\mathrm{M1}}_{\mathrm{HF}}(\mathbf{r}) - \rho^{\mathrm{M2}}_{\mathrm{HF}}(\mathbf{r})\right], \tag{4.16}$$

which captures polarization of density due to the (long-range) electronic correlation effects at *ab initio* level. Although CCSD is not chemically accurate for energies (often worse than MP2), this method is widely applied as a reference for electron density in closed-shell systems [23, 243, 275]. Why CCSD densities are more accurate than CCSD energies can be understood e.g. from the so-called Wigner's $2n+1$ rule in many-body perturbation theory (MBPT) [276]. This rule states that the $2n$-th and/or $(2n+1)$-th order perturbation contributions to a non-degenerate energy can be obtained from a knowledge of the wavefunction through n-th order [276]. In other words, having e.g. MP2 wave function is sufficient to obtain MP4-level energies. Similar considerations are valid in coupled cluster theory [276, 277], and hence the CCSD wave function (and density) can be considered as effectively having access to beyond-CCSD energies.

Nevertheless, to examine the effect of triples, essential to obtain chemical accuracy for correlation energies, we performed calculations of the coupled-cluster density including full triples (CCSDT)[2] for methane and neon dimers. The CCSDT densities were shown to be almost identical to the composite CCSDTQ densities for atoms, ions, and small molecules [279], and therefore CCSDT can be considered as a highly accurate reference method for densities. The obtained CCSDT–HF density differences for methane and neon dimers show only minor quantitative changes compared to the CCSD–HF counterparts (Figure 4.4). This validates using CCSD as the reference method for systems larger than methane or neon when CCSDT becomes prohibitively expensive.

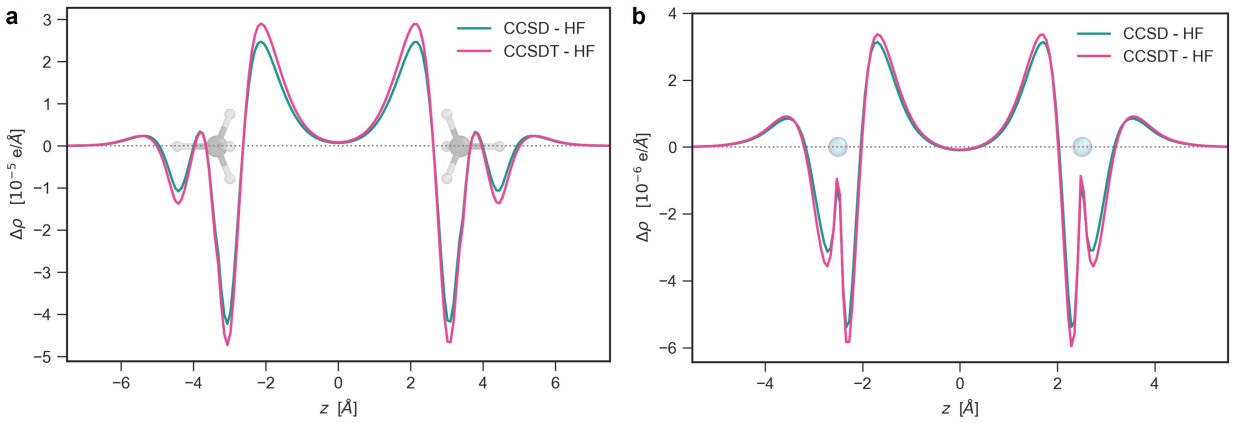

Figure 4.4: Effect of coupled-cluster triples on electron density deformation in **(a)** methane and **(b)** neon dimers. The plots show plane-averaged density polarization $\Delta\rho(z)$ as computed in MRCC using aug-cc-pVDZ basis sets, with the monomer geometries represented along the distance axis.

The choice of the mean-field reference also influences the resulting density polarization. From the quantum chemistry point of view, the HF method by definition provides a rigorous uncorrelated reference for coupled cluster methods. However, we should note that the CCSD–HF difference is generally not equivalent to dispersion-induced polarization and contains also other

[2] We computed CCSDT densities, since CCSD(T) gradients are not widely implemented in quantum chemistry programs. CCSD(T) densities should be very close to CCSD since non-iterative (T) correction does not affect T_1 and T_2 and makes only a small contribution to density [278].

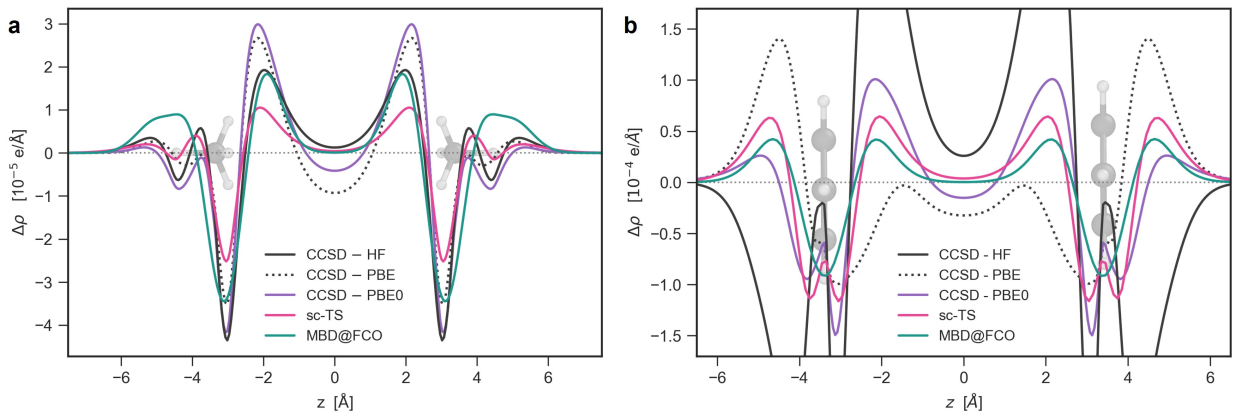

Figure 4.5: Effect of the mean-field reference method on electron density deformation in **(a)** methane and **(b)** benzene dimers. The plots show plane-averaged density polarization $\Delta\rho(z)$ as computed using CCSD, HF, PBE and PBE0 methods in PySCF with **(a)** aug-cc-pVQZ and **(b)** aug-cc-pVDZ basis sets, as well as $\Delta\rho(z)$ calculated with the sc-TS and MBD@FCO methods (see the text for further explanations). The monomer geometries are shown along the distance axis. In the case of benzene **(b)**, the extrema values for CCSD–HF are 3.17×10^{-4} and -2.95×10^{-4} e/Å (not shown in the plot for clarity).

effects, arising e.g. due to the differences between CCSD and HF description of induction interaction. We illustrate this in Figure 4.5 and Table 4.1 at the contrasting examples of methane and benzene dimers. We consider HF, PBE, and PBE0 as the mean-field references, and compute the respective differences with CCSD densities.

In methane dimer, the long-range interaction is dominated by vdW dispersion, hence the CCSD–HF difference is almost exclusively due to dispersion-induced polarization. This is also supported by the analysis of quadrupole moments (Table 4.1), which shows that there are no significant differences between HF, PBE0, and CCSD for the change in quadrupole moment due to intermonomer interactions. PBE and PBE0 functionals spill excessive charge into intermolecular space compared to CCSD, which leads to negative $\Delta\rho$ (Figure 4.5a) and is another manifestation of delocalization error (see Section 2.1.5). Therefore, the HF method seems to be the most reasonable reference method for the methane dimer.

This is in sharp contrast with parallel-displaced benzene dimer (at 6 Å distance), where CCSD–HF, CCSD–PBE, and CCSD–PBE0 produce qualitatively different density deformation profiles (Figure 4.5b). Compared to CCSD, PBE functional again spills too much charge into intermolecular space (negative $\Delta\rho$), which is transferred from the outer sides of the monomers (positive peaks in $\Delta\rho$). The HF density is, in contrast, too localized on monomers, which leads to overly large magnitudes of $\Delta\rho$ outside monomers. Hybrid PBE0 functional seems to deliver the most balanced description of benzene dimer's electronic structure among considered mean-field methods.

These findings can be rationalized by the analysis of quadrupole moments (Table 4.1), which shows a good agreement between PBE0 and CCSD for $Q_{zz}^{\text{dimer}} - 2Q_{zz}^{\text{mono}}$ (7% error), while the HF predictions are substantially worse (26% error). This suggests that in the case of benzene dimer, CCSD–HF deformation density is contaminated with the effects beyond vdW dispersion interactions, while the CCSD–PBE0 deformation density seems to be much closer to the dispersion-induced density deformation, which is highlighted by a good agreement with the sc-TS and MBD@FCO methods (Figure 4.5b).

To sum it up, filtering out the density deformation due to the vdW dispersion interactions using

methane	HF	PBE0	CCSD
Q_{zz}^{mono}	−8.4125	−8.3976	−8.3905
Q_{zz}^{dimer}	−16.8453	−16.8175	−16.8003
$Q_{zz}^{dimer} - 2Q_{zz}^{mono}$	**−0.0203**	**−0.0223**	**−0.0193**
benzene	**HF**	**PBE0**	**CCSD**
Q_{zz}^{mono}	−41.2886	−39.9845	−40.7897
Q_{zz}^{dimer}	−83.4515	−80.7073	−82.2708
$Q_{zz}^{dimer} - 2Q_{zz}^{mono}$	**−0.8743**	**−0.7383**	**−0.6914**

Table 4.1: Quadrupole moments of methane and benzene dimers as computed in Q-CHEM with aug-cc-pVQZ and aug-cc-pVDZ basis sets, respectively. The HF, PBE0, and CCSD values (in Debye·Å) of Q_{zz} component of quadrupole moment tensor (in Cartesian representation) are reported for the dimers and their monomers, together with the differences (in bold).

ab initio methods is a challenging task. In principle, an alternative approach would be considering the dispersion-induced wave function correction in Eq. (2.70) from SAPT. However, at the time of writing, we were not able to find a suitable code that could output those density corrections. Therefore, for small dispersion-dominated dimers, which we use to benchmark the MBD@FCO method in what follows, we opt for the CCSD–HF method as the reference for density deformations.

4.4.2 Results for Small Molecular Dimers

The six dispersion-dominated dimers were considered, including methane, pentane, ethene, neopentane, ethene-neopentane, and neopentane-pentane. The corresponding geometries were obtained from the S22×5 dataset [281] for methane and ethene and from the S66×8 [182] dataset for the remaining dimers, in all cases at largest available intermonomer distance. To

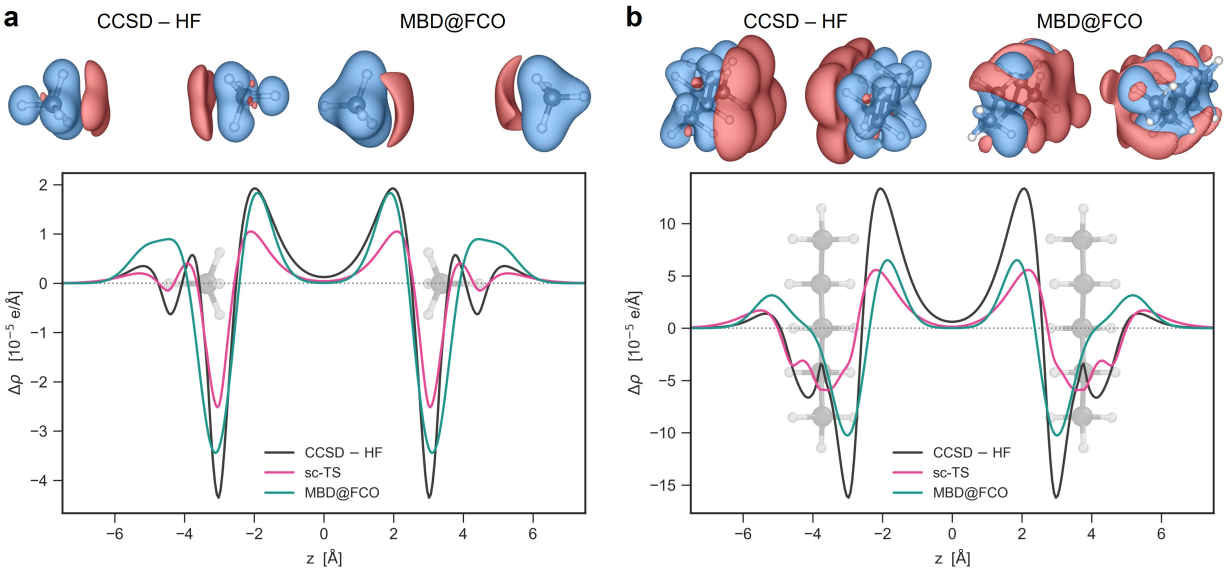

Figure 4.6: vdW-induced density polarization in in (a) methane and (b) pentane dimers, as computed by the CCSD–HF, sc-TS and MBD@FCO methods. The plots show plane-averaged density polarization $\Delta\rho(z)$ with the monomer geometries represented along the distance axis. The isosurfaces (top) illustrate $\Delta\rho(\mathbf{r})$ in the 3D space as calculated by CCSD−HF and MBD@FCO methods with the isovalues of 1.5×10^{-6} and $2 \times 10^{-6}\, e/\text{Å}^3$ for methane and pentane dimers, respectively. The red and blue colors denote the accumulation and depletion of charge density, correspondingly. The isosurface plots were produced using the VESTA software [280]. CCSD–HF calculations were performed using aug-cc-pVQZ and aug-cc-pVDZ basis sets for methane and pentane dimers, respectively.

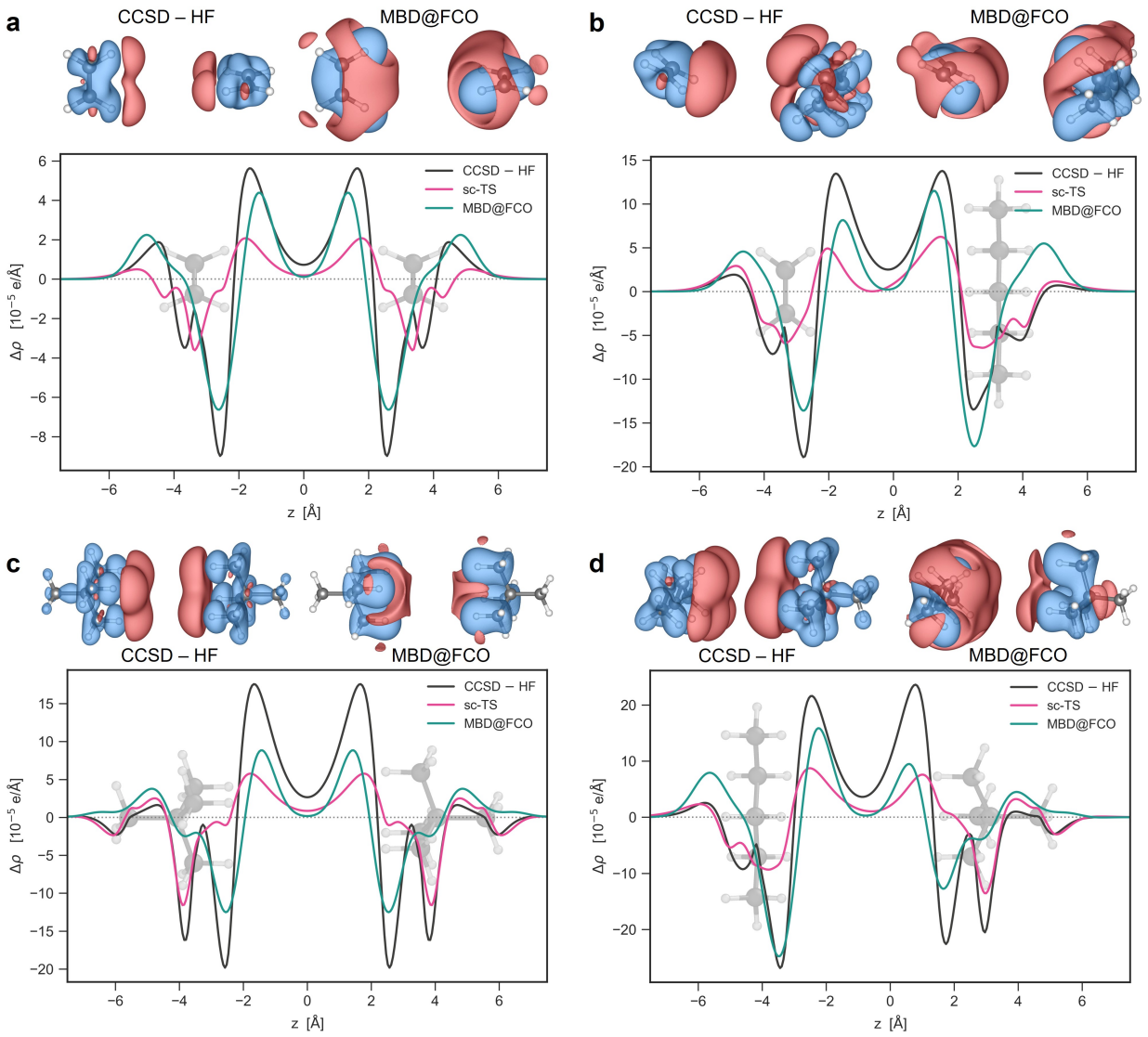

Figure 4.7: Electron density polarization induced by long-range correlations in **(a)** ethene, **(b)** ethene-pentane, **(c)** neopentane, and **(d)** pentane-neopentane dimers, as computed by the CCSD–HF, sc-TS and MBD@FCO methods. The plots show plane-averaged density polarization $\Delta\rho(z)$ with the monomer geometries represented along the distance axis. The isosurfaces illustrate $\Delta\rho(\boldsymbol{r})$ in the 3D space as calculated by CCSD−HF and MBD@FCO methods with the isovalues of **(a,d)** 4×10^{-6} and **(b,c)** 5×10^{-6} e/Å^3. CCSD–HF calculations were performed using **(a,b)** aug-cc-pVTZ and **(c,d)** aug-cc-pVDZ basis sets.

ensure that the computed density differences correspond to long-range electronic correlations, the geometries were modified, if needed, to provide intermolecular vacuum separation of at least 6 Å. Finally, they were rotated to align the z-axis with the line connecting the centers of monomers.

In Figure 4.6, we plot the plane-averaged density deformations $\Delta\rho(z) = \int\int \Delta\rho(\mathbf{r})\, dxdy$ for the CCSD–HF, sc-TS and MBD@FCO methods, as well as $\Delta\rho(\mathbf{r})$ isosurfaces for methane and pentane dimers. The results for the remaining four dimers are presented in Figure 4.7. The CCSD–HF, sc-TS, and MBD@FCO methods qualitatively agree in predicting charge redistribution due to electronic correlations. MBD@FCO accurately describes density polarization in the inter-monomer region, while discrepancies near the nuclei stem from the harmonic potential approximation in MBD, which differs from the full Coulomb potential. The self-consistent sc-

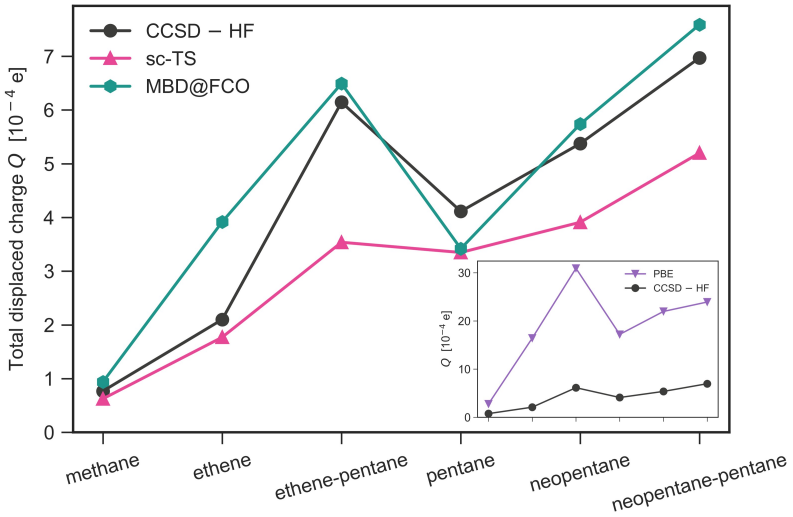

Figure 4.8: vdW-displaced charge in small molecular dimers as computed by the CCSD–HF, sc-TS and MBD@FCO methods. The dimers are ordered by the number of electrons. The inset shows the displaced charge computed using the bare PBE functional compared to the CCSD–HF reference values on a larger scale.

TS method agrees more closely with CCSD in the density tails, suggesting the importance of self-consistency, which is the focus of ongoing work on MBD@FCO. The isosurfaces of $\Delta\rho(\mathbf{r})$ displayed in Figures 4.6-4.7 witness that $\Delta\rho_{\text{pol}}(\mathbf{r})$ from MBD@FCO captures the key features of $\Delta\rho_{\text{CCSD}-\text{HF}}(\mathbf{r})$ redistribution also in 3D space. The remaining differences outside monomers can be attributed to the dipole coupling approximation used in the MBD model. In addition, the CCSD–HF difference includes contributions other than dispersion, e.g. the difference in electronic polarization between CCSD and HF. In methane and ethene dimers, this difference is small, while in relatively more polarizable dimers involving pentane or neopentane, this effect may contribute to the difference between CCSD and MBD (Figures 4.6-4.7).

The strength of vdW-induced polarization can be quantified using the total displaced charge, which we define as an integral of charge displacement over charge-accumulating or depleting regions, i.e.

$$Q_{\text{vdW}} = \int_{\Omega} \Delta\rho(\mathbf{r})\, d\mathbf{r}\,, \tag{4.17}$$

where Ω is the region with $\Delta\rho_{\text{pol}}(\mathbf{r}) > 0$. Equivalently, one can integrate over the negative $\Delta\rho(\mathbf{r})$ regions, which will give exactly the same magnitude but opposite sign because the integral of $\Delta\rho(\mathbf{r})$ over the whole space is zero by construction. This measure, Q_{vdW}, captures the total charge displaced due to the long-range vdW interactions and serves as a scalar measure of vdW polarization strength. The results summarized in Figure 4.8 exhibit a good agreement between MBD@FCO and CCSD–HF displaced charges. MBD@FCO closely follows the trend of CCSD–HF and has a significant error (86%) only for ethene dimer, with the other values being within the 22% range, resulting in the MARE of 12% if ethene dimer is excluded from statistics (24% if included). The sc-TS method, on the other hand, systematically underestimates Q_{vdW} with the MARE = 25%, with the maximal error of -42% for ethene-pentane dimer. Thus, MBD@FCO density provides an estimate of Q_{vdW} within 10-20%.

Furthermore, Ambrosetti *et al.* [168] have recently demonstrated that the fully-coupled MBD model can reproduce molecular optical excitation spectra at a semi-quantitative level. This ensures that the fully-coupled MBD method delivers a physically sound description of collective quasi-plasmonic fluctuation modes, the interplay of which creates a vdW-induced polarization of electronic density. Moreover, our results in Figure 4.6 align with the renowned Feynman's picture of vdW interactions: net charge accumulation between monomers causes an electrostatic force between this distorted electronic cloud and the nuclei of a given molecule, resulting

in effective intermolecular attraction [22, 282].

Overall, the results presented so far in this chapter validate MBD@FCO as a reliable method to account for the dispersion energy and dispersion-polarization of the density at the same time. This makes the MBD@FCO method unique compared to other widely used post-DFT vdW models, such as XDM or DFT-D3(4). The observed 10-20% errors are inherent to the atomic coarse-grained representation of charge oscillations and the dipole approximation applied in the MBD framework, and these errors do not affect qualitatively the observed effects. In what follows, we apply the FCO@MBD method to more extended and polarizable systems.

4.5 MBD@FCO Density Polarization Scales with System Size

To assess the extent of vdW charge density polarization in real systems, we examine two sets of large organic complexes. The first one consists of the 12 host-guest dimers from the S12L dataset [264]. The second, referred to as the L7+ dataset, includes 11 systems: seven large complexes from the L7 dataset [265] along with four additional systems – C_{60} buckyball dimer, parallel-displaced stacked dimers of circumcoronene (C3C3PD) and circum-circumcoronene (C4C4PD), and a DNA-ellipticine intercalation complex. In addition, we include a parallel-

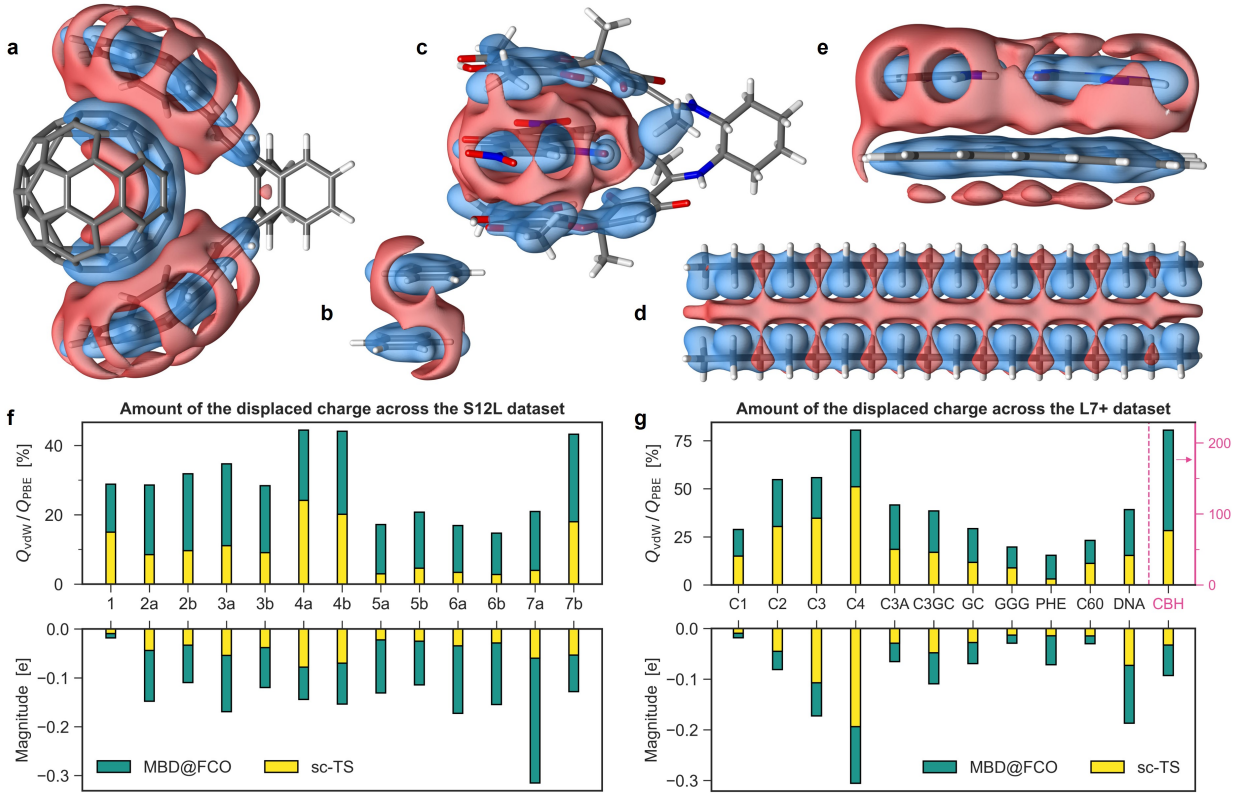

Figure 4.9: vdW-displaced charge across S12L and L7+ datasets. Isosurfaces of vdW density polarization between a complex and its isolated monomers for selected systems from S12L and L7+ datasets (with isovalue 3×10^{-5} a.u., unless otherwise noted): **(a)** C_{60} buckyball catcher (4a, S12L); **(b)** benzene dimer (1, S12L, 2×10^{-5} a.u.); **(c)** TNF "pincer" complex (3a, S12L, 6×10^{-5} a.u.); **(d)** octadecane dimer (CBH, L7+); **(e)** guanine-cytosine dimer on circumcoronene (C3GC, L7+). The red and blue colors denote accumulation and depletion of charge density, respectively. **(f,g)** The total vdW-displaced charge $-Q_{vdW}$ (bottom) and its ratio to the PBE-displaced charge Q_{PBE} (top) for **(f)** S12L and **(g)** L7+ datasets as computed by sc-TS and MBD@FCO methods. Note the separate y-axis for CBH in panel **(g)**.

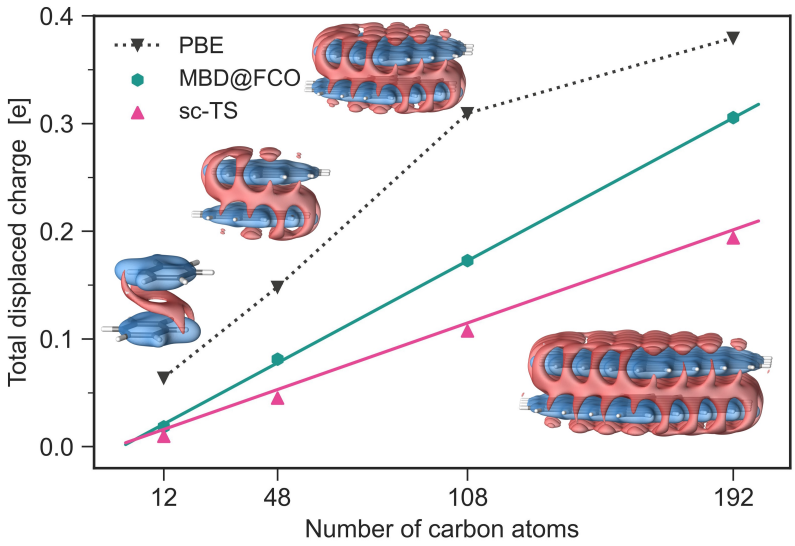

Figure 4.10: Linear scaling of vdW-displaced charge with system size. Dependence of the total displaced charge on the size of PAHs as computed by the three methods. Linear fits are shown for sc-TS and MBD@FCO to guide an eye. The insets display the 3×10^{-5} a.u. isosurfaces of the $\Delta\rho_{\mathrm{pol}}(\mathbf{r})$.

displaced stacked benzene dimer in both datasets (as complex '1' in S12L and C1C1PD in L7+), representing the prototypical $\pi-\pi$ interaction.

In total, we consider 24 unique systems, ranging from 24 to 240 atoms. For consistency, the atomic geometries and nomenclature follow the original publications [264, 265]. The additional geometries of the C_{60} dimer and the DNA-ellipticine complex were taken from Ref. [283], while the atomic geometries of C3C3PD and C4C4PD were optimized at the PBE+MBD@rsSCS level in FHI-AIMS [233] with 'tight' basis sets. The vdW-induced charge displacements, $\Delta\rho_{\mathrm{pol}}(\mathbf{r})$, were computed for all systems using both the sc-TS and MBD methods, and compared against the PBE functional. The choice of PBE is justified by its widespread use and relatively low computational cost compared to more accurate meta-GGA or hybrid functionals. Test calculations with the HF method for selected systems yielded qualitatively similar results, supporting the generality of our conclusions independent on a mean-field reference (see Appendix A6 and Figure A8).

Figure 4.9 summarizes the results, displaying isosurfaces of $\Delta\rho_{\mathrm{pol}}(\mathbf{r})$ from MBD@FCO for select systems. Large, polarizable monomers exhibit significant vdW-induced charge displacements toward the intermonomer regions, much more pronounced than in smaller dimers like methane or pentane considered above. As shown in Figure 4.9f-g (lower panels), Q_{vdW} ranges from 0.1 e to 0.3 e for large systems, whereas it is only about 0.01 e in the benzene dimer. These values are predicted by MBD@FCO, with the pairwise sc-TS method yielding much lower vdW-induced charge displacements. To contextualize these numbers, we compare Q_{vdW} with the analogous quantity Q_{PBE}, calculated using the bare PBE functional without vdW corrections. The ratio $Q_{\mathrm{vdW}}/Q_{\mathrm{PBE}}$ (shown in Figure 4.9f-g, upper panels) indicates that vdW-induced charge displacement can account for up to 80% of the PBE displacement in $\pi-\pi$ systems, and even exceeds the PBE value by a factor of 2 for the linear alkane chain (CBH). Notably, even at the sc-TS level, this ratio reaches 76 % for CBH, while MBD@FCO further increases vdW charge polarization by enabling delocalized collective density fluctuations along the chain (see Figure 4.9g).

To explore how vdW density polarization scales with system size, we consider a homologous sequence of polyaromatic hydrocarbon (PAH) dimers from the L7+ dataset: benzene (C1C1PD), coronene (C2C2PD), circumcoronene (C3C3PD), and circum-circumcoronene (C4C4PD). As illustrated in Figure 4.10, Q_{vdW} exhibits a perfect linear correlation with system size. The ratio $Q_{\mathrm{MBD}}/Q_{\mathrm{PBE}}$ reaches as high as 80 %, underscoring the importance of vdW density polarization in large aromatic systems. In contrast, Q_{PBE} shows signs of saturation in C4C4PD, with the value approximately 30 % lower than expected from linear extrapolation. This saturation is observed

for other semi-local and hybrid functionals as well (see Figure A8), indicating the possible limitations of these functionals in reproducing $\rho(\mathbf{r})$ for larger aromatic systems.

4.6 Enhanced Noncovalent Interaction Regions in Molecules and Folded Proteins

4.6.1 S12L and L7+ Complexes

In the framework of the NCI analysis [259, 260], the key quantity is the reduced density gradient (RDG):

$$s(\mathbf{r}) = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho(\mathbf{r})|}{\rho^{4/3}(\mathbf{r})} \,. \tag{4.18}$$

As demonstrated by Johnson *et al.* [259], noncovalent interactions are associated with regions where $s(\mathbf{r}) \to 0$, with stripe-like features in $s(\mathbf{r})$ vs. $\rho(\mathbf{r})$ plots serving as fingerprints for NCIs (see Figure 4.11a). NCIs can be distinguished as attractive or repulsive via the second eigenvalue λ_2 of the electron-density Hessian matrix ($\lambda_1 < \lambda_2 < \lambda_3$), where $\lambda_2 < 0$ indicates attraction, and $\lambda_2 > 0$ signifies repulsion [259, 260]. The product $\rho(\mathbf{r})\text{sign}(\lambda_2)$ effectively captures interaction strength, and NCIs can be visualized as 3D isosurfaces of $s(\mathbf{r})$. The gradient isosurfaces are colored according to $\rho(\mathbf{r})\text{sign}(\lambda_2)$, where blue indicates strong attractive interactions (such as hydrogen bonding), green represents weak vdW attractions, and red denotes steric repulsion.

To focus on non-covalent interactions, charge density cutoffs exclude covalent bonds and regions near nuclei, typically using $\rho_c = 0.05$ a.u. [259, 260]. NCI isosurfaces are defined over an s_{iso} range, usually $0.2 - 0.7$, with the specific value chosen to match NCI signatures from the 2D plots (see Figure 4.11a). An example NCI isosurface for the 7b host-guest complex, using $s_{iso} = 0.5$, is displayed in the inset, demonstrating the correspondence between the $s(\rho)$ diagram and the 3D NCI isosurface. Please see Appendix A7 for computational details on the NCI analysis.

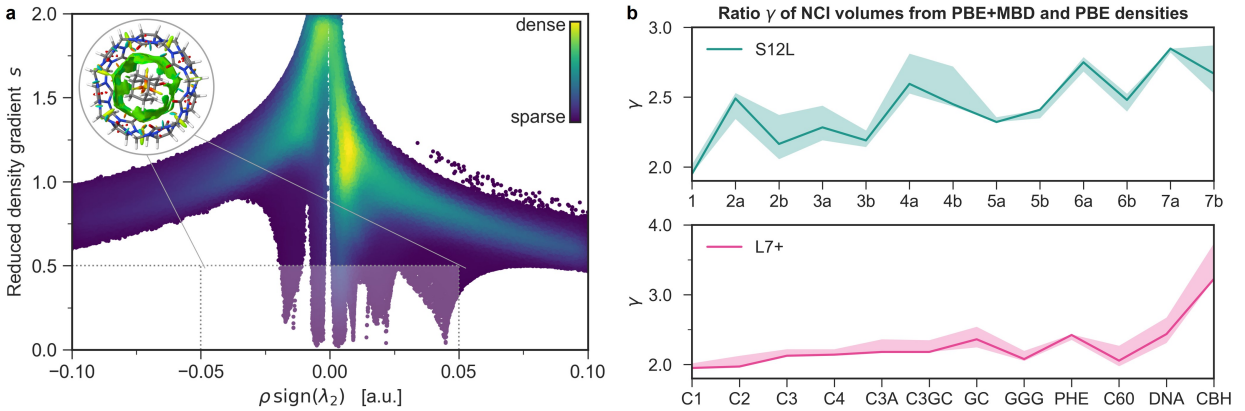

Figure 4.11: NCI analysis and NCI isosurface volumes explained. **(a)** Two-dimensional NCI plot computed from PBE+MBD@FCO density at the example of the 7b host-guest complex. Color coding indicates the statistical distribution of the data points on the graph: the brighter the color, the more data points are clustered there. The shaded region shows the subset of data points used to generate the NCI isosurface $s_{iso} = 0.5$ (the inset). **(b)** The ratio γ of NCI isosurface volumes computed from PBE+MBD@FCO and PBE densities for S12L (top) and L7+ (bottom) molecular sets. The lines display values computed at $s_{iso} = 0.5$, and shaded regions indicate the spread of γ when $s_{iso} = 0.4 - 0.6$.

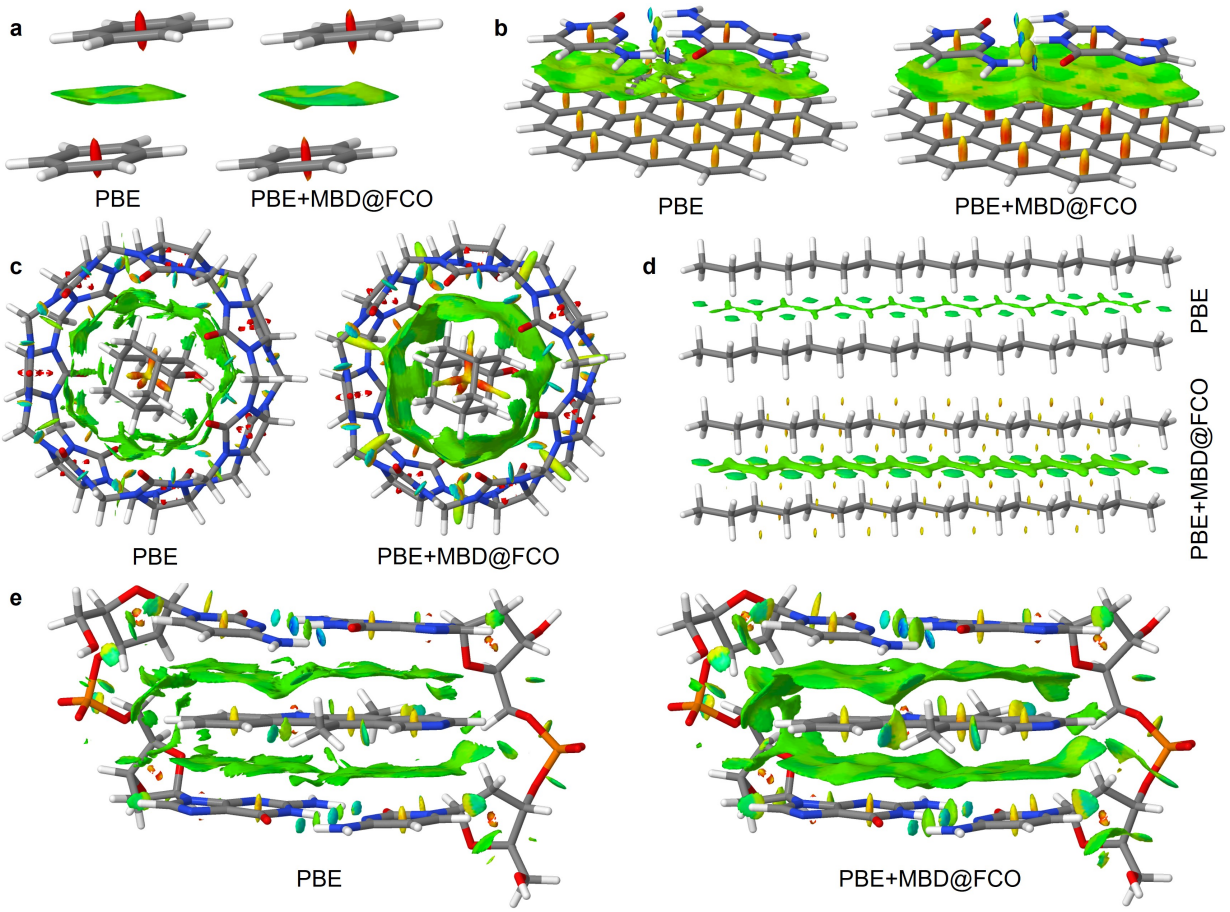

Figure 4.12: vdW contribution to the density makes NCI isosurfaces more connected. NCI isosurfaces of density gradient s (magnitude of s and coloring range are given in a.u.) calculated using PBE and PBE+MBD@FCO charge densities for **(a)** benzene dimer: $s = 0.5$, $[-0.02, 0.02]$; **(b)** C3GC (L7+): $s = 0.5$, $[-0.05, 0.04]$; **(c)** 7b (S12L): $s = 0.5$, $[-0.03, 0.03]$; **(d)** octadecane dimer (CBH, L7+): $s = 0.65$, $[-0.02, 0.02]$; **(e)** DNA-ellipticine: $s = 0.5$, $[-0.05, 0.05]$. Isosurfaces are plotted using the JMOL software.

We explore the impact of vdW polarization density by adding $\rho_{\text{pol}}(\mathbf{r})$ to the PBE density $\rho_{\text{PBE}}(\mathbf{r})$, yielding $\rho_{\text{tot}}(\mathbf{r}) = \rho_{\text{PBE}}(\mathbf{r}) + \rho_{\text{pol}}(\mathbf{r})$ (see Section 4.3 for explanations). NCI plots show significant changes when vdW polarization is included, as seen in Figure 4.12. The gradient isosurfaces reveal how vdW polarization thickens green vdW regions and elongates red steric repulsion zones. Larger systems, such as C3GC (Figure 4.12b) and DNA-ellipticine (Figure 4.12e), exhibit more connected and continuous vdW isosurfaces, providing a more coherent picture of vdW interactions. Additionally, repulsive regions in $\pi - \pi$ systems become more pronounced with MBD polarization, particularly in comparison to the benzene dimer.

The octadecane dimer (CBH, Figure 4.12d) offers a striking example. Without MBD, PBE density barely captures weak vdW attraction between chains, while small repulsive contacts between hydrogens are completely missed. These repulsive interactions appear with PBE+MBD@FCO, along with stronger, more widespread vdW attractions. Hydrogen bonds are also better resolved with MBD@FCO, as seen in C3GC and DNA-ellipticine.

To quantify these changes, we compute the total volume of NCI isosurfaces (V_{NCI}) using the NCIMILANO software [284]. The ratio $\gamma = V_{\text{NCI}}^{\text{PBE+MBD}} / V_{\text{NCI}}^{\text{PBE}}$ provides a metric reflecting the relative enhancement of interaction regions. We used $s_{\text{iso}} = 0.5$ for all systems and found minimal variation in γ within the typical NCI range of $s_{\text{iso}} = 0.4 - 0.6$, leaving overall trends unchanged (see Figure 4.11b).

The results for the S12L and L7+ datasets in Figure 4.11b show substantial increases in NCI volumes across all systems. The largest enhancement is found in alkane chains (CBH, γ = 3.23), consistent with vdW-driven interactions, while the smallest change occurs in the benzene dimer (γ = 1.95). The γ ratio shows a weak dependence on system size, increasing only slightly from 1.95 to 2.14 in the C1−C4 sequence of PAHs. On the other hand, γ is sensitive to interaction type, with non-vdW systems like water and formic acid dimers or bicyclo[2.2.2]octene (steric repulsion) [259] exhibiting γ values close to 1. This highlights the method's ability to distinguish vdW-driven ($\gamma = 2-3$) from non-vdW ($\gamma \sim 1$) interactions.

4.6.2 Results for the Fip35-WW Protein

As a representative biosystem, we examine the vdW polarization effects in the Fip35-WW protein, whose plasmon-like interactions with solvent were previously studied in detail [263]. From folding trajectories sampled in Ref. [285], we selected three representative atomic structures corresponding to unfolded, folded, and transition states. To focus on intra-protein interactions, water molecules and counter-ions were removed, and the resulting geometries were used for PBE and MBD calculations.

We first analyze the energetics, computing the relative energies of the three protein configurations (Figure 4.13). The semi-local PBE functional without long-range vdW interactions predicts a lower energy for the transition state than the folded state. In contrast, both MBD@rsSCS and MBD@FCO predict the folded state as the lowest-energy structure, highlighting the crucial stabilization provided by vdW dispersion. Notably, the MBD@FCO method, shown to give reliable prediction of dispersion energy for S12L systems in Figure 4.1, predicts a stabilization of 125 kcal/mol for the folded structure, driven by vdW forces. While these calculations neglect the influence of water and ions in physiological conditions, they still offer insight into the critical role of vdW dispersion in protein folding, in line with earlier findings [263, 286, 287].

Next, we examine the NCI isosurfaces (Figure 4.14). The general trends are consistent with those observed for S12L and L7+ molecules: the inclusion of MBD@FCO polarization density leads to smoother, more connected NCI isosurfaces, quantified by a volume enhancement factor $\gamma \approx 2.5$. To illustrate this, we zoom into the regions around selected residues (TRP8, ARG17, TYR20) of Fip35-WW. Using semi-local density, the NCI isosurfaces appear sparse and disconnected, complicating the interpretation of the vdW interactions near the residues. On the con-

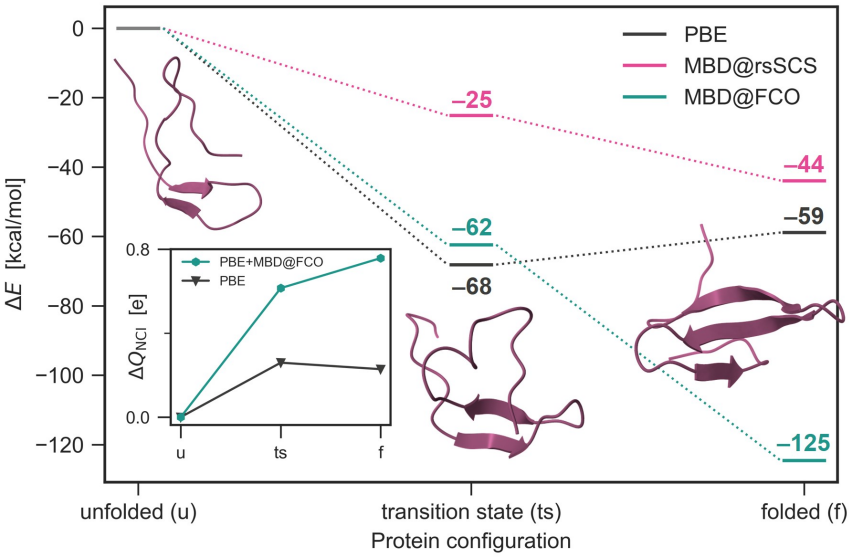

Figure 4.13: vdW interactions favour the folded configuration of the protein. The energy of Fip35-WW protein (relative to the unfolded state) along the folding trajectory, as calculated by PBE, MBD@rsSCS and MBD@FCO methods. The inset shows the change of NCI charge relative to the unfolded state, as computed by PBE and PBE+MBD@FCO methods.

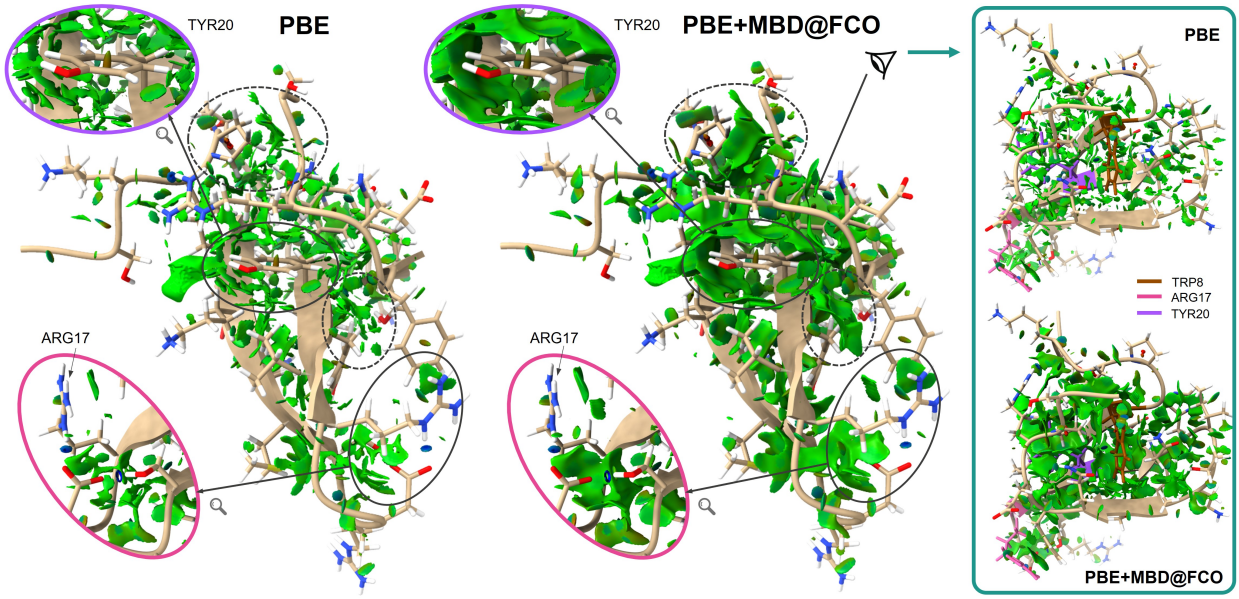

Figure 4.14: vdW-corrected vdW-corrected NCI analysis highlights interacting regions in the protein. NCI isosurfaces ($s = 0.5$) of Fip35-WW protein in the folded state computed using PBE and PBE+MBD@FCO density. Regions of the protein structure with the largest differences are highlighted in ellipsoids, with the vicinity of TYR20 and ARG17 residues additionally displayed in magnification. The right panel shows the view on Fip35-WW from the direction of TRP8 residue (in brown), as indicated by the side eye pictogram. For reference, ARG17 and TYR20 residues shown in the central panel are colored pink and purple, respectively. The images are created in the ChimeraX software [288] using the density gradients calculated with the NCIMILANO program [284].

trary, with PBE+MBD@FCO, the isosurfaces reveal significant vdW interactions between the residues. Many of them are encapsulated in green "vdW cages", which witness their additional stabilization by dispersion. Similar observations can be made for other regions of the protein structure.

These observations can be quantified by measuring the total charge Q_{NCI} contained in the NCI isosurfaces. Already in the unfolded (u) state, PBE+MBD@FCO density predicts much larger $Q^{u}_{\mathrm{NCI}} = 3.18\,e$, which is roughly 2.5 times the PBE prediction of $1.26\,e$. Along the folding trajectory, this interacting charge enhancement is further increased, as illustrated in the inset of Figure 4.13 by plotting $\Delta Q_{\mathrm{NCI}} = Q_{\mathrm{NCI}} - Q^{u}_{\mathrm{NCI}}$. Remarkably, the PBE+MBD@FCO charge shows a monotonous increase by up to $0.77\,e$ in the folded state, while PBE exhibits a maximum growth of only $0.26\,e$ in the transition state. For both methods trends in Q_{NCI} are consistent with energy ordering (Figure 4.13), which bridges the connection between the extent of interaction isosurfaces and their energy contributions.

Together with hydrogen bonding, hydrophobic interactions, and ion pairing, vdW dispersion is one of the main driving forces for protein folding [286, 287]. Our results underscore the importance of vdW dispersion and enable a better understanding of regions that are responsible for intraprotein interactions. In summary, incorporating dispersion-induced polarization enhances NCI features and facilitates their identification, particularly in larger, more complex biosystems. Our model can also treat interaction with solvent and charges, which are inherent for some amino acids (Asp, Glu, Lys, Arg) and surrounding ions. Electric fields from these charges can amplify dispersion interactions in amino acids by up to 35% [105]. The relative importance of electric-field-driven vdW effects in proteins needs further assessment, however, it might compete with salt bridges and cation-π interactions.

4.7 Discussion and Outlook

In this work, we investigated the shifts in charge density due to (long-range) vdW interactions in chemical and biological systems, using the MBD model [40] with an appropriate parametrization and dipole-coupling tensor. We demonstrated that while such polarization can be safely ignored in small molecules like benzene, it becomes significant in larger, extended systems from S12L and L7+ datasets, sometimes surpassing the magnitude of $\Delta\rho(\mathbf{r})$ from semi-local DFAs like for CBH alkane chain.

The vdW polarization also impacts chemical density-based analysis, as shown by the NCI index. Our results highlight how vdW polarization enhances the visualization of noncovalent interaction regions and requires a revision of our understanding of NCI surfaces. The Fip35-WW protein example illustrates that long-range vdW interactions are crucial for understanding protein folding, and thus vdW-induced density shifts are essential. Furthermore, vdW polarization may trigger higher-order interactions, both within proteins and with their solvent, via the dipole-correlated Coulomb singles mechanism, contributing up to 6 kJ/mol for S12L-like systems [26]. This effect is expected to increase in proteins due to their delocalized interactions [172, 263, 289, 290] and the presence of charged residues [105]. Hence, further exploration of the interplay between dispersion and polarization in complex systems is necessary.

The MBD@FCO method provides accurate dispersion energies and long-range correlation contributions to electronic density, making it a powerful vdW approach. Notably, this is achieved without empirical parameters, making the model fully self-contained. This underscores the importance of seamless treatment of long-range correlations in future DFAs. However, an open challenge remains: how to couple a vdW method like MBD@FCO with existing DFAs to avoid double-counting correlation effects at certain interatomic distances? The common approach of using a damping parameter introduces empirical adjustments that obscure the physical meaning of vdW energy [271–273]. One solution could be using "dispersionless" DFAs [291], such as those proposed by Pernal *et al.* [292], or mixing HF exchange with PBE correlation [273]. Another option is to obtain the range-separation parameter based on density instead of energy, minimizing errors by fitting the total DFA+MBD@FCO density to high-accuracy *ab initio* densities, such as those from CCSD for small molecules.

While our results were obtained using the semi-local PBE functional, the trends and conclusions apply to other DFAs, as confirmed by our comparison of HF and HF+MBD@FCO densities (see Appendix A6). From the perspective of NCI analysis, MBD@FCO densities can be computed and added to non-self-consistent atom-additive (promolecular) densities often used for large systems like proteins, improving the connectivity and smoothness of NCI isosurfaces. The only required inputs are atomic geometries and Hirshfeld volume ratios, or alternatives like charge population analysis ratios [152]. All the analyses presented here can be replicated using post-processing PYTHON scripts working with outputs from the LIBMBD code [160].

Our work demonstrates that the MBD model can be used to bridge accurate calculations of dispersion energy with efficient estimation of vdW-induced charge density, providing an intimate link between semi-local density-functional theory and nonlocal electron correlation energy models.

Chapter 5

Reduced Density Matrix Formalism for Many-Body Dispersion Method

All derivations in this chapter are done by A. Khabibrakhmanov, following the ideas and under the guidance of Dr. Matteo Gori.

Van der Waals (vdW) dispersion interactions are inherently many-body and nonlocal, arising from correlated electronic charge fluctuations. These interactions play a fundamental role in understanding non-covalent forces in chemical, biological, and materials systems. As discussed in Section 2.3, the accurate treatment of the vdW dispersion requires a sophisticated theory, accounting for the nonlocality of underlying electron correlations. The MBD model, as a coarse-grained approximation to the exact ACFD-RPA formalism, offers a good balance between computational cost, accuracy, and physical interpretability. In Chapter 4, we demonstrated the connection between the charge density of the coupled MBD modes and correlation-induced polarization of real electron density.

In the present chapter, we further develop these ideas and introduce a key formal advancement: the reduced-density matrix representation for the MBD model. This framework offers a fresh theoretical perspective on the MBD method, transitioning from a coarse-grained picture of fluctuating dipoles to continuous particle and energy density fields. Namely, this enables the definition of a local MBD energy density, $e_{\mathrm{MBD}}(\mathbf{r})$, which satisfies the integral relation

$$E_{\mathrm{MBD}} = \int_{\mathbb{R}^3} e_{\mathrm{MBD}}(\mathbf{r})\, d\mathbf{r}\,. \tag{5.1}$$

While $e_{\mathrm{MBD}}(\mathbf{r})$ is not uniquely defined, the proposed density-matrix framework offers a physically meaningful method to extract continuous spatial distributions from the coarse-grained MBD model.

The spatial localization of dispersion interactions has both theoretical and practical implications. The derived energy density, $\Delta e_{\mathrm{MBD}}(\mathbf{r})$, provides valuable insights into the spatial distribution of dispersion energy, as demonstrated by examples like the DNA-ellipticine complex. Mapping these interactions in real space makes it possible to identify functional groups or structural fragments most involved in dispersion stabilization, offering guidance for molecular design efforts in chemistry and biology. The visualized spatial extent of drudonic correlations over several angstroms, as seen in isosurface plots for DNA-ellipticine and C_{60} buckyball catcher complexes, underscores the nonlocal nature of vdW interactions and the importance of capturing these effects for accurate energy descriptions. These advanced visualization efforts go in the

same vein as the NCI analysis enhanced by dispersion-polarized densities in Chapter 4. Projecting MBD interactions onto real space also holds practical relevance for machine learning force fields (MLFFs). The ability to localize MBD energy contributions effectively enables the development of many-body-informed descriptors, enhancing the accuracy and applicability of ML models for systems where nonlocal vdW interactions play a crucial role [293].

Additionally, this chapter introduces a novel representation of the two-body MBD energy as a nonlocal correlation functional over dipole-coupled oscillator densities. Comparative analysis with the VV10 functional, a widely used nonlocal correlation model, reveals important similarities and distinctions in how these approaches capture nonlocal dispersion interactions. Despite differences in the spatial patterns of nonlocal correlations, both methods provide comparable long-range dispersion energy predictions, which align closely with reference SAPT values. Notably, the MBD@FCO two-body energy term exhibits surprisingly good agreement with SAPT even at short range, suggesting avenues for further refinement, such as improving the local polarizability functional in VV10 by leveraging insights from MBD@FCO.

In summary, this chapter provides a foundational step toward a more nuanced understanding of vdW dispersion interactions through a real-space density-matrix formalism, opening new possibilities for energy density analysis and advancements in density-functional development.

5.1 Reduced Density Matrix Formalism for MBD

The MBD Hamiltonian for a system of N coupled oscillators centered at $\mathbf{R}_A$, with frequencies ω_A, charges q_A and masses m_A reads

$$\hat{H}_{\mathrm{MBD}} = \sum_{A=1}^{N}\left[\frac{\hat{\mathbf{p}}_A^2}{2m_A} + \frac{m_A\omega_A^2}{2}(\mathbf{r}_A-\mathbf{R}_A)^2\right] + \frac{1}{2}\sum_{A\neq B} q_A q_B\,(\mathbf{r}_A-\mathbf{R}_A)^T\mathbf{T}_{AB}(\mathbf{r}_B-\mathbf{R}_B)\,, \tag{5.2}$$

where $\mathbf{T}_{AB}$ is the dipole interaction tensor. The corresponding Hamiltonian of N non-interacting QDOs is

$$\hat{H}_0 = \sum_{A=1}^{N}\left[\frac{\hat{\mathbf{p}}_A^2}{2m_A} + \frac{m_A\omega_A^2}{2}(\mathbf{r}_A-\mathbf{R}_A)^2\right]. \tag{5.3}$$

Denoting the ground state of the non-interacting QDO system as $|\mathbf{0}\rangle$ and the interacting (MBD) ground-state as $|\tilde{\mathbf{0}}\rangle$, the MBD interaction energy can be expressed as

$$E_{\mathrm{MBD}} = \langle\tilde{\mathbf{0}}|\hat{H}_{\mathrm{MBD}}|\tilde{\mathbf{0}}\rangle - \langle\mathbf{0}|\hat{H}_0|\mathbf{0}\rangle\,. \tag{5.4}$$

Let us introduce the MBD ground-state density matrix $\hat{\rho}_{\mathrm{MBD}} = |\tilde{\mathbf{0}}\rangle\langle\tilde{\mathbf{0}}|$ and the density matrix $\hat{\rho}_0 = |\mathbf{0}\rangle\langle\mathbf{0}|$ for the ground state of the non-interacting QDO system. We also define the shorthand notation for the set of coordinates $\underline{\mathbf{r}} = \{\mathbf{r}_1,...,\mathbf{r}_N\}$ and eigenstates of the N-body coordinate operator $|\underline{\mathbf{r}}\rangle \equiv |\mathbf{r}_1,...,\mathbf{r}_N\rangle$. Then, one can project the MBD ground state onto the eigenstates of the N-body coordinate operator using the identity relation for density operator $\mathbb{1}_{\mathbf{r}} = \int \mathrm{d}^N\mathbf{r}\,|\underline{\mathbf{r}}\rangle\langle\underline{\mathbf{r}}|$ with $\mathrm{d}^N\mathbf{r} = \prod_{A=1}^{N}\mathrm{d}\mathbf{r}_A$:

$$\begin{aligned}\langle\tilde{\mathbf{0}}|\hat{H}_{\mathrm{MBD}}|\tilde{\mathbf{0}}\rangle &= \langle\tilde{\mathbf{0}}|\mathbb{1}_{\mathbf{r}}\hat{H}_{\mathrm{MBD}}\mathbb{1}_{\mathbf{r}'}|\tilde{\mathbf{0}}\rangle = \langle\tilde{\mathbf{0}}|\int d^N\mathbf{r}\,|\underline{\mathbf{r}}\rangle\langle\underline{\mathbf{r}}|\hat{H}_{\mathrm{MBD}}\int d^N\mathbf{r}'\,|\underline{\mathbf{r}}'\rangle\langle\underline{\mathbf{r}}'|\tilde{\mathbf{0}}\rangle = \\ &= \iint d^N\mathbf{r}\,d^N\mathbf{r}'\,\langle\tilde{\mathbf{0}}|\underline{\mathbf{r}}\rangle\langle\underline{\mathbf{r}}|\hat{H}_{\mathrm{MBD}}|\underline{\mathbf{r}}'\rangle\langle\underline{\mathbf{r}}'|\tilde{\mathbf{0}}\rangle = \iint d^N\mathbf{r}\,d^N\mathbf{r}'\,\langle\underline{\mathbf{r}}|\hat{H}_{\mathrm{MBD}}|\underline{\mathbf{r}}'\rangle\langle\underline{\mathbf{r}}'|\hat{\rho}_{\mathrm{MBD}}|\underline{\mathbf{r}}\rangle\,.\end{aligned} \tag{5.5}$$

Performing similar calculations for $\langle\mathbf{0}|\hat{H}_0|\mathbf{0}\rangle$ and introducing a more compact notation $H(\underline{\mathbf{r}},\underline{\mathbf{r}}') = \langle\underline{\mathbf{r}}|\hat{H}|\underline{\mathbf{r}}'\rangle$, $\gamma^{(N)}(\underline{\mathbf{r}};\underline{\mathbf{r}}') = \langle\underline{\mathbf{r}}'|\,\hat{\rho}\,|\underline{\mathbf{r}}\rangle$, the MBD energy can be rewritten as

$$E_{\mathrm{MBD}} = \iint d^N\mathbf{r}\,d^N\mathbf{r}'\,\left[H_{\mathrm{MBD}}(\underline{\mathbf{r}},\underline{\mathbf{r}}')\gamma^{(N)}_{\mathrm{MBD}}(\underline{\mathbf{r}};\underline{\mathbf{r}}') - H_0(\underline{\mathbf{r}},\underline{\mathbf{r}}')\gamma^{(N)}_0(\underline{\mathbf{r}};\underline{\mathbf{r}}')\right], \tag{5.6}$$

where $\gamma^{(N)}(\underline{\mathbf{r}};\underline{\mathbf{r}}')$ is the *N-body density matrix*. On the other hand, we have the definition of MBD energy density (5.1). Therefore, to obtain the explicit expression for $e_{\mathrm{MBD}}(\mathbf{r})$, one should perform integration over all but three of the $6N$ coordinates in Eq. (5.6).

For this purpose, one needs to derive 1- and 2-body *reduced density matrices* (RDM). By definition, 1- and 2-RDM $\gamma_A^{(1)}$ and $\gamma_{AB}^{(2)}$ are obtained by tracing $\gamma^{(N)}(\underline{\mathbf{r}};\underline{\mathbf{r}}')$ over all other particles:

$$\gamma_A^{(1)}(\mathbf{r};\mathbf{r}') = \int \gamma^{(N)}(\underline{\mathbf{r}}_A;\underline{\mathbf{r}}'_A) \prod_{i\neq A} \delta(\mathbf{r}_i - \mathbf{r}'_i)\, d\mathbf{r}_i d\mathbf{r}'_i \,, \tag{5.7}$$

$$\gamma_{AB}^{(2)}(\mathbf{r},\mathbf{r}';\mathbf{r}'',\mathbf{r}''') = \int \gamma^{(N)}(\underline{\mathbf{r}\mathbf{r}'}_{AB};\underline{\mathbf{r}''\mathbf{r}'''}_{AB}) \prod_{i\neq A,B} \delta(\mathbf{r}_i - \mathbf{r}'_i)\, d\mathbf{r}_i d\mathbf{r}'_i \,, \tag{5.8}$$

where we introduced the shorthand notation for coordinates, explicitly tracking the indices of non-integrated variables in the N-body coordinate set:[1]

$$\begin{aligned}
\underline{\mathbf{r}}_A &= \{\mathbf{r}_1,\ldots,\mathbf{r}_{A-1},\mathbf{r},\mathbf{r}_{A+1},\ldots,\mathbf{r}_N\}\,; \quad \underline{\mathbf{r}}'_A = \{\mathbf{r}'_1,\ldots,\mathbf{r}'_{A-1},\mathbf{r}',\mathbf{r}'_{A+1},\ldots,\mathbf{r}'_N\}\,; \\
\underline{\mathbf{r}\mathbf{r}'}_{AB} &= \{\mathbf{r}_1,\ldots,\mathbf{r}_{A-1},\mathbf{r},\mathbf{r}_{A+1},\ldots,\mathbf{r}_{B-1},\mathbf{r}',\mathbf{r}_{B+1},\ldots,\mathbf{r}_N\}\,; \\
\underline{\mathbf{r}''\mathbf{r}'''}_{AB} &= \{\mathbf{r}'_1,\ldots,\mathbf{r}'_{A-1},\mathbf{r}'',\mathbf{r}'_{A+1},\ldots,\mathbf{r}'_{B-1},\mathbf{r}''',\mathbf{r}'_{B+1},\ldots,\mathbf{r}'_N\}\,.
\end{aligned}$$

As usually, we also define one- and two-body *particle density functions* $n_A^{(1)}(\mathbf{r})$ and $n_{AB}^{(2)}(\mathbf{r},\mathbf{r}')$ being diagonal elements of 1- and 2-RDM.

$$n_A^{(1)}(\mathbf{r}) = \gamma_A^{(1)}(\mathbf{r};\mathbf{r})\,, \quad n_{AB}^{(2)}(\mathbf{r},\mathbf{r}') = \gamma_{AB}^{(2)}(\mathbf{r},\mathbf{r}';\mathbf{r},\mathbf{r}')\,, \tag{5.9}$$

while the respective quantities for the whole system are obtained by summing over all drudons and their pairs:

$$\begin{aligned}
\gamma^{(1)}(\mathbf{r};\mathbf{r}') &= \sum_{A=1}^{N} \gamma_A^{(1)}(\mathbf{r};\mathbf{r}')\,; \quad \gamma^{(2)}(\mathbf{r},\mathbf{r}';\mathbf{r}'',\mathbf{r}''') = \sum_{A\neq B} \gamma_{AB}^{(2)}(\mathbf{r},\mathbf{r}';\mathbf{r}'',\mathbf{r}''')\,; \\
n^{(1)}(\mathbf{r}) &= \sum_{A=1}^{N} n_A^{(1)}(\mathbf{r})\,; \quad n^{(2)}(\mathbf{r},\mathbf{r}') = \sum_{A\neq B} n_{AB}^{(2)}(\mathbf{r},\mathbf{r}')\,.
\end{aligned}$$

Given these definitions, we can calculate MBD density matrices in an explicit form. For the N-body density matrix $\gamma_{\mathrm{MBD}}^{(N)}(\underline{\mathbf{r}};\underline{\mathbf{r}}')$, we have

$$\begin{aligned}
\gamma_{\mathrm{MBD}}^{(N)}(\underline{\mathbf{r}};\underline{\mathbf{r}}') = \langle\underline{\mathbf{r}}'|\tilde{\mathbf{0}}\rangle\langle\tilde{\mathbf{0}}|\underline{\mathbf{r}}\rangle = \Psi_{\mathrm{MBD}}(\underline{\mathbf{r}}')\Psi_{\mathrm{MBD}}^{*}(\underline{\mathbf{r}}) = \left[\left(\prod_{A=1}^{N} m_A^3\right)\left(\prod_{\alpha=1}^{3N}\frac{\tilde{\omega}_\alpha}{\pi\hbar}\right)\right]^{1/2} \times \\
\times \exp\left[-\frac{1}{2}\sum_{A,B=1}^{N}(\mathbf{r}'_A - \mathbf{R}_A)^T\mathbf{C}_{AB}(\mathbf{r}'_B - \mathbf{R}_B) - \frac{1}{2}\sum_{A,B=1}^{N}(\mathbf{r}_A - \mathbf{R}_A)^T\mathbf{C}_{AB}(\mathbf{r}_B - \mathbf{R}_B)\right],
\end{aligned} \tag{5.10}$$

where we used Eq. (4.8) for the ground-state MBD wave function. In the following derivations, we will omit the dependence on nuclear coordinates $\{\mathbf{R}_A\}$ for clarity, i.e. drudonic coordinates should be understood as differences with the respective nuclear coordinates as in (5.10); the same applies to primed variables. We will also drop transposition mark $(\cdot)^T$ (unless explicitly noted), and in all vector-matrix-vector products the left vector shall be understood as a row

[1] The necessity to keep track of integration indices arises due to *distinguishability* of quasiparticles in MBD (each drudon tethered to a particular nucleus), while electronic structure theory deals with *indistinguishable* fermions. Therefore, in standard textbooks on density matrix theory, the non-integrated coordinate is often fixed to be the first, and prefactors like N and $N(N-1)/2$, accounting for possible permutations, arise in front of the integrals in Eqs. (5.7) and (5.8), respectively [36].

vector, while the right one is a column vector. In addition, we will denote the normalization constant in (5.10) as Q. When arrived to the final answers, all omitted symbols will be restored in the formulas.

To perform integration in Eq. (5.7), it is convenient to separate variables in $\gamma^{(N)}(\underline{\mathbf{r}}_A;\underline{\mathbf{r}}'_A)$:

$$\begin{aligned}\gamma^{(N)}(\underline{\mathbf{r}}_A;\underline{\mathbf{r}}'_A) =& Q\exp\left[-\frac{1}{2}\sum_{i,j\neq A}\mathbf{r}_i\mathbf{C}_{ij}\mathbf{r}_j-\frac{1}{2}\sum_{j\neq A}\mathbf{r}\mathbf{C}_{Aj}\mathbf{r}_j-\frac{1}{2}\sum_{j\neq A}\mathbf{r}_j\mathbf{C}_{jA}\mathbf{r}-\frac{1}{2}\mathbf{r}\mathbf{C}_{AA}\mathbf{r}\right]\times\\ &\times\exp\left[-\frac{1}{2}\sum_{i,j\neq A}\mathbf{r}_i\mathbf{C}_{ij}\mathbf{r}_j-\frac{1}{2}\sum_{j\neq A}\mathbf{r}'\mathbf{C}_{Aj}\mathbf{r}_j-\frac{1}{2}\sum_{j\neq A}\mathbf{r}_j\mathbf{C}_{jA}\mathbf{r}'-\frac{1}{2}\mathbf{r}'\mathbf{C}_{AA}\mathbf{r}'\right]=\\ =&Q\exp\left[-\frac{1}{2}\mathbf{r}\mathbf{C}_{AA}\mathbf{r}-\frac{1}{2}\mathbf{r}'\mathbf{C}_{AA}\mathbf{r}'\right]\exp\left[-\sum_{j\neq A}(\mathbf{r}'+\mathbf{r})\mathbf{C}_{Aj}\mathbf{r}_j-\sum_{i,j\neq A}\mathbf{r}_i\mathbf{C}_{ij}\mathbf{r}_j\right],\end{aligned} \tag{5.11}$$

where we used the symmetry of $\mathbf{C}$ matrix. Thus, we have a product of terms quadratic in $\mathbf{r}_i$, linear in $\mathbf{r}_i$, and the term independent on integration variables. Therefore, introducing $\mathbf{F} = (\mathbf{r}'+\mathbf{r})^T\mathbf{C}_A$ and forming the complete square in the exponent, the integral for $\gamma^{(1)}_{\mathrm{MBD},A}$ can be computed as

$$\begin{aligned}\gamma^{(1)}_{\mathrm{MBD},A}(\mathbf{r};\mathbf{r}') &= Q\,e^{-\frac{1}{2}(\mathbf{r}'\mathbf{C}_{AA}\mathbf{r}'+\mathbf{r}\mathbf{C}_{AA}\mathbf{r})}\int\limits_{\mathbb{R}^{3N-3}}\exp\left[-\sum_{j\neq A}\mathbf{F}^{\mathrm{T}}_j\mathbf{r}_j-\sum_{i,j\neq A}\mathbf{r}_i\mathbf{C}_{ij}\mathbf{r}_j\right]\prod_{j\neq A}d\mathbf{r}_j=\\ &= Q\times\sqrt{\frac{\pi^{3N-3}}{\det\mathbf{C}^{/A}}}\times e^{-\frac{1}{2}(\mathbf{r}'\mathbf{C}_{AA}\mathbf{r}'+\mathbf{r}\mathbf{C}_{AA}\mathbf{r})}\times\exp\left[\frac{1}{4}(\mathbf{r}'+\mathbf{r})^T\mathbf{C}_A^T[\mathbf{C}^{/A}]^{-1}\mathbf{C}_A(\mathbf{r}'+\mathbf{r})\right],\end{aligned} \tag{5.12}$$

where we use the notation for the MBD correlation matrix as explained in Figure 4.2. Finally, introducing auxiliary 3×3 matrices $\mathbf{A}_{AA}$ and $\mathbf{B}_{AA}$

$$\mathbf{A}_{AA}=\mathbf{C}_{AA}-\frac{1}{2}\mathbf{B}_{AA}\,,\quad \mathbf{B}_{AA}=\mathbf{C}_A^T[\mathbf{C}^{/A}]^{-1}\mathbf{C}_A\,, \tag{5.13}$$

and recovering all omitted symbols and variables, we obtain:

$$\begin{aligned}\gamma^{(1)}_{\mathrm{MBD},A}(\mathbf{r};\mathbf{r}') = \frac{1}{\pi^{3/2}\Lambda_A^3}\exp\Big\{-\frac{1}{2}\Big[&(\mathbf{r}-\mathbf{R}_A)^T\mathbf{A}_{AA}(\mathbf{r}-\mathbf{R}_A)+\\ &+(\mathbf{r}'-\mathbf{R}_A)^T\mathbf{A}_{AA}\left(\mathbf{r}'-\mathbf{R}_A\right)\Big]+\frac{1}{2}(\mathbf{r}'-\mathbf{R}_A)^T\mathbf{B}_{AA}(\mathbf{r}-\mathbf{R}_A)\Big\}\,.\end{aligned} \tag{5.14}$$

Here, we introduced $\Lambda_A^3=\sqrt{\det\mathbf{C}^{/A}/\det\mathbf{C}}$ analogous to Eq. (4.10). Analogous calculations can be performed for non-interacting 1-RDM, $\gamma^{(1)}_{0,A}(\mathbf{r};\mathbf{r}')$, using the wave function of non-interacting QDOs (4.5). Those calculations are trivial to carry out, so here we present only the final answer:

$$\gamma^{(1)}_{0,A}(\mathbf{r};\mathbf{r}')=\frac{1}{\pi^{3/2}\lambda_A^3}\exp\left\{-\frac{1}{2}\left[(\mathbf{r}-\mathbf{R}_A)^T\lambda_A^{-2}(\mathbf{r}-\mathbf{R}_A)+(\mathbf{r}'-\mathbf{R}_A)^T\lambda_A^{-2}(\mathbf{r}'-\mathbf{R}_A)\right]\right\}\,, \tag{5.15}$$

with $\lambda_A=\sqrt{\hbar/m_A\omega_A}$. The corresponding one-body particle density functions are given by

$$\begin{aligned}n^{(1)}_{0,A}(\mathbf{r})&=\gamma^{(1)}_{0,A}(\mathbf{r};\mathbf{r})=\frac{1}{\pi^{3/2}\lambda_A^3}\exp\left\{-\lambda_A^{-2}(\mathbf{r}-\mathbf{R}_A)^2\right\}\,,\\ n^{(1)}_{\mathrm{MBD},A}(\mathbf{r})&=\gamma^{(1)}_{\mathrm{MBD},A}(\mathbf{r};\mathbf{r})=\frac{1}{\pi^{3/2}\Lambda_A^3}\exp\Big\{-(\mathbf{r}-\mathbf{R}_A)^T\mathbf{K}_{AA}(\mathbf{r}-\mathbf{R}_A)\Big\}\,,\end{aligned} \tag{5.16}$$

where $\mathbf{K}_{AA} = \mathbf{C}_{AA} - \mathbf{B}_{AA} = \mathbf{C}_{AA} - \mathbf{C}_A^T [\mathbf{C}^{/A}]^{-1} \mathbf{C}_A$ is the *reduced correlation matrix* of the A-th QDO in the MBD ground state. We note that the derived *particle* densities $n(\mathbf{r})$ are related to the *charge* densities $\rho(\mathbf{r})$ discussed in Chapter 4 via multiplication by charge, i.e. $\rho_i(\mathbf{r}) = q_i n_i^{(1)}(\mathbf{r})$.

To obtain 2-RDM, similar manipulations with correlation matrix indices have to be performed to separate the non-integrated variables $\mathbf{r}$, $\mathbf{r}'$, $\mathbf{r}''$ and $\mathbf{r}'''$:

$$\begin{aligned}\gamma^{(N)}(\underline{\mathbf{r}\mathbf{r}'}_{AB};\underline{\mathbf{r}''\mathbf{r}'''}_{AB}) = Q \times \exp\Bigg[-\frac{1}{2}\Bigg(&\sum_{i,j\neq A,B} \mathbf{r}_i \mathbf{C}_{ij} \mathbf{r}_j + \sum_{j\neq A,B} \mathbf{r}\mathbf{C}_{Aj}\mathbf{r}_j + \sum_{j\neq A,B} \mathbf{r}_j \mathbf{C}_{jA}\mathbf{r} + \\ &+ \sum_{j\neq A,B} \mathbf{r}'\mathbf{C}_{Bj}\mathbf{r}_j + \sum_{j\neq A,B} \mathbf{r}_j \mathbf{C}_{jB}\mathbf{r}' + \mathbf{r}\mathbf{C}_{AA}\mathbf{r} + \mathbf{r}'\mathbf{C}_{BB}\mathbf{r}' + \mathbf{r}\mathbf{C}_{AB}\mathbf{r}' + \mathbf{r}'\mathbf{C}_{BA}\mathbf{r}\Bigg)\Bigg] \\ &\times \exp\Big[\text{the same as above with } \mathbf{r}\to\mathbf{r}'',\ \mathbf{r}'\to\mathbf{r}'''\Big]\,.\end{aligned} \tag{5.17}$$

Benefiting from the symmetry of $\mathbf{C}$ and grouping similar terms together, we arrive at

$$\begin{aligned}\gamma^{(N)}(\underline{\mathbf{r}\mathbf{r}'}_{AB};\underline{\mathbf{r}''\mathbf{r}'''}_{AB}) = Q \times \exp\Bigg[&-\sum_{i,j\neq A,B} \mathbf{r}_i \mathbf{C}_{ij} \mathbf{r}_j - \sum_{j\neq A,B} \Big[(\mathbf{r}+\mathbf{r}'')\mathbf{C}_{Aj} + (\mathbf{r}'+\mathbf{r}''')\mathbf{C}_{Bj}\Big]\mathbf{r}_j \\ &- \frac{1}{2}(\mathbf{r},\mathbf{r}')\,\mathbf{G}_{AB}\begin{pmatrix}\mathbf{r}\\ \mathbf{r}'\end{pmatrix} - \frac{1}{2}(\mathbf{r}'',\mathbf{r}''')\,\mathbf{G}_{AB}\begin{pmatrix}\mathbf{r}''\\ \mathbf{r}'''\end{pmatrix}\Bigg]\,,\quad \mathbf{G}_{AB} = \begin{pmatrix}\mathbf{C}_{AA} & \mathbf{C}_{AB}\\ \mathbf{C}_{BA} & \mathbf{C}_{BB}\end{pmatrix}.\end{aligned} \tag{5.18}$$

where we used explicit matrix notation for the last two terms and introduced 6×6 matrix $\mathbf{G}_{AB}$. Following the same path as for integral in Eq. (5.12), one can obtain the explicit expression for 2-RDM as:

$$\begin{aligned}\gamma^{(2)}_{AB}(\mathbf{r},\mathbf{r}';\mathbf{r}'',\mathbf{r}''') = \frac{1}{\pi^3}\sqrt{\frac{\det\mathbf{C}}{\det\mathbf{C}^{/AB}}} \times \exp\Bigg[&-\frac{1}{2}(\mathbf{r},\mathbf{r}')\,\mathbf{G}_{AB}\begin{pmatrix}\mathbf{r}\\ \mathbf{r}'\end{pmatrix} \\ &- \frac{1}{2}(\mathbf{r}'',\mathbf{r}''')\,\mathbf{G}_{AB}\begin{pmatrix}\mathbf{r}''\\ \mathbf{r}'''\end{pmatrix} + \frac{1}{4}(\mathbf{r}+\mathbf{r}'',\mathbf{r}'+\mathbf{r}''')\,\mathbf{D}_{AB}\begin{pmatrix}\mathbf{r}+\mathbf{r}''\\ \mathbf{r}'+\mathbf{r}'''\end{pmatrix}\Bigg]\,,\end{aligned} \tag{5.19}$$

with 6×6 matrix $\mathbf{D}_{AB}$ defined as

$$\mathbf{D}_{AB} = \begin{pmatrix}\mathbf{C}_{A/B}^T[\mathbf{C}^{/AB}]^{-1}\mathbf{C}_{A/B} & \mathbf{C}_{A/B}^T[\mathbf{C}^{/AB}]^{-1}\mathbf{C}_{B/A}\\ \mathbf{C}_{B/A}^T[\mathbf{C}^{/AB}]^{-1}\mathbf{C}_{A/B} & \mathbf{C}_{B/A}^T[\mathbf{C}^{/AB}]^{-1}\mathbf{C}_{B/A}\end{pmatrix}. \tag{5.20}$$

The notation follows the same logic as before: the $(3N–6)\times(3N–6)$ matrix $\mathbf{C}^{/AB}$ is obtained from $\mathbf{C}$ by removing from it rows $\mathbf{C}_{A/B}^T$ and $\mathbf{C}_{B/A}^T$ (sized $3\times(3N–6)$ each) and columns $\mathbf{C}_{A/B}$ and $\mathbf{C}_{B/A}$ (sized $(3N–6)\times3$ each), which correspond to A-th and B-th QDO, respectively. Symbol / denotes that e.g. $\mathbf{C}_{A/B}$ is obtained from $\mathbf{C}_A$ by additionally removing the entries corresponding to the B-th QDO.

The two-body density function $n^{(2)}_{AB}(\mathbf{r},\mathbf{r}') = \gamma^{(2)}_{AB}(\mathbf{r},\mathbf{r}';\mathbf{r},\mathbf{r}')$ is then given by

$$\begin{aligned}n^{(2)}_{AB}(\mathbf{r},\mathbf{r}') = \frac{\pi^{-3}}{\Lambda^6_{AB}}\exp\Bigg[-(\mathbf{r},\mathbf{r}')\,\mathbf{J}_{AB}\begin{pmatrix}\mathbf{r}\\ \mathbf{r}'\end{pmatrix}\Bigg] = \frac{\pi^{-3}}{\Lambda^6_{AB}}\exp\Bigg[&-(\mathbf{r}-\mathbf{R}_A)^T\mathbf{J}^{11}_{AB}(\mathbf{r}-\mathbf{R}_A) \\ &- (\mathbf{r}'-\mathbf{R}_B)^T\mathbf{J}^{22}_{AB}(\mathbf{r}'-\mathbf{R}_B) - (\mathbf{r}-\mathbf{R}_A)^T\mathbf{J}^{12}_{AB}(\mathbf{r}'-\mathbf{R}_B) - (\mathbf{r}'-\mathbf{R}_B)^T\mathbf{J}^{21}_{AB}(\mathbf{r}-\mathbf{R}_A)\Bigg]\,,\end{aligned} \tag{5.21}$$

where we introduced $\Lambda^6_{AB} = \sqrt{\det \mathbf{C}'^{AB}/\det \mathbf{C}}$, and the 6×6 matrix $\mathbf{J}_{AB}$ is composed of 3×3 blocks $\mathbf{J}^{ij}_{AB}$ as

$$\mathbf{J}_{AB} = \mathbf{G}_{AB} - \mathbf{D}_{AB} = \begin{pmatrix} \mathbf{J}^{11}_{AB} & \mathbf{J}^{12}_{AB} \\ \mathbf{J}^{21}_{AB} & \mathbf{J}^{22}_{AB} \end{pmatrix}. \tag{5.22}$$

To compute the MBD energy density in Eq. (5.1), we also need to find $H(\underline{\mathbf{r}},\underline{\mathbf{r}}') = \langle \underline{\mathbf{r}}|\hat{H}|\underline{\mathbf{r}}'\rangle$. For convenience, we will calculate them separately for kinetic, potential, and dipole interaction energy terms in the Hamiltonian (5.2).

5.1.1 One-Body Terms in Hamiltonian

Kinetic energy

Let us denote $\hat{T}_A = \hat{\mathbf{p}}^2_A/2m_A$. Since $|\underline{\mathbf{r}}\rangle = |\mathbf{r}_1\rangle\otimes|\mathbf{r}_2\rangle\ldots\otimes|\mathbf{r}_N\rangle$ and $\langle\mathbf{r}'_i|\mathbf{r}_j\rangle = \delta_{ij}\delta(\mathbf{r}_j-\mathbf{r}'_i)$, we can write

$$\langle\underline{\mathbf{r}}|\hat{T}_A|\underline{\mathbf{r}}'\rangle = \langle\mathbf{r}_A|\hat{T}_A|\mathbf{r}'_A\rangle\prod_{i\neq A}\delta(\mathbf{r}_i-\mathbf{r}'_i)\,. \tag{5.23}$$

Using the well-known relation for matrix elements of momentum operator in coordinate representation $\langle\mathbf{r}|\hat{\mathbf{p}}^2|\mathbf{r}'\rangle = -\hbar^2\nabla^2_{\mathbf{r}}\delta(\mathbf{r}-\mathbf{r}')$, for the kinetic energy matrix element we have:

$$\begin{aligned}\langle\tilde{\mathbf{0}}|\hat{T}_A|\tilde{\mathbf{0}}\rangle = \iint d^N\mathbf{r}\,d^N\mathbf{r}'\,\langle\mathbf{r}_A|\hat{T}_A|\mathbf{r}'_A\rangle\left(\prod_{i\neq A}\delta(\mathbf{r}_i-\mathbf{r}'_i)\right)\gamma^{(N)}_{\mathrm{MBD}}(\underline{\mathbf{r}};\underline{\mathbf{r}}') = -\frac{\hbar^2}{2m_A}\times \\ \iint d\mathbf{r}d\mathbf{r}'\,\nabla^2_{\mathbf{r}}\delta(\mathbf{r}-\mathbf{r}')\gamma^{(1)}_{\mathrm{MBD},A}(\mathbf{r};\mathbf{r}') = \frac{\hbar^2}{2m_A}\int d\mathbf{r}\left[\nabla^2_{\mathbf{r}}\gamma^{(1)}_{\mathrm{MBD},A}(\mathbf{r};\mathbf{r}')\right]_{\mathbf{r}'=\mathbf{r}} = \int d\mathbf{r}\,t^{\mathrm{MBD}}_A(\mathbf{r})\,,\end{aligned} \tag{5.24}$$

where we used the property of delta-function derivative and introduced kinetic energy density $t^{\mathrm{MBD}}_A(\mathbf{r})$. The analogous equation holds for the non-interacting system. Substituting the expressions (5.14) and (5.15) for 1-RDM and performing elementary math, we obtain non-interacting and interacting kinetic energy densities:

$$\begin{aligned} t^0_A(\mathbf{r}) &= \frac{1}{\pi^{3/2}\lambda^3_A}\frac{\hbar\omega_A}{2}\left[3-\lambda_A^{-2}(\mathbf{r}-\mathbf{R}_A)^2\right]\exp\left[-\lambda_A^{-2}(\mathbf{r}-\mathbf{R}_A)^2\right], \\ t^{\mathrm{MBD}}_A(\mathbf{r}) &= \frac{1}{\pi^{3/2}\Lambda^3_A}\frac{\hbar\omega_A}{2}\left[3\left(\frac{\lambda_A}{\lambda^{\mathrm{eff}}_A}\right)^2-\lambda^2_A(\mathbf{r}-\mathbf{R}_A)^T\mathbf{K}^2_{AA}(\mathbf{r}-\mathbf{R}_A)\right]e^{-(\mathbf{r}-\mathbf{R}_A)^T\mathbf{K}_{AA}(\mathbf{r}-\mathbf{R}_A)}\,, \end{aligned} \tag{5.25}$$

where we introduced the effective length $\lambda^{\mathrm{eff}}_{\mathrm{A}} = \left[\frac{1}{3}\mathrm{Tr}\mathbf{A}_{AA}\right]^{-1/2}$ to retain the structural similarity between the two expressions.

Potential energy

The matrix of potential energy operator $\hat{\Pi}_A = m_A\omega^2_A\mathbf{r}^2_A/2$ is diagonal in coordinate representation:

$$\langle\underline{\mathbf{r}}|\hat{\Pi}_A|\underline{\mathbf{r}}'\rangle = \frac{m_A\omega^2_A}{2}(\mathbf{r}_A-\mathbf{R}_A)^2\langle\underline{\mathbf{r}}|\underline{\mathbf{r}}'\rangle = \frac{m_A\omega^2_A}{2}(\mathbf{r}_A-\mathbf{R}_A)^2\prod_{i=1}^N\delta(\mathbf{r}_i-\mathbf{r}'_i)\,. \tag{5.26}$$

Therefore, similar to Eq. (5.24), the ground-state potential energy can be expressed as

$$\langle\tilde{\mathbf{0}}|\hat{\Pi}_A|\tilde{\mathbf{0}}\rangle = \frac{m_A\omega^2_A}{2}\int d\mathbf{r}\,(\mathbf{r}-\mathbf{R}_A)^2 n^{(1)}_{\mathrm{MBD},A}(\mathbf{r}) = \int d\mathbf{r}\,\pi^{\mathrm{MBD}}_A(\mathbf{r})\,, \tag{5.27}$$

where the potential energy density $\pi_A^{\mathrm{MBD}}(\mathbf{r})$ was introduced. Obviously, the same holds for the non-interacting system, and substituting densities from Eq. (5.16) yields

$$\begin{aligned}\pi_A^0(\mathbf{r}) &= \frac{1}{\pi^{3/2}\lambda_A^3}\frac{\hbar\omega_A}{2}\lambda_A^{-2}(\mathbf{r}-\mathbf{R}_A)^2\exp\left[-\lambda_A^{-2}(\mathbf{r}-\mathbf{R}_A)^2\right],\\ \pi_A^{\mathrm{MBD}}(\mathbf{r}) &= \frac{1}{\pi^{3/2}\Lambda_A^3}\frac{\hbar\omega_A}{2}\lambda_A^{-2}(\mathbf{r}-\mathbf{R}_A)^2\exp\left[-(\mathbf{r}-\mathbf{R}_A)^T\mathbf{K}_{AA}(\mathbf{r}-\mathbf{R}_A)\right].\end{aligned} \tag{5.28}$$

5.1.2 Two-Body Term in MBD Hamiltonian

Since the dipole interaction operator $\hat{V}_{AB} = q_A q_B\,(\mathbf{r}_A-\mathbf{R}_A)^T\mathbf{T}_{AB}(\mathbf{r}_B-\mathbf{R}_B)$ is also diagonal in coordinate representation, it is straightforward to prove that

$$\langle\tilde{\mathbf{0}}|\hat{V}_{AB}|\tilde{\mathbf{0}}\rangle = q_A q_B\iint d\mathbf{r}d\mathbf{r}'\,(\mathbf{r}-\mathbf{R}_A)^T\mathbf{T}_{AB}(\mathbf{r}'-\mathbf{R}_B)\,n_{AB}^{(2)}(\mathbf{r},\mathbf{r}') = \int d\mathbf{r}\,v_{AB}(\mathbf{r})\,. \tag{5.29}$$

Unlike with the one-body terms, here interaction energy density $v_{AB}(\mathbf{r})$ cannot be identified immediately. The additional integration over $\mathbf{r}'$ has to be carried out using the explicit form of $n_{AB}^{(2)}(\mathbf{r},\mathbf{r}')$ in Eq. (5.21) (we omit the nuclear coordinates and transposition marks for clarity):

$$\begin{aligned}v_{AB}(\mathbf{r}) &= \frac{q_A q_B}{\pi^3\Lambda_{AB}^6}\int d\mathbf{r}'\,\mathbf{r}\mathbf{T}_{AB}\mathbf{r}'\exp\left[-\mathbf{r}\mathbf{J}_{AB}^{11}\mathbf{r}-\mathbf{r}'\mathbf{J}_{AB}^{22}\mathbf{r}'-\mathbf{r}\mathbf{J}_{AB}^{12}\mathbf{r}'-\mathbf{r}'\mathbf{J}_{AB}^{21}\mathbf{r}\right] =\\ &= Q_{AB}\times\exp\left[-\mathbf{r}\mathbf{J}_{AB}^{11}\mathbf{r}\right]\times\mathbf{r}\mathbf{T}_{AB}\int d\mathbf{r}'\mathbf{r}'\exp\left[-\mathbf{r}'\mathbf{J}_{AB}^{22}\mathbf{r}'-2\mathbf{r}\mathbf{J}_{AB}^{12}\mathbf{r}'\right],\end{aligned} \tag{5.30}$$

where we introduced $Q_{AB} = q_A q_B/\pi^3\Lambda_{AB}^6$, moved $\mathbf{r}'$-independent factors outside the integral, and used the symmetry of the $\mathbf{J}_{AB}$ matrix. By completing the exponent under the integral to the complete square, this Gaussian integral can be evaluated as

$$v_{AB}(\mathbf{r}) = \frac{Q_{AB}}{4}\sqrt{\frac{(2\pi)^3}{\det\mathbf{J}_{AB}^{22}}}\exp\left[-\mathbf{r}\mathbf{J}_{AB}^{11}\mathbf{r}\right]\times\mathbf{r}\mathbf{T}_{AB}\left[\mathbf{J}_{AB}^{22}\right]^{-1}\mathbf{M}^T\times\exp\left[\frac{1}{2}\mathbf{M}\left[\mathbf{J}_{AB}^{22}\right]^{-1}\mathbf{M}^T\right], \tag{5.31}$$

with $\mathbf{M} = -\mathbf{r}\mathbf{J}_{AB}^{12}/\sqrt{2}$ and $\mathbf{M}^T = -\mathbf{J}_{AB}^{21}\mathbf{r}/\sqrt{2}$, since $[\mathbf{J}_{AB}^{12}]^T = \mathbf{J}_{AB}^{21}$. Substituting this and recovering the dependence on nuclear coordinates and transposition marks where required, we finally obtain

$$v_{AB}(\mathbf{r}) = -\frac{q_A q_B}{\pi^{3/2}\tilde{\Lambda}_{AB}^3}(\mathbf{r}-\mathbf{R}_A)^T\mathbf{T}_{AB}^{\mathrm{eff}}(\mathbf{r}-\mathbf{R}_A)\exp\left[-(\mathbf{r}-\mathbf{R}_A)^T\Sigma_{AB}(\mathbf{r}-\mathbf{R}_A)\right], \tag{5.32}$$

where we introduced the effective length (cubed) $\tilde{\Lambda}_{AB}^3 = \sqrt{\det\mathbf{C}^{\prime AB}\cdot\det\mathbf{J}_{AB}^{22}/\det\mathbf{C}}$ and matrices

$$\mathbf{T}_{AB}^{\mathrm{eff}} = \mathbf{T}_{AB}\,[\mathbf{J}_{AB}^{22}]^{-1}\mathbf{J}_{AB}^{21}\,;\quad \Sigma_{AB} = \mathbf{J}_{AB}^{11}-\mathbf{J}_{AB}^{12}[\mathbf{J}_{AB}^{22}]^{-1}\mathbf{J}_{AB}^{21}\,. \tag{5.33}$$

5.2 MBD Energy Density Visualized in 3D Space

Using the results from the above, the MBD energy density can be split into one- and two-body parts,

$$e_{\mathrm{MBD}}(\mathbf{r}) = e_{\mathrm{MBD}}^{(1)}(\mathbf{r}) + e_{\mathrm{MBD}}^{(2)}(\mathbf{r})\,, \tag{5.34}$$

$$e_{\mathrm{MBD}}^{(1)}(\mathbf{r}) = \sum_{A=1}^{N}\left[t_A^{\mathrm{MBD}}(\mathbf{r}) - t_A^0(\mathbf{r})\right] + \sum_{A=1}^{N}\left[\pi_A^{\mathrm{MBD}}(\mathbf{r}) - \pi_A^0(\mathbf{r})\right], \tag{5.35}$$

$$e_{\mathrm{MBD}}^{(2)}(\mathbf{r}) = \frac{1}{2}\sum_{B\neq A}v_{AB}(\mathbf{r})\,, \tag{5.36}$$

as well as the MBD energy itself,

$$E_{\mathrm{MBD}} = E^{(1)}_{\mathrm{MBD}} + E^{(2)}_{\mathrm{MBD}} = \int d\mathbf{r}\, e^{(1)}_{\mathrm{MBD}}(\mathbf{r}) + \int d\mathbf{r}\, e^{(2)}_{\mathrm{MBD}}(\mathbf{r})\,. \tag{5.37}$$

We note that partitioning of the total MBD energy onto one- and two-body parts can also be done using the second-quantization approach (SQ-MBD) [172]. The advantage of our approach is that the derived $e_{\mathrm{MBD}}(\mathbf{r})$ provides information about the spatial distribution of dispersion energy in 3D space, representing it as an energy density scalar field. This offers a fresh perspective on the dispersion energy, which is typically considered as a single number.

Computational Details

The equations derived in Section 5.1 were implemented in a self-written PYTHON script using the output of the LIBMBD package [160]. The MBD@FCO variant developed in Chapter 4 was employed in the calculations. DFT-PBE computations with 'tight' settings in FHI-AIMS [233] were performed to parametrize the MBD Hamiltonian.

MBD energy densities were evaluated on a cube grid in real space. The grid resolution of 0.1 Å was found to be sufficiently fine, ensuring numerical convergence of the total MBD energy integral $E_{\mathrm{MBD}} = \int d\mathbf{r}\, e_{\mathrm{MBD}}(\mathbf{r})$ below 10^{-5} Ha.

5.2.1 Results for buckyball catcher and DNA-ellipticine

We illustrate this by considering the S12L and L7+ complexes from Chapter 4. As before, we focus on the dispersion interaction energy between two monomers, $\Delta E_{\mathrm{MBD}} = E^{\mathrm{dimer}}_{\mathrm{MBD}} - E^{\mathrm{mono1}}_{\mathrm{MBD}} - E^{\mathrm{mono2}}_{\mathrm{MBD}}$, resulting in the respective change in the MBD energy density,

$$\Delta e_{\mathrm{MBD}}(\mathbf{r}) = e^{\mathrm{dimer}}_{\mathrm{MBD}}(\mathbf{r}) - e^{\mathrm{mono1}}_{\mathrm{MBD}}(\mathbf{r}) - e^{\mathrm{mono2}}_{\mathrm{MBD}}(\mathbf{r})\,, \tag{5.38}$$

In Figure 5.1c, the MBD interaction energy density$\Delta e_{\mathrm{MBD}}(\mathbf{r})$ is shown for the '4a' C_{60} buckyball catcher: $\Delta e_{\mathrm{MBD}}(\mathbf{r})$ is negative on the monomers while being positive in between. This can be rationalized in terms of the changes in the spectrum of the MBD modes from isolated fragments to a dimer. As was shown in Ref. [25], the MBD modes spectrum of a dimer is broadened asymmetrically, with a slight shift towards lower energies, leading to the overall binding, as indicated by blue regions in Figure 5.1. However, also higher-energy modes appear in a dimer (see Figure 4d in Ref. [25]), which lead to regions with the positive $\Delta e_{\mathrm{MBD}}(\mathbf{r})$ (red in Figure 5.1).

Following Eqs. (5.34)–(5.36) and (5.38), the one- and two-body parts of $\Delta e_{\mathrm{MBD}}(\mathbf{r})$ can be defined accordingly. In Figure 5.1a-b, this decomposition of $\Delta e_{\mathrm{MBD}}(\mathbf{r})$ is displayed. The two-body dipole interaction term is negative everywhere in space, as expected, while the one-body term is mostly positive, contributing to binding only close to nuclei. The plots of $\Delta e^{(1)}_{\mathrm{MBD}}(\mathbf{r})$ and $\Delta e^{(2)}_{\mathrm{MBD}}(\mathbf{r})$ help to rationalize the total picture in Figure 5.1c, with the two-body term highlighting the parts of the complex contributing the most to its binding.

These qualitative observations from the isosurface plots are confirmed by a quantitative analysis of the binding energy density distributions across their full span (Figure 5.1, lower row). Namely, we analyze the population $N(e)de$ of a given energy density interval de, which is analogous to the density of states used in electronic structure theory and solid-state physics. The majority of the points in a `.cube` file for a given system belong to a vacuum around the system and therefore are associated with negligibly small values of Δe_{MBD}. To filter out this noise, we

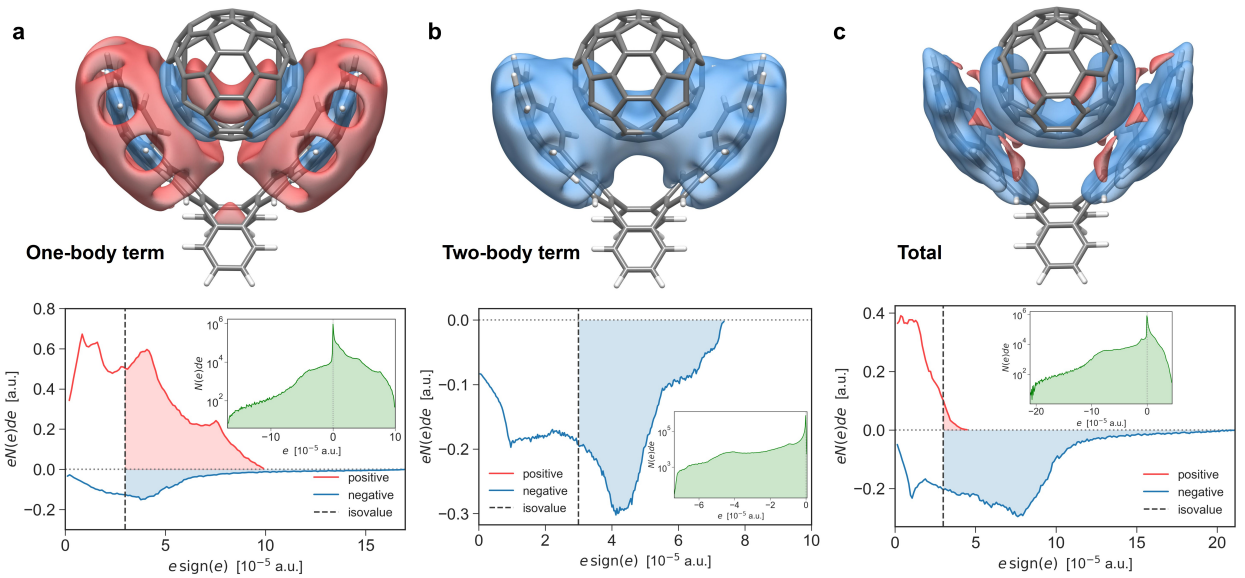

Figure 5.1: One- and two-body MBD energy densities and their sum for the buckyball catcher. Upper row: Isosurfaces of **(a)** $\Delta e^{(1)}_{\mathrm{MBD}}(\mathbf{r})$, **(b)** $\Delta e^{(2)}_{\mathrm{MBD}}(\mathbf{r})$, and **(c)** $\Delta e_{\mathrm{MBD}}(\mathbf{r}) = \Delta e^{(1)}_{\mathrm{MBD}}(\mathbf{r}) + \Delta e^{(2)}_{\mathrm{MBD}}(\mathbf{r})$. The shown isosurface values are $\pm 3 \times 10^{-5}$ a.u. The red and blue colors denote positive and negative energy density, respectively. **Lower row:** Contribution to the energy density $eN(e)de$ of the interval de with $N(e)de$ data points versus the absolute energy density value $e\,\mathrm{sign}(e)$. The dashed vertical line denotes the isovalue 3×10^{-5} a.u. used in the upper row isosurfaces, while red and blue shaded areas reflect the contribution to the MBD energy of the points encompassed by the respective isosurfaces. The insets show the population $N(e)de$ (green, in log-scale) of an interval de as a function of energy density e.

ignored the data points with $\Delta e_{\mathrm{MBD}} < 10^{-8}$ a.u., which affects the integrated MBD binding energy only at the level of 10^{-5} Ha (the same procedure applies to the one- and two-body parts). The remaining points were grouped in $n = 250$ bins, forming a histogram of $N(e)de$ shown as the insets in Figure 5.1 (lower row). Next, we evaluated the net contribution of a given interval de to the overall MBD energy as $eN(e)de$,[2] displayed as red and blue histograms in the lower row in Figure 5.1. These results clearly show that the one-body term has a positive (destabilizing) net contribution, while the two-body term is responsible for the overall stabilization of the complex at the MBD level. Such plots are also helpful for choosing the relevant isosurface value, e.g. using a rule of thumb that the encompassed data points (shown as shaded areas in Figure 5.1) represent the dominant (or at least substantial) part of the overall MBD energy (which is a sum of all histogram elements).

In Figure 5.2, we extend this analysis to a biologically relevant DNA-ellipticine intercalation complex [283]. Curiously, the energy density of the dipolar interaction exhibits selectivity. At an isosurface value of 4×10^{-5} a.u., the isosurface regions encompass only Cytosine (C) nucleobases (Figure 5.2b). To reveal additional contributions, extending to Guanine (G) nucleobases, one has to lower the isosurface value to roughly 3×10^{-5} a.u. The total MBD energy density is also concentrated more on Cytosines, which can be seen from the relatively larger isosurface volume around them compared to the Guanines (Figure 5.2c). While the chemical nature of this selectivity remains unclear at the time of writing, this observation indicates that the MBD energy density derived in this work might be a useful tool for chemical analysis in (bio)molecular systems. Our conclusions are further supported by the results for the C3GC and 3a TNF "pincer"

[2] This quantity has units of energy density, to get the energy units one has to multiply it by the elementary grid volume $dV = dxdydz$.

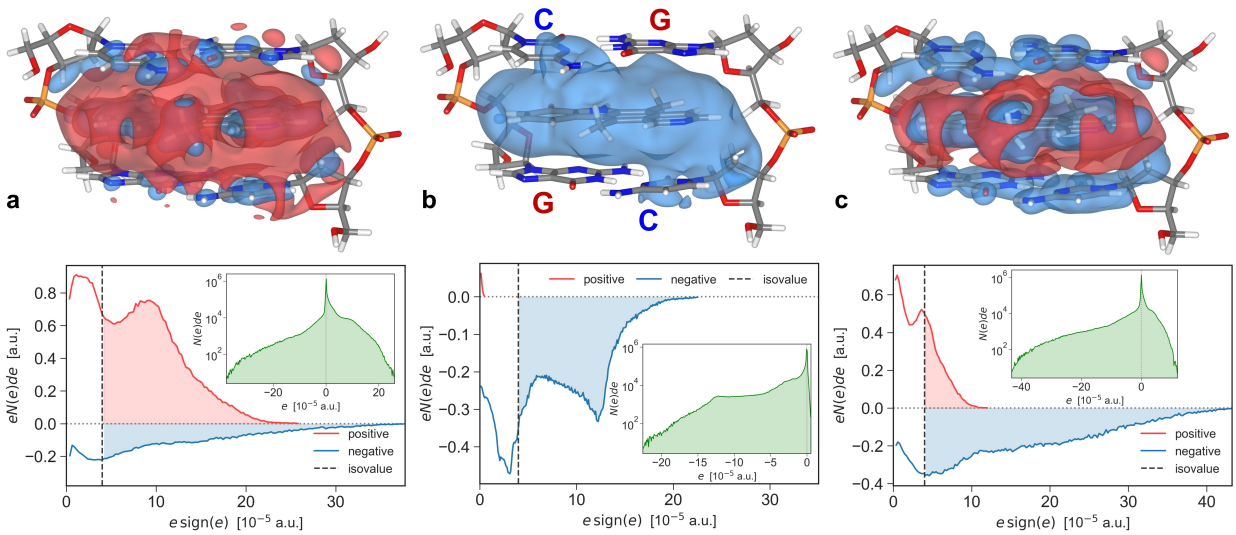

Figure 5.2: One- and two-body MBD energy densities and their sum for the DNA-ellipticine complex. The caption of Figure 5.1 applies, but with the isovalues $\pm 4 \times 10^{-5}$ a.u. In panel **(b)**, the nucleobases cytosine (C) and guanine (G) composing the given DNA fragment are highlighted.

complexes, presented in Appendix A8.

The proposed approach augments the existing density-based techniques, such as QTAIM [257] or NCI [259, 260], with the spatial distribution of the vdW dispersion energy. Moreover, it enables the analysis of e.g., energy density decay with intermonomer distance, and suggests using the dispersion energy density, which is a local intensive property, to train MLFFs capable of treating the vdW dispersion forces beyond the pairwise approximation. These and other applications form the scope for future work on this topic.

5.2.2 Connection Between Charge Density and Energy Density in MBD

By comparing the MBD density isosurfaces of the buckyball catcher in Figure 4.9 and its MBD energy density isosurfaces in Figure 5.1, one can spot the close resemblance in the shape of spatial distributions of $\Delta\rho_{\text{pol}}$ and Δe_{MBD}. For convenience, we explicitly compare the $\Delta\rho_{\text{MBD}}$ and Δe_{MBD} isosurfaces in Figure 5.3. Here, we demonstrate analytically that there is an actual connection between the two MBD observables.

Let us consider a dimer composed of identical QDOs A and B located at $\mathbf{R}_A$ and $\mathbf{R}_B$ as a toy model of e.g. an argon dimer. For the one-body density from Eq. (5.16), we have

$$n^{(1)}_{\text{MBD},A}(\mathbf{r}) = \frac{1}{\pi^{3/2}} \sqrt{\frac{\det \mathbf{C}}{\det \mathbf{C}_{BB}}} \exp\Big\{ -(\mathbf{r}-\mathbf{R}_A)^T \mathbf{K}_{AA} (\mathbf{r}-\mathbf{R}_A) \Big\}\,, \tag{5.39}$$

where $\mathbf{K}_{AA} = \mathbf{C}_{AA} - \mathbf{C}_{AB}[\mathbf{C}_{BB}]^{-1}\mathbf{C}_{BA}$, since $\mathbf{C}_A = \mathbf{C}_{BA}$, $\mathbf{C}_A^T = \mathbf{C}_{AB}$ and $\mathbf{C}^{/A} = \mathbf{C}_{BB}$ for the two-particle case. Analogously, for two-body density Eq. (5.21) gives

$$n^{(2)}_{AB}(\mathbf{r},\mathbf{r}') = \frac{\sqrt{\det \mathbf{C}}}{\pi^3} e^{-(\mathbf{r}-\mathbf{R}_A)^T \mathbf{C}_{AA} (\mathbf{r}-\mathbf{R}_A) - (\mathbf{r}'-\mathbf{R}_B)^T \mathbf{C}_{BB} (\mathbf{r}'-\mathbf{R}_B) - 2(\mathbf{r}-\mathbf{R}_A)^T \mathbf{C}_{AB} (\mathbf{r}'-\mathbf{R}_B)}\,. \tag{5.40}$$

Compared to the Eq. (5.21), the matrices $\mathbf{J}^{ij}_{AB}$ reduce just to the blocks of the correlation matrix, since in the two-particle case the integration over other degrees of freedom, which has led to more complex matrices $\mathbf{J}^{ij}_{AB}$, is not performed.

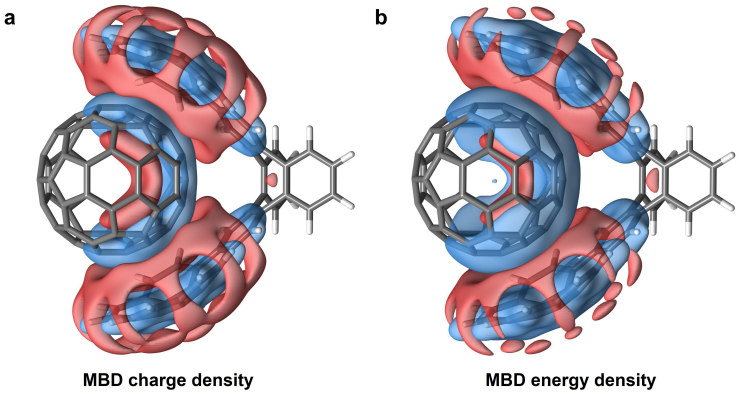

Figure 5.3: Connection between charge and energy densities in MBD. Isosurfaces of **(a)** MBD charge density (3×10^{-5} a.u.) and **(b)** MBD energy density (1.8×10^{-5} a.u.) for the '4a' buckyball catcher from the S12L dataset.

It is also easy to see that for a dimer the matrix $\Sigma_{AB} = \mathbf{J}^{11}_{AB} - \mathbf{J}^{12}_{AB}[\mathbf{J}^{22}_{AB}]^{-1}\mathbf{J}^{21}_{AB}$ reduces to $\Sigma_{AB} = \mathbf{C}_{AA} - \mathbf{C}_{AB}[\mathbf{C}_{BB}]^{-1}\mathbf{C}_{BA} = \mathbf{K}_{AA}$. Therefore, the 2-body energy density takes the form:

$$e^{(2)}_{AB}(\mathbf{r}) = -\frac{q_A q_B}{\pi^{3/2}}\sqrt{\frac{\det\mathbf{C}}{\det\mathbf{C}_{BB}}}(\mathbf{r}-\mathbf{R}_A)^T\mathbf{T}^{\text{eff}}_{AB}(\mathbf{r}-\mathbf{R}_A)\exp\left\{-(\mathbf{r}-\mathbf{R}_A)^T\mathbf{K}_{AA}(\mathbf{r}-\mathbf{R}_A)\right\}, \tag{5.41}$$

where the effective dipole tensor is now also simplified as $\mathbf{T}^{\text{eff}}_{AB} = \mathbf{T}_{AB}[\mathbf{C}_{BB}]^{-1}\mathbf{C}_{AB}$. There's also the second symmetric term with displacements centered around B:

$$e^{(2)}_{BA}(\mathbf{r}) = -\frac{q_A q_B}{\pi^{3/2}}\sqrt{\frac{\det\mathbf{C}}{\det\mathbf{C}_{AA}}}(\mathbf{r}-\mathbf{R}_B)^T\mathbf{T}^{\text{eff}}_{BA}(\mathbf{r}-\mathbf{R}_B)\exp\left\{-(\mathbf{r}-\mathbf{R}_B)^T\mathbf{K}_{BB}(\mathbf{r}-\mathbf{R}_B)\right\}, \tag{5.42}$$

and the total two-body contribution is given by their average $e^{(2)}_{\text{MBD}}(\mathbf{r}) = [e^{(2)}_{AB}(\mathbf{r}) + e^{(2)}_{BA}(\mathbf{r})]/2$. Now it is evident that two-body terms can be written as proportional to the respective one-body particle density, and the proportionality coefficient is the "dressed" dipole tensor:

$$e^{(2)}_{AB}(\mathbf{r}) = -q_A q_B(\mathbf{r}-\mathbf{R}_A)^T\mathbf{T}^{\text{eff}}_{AB}(\mathbf{r}-\mathbf{R}_A)\, n^{(1)}_{\text{MBD},A}(\mathbf{r}), \tag{5.43}$$

leading to the following expression for two-body contribution to the MBD ground-state energy:

$$\begin{aligned} E^{(2)}_{\text{MBD}} = \int e^{(2)}_{\text{MBD}}(\mathbf{r})\, d^3r = &-\frac{q_B}{2}\int(\mathbf{r}-\mathbf{R}_A)^T\mathbf{T}^{\text{eff}}_{AB}(\mathbf{r}-\mathbf{R}_A)\,\rho^{(1)}_{\text{MBD},A}(\mathbf{r})\,d\mathbf{r} - \\ &-\frac{q_A}{2}\int(\mathbf{r}-\mathbf{R}_B)^T\mathbf{T}^{\text{eff}}_{BA}(\mathbf{r}-\mathbf{R}_B)\,\rho^{(1)}_{\text{MBD},B}(\mathbf{r})\,d\mathbf{r}, \end{aligned} \tag{5.44}$$

where we absorbed one of the charges into ρ under an integral sign. As for the one-body term, its proportionality to $n^{(1)}_{\text{MBD},A}(\mathbf{r})$ follows immediately from Eqs. (5.25), (5.28) and (5.16). Moreover, their proportionality holds even in a general case, not only for a dimer. These findings rationalize the observed similarity of $\rho_{\text{MBD}}(\mathbf{r})$ and $e_{\text{MBD}}(\mathbf{r})$ spatial distributions.

5.3 Two-Body MBD Energy As a Nonlocal Correlation Functional

Based on Eq. (5.29), instead of the local two-body MBD energy density $e^{(2)}_{\text{MBD}}(\mathbf{r})$, we can introduce the *nonlocal* counterpart as

$$e^{(2)}_{\text{MBD}}(\mathbf{r},\mathbf{r}') = \frac{1}{2}\sum_{B\neq A} q_A q_B(\mathbf{r}-\mathbf{R}_A)^T\mathbf{T}_{AB}(\mathbf{r}'-\mathbf{R}_B)\, n^{(2)}_{AB}(\mathbf{r},\mathbf{r}'). \tag{5.45}$$

Based on the results of the previous section for a QDO dimer, in this model system we can factorize $n^{(2)}_{AB}(\mathbf{r},\mathbf{r}')$ in Eq. (5.40) using Eq. (5.39) as

$$n^{(2)}_{AB}(\mathbf{r},\mathbf{r}') = n^{(1)}_{A}(\mathbf{r})\, n^{(1)}_{B}(\mathbf{r}')\, \Gamma_{AB}(\mathbf{r},\mathbf{r}') \,, \tag{5.46}$$

where $\Gamma_{AB}(\mathbf{r},\mathbf{r}')$ is the correlation function having the explicit form:

$$\begin{aligned}\Gamma_{AB}(\mathbf{r},\mathbf{r}') = \sqrt{\frac{\det \mathbf{C}_{BB}}{\det \mathbf{K}_{BB}}} \times \exp\Big\{ &-(\mathbf{r}-\mathbf{R}_A)^T (\mathbf{C}_{AA}-\mathbf{K}_{AA})(\mathbf{r}-\mathbf{R}_A) \\ &- (\mathbf{r}'-\mathbf{R}_B)^T (\mathbf{C}_{BB}-\mathbf{K}_{BB})(\mathbf{r}'-\mathbf{R}_B) - 2(\mathbf{r}-\mathbf{R}_A)^T \mathbf{C}_{AB}(\mathbf{r}'-\mathbf{R}_B) \Big\} \,.\end{aligned} \tag{5.47}$$

In this derivation, we used the fact that $\det \mathbf{C} = \det \mathbf{C}_{AA} \cdot \det\left(\mathbf{C}_{BB} - \mathbf{C}_{BA}[\mathbf{C}_{AA}]^{-1}\mathbf{C}_{AB}\right) = \det \mathbf{C}_{AA} \cdot \det \mathbf{K}_{BB}$ as the determinant of a block matrix.

The performed factorization allows us to recast $E^{(2)}_{\mathrm{MBD}} = \iint d\mathbf{r} d\mathbf{r}'\, e^{(2)}_{\mathrm{MBD}}(\mathbf{r},\mathbf{r}')$, the two-body part of the MBD energy, in the form

$$E^{(2)}_{\mathrm{MBD}} = \frac{1}{2}\iint d\mathbf{r} d\mathbf{r}'\, \rho_A(\mathbf{r}) \Phi_{AB}(\mathbf{r},\mathbf{r}') \rho_B(\mathbf{r}') + \frac{1}{2}\iint d\mathbf{r} d\mathbf{r}'\, \rho_B(\mathbf{r}) \Phi_{BA}(\mathbf{r},\mathbf{r}') \rho_A(\mathbf{r}') \,, \tag{5.48}$$

where we used $\rho_i(\mathbf{r}) = q_i n^{(1)}_i(\mathbf{r})$ and introduced

$$\Phi_{AB}(\mathbf{r},\mathbf{r}') = (\mathbf{r}-\mathbf{R}_A)^T \mathbf{T}_{AB}(\mathbf{r}'-\mathbf{R}_B) \times \Gamma_{AB}(\mathbf{r},\mathbf{r}') \,. \tag{5.49}$$

Comparing Eq. (5.48) to the general form of nonlocal density functional (2.147), which we repeat here for clarity:

$$E_{\mathrm{c,nl}} = \frac{1}{2}\iint d\mathbf{r} d\mathbf{r}'\, \rho(\mathbf{r}) \Phi[\rho](\mathbf{r},\mathbf{r}') \rho(\mathbf{r}') \,, \tag{5.50}$$

we identify that $\Phi_{AB}(\mathbf{r},\mathbf{r}')$ defines the *coarse-grained nonlocal kernel*, capturing long-range correlation energy. Thus, the MBD model is equivalent to the nonlocal correlation functional, acting in the space of auxiliary *drudonic densities*. In the following, we compare it to a VV10 nonlocal functional of *electron density* for argon dimer.

5.3.1 Comparison to VV10 for Argon Dimer

Computational Details

The nonlocal part of the VV10 functional (without the βN term) is computed using the numerical integration quadratures [159]:

$$E^{\mathrm{VV10}}_{\mathrm{c}} = \frac{1}{2}\sum_A \sum_{i\in A} w_{Ai}\rho(\mathbf{r}_{Ai}) \sum_B \sum_{j\in B} w_{Bj}\rho(\mathbf{r}'_{Bj}) \Phi_{\mathrm{VV10}}(\mathbf{r}_{Ai},\mathbf{r}'_{Bj}) \,, \tag{5.51}$$

where w_{Ai} and w_{Bj} are the quadrature weights, and the grid points are $\mathbf{r}_{Ai} = \mathbf{R}_A + \mathbf{r}_i$, with $\mathbf{R}_A$ being the nuclear coordinate of A and $\mathbf{r}_i$ defining a one-center integration grid. For the visualization of the VV10 correlation kernel, Eq. (5.51) was implemented in PYTHON using the 'tight' atomic integration grids from the FHI-AIMS(v.221103) code. The density and density gradient norm, required to compute $E^{\mathrm{VV10}}_{\mathrm{c}}$ (see Section 2.3.3), were obtained from the PBE/aug-cc-pVQZ calculations in ORCA(v.5.0.3) [294] using the MULTIWFN(v.3.8) program [295].[3]

[3] We had to evaluate density and density gradient from ORCA to enable the consistent comparison with the VV10 energies from ORCA and MRCC. The VV10 functional is currently not implemented in FHI-AIMS; however, FHI-AIMS features printing of the integration grid points, unlike ORCA. MULTIWFN can evaluate the density and density gradient norm on a given set of points using the output of any quantum chemistry program.

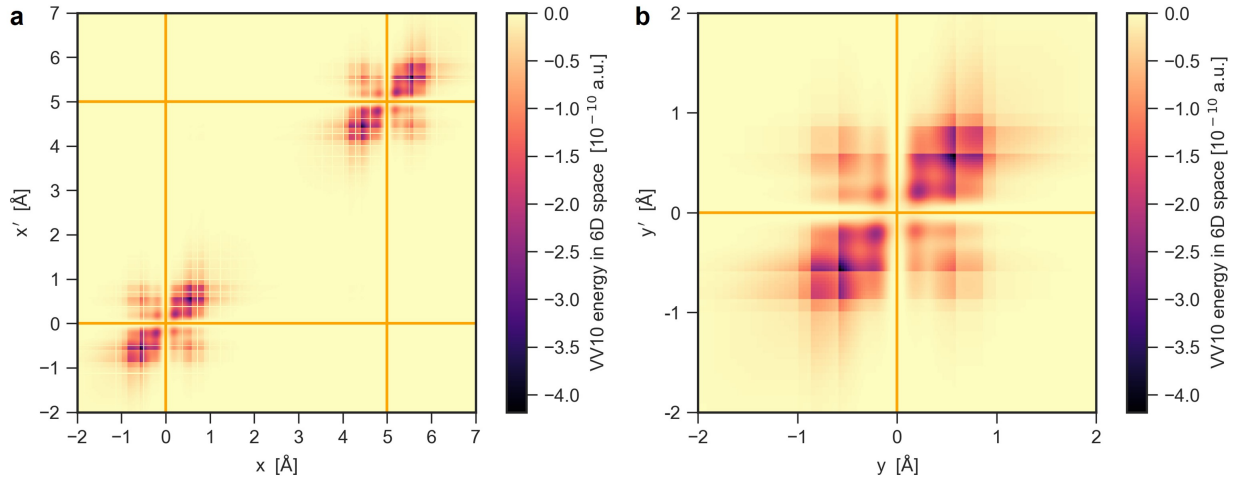

Figure 5.4: VV10 nonlocal correlation energy visualized for an argon dimer at 5 Å separation. Heatmap **(a)** corresponds to 2-dimensional plane (x, x') defined by $y = y' = z = z' = 0$ in 6-dimensional space, while panel **(b)** depicts (y, y') plane defined by $x = x' = z = z' = 0$. The nuclear coordinates are shown by orange lines. See the discussion in the text for more details.

The PYTHON implementation was validated by benchmarking the numerically integrated VV10 correlation energy E_c^{VV10} to the reference implementations in quantum-chemistry packages. The reference (non-self-consistent) nonlocal VV10 correlation energies were computed on top of the PBE functional and aug-cc-pVQZ basis sets in ORCA and MRCC. The VV10 parameters were fixed to their original values $C = 0.0093$ and $b = 5.9$ [159]. The discrepancies between the PYTHON, ORCA and MRCC values of E_c^{VV10} were found to be below 0.5 μHa, compared to the $E_c^{\text{VV10}} \approx -0.02$ Ha for argon dimer (without the positive βN_{elec} term, see Ref. [159]).

The geometry of the argon dimer was chosen so that atom A is at the coordinate origin, $\mathbf{R}_A = (0,0,0)$, while atom B is located at $\mathbf{R}_B = (R,0,0)$. As before, the MBD@FCO variant with the PBE/'tight'-based parametrization was employed for the MBD calculations using the LIBMBD library [160].

VV10 and MBD Nonlocal Energy Densities

In Figure 5.4, we show the nonlocal VV10 energy density $I_{\text{VV10}}(\mathbf{r}_i, \mathbf{r}'_j)$,

$$I_{\text{VV10}}(\mathbf{r}_i, \mathbf{r}'_j) = \frac{1}{2} w(\mathbf{r}_i)\, w(\mathbf{r}'_j)\, \rho(\mathbf{r}_i)\, \rho(\mathbf{r}'_j)\, \Phi_{\text{VV10}}(\mathbf{r}_i, \mathbf{r}'_j)\,, \tag{5.52}$$

for the argon dimer at $R = 5$ Å across two planes in the 6D-space. For clarity, in Eq. (5.52) we omitted atomic indices A, B compared to Eq. (5.51). The heatmaps in Figure 5.4 reveal the fine-grained pattern of the VV10 correlation, with the largest contributions concentrated close to the diagonal $\mathbf{r} = \mathbf{r}'$. The peak magnitudes (about -4.2×10^{-10} a.u.) of the VV10 energy density are observed at coordinates $x_{\max} \approx y_{\max} \approx 0.63$ Å (see Figure 5.4). From the symmetry of the system, $z_{\max} = y_{\max}$, which allows to estimate the distance to the nucleus as $R_{\max} \approx x_{\max}\sqrt{3} \approx 1.1$ Å. This roughly corresponds to the covalent radius of argon atom $R_{\text{cov}} = 1.06 \pm 0.1$ Å [296], while its vdW radius is $R_{\text{vdW}} = 1.88$ Å [102].

However, diagonal blocks highlighted by a color gradient in the heatmaps in Figure 5.4a correspond to a "local", or *intramonomer* part of electronic correlations captured by the VV10 functional, which do not contribute to the the interaction energy of a dimer. On the contrary, the antidiagonal blocks of the correlation heatmap, corresponding to the correlations between two

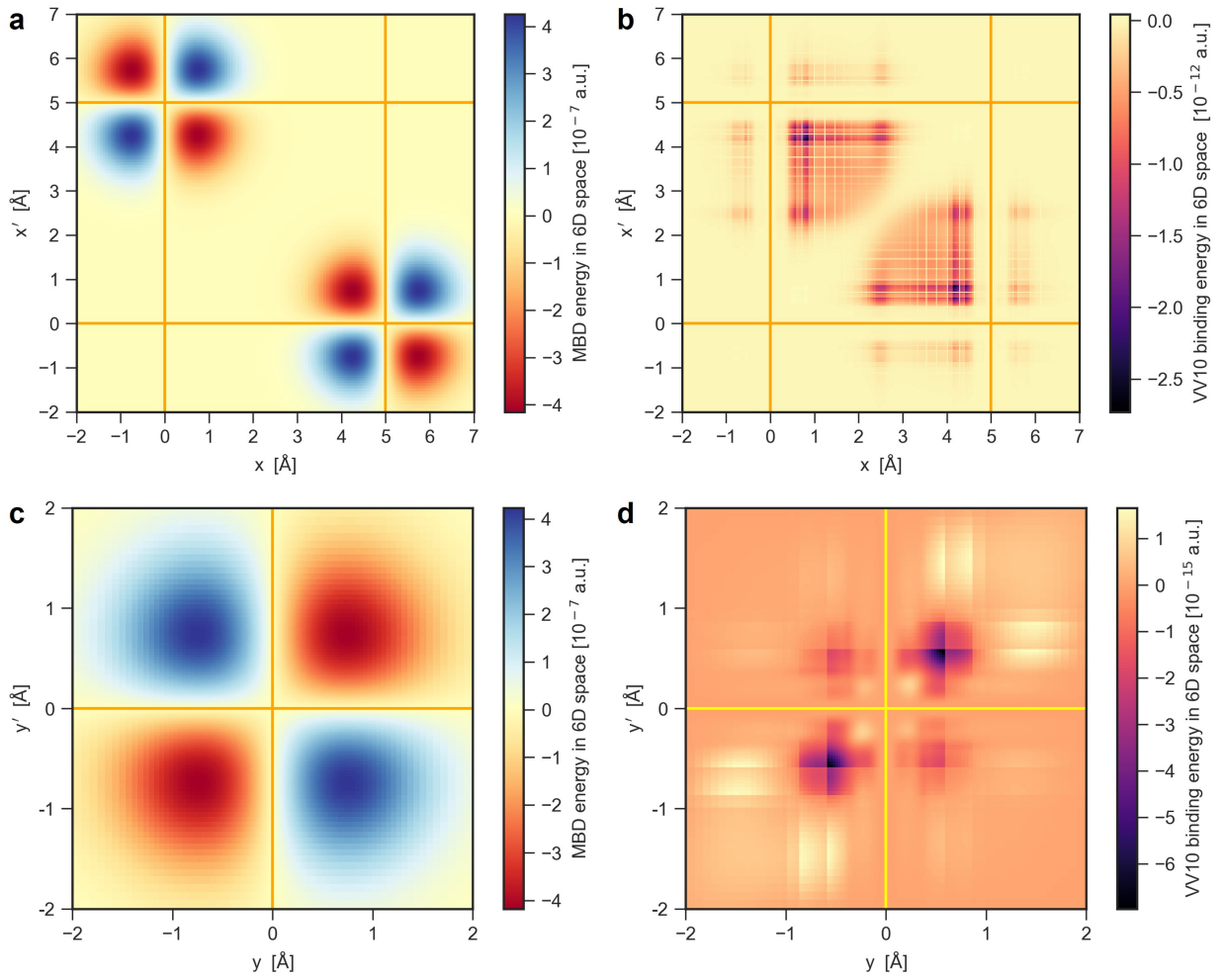

Figure 5.5: The nonlocal part of MBD and VV10 interaction energy. Heatmaps of **(a,c)** MBD and **(b,d)** VV10 *interaction* energy for argon dimer at 5 Å separation. The heatmaps **(a,b)** correspond to 2-dimensional plane (x, x') defined by $y = y' = z = z' = 0$ in 6-dimensional space, while the heatmaps **(c,d)** depict plane (y, y') defined by $x = x' = z = z' = 0$. The nuclear coordinates are shown by orange lines. See the discussion in the text for more details.

monomers, play a decisive role. In Figure 5.4a, those heatmap elements cannot be seen as they are, of course, much smaller than the intramonomer correlations. Therefore, we define the VV10 *interaction* energy density $\Delta I_{\mathrm{VV10}} = I^{\mathrm{D}}_{\mathrm{VV10}} - I^{\mathrm{M1}}_{\mathrm{VV10}} - I^{\mathrm{M2}}_{\mathrm{VV10}}$ (D - dimer, $M1, M2$ - monomers), to focus on the *intermonomer* contributions.

In Figure 5.5, we compare ΔI_{VV10} to the nonlocal MBD energy density ΔI_{MBD} in Eq. (5.48),[4]

$$\Delta I_{\mathrm{MBD}}(\mathbf{r}_i, \mathbf{r}'_j) = \frac{1}{2} w(\mathbf{r}_i) w(\mathbf{r}'_j) \sum_{B \neq A} q_A q_B (\mathbf{r}_i - \mathbf{R}_A)^T \mathbf{T}_{AB} (\mathbf{r}'_j - \mathbf{R}_B)\, n^{(2)}_{AB}(\mathbf{r}_i, \mathbf{r}'_j)\,, \tag{5.53}$$

evaluated at the same grid as ΔI_{VV10}. The defined VV10 and MBD nonlocal energy densities both describe the change in correlation energy due to interaction between monomers, which motivates their comparison. Heatmaps for both methods reveal that contributions to the interaction energy come from antidiagonal blocks, in agreement with the physical intuition. Remarkably, the VV10 functional mostly correlates density tails in the region between two atoms ($0 < x < 5$ Å, $0 < x' < 5$ Å) (Figure 5.5b), while the MBD correlation pattern exhibits quadrupolar

[4] In general, $\Delta I_{\mathrm{MBD}} = I^{\mathrm{D}}_{\mathrm{MBD}} - I^{\mathrm{M1}}_{\mathrm{MBD}} - I^{\mathrm{M2}}_{\mathrm{MBD}}$, but for atomic dimer $I^{\mathrm{M1}}_{\mathrm{MBD}} = I^{\mathrm{M2}}_{\mathrm{MBD}} = 0$ by construction.

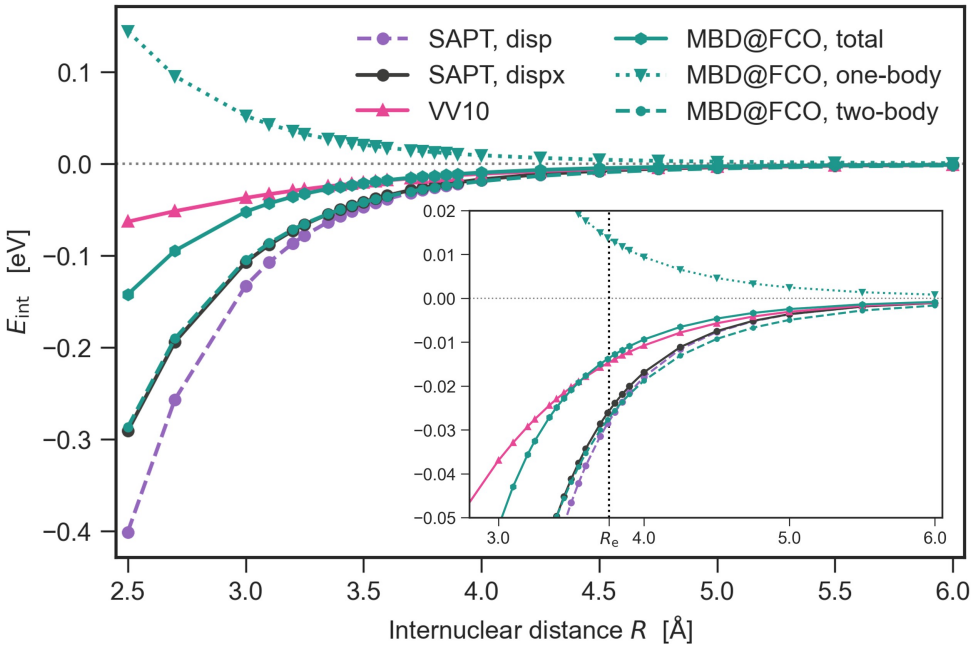

Figure 5.6: Dispersion energy curve of an argon dimer as computed by SAPT(CCSD) [8], VV10 and MBD@FCO methods. For the MBD@FCO method, the decomposition of the total energy onto the one- and two-body terms is displayed. The inset shows a zoom into the near-equilibrium region, with the equilibrium distance $R_e = 3.76$ Å marked by a vertical dotted line.

symmetry (panels (a) and (c)). The local symmetry of the VV10 correlations is more sophisticated, which can be seen in Figure 5.5d. Notably, the magnitude of ΔI_{VV10} within the (x, x') plane is three order of magnitude larger than within the (y, y') plane, although the magnitude of $I^{\mathrm{D}}_{\mathrm{VV10}}$ is the same across both planes (Figure 5.4). This speaks of the weak deformation of the PBE density perpendicular to the noncovalent dispersion "bond".

In turn, ΔI_{MBD} is five orders of magnitude greater than ΔI_{VV10}. Moreover, unlike the VV10 counterpart, ΔI_{MBD} does not manifest anisotropy, reaching the same magnitudes in both (x, x') and (y, y') planes. We associate these contrasts with the different meanings of the two-body nonlocal energy within the MBD and VV10 methods. In VV10, this energy directly corresponds to the total dispersion energy predicted by the method, while in MBD the two-body term has to be summed with the local one-body energy to obtain the dispersion energy. As we discussed above, the one-body term contributes overall positive energy, meaning that two-body energy in MBD has to be lower than in VV10 to produce almost the same dispersion energy at 5 Å (see Figure 5.6). The condition of a good agreement between the MBD and VV10 dispersion energy was the criterion to choose a separation of 5 Å, to make the comparison of the heatmaps as unbiased as possible.

However, if we compare MBD and VV10 dispersion energy curves in a wide range of distances, their behavior is drastically different (Figure 5.6). The VV10 dispersion energy manifests significant damping at shorter distances, which arises due to the choice of the parameter b in Eq. (2.155). Since VV10 is employed on top of the semi-local PBE functional, already capturing some portion of the correlation energy at intermediate distances, b is fine-tuned to avoid double-counting in this range [159], similar to the β parameter in the MBD@rsSCS or TS methods (see Section 2.3). On the other hand, MBD@FCO exposes much faster decay, albeit still slower than the reference SAPT counterparts $E^{(2)}_{\mathrm{disp}}$ ("SAPT, disp" in Figure 5.6) and $E^{(2)}_{\mathrm{dispx}} = E^{(2)}_{\mathrm{disp}} + E^{(2)}_{\mathrm{exch-disp}}$ ("SAPT, dispx" in Figure 5.6), computed at the SAPT(CCSD) level [8].

Surprisingly, if we focus only on the two-body part of the MBD@FCO energy, it agrees perfectly with $E^{(2)}_{\text{dispx}}$ in the whole range of distances, down to the ~70% equilibrium! The zoomed-in inset in Figure 5.6 reveals small discrepancies between $E^{(2)}_{\text{MBD}}$ and $E^{(2)}_{\text{dispx}}$ in the long-range tail of the energy. This remarkable agreement between two-body MBD energy and SAPT is found also in dimers of neon and krypton (see Appendix A9), meaning that it is not coincidental. This intriguing finding requires a thorough investigation, which is the focus of my current research but is beyond the scope of this thesis.

To sum up, this section illustrates the conceptual connection between the MBD model and nonlocal correlation functionals. This is the first step in this direction, which might eventually lead to the synergetic design of better vdW models, incorporating the strengths of both approaches.

5.3.2 Nonlocal MBD Correlations in 3D Space

As was shown above, the two-body energy term in the MBD model has interesting properties, being the analog of nonlocal correlation functionals and showing an excellent agreement with SAPT. The spatial distribution of $e^{(2)}_{\text{MBD}}(\mathbf{r},\mathbf{r}')$ was already illustrated in Figure 5.5 along the 2D slices in the 6D space. In this section, we aim to understand better the spatial distribution of $e^{(2)}_{\text{MBD}}(\mathbf{r},\mathbf{r}')$.

Alternative way to reduce the dimensionality of $e^{(2)}_{\text{MBD}}(\mathbf{r},\mathbf{r}')$ is to choose a fixed point $\mathbf{r}=\mathbf{R}$ in the 3D space and plot $e^{(2)}_{\text{MBD}}(\mathbf{R},\mathbf{r}')$ as the 3D scalar field. The reasonable choice for $\mathbf{R}$ is a coordinate

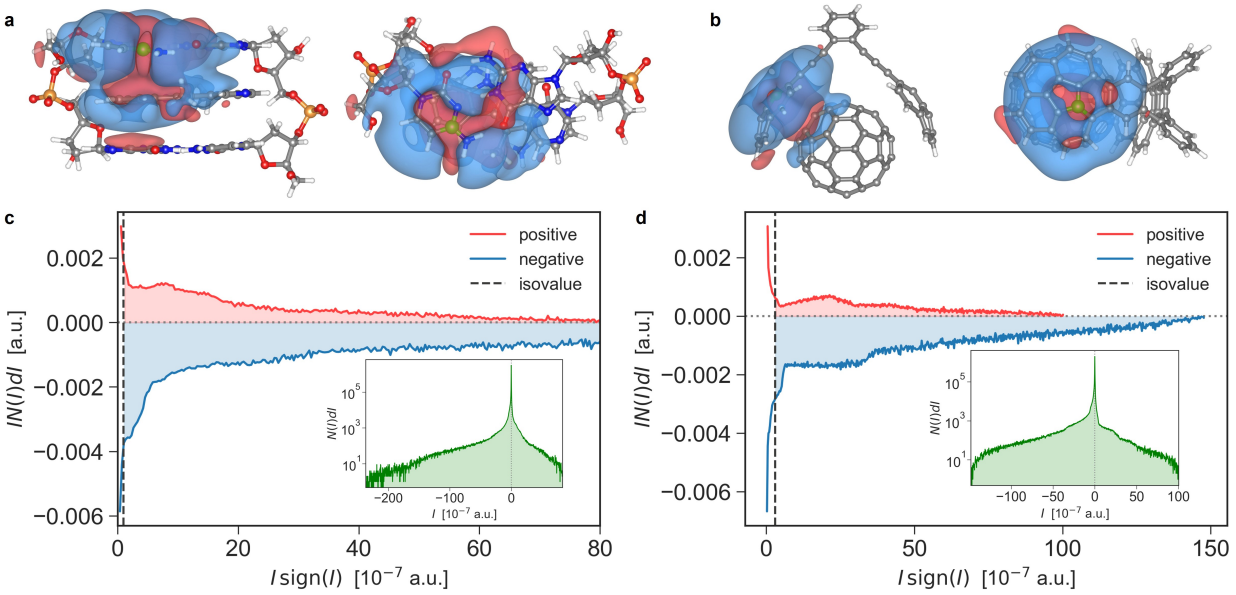

Figure 5.7: Nonlocality of dipolar correlations in MBD. (a,b) The isosurfaces of $I(\mathbf{r}') = e^{(2)}_{\text{MBD}}(\mathbf{R}_A,\mathbf{r}')$ for **(a)** DNA-ellipticine and **(b)** C_{60} buckyball catcher complexes. The isosurface values are **(a)** 10^{-7} a.u. and **(b)** 3×10^{-7} a.u. For each system, two view angles are presented, with the atom at $\mathbf{R}_A$ shown as a big green ball. As before, the red and blue colors denote, respectively, the positive and negative regions. **(c,d)** Contribution to the energy density $IN(I)dI$ of the interval dI with $N(I)dI$ data points versus the absolute energy density value $I\,\text{sign}(I)$ for **(c)** DNA-ellipticine and **(d)** C_{60} buckyball catcher. The dashed vertical line denotes the isovalues used in the corresponding isosurfaces above, while red and blue shaded areas reflect the contribution to the MBD energy of the points encompassed by those isosurfaces. The insets show the population $N(I)dI$ (green, in log-scale) of an interval dI as a function of the energy density I.

$\mathbf{R}_A$ of one of the nuclei in the system. In Figure 5.7, we illustrate the $e^{(2)}_{\mathrm{MBD}}(\mathbf{R}_A, \mathbf{r}')$ for the DNA-ellipticine **(a,c)** and C_{60} catcher complexes **(b,d)**. These plots are similar to Figures 5.1 and 5.2, but to emphasize that the quantity analyzed in Figure 5.7 has different meaning and units, we denote it as $I(\mathbf{r}') = e^{(2)}_{\mathrm{MBD}}(\mathbf{R}_A, \mathbf{r}')$.

The first interesting observation is that $e^{(2)}_{\mathrm{MBD}}(\mathbf{R}_A, \mathbf{r}')$ is positive in the vicinity of the reference point $\mathbf{R}_A$, indicating the anti-correlation between local dipoles. This is qualitatively similar to an exchange-hole around an electron in a many-electron system, albeit the exchange-hole has a completely different physical origin (see Section 2.3.2). Farther from $\mathbf{R}_A$, in the second coordination shell, $e^{(2)}_{\mathrm{MBD}}(\mathbf{R}_A, \mathbf{r}')$ becomes negative, indicating the regions, where the local dipoles are correlated the most to the local dipole at point $\mathbf{R}_A$. As can be seen from Figure 5.7a-b, the drudonic correlations leading to both positive and negative contributions to $e^{(2)}_{\mathrm{MBD}}(\mathbf{R}_A, \mathbf{r}')$ can extend as far as few angstroms. The panels **(c,d)** further reveal an important difference between $I(\mathbf{r}')$ and $\Delta e_{\mathrm{MBD}}(\mathbf{r})$ in Figure 5.1. Unlike for $\Delta e_{\mathrm{MBD}}(\mathbf{r})$, the smallest I have the largest contribution magnitude $IN(I)dI$, which motivates the choice of isosurface values in **(a,b)**.

To sum it up, the nonlocal energy density $e^{(2)}_{\mathrm{MBD}}(\mathbf{r}, \mathbf{r}')$ is a complex quantity, containing a plethora of information about the dipolar correlations in the system. The examples presented here are only the first steps towards extracting and understanding this information. The observed changing sign behavior of $e^{(2)}_{\mathrm{MBD}}(\mathbf{R}_A, \mathbf{r}')$ requires a more detailed study and deeper physical interpretation, which is the subject of future work.

5.4 Discussion and Outlook

Here, we introduced the density-matrix representation of the MBD model, which is the central advancement of this chapter. The derived MBD density-matrix formalism provides access to continuous observables in real space, such as one- and two-body particle and energy densities, thereby extending beyond the original coarse-grained representation of collective electronic response in the MBD model.

The utility of the proposed approach is demonstrated through numerous examples of the derived observables and their analysis. The local MBD energy density, $\Delta e_{\mathrm{MBD}}(\mathbf{r})$, reveals the spatial distribution of the dispersion energy, offering valuable insights for chemical analysis, as illustrated by the example of the DNA-ellipticine complex. The two-body energy term, $e^{(2)}_{\mathrm{MBD}}(\mathbf{r}, \mathbf{r}')$, sheds light on the intricate nature of dipolar correlations and offers a promising tool for understanding such interactions in complex molecular systems. The isosurface plots for DNA-ellipticine and C_{60} buckyball catcher complexes illustrate the spatial extent of dipolar correlations over several angstroms, emphasizing the inherent nonlocality of vdW dispersion interactions and the importance of capturing these effects for accurate energy descriptions.

The comparison between MBD@FCO and VV10 nonlocal energy distributions for the argon dimer further highlights this nonlocal perspective. Despite contrasting patterns of nonlocal correlations, both methods yield predictions for long-range dispersion energy that closely match reference SAPT values. Notably, the MBD@FCO two-body energy term also agrees well with SAPT at short range – an intriguing result that warrants further investigation. Refining the local polarizability functional (2.152) in VV10 based on the MBD@FCO polarizability density, which can be evaluated using the developed density-matrix representation, is a promising avenue for future work. These insights are essential for bridging the gap between coarse-grained and fine-grained representations of electronic correlations, providing a more nuanced understanding of vdW dispersion interactions.

Further exploration of the MBD density-matrix theory could also have broader implications for DFA design. A deeper understanding of dipolar quasiparticle correlations in the MBD ground state and their connection to real electron correlations might lead to new analytical asymptotic expressions for long-range correlation energy or novel constraints on the correlation functional in DFT.

In conclusion, this work represents a foundational step toward a more comprehensive understanding of nonlocal dipolar correlations in molecular systems. The visualization techniques and insights presented here open new possibilities for both theoretical advancements and practical applications. Future research could focus on refining local polarizability functionals, improving energy density descriptions, and uncovering novel connections between dipolar quasiparticles and electron correlations.

Chapter 6

Summary and Outlook

6.1 Digest

This section offers a concise overview of the key findings and insights obtained throughout this doctoral journey, while detailed discussions and conclusions can be found in the respective chapters.

This thesis advances the theoretical framework for describing van der Waals (vdW) interactions in complex chemical and biological systems by developing a universal vdW potential, elucidating the impact of vdW interactions on electronic charge density, and introducing a novel reduced density-matrix formalism for the many-body dispersion method. Our findings bridge fundamental physical principles with practical computational tools and provide a path forward for more accurate and efficient modeling of noncovalent interactions.

The vdW-QDO potential represents a significant step towards a universal and non-empirical vdW model. By leveraging vdW scaling laws, the quantum Drude oscillator (QDO) formalism, and the principle of corresponding states, we achieved a model parameterized by only two atomic descriptors. The vdW-QDO potential maintains the simplicity of the Lennard-Jones model while delivering accuracy comparable to the Tang-Toennies potential. Crucially, its robustness across the periodic table and its improved long-range behavior position it as a versatile tool for modeling extended systems where cumulative long-range interactions play a decisive role.

Our investigation into vdW-induced polarization of electron density highlights a deeper conceptual connection between vdW dispersion interactions and electronic structure. While charge density shifts induced by vdW interactions are often negligible in small molecules, they become increasingly significant in larger, more polarizable systems. These density redistributions influence chemical interaction descriptors, such as NCI analysis, and can have substantial implications for understanding protein folding and higher-order interactions in biological systems. The MBD@FCO method emerges as a consistent and parameter-free approach for capturing vdW-induced density changes, marking a notable advancement over empirical damping schemes commonly used in density functional approximations (DFAs).

Finally, a novel theoretical framework for the reduced density-matrix description of many-body dispersion interactions is presented. The proposed approach offers tools for detailed analysis of dipolar correlations and their visualization, which helps to enhance the understanding of dispersion interaction mechanisms in complex systems.

6.2 Perspective

This section highlights broader potential directions, challenges, and implications for further research, beyond the future perspectives outlined in the previous chapters.

Towards a Unified Framework for Nonbonded Force Fields

The vdW-QDO potential offers a compelling approach for modeling vdW interactions in both classical and machine-learning force fields. Its accurate representation of dispersion interactions has already been validated in applications such as the SO3LR machine-learning force field [33].

However, most current force fields treat nonbonded interactions through separate empirical components for vdW, electrostatics, and induction effects, often lacking a coherent theoretical foundation. In contrast, the QDO model provides a unified framework for treating these interactions (Section 2.2.3). While short-range electrostatic and induction effects were deliberately excluded in developing the vdW-QDO potential at this stage, their future inclusion remains a promising avenue for enhancing accuracy in polar systems. Importantly, the consistent treatment and parametrization of all nonbonded energy terms can potentially improve the overall accuracy of a force field [297].

Further improvements to the underlying physical models for interactions are also needed. Current models often overestimate repulsion energies, leading to inaccuracies in force-field predictions [297], as also observed with the vdW-QDO potential (Section 3.4). Developing many-atom exchange-repulsion potentials, possibly by extending the symmetrization of wavefunctions for dipole-interacting oscillators, offers a promising solution. Additionally, coupling force fields with machine-learning models presents an opportunity to overcome the rigidity of traditional parametric forms [235].

Advancing vdW Modeling in Density Functional Theory

Incorporating vdW physics within DFT remains challenging due to the inherent limitations of standard density functional approximations (DFAs). Existing methods, such as DFT-D3/4, XDM, and MBD, rely on empirically tuned damping functions that lack a rigorous physical basis. These functions must often compensate for deficiencies in the underlying DFA, leading to inconsistencies in the treatment of vdW interactions [271–273]. This issue becomes even more complex in metallic systems, where most DFAs inherently capture substantial long-range correlation effects as being designed to be exact for the homogeneous electron gas. These factors complicate the development of a well-defined and universally applicable combination of vdW models with DFAs, and no comprehensive solution has been proposed so far.

The MBD@FCO method developed in this work offers a promising path forward, providing dispersion energies closely aligned with SAPT benchmarks. Future research could focus on coupling MBD@FCO with DFAs e.g. in a density-driven manner, where range-separation parameters are determined based on reproducing reference electron density rather than empirical energy fits. Alternatively, the "dispersionless" DFAs, such as dlDF [292] or B86bPBE [291], could serve as foundations for the MBD@FCO correction. In general, the challenge of seamlessly combining (semi-)local DFT with long-range vdW models can hardly be tackled in isolation from the construction of better-balanced DFAs. This might be achieved, for instance, by imposing additional long-range constraints on exchange-correlation functionals. The remarkable agreement between SAPT dispersion and two-body MBD@FCO energy for the argon dimer (Fig-

ure 5.6) suggests that MBD@FCO is a promising vdW model for future DFA+vdW developments, warranting further investigation.

Nonlocal vdW density functionals (vdW-DF), on the other hand, offer an appealing alternative to explicitly range-separated DFA+vdW approaches, as they incorporate long-range correlation effects directly within the exchange-correlation functional. Although they offer a more physically motivated treatment than pairwise-additive methods, their computational cost remains high. Additionally, vdW-DFs also resort to range separation of correlation effects, albeit in an implicit form, and therefore suffer from similar issues with pure dispersion energy (see Figure 5.6). Overall, the vdW-DF approach is still in an early stage of development and needs further methodological refinement, particularly to account for many-body effects and to balance semi-local and nonlocal correlation in the mid-range [12]. A promising direction is combining vdW-DF and MBD approaches, as envisaged in Section 5.3.1, forming a fruitful area for future research.

In a broader context, metallic-like systems, where long-range plasmon-like density oscillations dominate, continue to pose a serious challenge for standard DFA+vdW approaches. Recent developments, such as the MBD+C method [169], have shown promise by incorporating charge hopping within the MBD framework. However, this approach requires further investigation and refinement. Ideally, partitioning the electron density to separately parameterize non-metallic regions while treating charge transfer independently may be a beneficial strategy. The MBD-NL method [21] follows this philosophy but disregards valuable information from the jellium-like regions of the electron density. Exploring how this discarded information can be harnessed could lead to significant improvements in predictive accuracy.

Generating High-Level Reference Data for vdW Benchmarking

The refinement of vdW models is often limited by the availability of high-quality reference data. While reliable *ab initio* C_8 and C_{10} dispersion coefficients are available for closed-shell elements, data for open-shell elements and molecular systems remain sparse. High-quality references for molecular C_8 and C_{10} coefficients are essential for benchmarking long-range vdW models, and in the absence of reliable experimental methods to extract these coefficients, *ab initio* calculations remain the only viable option.

Expanding the scope of high-level SAPT calculations to generate benchmark datasets is another critical direction. SAPT provides a unique benchmark method for the dispersion energy across all distance ranges, offering valuable guidance for vdW model development. However, only the S66×8 and S12L datasets have been covered by Hessellmann *et. al* [234, 270] at the SAPT-DFT level, with no data available for L7 and many other datasets. In addition, SAPT offers unique insights into higher-order interaction effects, such as dispersion-induced density polarization, though this capability is not widely implemented in existing codes, to our knowledge. Incorporating dispersion-induced polarization effects on energy through higher-order terms, such as $E^{(3)}_{\mathrm{disp-ind}}$, could guide the development of cheaper methods like Dipole-Correlated Singles (DCS) in MBD [26], which currently have to rely on empirical damping functions. The challenge is that these effects become pronounced only in sufficiently large systems, making SAPT calculations computationally expensive. Nevertheless, ongoing advancements in hardware and algorithms provide optimism for future progress.

Mining high-level interaction energies for large noncovalent complexes is equally vital for advancing the understanding of noncovalent interactions and benchmarking approximate methods, such as DFA+vdW. The recent “puzzling” benchmark comparing CCSD(T) and DMC meth-

ods for large molecules [162] sparked vigorous discussion within the community, prompting several follow-up benchmark studies [283, 298, 299]. These efforts, along with more specialized datasets – such as the recent QUID benchmark for ligand-pocket interactions [300] – are pushing the state-of-the-art and driving subsequent methodological advancements in the field.

Extensions to the MBD Framework

An interesting direction for future research is the inclusion of quadrupolar (and higher-order) multipolar couplings within the MBD framework. Recent efforts [165–167] lay the foundation for future advancements but do not provide conclusive insights into the relative importance of such couplings. This uncertainty largely stems from empiricism introduced due to the aforementioned lack of data on quadrupolar response properties, such as α_2 and C_8. The excellent agreement between SAPT dispersion and two-body MBD@FCO energy suggests that full dipole coupling between finite-width Gaussian dipoles might effectively capture higher-order multipole contributions. This intriguing result warrants closer examination and rationalization on a physical basis.

The impact of vdW interactions on electron density, as highlighted in this thesis, also deserves further investigation, particularly regarding their influence on observables such as energies and forces. Adapting the DCS framework [26] to the MBD@FCO parameterization offers a potential path forward here. On a technical level, implementing self-consistent MBD@FCO and integrating it into the LIBMBD package is a current research focus. However, obtaining meaningful results with the self-consistent MBD@FCO requires addressing the problem of combining the method with DFA, as noted above.

Closing Remarks

The integration of vdW physics into computational models remains a critical challenge in theoretical chemical physics. By advancing QDO-based force fields, improving DFT-vdW coupling, and generating high-quality benchmark datasets, future research can provide more accurate and efficient descriptions of noncovalent interactions. These efforts will be essential for tackling the complexities of large biomolecular systems and complex materials.

Appendices

Parts of Sec. A1–A4 and A5–A7 have been reproduced from the Supplementary Materials of Refs. [107] and [104], respectively.

A1 QDO-derived Damping Function

Here we present the derivation of the QDO damping function given by Eq. (3.24), following the approach used by Tang and Toennies to derive their renowned damping function [90, 206]. Asymptotic series expansion of dispersion energy is based on the assumption of non-overlapping densities. However, at typical interatomic separations in molecular systems, there is still a non-negligible overlap of electron densities. This overlap leads to short-range Pauli (exchange) repulsion, decreasing the density between interacting atoms and consequently leading to the reduced dispersion attraction between atoms, as can be understood using Feynman's arguments [22]. To account for this weakening of dispersion force, Tang and Toennies considered the correction to dispersion energy caused by the overlap [206]. They calculated this correction invoking the *semiclassical* Drude model [206], whereas here we employ the *quantum* Drude oscillator model for the same purpose.

The Hamiltonian of two non-interacting identical QDOs is

$$\hat{H}_0 = -\frac{\hbar^2}{2m}\nabla^2_{\mathbf{r}_1} - \frac{\hbar^2}{2m}\nabla^2_{\mathbf{r}_2} + \frac{k}{2}\mathbf{r}_1^2 + \frac{k}{2}\mathbf{r}_2^2\,,\; k = m\omega^2\,, \tag{A1}$$

where drudonic coordinates $\mathbf{r}_1 = (x_1, y_1, z_1)$ and $\mathbf{r}_2 = (x_2, y_2, z_2)$ have their origins, respectively, at nuclei A and B separated by distance $\mathbf{R}$. The multipolar couplings between the two non-overlapping QDOs can lead only to dispersion attraction between them. Let us now consider a coupling due to the short-range repulsion V_{rep}, which is of a given form. This repulsion is usually expressed as a function of internuclear distance R and therefore serves as the external potential for drudons arising from the symmetrization of the total wave function, as discussed in Chapter 3. To deduce the connection between interdrudonic and internuclear repulsive potentials, one can expand the former as

$$V_{\text{rep}}(|\mathbf{R}+\delta\mathbf{R}|) = V_{\text{rep}}(R) + \frac{dV_{\text{rep}}}{dR}\Delta R + \frac{1}{2}\frac{d^2V_{\text{rep}}}{dR^2}(\Delta R)^2 + \ldots\,, \tag{A2}$$

where $\delta\mathbf{R} = \mathbf{r}_1 - \mathbf{r}_2$, and $\Delta R = |\mathbf{R}+\delta\mathbf{R}| - R$. Assuming that $\mathbf{R}$ is in the z direction and $R \gg |\delta\mathbf{R}|$ (the two nuclei are separated by a distance much larger than the range of drudons' fluctuations) and keeping terms up to the second order, we get

$$\Delta R \approx \frac{1}{2R}\left[2R(z_1 - z_2) + (y_1 - y_2)^2 + (x_1 - x_2)^2\right]\,. \tag{A3}$$

Substituting $\delta\mathbf{R}$ into Eq. (A2) and again keeping terms up to the second order, we have

$$V_{\text{rep}}(|\mathbf{R}+\delta\mathbf{R}|) \approx V_{\text{rep}}(R) + \frac{dV_{\text{rep}}(R)}{dR}(z_1 - z_2) + \frac{1}{2R}\frac{dV_{\text{rep}}(R)}{dR}\left[(x_1-x_2)^2 + (y_1-y_2)^2\right] + \frac{1}{2}\frac{d^2V_{\text{rep}}(R)}{dR^2}(z_1-z_2)^2 . \tag{A4}$$

Thus, the (approximate) Hamiltonian of two repulsion-coupled QDOs reads

$$\hat{H}_{\text{int}} = -\frac{\hbar^2}{2m}\left(\nabla^2_{\mathbf{r}_1} + \nabla^2_{\mathbf{r}_2}\right) + \frac{k}{2}\left(\mathbf{r}_1^2 + \mathbf{r}_2^2\right) + V_{\text{rep}}(R) + \alpha(R)(z_1 - z_2) + \frac{\alpha(R)}{2R}\left[(x_1-x_2)^2 + (y_1-y_2)^2\right] + \frac{\beta(R)}{2}(z_1-z_2)^2 , \tag{A5}$$

with $\alpha(R) = dV_{\text{rep}}/dR$ and $\beta(R) = d^2V_{\text{rep}}/dR^2$ (for brevity, we omit the dependence of α and β on R in what follows). The above Hamiltonian represents a quadratic form in drudonic coordinates, and the following canonical transformation

$$\zeta_1 = \frac{z_1 - z_2}{\sqrt{2}} + \frac{\alpha\sqrt{2}}{k+2\beta}, \quad \zeta_2 = \frac{z_1+z_2}{\sqrt{2}}, \quad \xi_{1,2} = \frac{x_1 \mp x_2}{\sqrt{2}}, \quad \eta_{1,2} = \frac{y_1 \mp y_2}{\sqrt{2}}, \tag{A6}$$

brings it to the diagonal form

$$\hat{H}_{\text{int}} = -\frac{\hbar^2}{2m}\left(\nabla^2_{\tilde{\mathbf{r}}_1} + \nabla^2_{\tilde{\mathbf{r}}_2}\right) + \left(\frac{k}{2} + \frac{\alpha}{R}\right)\left(\xi_1^2 + \eta_1^2\right) + \left(\frac{k}{2} + \beta\right)\zeta_1^2 + \frac{k}{2}\tilde{\mathbf{r}}_2^2 + V_{\text{rep}}(R) - \frac{\alpha^2}{k+2\beta}, \tag{A7}$$

with $\tilde{\mathbf{r}}_1 = (\xi_1, \eta_1, \zeta_1)$, $\tilde{\mathbf{r}}_2 = (\xi_2, \eta_2, \zeta_2)$. The corresponding eigenfrequencies are

$$\omega_{\xi_1} = \omega_{\eta_1} = \sqrt{\frac{k}{m} + \frac{2\alpha}{mR}} \approx \omega\left[1 + \frac{\alpha}{kR}\right], \quad \omega_{\zeta_1} = \sqrt{\frac{k}{m} + \frac{2\beta}{m}} \approx \omega\left[1 + \frac{\beta}{k}\right], \tag{A8}$$

and $\omega_{\xi_2} = \omega_{\eta_2} = \omega_{\zeta_2} = \omega$. We ensure that the Taylor expansion made in Eq. (A8) delivers a good approximation for a wide range of distances, as illustrated by Figure A1. This reflects the fact that V_{rep} changes slowly compared to the harmonic potential. The interaction energy is

$$E_{\text{int}} = E - E_0 = \frac{\hbar}{2}\left(\omega_{\xi_1} + \omega_{\eta_1} + \omega_{\zeta_1} + \omega_{\xi_2} + \omega_{\eta_2} + \omega_{\zeta_2}\right) + V_{\text{rep}}(R) - \frac{\alpha^2}{k+2\beta} - 6\cdot\frac{\hbar\omega}{2} \approx \frac{\hbar\omega}{2}\left[\frac{2\alpha}{kR} + \frac{\beta}{k}\right] + V_{\text{rep}}(R) - \frac{\alpha^2}{k+2\beta} . \tag{A9}$$

Here, $V_{\text{rep}}(R)$ plays the role of external repulsive potential, whereas the remaining two terms represent the desired correction to dispersion energy caused by the presence of $V_{\text{rep}}(R)$. The second-order term $\alpha^2/(k+2\beta)$ should be omitted considering the first-order Taylor expansion for frequencies. Hence, for the (approximate) correction to dispersion energy, we finally have

$$\Delta V_{\text{disp}} = \frac{\hbar\omega}{2}\left[\frac{2\alpha}{kR} + \frac{\beta}{k}\right] . \tag{A10}$$

With our general expression for the QDO exchange energy $V_{\text{rep}} = Aq^2/R\exp(-m\omega R^2/2\hbar)$ given by Eq. (3.6), one can obtain

$$\begin{aligned} \alpha &= \frac{dV_{\text{rep}}}{dR} = -Aq^2\left[\frac{1}{R^2} + \frac{m\omega}{\hbar}\right]e^{-\frac{m\omega}{2\hbar}R^2}, \\ \beta &= \frac{d^2V_{\text{rep}}}{dR^2} = Aq^2\left[\frac{2}{R^3} + \frac{m\omega}{\hbar}\frac{1}{R} + \left(\frac{m\omega}{\hbar}\right)^2 R\right]e^{-\frac{m\omega}{2\hbar}R^2}, \\ \Delta V_{\text{disp}} &= \frac{Aq^2}{2R}\left[(\gamma R)^2 - 1\right]e^{-\frac{(\gamma R)^2}{2}}, \quad \gamma = \sqrt{\frac{m\omega}{\hbar}}, \end{aligned} \tag{A11}$$

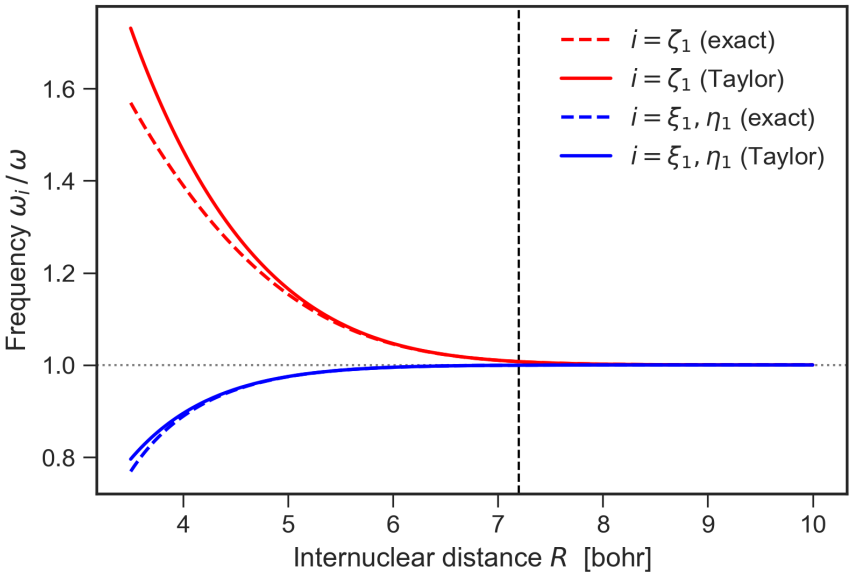

Figure A1: Eigenfrequencies of repulsion-coupled QDOs as a function of distance. ω_{ξ_1,η_1} (blue) and ω_{ζ_1} (red) from Eq. (A8) as described by the exact square root function (dashed) and the first-order Taylor expansion (solid) for Ar_2 dimer. The vertical dashed line denotes the equilibrium distance R_e. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

where the atomic units are used, with $k_e = 1/4\pi\varepsilon_0 = 1$ but $\hbar$ kept explicitly in the equations. The correction ΔV is positive, since $R^2(m\omega/\hbar) \gg 1$ at typical internuclear distances [102].

Now we use the correction ΔV_{disp} to derive our damping function, following the approach of Tang and Toennies [90]. By introducing $B = Aq^2/2$, the sum of dispersion energy and its repulsion correction can be written as

$$-\sum_{n\geq 3}\frac{C_{2n}}{R^{2n}} + \frac{B}{R}\left[(\gamma R)^2 - 1\right] e^{-\frac{(\gamma R)^2}{2}} = -\sum_{n\geq 3} f_{2n}(R)\frac{C_{2n}}{R^{2n}}\,, \tag{A12}$$

where on the left-hand side we have incorporated the desired damping function $f_{2n}(R)$. By introducing coefficients b_{2n} such that $\sum_n b_{2n} = B$, the damping function is extracted as

$$f_{2n}(R) = 1 - \frac{b_{2n}}{C_{2n}}\left(\gamma^2 R^{2n+1} - R^{2n-1}\right) e^{-\frac{(\gamma R)^2}{2}}\,. \tag{A13}$$

The coefficients b_{2n} are not uniquely defined. Therefore, from Eq. (A13) one can only assume that the general form of the damping function is

$$f_{2n}(R) = 1 - P_{2n+1}(\gamma, R) e^{-\frac{(\gamma R)^2}{2}}\,, \tag{A14}$$

where P_{2n+1} is the polynomial of the order $2n+1$. Since the damping function is dimensionless, it can only depend on the dimensionless combination of parameters related to the considered problem, which is uniquely defined by $x = \gamma R$. Thus, we can represent the damping function by

$$f_{2n}(x) = 1 - e^{-\frac{x^2}{2}} \sum_{k=0}^{2n+1} a_k x^k\,, \tag{A15}$$

To determine the polynomial coefficients a_k, we impose the natural boundary conditions

$$f_{2n}(R) \to 1\,,\ R \to \infty\,; \qquad f_{2n}(R) \to 0 + \mathcal{O}\left(R^{2n+1}\right)\,,\ R \to 0\,. \tag{A16}$$

Here, the first condition is fulfilled by default due to the functional form of Eq. (A15). The second condition enforces that all derivatives of $f_{2n}(R)$ up to order $2n+1$ must be zero at $R = 0$, which stems from the requirement that each term $f_{2n}(R)C_{2n}/R^{2n}$ in the damped dispersion series vanishes at $R = 0$ with a finite slope [90]. By the explicit calculation of n-th derivatives, we

found that to satisfy the above boundary conditions, all odd coefficients must be zero and only even powers of x survive, resulting in the uniquely defined damping function

$$f_{2n}(z) = 1 - e^{-\frac{(\gamma R)^2}{2}} \sum_{k=0}^{n} \frac{(\gamma R)^{2k}}{2^k \cdot k!} = 1 - e^{-z} \sum_{k=0}^{n} \frac{z^k}{k!}\,, \quad z = \frac{(\gamma R)^2}{2}\,. \tag{A17}$$

This damping function has exactly the same form as the Tang-Toennies damping function [90]

$$f_{2n}(z) = 1 - e^{-z} \sum_{k=0}^{2n} \frac{z^k}{k!}\,, \text{ where } z = bR\,, \tag{A18}$$

with the only difference in the limits of summation, in addition to the physical meaning of unitless variable ($z = bR$ for the TT damping function, with b stemming from the Born-Mayer repulsion term Ae^{-bR}). The difference is due to the distinct form of Pauli repulsion potential between the vdW-QDO and TT models, which delivers distinct results under applying the aforementioned boundary conditions at $R \to 0$. Interestingly, our quantum-mechanically derived damping function also matches well (but not exactly) with the one heuristically suggested by Slipchenko and Gordon [301]

$$f_{2n}(z) = 1 - e^{-z} \sum_{k=0}^{n} \frac{z^k}{k!}\,, \quad z = (\gamma R)^2\,, \tag{A19}$$

which they have obtained by a simple analogy to the Tang-Toennies damping function upon substitution of exponential overlap integral (of atomic densities) onto the Gaussian overlap integral (of oscillator densities). In Eq. (A19), we corrected the obvious misprint present in Ref. [301] where summation starts with $k = 1$. Another difference with our QDO damping function is that the heuristically obtained function of Slipchenko and Gordon takes $z = (\gamma R)^2$ as an argument instead of $(\gamma R)^2/2$, as obtained within our quantum-mechanical derivation.

A2 vdW-OQDO Parametrization

To set the value of product $m\omega$, we use the condition of balance of exchange repulsion and dispersion forces between two QDOs at the equilibrium in the dipole approximation

$$\frac{k_e q^2}{2}\left[\frac{1}{R_e^2} + \frac{m\omega}{\hbar}\right] e^{-\frac{m\omega}{2\hbar}R_e^2} = \frac{6C_6}{R_e^7}\,. \tag{A20}$$

In Ref. [102], it was shown that for typical R_e in noble-gas dimers $m\omega/\hbar$ is an order of magnitude larger than $1/R_e^2$, and therefore the term $1/R^2$ was neglected. This well works as the first-order approximation and leads to the quantum-mechanical scaling law for vdW radius Eq. (3.8). However, we found that for building the QDO-based vdW potential, this simplification introduces additional error in the location and depth of the potential minimum by making the energies and forces inconsistent. Therefore, here we use full Eq. (A20) to re-derive a consistent optimized QDO parametrization in analogy with Ref. [110].

Substituting $C_6 = \frac{3}{4}k_e^2\hbar\omega\alpha_1\frac{q^2}{m\omega^2}$ into Eq. (A20) and cancelling q^2, we obtain

$$k_e\left[\frac{m\omega}{\hbar} + \frac{1}{R_e^2}\right] e^{-\frac{m\omega}{2\hbar}R_e^2} = \frac{9\hbar}{m\omega}\frac{\alpha_1 k_e^2}{R_e^7}\,. \tag{A21}$$

Solving this for α_1 delivers

$$\alpha_1 = (4\pi\varepsilon_0)\, e^{-\frac{m\omega}{2\hbar}R_e^2}\left[\left(\frac{m\omega}{3\hbar}\right)^2 + \frac{m\omega}{9\hbar R_e^2}\right] R_e^7\,. \tag{A22}$$

From the other hand, for real atoms α_1 can be accurately recovered via R_e as [178]

$$\alpha_1 = \frac{(4\pi\varepsilon_0)}{a_0^4}\frac{\alpha_{\text{fsc}}^{4/3}}{128}R_e^7\,, \tag{A23}$$

where $\alpha_{\text{fsc}} = e^2/4\pi\varepsilon_0\hbar c \approx 1/137.036$ is the fine-structure constant. Combining Eqs. (A22) and (A23), one can obtain

$$e^{-\frac{m\omega}{2\hbar}R_e^2}\left[\left(\frac{m\omega}{3\hbar}\right)^2 + \frac{m\omega}{9\hbar R_e^2}\right] = \frac{1}{a_0^4}\frac{\alpha_{\text{fsc}}^{4/3}}{128}\,. \tag{A24}$$

This equation can be rearranged and rewritten in terms of dimensionless variables

$$a\cdot e^{bx} = 2x^2 + \frac{x}{b}\,, \quad x = \frac{m\omega a_0^2}{\hbar}\,, \tag{A25}$$

where

$$a = \frac{9\alpha_{\text{fsc}}^{4/3}}{64}\,, \quad b = \frac{R_e^2}{2a_0^2} = \frac{2(\alpha_1/4\pi\varepsilon_0)^{2/7}}{\alpha_{\text{fsc}}^{8/21}a_0^{6/7}}\,. \tag{A26}$$

The above transcendental equation is similar to its counterpart from Ref. [110], and it also has two solutions, one of which is very close to zero. As discussed in Ref. [110], this almost zero solution x_B does not have a simple physical interpretation, whereas the second (larger) root x_A of Eq. (A25) is useful to parametrize the QDO model. Thus, the full vdW-OQDO parametrization procedure can be summarized as

$$\omega = \frac{4C_6}{3\hbar\alpha_1^2 k_e^2}\,, \quad m = \frac{\hbar x_A}{\omega a_0^2}\,, \quad q = \sqrt{\alpha_1 m\omega^2}\,. \tag{A27}$$

The numerical values of the parameters for all noble-gas dimers are presented in Table A1.

For the damped vdW-QDO potential, the parametrization procedure has to be adapted accordingly. The force-balance equation in dipole approximation becomes

$$\frac{k_e q^2}{2}\left[\frac{1}{R_e^2} + \frac{m\omega}{\hbar}\right]e^{-\frac{m\omega}{2\hbar}R_e^2} = \frac{6f_6(R_e)C_6}{R_e^7} - \frac{f_6'(R_e)C_6}{R_e^6}\,. \tag{A28}$$

Repeating the manipulations performed for vdW-OQDO parametrization above, Eq. (A28) can be simplified and rewritten in terms of dimensionless variables in the form

$$a\left[e^{bx} - \left(\sum_{k=0}^{3}\frac{(bx)^k}{k!} + \frac{(bx)^4}{3\cdot 3!}\right)\right] = 2x^2 + \frac{x}{b}\,, \quad x = \frac{m\omega a_0^2}{\hbar}\,, \tag{A29}$$

where

$$a = \frac{18\alpha_1 a_0^4}{(4\pi\varepsilon_0)R_e^7}\,, \quad b = \frac{R_e^2}{2a_0^2}\,. \tag{A30}$$

In Eq. (A29), we used the explicit form of the QDO damping function (A17). Eq. (A29) has two solutions like Eq. (A25), and we use its larger root x_A^d to set the parameters of the oscillator according to Eq. (A27). Note that the ω parameter does not change due to the damping function, since it is fully defined by α_1 and C_6.

We note that to parametrize damped vdW-QDO potential for noble gases, Eq. (A23) can be used to make the transcendental equation (A29) dependent only on α_1. This leads to the same a and b coefficients (A26) as in the vdW-OQDO case. The numerical values of the damped vdW-OQDO parameters for all noble-gas dimers are presented in Table A1.

Table A1: Effective atomic polarizabilities α_1^{AB}, dispersion coefficients C_6^{AB} and vdW-OQDO parameters for 21 noble-gas dimers. α_1^{AB} and C_6^{AB} are calculated using Eqs. (3.21) and (3.19). Parameter sets $\{\omega, m, q\}$ and $\{\omega, m_d, q_d\}$ (in a.u.) correspond to the vdW-QDO potential without and with damping function, respectively.

Dimer	α_1^{AB}	C_6^{AB}	ω	m	q	m_d	q_d
He–He	1.38	1.46	1.02219	0.55810	0.89707	0.56760	0.90468
He–Ne	2.025	3.043	0.98941	0.49841	0.99400	0.50884	1.00434
He–Ar	6.24	9.513	0.32574	0.97613	0.80393	1.01360	0.81921
He–Kr	9.09	13.316	0.21487	1.27205	0.73066	1.33196	0.74767
He–Xe	14.34	19.265	0.12491	1.81468	0.63720	1.92445	0.65619
He–Rn	17.46	23.265	0.10175	2.05164	0.60901	2.19004	0.62922
Ne–Ne	2.67	6.38	1.19326	0.37164	1.18865	0.38066	1.20299
Ne–Ar	6.885	19.539	0.54960	0.55625	1.07555	0.57877	1.09710
Ne–Kr	9.735	27.225	0.38304	0.69398	0.99559	0.72791	1.01964
Ne–Xe	14.985	39.144	0.23243	0.95753	0.88044	1.01687	0.90731
Ne–Rn	18.105	47.232	0.19212	1.07013	0.84567	1.14379	0.87429
Ar–Ar	11.1	64.3	0.69583	0.36208	1.39498	0.38110	1.43115
Ar–Kr	13.95	91.100	0.62418	0.36734	1.41296	0.38922	1.45444
Ar–Xe	19.2	133.995	0.48465	0.41383	1.36611	0.44326	1.41386
Ar–Rn	22.32	162.193	0.43409	0.43337	1.35007	0.46689	1.40132
Kr–Kr	16.8	129.6	0.61224	0.34654	1.47727	0.36943	1.52526
Kr–Xe	22.05	191.715	0.52575	0.35968	1.48061	0.38732	1.53644
Kr–Rn	25.17	232.246	0.48879	0.36552	1.48258	0.39578	1.54272
Xe–Xe	27.3	285.9	0.51148	0.33725	1.55198	0.36648	1.61783
Xe–Rn	30.42	346.742	0.49960	0.32938	1.58144	0.35973	1.65270
Rn–Rn	33.54	420.6	0.49852	0.31624	1.62358	0.34706	1.70084

For group II elements, however, Eq. (A23) is not so accurate, since their dimers are not purely vdW bonded. Therefore, both α_1 and R_e should be used explicitly in Eq. (A30). To construct the vdW-QDO potential for these systems, we used only the oscillator parameters for Sr_2. Nevertheless, for completeness, the damped vdW-OQDO parameters for all group II dimers are presented in Table A2.

Dimer	ω	m_d	q_d
Mg_2	0.16445	1.10568	1.46011
Ca_2	0.11458	1.17281	1.55535
Sr_2	0.10639	1.06709	1.54334
Ba_2	0.09198	1.06133	1.56703
Zn_2	0.32010	0.65638	1.61268
Cd_2	0.43226	0.46174	1.99216
Hg_2	0.45481	0.50101	1.87435

Table A2: The damped vdW-OQDO parameters for group II dimers (in a.u.) obtained using the reference $\{\alpha_1, C_6\}$ from Table 3.4

A3 Errors for Two Lennard-Jones Parametrizations of Noble Gases

In this section, we report the Lennard-Jones parameters and errors of the corresponding LJ potentials for noble-gas dimers. We considered two sets of LJ parameters. The first one, denoted as LJ1 throughout the manuscript, is taken from Ref. [302], where $(\sigma_1, \varepsilon_1)$ are compiled from different sources. These parameters were fitted to reproduce the thermodynamic data for noble gases. The second set of parameters denoted as LJ2 was created by ourselves. We used the reference *ab initio* values of R_e and D_e compiled in Ref. [86] to determine $(\sigma_2, \varepsilon_2)$ using the obvious relations $\sigma_2 = R_e \times 2^{-1/6}$, $\varepsilon_2 = D_e$. The numerical values for both parameter sets are presented in Table A3.

Table A3: The two considered sets of Lennard-Jones parameters for noble gases: LJ1 $(\sigma_1, \varepsilon_1)$ [302] and LJ2 $(\sigma_2, \varepsilon_2)$ obtained by ourselves. The units are Bohr for σ and meV for ε.

Element	σ_1	ε_1	σ_2	ε_2
He	4.838	0.8807	4.996	0.9475
Ne	5.197	3.0678	5.194	3.6324
Ar	6.425	10.341	6.334	12.319
Kr	6.803	14.736	6.761	17.310
Xe	7.748	18.958	7.370	24.126

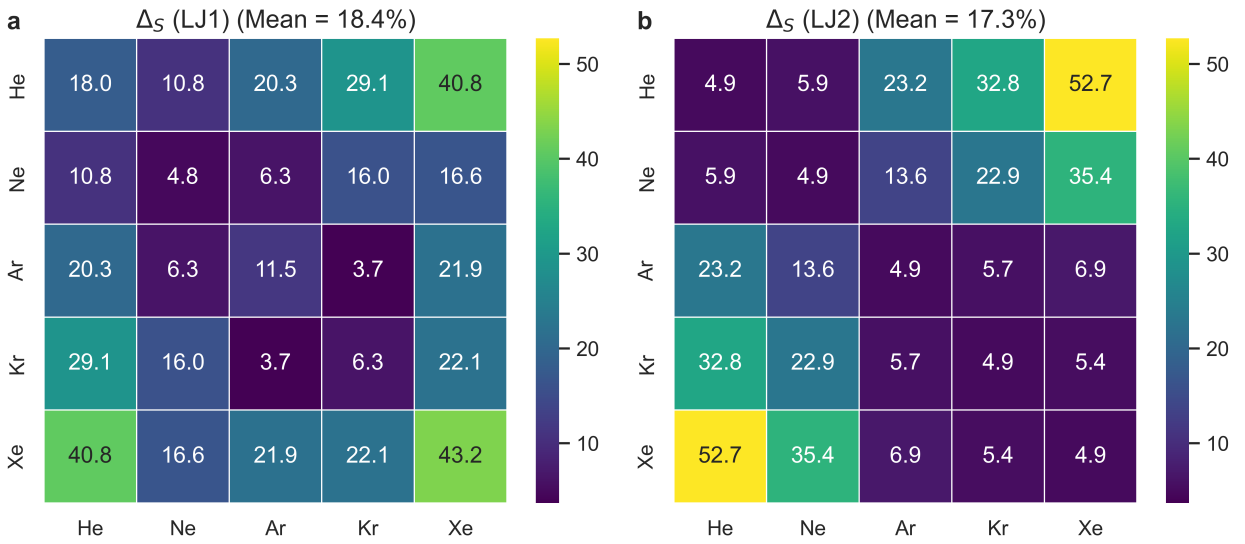

Figure A2: Assessment of the LJ potential accuracy. Δ_S metric (in %) calculated for two LJ parametrizations relative to the TTS potential [86]. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

In Figure A2 we show the error of LJ1 and LJ2 potentials expressed in terms of the Δ_S metrics, as described in Section 3.2. As one can see, the LJ2 parametrization, designed to reproduce correct $\{R_e, D_e\}$, works well for dimers composed of atoms possessing similar vdW radius (He, Ne is one group, and Ar, Kr, Xe is another). However, for dimers mixed between the two groups, errors of LJ2 increase drastically. This indicates the lack of flexibility of the LJ potential as well as the limited accuracy of Lorentz-Berthelot combination rules.

A4 vdW-QDO Potential Curves for Molecular Dimers

Here we present our results for the eight molecular dimers considered in our study. They allow us to conclude that the vdW-QDO potential works well in predicting dispersion energy between aliphatic molecules, but it has difficulties with treating benzene. We attribute this to a high

anisotropy of the benzene polarizability tensor (longitudinal component is twice as large as the transverse one [11]), which means that dispersion interaction requires many-body treatment in this case.

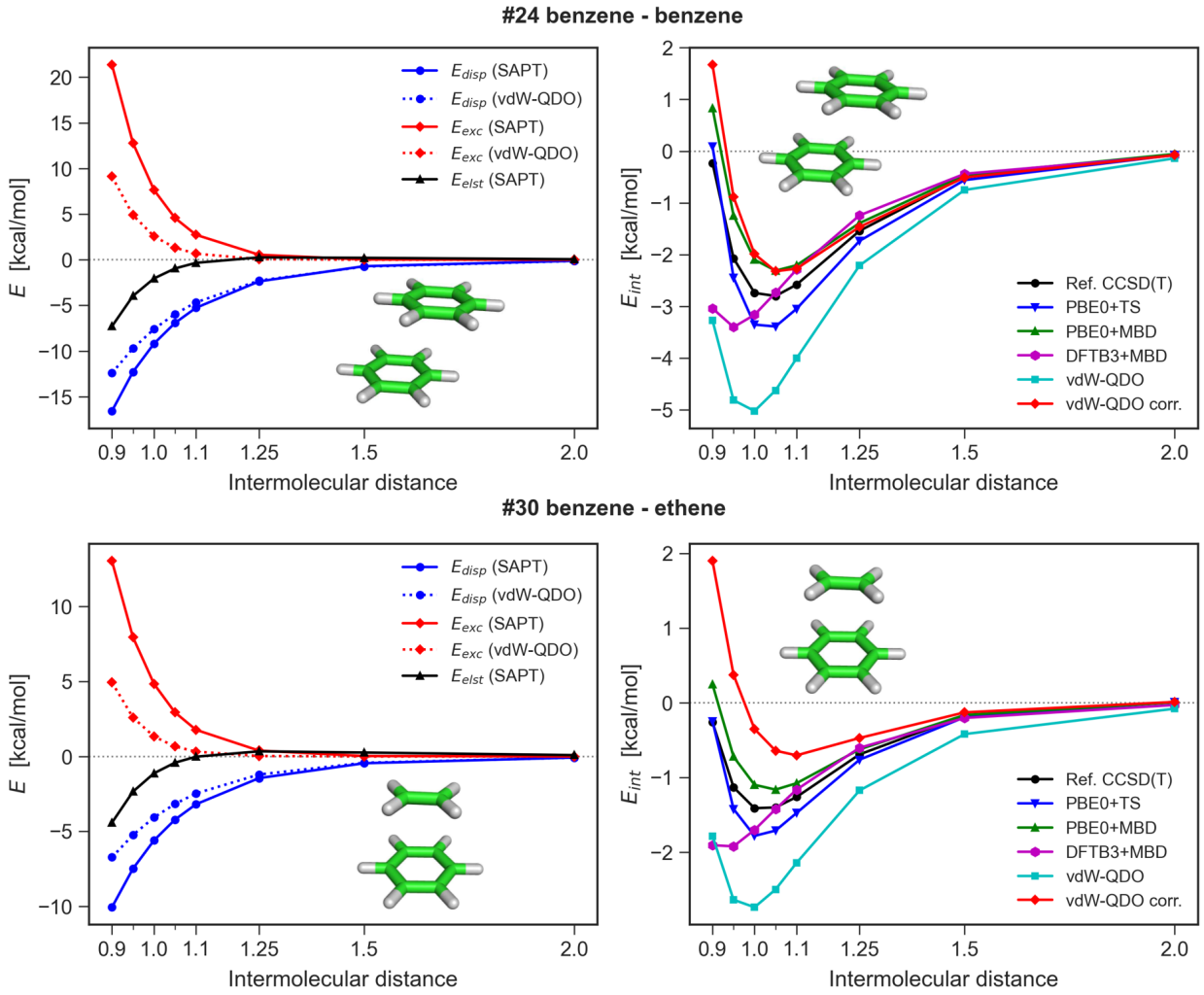

Figure A3: Interaction energy curves for benzene and benzene-ethene dimers. Left: Dispersion (blue) and exchange (red) contributions to the interaction energy of molecular dimers (shown as insets) calculated by SAPT-DFT(solid lines) and damped vdW-QDO potential (dotted lines). In addition, an electrostatic term from SAPT-DFT is displayed in black. The presented SAPT-DFT results correspond to the calculations from Ref. [234] with the ALDA xc-kernel and pVQZ basis sets. **Right:** Interaction energy curves of molecular dimers as calculated by different methods: reference CCSD(T) [182] (black circles); PBE0+TS (blue) and PBE0+MBD (green) corrections; DFTB3+MBD (magenta); damped vdW-QDO potential (cyan); SAPT-corrected vdW-QDO potential (red). Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

We also observe that the accuracy of the vdW-QDO method is higher for more spherical molecules. The best results for dispersion energy are obtained for the dimer of neopentane, which is the closest to the spherical symmetry among the considered molecules.

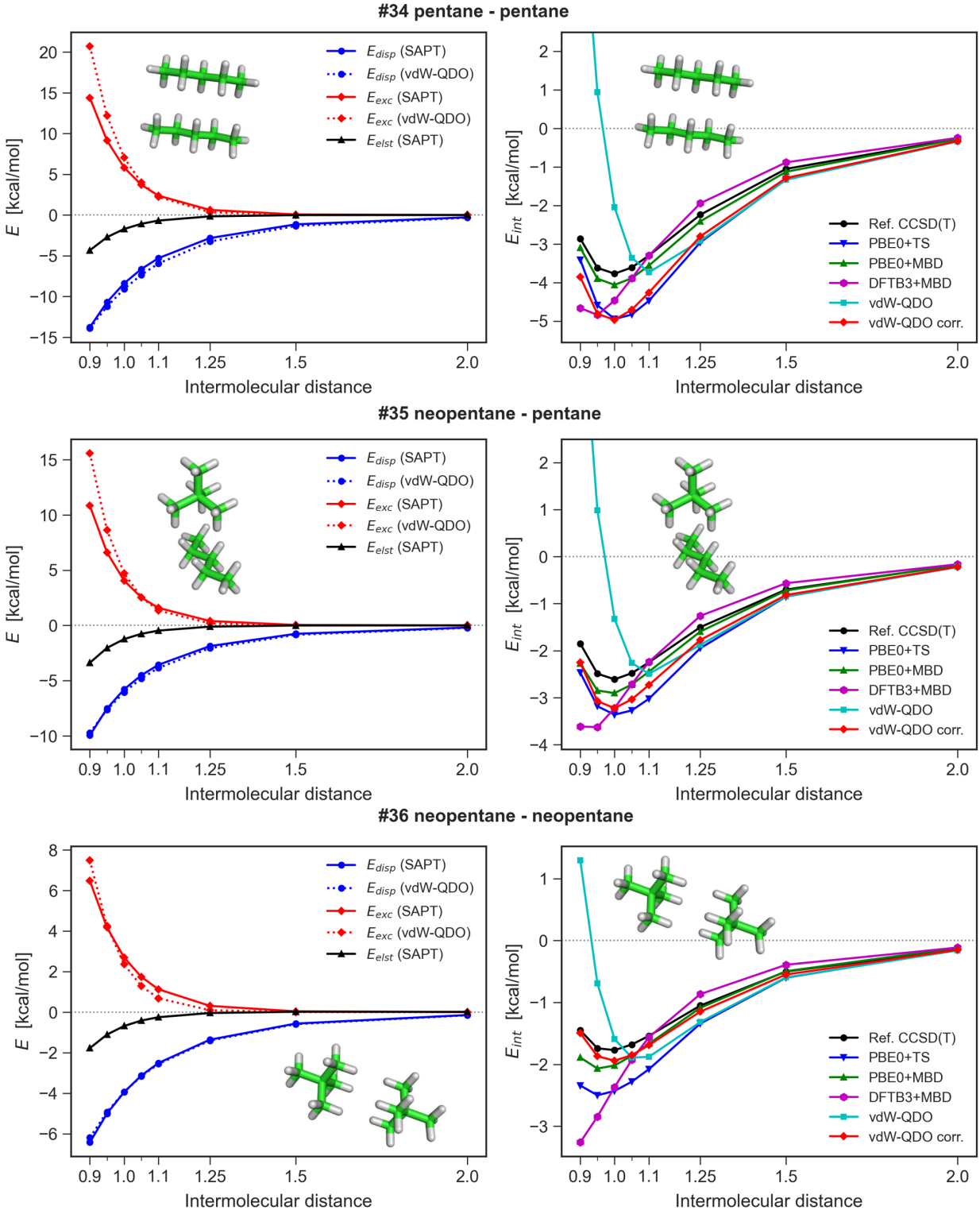

Figure A4: Interaction energy curves for pentane-pentane, neopentane-pentane, and neopentane-neopentane dimers. See the caption to Figure A3 for detailed explanations. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

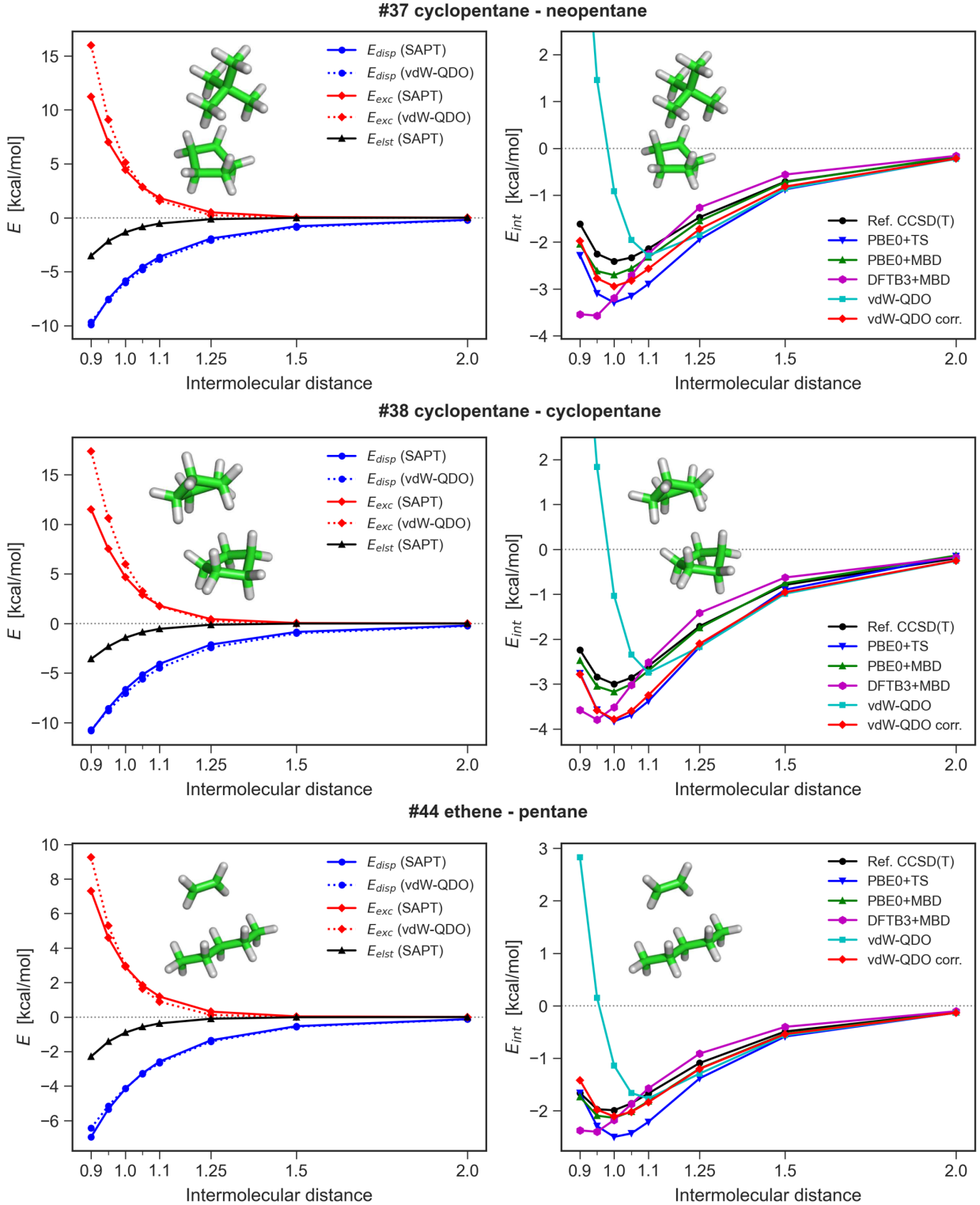

Figure A5: Interaction energy curves for cyclopentane-neopentane, cyclopentane-cyclopentane, and ethene-pentane dimers. See the caption to Figure A3 for detailed explanations. Reproduced from Ref. [107] under the CC-BY 4.0 license. © 2023 The Authors. Published by American Chemical Society.

A5 CCSD–HF Calculations: Technical Details, Additional Results

We examined basis set convergence of the CCSD−HF density differences by performing calculations with aug-cc-pVXZ basis sets, where X = D, T, Q for methane dimer and X = D, T for ethene and ethene-pentane dimers (Figure A6). The observed variations are rather small in all three cases. In methane, the TZ basis gives practically the same result as QZ, while DZ only slightly overpolarizes the dimer. For ethene and ethene-pentane dimer, DZ basis also shows the close agreement with TZ, which was the largest affordable therein. Overall, our results prove that the aug-cc-pVDZ basis set already gives sufficiently converged results, which justifies its use for pentane and neopentane dimers. Table A4 summarizes the information about the basis sets used for production calculations.

Additionally, for the methane dimer, we performed a cross-validation of the PySCF results with Q-Chem [268] and MRCC. In Q-Chem, fully relaxed (including orbital relaxation terms) CCSD density was computed. This is in contrast to PySCF and MRCC, where unrelaxed CCSD density was evaluated. Despite these minor differences in methodology, excellent agreement was obtained between the three codes (Figure A6b), ensuring the reproducibility of our results. Also, this test ensures that the orbital relaxation effects are rather small and can be safely neglected in the context of this work.

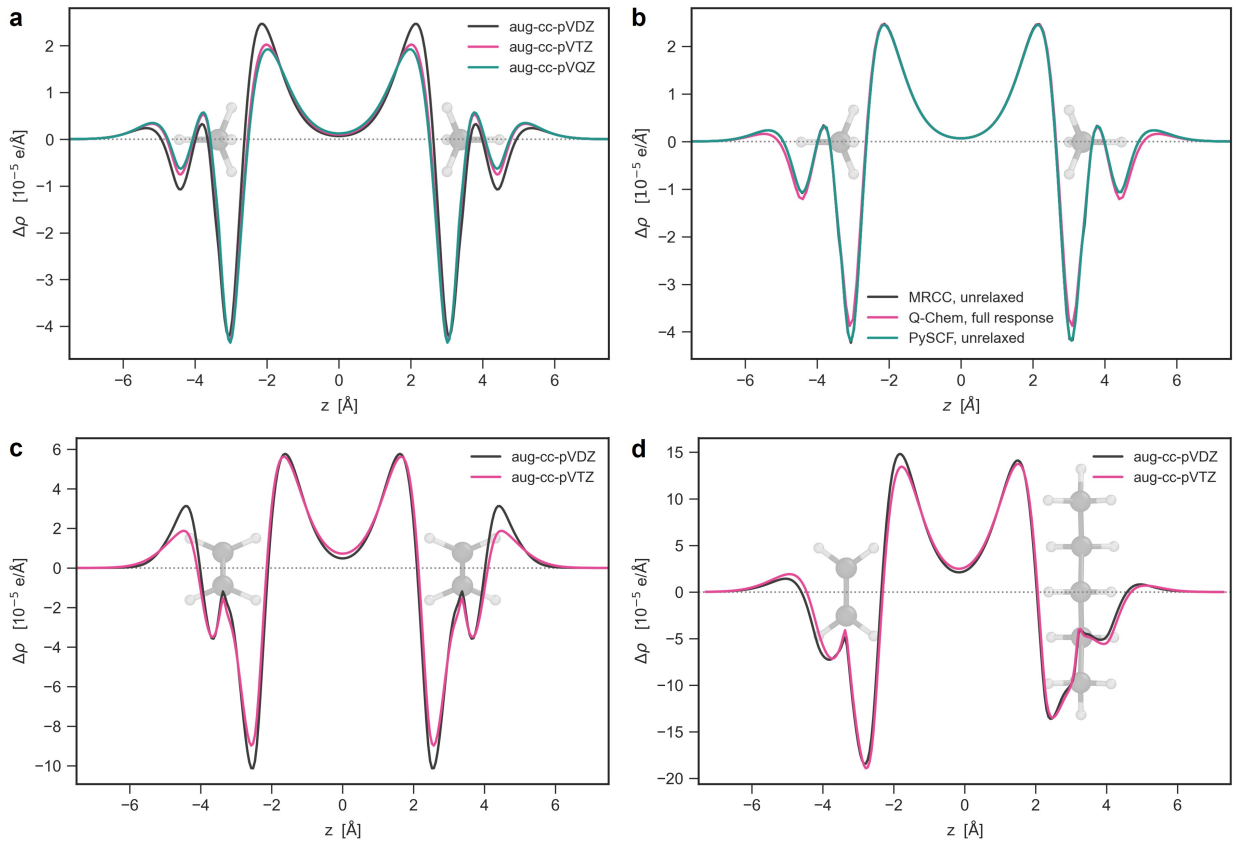

Figure A6: Basis set convergence of CCSD–HF density differences $\Delta\rho_{\mathrm{CCSD-HF}}(z)$ computed using **(a,c,d)** PySCF for **(a)** methane dimer, **(c)** ethene dimer, and **(d)** ethene-pentane dimer. **(b)** Benchmark of MRCC, Q-Chem and PySCF codes for $\Delta\rho_{\mathrm{CCSD-HF}}(z)$ at the example of methane dimer with aug-cc-pVDZ basis.

The calculations of plane-averaged $\Delta\rho_{\mathrm{CCSD-HF}}(z)$ have to be done using electron densities represented on a real-space grid in `.cube` format; therefore, sufficient numerical precision of data in these files has to be provided. The default `.cube` printing precision in PySCF is 5 digits after

Dimer	N_{elec}	Basis set	N_{basis}
methane	20	aug-cc-pVQZ	528
ethene	32	aug-cc-pVTZ	368
ethene-pentane	58	aug-cc-pVTZ	690
pentane-pentane	84	aug-cc-pVDZ	446
pentane-neopentane	84	aug-cc-pVDZ	446
neopentane-neopentane	84	aug-cc-pVDZ	446

Table A4: Basis sets used in CCSD–HF density calculations. Number of electrons in dimers N_{elec} and the corresponding number of basis functions N_{basis} are shown.

the comma, which was observed to lead to a numerical noise when computing $\Delta\rho_{\text{CCSD-HF}}(z)$ (Figure A7). It was found that this noise increases with increasing basis set cardinality and system size. We revealed that outputting the density to the `.cube` file with 9-digit precision eliminates the noise, and the source code of PySCF (`cubegen.py`) was modified accordingly. This setting was used for all PySCF calculations. For comparison, the default `.cube` printing precision is 12 digits in Q-Chem and 8 digits in FHI-aims. The cuboid grid spacing was set to 0.1 Bohr to ensure sufficient sampling of the density.

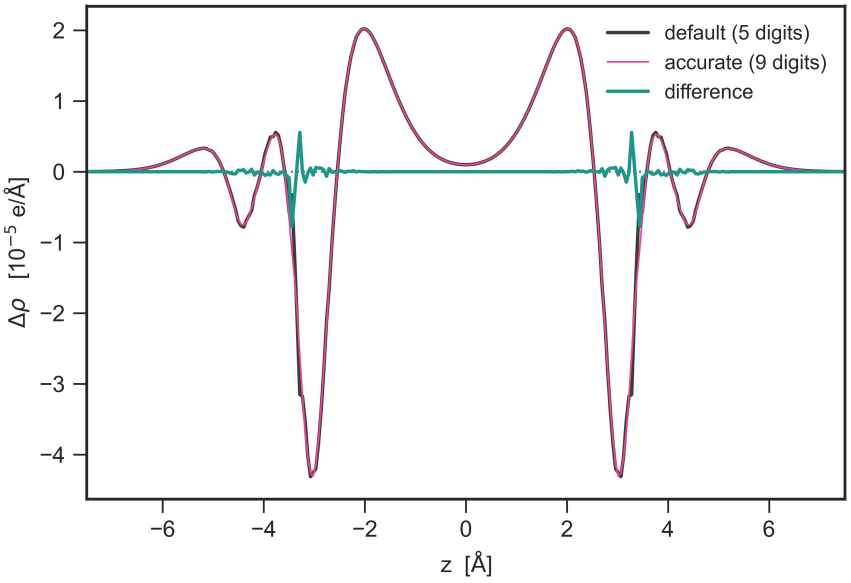

Figure A7: Effect of the density printing precision on the CCSD–HF difference. $\Delta\rho_{\text{CCSD-HF}}(z)$ for methane dimer (aug-cc-pVTZ basis) with different .cube file precision.

A6 Results with the Hartree-Fock and DFA Densities

In this section, we present the results for charge displacements in selected systems as calculated by the HF method and several DFAs. Figure A8a,b shows the ratio of vdW-displaced charge Q_{vdW} to its PBE and HF counterparts, Q_{PBE} and Q_{HF}, for the six selected systems. Overall, Q_{HF} is about 10 % larger than Q_{PBE}, which modifies the ratio accordingly. Nevertheless, this does not change the conclusions made in Chapter 4 based on the PBE densities.

For the C1–C4 sequence of polyaromatic hydrocarbons (PAH), the additional calculations of charge displacement were carried out with LDA, PBE0, and B3LYP xc-functionals as well as with the Hartree-Fock (HF) method. The computational settings were kept the same as described for the PBE case in Chapter 4. The results obtained (Figure A8c) witness that both (semi-)local and hybrid DFAs exhibit similar saturation of displaced charge moving from C1 to C4, and this is not specific to the PBE functional. The only different case is the HF method, showing almost linear behavior.

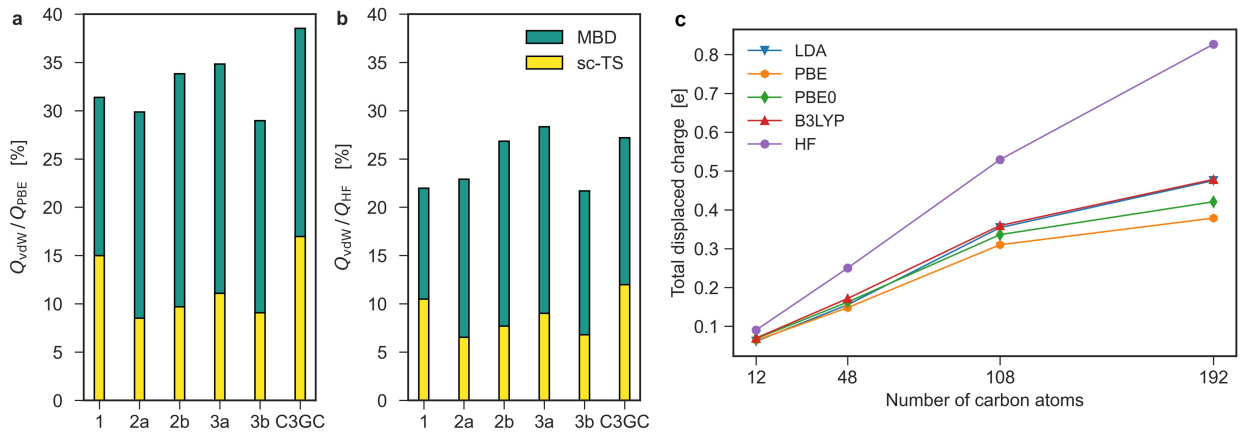

Figure A8: Interaction-displaced charges predicted by the HF and PBE methods. **(a,b)** The ratio of vdW-displaced charge Q_{vdW} as computed by sc-TS and MBD methods to **(a)** PBE-displaced charge Q_{PBE} and **(b)** HF-displaced charge Q_{HF} for the six selected systems. **(c)** The system size dependence of total displaced charge in the C1–C4 sequence of PAHs as computed by various mean-field methods.

To show that our conclusions regarding the effect of vdW-induced polarization on the NCI analysis are also not specific to the choice of DFA, we performed test calculations using the HF densities. The HF and HF+MBD results displayed in Figure A9 manifest essentially the same differences as in the case of PBE and PBE+MBD (Figure 4.12), supporting our point.

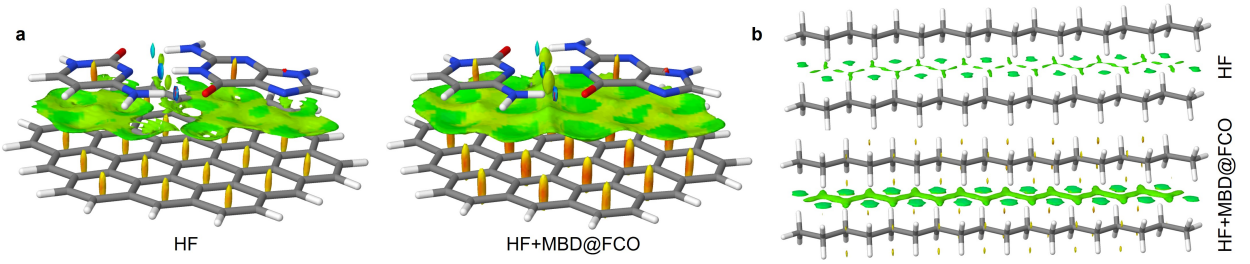

Figure A9: The NCI analysis based on the HF densities. NCI isosurfaces of density gradient s (magnitude of s and coloring range are given in a.u.) calculated using HF and HF+MBD charge densities for **(a)** C3GC (L7+): $s = 0.5$, $[-0.05, 0.04]$; **(b)** CBH (L7+): $s = 0.65$, $[-0.02, 0.02]$.

A7 NCI Analysis: Computational Details

The densities computed using the PBE functional in FHI-AIMS were output in the `.cube` 3D grid format. Analytical MBD density polarization from Eq. (4.10) was evaluated on the 3D grid identical to the PBE case using the self-written PYTHON script relying on the LIBMBD library [160]. The total PBE+MBD densities were obtained as a sum of PBE and MBD densities on the grid. The grids had a step of $h = 0.1$ Å and covered the whole system plus 7 Bohrs of vacuum as measured from the outermost atoms of the system.

To perform the NCI analysis of PBE and PBE+MBD densities, the density-based code NCIMILANO [284] was used. The code was modified by its author Dr. Gabriele Saleh upon our request to make it compatible with the `.cube` formatting produced by FHI-AIMS. NCIMILANO uses the 4th-order finite-difference formulas to compute derivatives of electron density [284] and also evaluates volume and charge of NCI isosurfaces within the fixed range of $s_{\text{iso}} = 0.1, 0.2, 0.3, \ldots, 1.0$, thus fully encompassing the typical noncovalent interactions range. While often rather fine grids are employed for NCI analysis (0.05-0.1 Bohr and less) [259, 284], for large systems this

quickly explodes `.cube` file size, making the following analysis time- and resource-consuming. The chosen $h = 0.1$ Å spacing was found to be a fair compromise delivering a sufficiently fine representation of density and its derivatives while maintaining a manageable file size. This was ensured by test calculations employing twice finer grids with $h = 0.05$ Å, in which no significant differences both in NCI isosurface plots and NCI volume ratios were observed compared to 0.1 Å step size, as summarized in Table A5.

Systems	$h = 0.1$ Å	$h = 0.05$ Å
benzene	1.950	2.026
7b ADOH-CB7	2.669	2.663
C3GC	2.180	2.210

Table A5: Effect of the grid spacing on the NCI volumes. Ratio $\gamma = V_{\mathrm{NCI}}^{\mathrm{PBE+MBD}} / V_{\mathrm{NCI}}^{\mathrm{PBE}}$ for selected systems computed using grids with 0.1 and 0.05 Å step.

A8 Additional Results on MBD Energy Density

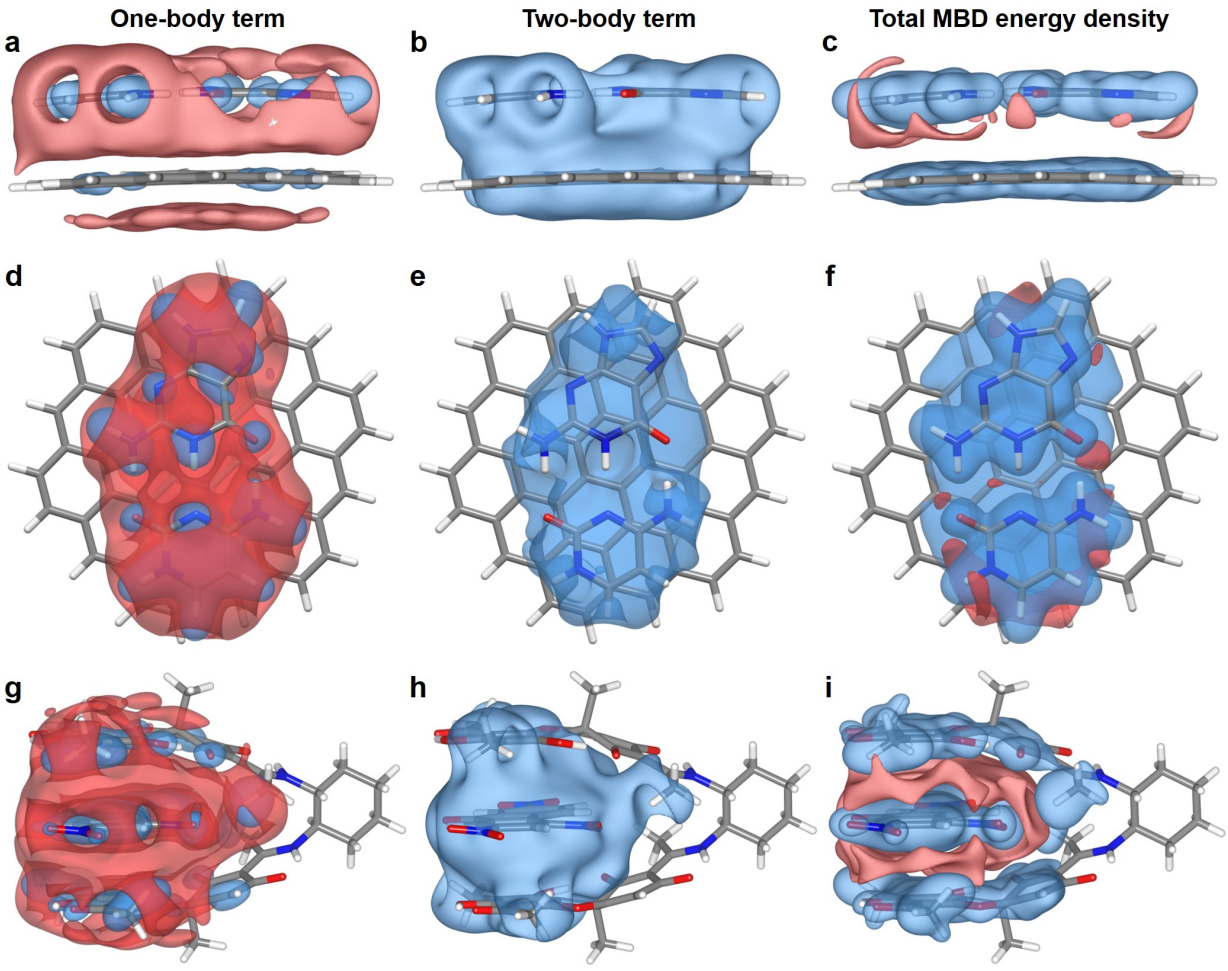

Figure A10: One- and two-body MBD energy densities and their sum for the C3GC and 3a TNF "pincer" complexes. (a-f) The isosurfaces (4×10^{-5} a.u.) showing the side **(a-c)** and top **(d-f)** views on the C3GC dimer from the L7 dataset. **(g-i)** The isosurfaces (4×10^{-5} a.u.) for the 3a TNF "pincer" from the S12L dataset. The red and blue colors denote positive and negative values, respectively.

A9 Dispersion Energy Curves for Neon and Krypton

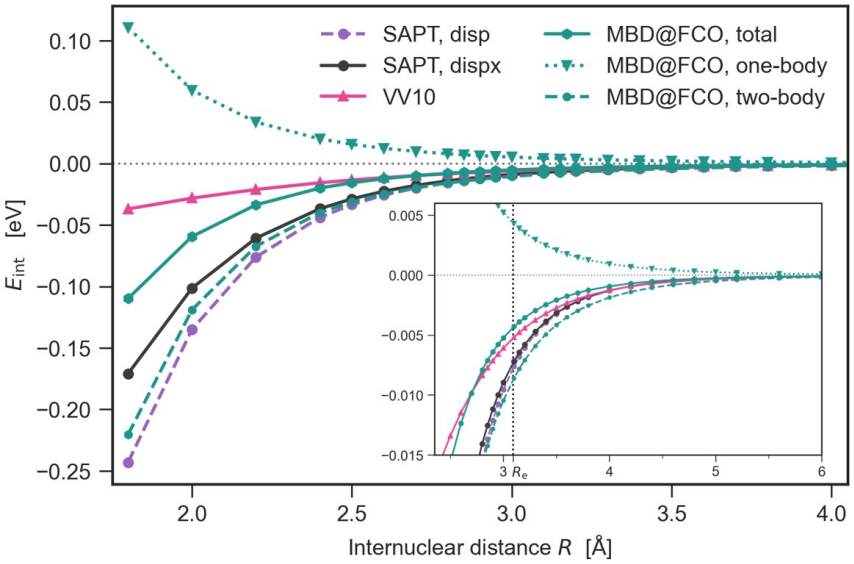

Figure A11: Dispersion energy curve of neon dimer as computed by the SAPT(CCSD) [8], VV10 and MBD@FCO methods. For MBD@FCO, the decomposition of the total energy onto the one- and two-body parts is displayed. The inset shows a zoom into the near-equilibrium region, with $R_e = 3.09$ Å marked by a vertical dotted line.

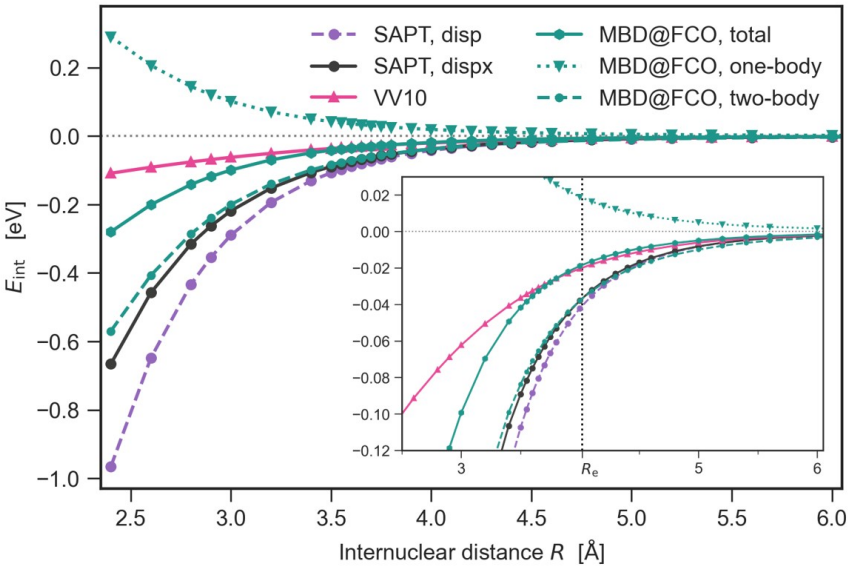

Figure A12: Dispersion energy curve of krypton dimer. The caption of Figure A11 applies, with the corresponding $R_e = 4.02$ Å.

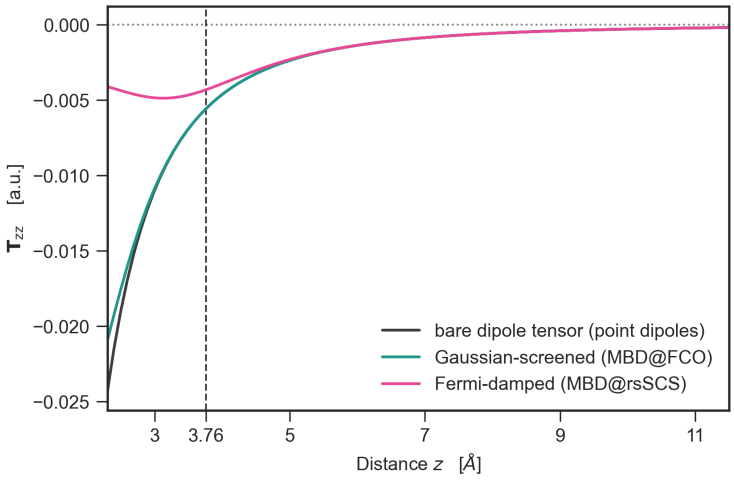

Figure A13: Dipole tensor in an argon dimer with various damping schemes. The $\mathbf{T}_{zz}$ element of a dipole tensor in argon dimer as a function of interatomic separation z for the cases of no damping (black), natural Gaussian damping (teal), and Fermi damping (pink). $R_e = 3.76$ Å is shown with a vertical dashed line.

Bibliography

[1] J. D. Van der Waals, *Over de Continuiteit van den Gas-en Vloeistoftoestand*, Ph.D. thesis, University of Leiden (1873).

[2] J. D. Van der Waals, *Verhand. Kon. Akad.* **20**, 1 (1880).

[3] K. Tang and J. P. Toennies, *Angew. Chem. Int. Ed.* **49**, 9574 (2010).

[4] J. S. Rowlinson, *Cohesion: a Scientific History of Intermolecular Forces* (Cambridge University Press, 2002).

[5] S. Wang, *Phys. Z.* **28**, 663 (1927).

[6] F. London, *Zeitschrift für Physik* **63**, 245 (1930).

[7] J. N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, 2011).

[8] L. Shirkov and V. Sladek, *J. Chem. Phys.* **147**, 174103 (2017).

[9] A. Stone, *The Theory of Intermolecular Forces* (Oxford University Press, 2013).

[10] I. G. Kaplan, *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials* (John Wiley & Sons, 2006).

[11] J. Hermann, R. A. DiStasio Jr., and A. Tkatchenko, *Chem. Rev.* **117**, 4714 (2017).

[12] M. Stöhr, T. Van Voorhis, and A. Tkatchenko, *Chem. Soc. Rev.* **48**, 4118 (2019).

[13] P. A. M. Dirac, *Proc. R. Soc. A* **123**, 714 (1929).

[14] J. Hoja, H.-Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio Jr., and A. Tkatchenko, *Sci. Adv.* **5**, eaau3338 (2019).

[15] D. Firaha, Y. M. Liu, J. van de Streek, K. Sasikumar, H. Dietrich, J. Helfferich, L. Aerts, D. E. Braun, A. Broo, A. G. DiPasquale, *et al.*, *Nature* **623**, 324 (2023).

[16] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, *Nature* **596**, 583 (2021).

[17] See https://www.nobelprize.org/prizes/chemistry/2024/press-release/.

[18] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, *Chem. Rev.* **121**, 9816 (2021).

[19] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, *Chem. Rev.* **116**, 5105 (2016).

[20] E. R. Johnson, in *Non-Covalent Interactions in Quantum Chemistry and Physics* (Elsevier, 2017) pp. 169–194.

[21] J. Hermann and A. Tkatchenko, *Phys. Rev. Lett.* **124**, 146401 (2020).

[22] R. P. Feynman, *Phys. Rev.* **56**, 340 (1939).

[23] N. Ferri, R. A. DiStasio Jr., A. Ambrosetti, R. Car, and A. Tkatchenko, *Phys. Rev. Lett.* **114**, 176802 (2015).

[24] N. Ferri, A. Ambrosetti, and A. Tkatchenko, *Phys. Rev. Mater.* **1**, 026003 (2017).

[25] J. Hermann, D. Alfè, and A. Tkatchenko, *Nat. Commun.* **8**, 14052 (2017).

[26] M. Stöhr, M. Sadhukhan, Y. S. Al-Hamdani, J. Hermann, and A. Tkatchenko, *Nat. Commun.* **12**, 137 (2021).

[27] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, *J. Comput. Chem.* **26**, 1668 (2005).

[28] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, *et al.*, *J. Comput. Chem.* **30**, 1545 (2009).

[29] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4**, 435 (2008).

[30] J. E. Jones, *Proc. R. Soc. A* **106**, 463 (1924).

[31] O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda, and K.-R. Müller, *Nat. Commun.* **12**, 7273 (2021).

[32] O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. Medrano Sandonas, J. T. Berryman, A. Tkatchenko, and K.-R. Müller, *Sci. Adv.* **10**, eadn4397 (2024).

[33] A. Kabylda, J. T. Frank, S. S. Dou, A. Khabibrakhmanov, L. M. Sandonas, O. T. Unke, S. Chmiela, K.-R. Müller, and A. Tkatchenko, *ChemRxiv: 10.26434/chemrxiv-2024-bdfr0* (2024).

[34] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989).

[35] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press USA, New York, 1994).

[36] J. Toulouse, in *Density Functional Theory: Modeling, Mathematical Analysis, Computational Methods, and Applications* (Springer, 2022) pp. 1–90.

[37] B. Jeziorski, R. Moszynski, and K. Szalewicz, *Chem. Rev.* **94**, 1887 (1994).

[38] K. Szalewicz and B. Jeziorski, *J. Mol. Model.* **28**, 273 (2022).

[39] A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.* **102**, 073005 (2009).

[40] A. Tkatchenko, R. A. DiStasio Jr., R. Car, and M. Scheffler, *Phys. Rev. Lett.* **108**, 236402 (2012).

[41] A. Ambrosetti, A. M. Reilly, R. A. DiStasio Jr., and A. Tkatchenko, *J. Chem. Phys.* **140**, 18A508 (2014).

[42] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).

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

[44] L. H. Thomas, *Math. Proc. Camb. Philos. Soc.* **23**, 542 (1927).

[45] E. Fermi, *Rend. Accad. Naz. Lincei* **6**, 32 (1927).

[46] X. Ren, P. Rinke, C. Joas, and M. Scheffler, *J. Mater. Sci.* **47**, 7447 (2012).

[47] R. Kubo, *Rep. Prog. Phys.* **29**, 255 (1966).

[48] P. Nozières and D. Pines, *Theory Of Quantum Liquids*, Vol. 1 (Benjamin, 1966).

[49] L. Landau and E. Lifschitz, *Statistical Physics*, Vol. 5 of *Course of Theoretical Physics* (Pergamon Press, 1980).

[50] S. L. Adler, *Phys. Rev.* **126**, 413 (1962).

[51] N. Wiser, *Phys. Rev.* **129**, 62 (1963).

[52] E. K. U. Gross and W. Kohn, *Phys. Rev. Lett.* **55**, 2850 (1985).

[53] E. Runge and E. K. U. Gross, *Phys. Rev. Lett.* **52**, 997 (1984).

[54] J. Toulouse, F. Colonna, and A. Savin, *Phys. Rev. A* **70**, 062505 (2004).

[55] J. Toulouse, I. C. Gerber, G. Jansen, A. Savin, and J. G. Ángyán, *Phys. Rev. Lett.* **102**, 096404 (2009).

[56] J. P. Perdew and K. Schmidt, *AIP Conf. Proc.* **577**, 1 (2001).

[57] P. A. Dirac, *Math. Proc. Camb. Philos. Soc.* **26**, 376 (1930).

[58] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).

[59] J. P. Perdew and Y. Wang, *Phys. Rev. B* **45**, 13244 (1992).

[60] T. Chachiyo, *J. Chem. Phys.* **145**, 021101 (2016).

[61] A. J. Cohen, P. Mori-Sánchez, and W. Yang, *Chem. Rev.* **112**, 289 (2012).

[62] K. R. Bryenton, A. A. Adeleke, S. G. Dale, and E. R. Johnson, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **13**, e1631 (2023).

[63] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).

[64] J. Sun, B. Xiao, Y. Fang, R. Haunschild, P. Hao, A. Ruzsinszky, G. I. Csonka, G. E. Scuseria, and J. P. Perdew, *Phys. Rev. Lett.* **111**, 106401 (2013).

[65] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.* **91**, 146401 (2003).

[66] J. Sun, A. Ruzsinszky, and J. P. Perdew, *Phys. Rev. Lett.* **115**, 036402 (2015).

[67] A. D. Becke, *J. Chem. Phys.* **98**, 1372 (1993).

[68] A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).

[69] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.* **98**, 11623 (1994).

[70] A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).

[71] C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).

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

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

[74] J. Hoja, A. M. Reilly, and A. Tkatchenko, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **7**, e1294 (2017).

[75] K. Patkowski, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **10**, e1452 (2020).

[76] H. Longuet-Higgins, *Discuss. Faraday Soc.* **40**, 7 (1965).

[77] E. Zaremba and W. Kohn, *Phys. Rev. B* **13**, 2270 (1976).

[78] M. Jaszunski and R. McWeeny, *Mol. Phys.* **55**, 1275 (1985).

[79] A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, *J. Chem. Phys.* **123**, 214103 (2005).

[80] B. Jeziorski, K. Szalewicz, and G. Chałasiński, *Int. J. Quantum Chem.* **14**, 271 (1978).

[81] K. Szalewicz, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 254 (2012).

[82] T. Korona, in *Recent Progress in Coupled Cluster Methods: Theory and Applications* (Springer Netherlands, Dordrecht, 2010) pp. 267–298.

[83] A. Hesselmann, G. Jansen, and M. Schütz, *J. Chem. Phys.* **122**, 014103 (2005).

[84] J. G. Ángyán, G. Jansen, M. Loss, C. Hättig, and B. A. Heß, *Chem. Phys. Lett.* **219**, 267 (1994).

[85] K. T. Tang and J. P. Toennies, *J. Chem. Phys.* **118**, 4976 (2003).

[86] X. Sheng, J. P. Toennies, and K. Tang, *Phys. Rev. Lett.* **125**, 253402 (2020).

[87] X. W. Sheng and K. T. Tang, *Phys. Chem. Chem. Phys.* **23**, 7748 (2021).

[88] A. Derevianko, S. G. Porsev, and J. F. Babb, *Atom. Data Nucl. Data* **96**, 323 (2010).

[89] J. Jiang, J. Mitroy, Y. Cheng, and M. Bromley, *Atom. Data Nucl. Data* **101**, 158 (2015).

[90] K. T. Tang and J. P. Toennies, *J. Chem. Phys.* **80**, 3726 (1984).

[91] R. Buckingham, *Trans. Faraday Soc.* **54**, 453 (1958).

[92] K. T. Tang, J. P. Toennies, and C. L. Yiu, *Int. Rev. Phys. Chem.* **17**, 363 (1998).

[93] P. Li, J. Ren, N. Niu, and K. Tang, *J. Phys. Chem. A* **115**, 6927 (2011).

[94] R. A. DiStasio Jr., V. V. Gobre, and A. Tkatchenko, *J. Phys. Condens. Matter* **26**, 213202 (2014).

[95] A. M. Reilly and A. Tkatchenko, *Chem. Sci.* **6**, 3289 (2015).

[96] F. Wang and K. D. Jordan, *J. Chem. Phys.* **114**, 10717 (2001).

[97] T. Sommerfeld and K. D. Jordan, *J. Phys. Chem. A* **109**, 11531 (2005).

[98] T. W. Whitfield and G. J. Martyna, *Chem. Phys. Lett.* **424**, 409 (2006).

[99] A. P. Jones, J. Crain, V. P. Sokhan, T. W. Whitfield, and G. J. Martyna, *Phys. Rev. B* **87**, 144103 (2013).

[100] M. Sadhukhan and F. R. Manby, *Phys. Rev. B* **94**, 115106 (2016).

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

[102] D. V. Fedorov, M. Sadhukhan, M. Stöhr, and A. Tkatchenko, *Phys. Rev. Lett.* **121**, 183401 (2018).

[103] O. Vaccarelli, D. V. Fedorov, M. Stöhr, and A. Tkatchenko, *Phys. Rev. Res.* **3**, 033181 (2021).

[104] A. Khabibrakhmanov, M. Gori, C. Müller, and A. Tkatchenko, *ChemRxiv: 10.26434/chemrxiv-2024-lf3nr* (2024).

[105] A. Kleshchonok and A. Tkatchenko, *Nat. Commun.* **9**, 3017 (2018).

[106] M. Sadhukhan and A. Tkatchenko, *Phys. Rev. Lett.* **118**, 210402 (2017).

[107] A. Khabibrakhmanov, D. V. Fedorov, and A. Tkatchenko, *J. Chem. Theory Comput.* **19**, 7895 (2023).

[108] P. L. Silvestrelli and A. Ambrosetti, *J. Chem. Phys.* **150**, 164109 (2019).

[109] F. S. Cipcigan, J. Crain, V. P. Sokhan, and G. J. Martyna, *Rev. Mod. Phys.* **91**, 025003 (2019).

[110] S. Góger, A. Khabibrakhmanov, O. Vaccarelli, D. V. Fedorov, and A. Tkatchenko, *J. Phys. Chem. Lett.* **14**, 6217 (2023).

[111] W. Heitler and F. London, *Z. Phys.* **44**, 455 (1927).

[112] S. Y. Buhmann, *Dispersion Forces I*, Vol. 247 of *Springer Tracts in Modern Physics* (Springer, 2013).

[113] S. Buhmann, *Dispersion Forces II*, Vol. 248 of *Springer Tracts in Modern Physics* (Springer, 2013).

[114] L. M. Woods, D. A. R. Dalvit, A. Tkatchenko, P. Rodriguez-Lopez, A. W. Rodriguez, and R. Podgornik, *Rev. Mod. Phys.* **88**, 045003 (2016).

[115] R. Nimalakirthi and K. Hunt, *J. Chem. Phys.* **98**, 3066 (1993).

[116] B. Axilrod and E. Teller, *J. Chem. Phys.* **11**, 299 (1943).

[117] Y. Muto, *Proc. Phys. Math. Soc. Jpn.* **17**, 629 (1943).

[118] E. Lifschitz, *Sov. Phys. JETP* **2**, 73 (1956).

[119] J. F. Dobson, *Int. J. Quantum Chem.* **114**, 1157 (2014).

[120] A. Ambrosetti, N. Ferri, R. A. DiStasio Jr., and A. Tkatchenko, *Science* **351**, 1171 (2016).

[121] P. Hauseux, T.-T. Nguyen, A. Ambrosetti, K. S. Ruiz, S. P. Bordas, and A. Tkatchenko, *Nat. Commun.* **11**, 1651 (2020).

[122] P. Hauseux, A. Ambrosetti, S. P. A. Bordas, and A. Tkatchenko, *Phys. Rev. Lett.* **128**, 106101 (2022).

[123] H. B. G. Casimir and D. Polder, *Phys. Rev.* **73**, 360 (1948).

[124] J. Hermann, *Towards Unified Density-Functional Model of van der Waals Interactions*, Ph.D. thesis, Humboldt-Universität zu Berlin (2018).

[125] S. Grimme, *J. Comput. Chem.* **25**, 1463 (2004).

[126] S. Grimme, *J. Comput. Chem.* **27**, 1787 (2006).

[127] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010).

[128] J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **10**, 6615 (2008).

[129] S. Grimme, S. Ehrlich, and L. Goerigk, *J. Comput. Chem.* **32**, 1456 (2011).

[130] A. Otero-De-La-Roza and E. R. Johnson, *J. Chem. Phys.* **138**, 174109 (2013).

[131] A. M. Reilly and A. Tkatchenko, *J. Chem. Phys.* **139**, 024705 (2013).

[132] J. Řezáč, Y. Huang, P. Hobza, and G. J. Beran, *J. Chem. Theory Comput.* **11**, 3065 (2015).

[133] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, *J. Chem. Phys.* **150**, 154122 (2019).

[134] A. D. Becke and E. R. Johnson, *J. Chem. Phys.* **122**, 154104 (2005).

[135] A. D. Becke and E. R. Johnson, *J. Chem. Phys.* **127**, 154108 (2007).

[136] A. D. Becke and M. R. Roussel, *Phys. Rev. A* **39**, 3761 (1989).

[137] A. D. Becke and E. R. Johnson, *J. Chem. Phys.* **124**, 014104 (2006).

[138] A. Otero-De-La-Roza and E. R. Johnson, *J. Chem. Phys.* **136**, 174109 (2012).

[139] A. J. Price, A. Otero-de-la Roza, and E. R. Johnson, *Chem. Sci.* **14**, 1252 (2023).

[140] A. Koide, *J. Phys. B: Atom. Mol. Phys.* **9**, 3173 (1976).

[141] J. G. Ángyán, *J. Chem. Phys.* **127**, 024108 (2007).

[142] J. Ángyán, *Int. J. Quantum Chem.* **109**, 2340 (2009).

[143] P. W. Ayers, *J. Math. Chem.* **46**, 86 (2009).

[144] A. Heßelmann, *J. Chem. Phys.* **130**, 084104 (2009).

[145] K. Tang, *Phys. Rev.* **177**, 108 (1969).

[146] S. Góger, M. R. Karimpour, and A. Tkatchenko, *J. Chem. Theory Comput.* **20**, 6621 (2024).

[147] P. Bultinck, C. Van Alsenoy, P. W. Ayers, and R. Carbó-Dorca, *J. Chem. Phys.* **126**, 144111 (2007).

[148] T. Bučko, S. Lebègue, J. Hafner, and J. G. Ángyán, *J. Chem. Theory Comput.* **9**, 4293 (2013).

[149] T. Bučko, S. Lebègue, J. G. Ángyán, and J. Hafner, *J. Chem. Phys.* **141**, 034114 (2014).

[150] T. Gould, S. Lebègue, J. G. Ángyán, and T. Bučko, *J. Chem. Theory Comput.* **12**, 5920 (2016).

[151] V. G. Ruiz, W. Liu, E. Zojer, M. Scheffler, and A. Tkatchenko, *Phys. Rev. Lett.* **108**, 146103 (2012).

[152] M. Stöhr, G. S. Michelitsch, J. C. Tully, K. Reuter, and R. J. Maurer, *J. Chem. Phys.* **144**, 151101 (2016).

[153] P. P. Poier, T. Jaffrelot Inizan, O. Adjoua, L. Lagardere, and J.-P. Piquemal, *J. Phys. Chem. Lett.* **13**, 4381 (2022).

[154] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, *Phys. Rev. Lett.* **92**, 246401 (2004).

[155] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, *Phys. Rev. B* **82**, 081101 (2010).

[156] O. A. Vydrov and T. Van Voorhis, *J. Chem. Phys.* **130**, 104105 (2009).

[157] O. A. Vydrov and T. Van Voorhis, *Phys. Rev. Lett.* **103**, 063004 (2009).

[158] O. A. Vydrov and T. Van Voorhis, *Phys. Rev. A* **81**, 062708 (2010).

[159] O. A. Vydrov and T. Van Voorhis, *J. Chem. Phys.* **133**, 244103 (2010).

[160] J. Hermann, M. Stöhr, S. Góger, S. Chaudhuri, B. Aradi, R. J. Maurer, and A. Tkatchenko, *J. Chem. Phys.* **159**, 174802 (2023).

[161] A. Mayer, *Phys. Rev. B* **75**, 045407 (2007).

[162] Y. S. Al-Hamdani, P. R. Nagy, A. Zen, D. Barton, M. Kállay, J. G. Brandenburg, and A. Tkatchenko, *Nat. Commun.* **12**, 3927 (2021).

[163] A. Ambrosetti, D. Alfè, R. A. DiStasio Jr., and A. Tkatchenko, *J. Phys. Chem. Lett.* **5**, 849 (2014).

[164] M. Kim, W. J. Kim, T. Gould, E. K. Lee, S. Lebègue, and H. Kim, *J. Am. Chem. Soc.* **142**, 2346 (2020).

[165] D. Massa, A. Ambrosetti, and P. L. Silvestrelli, *Electron. Struct.* **3**, 044002 (2021).

[166] D. Massa, A. Ambrosetti, and P. L. Silvestrelli, *J. Chem. Phys.* **154**, 224115 (2021).

[167] P. P. Poier, O. Adjoua, L. Lagardère, and J.-P. Piquemal, *J. Phys. Chem. Lett.* **14**, 1609 (2023).

[168] A. Ambrosetti, P. Umari, P. L. Silvestrelli, J. Elliott, and A. Tkatchenko, *Nat. Commun.* **13**, 813 (2022).

[169] J. F. Dobson and A. Ambrosetti, *J. Chem. Theory Comput.* **19**, 6434 (2023).

[170] P. P. Poier, L. Lagardere, and J.-P. Piquemal, *J. Chem. Theory Comput.* **18**, 1633 (2022).

[171] P. P. Poier, T. Jaffrelot Inizan, O. Adjoua, L. Lagardere, and J.-P. Piquemal, *J. Phys. Chem. Lett.* **13**, 4381 (2022).

[172] M. Gori, P. Kurian, and A. Tkatchenko, *Nat. Commun.* **14**, 8218 (2023).

[173] R. A. Buckingham, *Proc. R. Soc. A* **168**, 264 (1938).

[174] J. Barker, R. Fisher, and R. Watts, *Mol. Phys.* **21**, 657 (1971).

[175] B. L. Blaney and G. E. Ewing, *Annu. Rev. Phys. Chem.* **27**, 553 (1976).

[176] K. Tang and J. Toennies, *Mol. Phys.* **106**, 1645 (2008).

[177] J. C. Slater, *J. Chem. Phys.* **43**, S11 (1965).

[178] A. Tkatchenko, D. V. Fedorov, and M. Gori, *J. Phys. Chem. Lett.* **12**, 9488 (2021).

[179] J. De Boer and A. Michels, *Physica* **5**, 945 (1938).

[180] J. De Boer, *Physica* **14**, 139 (1948).

[181] D. A. McQuarrie and J. D. Simon, *Physical Chemistry: a Molecular Approach*, Vol. 1 (University Science Books, 1997).

[182] J. Řezáč, K. E. Riley, and P. Hobza, *J. Chem. Theory Comput.* **7**, 2427 (2011).

[183] W. L. Bade, *J. Chem. Phys.* **27**, 1280 (1957).

[184] B. Thole, *Chem. Phys.* **59**, 341 (1981).

[185] A. Jones, *Quantum Drude Oscillators for Accurate Many-Body Intermolecular Forces*, Ph.D. thesis, The University of Edinburgh (2010).

[186] A. Jones, F. Cipcigan, V. P. Sokhan, J. Crain, and G. J. Martyna, *Phys. Rev. Lett.* **110**, 227801 (2013).

[187] V. P. Sokhan, A. P. Jones, F. S. Cipcigan, J. Crain, and G. J. Martyna, *Proc. Natl. Acad. Sci. USA* **112**, 6341 (2015).

[188] M. R. Karimpour, D. V. Fedorov, and A. Tkatchenko, *Phys. Rev. Res.* **4**, 013011 (2022).

[189] M. R. Karimpour, D. V. Fedorov, and A. Tkatchenko, *J. Phys. Chem. Lett.* **13**, 2197 (2022).

[190] A. Ambrosetti and P. L. Silvestrelli, *Phys. Rev. B* **85**, 073101 (2012).

[191] E. Harder, V. M. Anisimov, I. V. Vorobyov, P. E. M. Lopes, S. Y. Noskov, A. D. MacKerell, and B. Roux, *J. Chem. Theory Comput.* **2**, 1587 (2006).

[192] P. E. M. Lopes, J. Huang, J. Shim, Y. Luo, H. Li, B. Roux, and A. D. MacKerell, *J. Chem. Theory Comput.* **9**, 5430 (2013).

[193] A. N. S. Adluri, J. N. Murphy, T. Tozer, and C. N. Rowley, *J. Phys. Chem. B* **119**, 13422 (2015).

[194] P. P. Poier, L. Lagardère, and J.-P. Piquemal, *J. Chem. Theory Comput.* **18**, 1633 (2022).

[195] H. Muhli, X. Chen, A. P. Bartók, P. Hernández-León, G. Csányi, T. Ala-Nissila, and M. A. Caro, *Phys. Rev. B* **104**, 054106 (2021).

[196] M. Born and J. E. Mayer, *Z. Phys.* **75**, 1 (1932).

[197] R. Hellmann, E. Bich, and E. Vogel, *Mol. Phys.* **106**, 133 (2008).

[198] E. Duman and B. Smirnov, *Opt. Spectrosc* **29**, 229 (1970).

[199] U. Kleinekathöfer, K. Tang, J. Toennies, and C. Yiu, *J. Chem. Phys.* **103**, 6617 (1995).

[200] L. Salem, *Proc. R. Soc. A* **264**, 379 (1961).

[201] J. A. Rackers and J. W. Ponder, *J. Chem. Phys.* **150**, 084104 (2019).

[202] K. Tang and J. P. Toennies, *Z. Phys. D* **1**, 91 (1986).

[203] M. J. Van Vleet, A. J. Misquitta, A. J. Stone, and J. R. Schmidt, *J. Chem. Theory Comput.* **12**, 3851 (2016).

[204] A. Bondi, *J. Phys. Chem.* **68**, 441 (1964).

[205] J. Farrar, T. Schafer, and Y.-T. Lee, *AIP Conf. Proc.* **11**, 279 (1973).

[206] K. T. Tang and J. P. Toennies, *J. Chem. Phys.* **66**, 1496 (1977).

[207] W. Cencek, M. Przybytek, J. Komasa, J. B. Mehl, B. Jeziorski, and K. Szalewicz, *J. Chem. Phys.* **136**, 224303 (2012).

[208] K. Patkowski and K. Szalewicz, *J. Chem. Phys.* **133**, 094304 (2010).

[209] B. Jäger, R. Hellmann, E. Bich, and E. Vogel, *J. Chem. Phys.* **144**, 114304 (2016).

[210] R. Hellmann, B. Jäger, and E. Bich, *J. Chem. Phys.* **147**, 034304 (2017).

[211] J. López Cacheiro, B. Fernández, D. Marchesan, S. Coriani, C. Hättig, and A. Rizzo, *Mol. Phys.* **102**, 101 (2004).

[212] T. P. Haley and S. M. Cybulski, *J. Chem. Phys.* **119**, 5487 (2003).

[213] A. Shee, S. Knecht, and T. Saue, *Phys. Chem. Chem. Phys.* **17**, 10978 (2015).

[214] J. Delhommelle and P. Millié, *Mol. Phys.* **99**, 619 (2001).

[215] W. Song, P. J. Rossky, and M. Maroncelli, *J. Chem. Phys.* **119**, 9145 (2003).

[216] D. Boda and D. Henderson, *Mol. Phys.* **106**, 2367 (2008).

[217] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, *Crit. Rev. Solid State Mater. Sci.* **39**, 1 (2014).

[218] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, *et al.*, *Science* **351**, aad3000 (2016).

[219] D. Yang, P. Li, and K. Tang, *J. Chem. Phys.* **131**, 154301 (2009).

[220] P. Li, W. Xie, and K. Tang, *J. Chem. Phys.* **133**, 084308 (2010).

[221] G. Yin, P. Li, and K. Tang, *J. Chem. Phys.* **132**, 074303 (2010).

[222] L. Wei, P. Li, L. Qiao, and K. Tang, *J. Chem. Phys.* **139**, 154306 (2013).

[223] L. Wei, P. Li, and K. Tang, *Chem. Phys. Lett.* **635**, 285 (2015).

[224] S. G. Porsev and A. Derevianko, *J. Exp. Theor. Phys.* **102**, 195 (2006).

[225] H. Ladjimi and M. Tomza, *Phys. Rev. A* **109**, 052814 (2024).

[226] P. Schwerdtfeger and J. K. Nagle, *Mol. Phys.* **117**, 1200 (2019).

[227] K. Zaremba-Kopczyk and M. Tomza, *Phys. Rev. A* **104**, 042816 (2021).

[228] E. Pahl, D. Figgen, C. Thierfelder, K. A. Peterson, F. Calvo, and P. Schwerdtfeger, *J. Chem. Phys.* **132**, 114301 (2010).

[229] E. Tiesinga, S. Kotochigova, and P. S. Julienne, *Phys. Rev. A* **65**, 042722 (2002).

[230] O. Allard, A. Pashov, H. Knöckel, and E. Tiemann, *Phys. Rev. A* **66**, 042503 (2002).

[231] A. Stein, H. Knöckel, and E. Tiemann, *Phys. Rev. A* **78**, 042508 (2008).

[232] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, *et al.*, *J. Chem. Phys.* **152**, 124101 (2020).

[233] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, *Comput. Phys. Commun.* **180**, 2175 (2009).

[234] A. Heßelmann, *J. Chem. Theory Comput.* **14**, 1943 (2018).

[235] M. Stöhr, L. Medrano Sandonas, and A. Tkatchenko, *J. Phys. Chem. Lett.* **11**, 6835 (2020).

[236] R. Podeszwa, K. Pernal, K. Patkowski, and K. Szalewicz, *J. Phys. Chem. Lett.* **1**, 550 (2010).

[237] J. Westermayr, S. Chaudhuri, A. Jeindl, O. T. Hofmann, and R. J. Maurer, *Digit. Discov.* **1**, 463 (2022).

[238] J. Hoja, L. Medrano Sandonas, B. G. Ernst, A. Vazquez-Mayagoitia, R. A. DiStasio Jr., and A. Tkatchenko, *Sci. Data* **8**, 43 (2021).

[239] L. Medrano Sandonas, D. Van Rompaey, A. Fallani, M. Hilfiker, D. Hahn, L. Perez-Benito, J. Verhoeven, G. Tresadern, J. Kurt Wegner, H. Ceulemans, *et al.*, *Sci. Data* **11**, 742 (2024).

[240] T. Plé, O. Adjoua, L. Lagardère, and J.-P. Piquemal, *J. Chem. Phys.* **161**, 042502 (2024).

[241] See https://github.com/thomasple/FeNNol for the FeNNol code.

[242] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, *Comp. Phys. Comm.* **271**, 108171 (2022).

[243] M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, and K. A. Lyssenko, *Science* **355**, 49 (2017).

[244] M.-C. Kim, E. Sim, and K. Burke, *Phys. Rev. Lett.* **111**, 073003 (2013).

[245] S. Vuckovic, S. Song, J. Kozlowski, E. Sim, and K. Burke, *J. Chem. Theory Comput.* **15**, 6636 (2019).

[246] E. Sim, S. Song, S. Vuckovic, and K. Burke, *J. Am. Chem. Soc.* **144**, 6625 (2022).

[247] M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, and K. Burke, *Nat. Commun.* **11**, 5223 (2020).

[248] O. Unke, M. Bogojeski, M. Gastegger, M. Geiger, T. Smidt, and K.-R. Müller, *Adv. Neural Inf. Process. Syst.* **34**, 14434 (2021).

[249] T. Koker, K. Quigley, E. Taw, K. Tibbetts, and L. Li, *npj Comput. Mater.* **10**, 161 (2024).

[250] S. Dick and M. Fernandez-Serra, *Nat. Commun.* **11**, 3509 (2020).

[251] J. A. Rackers, L. Tecot, M. Geiger, and T. E. Smidt, *Mach. Learn.: Sci. Technol.* **4**, 015027 (2023).

[252] L. Cheng, P. B. Szabó, Z. Schätzle, D. Kooi, J. Köhler, K. J. Giesbertz, F. Noé, J. Hermann, P. Gori-Giorgi, and A. Foster, *arXiv: 2409.01306* (2024).

[253] S. N. Steinmann, C. Piemontesi, A. Delachat, and C. Corminboeuf, *J. Chem. Theory Comput.* **8**, 1629 (2012).

[254] P. Politzer, P. R. Laurence, and K. Jayasuriya, *Environ. Health Perspect.* **61**, 191 (1985).

[255] A. D. Becke and K. E. Edgecombe, *J. Chem. Phys.* **92**, 5397 (1990).

[256] B. Silvi and A. Savin, *Nature* **371**, 683 (1994).

[257] R. Bader, *Atoms in Molecules: A Quantum Theory* (Oxford Science Publications, 1990).

[258] R. F. Bader, *Chem. Rev.* **91**, 893 (1991).

[259] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang, *J. Am. Chem. Soc.* **132**, 6498 (2010).

[260] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, and W. Yang, *J. Chem. Theory Comput.* **7**, 625 (2011).

[261] J. Contreras-García, R. A. Boto, F. Izquierdo-Ruiz, I. Reva, T. Woller, and M. Alonso, *Theor. Chem. Acc.* **135**, 1 (2016).

[262] E. Pastorczak and C. Corminboeuf, *J. Chem. Phys.* **146**, 120901 (2017).

[263] M. Stöhr and A. Tkatchenko, *Sci. Adv.* **5**, eaax0024 (2019).

[264] S. Grimme, *Chem. Eur. J.* **18**, 9955 (2012).

[265] R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, and P. Hobza, *J. Chem. Theory Comput.* **9**, 3364 (2013).

[266] R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992).

[267] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, *et al.*, *J. Chem. Phys.* **153**, 024109 (2020).

[268] E. Epifanovsky, A. T. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, *et al.*, *J. Chem. Phys.* **155**, 084801 (2021).

[269] M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, B. Hégely, *et al.*, *J. Chem. Phys.* **152**, 074107 (2020).

[270] A. Heßelmann and T. Korona, *J. Chem. Phys.* **141**, 094107 (2014).

[271] M. Shahbaz and K. Szalewicz, *Phys. Rev. Lett.* **121**, 113402 (2018).

[272] M. Shahbaz and K. Szalewicz, *Theor. Chem. Acc.* **138**, 25 (2019).

[273] M. Gray and J. Herbert, *ChemrXiv: 10.26434/chemrxiv-2023-xgr3q-v3* (2024).

[274] M. Stöhr, *van der Waals Dispersion Interactions in Biomolecular Systems: Quantum-Mechanical Insights and Methodological Advances*, Ph.D. thesis, University of Luxembourg (2020).

[275] K. R. Brorsen, Y. Yang, M. V. Pak, and S. Hammes-Schiffer, *J. Phys. Chem. Lett.* **8**, 2076 (2017).

[276] V. Kvasnička, V. Laurinc, and S. Biskupič, *Mol. Phys.* **39**, 143 (1980).

[277] S. A. Kucharski and R. J. Bartlett, in *Advances in Quantum Chemistry*, Vol. 18 (Elsevier, 1986) pp. 281–344.

[278] D. S. Ranasinghe, A. Perera, and R. J. Bartlett, *J. Chem. Phys.* **147**, 204103 (2017).

[279] P. D. Mezei, G. I. Csonka, and M. Kállay, *J. Chem. Theory Comput.* **13**, 4753 (2017).

[280] K. Momma and F. Izumi, *J. Appl. Crystallogr.* **44**, 1272 (2011).

[281] L. Gráfová, M. Pitoňák, J. Řezáč, and P. Hobza, *J. Chem. Theory Comput.* **6**, 2365 (2010).

[282] K. Hunt, *J. Chem. Phys.* **92**, 1180 (1990).

[283] C. Villot, F. Ballesteros, D. Wang, and K. U. Lao, *J. Phys. Chem. A* **126**, 4326 (2022).

[284] G. Saleh, L. Lo Presti, C. Gatti, and D. Ceresoli, *J. Appl. Crystallogr.* **46**, 1513 (2013).

[285] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, *et al.*, *Science* **330**, 341 (2010).

[286] K. A. Dill and J. L. MacCallum, *Science* **338**, 1042 (2012).

[287] R. W. Newberry and R. T. Raines, *ACS Chem. Biol.* **14**, 1677 (2019).

[288] E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin, *Protein Sci.* **30**, 70 (2021).

[289] I. Nardecchia, J. Torres, M. Lechelon, V. Giliberti, M. Ortolani, P. Nouvel, M. Gori, Y. Meriguet, I. Donato, J. Preto, *et al.*, *Phys. Rev. X* **8**, 031061 (2018).

[290] M. Lechelon, Y. Meriguet, M. Gori, S. Ruffenach, I. Nardecchia, E. Floriani, D. Coquillat, F. Teppe, S. Mailfert, D. Marguet, *et al.*, *Sci. Adv.* **8**, eabl5855 (2022).

[291] A. J. Price, K. R. Bryenton, and E. R. Johnson, *J. Chem. Phys.* **154**, 230902 (2021).

[292] K. Pernal, R. Podeszwa, K. Patkowski, and K. Szalewicz, *Phys. Rev. Lett.* **103**, 263201 (2009).

[293] H. Muhli, T. Ala-Nissila, and M. A. Caro, *arXiv: 2407.06409* (2024).

[294] F. Neese, *WIRES Comput. Molec. Sci.* **12**, e1606 (2022).

[295] T. Lu and F. Chen, *J. Comput. Chem.* **33**, 580 (2012).

[296] B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán, and S. Alvarez, *Dalton Trans.* , 2832 (2008).

[297] C. Castillo-Orellana, F. Heidar-Zadeh, and E. Vöhringer-Martinez, *ChemrXiv: 10.26434/chemrxiv-2025-ttjb3* (2025).

[298] F. Ballesteros, S. Dunivan, and K. U. Lao, *J. Chem. Phys.* **154**, 154104 (2021).

[299] K. U. Lao, *J. Chem. Phys.* **161**, 234103 (2024).

[300] M. Puleva, L. M. Sandonas, B. Lőrincz, J. Charry, D. M. Rogers, P. R. Nagy, and A. Tkatchenko, *ChemrXiv: 10.26434/chemrxiv-2025-f6615* (2025).

[301] L. V. Slipchenko and M. S. Gordon, *Mol. Phys.* **107**, 999 (2009).

[302] B. A. Mamedov and E. Somuncu, *J. Mol. Struct.* **1068**, 164 (2014)