

Electronic Structure Calculations and Machine Learning

Citation for published version (APA):

Çaylak, O. (2023). *Electronic Structure Calculations and Machine Learning: from Theory to Multiscale Applications*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Mathematics and Computer Science]. Eindhoven University of Technology.

Document status and date:

Published: 05/12/2023

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Electronic Structure Calculations and Machine Learning

from Theory to Multiscale Applications

Onur Çaylak

Copyright © 2023 by O. Çaylak. All Rights Reserved.

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

Thesis title: Electronic Structure Calculations and Machine Learning,
from Theory to Multiscale Applications / by O. Çaylak
Technische Universiteit Eindhoven, 2023. Proefschrift.

Cover design by Onur Çaylak

A catalogue record is available from the Eindhoven University of Technology Library

Keywords: Quantum Machine Learning, Electronic Structure Calculations

The work in this thesis has been sponsored by NWO.

Electronic Structure Calculations and Machine Learning: from Theory to Multiscale Applications

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr. S.K. Lenaerts, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op
dinsdag 5 december 2023 om 13:30 uur

door

ONUR ÇAYLAK

geboren te VENRAY

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter: prof. dr. M.G.J. van den Brand

promotor: dr. B. Baumeier

co-promotor: prof. dr. O.A. von Lilienfeld (University of Toronto)

leden: prof. dr. K. P. L. Veroy-Grepl

prof. dr. S. Calero

dr. M. Rupp (Luxemburg Institute of Science and Technology)

prof. dr. T. Bereau (Heidelberg University)

Het onderzoek dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

Abstract

Understanding the behavior of electrons at sub-nanometer scales is pivotal for the functionality of diverse materials, underpinning crucial physical, chemical, and biological processes. Central to this behavior lies the vast chemical compound space. Within this expanse, the number of feasible small organic molecules is astoundingly estimated between 10^{60} and 10^{200} . A grasp of the relationship between compound structures and their properties can enable the prediction of novel compound behaviors, eliminating the need for exhaustive physical testing.

This research explores the connection between electronic structure calculations and machine learning. We delve into the application of the many-body Green's function (*GW*-BSE) method within Electronic Structure Theory, targeting the optimized geometry of molecules in excited states. The potential of the Wasserstein metric within Machine Learning Theory is explored as an alternative metric addressing atom-index invariance challenges in adjacency matrix-based representations. We also introduce the use of a machine learning surrogate model within Hybrid *GW*-BSE/ML Models, proficient at simplifying the complex steps of *GW* calculations and integrating with quantum-classical embedding. Further, a framework for a surrogate neural network model within the context of Hybrid Multiscale Models is presented, emphasizing the inter-molecular electronic coupling in disordered material morphologies.

List of Publications

Publications marked with ■ (☒) are used fully (in part) in this thesis.

- **Electronic Excitations in Complex Molecular Environments: Many-Body Green's Functions Theory in VOTCA-XTP**
Jens Wehner, Lothar Brombacher, Joshua Brown, Christoph Juhghans, Onur Çaylak, Yuriy Khalak, Pranav Madhikar, Gianluca Tirimbò, Björn Baumeier
Journal of Chemical Theory and Computation **14**, 6253 (2018)
- **Evolutionary approach to constructing a deep feedforward neural network for prediction of electronic coupling elements in molecular materials**
Onur Çaylak, Anil Yaman, Björn Baumeier
Journal of Chemical Theory and Computation **15**, 1777 (2019)
- **Wasserstein metric for improved quantum machine learning with adjacency matrix representations**
Onur Çaylak, O. Anatole von Lilienfeld, Björn Baumeier
Machine Learning: Science and Technology **1**, 03LT01 (2020)
- **A Kernel based Machine Learning Approach to Computing Quasi-particle Energies within Many-Body Green's Functions Theory**
Gianluca Tirimbo, Onur Çaylak, Björn Baumeier
NeurIPS, Machine Learning for Molecules (2020)

- **Glassy dynamics from generalized mode-coupling theory: existence and uniqueness of solutions for hierarchically coupled integro-differential equations**
Rutger A. Biezemans, Simone Ciarella, Onur Çaylak, Björn Baumeier, Lisbeth M.C. Janssen
Journal of Statistical Mechanics: Theory and Experiment **10** 103301 (2020)
- **Existence and uniqueness of solutions to the time-dependent Kohn-Sham equations coupled with classical nuclear dynamics**
Björn Baumeier, Onur Çaylak, Carlo Mercuri, Mark Peletier, Georg Prokert, Wouter Scharpach
arXiv preprint arXiv:2011.10542, 2020
- **Excited-state electronic structure of molecules using many-body Green's functions: Quasiparticles and electron–hole excitations with VOTCA-XTP**
Gianluca Tirimbò, Vivek Sundaram, Onur Çaylak, Wouter Scharpach, Javier Sijen, Christoph Junghans, Joshua Brown, F Zapata Ruiz, Nicolas Renaud, Jens Wehner, Björn Baumeier
Journal of Chemical Physics **152**, 114103 (2020)
- **Excited-state geometry optimization of small molecules with Many-Body Green's Functions Theory**
Onur Çaylak, Björn Baumeier
Journal of Chemical Theory and Computation **17**, 879 (2021)
- **Machine Learning of Quasiparticle Energies in Molecules and Clusters**
Onur Çaylak, Björn Baumeier
Journal of Chemical Theory and Computation **17**, 4891 (2021)

Meta-control of social learning strategies

Anil Yaman, Nicolas Bredeche, Onur Çaylak, Joel Z Leibo, Sang Wan Lee

PLOS Computational Biology, 18(2): e1009882 (2021)

Contents

Abstract	1
List of Publications	2
1 Introduction	8
2 Many-Electron Quantum Mechanics for Molecules	19
2.1 Many-electron wave-function approaches	21
2.1.1 Coupled Cluster Theory	23
2.1.2 Configuration interaction	24
2.1.3 Quantum Monte Carlo	25
2.2 Density-based Methods and Extensions	25
2.2.1 Density Functional Theory	26
2.2.2 Time dependent Density Functional Theory	28
2.3 Many-body Green's function theory	31
2.3.1 Implementation of <i>GW</i> -BSE using Gaussian-type orbitals	36
2.3.2 Frequency dependence of the self-energy	38
3 Excited-state geometry optimization of small molecules with Many-Body Green's Functions Theory	42
3.1 Introduction	44
3.2 Results	47
3.2.1 Carbon monoxide	47

3.2.2	Acetone	51
3.2.3	Influence of the ground state gradient	57
3.2.4	Acrolein	60
3.2.5	Methylenecyclopropene	63
3.2.6	Discussion	66
3.3	Summary	68
4	Wasserstein metric for improved quantum machine learning with adjacency matrix representations	70
4.1	Introduction	71
4.2	Method	73
4.3	Results and discussion	77
4.4	Conclusions	81
5	Δ-Machine Learning of Quasiparticle Energies in Molecules and Clusters	83
5.1	Introduction	84
5.2	Methodological Background	87
5.2.1	Δ-Machine Learning	87
5.3	Results	88
5.3.1	Orbital-sensitive descriptors	88
5.3.2	QM8 Molecules	91
5.3.3	Water Monomers and Dimers	97
5.3.4	Acetone in Water	99
5.4	Summary	104
6	An Evolutionary Approach to Constructing a Deep Feedforward Neural Network for Prediction of Electronic Coupling Elements in Molecular Materials	106
6.1	Introduction	107
6.2	Multiscale Ab initio Model	111

6.3 Machine Learning Model	113
6.3.1 Data representation	113
6.3.2 Deep Feedforward Neural Networks and Evolutionary Algorithms	115
6.3.3 Construction of a deep FFNN for prediction of electronic coupling elements	118
6.4 Results	120
6.5 Summary	126
7 Discussion & Outlook	128
Summary	136
Acknowledgement	139
About the author	140
References	140
Index	169

Chapter 1

Introduction

The behavior of electrons at their fundamental sub-nanometer scale governs the functionality of many natural and synthetic materials that shape our world. From a protein’s role in enabling life-sustaining processes [1], to a semiconductor’s ability to conduct charge [2–4], or a catalyst’s capacity to accelerate chemical reactions [5], knowledge about the electrons’ properties provides an understanding of how countless physical, chemical, and biological processes work on a fundamental level. With this understanding also come, ideally, strategies for the rational design of new, improved materials for specific application purposes: from increased efficiencies of photovoltaic systems, to higher energy storage capacity in batteries, or more efficient synthetic processes, to name only a few.

The search for such improved materials is often also referred to as optimizations in, or explorations of, *chemical compound space*, or chemical space [6]. The term refers to a multidimensional hypothetical space that encompasses all possible chemical compounds. These compounds are formed by various combinations of atoms, each with its own unique arrangement and connectivity. It is noteworthy that chemical compound space includes not only the known and documented compounds but also

the countless potential compounds that have not yet been synthesized or discovered. Consequently, the number of possible chemical compounds is astronomically large due to the various types of atoms, their different arrangements, and the potential for multiple levels of structural complexity. The estimated number of small organic molecules alone is in the order of 10^{60} to 10^{200} . Chemical compounds can vary in terms of their structural features, such as the types of atoms present, the way atoms are connected (bonding patterns), and the spatial arrangement of atoms. This diversity leads to a wide range of chemical properties and behaviors.

Ideally, exploring chemical compound space allows scientists to identify patterns and relationships between the structure of a compound and its properties. This insight is crucial for predicting the behavior of new compounds without having to synthesize and test them physically. To obtain this insight, both high-throughput synthesis and computational approaches, or a combination of both, play a role. This thesis will focus on several aspects related to computational approaches, in particular the calculation of electronic properties on the quantum level, also known as *Electronic Structure*.

Electronic Structure Theory is fundamentally nothing else than the application of quantum mechanics, as initially formulated in the early 20th century [7], to many-electron systems. However, already at an early stage it became apparent that the problem of having interacting electrons makes finding exact solutions to the electronic Schrödinger equation (more about this in Chapter 2) practically impossible.

Instead, many approximate methods have been devised and scrutinized, many of which are based on effective single-particle theories. In such approaches, it is assumed that the property of interest (either the

many-electron wavefunction or the electronic density, depending on the specifics) can be constructed with the help of a set of, in some sense, surrogate single-electron wavefunctions, which in turn can be obtained as solutions to a set of single-particle Schrödinger-like equations. One famous approach that builds the many-electron wavefunction in this way is the Hartree-Fock method [8] (see Section 2.1 for details). While this method laid the groundwork for modeling interacting electron systems, its inherent approximations caused limitations in the quality of predictions when compared to experimental references. With a view on exploring chemical compound space with computational methods, this lack of accuracy and predictive power is a major disadvantage. More sophisticated methods have been developed, which aim to build improved models upon the Hartree-Fock baseline (such as coupled-cluster theory or configuration interaction, see Chapter 2). While this class of methods generally improves the predictions of Hartree-Fock, they come at an increasingly larger computational cost, which make their large-scale applications, even with current computational resources, prohibitive.

The term *accuracy* has been used rather loosely up to now. To be more precise, what one generally aims for is the so-called *chemical accuracy*. Chemical accuracy is a term used in computational chemistry to describe the level of accuracy or precision that is required in the results of calculations to reliably predict or describe chemical phenomena. In more technical terms, achieving chemical accuracy typically means that the calculated properties of a system (such as energy, bond lengths, reaction barriers, etc.) are within a certain range of deviation from the corresponding experimental values. This range of deviation is usually on the order of a few kcal/mol for energy calculations or a few tenths of an angstrom for bond lengths and other structural properties. The concept of chemical accuracy is important because it indicates whether a com-

putational method is capable of providing results that are on par with or close to experimental data. When a computational method is said to have achieved chemical accuracy, it means that its predictions are accurate enough to be useful in guiding experimental work, designing new molecules, predicting reaction pathways, and making other important decisions in chemical research and development.

Taking a different view on the many-electron problem, Density Functional Theory (DFT) emerged as a revolutionary paradigm shift in the 1960s. Instead of working explicitly with the many-electron wavefunction, DFT focuses on the electron density distribution, significantly reducing computational complexity. The Hohenberg-Kohn theorems [9] provided a rigorous, and most notably *exact*, foundation, linking the electron density in the ground state to the external potential (e.g., from nuclei or other external sources). As will be described in more detail in Section 2.2.1, Kohn and Sham [10] introduced a system of effective single-particles to give some practicality to the statements of the theorems and to allow for the construction of the many-electron density from these single-electron wavefunctions. The resulting set of equations contain a so-called *exchange-correlation* term, which includes in simple terms all quantum-mechanical electron-electron interactions beyond the classical electrostatic ones. However, an exact expression for the exchange-correlation energy remained elusive, leading to the development of approximations.

Even though these approximate functionals are in general of varying quality depending on the actual system they are applied to, they enabled the study of larger systems, from molecules to solids. The success of DFT stems from its balance between accuracy and computational efficiency, and DFT has become the work horse of electronic structure calculations, for better or worse. Notoriously difficult are, e.g., materials

with strongly correlated electrons [11], long-range charge transfer [12], or weakly-bonded van der Waals systems [13]. In some cases, even the dissociation limits of some small diatomic molecules is predicted qualitatively wrong [14]. By construction, DFT generally does not offer insights beyond the ground state, so reliable predictions of electronic excitations, either charged excitations from the addition or removal of an electron to/from the system, or neutral excitations, e.g., after absorption of a photon, are outside its scope. The latter can in principle be addressed by using a time-dependent formulation of DFT. However, this TDDFT inherits some of the limitations that plagues ground state DFT from the use of approximate functionals [12]. From the perspective of exploring chemical compound space for the design of materials, it is questionable whether DFT and TDDFT on their own provide the required accuracy and reliability.

As a contrast, or rather alternative, to the above *first-principles* models, *data-driven* models or machine learning (ML) has recently found numerous applications in electronic structure calculations[15–23]. For instance, ML models can be trained to predict potential energy surfaces and interatomic force fields, enabling faster and more accurate molecular dynamics simulations. This reduces the computational cost of exploring complex chemical reactions and materials behavior. ML techniques can also correct for the limitations of DFT calculations. By learning from high-level quantum mechanical calculations, ML models can provide accurate corrections to DFT results (Δ -ML), improving the accuracy of electronic structure predictions. An extremely attractive application of ML models lies in the prediction of various molecular properties, such as dipole moments, polarizabilities, binding energies, or excitation energies solely from information about the chemical composition and a material’s geometric structure. This could prove especially useful for

screening large datasets of molecules for specific applications like drug discovery.

In all the cases above, ML models act as *surrogate models* for expensive calculations, reducing the need for repeated high-level quantum mechanical simulations and significantly speeding up the overall computational process. Such a process can not only involve the prediction of a single property of interest, for example, but also involve multiple steps as in multiscale or multiphysics approaches. For these approaches, it is particularly attractive to devise hybrid models that integrate physics-based first-principles with ML techniques. Done right, these models leverage the strengths of both approaches to improve predictive accuracy.

As with the first-principles models, there is a variety of different elements that constitute a machine learning model. Researchers need to make choices about the class of ML is most suitable for the problem at hand, and how to represent the data best, all in the search for data-efficient high-accuracy predictions. To be more specific, kernel-based machine learning and neural networks are two distinct approaches within the field of machine learning, each with its own characteristics, advantages, and limitations. In simple terms, kernel methods focus on transforming data into a higher-dimensional feature space to make it more amenable to linear separation or analysis. They rely on the notion of a kernel function, which measures similarity between data points in the original space or the transformed space. In contrast, neural networks consist of interconnected layers of artificial neurons (nodes). Each neuron applies an activation function to its weighted inputs and passes the result to the next layer. Neural networks can capture complex, non-linear relationships in data through their layered structure and learned weights.

As for the representation of data, a plethora of descriptors have been used specifically for machine learning of electronic structure with both kernel method and neural networks. Typically, these descriptors contain atomic features (atomic number, radii, effective charges, positions), information about the chemical environment (local chemical environment, coordination number), or bonding information (bond length, angles, torsions), in either very explicit ways or implicitly. While the details of the representation (other than uniqueness) are less crucial for artificial neural networks, the specific definition of how a chemical system is being specified is known to dramatically affect the learning efficiency of kernel based models. Namely, encoding the right physics, such as translational or atom-index invariance, in the representation results in systematic reduction of quantum data needs for achieving the same pre-defined predictive accuracy [24]. Especially, the lack of atom-index invariance has been a notorious problem for the class of adjacency matrix based descriptors together with the commonly used Euclidean norms in the kernel.

Against this larger background of open problems both in electronic structure calculations themselves and machine learning approaches for it, this thesis aims at developing and scrutinizing improved models in the individual areas, but most importantly in hybrid models. Particular (but not exclusive) attention is given to an electronic structure method, in which both charged and neutral electronic excitations are described within a framework based on many-body perturbation theory employing Green's functions, known as *GW*-BSE (see Section 2.3). This method is interesting because it provides single- and two-particle excitation energies within a rigorous theoretical setting, while still based on a DFT reference calculation. However, it comes at a significantly increased computational cost, albeit not as much as higher-order wavefunction

methods. This, in turn, makes it particularly interesting to investigate ways, how to use ML to accelerate not only plain *GW*-BSE calculations but also hybrid quantum-classical workflows, which include *GW*-BSE as the choice for the quantum method.

All in all, in this thesis problems of electronic structure calculations and machine learning will be tackled from four distinct angles:

1. From the **angle of electronic structure theory**: evaluation of the *GW*-BSE method for geometry optimization of small molecules in excited states.
2. From the **angle of machine learning theory**: exploration of the use of the Wasserstein metric as an alternative to standard metrics to resolve the problems arising from lack of atom-index invariance with adjacency matrix based representations in kernel methods.
3. From the **angle of hybrid *GW*-BSE/ML models**: development of a surrogate ML model for the most time-consuming step in *GW* calculations and its integration into full workflows, including quantum-classical embedding.
4. From the **angle of hybrid multiscale models**: development of a strategy to build a surrogate neural network model for inter-molecular electronic coupling elements in disordered material morphologies accounting for small details in mutual molecular orientations and relevant target property values distributed over several orders of magnitude.

In what follows, the first part of the thesis will provide a concise overview of the Theoretical Framework, including both the basic notions and governing equations for Many-Electron Quantum Mechanics for Molecules in Chapter 2. The first of the results chapters, Chapter 3,

contains a benchmark study of gas phase geometry optimizations in the excited states of carbon monoxide, acetone, acrolein, and methylenecyclopropene using GW -BSE employing numerical gradients. The influence of several typical approximations in the GW -BSE framework (see Section 2.3) is scrutinized: use of one-shot G_0W_0 or eigenvalue self-consistent $\text{ev}GW$, employing a fully-analytic approach or plasmon-pole model for the frequency dependence of the electron self-energy, or performing the BSE step within the Tamm–Dancoff approximation. The obtained geometries are compared to reference results from multireference perturbation theory (CASPT2), variational Monte Carlo (VMC), second-order approximate coupled cluster (CC2), and time-dependent density-functional theory (TDDFT).

The focus of Chapter 4 lies on studying the use of the Wasserstein metric to measure distances between molecules represented by the atom index dependent adjacency matrix representations, used in kernel ridge regression based supervised learning. Particular attention is being given to the fact that this metric provides a similarity measure that introduces index-invariance in such models. Systematic improvements are sought on training efficiency and predicting smoothness using the examples of both molecular distortions and a standard dataset containing tens of thousands organic molecules (QM9, [25]).

In Chapter 5, a Δ -Machine Learning approach is developed for the prediction of GW quasiparticle energies and photoelectron spectra of molecules and clusters, using orbital-sensitive graph-based representations in kernel ridge regression based supervised learning. To achieve this, adjacency matrix representations are made orbital-sensitive by augmenting them with atom-centered orbital charges and Kohn–Sham orbital energies, which are both readily available from baseline calculations on the level of DFT. First, effects of different constructions of

the orbital-sensitive representations (OSR) on the prediction of frontier orbital energies of 22K molecules of the QM8 dataset will be scrutinized, and the possibility to predict the full photoelectron spectrum of molecules within the dataset using a single model explored. Second, focus will be on the capability of the OSR-based Δ -ML model to capture effects of intra- and intermolecular conformations in application to water monomers and dimers. Finally, we consider the embedding of the ML approach into multiscale simulation workflows as a hybrid first-principles/ML model, using as prototypical example the prediction of solvatochromic shifts of quasiparticle and electron-hole excitation energies of solvated acetone employing a setup combining Molecular Dynamics, DFT, the GW approximation and the Bethe–Salpeter Equation.

In the final results chapter, Chapter 6, another hybrid first-principles and machine learning model is considered, aiming at the simulation of charge mobilities in disordered organic semiconductors. In the standard multiscale workflow, combinations of quantum and classical methods are employed to determine the physical quantities entering rates of electron transfer between two molecules, depending on mutual orientations, and mobility is finally determined from dynamics on a graph. One of the physical quantities entering this approach is the electronic coupling element, which needs to be computed from quantum calculations of individual monomers and the pair involved in the transfer process. Even on DFT level, doing this for around 20k pairs is a computational bottleneck, especially considering that in some cases much larger simulation sizes need to be considered. To remove this bottleneck and make much larger system sizes accessible, the development of a general framework for the construction of a deep feedforward neural network (FFNN) to predict distance and orientation dependent electronic coupling elements in disordered molecular materials is presented. Herein, a particular point of

attention is the automatized selection of an optimal architecture of the artificial neural network within a predefined search space with the help of an evolutionary algorithm. As will be shown, the fact that electronic coupling elements in disordered material morphologies are very sensitive to molecular orientations and relevant target property values are distributed over several orders of magnitude, makes simple minimizing the model error with stochastic gradient descent based backpropagation insufficient. Instead, additional guidance to the evolutionary algorithm is provided by simultaneous maximization of a model fitness that takes into account additional physical properties, such as the field-dependent carrier mobility. As a prototypical system, we consider hole transport in amorphous tris(8-hydroxyquinolinato)aluminum.

Finally, this thesis will be concluded by a Discussion & Outlook in Chapter 7.

Chapter 2

Many-Electron Quantum Mechanics for Molecules

Later sections of this chapter are in part based on the papers:

J. Wehner, L. Brombacher, J. Brown, C. Junghans, O. Çaylak, Y. Khalak, P. Madhikar, G. Tirimbò, B. Baumeier, Journal of Chemical Theory and Computation **14**, 6253 (2018)

and

G. Tirimbò, V. Sundaram, O. Çaylak, W. Scharpach, J. Sijen, C. Junghans, J. Brown, F. Zapata Ruiz, N. Renaud, J. Wehner, B. Baumeier, Journal of Chemical Physics **152**, 114103 (2020).

The quantum-mechanical treatment of structural and electronic properties of molecules and molecular complexes is based on the many-body Schrödinger equation. Analytical solutions to this equation can be found for only a few systems, highlighting the need for the use of suitable approximations combined with seeking their solutions numerically. Throughout this thesis, we will encounter different approximations to the full many-electron and many-nuclear problem, analyze some of them (in Chapter 3), and use their results as references in several Ma-

chine Learning settings (in Chapters 4–6). The purpose of this chapter is to give a concise overview about the various approximations to the full Schrödinger equation with the aim of understanding the respective strengths and weaknesses from a computational perspective.

In the following, we consider a molecular system consisting of M atoms and N electrons. In the absence of time-dependent external potentials and combined with the Born–Oppenheimer approximation [26], electronic and nuclear degrees of freedom are decoupled and the nuclei only affect the electronic system as a fixed external potential. The (non-relativistic) many-electron Hamiltonian then reads (in atomic units)

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 - \sum_{i=1}^N \sum_{\alpha=1}^M \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \sum_{i < j}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{\beta < \alpha}^M \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\beta - \mathbf{R}_\alpha|}, \quad (2.1)$$

comprising in order of appearance the kinetic energy operator of the electrons, the attractive electron-nuclei interaction (Z_α is the nuclear charge of atom α), the electron-electron repulsion, and finally the repulsion among the nuclei in the system. Throughout this thesis, electronic coordinates are labeled with lowercase ($\{\mathbf{r}_i\}$) while capital letters are used to denote nuclear coordinates ($\{\mathbf{R}_\alpha\}$), unless explicitly stated otherwise. Note that we will interchangeably use the Dirac (bra-ket) and real-space wave function notation for all the equations. In principle, knowing the Hamiltonian, the exact anti-symmetric wave function $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N; \{\mathbf{R}_\alpha\})$ of the many-electron system is the solution of the stationary Schrödinger equation

$$\hat{H} |\Psi\rangle = E |\Psi\rangle. \quad (2.2)$$

As mentioned above, analytic solutions are not available except for minimal cases, such as cases with $N = 1$. One might be tempted to seek

a numerical solution by discretization of the wave function on a spatial grid. However, even with an extremely coarse grid of only 20 grid points in each direction, 20^{3N} values are needed. For a 10 electron system, like water for instance, this naive approach would require more than 10^{39} values to be stored. As a consequence of the impracticability of such an approach, alternative methods have been pursued. In the remainder of this chapter, we consider briefly two different classes of approximations: methods that use the many-electron wave function (Section 2.1) and methods that are based on the much lower-dimensional electronic density (Section 2.2). For a more exhaustive perspective of these methods and numerical techniques, we refer the reader to Refs. [27, 28].

2.1 Many-electron wave-function approaches

The problem of solving (2.2) with the Hamiltonian (2.1) stem from the electron-electron interaction term. Without it, the electronic N -electron Hamiltonian can be written as a sum of commuting Hamiltonians of single-particles, and the N -electron wave function is simply a product of respective single-electron wave functions. However, such a product wave function does not take the Fermionic nature of the electrons into account, which imposes anti-symmetry of the wave function with respect to particle exchange, i.e., $\Psi(\dots, \mathbf{r}_i, \mathbf{r}_j, \dots) = -\Psi(\dots, \mathbf{r}_j, \mathbf{r}_i, \dots)$. This can be enforced by using a single Slater determinant of single particle

wave functions:

$$\Psi(\{\mathbf{r}_i\}) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) & \cdots & \phi_N(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) & \cdots & \phi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{r}_N) & \phi_2(\mathbf{r}_N) & \cdots & \phi_N(\mathbf{r}_N) \end{vmatrix}. \quad (2.3)$$

From the variational principle, it follows that the energy of a quantum system, evaluated with any normalized trial wave function $|\Psi_t\rangle$ will always approximate the true ground state energy E_0 from above, i.e., $E_0 \leq E[\Psi_t] = \langle \Psi_t | \hat{H} | \Psi_t \rangle$. It implies that one can use the Slater determinant (2.3) as the trial wave function and derive a set of N coupled equations for the N single particle orbitals from a minimization of $E[\Psi_t]$ under the constraint of normalization of the effective single-particle functions. This approach is known as the *Hartree–Fock* method and the minimization procedure yields equations for the associated orbitals $\{\phi_i^{\text{HF}}\}$ which read

$$\hat{H}^{\text{HF}} |\phi_i^{\text{HF}}\rangle = [\hat{H}_0 + \hat{\Sigma}_{\text{x}}] |\phi_i^{\text{HF}}\rangle = \epsilon_i^{\text{HF}} |\phi_i^{\text{HF}}\rangle. \quad (2.4)$$

Here, $\hat{H}_0 = \hat{T}_0 + \hat{V}_{\text{ext}} + \hat{V}_{\text{H}}$, with \hat{T}_0 being the kinetic energy, \hat{V}_{ext} an external potential, $\hat{V}_{\text{H}} = \int n_0(\mathbf{r}) v_{\text{C}}(\mathbf{r}, \mathbf{r}') d\mathbf{r}$ the classical Hartree potential with v_{C} the bare Coulomb interaction. The exchange operator $\hat{\Sigma}_{\text{x}}(\mathbf{r}, \mathbf{r}') = - \sum_i \frac{\phi_i^{\text{HF}}(\mathbf{r}) \phi_i^{\text{HF}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}$ defines the energy due to the antisymmetry of the total N -electron wave function, also called electron exchange. Using the $|\phi^{\text{HF}}\rangle$ to determine the total ground state energy, one has observed that the HF method yields total energies or derived properties with significant deviations from experimental data. This is attributed to the fact that the explicit electron-electron interaction in the N -electron Schrödinger equation has been recast as an effective (mean-field) potential acting on

single electrons, comprising a classical contribution (the Hartree term) and an explicitly quantum term (the exchange term). The latter is, however, not a complete representation of all quantum effects but merely of that resulting from the anti-symmetry condition. Whatever is missing beyond this exchange is also referred to as correlation.

A number of approaches, called post-Hartree–Fock methods, have been devised to include electron correlation to the multi-electron wave function and also allow for the construction of excited state wave functions, as will be summarized below.

2.1.1 Coupled Cluster Theory

Multi-electron wave functions that are based on the basic Hartree–Fock molecular orbital method but that include correlation effects can be obtained by an expansion of the kind

$$|\psi\rangle = \exp(T) |\Psi_0^{\text{HF}}\rangle, \quad (2.5)$$

where $T = \sum_{i=1} T_i$ the so called cluster operator and T_i the i -th excitation operator ($i = 1$ single excitation, $i = 2$ double excitation and so on). This *Coupled Cluster* (CC) formulation is an – at least in principle – exact solution to the time-independent Schrödinger equation Eq. (2.2). In practice, however, this choice turns the originally Hermitian problem into a non-Hermitian one, and as a consequence, the total energy is non-variational. In small systems, this problem can be overcome using variationally optimized coupled cluster [29–32]. From a pragmatic point of view, it is a general drawback of most coupled-cluster implementations is their computational cost. CC methods scale at best as $\mathcal{O}(N^p)$, where p is a relatively high power, e.g., $p = 7$ for coupled cluster including sin-

gle and double excitations (CCSD), with triples treated perturbatively (CCSD(T)). In spite of this, CC has been successfully used not only for ground state but also for excited state calculations [33–35].

2.1.2 Configuration interaction

Another post-Hartree-Fock variational method for quantum chemical multi-electron systems is known as *Configuration Interaction* (CI). The term *configuration* indicates that instead of single Slater determinants as in Hartree–Fock, so called configuration state functions (CSFs) $|\Psi_I\rangle$ are used for the expression of wave function. That is, the many-electron wave function $|\Psi\rangle$ is a linear combination of the form

$$|\Psi\rangle = \sum_{I=0} c_I |\Psi_I\rangle, \quad (2.6)$$

and the CSFs are built as Slater determinants from Hartree–Fock spin orbitals. If only one spin orbital differs, e.g., by replacing one occupied with one unoccupied orbital, such a CSF is a single excitation determinant. For two differing orbitals, this is a double excitation, and so on. In a *full CI* procedure, the expansion includes all possible CSFs and provides an exact solution to the electronic Schrödinger equation within the space spanned by the one-particle basis set. However, as in the case of CC, the method is limited in practice to relatively small systems due to long CPU time and large memory requirements.

2.1.3 Quantum Monte Carlo

The idea behind Quantum Monte Carlo (QMC) is based on the stochastic estimation of the energy of a trial wave function according to

$$E_{\text{trial}} = \frac{\int \psi_{\text{trial}}(\mathbf{r}) \hat{H} \psi_{\text{trial}}(\mathbf{r}) d^{3N} \mathbf{r}}{\int \psi_{\text{trial}}(\mathbf{r}) \psi_{\text{trial}}(\mathbf{r}) d^{3N} \mathbf{r}}, \quad (2.7)$$

with the minimal energy corresponding to the ground state energy, and the associated trial wave function the approximation to the many-electron ground state wave function. Refined methods exist (for the calculation of ground and also excited-states), such as Variational Monte Carlo (VMC) [36], in which a rigorous variational principle is applied to both ground and excited state wave functions. From a computational perspective, QMC methods typically scale with system-size as $\mathcal{O}(N^3)$, albeit with a large prefactor.

2.2 Density-based Methods and Extensions

In the previous section, we have seen that treating the many-electron Schrödinger equation with explicit $3N$ electronic variables is tremendously challenging, and often only possible for relatively small systems at substantial computational cost. Upon closer inspection of the electronic Hamiltonian in (2.1) one realizes that it comprises only one- and two-electron operators, that is, operators that act on either a single electronic coordinate or two. Does one then really need the N -electron wave function?

Suppose we have this N -electron wave function, we can define the second-order density matrix as

$$P_2(\mathbf{r}'_1, \mathbf{r}'_2; \mathbf{r}_1, \mathbf{r}_2) = \frac{N(N-1)}{2} \times \int \Phi^*(\mathbf{r}'_1, \mathbf{r}'_2, \dots, \mathbf{r}_N) \Phi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) d^3\mathbf{r}_3 \dots d^3\mathbf{r}_N. \quad (2.8)$$

Specifically, its diagonal elements $\bar{P}_2(\mathbf{r}_1, \mathbf{r}_2) = P_2(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1, \mathbf{r}_2)$ are the *two-particle density matrix*. From the definition in (2.9), one can also write a first-order density matrix in terms of P_2 as

$$P_1(\mathbf{r}'_1; \mathbf{r}_1) = \frac{2}{N-1} \int P_2(\mathbf{r}'_1, \mathbf{r}_2; \mathbf{r}_1, \mathbf{r}_2) d^3\mathbf{r}_2. \quad (2.9)$$

In this case, the diagonal element $\bar{P}_1(\mathbf{r}_1) = P_1(\mathbf{r}_1; \mathbf{r}_1)$ is the *charge density*. Instead of determining the total energy from as the expectation value of the electronic Hamiltonian with the full $3N$ -dimensional wave function ψ , one can obtain the same using the two-particle density matrix and the charge density

$$E = \int (\hat{T}_{\text{el}} + \hat{V}_{\text{nuc-el}}) \bar{P}_1(\mathbf{r}_1) d^3\mathbf{r}_1 + \int \hat{V}_{\text{el-el}} \bar{P}_2(\mathbf{r}_1, \mathbf{r}_2) d^3\mathbf{r}_1 d^3\mathbf{r}_2. \quad (2.10)$$

Note that this evaluation requires only information about a six-dimensional object.

2.2.1 Density Functional Theory

Hohenberg and Kohn took this even further when they realized for finding the ground-state energy \bar{P}_2 is not even required. Instead, it can be shown that the ground state is *completely* determined by the charge den-

sity $n(\mathbf{r}) = \bar{P}_1(\mathbf{r})$ alone. Two theorems relate the ground state to the electron density:

1. The density n_0 , which minimizes the ground-state energy, uniquely determines the external potential $\hat{V}_{\text{ext}}(\mathbf{r})$ acting on the electronic system. The ground state Ψ_0 is a one-to-one functional of the particle density $n(\mathbf{r})$.

(Note that in this theorem, external potentials are considered equivalent if they only differ an additional constant which is independent of space.)

2. The energy functional

$$E[n] = \int v_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + \langle \Psi | \hat{T} + \hat{V}_{\text{el-el}} | \Psi \rangle \quad (2.11)$$

obeys a variational principle with respect to the particle density $n(\mathbf{r})$ and is minimal for the ground-state density $n_0(\mathbf{r})$:

$$E_0 = E[n_0] \leq E[n]. \quad (2.12)$$

While theoretically fascinating, the application of the Hohenberg–Kohn theorems is not possible directly. A seemingly convenient strategy would be to minimize (2.11) via variation of the functional with respect to $n(\mathbf{r})$. However, as there is no general functional form of the kinetic energy $T[n]$, this is infeasible. Kohn and Sham [37] have suggested using a fictitious system of N single electrons instead of the interacting N electrons. These single electrons then interact only via an effective potential derived from the density. Reminiscent of the single-electron orbitals in Hartree–Fock theory, one can write the density as a sum of contributions from occupied single-electron functions (Kohn–Sham orbitals), i.e., $n(\mathbf{r}) = \sum_{i=1}^{\text{occ}} |\phi_i^{\text{KS}}(\mathbf{r})|^2$. Splitting off the classical Hartree interaction of a

charge density from the full electron-electron interaction, the remaining unknown quantum contribution and the differences in kinetic energies between the true and fictitious system are subsumed in the so called *exchange-correlation* energy functional $E_{\text{xc}}[n]$. This reformulation now allows to minimize the equivalent of (2.11) with respect to the single-particle orbitals $\phi_i^{\text{KS}}(\mathbf{r})$ under the constraint of orthonormality, leading to the Kohn–Sham (KS) equations

$$\hat{H}^{\text{KS}} |\phi_i^{\text{KS}}\rangle = [\hat{H}_0 + \hat{V}_{\text{xc}}[n]] |\phi_i^{\text{KS}}\rangle = \varepsilon_i^{\text{KS}} |\phi_i^{\text{KS}}\rangle. \quad (2.13)$$

Here, \hat{H}_0 is the same as in (2.4) and $\hat{V}_{\text{xc}}[n]$ is the exchange-correlation (xc) potential, formally containing all many-body effects exactly. However, as the explicit functional form is not known, approximate exchange-correlation functionals need to be employed. A thorough discussion of specific functional choices is beyond the scope of this work and refer the reader to Refs. [38, 39]. With the caveat of the varying quality of the different functional choices, DFT generally enables calculations of the ground-state density and total-energy derived properties (atomic structure, lattice parameters, total energy, phase stability, electronic density, elastic constants, and phonon frequencies). However, the KS eigenvalues ε^{KS} in (2.13) cannot be considered as *true* electronic excitations, as experimentally measurable in direct or inverse photoemission experiments. Also, because DFT is an effective independent particle theory, coupling of electrons and holes, e.g., after photoexcitation, is not accounted for.

2.2.2 Time dependent Density Functional Theory

Treating neutral excitations of the electronic ground state within the broader framework of DFT requires a time-dependent extension, called

time-dependent density functional theory (TD-DFT) [40]. A theorem by Runge and Gross relates the time-evolution of a many-electron system in the presence of a time-dependent external potential to the time-dependent electron density. With very similar arguments (though very different mathematical formalism) one can adapt the time-independent Kohn-Sham scheme introduced in Section 2.2.1, and similarly collect all exchange and correlation effects in the time-dependent exchange-correlation potential $V_{\text{xc}}(\mathbf{r}, t)$.

If one is interested in the excitation energies of the system, the full treatment of time-dependent Kohn–Sham equations is not necessary and one can instead use a linear response formulation of TD-DFT [41]. Key ingredient to it is the susceptibility $\chi(\mathbf{x}, \mathbf{x}')$ (with $\mathbf{x} = (\mathbf{r}, t)$ and $\mathbf{x}' = (\mathbf{r}', t')$ space-time variables), which is defined as the linear response kernel between the variations of the electron density $\delta n(\mathbf{x}) = n(\mathbf{x}) - n_0(\mathbf{r})$ with respect to a local external perturbation $U(\mathbf{x}')$:

$$\delta n(\mathbf{x}) = \int \chi(\mathbf{x}; \mathbf{x}') U(\mathbf{x}') d\mathbf{x}'. \quad (2.14)$$

One can relate this susceptibility to the susceptibility of independent particles χ_0 (a response function whose poles are the KS energies) via a Dyson-like equation

$$\begin{aligned} \chi(\mathbf{r}, \mathbf{r}', \omega)^{-1} &= \chi_0(\mathbf{r}, \mathbf{r}'; \omega)^{-1} + K^{\text{DFT}}(\mathbf{r}, \mathbf{r}', \omega) \\ &= \chi_0(\mathbf{r}, \mathbf{r}'; \omega)^{-1} + f_{\text{Hxc}}(\mathbf{r}, \mathbf{r}'; \omega), \end{aligned} \quad (2.15)$$

with K^{DFT} the so-called TD-DFT kernel, expressed as $f_{\text{Hxc}} = v_c(\mathbf{r}, \mathbf{r}') + f_{\text{xc}}(\mathbf{r}, \mathbf{r}', \omega)$ with f_{xc} the exchange-correlation kernel. The susceptibility $\chi(\mathbf{r}, \mathbf{r}', \omega)$ has poles (and equivalently $\chi(\mathbf{r}, \mathbf{r}', \omega)^{-1}$ has zeroes) at the sought excitation energies Ω_S .

Casida showed that if the kernels are frequency-independent, one can use the ground state KS orbitals to form a transition product basis from occupied v to unoccupied c states to express the actual transitions with Ω_S [42] by solving the matrix equations

$$\begin{pmatrix} \underline{\mathbf{H}}^{\text{res}} & \underline{\mathbf{K}} \\ -\underline{\mathbf{K}} & -\underline{\mathbf{H}}^{\text{res}} \end{pmatrix} \begin{pmatrix} \mathbf{A}^S \\ \mathbf{B}^S \end{pmatrix} = \Omega_S \begin{pmatrix} \mathbf{A}^S \\ \mathbf{B}^S \end{pmatrix} \quad (2.16)$$

Here, $\underline{\mathbf{H}}^{\text{res}}$ and $\underline{\mathbf{K}}$ are matrices on their own in the vc -transition basis, with

$$\begin{aligned} H_{vc,v'c'}^{\text{res}} &= H_{vc,v'c'}^D + K_{vc,v'c'}^X + K_{vc,v'c'}^{\text{xc}}, \\ K_{cv,v'c'} &= K_{cv,v'c'}^X + K_{cv,v'c'}^{\text{xc}}, \end{aligned} \quad (2.17)$$

and

$$\begin{aligned} H_{vc,v'c'}^D &= (\epsilon_c - \epsilon_v) \delta_{vv'} \delta_{cc'}, \\ K_{vc,v'c'}^X &= \int \phi_c^*(\mathbf{r}) \phi_v(\mathbf{r}) v_C(\mathbf{r}, \mathbf{r}') \phi_{c'}(\mathbf{r}') \phi_{v'}^*(\mathbf{r}') d^3\mathbf{r} d^3\mathbf{r}', \\ K_{vc,v'c'}^{\text{xc}} &= \int \phi_c^*(\mathbf{r}) \phi_v(\mathbf{r}) f_{\text{xc}}(\mathbf{r}, \mathbf{r}') \phi_{c'}(\mathbf{r}') \phi_{v'}^*(\mathbf{r}') d^3\mathbf{r} d^3\mathbf{r}'. \end{aligned} \quad (2.18)$$

Note that in (2.16), excitations with $\Omega_S > 0$ are paired with de-excitations $\Omega_S < 0$. In the Tamm–Dancoff approximation (TDA), contributions arising from de-excitations are neglected by setting the offdiagonal terms $\underline{\mathbf{K}} = 0$. Thereby, the generalized non-Hermitian eigenvalue problem in (2.16) transforms into a standard, Hermitian eigenvalue problem

$$\underline{\mathbf{H}}^{\text{res}} \mathbf{A}_S^{\text{TDA}} = \Omega_S^{\text{TDA}} \mathbf{A}_S^{\text{TDA}}. \quad (2.19)$$

As in the case for the ground state calculations, approximations are needed to perform numerical calculations as the exact time-dependent exchange-correlation action functional (from which V_{xc} and f_{xc} are derived) is not known and results from TD-DFT calculations show sensitivity on the choice of the exchange-correlation functional.

2.3 Many-body Green's function theory

As described in the previous section, TD-DFT provides access to neutral electronic excitations, with the drawback of a dependence on the quality of the used exchange-correlation functional approximation. Excitations in which the number of electrons in the system is changed, i.e., the addition ($N \rightarrow N + 1$) or removal ($N \rightarrow N - 1$) of a single electron to/from the system, are not described. Recall from Section 2.2.1 that the Kohn–Sham eigenvalues are formally introduced only as Lagrange parameters for ensuring the orthonormality of the single-particle wave functions. In practice, the interpretation of the ϵ^{KS} as excitation energies works, at best, qualitatively. The presence of spurious electronic self-interaction inherent in the use of all approximate exchange-correlation functionals, for instance, manifests itself in significant underestimation of the energies compared to experiment, and it is not unusual to encounter errors of 50% in calculated HOMO-LUMO gaps in molecules (or the fundamental band gap in solids).

An alternative approach to treat electronic excitations, both charged and neutral, with respect to the reference ground state exploits the concept of Green's functions in a many-body setting. In simple terms, solutions to the interacting electron system are related to those of the (solvable) non-interacting one. In this framework, electron addition or

removal describes the process involving not only the excited electrons but also the collective response of the other electrons, also referred to as the excitation of a quasiparticle.

The energies of these quasiparticles are given the poles of the interacting one-electron Green's function and can be obtained from the Dyson equation [43–46]

$$\left[\hat{H}_0 + \hat{\Sigma}(\varepsilon_i^{\text{QP}}) \right] |\phi_i^{\text{QP}}\rangle = \varepsilon_i^{\text{QP}} |\phi_i^{\text{QP}}\rangle , \quad (2.20)$$

where $|\phi_i^{\text{QP}}\rangle$ are the quasiparticle wavefunctions. The operator $\hat{\Sigma}(\cdot)$ is the self-energy operator, which describes the exchange-correlation many-body effects and in loose terms replaces the exchange-correlation potential in the Kohn–Sham equations (2.13). Within the GW approximation, the self-energy be expressed as a convolution of the one-particle Green's function $G(\mathbf{r}, \mathbf{r}', \omega)$ with the screened Coulomb interaction

$$W(\mathbf{r}, \mathbf{r}', \omega) = \epsilon^{-1}(\mathbf{r}, \mathbf{r}', \omega) v_c(\mathbf{r}, \mathbf{r}') = \frac{\epsilon^{-1}(\mathbf{r}, \mathbf{r}', \omega)}{|\mathbf{r} - \mathbf{r}'|}, \quad (2.21)$$

where ϵ^{-1} is the inverse dielectric function calculated in the random-phase approximation [47]. The non-local, energy-dependent self-energy operator can thenbe explicitly written as

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int d\omega' G(\mathbf{r}, \mathbf{r}', \omega + \omega') W(\mathbf{r}, \mathbf{r}', \omega'). \quad (2.22)$$

The self-energy $\Sigma = iG\bar{W}$ can be separated into its bare exchange part $\Sigma_x = iGv_c$ and its correlation part $\Sigma_c = iG\widetilde{W}$, where $\widetilde{W} = W - v_c$. With that, the frequency integral in 2.22, can be evaluated fully analytically by calculating the reducible polarizability in terms of an eigenvalue decomposition of the random phase approximation (RPA) Hamiltonian. See, e.g., Refs. [48–51] for details. While being exact, the scaling of the

diagonalization of the RPA Hamiltonian makes the application of this approach infeasible for large systems. In its place, the frequency integration can be approximated within a generalized plasmon-pole model (PPM) [52, 53].

Instead of attempting a direct solution of (2.20), quasiparticle wavefunctions $|\phi_i^{\text{QP}}\rangle$ are conventionally expanded in terms of the Kohn-Sham wavefunctions of a preceding DFT calculation. Together with simultaneous addition and subtraction of V_{xc} in the quasiparticle Hamiltonian on the left-hand side of (2.20), this allows to write

$$H_{ij}^{\text{QP}}(E) = \epsilon_i^{\text{KS}} \delta_{ij} + \langle \phi_i^{\text{KS}} | \hat{\Sigma}(E) - \hat{V}_{\text{xc}} | \phi_j^{\text{KS}} \rangle. \quad (2.23)$$

If one further approximates the expansion by setting $|\phi_i^{\text{QP}}\rangle = |\phi_i^{\text{KS}}\rangle$, one can obtain the quasiparticle energies perturbatively as the diagonals of (2.23)

$$\begin{aligned} \epsilon_i^{\text{QP}} &= \epsilon_i^{\text{KS}} + \Delta\epsilon_i^{\text{GW}} \\ &= \epsilon_i^{\text{KS}} + \langle \phi_i^{\text{KS}} | \hat{\Sigma}(\epsilon_i^{\text{QP}}) - \hat{V}_{\text{xc}} | \phi_i^{\text{KS}} \rangle. \end{aligned} \quad (2.24)$$

From (2.24) one can see that $\Delta\epsilon_i^{\text{GW}}$ itself depends on ϵ_i^{QP} , requiring a fixed-point procedure for finding a solution. Also, as the ϵ_i^{QP} enter both the energy-dependent microscopic dielectric function determined within the RPA as well as in the Green's function in the expression for the self-energy Eq. 2.22, different variants of the GW approximation can be defined, depending on if G or W are updated and if so, how often. Performing no update is known as a *one-shot* G_0W_0 calculation. In the so called ev GW approach, the quasiparticle energies are updated in the evaluation of both G and W forming Σ until self-consistency is reached. The final obtained quasiparticle energies then can be seen as the true electron addition or removal energies, respectively.

Neutral electronic excitations, such as the ones obtainable by the linear response formulation of TD-DFT described above, can be described within Many-Body Green's Functions theory. On top of the quasiparticle picture just discussed, such excitations can be seen as the formation coupled electron-hole pairs, described using a product basis of QP wave functions, i.e.,

$$\chi_S(\mathbf{r}_e, \mathbf{r}_h) = \sum_v^{\text{occ}} \sum_c^{\text{unocc}} \sum_{\sigma\sigma'} A_{vc,\sigma\sigma'}^S \phi_{c,\sigma'}(\mathbf{r}_e) \phi_{v,\sigma}^*(\mathbf{r}_h) + B_{vc,\sigma\sigma'}^S \phi_{v,\sigma'}(\mathbf{r}_e) \phi_{c,\sigma}^*(\mathbf{r}_h), \quad (2.25)$$

where \mathbf{r}_e (\mathbf{r}_h) is for the electron (hole) coordinate, and we drop the label QP for clarity. The $A_{vc,\sigma\sigma'}$ ($B_{vc,\sigma\sigma'}$) are expansion coefficients of the excited state wave function in terms of resonant (anti-resonant) transitions between QP occupied (occ.) states v and unoccupied (unocc.) c with spin σ and σ' , respectively. They can be obtained as solutions of the Bethe–Salpeter Equation (BSE) in the form of an effective two-particle Hamiltonian problem

$$\begin{pmatrix} \underline{\mathbf{H}}^{\text{res}} & \underline{\mathbf{M}} \\ -\underline{\mathbf{M}} & -\underline{\mathbf{H}}^{\text{res}} \end{pmatrix} \begin{pmatrix} \mathbf{A}^S \\ \mathbf{B}^S \end{pmatrix} = \Omega_S \begin{pmatrix} \mathbf{A}^S \\ \mathbf{B}^S \end{pmatrix}. \quad (2.26)$$

Note the structural similarity of the Bethe–Salpeter equation to the Casida formulation of linear-response TD-DFT (2.16). We will see below that there are, however, conceptual difference in how the elements of the matrix blocks are determined. In case of negligible spin-orbit coupling, it can be shown that the BSE Hamiltonian (as well as the one in (2.16)) has block structure in terms of the spin combination of the electron and hole states [54]. As a consequence, (2.26) can be transformed into two independent Hamiltonians for singlet and triplet excitations, respectively, and the explicit spin variables are not needed. Specifically, the corre-

sponding matrix elements of $\underline{\mathbf{H}}^{\text{res}}$ and $\underline{\mathbf{M}}$ are

$$\underline{H}_{vc,v'c'}^{\text{res}} = D_{vc,v'c'} + \kappa M_{vc,v'c'}^x + M_{vc,v'c'}^d \quad (2.27)$$

$$\underline{M}_{cv,v'c'} = \kappa M_{cv,v'c'}^x + M_{cv,v'c'}^d, \quad (2.28)$$

where $\kappa = 2$ (0) for spin singlet (triplet) excitations, and

$$D_{vc,v'c'} = (\epsilon_c - \epsilon_v)\delta_{vv'}\delta_{cc'}, \quad (2.29)$$

$$\begin{aligned} M_{vc,v'c'}^d = & - \int d^3\mathbf{r}_e d^3\mathbf{r}_h \phi_c^*(\mathbf{r}_e) \phi_{c'}(\mathbf{r}_e) W(\mathbf{r}_e, \mathbf{r}_h, \omega = 0) \\ & \times \phi_v(\mathbf{r}_h) \phi_{v'}^*(\mathbf{r}_h), \end{aligned} \quad (2.30)$$

$$\begin{aligned} M_{vc,v'c'}^x = & \int d^3\mathbf{r}_e d^3\mathbf{r}_h \phi_c^*(\mathbf{r}_e) \phi_v(\mathbf{r}_e) v_c(\mathbf{r}_e, \mathbf{r}_h) \\ & \times \phi_{c'}(\mathbf{r}_h) \phi_{v'}^*(\mathbf{r}_h). \end{aligned} \quad (2.31)$$

The terms labeled with D arise from free interlevel transition between occupied and empty quasiparticle states, the *direct interaction* M^d (Eq. 2.30) is responsible for the binding of the electron-hole pair and is based on the attractive, but screened, interaction W (in the static approximation $\omega = 0$) between electron and hole. The repulsive *exchange interaction* M^x (Eq. 2.31) is responsible for the singlet-triplet splitting.

In systems for which the elements of the off-diagonal blocks $\underline{\mathbf{K}}$ in Eq. (2.26) are negligible, it is legitimate to use the Tamm–Dancoff Approximation (cf., (2.19)) for the BSE [55], in which the electron-hole amplitude is formed by resonant transitions from occupied v to unoccupied c states only:

$$\chi_S^{\text{TDA}}(\mathbf{r}_e, \mathbf{r}_h) = \sum_v^{\text{occ}} \sum_c^{\text{unocc}} A_{\text{TDA}, vc}^S \phi_c(\mathbf{r}_e) \phi_v^*(\mathbf{r}_h). \quad (2.32)$$

The effective Hamiltonian reduces to the upper diagonal block of Eq. (2.26):

$$\underline{\mathbf{H}}^{\text{res}} \vec{A}_{\text{TDA}}^S = \Omega_S^{\text{TDA}} \vec{A}_{\text{TDA}}^S. \quad (2.33)$$

The TDA is known to reduce triplet instabilities [56, 57]. On the other hand, if the coupling between resonant and anti-resonant parts is significant, and its neglect can cause deviations of several 0.1 eV from results obtained with the full approach [58], in particular for small molecules.

2.3.1 Implementation of GW -BSE using Gaussian-type orbitals

As the GW -BSE method is not as commonly implemented and available in software packages compared to, e.g., DFT, we close this chapter with a few details on its numerical implementation in the VOTCA-XTP library.

VOTCA-XTP uses Gaussian-type orbitals (GTOs)

$$\varphi_\alpha(\mathbf{r}) = \varphi_{\{\ell, m_\ell, \gamma, I\}}(\mathbf{r}) = N_{\ell\gamma} Y_{\ell m_\ell} |\mathbf{r} - \mathbf{R}_I|^l \exp(-\gamma |\mathbf{r} - \mathbf{R}_I|^2) \quad (2.34)$$

with decay constant γ centered around atom position \mathbf{R}_I , where $Y_{\ell m_\ell}$ are spherical harmonics with angular momentum number ℓ and magnetic quantum numbers m_ℓ , and $N_{\ell\gamma}$ is a normalization constant, to expand the one- and two-point quantities involved in both DFT and GW -BSE steps. Specifically, the KS states in Eq. (2.13) are expressed using these basis functions in real space as

$$\phi_i^{\text{KS}}(\mathbf{r}) = \sum_\alpha c_\alpha^i \varphi_\alpha(\mathbf{r}) \quad (2.35)$$

turning Eqs. 2.13, 2.23, and 2.26 into (generalized) eigenvalue problems in matrix form. VOTCA-XTP evaluates the integrals of the respective operators over the Gaussian basis functions using the modified re-

cursive algorithms by Obara and Saika [59, 60] for contracted Gaussian basis functions with $l \leq 4$.

Of particular importance is the computation of 4-center repulsion integrals over the GTOs

$$(\alpha\beta|\alpha'\beta') = \iint d^3\mathbf{r} d^3\mathbf{r}' \frac{\varphi_\alpha(\mathbf{r})\varphi_\beta(\mathbf{r})\varphi_{\alpha'}(\mathbf{r}')\varphi_{\beta'}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (2.36)$$

which scales with N_b^4 (with N_b the number of basis functions) and occurs in the KS Hamiltonian term \hat{V}_H and in the self-energy. The set of N_b^2 unique product functions $\varphi_\alpha(\mathbf{r})\varphi_\beta(\mathbf{r})$ can be approximated by a smaller auxiliary basis containing only $N_{\text{aux}} = 3N_b$ to $5N_b$ functions ξ_μ . This reduces the scaling from N_b^4 to N_b^3 by rewriting the 4-center integrals as a combination of 3-center and 2-center repulsion integrals [61]:

$$(\alpha\beta|\alpha'\beta') \approx \sum_{\mu,\nu} (\alpha\beta|\mu)(\mu|\nu)^{-1}(\nu|\alpha'\beta'), \quad (2.37)$$

where $(\mu|\nu)^{-1}$ is an element of the inverse of the 2-center repulsion matrix

$$(\mu|\nu) = \iint d^3\mathbf{r} d^3\mathbf{r}' \xi_\mu(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \xi_\nu(\mathbf{r}') \quad (2.38)$$

and $(\alpha\beta|\mu)$ is an element of the 3-center repulsion tensor

$$(\alpha\beta|\mu) = \iint d^3\mathbf{r} d^3\mathbf{r}' \varphi_\alpha(\mathbf{r})\varphi_\beta(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \xi_\mu(\mathbf{r}'). \quad (2.39)$$

The expression in Eq. (2.37) appears formally as the insertion of a resolution-of-identity (RI) with metric $(\nu|\mu)^{-1}$.

Within the RI approximation, the elements of the QP Hamiltonian in the basis of KS states contain $\Sigma_{mn}(E) = \langle \phi_m^{\text{KS}} | \hat{\Sigma}(E) | \phi_n^{\text{KS}} \rangle$ (Eqs. 2.23

and 2.24), which are determined as

$$\Sigma_{mn}(E) = \sum_{\mu,\nu} \sum_l I_\mu^{ml} I_\nu^{nl} \frac{i}{2\pi} \int d\omega \frac{e^{i\omega\theta} \epsilon_{\mu\nu}^{-1}(\omega)}{E + \omega - \epsilon_l \pm i\eta}, \quad (2.40)$$

where the factor with $\theta \rightarrow 0^+$ ensures convergence of the integral, and the imaginary perturbations $\pm\eta$ avoid singularities on the real axis, where the plus (minus) is taken when l is occupied (unoccupied). Further,

$$I_\mu^{ml} = \sum_v (\mu|v)^{-1/2} \sum_{\alpha,\beta} c_\alpha^m c_\beta^l (\alpha\beta|v) = \sum_v (\mu|v)^{-1/2} M_v^{ml} \quad (2.41)$$

and

$$\epsilon_{\mu\nu}(\omega) = \delta_{\mu\nu} - 2 \sum_m^{\text{occ}} \sum_l^{\text{unocc}} I_\mu^{ml} I_\nu^{ml} \left[\frac{1}{\omega - (\epsilon_m - \epsilon_l) + 2i\eta} - \frac{1}{\omega + (\epsilon_m - \epsilon_l) - 2i\eta} \right]. \quad (2.42)$$

is called the dielectric matrix. In the G_0W_0 approach, we take the KS energies $\epsilon_i = \epsilon_i^{\text{KS}}$, whereas in the $\text{ev}GW$ approach, we take the QP energies $\epsilon_i = \epsilon_i^{\text{QP}}$. Currently, VOTCA-XTP pre-calculates all integrals at the start of the calculation and keeps I_μ^{ml} in memory.

2.3.2 Frequency dependence of the self-energy

The frequency integration, in Eq. (2.22), is one of the major difficulties in a GW calculation. Although it is possible to perform a numerical integration, this is likely unstable, since the integrand needs to be evaluated in regions in which it is ill-behaved. VOTCA-XTP offers different alternatives for an approximate or exact integration, summarized in the following. We present three methods: one that is exact yet takes too

much computational effort, one that is exact and is reduced in scaling, and one that is approximate yet very well-scaled.

In the GW approach, it is customary to separate the self-energy $\Sigma = iG\mathcal{W}$ into its bare exchange part $\Sigma_x = iGv_c$ and its correlation part $\Sigma_c = iG\widetilde{\mathcal{W}}$, where $\widetilde{\mathcal{W}} = \mathcal{W} - v_c$. In the following, several approaches to determine the correlation part of Eq. (2.40) are presented.

Fully Analytical Approach (FAA)

The integral in Eq. (2.40) can be evaluated analytically, yielding an exact expression of the correlation part of the self-energy. Its evaluation requires the calculation of the reducible polarizability \hat{P} . We can express it in terms of an eigenvalue decomposition of the RPA Hamiltonian \hat{H}^{RPA}

$$\hat{P}(\omega) = [\hat{H}^{\text{RPA}} - \omega]^{-1} = \sum_S \frac{|\chi_S\rangle\langle\chi_S|}{\Omega_S - \omega}, \quad (2.43)$$

where the RPA Hamiltonian obeys a BSE as in Eq. (2.26) with $\kappa = -1$ and $K^d = 0$ in Eqs. 2.27 and 2.28.

We can apply analytic continuation to the complex plane and contour deformation techniques to the convolution Eq. (2.22) [62, 63]. The resulting matrix entries of the correlation part of the self-energy are given by

$$\Sigma_{c,mn}(E) = 2 \sum_{l,S} \frac{R_{ml}^S R_{nl}^S}{E - \epsilon_l \pm (\Omega_S - i\eta)}, \quad (2.44)$$

where \pm denotes $+$ ($-$) for l occupied (unoccupied), and the factor 2 accounts for spin degeneracy. The residues R_{mn}^S are calculated as

$$R_{mn}^S = \sum_{\mu,\nu} \sum_v^{\text{occ}} \sum_c^{\text{unocc}} I_\mu^{mn} I_\nu^{vc} (A_{vc}^S + B_{vc}^S). \quad (2.45)$$

While this approach is analytically exact, it is not feasible for large systems as the diagonalization of \hat{H}^{RPA} scales as N^6 in computational effort and N^4 in memory required.[64]

Plasmon-Pole Model (PPM)

Instead of a formally exact treatment of the frequency dependence, it can be approximated within a generalized plasmon-pole model (PPM) [65, 66]. The dielectric matrix can be expressed in terms of its eigenvalues $\lambda_{\mu'}$ and eigenvectors $\Phi_{\mu'}$ as

$$\epsilon_{\mu\nu}(\omega) = \sum_{\mu'} \Phi_{\mu'}^\mu(\omega) \lambda_{\mu'}(\omega) \Phi_{\mu'}^\nu(\omega), \quad (2.46)$$

In the PPM, eigenvectors are assumed to be frequency-independent, so only the eigenvalues $\lambda_{\mu'}$ depend on ω . In particular, this approximate dependence reads

$$\lambda_{\mu'}^{-1}(\omega) \approx 1 + \frac{z_{\mu'} \omega_{\mu'}}{2} \left[\frac{1}{\omega - (\omega_{\mu'} - i\eta)} - \frac{1}{\omega + (\omega_{\mu'} - i\eta)} \right]. \quad (2.47)$$

Here, $z_{\mu'}$ denotes the plasmon-pole weight and $\omega_{\mu'}$ denotes the plasmon-pole frequency. These two model parameters are found by fitting the plasmon-pole model to the exact dielectric function [67], as shown in Eq. (2.42), for the frequencies $\omega = 0$ and $\omega = iE_0$, with E_0 an additional model parameter, typically $E_0 = 0.5 \text{ Ha}$. The correlation part of the self-

energy results from the second term of Eq. (2.47), and its matrix entries are obtained as

$$\Sigma_{c,mn}(E) = 2 \sum_{l,\mu'} \frac{1}{4} \frac{z_{\mu'} \omega_{\mu'} I_{\mu'}^{ml} I_{\mu'}^{nl}}{E - \varepsilon_l \pm \omega_{\mu'}}, \quad (2.48)$$

where \pm denotes $+$ ($-$) for l occupied (unoccupied), and the factor 2 accounts for spin.

Chapter 3

Excited-state geometry optimization of small molecules with Many-Body Green's Functions Theory

This chapter is based on the paper: O. Çaylak and B. Baumeier, J. Chem. Theory Comput. **17**, 879 (2021).

In this thesis, which focuses on computational methods in quantum chemistry with an emphasis on electronic structure calculations, traditional methodologies in the sense of first-principles models form an essential foundation. Particularly, before the application of machine learning-based strategies, it is essential to understand such techniques, especially when they involve tasks that are computationally intense, such as computing numerical gradients needed, e.g., for use in optimization problems. Given this, this chapter is crafted with two main objectives. First, we aim to provide an in-depth exploration of the optimization of molecular geometries with *GW*-BSE, which also serves as a benchmark study, showing both inherent challenges and accuracies in various scenarios. We will highlight the intricacies of calculations, em-

phasizing the computational demand and its implications. Second, this chapter serves as a precursor, laying down the groundwork for the subsequent sections where machine learning emerges as an alternative to traditional methods.

We present a benchmark study of gas phase geometry optimizations in the excited states of carbon monoxide, acetone, acrolein, and methylenecyclopropene using many-body Green's functions theory within the *GW* approximation and the Bethe–Salpeter Equation (BSE) employing numerical gradients. We scrutinize the influence of several typical approximations in the *GW*-BSE framework: using of one-shot G_0W_0 or eigenvalue self-consistent ev*GW*, employing a fully-analytic approach or plasmon-pole model for the frequency dependence of the electron self-energy, or performing the BSE step within the Tamm–Dancoff approximation. The obtained geometries are compared to reference results from multireference perturbation theory (CASPT2), variational Monte Carlo (VMC), second-order approximate coupled cluster (CC2), and time-dependent density-functional theory (TDDFT). We find overall a good agreement of the structural parameters optimized with the *GW*-BSE calculations with CASPT2, with an average relative error of around 1 % for the G_0W_0 and 1.5 % for the ev*GW* variants based on a PBE0 ground state, respectively, while the other approximations have negligible influence. The relative errors are also smaller than those for CC2 and TDDFT with different functionals and only larger than VMC, indicating that the *GW*-BSE method does not only yield excitation energies but also geometries in good agreement with established higher-order wavefunction methods.

3.1 Introduction

Electronically excited molecules play a pivotal role in a wide range of processes and applications, e.g., from photosynthesis to light-conversion processes in organic opto-electronic devices [68–70], as probes in fluorescent spectroscopy [71], or as intermediates in catalytic reactions [5]. From a computational perspective, excited state properties are conventionally studied using either post-Hartree Fock methods, such as different version of Configuration Interaction or Coupled Cluster (CC) approaches, or time-dependent formulations of density-functional theory (TDDFT). However, the accuracy of these methods can sensitively depend on both the type of excitations (e.g., local vs. charge-transfer type) studied, and the details of the level of theory used, such as order of allowed excitations in CC or the functional in TDDFT [72].

Inspired by its successes in the solid-state community, the use of many-body Green’s functions theory [64, 73–75] in the *GW* approximation with the Bethe–Salpeter Equation (BSE) has attracted increasing attention for the calculation of electronic excitation in molecular systems in past years. [63, 76–86]. Starting from an N -electron (DFT) ground state, quasiparticle excitations for the removal and addition of an electron are calculated first in the *GW* step. Subsequently, the BSE is used to describe neutral excitations as coupled electron-hole pairs based on quasi-particle product functions [75, 87]. This formulation is similar to linear-response TDDFT in the Casida formulation [88] with comparable computational cost, and can be used to predict excitation energies for different types of excitations with good accuracy [78, 79].

Focus of the studies based on *GW*-BSE have nearly exclusively focused on excitation energies based on ground-state geometries, i.e., absorption energies. Only a few efforts have been directed at investiga-

tions of geometric relaxations after excitation and/or emission energies. The total energy of the excited state S depending on nuclear coordinates \mathbf{R} is given as $E_S(\mathbf{R}) = E_0(\mathbf{R}) + \Omega_S(\mathbf{R})$, where E_0 is the ground state energy and Ω_S the excitation energy. Taking its gradient with respect to the nuclear coordinates, $\partial_{\mathbf{R}} E_S$, shows that it is a sum of the standard gradient of the ground state energy and the gradient of the excitation energy, $\partial_{\mathbf{R}} \Omega_S$. Ismali-Beigel and Louie [89] discussed that employing the Hellman-Feynman theorem, the latter contribution can be evaluated as

$$\partial_{\mathbf{R}} \Omega_S = \langle S | \partial_{\mathbf{R}} \hat{H}^{\text{BSE}} | S \rangle, \quad (3.1)$$

where \hat{H}^{BSE} is the two-particle Hamiltonian of the BSE for the electron-hole pair (see Chapter 2.3). This Hamiltonian contains the contributions from the single quasi-particles as given by the *GW* approximation and their interactions. The central quantity in *GW*-BSE is the non-local, energy-dependent electron self-energy operator [90], which contains many-body effects, or electron correlation. It is given as the convolution in the frequency domain of the one-electron Green's function G and a screened Coulomb interaction W typically evaluated using the Random-Phase Approximation (RPA). Due to its form no exact analytic expressions are available for the evaluation of Eq. (3.1), and one has in practice to rely on numerical gradients instead. This complication is further exacerbated by the fact that different approaches can be used both on the *GW* and BSE levels: the *GW* convolution can be performed with exact fully-analytic (FAA) methods [63, 64, 86] or simplifications, such as a Plasmon-Pole Model [90] (PPM). The self-energy operator also implicitly depends on the quasi-particle energies ϵ_i^{QP} via G and W . In the $G_0 W_0$ approach, both quantities are constructed using the Kohn-Sham (KS) single-particle energies ϵ_i^{KS} of the underlying DFT

calculation. Furthermore, the obtained $\varepsilon_i^{\text{QP}}$ can be used in the construction of an updated self-energy operator, followed by the determination of improved quasi-particle energies. This procedure can either be repeated n times ($G_n W_n$) or until self-consistency in the energies is reached (ev*GW*). On top of this, the BSE can either be used in its full form including resonant-antiresonant coupling terms or within the Tamm–Danoff Approximation (TDA). Benchmarking excitation energies [81, 85, 91] has revealed that best accuracy with respect to the theoretical best estimate is achieved with the ev*GW*-BSE/full/FAA variant, with the underlying DFT ground state calculation performed with a hybrid functional, typically PBE0 [92]. A systematic analysis of the accuracy of excited state geometries using *GW*-BSE in general and the influence of the specific methodological choices mentioned above intrinsic to the *GW*-BSE steps in particular has not been performed, to date.

Here, we give a benchmark of optimized excited state geometries for a set of standard small molecules comprising carbon monoxide, acetone, *s*-cis- and *s*-trans acrolein, and methylenecyclopropene (MCP) in different symmetries and covering both $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ excitations [93–95]. We briefly discuss the suitability of the various quasi-particle self-consistency choices in the *GW* step for the calculation of numerical gradients by inspection of the one-dimensional excited-state potential energy surface (PES) of CO and also highlight the influence of the different self-energy models and the BSE types. Then, we focus on eight variants of the *GW*-BSE approach for the rest of molecular set and evaluate them against each other. The obtained geometries are compared to reference data from CASPT2, variational Monte-Carlo (VMC), CC2, and TDDFT with the PBE0 [92], CAM-B3LYP [96], and M06-2X [97] functionals, respectively, in order to assess how reliable *GW*-BSE based excited states geometries are.

3.2 Results

In the following, if not explicitly stated otherwise, all calculations have been performed using the cc-pVTZ basis set [98] with the optimized RI-basis from Ref. [99], and the hybrid PBE0 functional [92] on the DFT level. All molecular orbitals are used for the calculation of the screening in the RPA step of *GW* and in the product basis of the electron-hole wave functions in the BSE. For G_0W_0 and $G_{50}W_{50}$, we determine the quasiparticle corrections to all molecular orbital energies, while in ev*GW*, only the lowest $2n_{\text{occ}}$ (all n_{occ} doubly occupied and the lowest n_{occ} unoccupied orbitals) are explicitly corrected and the higher levels are scissors shifted according to the highest absolute quasiparticle correction among the explicitly corrected unoccupied orbitals.

The Z-matrix coordinates of a molecule are used for *GW*-BSE geometry optimizations. The geometry is updated according to numerical gradients of the excited state energy with respect to internal coordinates using the Broyden–Fletcher–Goldfarb–Shanno method. The numerical gradients are evaluated by computing the central difference with displacements of 10^{-3} a_B in the bond lengths and 10^{-2} deg in the bond and dihedral angles. The algorithm begins at the optimized DFT ground-state geometry and proceeds iteratively until the forces are smaller than 10^{-3} Hartree/Bohr and 10^{-4} Hartree/deg for the bond lengths and (bond and dihedral) angles, respectively.

3.2.1 Carbon monoxide

As described in Chapter 2.3, *GW*-BSE calculations can be performed in different forms depending on the treatment of the frequency depen-

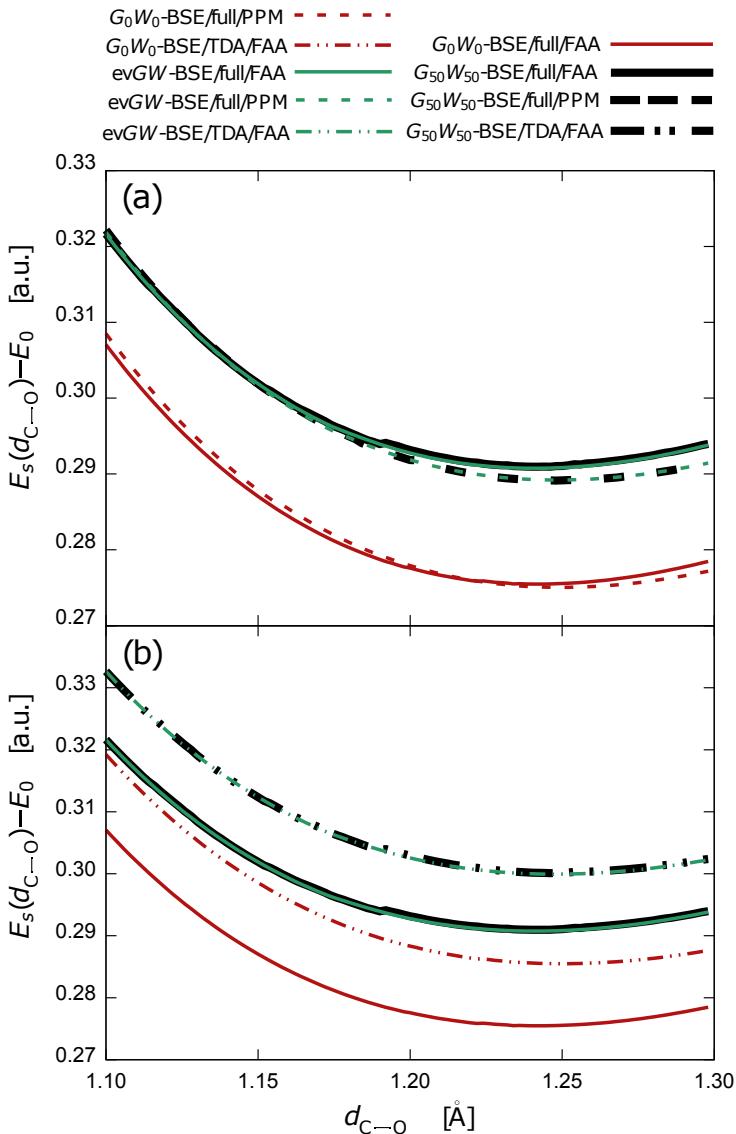


Figure 3.1: Total energy of the $A^1\Pi$ excited state in CO as a function of bond length d_{C-O} , relative to the energy of the optimized ground state E_0 . (a): Results for the full BSE with G_0W_0 - (red), $G_{50}W_{50}$ - (black), and $evGW$ -BSE (green) variants using the cc-pVTZ basis and the PBE0 hybrid functional in the underlying DFT calculation, employing both the FAA (full lines) and PPM (dashed lines) for the frequency-dependence of the self-energy. (b): Results for the G_0W_0 - (red), $G_{50}W_{50}$ - (green), and $evGW$ -BSE (blue) variants with the FAA using the full BSE (full lines) and the TDA (dashed lines).

Table 3.1: Optimized bond length (in Å) in the $A^1\Pi$ excited state of carbon monoxide, as obtained from different *GW*-BSE variants with the cc-pVTZ basis and the PBE0 hybrid functional in the underlying DFT calculation.

	full BSE		TDA	
	FAA	PPM	FAA	PPM
G_0W_0 -BSE	1.241	1.249	1.249	1.256
$G_{50}W_{50}$ -BSE	1.242	1.248	1.247	1.256
ev <i>GW</i> -BSE	1.241	1.248	1.248	1.255
G_0W_0 -BSE ¹				1.26
CIS ²		1.213		
CIS(D) ²		1.263		
EOM-CCSD ²		1.224		
Exp. ³		1.235		

¹from Ref. [89], plane wave basis

²from Ref. [100], pVTZ+ basis

³from Ref. [101]

dence of the self-energy (FAA vs. PPM), the level of self-consistency (G_0W_0 vs G_nW_n vs ev*GW*), both related to the *GW* step, and the use of the full BSE (Eq. (2.26)) vs the TDA (Eq. (2.33)). To illustrate the effects of the different approximations on the obtained excited state total energies and optimized geometries, we consider in this section the one-dimensional potential energy surface of the excited state $A^1\Pi$ in carbon monoxide as a function of carbon-oxygen distance ($d_{\text{C-O}}$). All 60 molecular orbitals are included in the RPA calculation and in the determination of quasi-particle energies in G_0W_0 and $G_{50}W_{50}$, respectively. For ev*GW* explicit corrections are computed for the lowest 14 molecular orbitals. In all cases the BSE product space is formed with 7 occupied and 53 unoccupied orbitals.

Figure 3.1(a) shows the influence of the different levels at the GW step. Results obtained with the FAA (TDA) are shown as solid (dashed) lines, while G_0W_0 is indicated by red, $G_{50}W_{50}$ by green, and ev GW by blue. Considering the FAA, as expected, the G_0W_0 curve is obtained as about 0.12 Hartree lower in energy than with the iterative approaches, mainly due to the too small quasiparticle gap in the one-shot version. Both $G_{50}W_{50}$ and ev GW show unsurprisingly very similar results. However, at a distance of 1.19 Å, the $G_{50}W_{50}$ curve exhibits a faint step. This is due to the fact that, when all molecule orbital energies are quasiparticle-corrected, states with high energy are characterized by a highly structured frequency-dependence of the self-energy. As a consequence, finding solutions to the quasi-particle equation Eq. (2.24) at the required accuracy is numerically difficult and even more so due to the existence of multiple possible solutions over a small energy interval. This problem is avoided in the ev GW calculation by restricting the number of explicitly quasi-particle corrected orbitals, as mentioned above.

Employing the PPM (dashed lines in Figure 3.1), yields excited state PESs with little difference to their FAA counterparts, and the same observations regarding the difference of G_0W_0 to the iterative methods can be made as in the FAA case. However, there is no visually discernible difference between $G_{50}W_{50}$ and the ev GW results, which can be attributed to the less structured self-energy in the PPM. In Figure 3.1(b), we compare the the same $A^1\Pi$ excited state PES obtained when using the TDA instead of the full BSE on top of the different GW versions with the FAA. Generally, the TDA energies are roughly 0.1 Hartree higher, though the curvature of the PESs are very similar.

Finally, we summarize in Table 3.1 the optimal carbon-oxygen bond length obtained by the various methods, compared to literature. In all case, we can see that the different GW -BSE methods hardly influence

the location of the minimum (no deviation larger than 0.002 Å). The use of the PPM leads to a slightly enlarged bond (up to 0.008 Å for the full BSE and 0.009 Å for the TDA). Similarly, using the TDA for the same self-energy method also leads to bond lengths being less than 0.01 Å shorter than for the full BSE. Our G_0W_0 -BSE/TDA/PPM result (1.256 Å) is also in agreement with the value of 1.26 Å reported from a calculation with a plane wave basis [89]. The optimized bond length of the formally most precise method with the fewest approximations, ev GW -BSE/full/FAA, of 1.241 Å is close to the experimentally determined value of 1.235 Å.

The above results for CO were obtained by finely scanning the excited state total energy as a function of its bond length. For molecules with more degrees of freedom such a procedure is impractical and instead energy-gradient based optimization algorithms need to be employed. Due to the lack of analytic expression for the gradients of the respective GW -BSE methods, numerical differentiation has to be used. The steps observed in the total energies of CO with the $G_{50}W_{50}$ method clearly indicates that this variant is unsuitable for yielding an accurate gradient. In the following, we will therefore focus on the evaluation of G_0W_0 - and ev GW -based BSE approaches for the geometry optimization of a set of small molecules in excited states.

3.2.2 Acetone

We start with the optimization of the $n \rightarrow \pi^*$ excited state of acetone ($1^1A''$) with C_s symmetry imposed. This structure is due to the pyramidalization of the central carbon atom out of the molecular plane, and it allows to address the quality of predictions of bond lengths, bond angles, and dihedrals in a single small molecule. All 204 molecular orbitals are included in the RPA calculation and in the determination

Excited-state geometry optimization with ***GW*-BSE**

Table 3.2: Optimized bond lengths (in Å), angles and dihedrals (in degrees) of the $n \rightarrow \pi^*$ excited state ($1^1A''$) of acetone in C_s symmetry. Results from the different GW -BSE variants are obtained with the cc-pVTZ basis and the PBE0 hybrid functional in the underlying DFT calculation. Deviations Δ , with respect to the CASPT2 reference are given in parentheses.

	C=O	C-C	$\theta(C-C-C)$	$\Theta(H-C-C=O)$
G_0W_0 -BSE/full/FAA	1.327 (-0.023)	1.495 (-0.001)	113.63 (+0.88)	49.83 (-2.42)
ev GW -BSE/full/FAA	1.302 (-0.048)	1.504 (+0.008)	114.58 (+1.83)	50.09 (-2.16)
G_0W_0 -BSE/TDA/FAA	1.321 (-0.029)	1.492 (-0.004)	114.16 (+1.41)	49.99 (-2.26)
ev GW -BSE/TDA/FAA	1.303 (-0.047)	1.502 (+0.006)	114.61 (+1.86)	49.68 (-2.67)
G_0W_0 -BSE/full/PPM	1.327 (-0.023)	1.494 (-0.002)	113.87 (+1.12)	49.75 (-2.50)
ev GW -BSE/full/PPM	1.308 (-0.042)	1.497 (+0.001)	114.27 (+1.52)	49.65 (-2.60)
G_0W_0 -BSE/TDA/PPM	1.321 (-0.029)	1.490 (-0.006)	114.44 (+1.69)	49.04 (-3.21)
ev GW -BSE/TDA/PPM	1.308 (-0.042)	1.496 (+0.000)	114.43 (+1.68)	49.33 (-2.92)
CASPT2 ¹	1.350	1.496	112.75	52.25
CC2 ¹	1.404 (+0.054)	1.477 (-0.019)	112.63 (-0.12)	55.37 (+3.12)
VMC ^{1,2}	1.344 (-0.006)	1.489 (-0.007)	112.52 (-0.23)	52.16 (-0.09)
TDDFT/PBE0 ¹	1.301 (-0.049)	1.493 (-0.003)	114.91 (+2.16)	51.50 (-0.75)
TDDFT/CAM-B3LYP ¹	1.295 (-0.055)	1.504 (+0.008)	114.84 (+2.09)	51.38 (-0.87)
TDDFT/M06-2X ¹	1.288 (-0.062)	1.516 (+0.020)	114.90 (+2.15)	50.10 (-2.15)

¹from Ref. [102]

²with pseudopotentials and a pVQZ type basis of H and pVTZ type basis for C and O

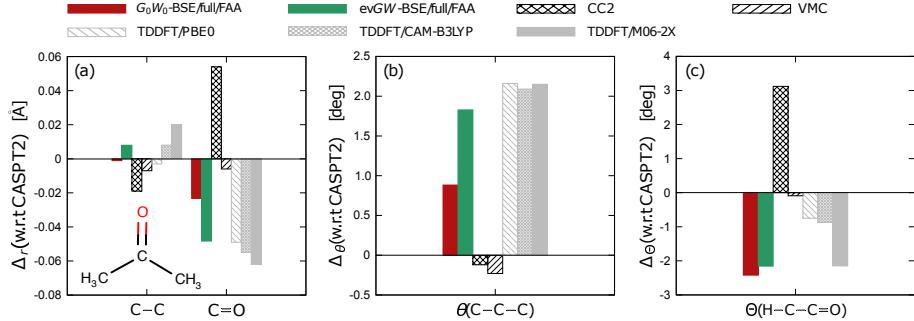


Figure 3.2: Deviations with respect to the CASPT2 reference of bond lengths (in \AA), angle, and dihedral (in degrees) of acetone in bent C_s symmetry as resulting from G_0W_0 -BSE/full/FAA (red) and ev GW -BSE/full/FAA (green) optimizations of the $n \rightarrow \pi^*$ excited state, as well as CC2, VMC, and TDDFT results.

of quasi-particle energies in G_0W_0 . For ev GW explicit corrections are computed for the lowest 32 molecular orbitals, while the BSE product space is formed with 16 occupied and 172 unoccupied orbitals. The resulting optimized structural parameters are listed in Table 3.2, together with reference data from the literature. Deviations from the respective results obtained with CASPT2 are given in parentheses. Note that we chose CASPT2 as a reference for the analysis since it has been shown to compare favorably with CC3 [103] and since it has previously been used in a similar study [102], which makes the data readily available.

Across the eight different GW -BSE variants, the obtained C=O (C–C) bond lengths agree within 0.048 \AA (0.008 \AA) or better with the CASPT2 reference. The quality of the predictions is slightly better than that of TDDFT or CC2, but overall a bit worse than VMC. Similar observations can be made for the C–C–C angle, where the GW -BSE optimizations yield values between 0.88° and 1.86° larger than CASPT2. For comparison, the deviations of the TDDFT results from this reference are around

2.1° , while it is -0.12° for CC2 and -0.23° in VMC. All *GW*-BSE variants consistently underestimate the H–C–C=O dihedral angle in a range of 2.16° to 3.21° . Interestingly, this is very similar to what is obtained for TDDFT/M06-2X but larger than TDDFT with PBE0 and CAM-B3LYP functionals.

From the data in Table 3.2 it is apparent that while there is some variation in the results of the *GW*-BSE methods, it is mostly similar in variation to the variations of TDDFT depending on the choice of the functional. We focus therefore in the following on the two approaches with the least approximations: G_0W_0 -BSE/full/FAA (red) and ev*GW*-BSE/full/FAA (green) as visualized in Figure 3.2. It is interesting to note that the one-shot approach yields better bond lengths and angles than the version with eigenvalue self-consistency, while for the dihedral both are comparable. This is in contrast to what is known for the excitation energies from the ground state structure, where ev*GW* is generally considered to be superior to G_0W_0 .

When imposing C_{2v} symmetry, the structure of acetone in its $n \rightarrow \pi^*$ excited state is planar. Based on the findings on the bent C_s geometry, we restrict the explicit discussion of *GW*-BSE optimized geometries to those obtained with G_0W_0 -BSE/full/FAA and ev*GW*-BSE/full/FAA, and compare them to reference methods. Structural parameters for all *GW*-BSE variants are summarized in Table 3.3. Figure 3.3(a) shows the deviations of the optimized carbon–carbon bond length with respect to the CASPT2 results. Overall, we notice a behavior consistent with that of the C_s structure (see Figure 3.2): In both G_0W_0 -BSE/full/FAA and ev*GW*-BSE/full/FAA, the C–C bond length of 1.484 \AA and 1.490 \AA , respectively, are in very close agreement with the reference of 1.487 \AA . Deviations for the C=O bond are slightly larger with the G_0W_0 -BSE/full/FAA underes-

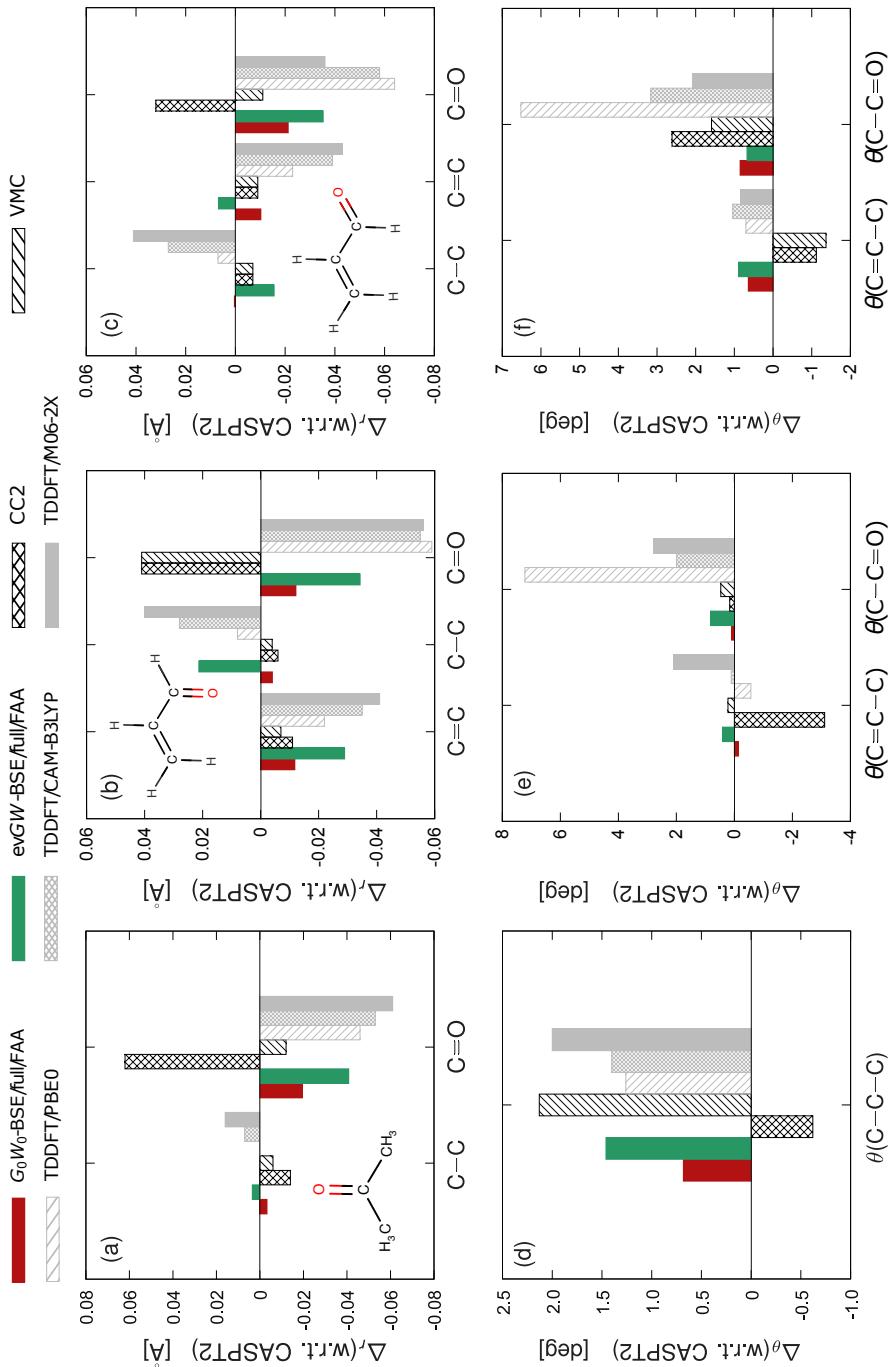


Figure 3.3: Deviations with respect to CASPT2 reference of bond lengths (in Å) and angles (in degrees) of acetone in planar C_{2v} symmetry ((a) and (d)), s-cis ((b) and (e)), and s-trans acrolein ((c) and (f)) as resulting from G_0W_0 -BSE/full/FAA (red) and ev GW -BSE/full/FAA (green) $n \rightarrow \pi^*$ excited state optimizations, as well as CC2, VMC, and TDDFT results. See insets for atom definitions.

Table 3.3: Optimized bond lengths (in Å) and angles (in degrees) of the $n \rightarrow \pi^*$ excited state (1^1A_2) of acetone in C_{2v} symmetry. Results from the different *GW*-BSE variants are obtained with the cc-pVTZ basis and the PBE0 hybrid functional in the underlying DFT calculation.

	C=O	C–C	$\theta(\text{C–C–C})$
G_0W_0 -BSE/full/FAA	1.341	1.484	117.54
ev <i>GW</i> -BSE/full/FAA	1.319	1.490	118.32
G_0W_0 -BSE/TDA/FAA	1.340	1.482	117.43
ev <i>GW</i> -BSE/TDA/FAA	1.319	1.491	118.51
G_0W_0 -BSE/full/PPM	1.339	1.485	117.50
ev <i>GW</i> -BSE/full/PPM	1.324	1.486	118.20
G_0W_0 -BSE/TDA/PPM	1.337	1.486	117.73
ev <i>GW</i> -BSE/TDA/PPM	1.323	1.487	118.29
CASPT2 ¹	1.360	1.487	116.86
CC2 ¹	1.422	1.473	116.24
VMC ^{1,2}	1.348	1.481	118.99
TDDFT/PBE0 ¹	1.314	1.487	118.12
TDDFT/CAM-B3LYP ¹	1.307	1.494	118.26
TDDFT/M06-2X ¹	1.299	1.503	118.86

¹from Ref. [102]

²with pseudopotentials and a pVDZ type basis of H and pVTZ type basis for C and O

timating the reference value by 0.019 Å, and ev*GW*-BSE/full/FAA by 0.041 Å. Notably, these bond length deviations are slightly larger than in VMC, but smaller than what is found for CC2 and TDDFT with various functionals. The same observation holds for the C–C–C angle as shown in Figure 3.3(d), which is overestimated in both G_0W_0 -BSE/full/FAA (by 0.68°) and ev*GW*-BSE/full/FAA (by 1.46°). Especially the former deviation is smaller than those of TDDFT and in absolute value close to VMC.

3.2.3 Influence of the ground state gradient

As mentioned in the introduction, the excited-state *GW*-BSE gradient with respect to the nuclear coordinates, $\partial_{\mathbf{R}}E_S$, is composed of the standard gradient of the ground state energy $\partial_{\mathbf{R}}E_0$ and the gradient of the excitation energy $\partial_{\mathbf{R}}\Omega_S$. The previous section focused on the influence of different methods to obtain Ω_S for the numerical derivative when the ground state is evaluated with DFT and the PBE0 functional. The choice of this specific functional is motivated by the fact that it has been shown to be very reliable for the calculation of vertical excitation energies within *GW*-BSE [85, 91].

In order to scrutinize the influence of the ground state gradient, we here first repeat the optimization of the excited state of acetone in C_s symmetry with the PBE functional and restrict ourselves to the G_0W_0 -BSE/full/FAA and ev*GW*-BSE/full/FAA versions. The respective results are given in Table 3.4. As a general observation, PBE is known to overestimate bond lengths in the ground state, while the hybrid PBE0 underestimates them, and we can see a similar trend in the optimized excited state geometries. For instance, the C=O bond is elongated by 0.053 Å in G_0W_0 -BSE/full/FAA compared to the PBE0 result and by 0.019 Å in

Table 3.4: Optimized bond lengths (in Å), angles and dihedrals (in degrees) of the $n \rightarrow \pi^*$ excited state ($1^1A''$) of acetone in C_s symmetry. Results from the G_0W_0 -BSE/full/FAA and ev GW -BSE/full/FAA are obtained with the cc-pVTZ basis for different evaluations of the ground state energy (PBE0, PBE, RPA) in the underlying DFT calculation. Deviations Δ_i with respect to the CASPT2 reference are given in parentheses.

	C=O	C-C	$\theta(C-C-C)$	$\Theta(H-C-C=O)$
G_0W_0 -BSE/full/FAA@PBE0	1.327 (-0.023)	1.495 (-0.001)	113.63 (+0.88)	49.83 (-2.42)
G_0W_0 -BSE/full/FAA@PBE	1.380 (+0.030)	1.481 (-0.015)	112.97 (+0.22)	49.31 (-2.94)
G_0W_0 -BSE/full/FAA@RPA	1.335 (-0.015)	1.496 (+0.000)	113.33 (+0.58)	50.25 (-2.00)
ev GW -BSE/full/FAA@PBE0	1.302 (-0.048)	1.504 (+0.008)	114.58 (+1.83)	50.09 (-2.16)
ev GW -BSE/full/FAA@PBE	1.321 (-0.029)	1.515 (+0.019)	114.27 (+1.52)	50.09 (-2.16)
ev GW -BSE/full/FAA@RPA	1.309 (-0.041)	1.506 (+0.010)	113.93 (+1.18)	50.03 (-2.22)

evGW-BSE/full/FAA. We observe small variations also for bond and dihedrals angles.

As an alternative, we also consider different evaluation of the ground state energy. Based on the DFT orbitals, the total energy is not determined using the DFT exchange-correlation functional but full Hartree–Fock-like exchange and RPA correlation energy, i.e., $E_0 = E_0^{\text{DFT}} - E_{\text{xc}} + E_{\text{x}} + E_{\text{c}}^{\text{RPA}}$. Specifically, the correlation is given as [104]

$$E_{\text{c}}^{\text{RPA}} = \frac{1}{2} \text{tr}(\underline{\mathbf{M}}^{1/2} - \underline{\mathbf{H}}^{\text{RPA,res}}) \quad (3.2)$$

where $\underline{\mathbf{M}} = (\underline{\mathbf{H}}^{\text{RPA,res}} - \underline{\mathbf{K}}^{\text{RPA}})^{1/2} (\underline{\mathbf{H}}^{\text{RPA,res}} + \underline{\mathbf{K}}^{\text{RPA}}) (\underline{\mathbf{H}}^{\text{RPA,res}} - \underline{\mathbf{K}}^{\text{RPA}})^{1/2}$ and $\underline{\mathbf{H}}^{\text{RPA,res}}$ ($\underline{\mathbf{K}}^{\text{RPA}}$) is the resonant (anti-resonant) part of the RPA Hamiltonian with $\kappa = -1$ and $K^d = 0$ in Eqs. 2.27 and 2.28. This approach can formally be considered the limit of a BSE ground state energy [105, 106] with an unscreened Coulomb potential. The resulting optimized excited-state structural parameters using PBE0 orbitals in E_{x} and $E_{\text{c}}^{\text{RPA}}$ for acetone C_s are also given in Table 3.4. Some smaller changes are noted with respect to the calculations with the standard PBE0 ground state energy. In G_0W_0 -BSE/full/FAA (*evGW*-BSE/full/FAA), the C=O bond extends by 0.008 Å (0.007 Å), the C–C bond by 0.001 Å (0.002 Å). The C–C–C angle is minimally reduced by 0.30° (0.65°) for G_0W_0 -BSE/full/FAA (*evGW*-BSE/full/FAA), and similarly small variations are seen for the H–C–C=O dihedral.

Evidently, the specific choice of method for evaluating the ground state energy influences the excited state optimization to some extent. From the obtained structural parameters, no clear preference for either of the methods can be determined. In order to facilitate a systematic comparison with literature data, in particular with TDDFT and the various hybrid functionals, we will focus on evaluating the ground state

energy with the standard DFT energy and the PBE0 functional for the remainder of this work.

3.2.4 Acrolein

The preceding analysis of acetone in the $n \rightarrow \pi^*$ excited state has indicated that the G_0W_0 -BSE/full/FAA method provides optimized geometries closer to those of the CASPT2 reference than the ev GW -BSE/full/FAA version. To scrutinize whether this notion also holds in more general cases, we now consider the same type of excitation in *s*-cis and *s*-trans configurations of acrolein and optimize the respective structures in a planar geometry. For both configurations, all 176 molecular orbitals are included in the RPA calculation and in the determination of quasi-particle energies in G_0W_0 . In ev GW explicit corrections are computed for the lowest 30 molecular orbitals, while the BSE product space is formed with 15 occupied and 161 unoccupied orbitals.

Detailed structural parameters of the optimized structure for *s*-cis acrolein as resulting from all GW -BSE methods are summarized in Table 3.5. In Figure 3.3, we show the deviations of the bond lengths (panel (b)) and angles (panel (e)) with respect to the CASPT2 reference. Overall, we notice that, as in the case of acetone, the G_0W_0 -BSE/full/FAA method yields structures in closer agreement with CASPT2 than ev GW -BSE/full/FAA. For instance, the C=C bond length is shorter by 0.012 Å in G_0W_0 -BSE/full/FAA but shorter by 0.029 Å in ev GW -BSE/full/FAA. Similar ratio between the two methods is also found for the C–C and C=O bond. In fact, the deviations seen with G_0W_0 -BSE/full/FAA are very close to those obtained with CC2 and VMC, respectively. The ev GW -BSE/full/FAA results for the bonds, on the other hand, are comparable

Table 3.5: Optimized bond lengths (in Å) and angles (in degrees) of the $n \rightarrow \pi^*$ excited state ($1^1A''$) of *s*-cis acrolein. Results from the different *GW*-BSE variants are obtained with the cc-pVTZ basis and the PBE0 hybrid functional in the underlying DFT calculation.

	C=O	C–C	C=C	$\theta(\text{C–C–C})$	$\theta(\text{C–C–O})$
G_0W_0 -BSE/full/FAA	1.321	1.374	1.379	124.39	123.94
ev <i>GW</i> -BSE/full/FAA	1.299	1.399	1.362	124.94	124.66
G_0W_0 -BSE/TDA/FAA	1.313	1.381	1.375	124.41	125.00
ev <i>GW</i> -BSE/TDA/FAA	1.298	1.399	1.362	124.74	124.63
G_0W_0 -BSE/full/PPM	1.313	1.378	1.379	124.24	124.42
ev <i>GW</i> -BSE/full/PPM	1.303	1.398	1.363	124.63	124.14
G_0W_0 -BSE/TDA/PPM	1.311	1.377	1.381	123.93	124.73
ev <i>GW</i> -BSE/TDA/PPM	1.301	1.398	1.363	124.42	124.15
CASPT2 ¹	1.333	1.378	1.391	124.53	123.84
CC2 ¹	1.374	1.372	1.380	121.42	124.00
VMC ^{1,2}	1.320	1.373	1.386	124.46	125.16
TDDFT/PBE0 ¹	1.274	1.386	1.369	123.96	131.06
TDDFT/CAM-B3LYP ¹	1.278	1.403	1.356	124.64	125.83
TDDFT/M06-2X ¹	1.277	1.418	1.350	124.09	126.62

¹from Ref. [102]²with pseudopotentials and a pVDZ type basis of H and pVTZ type basis for C and O

Table 3.6: Optimized bond lengths (in Å) and angles (in degrees) of the $n \rightarrow \pi^*$ excited state (1A_2) of *s*-trans acrolein. Results from the different *GW*-BSE variants are obtained with the cc-pVTZ basis and the PBE0 hybrid functional in the underlying DFT calculation.

	C=O	C–C	C=C	$\theta(\text{C–C–C})$	$\theta(\text{C–C–O})$
G_0W_0 -BSE/full/FAA	1.315	1.377	1.382	123.90	125.39
ev <i>GW</i> -BSE/full/FAA	1.301	1.361	1.399	124.15	125.21
G_0W_0 -BSE/TDA/FAA	1.312	1.376	1.382	124.50	125.68
ev <i>GW</i> -BSE/TDA/FAA	1.300	1.363	1.398	123.94	125.58
G_0W_0 -BSE/full/PPM	1.315	1.379	1.378	123.33	125.76
ev <i>GW</i> -BSE/full/PPM	1.306	1.364	1.396	123.88	124.77
G_0W_0 -BSE/TDA/PPM	1.311	1.382	1.378	123.02	125.78
ev <i>GW</i> -BSE/TDA/PPM	1.305	1.365	1.396	123.77	124.78
CASPT2 ¹	1.336	1.377	1.392	123.26	124.54
CC2 ¹	1.368	1.370	1.383	122.14	127.15
VMC ^{1,2}	1.327	1.368	1.383	122.57	125.82
TDDFT/PBE0 ¹	1.272	1.384	1.369	123.96	131.06
TDDFT/CAM-B3LYP ¹	1.278	1.404	1.353	124.30	127.70
TDDFT/M06-2X ¹	1.300	1.418	1.349	124.09	126.62

¹from Ref. [102]

²with pseudopotentials and a pVDZ type basis of H and pVTZ type basis for C and O

to those registered for TDDFT with different functionals and on average slightly better. For the two angles in Figure 3.3(f), the same general observations can be made.

For *s*-trans acrolein (see full data in Table 3.6), the deviations of the two *GW*-BSE methods, as well as those of CC2, VMC, and the TDDFT variants with respect to CASPT2 are shown in panels (c, bonds) and (f, angles) of Figure 3.3. Among the three bonds, we note the smallest deviation for the C–C in G_0W_0 -BSE/full/FAA, while the C=C (C=O) bond is underestimated by 0.010 Å (0.019 Å) with respect to CASPT2. We note again that ev*GW*-BSE/full/FAA results are slightly worse in comparison but overall better than TDDFT. For the two angles in Figure 3.3(f), a behavior similar to that of *s*-cis acrolein is visible: both shown *GW*-BSE variants yield results that are comparable to those of CC2 and VMC, and in case of the C–C=O angle significantly better than TDDFT with the various functionals.

3.2.5 Methylenecyclopropene

Having established in the previous sections that the excited state geometries optimized with the G_0W_0 -BSE/full/FAA and ev*GW*-BSE/full/FAA methods provide structural parameters in good agreement with reference data for $n \rightarrow \pi^*$ transitions in a variety of small molecules, we now turn to the analysis of a $\pi \rightarrow \pi^*$ transition in methylenecyclopropene (MCP, see inset of Figure 3.4). In general, it is to be expected that a twist of 90° around the C=CH₂ group will stabilize the state with a diradical character, which is known to be problematic for CC2 and TDDFT. To facilitate a consistent comparison, we impose planar C_{2v} symmetry on the excited state structure. In the *GW*-BSE calculations, all 176 molecular orbitals are included in the RPA calculation and in the determination of quasi-particle energies in G_0W_0 . For ev*GW* explicit corrections are computed for the lowest 28 molecular orbitals, while the BSE product space is formed with 14 occupied and 162 unoccupied orbitals, respec-

Table 3.7: Optimized bond lengths (in Å) and angles (in degrees) of the $\pi \rightarrow \pi^*$ excited state (1^1B_2) of methylenecyclopropene. Results from the different *GW*-BSE variants are obtained with the cc-pVTZ basis and the PBE0 hybrid functional in the underlying DFT calculation.

	C ₁ =C ₂	C ₁ -C ₃	C ₄ =C ₃	$\theta(C_2=C_1-C_3)$	$\theta(C_1-C_3=C_4)$
<i>G</i> ₀ <i>W</i> ₀ -BSE/full/FAA	1.455	1.340	1.524	146.31	55.69
ev <i>GW</i> -BSE/full/FAA	1.436	1.349	1.504	146.39	56.13
<i>G</i> ₀ <i>W</i> ₀ -BSE/TDA/FAA	1.462	1.350	1.492	146.57	56.22
ev <i>GW</i> -BSE/TDA/FAA	1.445	1.348	1.488	146.51	56.51
<i>G</i> ₀ <i>W</i> ₀ -BSE/full/PPM	1.448	1.349	1.513	146.26	55.72
ev <i>GW</i> -BSE/full/PPM	1.433	1.350	1.450	146.45	56.26
<i>G</i> ₀ <i>W</i> ₀ -BSE/TDA/PPM	1.455	1.348	1.485	146.55	56.58
ev <i>GW</i> -BSE/TDA/PPM	1.440	1.350	1.484	146.56	56.61
CASPT2 ¹	1.461	1.360	1.496	146.64	56.64
CC2 ¹	1.457	1.349	1.512	145.90	55.90
VMC ^{1,2}	1.456	1.351	1.483	146.66	56.66
TDDFT/PBE0 ¹	1.432	1.349	1.485	146.62	56.62
TDDFT/CAM-B3LYP ¹	1.419	1.344	1.486	146.43	56.43
TDDFT/M06-2X ¹	1.424	1.347	1.483	146.62	56.61

¹from Ref. [102]

²with pseudopotentials and a pVDZ type basis of H and pVTZ type basis for C and O

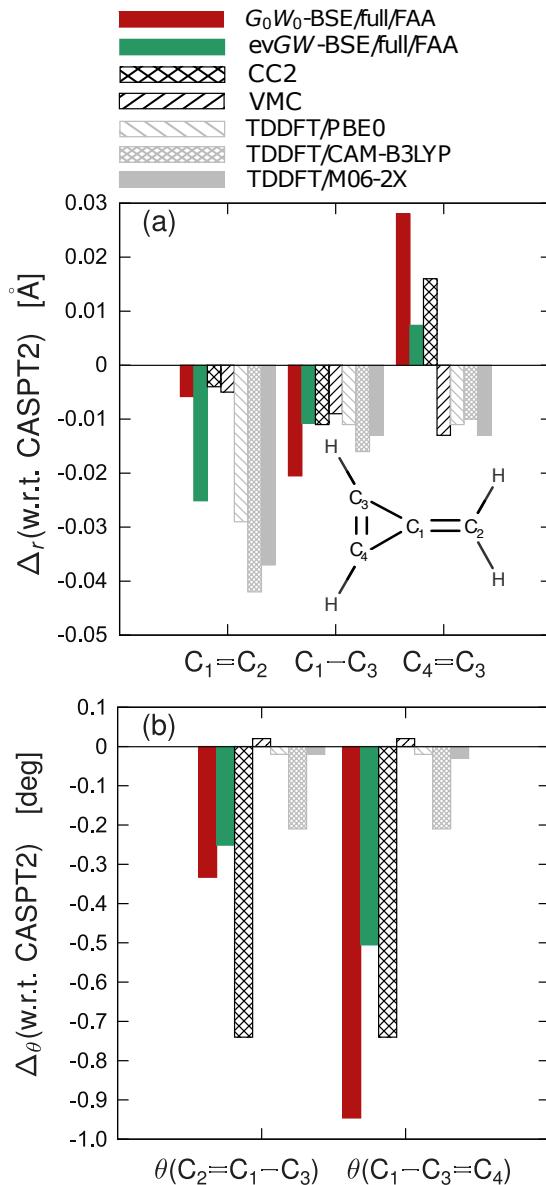


Figure 3.4: Deviations with respect to the CASPT2 reference of bond lengths (in Å) and angles (in degrees) of MCP as resulting from G_0W_0 -BSE/full/FAA (red) and ev GW -BSE/full/FAA (green) optimizations of the $\pi \rightarrow \pi^*$ excited state with C_{2v} symmetry, as well as CC2, VMC, and TDDFT results.

tively. Full data is listed in Table 3.7 and the deviations of the optimized bond length and angles from the CASPT2 reference are shown in Figure 3.4.

We note that we label the bonds in the following, and as shown as the inset in Figure 3.4, according to the double/single bond character in the ground state. In the excited state, this character is inverted for C₁–C₃ and C₄=C₃ in the *GW*-BSE optimizations, in agreement with the literature. In general, the bond length deviations in panel (a) reveal a satisfying agreement: for G₀W₀-BSE/full/FAA, we note that the length of the C₁=C₂ bond of 1.455 Å is very close to the prediction of CC2 and VMC, and with that significantly better than TDDFT. On the other hand, the ev*GW*-BSE/full/FAA result exhibits a deviation of -0.025 Å very similar to that of TDDFT/PBE0. Interestingly, the trend of the G₀W₀-BSE/full/FAA results being closer to the CASPT2 reference that we observed for the $n \rightarrow \pi^*$ excited states does not appear to hold for the C₁–C₃ and C₄=C₃ bonds of MCP. Instead, the deviations of the ev*GW*-BSE/full/FAA method are significantly smaller. For C₁–C₃ it is only -0.011 Å compared to -0.020 Å, and 0.007 Å compared to 0.028 Å for C₄=C₃, respectively. Also for the angles as shown in Figure 3.4(b), our results indicate that the ev*GW*-BSE/full/FAA variant yields more reliable excited state structural properties than G₀W₀-BSE/full/FAA.

3.2.6 Discussion

From the analysis of the excited state geometry optimizations for the different molecules in Chapter 3.2, it has become clear that *GW*-BSE in general provides structural parameters in good agreement with reference data from high level methods. It has also become apparent that the G₀W₀-BSE/full/FAA variant seemed to yield better agreement with

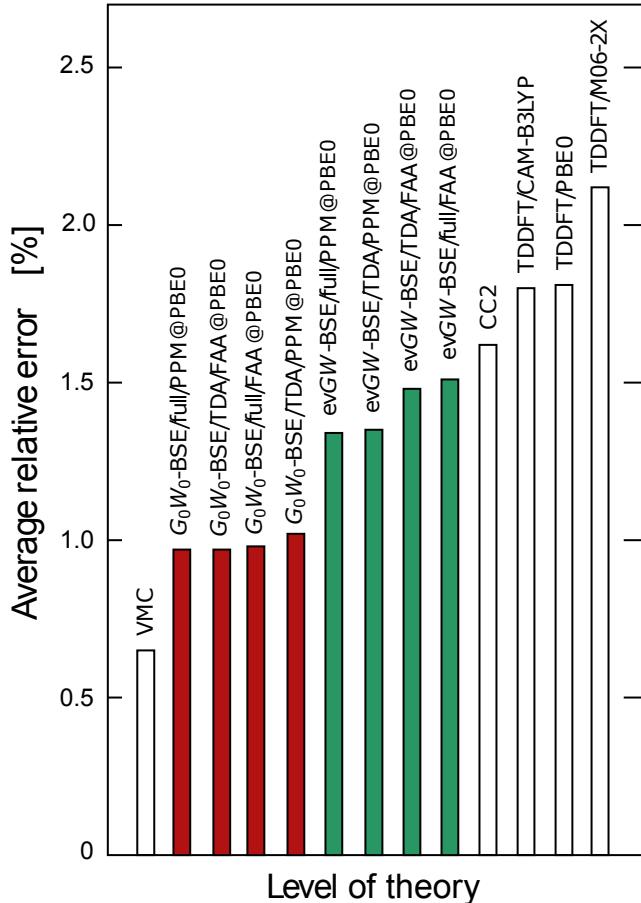


Figure 3.5: Methods ranked according to their average relative error (in %) with respect to the CASPT2 reference.

CASPT2 data than the ev GW -BSE/full/FAA method, which is usually considered to be preferred for accuracy in predicting excitation energies. However, the trend is not always clear, and in particular for $\pi \rightarrow \pi^*$ excitations, considering self-consistency in the GW quasiparticle energies seems to perform better than the one-shot approach. Since we restricted the explicit analysis mostly to the full BSE and FAA for the frequency

dependence of the self-energy, and to give an overall ranking of the different methods, we calculate the mean relative error of all structural parameters with respect to CASPT2 for all the eight different *GW*-BSE combination we have considered.

These average relative errors are shown in Figure 3.5 compared to VMC, CC2, and TDDFT with three functionals. Clearly, all four G_0W_0 -BSE methods studied in this work exhibit an overall error around 1 %, nearly independent of the approximations. In contrast, the use of the ev*GW*-BSE methods for the geometry optimization yield errors of about 1.4-1.5 %. It is interesting to note that the two variants employing the PPM are slightly better than those based in the FAA. While the smallest overall error with respect to CASPT2 is found for VMC, we also see that all *GW*-BSE methods perform slightly better than CC2 or TDDFT with the CAM-B3LYP, PBE0, or M06-2X functionals, respectively.

3.3 Summary

We have presented the results of gas phase geometry optimizations in the excited states of a set of small molecules using many-body Green's functions theory within the *GW* approximation and the Bethe-Salpeter Equation. Our results show that, overall, one-shot G_0W_0 yields geometries in slightly better agreement with CASPT2 reference than ev*GW*-based variants when used with a PBE0 ground state. The use of a plasmon-pole model or the Tamm-Danoff approximation only have a negligible influence on the quality of the obtained structural parameters. We find relative errors smaller than for CC2 and TDDFT with different functionals and only larger than VMC. Our results indicate that *GW*-BSE

can yield geometries in good agreement with established higher-order wave-function methods for small molecules.

A fundamental disadvantage that the *GW*-BSE approach shares with most excited state electronic structure methods is the high computational cost involved. So while from the angle of plain electronic structure theory, the scrutiny of *GW*-BSE has revealed very good predictive power of the method, its practical applicability is limited particularly when numerical gradients need to be calculated. This brings up the idea of using machine learning to speed up computationally heavy calculations. In the following chapters, we will look into these challenges in more detail, study representations, construct ML models, and apply them in hybrid models.

Chapter 4

Wasserstein metric for improved quantum machine learning with adjacency matrix representations

This chapter is based on the letter: O. Çaylak, O. A. von Lilienfeld, and B. Baumeier, *Mach. Learn.: Sci. Technol.* **1**, 03LT01 (2020).

Building on our earlier discussion about the complexities and challenges of first-principles methods, we will show that alternative approaches like machine learning models offer promising results. Key among the challenges identified was the need for effective construction of ML models. One aspect that is central to constructing kernel methods is the choice of the right metric to measure the similarities between molecular representations. This brings us to the heart of this chapter, where we explore the potential of the Wasserstein metric.

We study the Wasserstein metric to measure distances between molecules represented by the atom index dependent adjacency “Coulomb” matrix, used in kernel ridge regression based supervised learning. Result-

ing machine learning models of quantum properties, a.k.a. quantum machine learning models, exhibit improved training efficiency and result in smoother predictions of molecular distortions. We first illustrate smoothness for the continuous extraction of an atom from some organic molecule. Learning curves, quantifying the decay of the atomization energy's prediction error as a function of training set size, have been obtained for tens of thousands of organic molecules drawn from the QM9 data set. In comparison to conventionally used metrics (L_1 and L_2 norm), our numerical results indicate systematic improvement in terms of learning curve off-set for random as well as sorted (by norms of row) atom indexing in Coulomb matrices. Our findings suggest that this metric corresponds to a favorable similarity measure which introduces index-invariance in any kernel based model relying on adjacency matrix representations.

4.1 Introduction

The application of machine learning (ML) to atomistic simulation has been gaining traction over recent years [15–23]. Kernel ridge regression (KRR) models of quantum properties (Q) applicable throughout chemical compound space (CCS) was established in 2012 [107], and has been growing ever since [108–112]. See Refs. [113, 114] for more details and references about the methodology known as Quantum Machine Learning (QML). By now, QML has become a viable and popular approach for generating surrogate property models enabling rapid estimates of relevant molecular and materials properties, holding great promise for computational materials and molecular design [115, 116], as recently exemplified for the discovery of nearly ninety stable crystal candidates in the Elpasolite structure [117].

When setting up standard KRR based QML models of some quantum property P (aka “label”) [110], three fundamental choices must be made, (i) the representation \mathbf{x} (aka “feature”), (ii) the kernel function (k), and (iii) the metric ($\text{dist}(\cdot, \cdot)$), such that

$$P(\mathbf{x}) = \sum_i^N \beta_i k(\text{dist}(\mathbf{x}, \mathbf{x}_i)), \quad (4.1)$$

where N and $\{\beta_i\}$ correspond to number and regression coefficients of training instances, respectively. The representation of a chemical system is known to play an important role. For example, when using incomplete representations (or non-unique), proof was given that QML models can produce absurd results [118]. While the details of the representation (other than uniqueness) are less crucial for artificial neural networks, the specific definition of how a chemical system is being specified is known to dramatically affect the learning efficiency of KRR based QML models. Namely, encoding the right physics, such as translational or atom-index invariance, in the representation results in systematic reduction of quantum data needs for achieving the same pre-defined predictive accuracy [24]. This is of particular interest since QML models are typically trained within scarce data regimes due to (a) the immense computational (or experimental) cost for generating labels and (b) the tremendous scale of CCS [6, 119].

Due to their obvious impact on QML model performance, it is not surprising that substantial efforts have been made to improve conventional representations. For example, using atomization energies of organic molecules stored in the QM9 dataset [25], various benchmark results have been obtained including as representations the Coulomb matrix and BOB (2015) [120], BAML (2016) [24], HDAD (2017) [121], constant-size-descriptors (2018) [122], SLATM [123] (2017), SOAP (2017) [113],

FCHL (2018) [124], MBD (2018) [125], and wavelets (2018) [126]). See Ref. [23] for a joint graphical illustration of learning curves coming from these, as well as artificial neural network based, models. Apart from the representation, the choice of kernel function is also known to affect the performance of the QML model, as shown in Refs. [110, 113, 127–129].

Within this chapter, we focus on the aforementioned third choice (iii): The metric. More specifically, previous studies have predominantly relied on Euclidean or Manhattan norms as a metric. This choice can be questioned when it comes to atom index dependence of adjacency matrices, such as the Coulomb matrix (CM) [107], possibly resulting in discontinuities in the surrogate model due to displacement (or alchemical change [119]) of the nucleus. In this chapter, we discuss how such spurious artifacts are resolved by using a more sophisticated, distribution based measure: The Wasserstein metric [130].

4.2 Method

The application of Kernel Ridge Regression (KRR) models to predicting molecular quantum properties has been very successful over recent years [131–135]. The main idea relies on constructing a kernel matrix with a kernel function k that can quantitatively measure similarity between molecular representations \mathbf{x}_i and \mathbf{x}_j , which are vector representations that encode the molecular physics [136–138]. The Laplacian kernel function, for example, is described as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_1}{\sigma}\right). \quad (4.2)$$

In context of QM, the goal of KRR is to map an input molecular representation \mathbf{x} to a target quantum property p . Such a mapping is given by

$$p(\mathbf{x}) = \sum_{n=1}^N \alpha_n k(\mathbf{x}, \mathbf{x}_n), \quad (4.3)$$

where α_n stand for the n -th regression coefficient, while \mathbf{x}_n being the n -th training sample. The learning process within the KRR framework corresponds to obtaining the regression coefficients vector α for a given reference property vector $\mathbf{p}^{\text{train}}$, a kernel matrix \mathbf{K} , and a regularization coefficient λ by evaluating

$$\alpha = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{p}^{\text{train}}, \quad (4.4)$$

where \mathbf{I} is an identity matrix. Additionally, the so-called hyperparameters λ and σ are optimized using the mean-absolute-error (MAE) metric, where all optimizations are performed with 5-fold cross-validation.

The CM is an adjacency matrix with diagonal and off-diagonal terms corresponding to approximate free atom and nuclear repulsion contributions to the total potential energy of a molecule, respectively [107]. Its adaptation to crystal representations was published subsequently [139]. Its creation was motivated by the fact that it is unique for fixed molecular charges up to permutation of atoms, and that first-principles calculations also require only nuclear coordinates \mathbf{R}_I and nuclear charges Z_I as input.

$$C_{IJ} = \begin{cases} 0.5 Z_I^{2.4} & I = J \\ \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & I \neq J \end{cases}. \quad (4.5)$$

As the CM is invariant to 3D translations and rotations of a molecule, it intrinsically ensures that the molecule’s potential energy is constant under those transformations. Among the early problems identified for the general application of the Coulomb matrix is the index dependence. Sorting the Coulomb matrix such that $||C_m|| \geq ||C_{m+1}||$ for all m , where C_m is the m -th row, renders the CM bijective up to rotation and translation. While appealing for its simplicity and still in use for various applications [140, 141], the use of the sorted CM must be cautioned when applied to situations in which smooth geometrical (or alchemical) changes are under consideration. Such applications include training throughout CCS and subsequent prediction of energies in molecular dynamics trajectories, or geometry relaxations energies, when sorting can lead to a swapping of indices in the vectorized forms of the CM, and sudden atomic index reassessments between test and query system. This *indexing problem* manifests itself in discontinuities in predictions and/or the need for a large number of data points for training the respective models.

Here, rather than attempting to resolve the indexing problem through ever more sophisticated representations, for example using atom centered symmetry functions [142], SOAP [143], HDAD [121], SLATM [123], MBD [125], or FCHL [124], we investigate if this issue can also be resolved by using a different metric, capable of alleviating the sudden reassignment occurring within L_1 or L_2 norms. In particular, the Wasserstein norm of order 1, which is denoted by W_1 , is a natural way to compare two probability distributions p and q [130]. This norm is widely used in various fields, like machine learning, image processing, and signal processing [144–146].

As illustrated in Fig. 4.1, the Wasserstein metric is the minimal amount of work needed to transform one distribution into another; work being

defined as the amount of distribution times the distance it has to be transported. There are many different ways of transporting an amount of distribution from a region x of p into a region y of q . The set of all possible transport plans to move p into q is denoted by Γ . Hence, computing the distance between two distributions can be formulated as an optimization problem where the aim is to find that transport plan $\gamma \in \Gamma$ such that the total amount of work is minimal. The Wasserstein [130] metric is expressed as

$$\begin{aligned} W_1(p, q) &= \inf_{\gamma \in \Gamma} \int_{\mathbb{R}^d \times \mathbb{R}^d} d\gamma(x, y) |x - y| \\ &= \int_{\mathbb{R}} dt \left| P^{-1}(t) - Q^{-1}(t) \right|, \end{aligned} \quad (4.6)$$

where the right-hand side equation was shown to hold in Ref. [147] with P and Q being the cumulative distribution functions of p and q . The vectorized two-dimensional Coulomb matrix representation of a molecule can be used in Eq. (4.6). In practice, the evaluation is done by using empirical cumulative distribution functions (cdf). The integral in Eq.(3) becomes the absolute value of the difference of two distinct cdf's. So, in worst case scenario it behaves like $\mathcal{O}(2n \log(n))$. From now on, we set the L_1 and L_2 based kernel functions k to be the well known Laplacian and Gaussian kernel, respectively. Additionally, we define the Wasserstein-based kernel as

$$k(\mathbf{x}, \mathbf{y}) = e^{-\alpha W_1(\mathbf{x}, \mathbf{y})}. \quad (4.7)$$

Here, we note that other kernel functions or representations could be used in combination with the Wasserstein metric just as well.

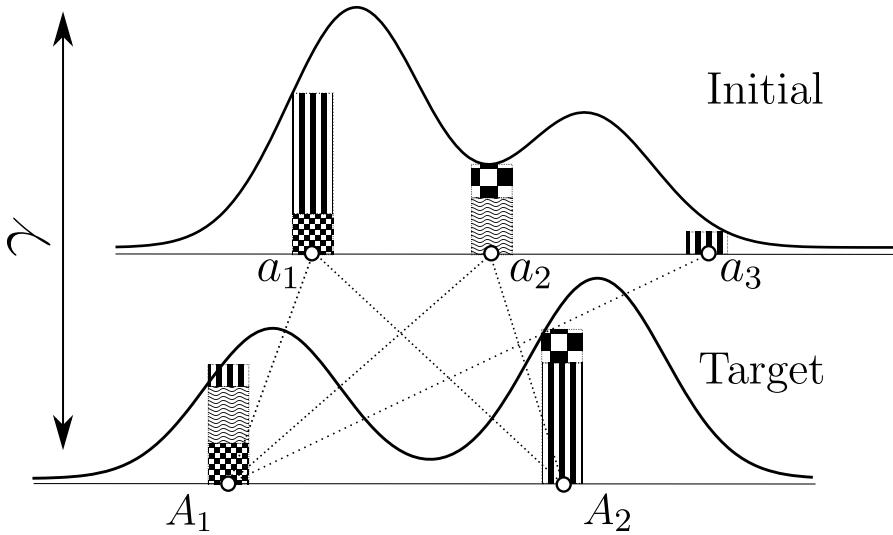


Figure 4.1: Illustrating some transport plan (not necessarily the optimal) between some initial and target distributions. The patterned boxes indicate how an amount of distribution is redistributed from an initial set of points a_1, a_2 , and a_3 to points A_1 and A_2 .

All QML models of atomization energies of QM9 molecules [25] were trained using KRR with 5-fold cross-validation for hyperparameter optimization, and tested on 2000 out-of-sample molecules.

4.3 Results and discussion

First, we illustrate the issue of smoothness by subjecting an innocent organic molecule to drastic distortions. More specifically, and as shown in Fig. 4.2, consider the energy E as a function of continuous displacement d (in steps of 0.01 Å) of some central carbon atom along an axis orthogonal to the molecular plane. The molecule used in this example is drawn

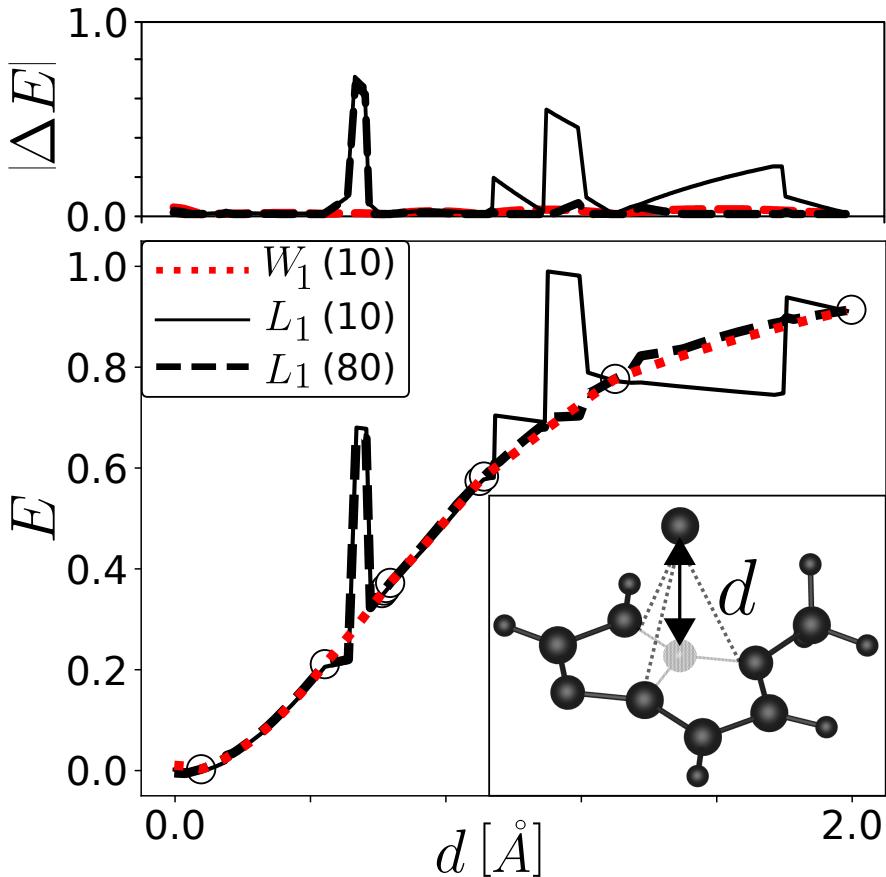


Figure 4.2: Toy model illustrating the effect of atom index invariance (W_1) and lack thereof (L_1). As a model, a normalized atom pair-wise interatomic Lennard Jones potential energy estimate has been used, giving E as a function of displacement d (in steps of 0.01 Å) of a central atom, as shown in the inset. Randomly selected training data points are indicated as open circles. CM based QML predictions with the Wasserstein kernel W_1 trained on ten data points (dotted line) are smooth, while discontinuities are visible for the L_1 norm for small (solid) and large (dashed) training sets. $|\Delta E|$ indicates the energy difference between the true and the predicted value.

randomly from the QM9 dataset.¹ Corresponding Lennard-Jones potential is calculated using the LJ formula where all parameters are set to 1 independent of atom type. Furthermore, we have normalized the curves with respect to the highest energy of obtained results. The energies are smooth, while CM based QML model predictions (after training on 10 instances drawn at random) using L_1 are discontinuous. Even after increasing the training set for the L_1 model to 80 instances, the discontinuity at $d \approx 0.5$ is retained, indicating that lack of learning. It is non trivial to figure out what the exact cause is for the initial discontinuity, mainly due to the highly dependent variables in the CM representation. The Carbon atom that is being extracted is bonded to two Nitrogen atoms and one Carbon atom. However, there is another Carbon-Nitrogen bond in the molecule with initially a longer bond length that is not bonded to the extracted Carbon atom. The extraction of the Carbon atom stretches the bonds and around $d \approx 0.5$ the order swaps, which could be the reason for the first discontinuity in the curve. As pointed out above, the main benefit of Eq. (4.7) is that W_1 is invariant under permutations of row- and column-indices of the adjacency matrix. When applied to the model system from Fig. 4.2, which showed the typical indexing problem with the Laplacian kernel and L_1 -norm, the W_1 metric yields smooth energies as a function of displacement. Two aspects are noteworthy: First, it is apparent that the predictions are smooth (differentiable) across the full range of displacements and do not exhibit any discontinuities that affected the standard kernels in combination with the sorted Coulomb matrix. Second, a training data set including only 10 reference points is sufficient to give accurate results. These numerical results clearly indicate that the use of the Wasserstein metric in the kernel cures the indexing problem and alleviates the associated low prediction quality and data efficiency.

¹dsgdb9nsd_030000.xyz

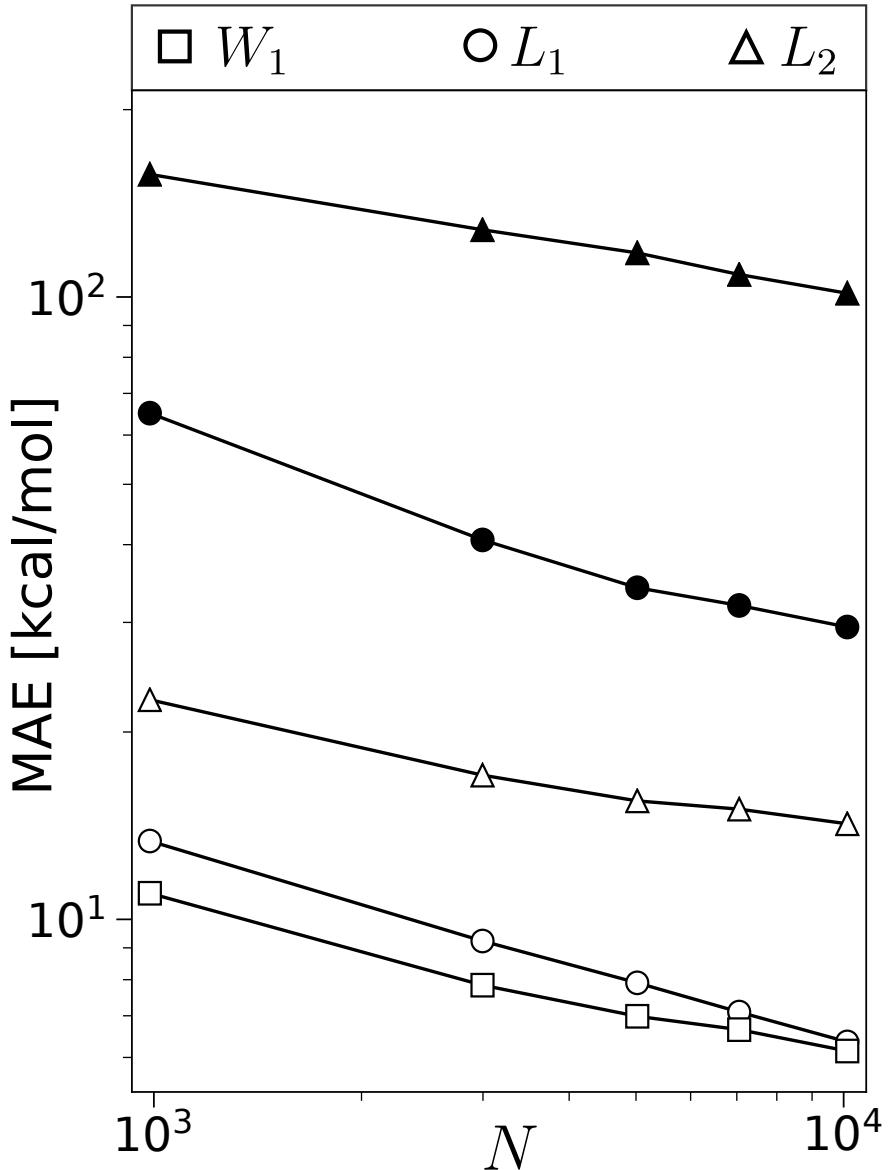


Figure 4.3: Learning curves for DFT/B3LYP/6-31G(2df,p) atomization energies of QM9 molecules for various kernels and representations. Circles and triangles indicate the mean absolute error (MAE) obtained with L_1 and L_2 based kernels, respectively, using randomized (filled symbols) and sorted (open symbols) Coulomb matrix representation. The Wasserstein metric is permutation invariant and the MAE obtained with it is given by squares.

To further investigate the performance of Wasserstein based kernels Eq. (4.7) in QML models, we have turned to the classic benchmark of atomization energies of organic molecules in the QM9 data set [25]. After training on up to 10k molecules, randomly sampled from the entire QM9 data set, learning curves are presented in Fig. 4.3. Results for L_2 and L_1 based kernels are in line with observations made in Ref. [110]: For sorted as well as for randomly indexed CMs the L_1 based QML model exhibits lower off-sets than the L_2 based model. Not surprisingly, use of sorted CMs also results in smaller off-sets than for random CMs. However, the W_1 based metric results in the same learning curve after being applied to random as well as to sorted CMs. Its overall learning curve off-set and slope indicates same (even slightly better) performance as the L_1 norm applied to the sorted CM, reaching \sim 6 kcal/mol after training on 10k instances.

4.4 Conclusions

Considering the findings for the continuous atomic displacement as well as the QM9 molecules, the Wasserstein metric enables the generation of QML models which achieve (a) data-efficient learning and (b) smooth target function estimates. While all our numerical evidence has relied on the CM, we emphasize that the observed solution of the indexing problem and the simultaneous improvement of the predictions by using the Wasserstein kernel is not inherently specific to the CM representation. In fact, the Wasserstein metric can readily be tested with other QML models relying on any graph-based representation. This could be particularly relevant in the context of recent work on learning force-fields or electronic properties, relying on inverse distances rather than the CM representation. [148–151]

To summarize, we have presented the Wasserstein metric as an index-invariant way to measure distances between molecular graph-based representations. Our numerical findings indicate that the resulting QML models combine smoothness with data efficiency in learning. Future work will explore the various possible combinations of kernel functions, Wasserstein metric, and representations other than the CM. All methods used in this work are now also available in the QMLCODE package. [152]

Building upon our insights into the potential of the Wasserstein metric in the context of quantum machine learning models, we must also highlight the significance of molecular representation in dictating the quality and accuracy of predictions. While the Wasserstein metric measures distances between molecules, it is the very construction and choice of these representations that can largely influence outcomes. The essence of achieving optimum predictive capabilities lies not only in the effective metric choice but also in the robust and intricate representation of the molecules themselves. In this light, the ensuing chapter delves deeper into the nuances of molecular representation. Introducing the Δ -Machine Learning approach, we explore its efficacy in predicting *GW* quasiparticle energies and photoelectron spectra, all while emphasizing the vital role played by various orbital-sensitive graph-based representations in achieving this feat.

Chapter 5

Δ-Machine Learning of Quasiparticle Energies in Molecules and Clusters

This chapter is based on the paper: O. Çaylak and B. Baumeier, J. Chem. Theory Comput. **17**, 4891 (2021).

As we have underscored the critical nature of the Wasserstein metric and its application in quantum machine learning, it becomes abundantly clear that the methods we employ to represent molecular structures and interactions are essential. The challenge is not just about selecting the right metric but intricately intertwining it with a representation that captures the essence of the molecular structure. In this pursuit of optimal representation, we turn our attention to the potential of orbital-sensitive graph-based representations. With this context, we transition into the specifics of the Δ-Machine Learning approach.

We present a Δ-Machine Learning approach for the prediction of GW quasiparticle energies (ΔMLQP) and photoelectron spectra of molecules and clusters, using orbital-sensitive graph-based representations in kernel ridge regression based supervised learning. Coulomb matrix, Bag-

of-Bonds, and Bonds-Angles-Torsions representations are made orbital-sensitive by augmenting them with atom-centered orbital charges and Kohn–Sham orbital energies, which are both readily available from baseline calculations on the level of density-functional theory (DFT). We first illustrate the effects of different constructions of the orbital-sensitive representations (OSR) on the prediction of frontier orbital energies of 22K molecules of the QM8 dataset, and show that it is possible to predict the full photoelectron spectrum of molecules within the dataset using a single model with a mean-absolute error below 0.1 eV. We further demonstrate that the OSR-based Δ MLQP captures the effects of intra- and intermolecular conformations in application to water monomers and dimers. Finally, we show that the approach can be embedded in multiscale simulation workflows, by studying the solvatochromic shifts of quasiparticle and electron-hole excitation energies of solvated acetone in a setup combining Molecular Dynamics, DFT, the GW approximation and the Bethe–Salpeter Equation. Our findings suggest that the Δ MLQP model allows to predict quasiparticle energies and photoelectron spectra of molecules and clusters with GW accuracy at DFT cost.

5.1 Introduction

Fundamental insights gained by computational analysis of electronically excited states of molecular systems can help improving the design of molecular materials and plays therefore a vital role in material science. However, obtaining quantitative predictions is challenging as traditional methods either come at insufficient accuracy, e.g., due to the lack of correlation in interpreting orbital energies of Hartree–Fock, or at the price of high computational costs, as for coupled cluster methods, quantum Monte Carlo, or Green’s function approaches. Hence, the incorporation

of quantum machine learning (QML) has been gaining great traction over recent years. QML based surrogate property models have become a popular alternative approach for their fast, reliable, and accurate predictions of molecular and material properties [133, 135, 138, 153–164].

The main advantage of ML models is that they allow predictions of molecular properties with improved efficiency at a lower computational cost compared to traditional quantum chemistry approaches. Method development in the field of QML is progressing rapidly and it is increasingly influencing traditional methods [134, 138, 165–168]. Developments in molecular representations and QML models have paved the way for predicting energetic, electronic, and thermodynamic properties, such as atomization energies, dipole moments, polarizabilities, and harmonic frequencies [132, 137, 169].

QML of excited states of molecules has remained difficult in comparison. Recent work [168, 170, 171] has achieved promising results for predictions of single frontier orbital (highest or lowest molecular orbital) energies. However, some applications, such as the evaluation of direct or inverse photoelectron spectra, require predictions for a wider range of orbitals simultaneously, with sensitivity to conformational details of the actual molecules and/or a complex embedding environment. Circumventing the need to build separate models for each state of interest and the associated difficulties in finding unique characterizations of multiple orbitals across a wide range of chemical space can be achieved, for instance, using neural networks that learn the required information implicitly from orbital-free representations [157, 172], or by capturing capture both structural and orbital details in the representation.

In this chapter we show a way of augmenting existing molecular cartesian coordinates-based representations with orbital-specific infor-

mation from density-functional theory (DFT) which allows us to individually predict quasiparticle energies on the level of Many-Body Green's Functions Theory in the GW approximation and to calculate full photoelectron spectra of molecules and clusters with a single target kernel ridge regression Δ -Machine Learning model (Δ MLQP). We adopt the Δ -ML approach [132] as its concept of learning the corrections to a certain baseline property matches directly the way in which quasiparticle corrections are obtained perturbatively to the Kohn–Sham orbital energies, i.e., $\epsilon_i^{\text{QP}} = \epsilon_i^{\text{KS}} + \Delta\epsilon_i^{GW}$. Within Δ MLQP, we consider specifically the orbital-sensitive augmentation of Coulomb matrix (CM), Bag-of-Bonds (BoB), and Bonds-Angles-Torsions (BAT) representations by a combination of atom-centered orbital charges and Kohn–Sham orbital energies, which are all easily accessible from standard DFT ground state calculations [136, 137, 154, 158]. We illustrate the effect of different methodological choices, such as the determination of the orbital charges as Mulliken populations [173] or from a Gaussian Distributed Multipole Analysis [174] (GDMA), on the prediction of the full quasiparticle spectra of molecules in the QM8 dataset. To scrutinize that orbital-sensitive representations (OSR) with multiple orbitals are also capable of resolving the effects of intra- and intermolecular conformations, we also study the photoelectron spectra of water monomers and dimers, taken from the H₂O-13 dataset [175]. Finally, we study the use of the Δ MLQP predicted, conformational-sensitive quasiparticle energies embedded in calculations of electron-hole excitations on the level of the Bethe–Salpeter Equation (BSE) for the prototypical example of acetone in aqueous solution, and compare the solvatochromic shifts obtained from BSE@ GW and BSE@ML.

5.2 Methodological Background

5.2.1 Δ-Machine Learning

The typical workflow of a supervised machine learning based task is to map an input vector, e.g., a vectorized form of a representation of a molecular geometry, to a target property, such as an atomization or excitation energy, calculated with a high-accuracy method. In some cases, however, it is computationally infeasible to obtain such high-accuracy data for the large number of training/testing inputs required to achieve satisfactory learning results. As a consequence, such *direct* ML approaches can be limited to low-accuracy targets. To circumvent the problem of requiring large high-accuracy datasets, Δ-ML aims to predict the highly accurate target property at the same cost of the computationally cheaper methods, which is often referred to as the baseline property [176, 177]. This approach is typically more data efficient than direct ML, since the computationally expensive high-accuracy simulations are needed only for a considerably smaller subset to obtain a certain predictive power [131, 132].

The accurate target property is labeled as p^t and is obtained by

$$p^t(\mathbf{x}) = p^b(\mathbf{x}) + \Delta_b^t(\mathbf{x}) \quad (5.1)$$

$$= p^b(\mathbf{x}) + \sum_{n=1}^N \alpha_n k(\mathbf{x}, \mathbf{x}_n), \quad (5.2)$$

with p^b being the baseline property and α_n being the n -th regression coefficient of a KRR model that is trained to predict the difference between the target and baseline property, i.e. $p^t - p^b$. The Δ-ML model has been used in various applications and has shown to be powerful in not only

saving computational time, but also achieving much higher precision compared to traditional machine learning approaches [132, 177].

5.3 Results

We illustrate the applicability of the orbital-sensitive Δ machine learning model for predicting full quasiparticle spectra by reporting the predictive performance on QM8 molecules and water monomers/dimers for molecular cartesian coordinates-based representations. Subsequently, to obtain excitation energies of acetone and acetone in water, the proposed approach is used as a surrogate model to evaluate BSE. The results are then used to study solvatochromic shifts and benchmark it against experimental data.

5.3.1 Orbital-sensitive descriptors

Figure 5.1(a) shows a schematic overview of the Δ MLQP approach as a surrogate model to predict quasiparticle energies, thereby bypassing the computationally expensive first-principles GW step. It is based on the idea of learning the nonlinear transformations from Kohn–Sham energies to quasiparticle energies, motivated by the fact that for all orbitals i

$$\epsilon_i^{\text{QP}} = \epsilon_i^{\text{KS}} + \Delta\epsilon_i, \quad (5.3)$$

corresponds directly to the form of Eq. 5.1. This allows us to identify ϵ_i^{KS} with the baseline property $p^b(\mathbf{x})$ and to learn the quasiparticle corrections $\Delta\epsilon_i = \langle \phi_i^{\text{KS}} | \hat{\Sigma}(\epsilon_i^{\text{QP}}) - \hat{V}_{\text{xc}} | \phi_i^{\text{KS}} \rangle$ as in Eq. 2.24.

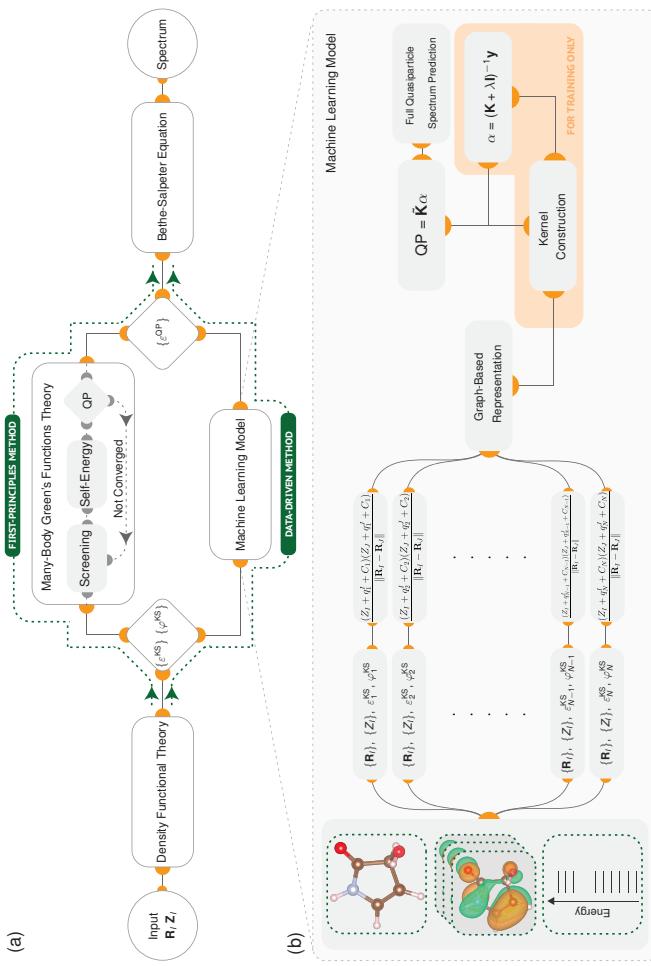


Figure 5.1: A schematic overview of first-principles and data-driven routes for the calculation of GW quasiparticle energies and their embedding into a $BSE@DFT-GW/ML$ workflow. (a) Atomic charges and atomic coordinates are used as input for DFT calculations. DFT generates Kohn-Sham energies and wavefunctions, which are being used as input for either quasiparticle calculations in the first-principles route or Δ -ML model in the data-driven route. The output of either route is then used as input for BSE calculations to output the excitation energy spectra. The machine learning block consist of two operations. (b) Details of the Δ MLQP model with orbital-sensitive representations. First, the DFT output together with geometric information is transformed into a molecular representation. Second, the resulting vector is used to build up a kernel matrix to predict Δ QP energies, which is then are added to KS energies and used as output.

To be able to replace all orbital-dependent $\Delta\epsilon$ by a single Laplacian kernel based (or any other) machine learning model requires a representation that incorporates orbital information. Traditional graph-based representations, like the Coulomb Matrix (CM) and BoB [137, 154, 158], which solely rely on the sets of atomic positions $\{\mathbf{R}\}$ and nuclear charges $\{Z\}$ lack such information and have therefore no injectivity for predicting full spectra of molecules.

We propose an extension to molecular cartesian coordinates-based representation that includes information about the Kohn–Sham orbital energies and wavefunctions, and will refer to the extended version as *orbital-sensitive representation* n-OSR, where n stands for the number of included orbitals. To this end, we map the $\{\varphi_k^{\text{KS}}(\mathbf{r})\}$ to a set effective orbital-dependent atomic charges $\{q_k\}$ and add those with its (rescaled) Kohn–Sham energy $C_k = \zeta \epsilon_k^{\text{KS}}$, where $[\zeta] = \text{e/Hartree}$, to the nuclear charges. For example, the Coulomb Matrix representation for a molecule at electronic state k can be extended as

$$\text{CM}_k = \begin{cases} \frac{(Z_I + q_k^I + C_k)(Z_J + q_k^J + C_k)}{\|\mathbf{R}_I - \mathbf{R}_J\|}, & \text{for } I \neq J \\ \frac{1}{2}(Z_I + q_k^I + C_k)^2, & \text{for } I = J. \end{cases} \quad (5.4)$$

A modification as in Eq. 5.4 allows us to introduce the missing injectivity for multi-state predictions within a single model, as indicated in Fig. 5.1(b). It is a particularly attractive choice as all ingredients for this modification are readily available from the DFT baseline calculations, and its very simple, physically interpretable form is easy to implement. The same idea of incorporating orbital-sensitivity can be applied to all \mathbf{R} and Z dependent representations, such as BoB or BAT. In the following, we will evaluate the above extension of molecular cartesian coordinates-

based representations, their dependence on the choice of different methods to obtain the $\{q_k\}$, in the use of ΔMLQP.

5.3.2 QM8 Molecules

We use molecular geometries from the QM8 data set, which contains more than 20000 synthetically feasible small organic molecules with up to eight CONF atoms [131, 169]. All quantum-mechanical calculations have been performed with the VOTCA-XTP package [50, 178]. For each molecule, we first perform DFT ground state calculations with the PBE0 hybrid functional [179], the def2-TZVP basis set [180], and an optimized auxiliary basis for the resolution-of-identity techniques [181]. Orbital-dependent atomic charges are determined from a Gaussian Distributed Multipole Analysis (GDMA) [174] or from Mulliken populations [173]. Eigenvalue selfconsistent ($evGW$) quasiparticle energies are then determined for the lowest $2N_{occ}$ states (excluding the core levels), where N_{occ} is the number of occupied levels. All orbitals are included in the RPA step and not explicitly corrected higher levels are scissors shifted according to the highest absolute quasiparticle correction among the explicitly corrected unoccupied orbitals. The frequency integration in Eq. 2.22 is performed using the FAA.

In Fig. 5.2 we report the performance measures, such as correlations and learning curves, for predicting individual HOMO energies (Fig. 5.2(a,d)), individual LUMO energies (Fig. 5.2(b,e)), and predicting simultaneously both HOMO and LUMO energies (Fig. 5.2(c,f)) of QM8 molecules with various Coulomb matrix (CM), bag-of-bonds (BoB), and bonds-angles-torsions (BAT) representations and extensions based on GDMA orbital charges. In all cases, here and in the following sections, we used $\sigma = 800$ and $\lambda = 10^{-8}$ in the KRR models, and $\zeta = 1$ e/Hartree. The hyperparam-

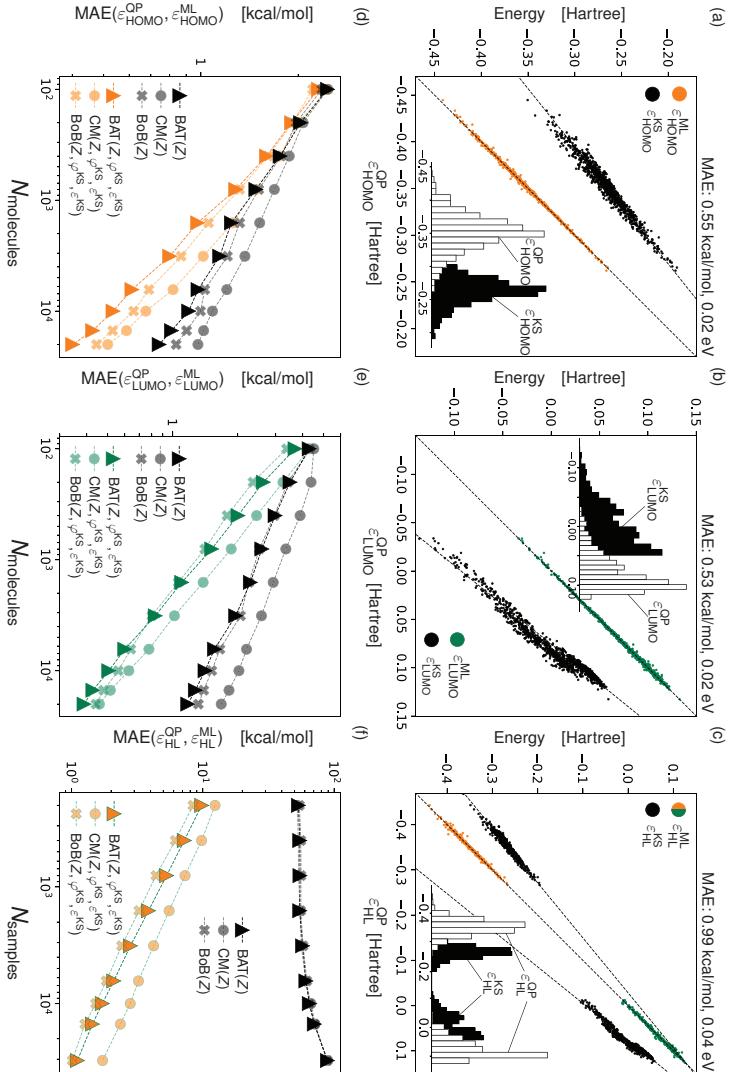


Figure 5.2: Correlation plots and learning curves for Δ -Machine Learning HOMO, LUMO, and HOMO-LUMO energies of QM8 molecules. (a), (b) Correlation plots for single orbital energy level predictions with 1-OSR models, where KS/QP HOMO and LUMO energies are visualized in black and Δ -ML/QP HOMO and LUMO energies in orange and green, respectively. (c) A correlation plot for the simultaneous prediction of HOMO and LUMO energies in an 2-OSR model, where KS/QP HOMO/LUMO energies are represented in black and Δ -ML in orange/green. The inset in each correlation plot shows the histogram of QP (white) and KS (black) energies. (d)-(f) HOMO, LUMO, and HOMO-LUMO Δ -ML learning curves for QP predictions with various orbital independent (black) and orbital dependent representations (orange, green, and orange/green).

eters were optimized for the QM8 dataset with all considered representations and for the later use cases, it was checked that variations of this optimum have negligible effect on the predictions.

The correlation and distribution plots display the expected nonlinear shift between KS and QP energies. As expected, the HOMO (LUMO) ϵ^{KS} are consistently above (below) the corresponding ϵ^{QP} . The 1-OSR Δ -ML models as in Fig. 5.2(a,b) are constructed using frontier orbitals of 18000 molecules as training set and 2000 as testing set and are based on the extended BAT representations. They transform KS energies to QP energies with a mean absolute error of 0.02 eV, respectively. The corresponding learning curves in Fig. 5.2(d,e) show systematic decay in error with increasing number of training samples N_{samples} (here, each sample corresponds to one molecule) for both standard and extended representations. However, it is also clearly visible that even for a 1-OSR model, the inclusion of information about the KS orbital energies and wavefunctions via the partial charges systematically improves the MAE at fixed training set size in all cases.

We proceed with discussing the first 2-OSR Δ MLQP model, trained on a mixed set of 30000 HOMOs and LUMOs, with testing performed on 5000. The results of a model simultaneously predicting HOMO and LUMO quasiparticle energies is shown in Fig. 5.2(c). Using orbital-sensitive BAT we note a combined MAE of 0.04 eV. The associated learning curves in Fig. 5.2(f) show clear improvements of the MAE with increasing number of samples. Note that the training set used to predict multiple states simultaneously contains for each molecule multiple states and the number of samples does therefore not correspond to the number of molecules. Additionally, unmodified representations fail in predicting both targets at the same time, as expected. Overall,

the orbital-sensitive representations are more data efficient and allow double-state predictions.

Based on the promising performance of the 2-OSR Δ -ML model, we apply the same method to predicting the full quasiparticle spectra of QM8 molecules in Multi-OSR Δ MLQP. To do this, we first shuffle the dataset comprising all considered quasiparticle levels of all molecules, and then select a random subset with 30000 samples as training and 5000 as testing set. In Fig. 5.3(a) the ML-QP correlation plot is shown based on application to 1000 out-of-samples orbitals. Good agreement of the distributions can be observed over an energy range covering around 2 Hartree, where a mean absolute error of 0.06 eV provides a good example of the predictive capabilities of the Multi-OSR Δ -ML model.

Figure 5.3(b) shows the learning curves for standard and orbital-sensitive representations. While orbital-sensitive representations allow us to predict the entire QP spectra of QM8 molecules, representations without orbital information fail to learn and hence we cannot report systematic improvement with increasing number of training samples. How this translates to a more practical example is shown in Fig. 5.3(c). A density of states (DOS) plot of a randomly chosen QM8 molecule (ID 014520 in the QM8 data set) based on KS, QP, and ML, respectively, is shown. The difference between the KS HOMO-LUMO gap and the QP HOMO-LUMO gap equates to approximately 4 eV, while the difference between the full QP spectra and ML spectra is 0.07 eV (MAE of all predicted levels). The DOS (with a Gaussian broadening of 0.022 Hartree) based on Δ MLQP (solid line) is practically indistinguishable from the one based on explicit QP energies (shaded area).

Finally, we compare in Tab. 5.1 the MAE of predictions of quasiparticle energies obtained with generic and different orbital-sensitive repre-

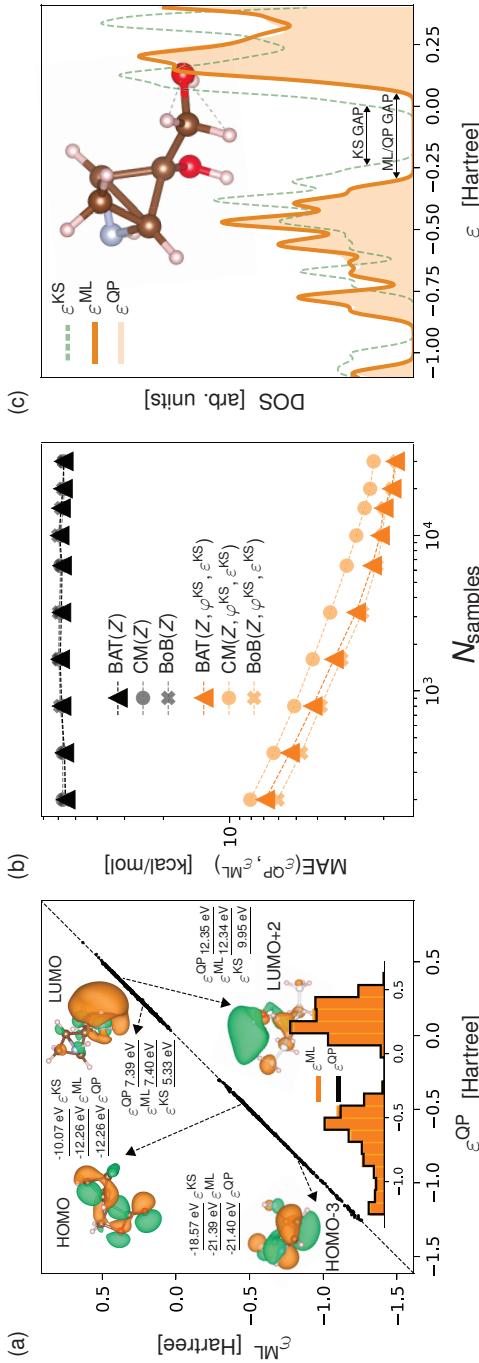


Figure 5.3: Multi-OSR Δ -ML of full quasiparticle spectra of QM8 molecules with a single model. (a) Correlation plot of Δ -ML/QP energies, where the arrows are pointing at randomly chosen molecules. The inset shows the histogram of ML (orange) and QP (black) energies. (b) Learning curves for QP energy predictions with various orbital independent (black) and orbital dependent representations (orange). (c) The density of states of a randomly chosen QM8 molecule, where the shaded area (light orange) represents the QP energies, while the green and orange lines describe the KS and ML energies, respectively.

Table 5.1: Mean Absolute Errors (in eV) of Predicting ev GW Quasiparticle Energies by Kernel Ridge Regression using standard CM, BoB and BAT representations, as well as our orbital-sensitive extension based on GDMA and Mulliken orbital charges, respectively. The best performing models are marked in bold.

	HOMO	LUMO	2-OSR	Multi-OSR
CM	0.043	0.073	—	—
BoB	0.036	0.057	—	—
BAT	0.032	0.051	—	—
CM-GDMA	0.022	0.020	0.046	0.104
BoB-GDMA	0.021	0.019	0.029	0.083
BAT-GDMA	0.017	0.017	0.034	0.089
CM-Mulliken	0.031	0.038	0.109	0.155
BoB-Mulliken	0.026	0.033	0.070	0.108
BAT-Mulliken	0.024	0.030	0.072	0.107

sentations. As also apparent from Fig. 5.2(d,e), the addition of orbital information into the representation reduces the MAE even for the single-orbital models by up to 50 %. Differences between CM, BoB, and BAT are very small in these cases. Regarding the use of different techniques to obtain effective orbital charges, we note that GDMA yields ~ 0.010 eV lower MAEs than Mulliken populations, and the overall lowest MAE is obtained for the BAT-GDMA combination. Moving to the 2-OSR- and Multi-OSR models, BoB-GDMA performs slightly better than BAT-GDMA and CM-GDMA. In the 2-OSR model comprising HOMO and LUMO, using Mulliken charges leads to a doubling of the MAE, albeit still lower than 0.1eV. For Multi-OSR, the MAE are expectedly a bit higher due to the high-dimensionality of the data but the differences between the use of GDMA and Mulliken charges are relatively smaller, in particular

for BoB and BAT representations. While the Mulliken population based orbital-sensitive representations appear to yield a slightly higher MAE as compared to the GDMA-based ones, it should be stressed that the latter come with a higher computational cost, noticeable in particular for larger systems (see also Section 5.3.4), and are not widely available for standard DFT applications.

5.3.3 Water Monomers and Dimers

In the QM8 example we focused on predicting the full QP spectra of single molecules. In this example we want to show that a single Δ-ML model can predict the full QP spectra of water dimers and monomers simultaneously, i.e., that ΔMLQP is sensitive to intermolecular conformations as well.

The molecular geometries used in this application originate from the H₂O-13 dataset [175] that consists of 2000 water dimers with O-O distances less than 4.5 Å obtained from an MD simulation. All *GW* properties used in the training and testing set were calculated as mentioned for the QM8 data in the previous section. To build a Multi-OSR ΔMLQP model, we first consider not only the 2000 dimer structures but also extract 4000 monomers from them. In water dimers (monomers) 18 quasiparticle states are taken into account, so the orbital dataset we have used contains 52000 samples in total. From this set, 30000 randomly selected samples are used for training and 5K for testing.

Fig. 5.4(a) shows the QP-KS correlation (black) and QP-ML correlation (orange/green), respectively. The Multi-OSR ΔMLQP approach based on BAT is able to transform the Kohn-Sham energies to quasiparticle energies with great accuracy (MAE 0.03 eV). In Fig. 5.4(b), we plot the DOS of a randomly chosen water dimer and its constituent

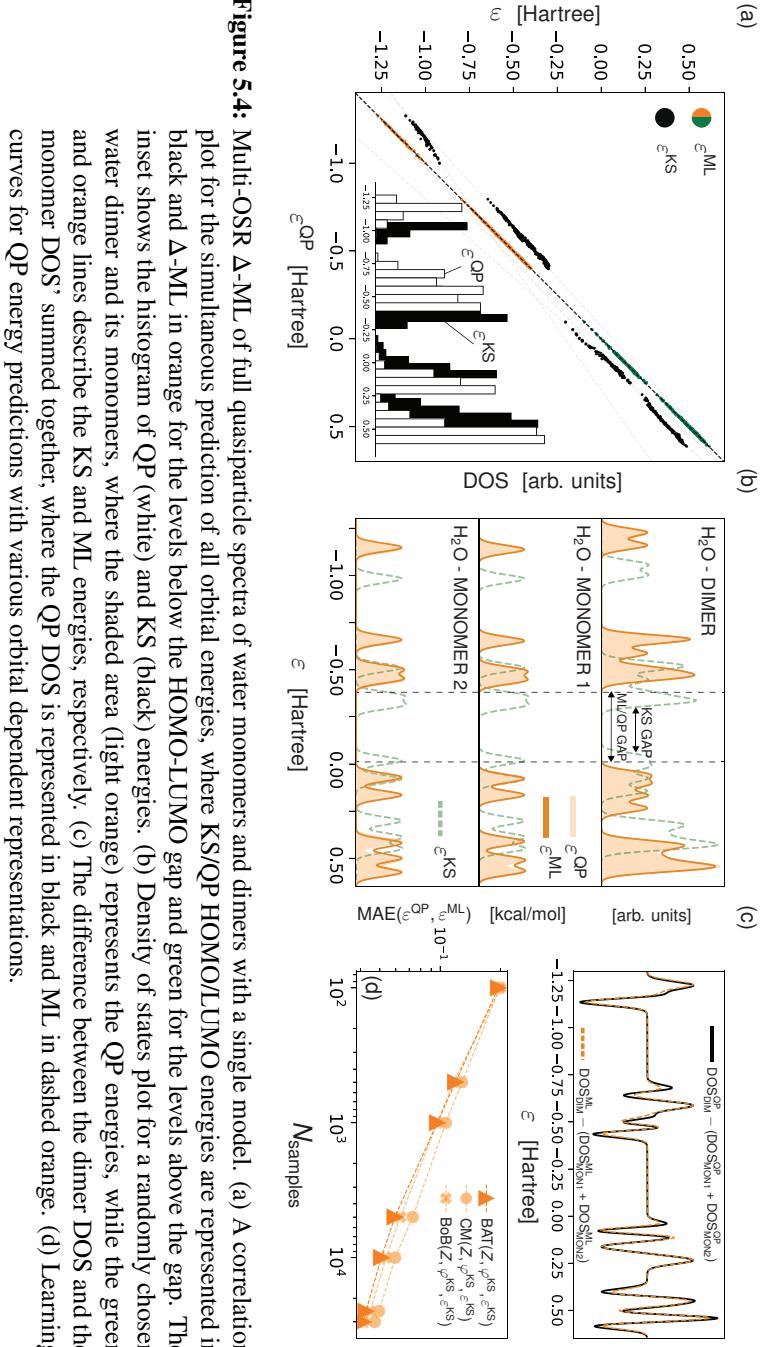


Figure 5.4: Multi-OSR Δ -ML of full quasiparticle spectra of water monomers and dimers with a single model. (a) A correlation plot for the simultaneous prediction of all orbital energies, where KS/QP HOMOLUMO energies are represented in black and Δ -ML in orange for the levels below the HOMO-LUMO gap and green for the levels above the gap. The inset shows the histogram of QP (white) and KS (black) energies, where the shaded area (light orange) represents the water dimer and its monomers, where the shaded area (light orange) represents the QP energies, while the green and orange lines describe the KS and ML energies, respectively. (c) The difference between the dimer DOS and the monomer DOS' summed together, where the QP DOS is represented in black and ML in dashed orange. (d) Learning curves for QP energy predictions with various orbital dependent representations.

monomers broadened by 0.022 Hartree. On the shown scale, the small differences among the two monomers are hard to distinguish, independent of the method. More importantly, it is apparent that the ΔMLQP captures the respective openings of the HOMO-LUMO gaps and energy dependent quasiparticle corrections, as well as the effects of intermolecular interactions in the dimer conformation. This point is emphasized by the analysis of the difference between the actual dimer DOS and a simple superposition of the two monomer DOS as in the top panel of Fig. 5.4(c), evaluated based on the explicit QP energies (solid black) and the one predicted by ΔMLQP (dashed). These differences reveal the shifts of the coupled dimer energy levels due to intermolecular interactions and the effects are captured by ΔMLQP at very good accuracy not only near the gap but even for, e.g., the deep O_{2s} levels. The lower panel in Fig. 5.4(c) shows monotonous decay of the prediction errors as a function of training set size, where orbital sensitive BoB and BAT perform slightly better than the orbital sensitive CM. Again, we see from the correlation plots, DOS and the learning curves indicate that the full QP spectra of water dimers and monomers can be accurately reproduced with just a single delta machine learning model.

5.3.4 Acetone in Water

We now consider an example of even more complex molecular clusters: aqueous acetone. The increased complexity stems from combining two different molecular species and considering more molecules in the clusters, leading to a very high-dimensional problem as the number of states and conformations increase dramatically. Specifically, the choice for aqueous acetone is motivated by the fact that it is known to exhibit a solvatochromic shift of the lowest coupled electron-hole excitation energy of ∼ 0.2 eV[182–184], a combined effect of similar shifts to the

individual quasiparticle energies and modified screening of the electron-hole interaction in water. From the perspective of our Δ MLQP model, this poses the additional question of whether its predictions are accurate enough in such a case to embed them into the calculation of the electron-hole excitation energies via the Bethe-Salpeter Equation (BSE@ML vs BSE@*GW*), as noted in the workflow scheme in Fig. 5.1(a).

To answer this question, we first generate structural data by performing classical Molecular Dynamics simulations of a single acetone in 219 water molecules using an OPLS-AA type forcefield for acetone, automatically generated by LigParGen [185], and the TIP3P model for water [186]. Geometric mixing rules for Lennard-Jones diameters and energies were used for atoms of different species [187]. Non-bonded interactions between atom pairs within a molecule separated by one or two bonds were excluded. Interaction was reduced by a factor of 1/2 for atoms separated by three bonds and more. Simulations were run using GROMACS version 2019.6 [188]. A 0.9 nm cutoff was employed for the real space part of electrostatics and Lennard-Jones interactions. The long-range electrostatics were calculated using particle-mesh Ewald (PME) [189] with the reciprocal-space interactions evaluated on a 0.18 nm grid with cubic interpolation of order 4. An initial configuration was prepared in a cubic box of size 2 nm and energy minimized using the steepest descent algorithm, followed by a 6 ns simulation in constant particle number, volume and temperature (NVT) ensemble. Temperature was kept constant at 300 K using the stochastic velocity rescaling thermostat [190] with time constant 0.5 ps, and the velocity-Verlet algorithm [191] was employed to integrate the equations of motions with a 1 fs time step. Simulations were then continued for 200 ns in constant particle number, pressure and temperature (NpT) ensemble at 300 K and 1 bar controlled by the Berendsen [192] barostat

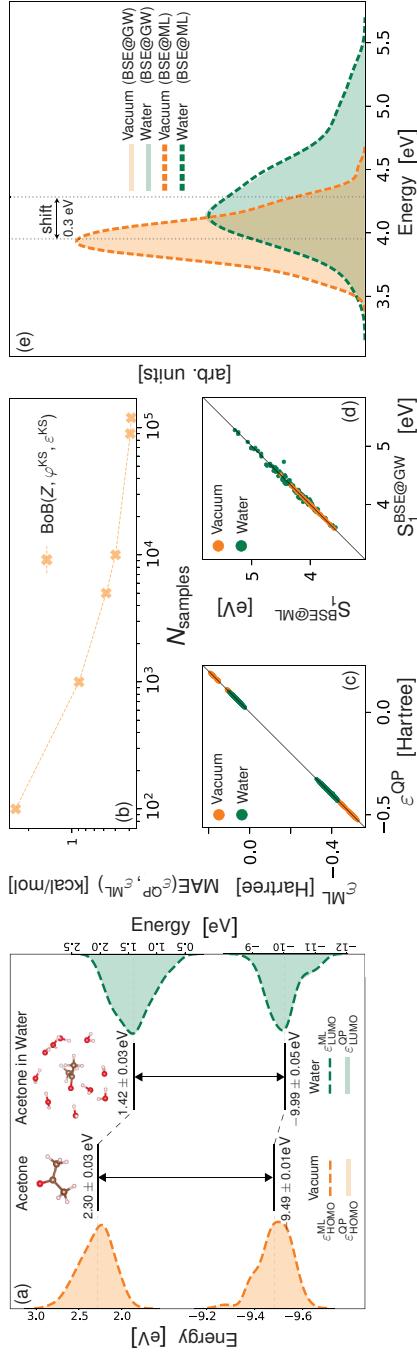


Figure 5.5: Multi-OSR Δ -ML of full quasiparticle spectra of acetone and acetone in water with a single model and application to calculating $n \rightarrow \pi^*$ excitation energies with BSE. (a) The black bars are a visual representation of the HOMO-LUMO gap in acetone in vacuum (orange) and acetone in water (green) with ML predictions. The shaded areas represent the HOMO and LUMO distributions as resulting from explicit $\text{ev}GW$ calculations, while the one resulting from the 2-OSR Δ MLQP model are shown as dashed lines. (b) Prediction errors as a function of training set size for the 8-OSR-BoB model to be used in BSE@ML calculations. (c) QP-ML quasiparticle energy correlation for 8-OSR-BoB. (d) Correlation between S_1 ($n \rightarrow \pi^*$) energies from BSE@GW and BSE@ML. (e) S_1 energy distributions of a single acetone (orange) and acetone in water (green) from BSE@GW and BSE@ML.

with a coupling time constant of 2.0 ps. From the last 100 ns of this run, 10000 snapshots at a time interval of 10 ps are extracted and clusters containing acetone and the ten water molecules closest to it are selected for the GW -BSE calculations. With this choice we ascertain that the first solvation shell is included in the cluster, contributing the strongest to the expected solvatochromic shifts. Note that we are not strictly targeting quantitative accuracy of the actual excitation energies compared to experiment with this study, but to demonstrate the internal consistency between BSE@ GW and BSE@ML in our Multi-OSR Δ MLQP approach.

GW -BSE calculations are performed on the selected clusters without and with the water molecules included, following the same protocol as before, with the exception of the treatment of the frequency-dependence in Eq. 2.22, for which we employ here a two-parameter generalized Plasmon-Pole Model [52, 53, 193]. Explicit quasiparticle corrections are determined for the four highest occupied and four lowest unoccupied molecular orbitals, while the full single-particle spectrum is included in the formation of the product basis for the BSE.

We first construct as a proof-of-concept a 2-OSR-BoB model including HOMO and LUMO in the absence and presence of a water solvation shell. From the total of 80000 samples, we select 5000 for training and 1000 for testing, ensuring that the sets have an equal amount of data for both HOMO and LUMO, with and without water, respectively. The energy distributions as obtained from the explicit calculations of ϵ^{QP} in vacuum (in water), calculated on 400 (100 of each case) out of sample data points, are shown as filled orange (green) areas in Fig. 5.5(a). Indicated are also the means and their error, showing a distinct lowering of 0.50 eV and 0.92 eV for the HOMO and LUMO, respectively, in the presence of water, thereby decreasing the HOMO-LUMO gap by 0.42 eV. Compar-

ing the respective distributions obtained from the 2-OSR Δ MLQP model shown as dashed lines in Fig. 5.5(a), hardly any differences can be observed.

As mentioned above, the determination of electron-hole excitation energies with BSE@*GW* requires inclusion of eight explicitly corrected orbitals near the gap. For this purpose, we now build an 8-OSR Δ MLQP model, for which Fig. 5.5(b) shows the learning curves. With more than 10K samples, a MAE of lower than 0.5 kcal/mol can be achieved. Figure 5.5(c) shows the corresponding correlation between explicitly calculated QP energies, and the ones from 8-OSR Δ MLQP, for which we have selected from the full dataset comprising 320K orbitals, 120K for training and 10K for testing. Clearly, the Δ MLQP model again provides excellent predictions for the eight different orbital energies in vacuum and solution, respectively. In the following step, these predicted quasiparticle energies for the eight explicitly corrected orbitals are used as input for the BSE. The remaining occupied (unoccupied) single-particle energies are scissors shifted according to the highest absolute quasiparticle correction among the explicitly corrected occupied (unoccupied) orbitals, as in the BSE@*GW* reference. In Fig. 5.5(d) we show the correlation between the determined S_1 energies of the $n \rightarrow \pi^*$ transition. Two interesting aspects should be noted: the subset of results for BSE@ML for the vacuum structure (orange) appear to agree better than the one for the acetone-water clusters (green), which is not surprising due to the bigger conformational space of the latter. More importantly, however, one can clearly see a systematic shift of the S_1 energies in aqueous solution to higher energies compared to in vacuo – equally obtained for both BSE@*GW* and BSE@ML – qualitatively in line with the experimental observations.

Finally, we show in Fig. 5.5(e) distributions of the $n \rightarrow \pi^*$ excitation energies of acetone as obtained by the BSE@*GW* (filled curves) and BSE@ML (dashed lines) approaches in vacuum (orange) and in water (green), respectively. It is evident that both methods predict both a broadening of the distribution upon solvation, as well as a shift to higher energies. Differences of the distributions on BSE@*GW* and BSE@ML levels are minuscule. The predicted solvatochromic shift of the mean of the distributions amount to 0.30 eV, as indicated by the dashed lines in Fig. 5.5(e). From peak-to-peak, the shift is 0.13 eV.

5.4 Summary

We have introduced orbital-sensitive augmentations of molecular cartesian coordinates-based representations in Δ -machine learning of full quasiparticle excitation energies in molecules and clusters. The proposed Δ MLQP approach is capable of predicting the *GW* energies of multiple orbitals across multiple molecules and/or intra- and inter molecular conformations in a single kernel ridge regression based supervised learning model. We have demonstrated this in application to the QM8 molecular dataset and to water monomers and dimers. Furthermore, it has been shown that a single orbital-sensitive Δ -ML model for quasiparticle energies can be embedded in multiscale simulation workflows, showcased in the prediction of solvatochromic shifts of excitation energies in aqueous acetone.

Having explored the potential of orbital-sensitive representations in the realm of Δ -machine learning, we have seen how intricacies of molecular structures can be adeptly captured and translated into meaningful predictions for quasiparticle energies. However, the landscape of molec-

ular prediction is vast, and there remains an ever-growing potential to further our understanding using other advanced machine learning methodologies. In particular, deep learning paradigms, such as feed-forward neural networks (FFNNs), offer another view to molecular interactions, especially when delving into electronic dynamics and couplings. While kernel ridge regression brought us closer to the molecular intricacies, FFNNs will be used in the next chapter to navigate the molecular nanomaterial dynamics.

Chapter 6

An Evolutionary Approach to Constructing a Deep Feedforward Neural Network for Prediction of Electronic Coupling Elements in Molecular Materials

This chapter is based on the paper: O. Çaylak, A. Yaman, and B. Baumeier, *J. Chem. Theory Comput.* **15**, 1777 (2019).

We present a general framework for the construction of a deep feed-forward neural network (FFNN) to predict distance and orientation dependent electronic coupling elements in disordered molecular materials. An evolutionary algorithm automatizes the selection of an optimal architecture of the artificial neural network within a predefined search space. Systematic guidance, beyond minimizing the model error with stochastic gradient descent based backpropagation, is provided by simultaneous maximization of a model fitness that takes into account ad-

ditional physical properties, such as the field-dependent carrier mobility. As a prototypical system, we consider hole transport in amorphous tris(8-hydroxyquinolinato)aluminum. Reference data for training and validation is obtained from multiscale ab initio simulations, in which coupling elements are evaluated using density-functional theory, for a system containing 4096 molecules. The Coulomb matrix representation is chosen to encode the explicit molecular pair coordinates into a rotation and translation invariant feature set for the FFNN. The final optimized deep feedforward neural network is tested for transport models without and with energetic disorder. It predicts electronic coupling elements and mobilities in excellent agreement with the reference data. Such a FFNN is readily applicable to much larger systems at negligible computational cost, providing a powerful surrogate model to overcome the size limitations of the ab initio approach.

6.1 Introduction

Dynamics of electronic excitations drive the functionality of molecular nanomaterials in many energy applications, e.g., in organic photovoltaics, photocatalysis, thermoelectricity, or energy storage. These dynamics are governed not only by the chemical structure, architecture, or electronic structure of molecular building blocks but also by the local and global morphology of the materials and molecular interactions on the mesoscale [194–196]. It is essential to understand how elementary dynamic processes, such as electron or energy transfer, emerge from an interplay of morphology and electronic structure. Such fundamental insight will eventually allow for controlling the above processes and pave the way for a rational design of molecular materials. While macroscale

information can be obtained experimentally, zooming into the electronic dynamics at an (sub)atomic scale is nearly impossible [197].

Computational modeling of, e.g., charge dynamics can provide valuable insight in this situation. The Gaussian disorder model and its various extensions [198–201] have been used to study general aspects of transport, such as temperature or carrier density dependence [202, 203]. For material specificity, they require information about the width of the density of states, which is typically obtained by fitting to macroscale or device-scale observables, for instance, a current-voltage curve. These more descriptive models cannot provide detailed information about underlying intermolecular processes.

In contrast, bottom-up simulations of charge and exciton dynamics in large-scale morphologies aim to explicitly zoom in to elementary charge transfer reactions at molecular level and predict the mesoscale charge dynamics using multiscale strategies which link quantum and supramolecular scales [204–206]. Such approaches have shown a remarkable level of predictiveness [196, 207–209] but come at the price of high computational costs. They typically involve the determination of bi-molecular electron transfer rates in explicit material morphologies, which in turn requires the calculation of intermolecular electronic coupling elements [210, 211], or transfer integrals, of the form

$$J_{ij} = \langle \Psi_i | \hat{H} | \Psi_j \rangle , \quad (6.1)$$

where $|\Psi_i\rangle$ and $|\Psi_j\rangle$ are diabatic states of molecules i and j , respectively, and \hat{H} is the Hamiltonian of the coupled system. Practical evaluation of Eq. (6.1) relies on quantum-mechanical information about the relevant electronic states of the two individual monomers, as well as of the dimer formed by them. Based on density-functional theory (DFT)

the necessary calculations for a morphology consisting of a few thousand molecules of moderate size can consume several hundreds of days of CPU time, even with techniques optimized for large-scale systems. Many relevant materials or processes, e.g., the system-size dependence of carrier mobility in dispersive transport, realistic carrier densities, or disordered interfaces in heterojunctions, can only be studied using significantly larger systems that are currently not accessible to multiscale models.

Surrogate stochastic models have been developed to overcome some of these limitations [212, 213]. They represent the molecular morphology by (random) point patterns and assign coupling elements between them by drawing from distributions with distance-dependent means and standard deviations, fitted to microscopic data. These models manage to reproduce, e.g., the field-dependence of the mobility stochastically, i.e., averages over several realizations are required to obtain a comparable behavior. While the generation of larger surrogate systems is computationally inexpensive, information about the molecular details is lost and the models are not transferable to interfaces or heterostructures.

In this chapter, we develop an alternative surrogate model which allows application to system sizes currently inaccessible to the multiscale ab initio approach while retaining its molecular level details. Focus is on removing the computational bottleneck associated with the explicit quantum-mechanical evaluations of electronic couplings using Eq. (6.1) with the help of Machine Learning (ML). ML has attracted considerable interest as a tool to save computational costs in large scale calculations [214] or in exploring chemical space, e.g., by predicting material properties [215–218]. Different models differ by the representation of the molecular information (features) and the choice of a suitable ML al-

gorithm, and their combinations have been optimized accordingly [216, 219].

For our goal of predicting electronic coupling elements, an appropriate data representation must accurately take distance and mutual orientations between two molecules of the same type, as given by the explicit atomic coordinates, into account. The machine learning algorithm must be capable of reliably predicting values of J_{ij} that can easily span several orders of magnitude, in particular in amorphous molecular materials. In this situation we turn to biologically inspired computational models known as artificial neural networks (ANNs) [220]. However, the construction of an appropriate network architecture is not trivial and may require a trial-and-error approach. We deal with this problem in a systematic way by using search algorithms such as evolutionary algorithms (EA) [221]. The advantage of using an EA approach to constructing a neural network is that it not only minimizes the model error but is also capable of taking into account additional physical principles providing systematic guidance to designing architectures.

Specifically, we employ such an evolutionary method to construct a multi-layered (deep) feedforward neural network (FFNN) for the prediction of electronic coupling elements based on the Coulomb Matrix representation[219] of molecular orientations. As a prototypical system, we consider an amorphous morphology of tris(8-hydroxyquinolinato)aluminum (Alq_3). An ab initio model of hole transport explicitly determined for a system containing 4096 molecules serves as a reference for the training of the FFNN. The electric-field dependence of the hole mobility is used as additional fitness parameter in the evolutionary algorithm. The final FFNN model provides inexpensive predictions of J_{ij} with which hole mobilities are obtained in excellent agreement with the ab initio data, both without and with energetic disorder. We demonstrate that it is

readily applicable to systems of larger size containing 8192 and 32768 molecules, respectively.

This chapter is organized as follows: In Chapter 6.2, we briefly recapitulate steps in the multiscale ab initio model to obtain the reference data. Details about the data representation and processing, the evolutionary approach for constructing the feedforward neural network including the definition of its fitness are given in Chapter 6.3. Validation of the model and prediction results are discussed in Chapter 6.4. A brief summary concludes the chapter.

6.2 Multiscale Ab initio Model

In what follows, we briefly summarize the steps of the multiscale model of charge transport in disordered molecular materials, needed to generate the ab initio reference for the FFNN. More in-depth discussion of the procedures can be found in Ref. [204]. The starting point is the simulation of an atomistic morphology using classical Molecular Dynamics (MD). 4096 Alq₃ molecules are arranged randomly in a cubic box, which is then first equilibrated above the glass transition temperature in an NpT ensemble at $T = 700\text{ K}$ and $p = 1\text{ bar}$, and subsequently annealed to $T = 300\text{ K}$. The Berendsen barostat [222] with a time constant of 2.0 ps and the velocity rescaling thermostat [223] with a time constant of 0.5 ps are used throughout. All calculations make use of an OPLS-based force field specifically developed [224] for Alq₃ and are performed using Gromacs [225].

In the obtained room-temperature morphology, the centers of mass of the molecules define the hopping sites for charge carriers. Pairs of molecules for which any intermolecular distance of the respective 8-

hydroxyquinoline ligands is less than 0.8 nm are added to a neighborlist of possible charge transfer pairs. Transfer rates between two molecules i and j are calculated within the high-temperature limit of non-adiabatic transfer theory [226] using the Marcus expression, which is given by

$$\omega_{ij} = \frac{2\pi}{\hbar} \frac{J_{ij}^2}{\sqrt{4\pi\lambda_{ij}k_B T}} \exp\left[-\frac{(\Delta G_{ij} - \lambda_{ij})^2}{4\lambda_{ij}k_B T}\right], \quad (6.2)$$

where \hbar is the reduced Planck constant, T the temperature, λ_{ij} the reorganization energy, k_B the Boltzmann's constant, ΔG_{ij} the free energy difference between initial and final states, and J_{ij} the electronic coupling element, as defined in Eq. (6.1). A hole reorganization energy of 0.23 eV is obtained from DFT calculations with the B3LYP functional [227] and a triple- ζ basis set [228]. Site energies E_i are determined from a classical atomistic model that takes into account the effects of local electric fields and polarization in the morphology [229, 230]. Their distribution is approximately Gaussian with $\sigma = 0.20$ eV. The influence of an externally applied electric field \mathbf{F} is added to the site energy difference as $\Delta G_{ij} = \Delta E_{ij} + \Delta E_{ij}^{\text{ext}} = \Delta E_{ij} + e\mathbf{F}\mathbf{r}_{ij}$, where \mathbf{r}_{ij} is the distance vector between molecules i and j . Electronic coupling elements are calculated using a dimer-projection technique based on DFT [211] with the Perdew-Burke-Ernzerhof functional [231] and the triple- ζ basis. All DFT calculations are performed with the Orca software package [232], while the VOTCA package[204, 233] is used for all charge transport related steps.

The molecular centers of mass and the hopping rates between the molecules can be interpreted as the vertices and edges of a weighted directed graph. In a system with only one charge carrier, the time evolution of the occupation probabilities P_i are described by the Kolmogorov

forward equation (Master equation)

$$\frac{dP_i}{dt} = \sum_j P_j \omega_{ji} - \sum_j P_i \omega_{ij}. \quad (6.3)$$

However, in this work, we are interested in a system that is in a steady state. This restriction allows us to write Eq. (6.3) as

$$\sum_j P_j \omega_{ji} - \sum_j P_i \omega_{ij} = 0 \Rightarrow \mathbf{W}\mathbf{p} = 0. \quad (6.4)$$

Here, the matrix \mathbf{W} can be constructed from the Marcus rates ω_{ij} . The field-dependent mobility $\mu(\mathbf{F})$ can be obtained from steady state occupation probabilities via the relation

$$\langle \mathbf{v} \rangle = \sum_i \sum_j P_j \omega_{ji} (\mathbf{r}_i - \mathbf{r}_j) = \mu(\mathbf{F}) \mathbf{F}, \quad (6.5)$$

where $\langle \mathbf{v} \rangle$ is the average velocity.

6.3 Machine Learning Model

6.3.1 Data representation

Explicit structural information of molecular pair geometries is extracted from MD simulations and used to construct the features of the dataset. Featurization is the process of encoding molecules into vectors, where each vector gets a label, in this case $\log_{10}[(J_{ij}/\text{eV})^2]$.

Coupling elements between molecular pairs are translation and rotation invariant, which is not accounted for in the plain atom coordinates \mathbf{R}_i . The Coulomb matrix (CM) [216, 219, 234] representation is capa-

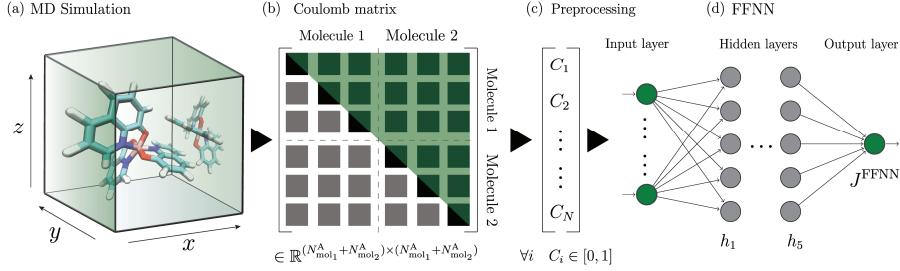


Figure 6.1: Overview of the data flow from raw molecular dynamics information to a neural network input. (a) Explicit atomic coordinates of molecular pairs is extracted from an MD snapshot. (b) A symmetric Coulomb matrix with dimension given by the sum of the number of atoms per molecule $(N_{\text{mol}_1}^A + N_{\text{mol}_2}^A)^2$ is constructed. (c) To only keep relevant and non-redundant information in the vectorized form of the Coulomb matrix, preprocessing techniques such as feature selection (upper triangle) and data scaling (to $[0, 1]$) are introduced. (d) The final vectorized CM enters a feedforward neural network with the hidden layers h_1, \dots, h_5 to predict the electronic coupling elements J_{ij}^{FFNN} .

ble of capturing this invariance and is used in the following to encode crucial information into the dataset's features. For every molecular pair the entries c_{ij} of the corresponding Coulomb matrix \mathbf{C} are computed according to

$$c_{ij} := \begin{cases} \frac{1}{2} Z_i^{2.4} & \text{if } i = j, \\ \frac{Z_i Z_j}{\|\mathbf{R}_i - \mathbf{R}_j\|} & \text{if } i \neq j, \end{cases} \quad (6.6)$$

where Z_i is the nuclear charge of atom i . Figure 6.2 illustrates the building blocks of the CM representation applied to a pair of molecules. One obtains a symmetric matrix that consists of four equally sized block matrices. The upper-right block indicates the inter-molecular orientations, whereas the upper-left and lower-right matrices represent the intra-molecular conformations.

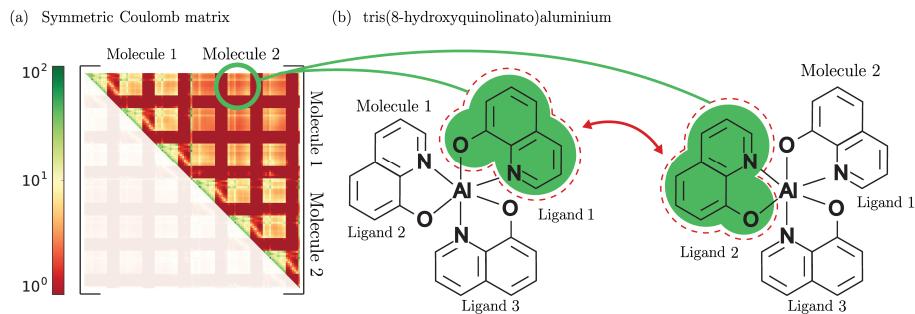


Figure 6.2: Visualization of a Coulomb matrix for molecular pairs. (a) The Coulomb matrix contains values corresponding to the inter- and intramolecular electrostatic interactions. Lowest values (dark red) correspond to the relations of hydrogen atoms, whereas interactions among heavy atoms of the ligands lead to values near 10 (yellow). (b) Schematic representation of how the arrangement of two Alq_3 molecules is encoded in specific regions of the upper right block of the CM.

Before being usable as input for the FFNN, the calculated CMs must be preprocessed, as illustrated in Figure 6.1(b,c). This step involves the removal of the lower triangular entries including the diagonal elements, and scaling of the values to the interval $[0, 1]$. Subsequent vectorization yields the instances of the descriptor space as input to an artificial neural network.

6.3.2 Deep Feedforward Neural Networks and Evolutionary Algorithms

Artificial neural networks consist of a number of artificial neurons typically arranged as layers with specific connectivity referred as topology. Among the ANNs, feed forward neural networks arrange a certain number of consecutive layers of neurons where each neuron in each layer is directionally connected to all neurons within the next layer. The ac-

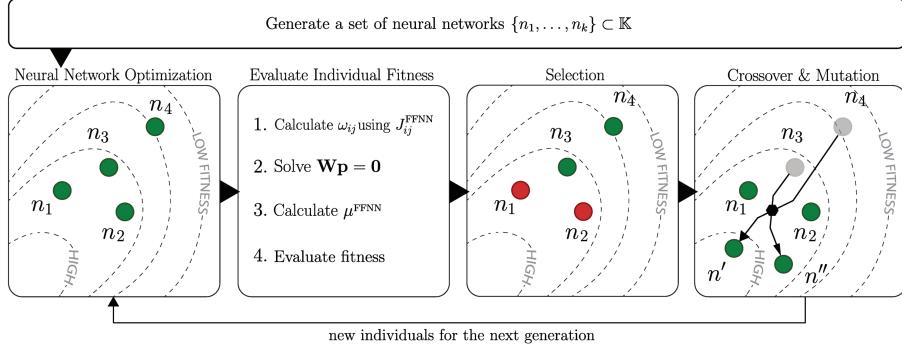


Figure 6.3: Flow chart of the evolutionary algorithm used for optimizing the deep FFNN. The initialization process consists of generating $k = 30$ arbitrary FFNN architectures $\{n_1, \dots, n_k\}$. (a) The weights of the networks are then optimized with a backpropagation algorithm. (b) For every optimized network n_i the predicted electronic coupling elements J_{ij}^{FFNN} are used in Eq. (6.2) to determine the matrix \mathbf{W} . After solving the stationary Master equation (Eq. (6.4)) and calculation of the field-dependent mobility, the fitness of the architectures is evaluated. (c) The architectures are then ordered based on their fitnesses and selected according to the roulette wheel principle. (d) After applying the crossover and mutation operators a new generation of feedforward neural networks is generated and the whole process is repeated until one of the stopping criteria is satisfied.

tivation a_i^m of neuron i in layer m is computed from the activations of neurons in the $m - 1$ -th layer according to

$$a_i^m = f \left(\sum_{j=0} v_{ij}^{m-1,m} \cdot a_j^{m-1} \right), \quad (6.7)$$

where $v_{i,j}$ is the weight of the connection between the neurons and f is an activation functions. For our purposes this activation function is given by a sigmoid function.

One of the conventional ways of training the FFNNs is the backpropagation algorithm with stochastic gradient descent. However, the number

of layers and the number of neurons per each layer should be defined before the training. These parameters are referred as hyper-parameters and play an important role in the performance of the networks. Although there are some “rule of thumb” guidelines established based on empirical studies, the selection of a proper set of hyper-parameter settings may require lots of expert knowledge and/or trial-and-error. This can be avoided by search algorithms like Genetic Algorithms (GAs). GAs are population based global search algorithms inspired by biological evolution [221]. The research field known as Neuroevolution employs evolutionary computing approaches to optimize artificial neural networks [235]. Adopting the terminology from biology, the genetic material of a population of individuals encode parameters of the ANNs. The encoding depends on the parameters of the ANNs to be optimized (topology and/or weights) [236, 237]. The workflow of such a genetic algorithm is shown schematically in Figure 6.3. It starts with randomly initializing a population of individuals (Figure 6.3(a)), where each individual is evaluated and assigned a fitness value (Figure 6.3(b)) to measure its performance. The main part of the algorithm performs selection, crossover, and mutation operations aimed at iteratively improving the fitness values. The selection operator (Figure 6.3(c)) selects the individuals with better fitness values to construct a next generation of individuals. One of the most commonly used selection operators is roulette wheel selection, in which individuals are selected with a probability proportional to their fitness values. The stochasticity of this selection process may occasionally cause the best individuals to disappear from the population. It can be combined with the elitist selection scheme, which selects the top ℓ best ranked individuals, such as n_1 and n_2 in Figure 6.3(c), and transfers them unchanged directly to the next generation. The crossover operator combines two individuals selected by the roulette wheel operator (parents, n_3 and n_4 in Figure 6.3(d)), to gener-

ate two new individuals (offspring, n' and n''). In particular, the 1-point crossover operator selects, with a probability of p_c , a random point to copy two different parts of two parents to generate offspring. Subsequently, a mutation operator flips the bit value in components of the binary representation of the offspring individuals randomly with a small probability p_m . Overall, the GA is run for a certain number of iterations or until a satisfactory solution, defined, e.g., by a threshold fitness value, is found.

6.3.3 Construction of a deep FFNN for prediction of electronic coupling elements

In this work, we use a genetic algorithm to optimize the topology parameters (the number of hidden layers and the number of neurons in each hidden layer) of the feedforward neural networks. Each individual $n_i \in \mathbb{K}$ in the population is represented as five dimensional strings, where

$$\mathbb{K} = \{n := (i_1, \dots, i_5) \mid i_j \in \mathcal{N} := \{0, 50, 100, \dots, 1000\}\}, \quad (6.8)$$

which encodes an ANN topology. Therefore, the maximum number of hidden layers a network can have is set to 5, and the number of neurons each hidden layer is taken from the set of 21 discrete values. We limit the FFNN topologies in this manner to reduce the search space and computational complexity. Consequently, our genetic algorithm aims to find the optimum model settings in $21^5 = 4084101$ number of possible networks.

The FFNNs are trained on a training dataset using backpropagation to minimize the error between the target $\log_{10}[(J_{ij}^{\text{DFT}}/\text{eV})^2]$ (actual la-

bels of the input data) and predicted outputs $\log_{10}[(J_{ij}^{\text{FFNN}}/\text{eV})^2]$. We use three distinct snapshots extracted from different MD simulations for training and validation. Each snapshot contains 4096 Alq₃ molecules with approximately 24000 pairs in the neighbor list as described in Chapter 6.2. The first snapshot is used to optimize the weights of a chosen neural network, while the second snapshot is used for selecting the neural network architectures based on their fitness values. The third dataset is used to analyze the performance of the constructed final neural network on completely unseen data.

The fitness value of a given feedforward neural network architecture is determined by evaluating the mean squared error of the difference between the mobility μ^{FFNN} and the reference mobility μ^{DFT}

$$\Xi = \left(\frac{1}{N_F} \sum_{k=1}^{N_F} (\mu_k^{\text{DFT}} - \mu_k^{\text{FFNN}})^2 \right)^{-1}, \quad (6.9)$$

where N_F stands for the number of field values. For each $|\mathbf{F}| = b \cdot 10^7 \text{ V/m}$ with $b \in \{3, 4, 5, 7, 9\}$, the mobility is obtained as an average over $\pm x, \pm y, \pm z$ directions of the applied field.

Our GA starts with randomly initializing 30 individuals and evaluating the fitness of the constructed FFNNs, respectively, as illustrated in Figure 6.3. We use the roulette wheel selection with the elitism of $\ell = 2$ and standard 1-point crossover operators with the probability of $p_c = 1$. The mutation operator selects a random component of a string with a probability $p_m = 0.1$, and replaces it with randomly selected value from \mathcal{N} . The probability of selecting $0 \in \mathcal{N}$ for the mutation is 0.3, and the rest of the values shares 0.7 with equal probabilities to encourage smaller networks. The GA was run until there is no fitness improvement for 20 generations.

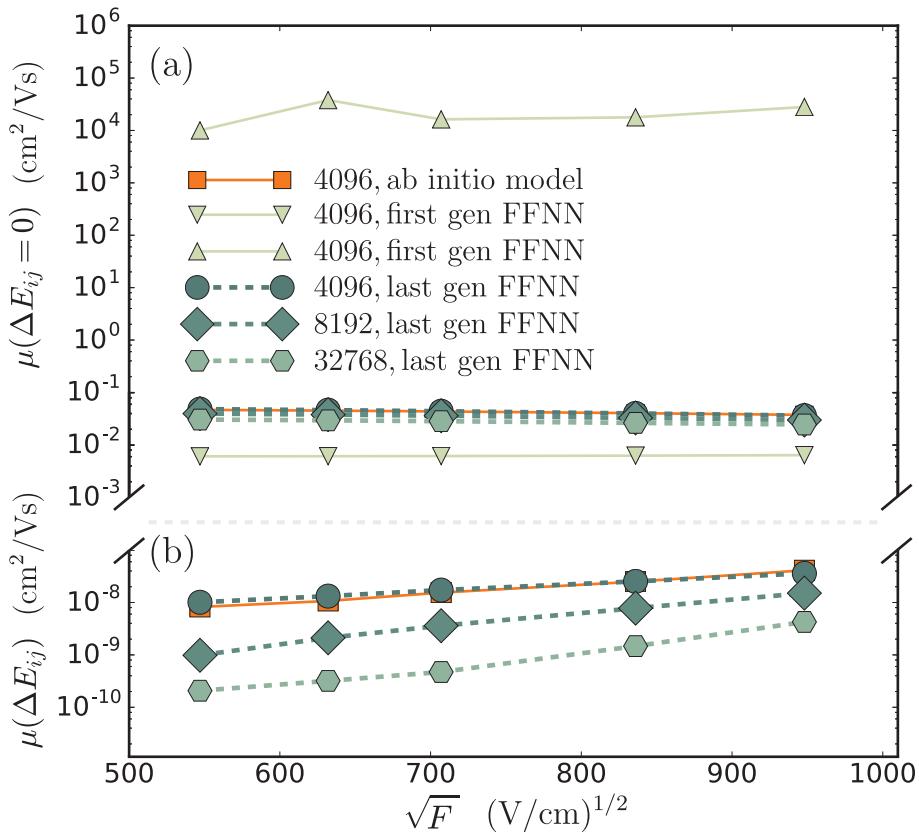


Figure 6.4: Field-dependent hole mobility (Poole-Frenkel plot) of Alq_3 , for systems containing 4096, 8192 and 32768 molecules. In (a) the mobility μ for the disorder-free case, i.e. $\Delta E_{ij} = 0$, is given, whereas (b) illustrates the mobility μ for the case with disorder, i.e., $\Delta E_{ij} \neq 0$.

6.4 Results

In Figure 6.4 we show the calculated field-dependence of the hole mobility in Alq_3 for various different models (a) without ($\mu(\Delta E_{ij} = 0)$) and (b) with ($\mu(\Delta E_{ij})$) energetic disorder taken into account in Eq. (6.2). The data indicated by the orange squares has been obtained by the ab

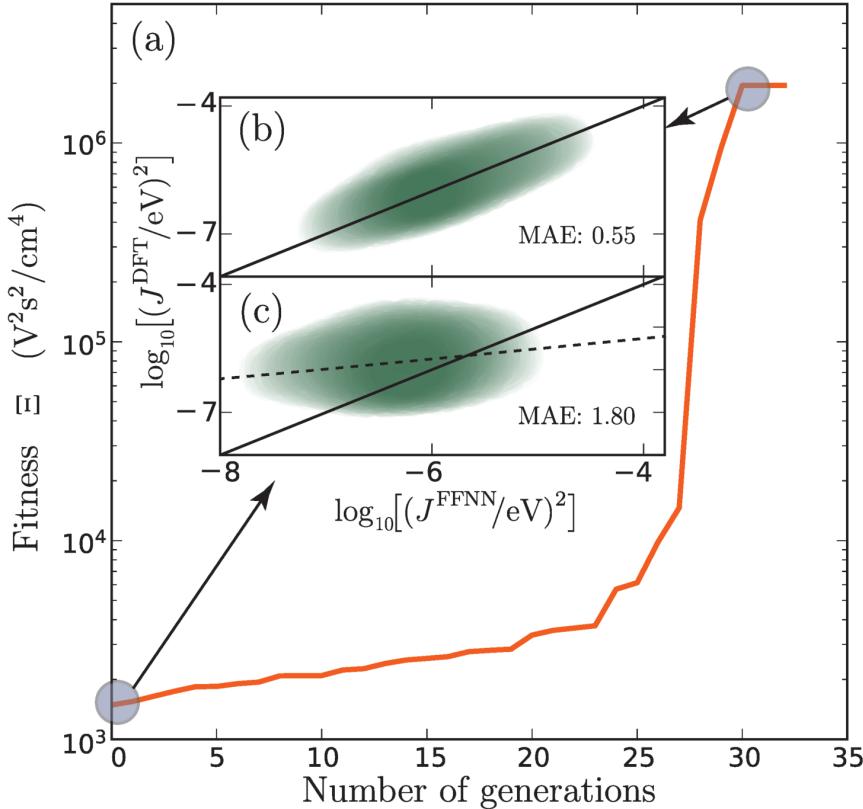


Figure 6.5: (a) Fitness evolution of the best performing feedforward neural network in each generation, showing a slow growth followed by rapid improvement reminiscent of punctuated equilibrium. (b) Correlation of predicted and reference data for the coupling elements of the final optimal FFNN model, compared to (c) the one in the first generation.

initio model as described in Chapter 6.2 and serves as the reference for the FFNN model.

In particular, we chose the disorder-free case as in Figure 6.4(a) in the evaluation of the fitness (Eq. (6.9)) during the evolutionary FFNN optimization. Here, the rates and concomitantly the mobility are solely

determined by the topological connectivity of the charge transporting network given by the electronic coupling elements [238]. The reference mobility has a minimally negative slope with increasing field strength, which is attributed to the system being in the inverted Marcus regime for $\Delta E_{ij} = 0$. Light green triangles show $\mu^{\text{FFNN}}(F)$ as it results from two individuals in the first generation FFNN, with vastly different performances. The first one yields completely unphysical behavior with mobilities on the order of $10^4 \text{ cm}^2/\text{Vs}$, about five orders of magnitude larger than the ab initio reference. In comparison, the second model is much closer but underestimates $\mu^{\text{DFT}}(F)$ consistently by about a factor of 10. While this agreement appears acceptable, a closer inspection of Figure 6.5(a) reveals a low fitness value ($\Xi = 1.5 \cdot 10^3 \text{ V}^2\text{s}^2/\text{cm}^4$). The predicted $\log_{10}[(J^{\text{FFNN}}/\text{eV})^2]$ show a MAE of 1.80 and are only weakly correlated to the DFT reference, as can be seen in Figure 6.5(c). From this starting point, the evolution of the FFNN results in an initially slowly increasing fitness. Going through 25 generations the fitness only improves by a factor of two. This slow growth is followed by a rapid fitness evolution that takes place within only six generations, after which it stops quickly and the process ends up in an equilibrium situation. Such a phenomenon of instantaneous change is not unique to our evolutionary FFNN and it has also been observed in evolutionary biology with similar patterns in the fossil records known as punctuated equilibrium [239]. The last generation FFNN consists of two layers with 800 and 550 neurons, respectively, and is characterized by a fitness value of $\Xi = 2 \cdot 10^6 \text{ V}^2\text{s}^2/\text{cm}^4$, an improvement of three orders of magnitude over the first model. This is also reflected by the obtained hole mobility (circles) in Figure 6.4(a), which is practically indistinguishable from the ab initio reference, see also Table 6.1. The MAE is reduced to 0.55 and the correlation in Figure 6.5(b) is clearly improved.

With a FFNN model at hand that shows good characteristics and performs well for the disorder-free case, we now evaluate its applicability in the realistic scenario with energetic disorder taken into account. This constitutes an important test for the FFNN model of the electronic coupling elements. While optimization of the model based on the disorder-free case should, ideally, predict the connectivity of the charge transporting network accurately, it does so for a completely flat energy landscape. It cannot be excluded a priori that coupling elements that are below the percolation threshold ($J^2 < 4 \cdot 10^{-7} \text{ eV}^2$), and hence insignificant for $\mu(\Delta E_{ij} = 0)$, provide low-probability, but crucial, additional pathways to avoid or escape low-energy regions (traps) in the $\Delta E_{ij} \neq 0$ case. With the large energetic disorder of $\sigma = 0.20 \text{ eV}$ the ab initio reference model exhibits a mobility reduction of about six orders of magnitude, see orange squares in Figure 6.4(b), and a noticeable positive field-dependence known as Poole-Frenkel behavior with $\mu(F) = \mu_0 \exp(\alpha \sqrt{F})$. The FFNN model reproduces this behavior extremely well, with an observed maximum error of $5.0 \cdot 10^{-9} \text{ cm}^2/\text{Vs}$ and a MAE of $2.7 \cdot 10^{-9} \text{ cm}^2/\text{Vs}$, which are both smaller than the average standard error of the mean of μ^{DFT} of $5.4 \cdot 10^{-9} \text{ cm}^2/\text{Vs}$, see Table 6.1. Poole-Frenkel slopes of $\alpha^{\text{DFT}} = 4.2 \cdot 10^{-3} (\text{cm}/\text{V})^{1/2}$ and $\alpha^{\text{FFNN}} = 3.2 \cdot 10^{-3} (\text{cm}/\text{V})^{1/2}$ are in reasonable agreement with each other, taking the bounds of the respective errors of the individual mobility values as given in Table 6.1 into account.

Based on this comparison, we conclude that the FFNN provides very reliable predictions of electronic coupling elements in Alq₃ over a wide range of magnitudes taking into account explicit details of molecular orientations in large-scale morphologies. It also comes at a significantly reduced computational cost compared to the ab initio model. For a single frame containing N molecules and (on average) κN hopping pairs,

Table 6.1: Hole mobilities (in cm^2/Vs) of Alq₃ for different values of externally applied electric-field. Results for cases both without ($\Delta E_{ij} = 0$) and with energetic disorder ($\Delta E_{ij} \neq 0$) are given for the three different system sizes considered, as obtained from DFT and FFNN based coupling elements, respectively.

		$3 \cdot 10^7 \text{ V/m}$	$4 \cdot 10^7 \text{ V/m}$	$5 \cdot 10^7 \text{ V/m}$	$7 \cdot 10^7 \text{ V/m}$	$9 \cdot 10^7 \text{ V/m}$
without disorder $\Delta E_{ij} = 0$						
4096	DFT	$4.7 \cdot 10^{-2} \pm 2.1 \cdot 10^{-4}$	$4.6 \cdot 10^{-2} \pm 1.7 \cdot 10^{-4}$	$4.4 \cdot 10^{-2} \pm 2.0 \cdot 10^{-4}$	$4.2 \cdot 10^{-2} \pm 2.1 \cdot 10^{-4}$	$4.1 \cdot 10^{-2} \pm 4.3 \cdot 10^{-4}$
4096	FFNN	$4.6 \cdot 10^{-2} \pm 1.8 \cdot 10^{-4}$	$4.5 \cdot 10^{-2} \pm 2.1 \cdot 10^{-4}$	$4.3 \cdot 10^{-2} \pm 2.5 \cdot 10^{-4}$	$4.0 \cdot 10^{-2} \pm 3.0 \cdot 10^{-4}$	$3.7 \cdot 10^{-2} \pm 4.3 \cdot 10^{-4}$
8192	FFNN	$3.9 \cdot 10^{-2} \pm 3.6 \cdot 10^{-4}$	$3.8 \cdot 10^{-2} \pm 3.8 \cdot 10^{-4}$	$3.6 \cdot 10^{-2} \pm 4.0 \cdot 10^{-4}$	$3.3 \cdot 10^{-2} \pm 4.3 \cdot 10^{-4}$	$3.0 \cdot 10^{-2} \pm 4.4 \cdot 10^{-4}$
32769	FFNN	$3.0 \cdot 10^{-2} \pm 5.1 \cdot 10^{-5}$	$2.9 \cdot 10^{-2} \pm 6.0 \cdot 10^{-5}$	$2.8 \cdot 10^{-2} \pm 6.7 \cdot 10^{-5}$	$2.6 \cdot 10^{-2} \pm 7.6 \cdot 10^{-5}$	$2.4 \cdot 10^{-2} \pm 7.8 \cdot 10^{-5}$
with disorder $\Delta E_{ij} \neq 0$						
4096	DFT	$8.2 \cdot 10^{-9} \pm 1.5 \cdot 10^{-9}$	$1.1 \cdot 10^{-8} \pm 2.3 \cdot 10^{-9}$	$1.5 \cdot 10^{-8} \pm 4.4 \cdot 10^{-9}$	$3.0 \cdot 10^{-8} \pm 8.8 \cdot 10^{-9}$	$4.1 \cdot 10^{-8} \pm 9.8 \cdot 10^{-9}$
4096	FFNN	$1.0 \cdot 10^{-8} \pm 4.6 \cdot 10^{-10}$	$1.3 \cdot 10^{-8} \pm 1.2 \cdot 10^{-9}$	$1.7 \cdot 10^{-8} \pm 1.9 \cdot 10^{-9}$	$2.5 \cdot 10^{-8} \pm 2.6 \cdot 10^{-9}$	$3.6 \cdot 10^{-8} \pm 4.2 \cdot 10^{-9}$
8192	FFNN	$9.8 \cdot 10^{-10} \pm 4.6 \cdot 10^{-10}$	$2.1 \cdot 10^{-9} \pm 1.2 \cdot 10^{-9}$	$3.6 \cdot 10^{-9} \pm 1.9 \cdot 10^{-9}$	$7.8 \cdot 10^{-9} \pm 2.6 \cdot 10^{-9}$	$1.5 \cdot 10^{-8} \pm 4.2 \cdot 10^{-9}$
32769	FFNN	$2.1 \cdot 10^{-10} \pm 1.2 \cdot 10^{-10}$	$3.2 \cdot 10^{-10} \pm 1.7 \cdot 10^{-10}$	$4.7 \cdot 10^{-10} \pm 2.8 \cdot 10^{-10}$	$1.5 \cdot 10^{-9} \pm 1.1 \cdot 10^{-9}$	$4.3 \cdot 10^{-9} \pm 3.0 \cdot 10^{-9}$

the total CPU time for the coupling elements is $T_N = \kappa NT_{\text{coupl}}$, where T_{coupl} is a typical CPU time per coupling element. Using the DFT based method as described in Chapter 6.2 consumes about $T_{\text{coupl}} = 20$ min on one thread of an Intel(R) Xeon(R) CPU E7-4830 v4 @ 2.00GHz for Alq₃. With $\kappa \approx 5.5$, one obtains $T_{4096} = 312$ d. These calculations are however easily parallelizable in high-throughput settings, so that transport simulations can be performed within an acceptable total simulation time of, e.g., one week. For the 4096 molecule system of Alq₃, this can be achieved by using about 45 threads simultaneously. It is apparent that due to the linear scaling of T_N with the number of molecules, studying even slightly larger systems (which might be necessary when transport properties are system-size dependent or to investigate realistic carrier concentrations) implies a linear increase in either total simulation time or number of simultaneously used threads. Note that we are not addressing issues related to memory or storage. Compared to the cost of the DFT calculations of the reference data, the cost for training the FFNN is small. Within the search space, training of a single neural network, making use of a Nvidia Titan Xp GPU (3840 CUDA cores running at 1.6 GHz), takes about 5 minutes. Solving the Master equation (Eq. (6.4)) for fitness evaluation as part of the EA approach was performed on a single CPU thread in a few seconds. The complete training procedure as in Figure 6.5 takes in total about three days. With the trained FFNN at hand, the evaluation of a coupling element is practically instantaneous, which removes the above costs and restrictions of the ab initio model.

To demonstrate the applicability of the FFNN in this context, we have also simulated Alq₃ morphologies containing 8192 and 32769 molecules, respectively, following the same procedure as before, except for the calculation of the J_{ij}^{DFT} , which would have taken $T_{8192} = 624$ d and $T_{32769} = 2496$ d. Applying the FFNN model, the hole mobilities are cal-

culated and the results are shown in Figure 6.4 and in Table 6.1. In the disorder-free case (Figure 6.4(a)), the mobility is as expected practically independent of system size. With energetic disorder taken into account, the situation in Figure 6.4(b) is markedly different. Doubling the system size from 4096 to 8192 molecules lowers the mobility by about one order of magnitude, while another quadrupling further reduces the mobility by the same amount. Such a behavior is indicative of dispersive transport [240] and is related to the fact that the mean transport energy of the charge carrier depends on the system size.

All in all, the FFNN constructed with the evolutionary approach described in this work based on fitness evaluation in the $\Delta E_{ij} = 0$ case has not only proven to work well for the more realistic, unseen $\Delta E_{ij} \neq 0$ simulation but also in application to larger systems that are inaccessible to the standard ab initio model.

6.5 Summary

To summarize, we have presented a general framework for the construction of a deep feedforward neural network to predict electronic coupling elements. The final FFNN model constructed for an amorphous organic semiconductor, tris-(8-hydroxyquinoline)aluminum, showed good agreement with ab initio reference data with and without energetic disorder. Additionally, we have shown that the final model is applicable to larger systems, which makes the presented approach a promising candidate to overcome system size limitations inherent to computationally expensive multiscale approaches.

Looking ahead, a logical continuation would be to tailor this approach to diverse material systems, extending its applicability and re-

fining its precision. Furthermore, as we go into this field of research, it becomes important to investigate the integration of quantum mechanical principles into our models. This could potentially yield more robust, accurate, and physically-grounded predictive tools, capable of addressing even more intricate challenges in the field.

Chapter 7

Discussion & Outlook

Throughout this thesis, several open problems in electronic structure calculations and machine learning approaches have been explored by developing and scrutinizing improved models in both areas individually (Chapters 3 and 4, respectively) and in hybrid first-principles and ML models (Chapters 5 and 6).

As outlined in the Introduction (Chapter 1), problems of electronic structure calculations and machine learning have been tackled from four distinct angles:

1. From the **angle of electronic structure theory**:

In Chapter 3, we have presented the results of gas phase geometry optimizations in the excited states of a set of small molecules (carbon monoxide, acetone, acrolein, and methylenecyclopropene) using *GW*-BSE employing numerical gradients. Our results show that, overall, one-shot G_0W_0 yields geometries in slightly better agreement with CASPT2 reference (overall error around 1 %) than ev*GW*-based variants (errors of about 1.4-1.5 %) when used with a PBE0 ground state. The use of a plasmon-pole model or the Tamm-Dancoff approximation only have a negligible influence on

the quality of the obtained structural parameters. We find relative errors smaller than for CC2 and TDDFT with different functionals and only larger than VMC. Our results indicate that *GW*-BSE can yield geometries in good agreement with established higher-order wave-function methods for small molecules.

2. From the **angle of machine learning theory**:

In Chapter 4, we have presented the Wasserstein metric as an index-invariant way to measure distances between molecular graph-based representations. The objective here was to resolve the problems arising from lack of atom-index invariance with adjacency matrix based representations in kernel methods. The use of the Wasserstein metric has been scrutinized using a model of the continuous atomic displacement in a test molecule as well as the prediction of atomization energies in the QM9 molecule dataset. We find that the Wasserstein metric enables the generation of QML models which achieve (a) data-efficient learning and (b) smooth target function estimates. We have contributed this new method to the QMLCODE package. [152]

3. From the **angle of hybrid *GW*-BSE/ML models**:

The focus of Chapter 5 was on the development of a surrogate ML model for the most time-consuming step in *GW* calculations, namely the calculation of the quasiparticle energy corrections, and its integration into full workflows, including quantum-classical embedding. To this end, we have introduced orbital-sensitive augmentations of molecular cartesian coordinates-based representations in Δ -machine learning of full quasiparticle excitation energies in molecules and clusters. The proposed Δ MLQP approach is capable of predicting the *GW* energies of multiple orbitals across multiple molecules and/or intra- and inter molecular conforma-

tions in a single kernel ridge regression based supervised learning model. We have demonstrated this in application to the QM8 molecular dataset and to water monomers and dimers. Furthermore, it has been shown that a single orbital-sensitive Δ -ML model for quasiparticle energies can be embedded in multiscale simulation workflows, showcased in the prediction of solvatochromic shifts of excitation energies in aqueous acetone.

4. From the **angle of hybrid multiscale models:**

In the final results Chapter 6, we developed an optimization strategy for a surrogate neural network model integrated in a first-principles workflow for the simulation of charge-carrier dynamics in disordered molecular semiconductors. Herein, the explicit determination of inter-molecular electronic coupling elements is a computational bottleneck that severely restricts the system sizes accessible to the first-principles model. Specifically, we focused on the development of a general framework for the construction of a deep feedforward neural network to predict distance and orientation dependent electronic coupling elements in disordered molecular materials. Using hole transport properties in amorphous tris(8-hydroxyquinolinato)aluminum as a prototypical system, we showed that because of the sensitivity of electronic coupling elements in disordered material morphologies to molecular orientations to the extent that target property values are distributed over several orders of magnitude even for to the eye rather similar molecular conformations, makes simple minimizing the model error with stochastic gradient descent based backpropagation insufficient. Our final model optimization strategy instead incorporates additional guidance to the evolutionary algorithm by simultaneous maximization of a model fitness that takes into account additional physical

properties, such as the field-dependent carrier mobility. The hybrid first-principles/ML multiscale model predicts such field-dependence mobilities for several physical setting in good agreement with the plain first-principles reference data. Additionally, we have demonstrated that the final model is applicable to larger systems, which makes the presented approach a promising candidate to overcome system size limitations inherent to computationally expensive multiscale approaches.

To summarize, throughout this thesis, we explored from several angles theoretical constructs and practical applications linking electronic structure calculations and machine learning models. A common aspect becoming evident from these angles is that to harmoniously blend the algorithms of machine learning with the precision of physical models requires the incorporation of physical principles into, e.g., data representation and optimization. While modifying data representation to embed, at least in part, physical concepts, kernel methods and neural networks as discussed in this thesis are general tools, and at no stage of the process, explicit use is made of the fact that the to-be-predicted properties are quantum properties.

There has been a growing interest towards bridging this divide by integrating physics-based nuances into statistical models[241–245]. Considerable attention is directed towards approaches which the emphasis is on the Schrödinger equation directly, with emphasis on the actual (partial) differential equation instead the observable, e.g., an energy. Typically, neural networks of varying sophistication are used as solvers for the Schrödinger equation, having to account, i.a., for the fermionic nature of the electrons or the square-integrability of the wave function(s).

From the work in this thesis, an interesting observation that transpires may offer another, maybe complementary, perspective on the issue of predicting quantum properties. We realize in various instances that there is a clear relationship between the to-be-predicted observables as expectation values of quantum-mechanical operators and non-linear regression in multidimensional dimensional spaces. Specifically, quantum mechanics and (kernel-based) machine learning both rely on concepts from linear algebra such as Hilbert spaces, inner products and linear operators.

To illustrate this notion, let $|\psi\rangle$ be a wavefunction that describes a quantum state of an atomistic system with associated property p , $\langle\psi|\psi\rangle = 1$, and let \mathbf{A} be the corresponding operator with $\mathbf{A}|\psi_i\rangle = \lambda_i|\psi_i\rangle$ for any of its pure eigenstates ψ_i , and hence $\langle\psi_i|\mathbf{A}|\psi_i\rangle = \lambda_i$. Since \mathbf{A} corresponds to an observable, it is linear and Hermitian; the eigenstates of an Hermitian operator form an orthonormal basis, $|\psi\rangle = \sum_i \beta_i |\psi_i\rangle$, with $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. From the superposition principle, for any state $|\psi\rangle$

$$\langle\psi|\mathbf{A}|\psi\rangle = \sum_i \beta_i \lambda_i \langle\psi|\psi_i\rangle = \sum_{i,j} \beta_i \beta_j \lambda_i \langle\psi_i|\psi_j\rangle = \sum_i \beta_i^2 \lambda_i = p_\psi. \quad (7.1)$$

Eq. 7.1 describes the computations of a property y of wave functions $|\psi\rangle$ as the expectation $\langle\psi|\mathbf{A}|\psi\rangle$ of a corresponding operator \mathbf{A} . This operator is a quadratic form acting on a very high-dimensional object $|\psi\rangle$. To compute p_ψ , not all degrees of freedom of $|\psi\rangle$ might be necessary. For example, the Hohenberg–Kohn theorems show that for electronic ground state properties, the electron density, an object of much lower dimensionality, is sufficient.

Motivated by these observations, we explicitly transfer the wave functions $|\psi\rangle$ to a lower-dimensional space $\mathcal{X} = \mathbb{R}^d$, $|\psi\rangle \mapsto \vec{x}$. The

dimensions of \mathcal{X} can be seen as basis functions (“features”) describing wave functions. Let $\vec{b}_1, \dots, \vec{b}_d \in \mathcal{X}$ be the eigenvectors of \mathbf{A} with corresponding eigenvalues λ_i . Any $\vec{x} \in \mathcal{X}$ can then be decomposed as $\vec{x} = \sum_{i=1}^d \gamma_i \vec{b}_i$, and $\vec{b}_i^T \vec{b}_j = \delta_{i,j}$. In \mathcal{X} , 7.1 becomes the predicting equation

$$\vec{x}^T \mathbf{A} \vec{x} = \sum_{i=1}^d \gamma_i \vec{x}^T \mathbf{A} \vec{b}_i = \sum_{i=1}^d \gamma_i \lambda_i \vec{x}^T \vec{b}_i = \sum_{i,j=1}^d \gamma_i \gamma_j \lambda_i \vec{b}_i^T \vec{b}_j = \sum_{i=1}^d \gamma_i^2 \lambda_i = p(\vec{x}). \quad (7.2)$$

We can further set $|\psi\rangle \mapsto \vec{x} \mapsto \phi(\vec{x})$, where $\phi(\vec{x})$ is a feature function in feature space \mathcal{F} , with \vec{x} as input space elements, and estimate operator \mathbf{A} , such that property p_ϕ is given by $p_\phi = \langle \phi(\vec{x}) | \mathbf{A} | \phi(\vec{x}) \rangle$. By expanding feature functions using orthonormal basis vectors in feature space $\tilde{\phi}_1, \dots, \tilde{\phi}_d$ with $\text{Span}\{\tilde{\phi}_1, \dots, \tilde{\phi}_d\} = \mathcal{F}$, one finds that:

$$\tilde{p}(\vec{x}) = \langle \phi(\vec{x}) | \mathbf{A} | \phi(\vec{x}) \rangle_F = \phi(\vec{x})^\top \mathbf{A} \phi(\vec{x}) \quad (7.3)$$

$$= \sum_{k=1}^d \lambda_k \phi(\vec{x})^\top \tilde{\phi}_k \tilde{\phi}_k^\top \phi(\vec{x}) \quad (7.4)$$

$$= \sum_{k=1}^d \lambda_k \phi(\vec{x})^\top \left(\sum_{m=1}^l \beta_{km} \phi(\vec{x}_m) \right) \left(\sum_{p=1}^l \beta_{kp} \phi(\vec{x}_p)^\top \right) \phi(\vec{x}) \quad (7.5)$$

$$= \sum_{k=1}^d \sum_{m,p=1}^l \lambda_k \beta_{km} \beta_{kp} \phi(\vec{x})^\top \phi(\vec{x}_m) \phi(\vec{x}_p)^\top \phi(\vec{x}) \quad (7.6)$$

$$= \sum_{k=1}^d \sum_{m,p=1}^l \lambda_k \beta_{km} \beta_{kp} \langle \phi(\vec{x}) | \phi(\vec{x}_m) \rangle_F \langle \phi(\vec{x}_p) | \phi(\vec{x}) \rangle_F \quad (7.7)$$

$$= \sum_{k=1}^d \sum_{m,p=1}^l \lambda_k \beta_{km} \beta_{kp} k(\vec{x}, \vec{x}_m) k(\vec{x}_p, \vec{x}) \quad (7.8)$$

$$= \sum_{k=1}^d \lambda_k \left(\sum_{m=1}^l \beta_{km} k(\vec{x}, \vec{x}_m) \right) \left(\sum_{m=1}^l \beta_{km} k(\vec{x}, \vec{x}_m) \right) \quad (7.9)$$

$$= \sum_{k=1}^d \lambda_k \left(\sum_{m=1}^l \beta_{km} k(\vec{x}, \vec{x}_m) \right)^2 \quad (7.10)$$

At this moment, we speculate that such a *quadratic* kernel form for the estimation of the property in Eq. 7.10 may have an advantage compared to standard KRR applied to quantum mechanics as it is more closely related via the respective Hilbert space arguments. It is conceivable that such an approach may be more data-efficient. At the same time, the formulation has disadvantage that the number of hyperparameters that need to be optimized is significantly larger than for standard KRR, similar to what is usually encountered for neural networks. Nevertheless, the formulation is intriguing and might inspire further research into model training, training efficiency, and to explore intricacies of the

input/feature spaces. Such studies could not only further validate the robustness and applicability of the models based on physical concepts, but also open up ways for more research and applications in the field of quantum machine learning.

Summary

Electronic Structure Calculations and Machine Learning: from Theory to Multiscale Applications

In this dissertation, we start by delving into the theory of many-electron quantum mechanics for molecular structures. A concise yet comprehensive overview is given, highlighting the most prevalent methodologies that have gained significant traction within the computational quantum chemistry community.

Following this foundational exploration, the theoretical foundations and practical applications of quantum machine learning (QML) models to both isolated molecules and clusters are systematically addressed.

Initially, we probe into the *GW*-BSE theory, presenting a comprehensive benchmark study on gas phase geometry optimizations in the excited states of small molecules by employing numerical gradients.

Subsequently, our focus shifts to the novel utilization of the Wasserstein metric for measuring distances between molecules, represented by the atom index dependent adjacency Coulomb matrix within the scope of kernel ridge regression-based supervised learning. The QML models developed exhibit improved training efficiency and yield enhanced continuity in energy predictions related to molecular deformations. An

illustration of this continuity is showcased in scenarios involving the extraction of an atom from an organic molecule. To provide a quantitative dimension, learning curves, which articulate the reduction of the atomization energy's prediction error relative to training set dimensions, have been obtained for tens of thousands of organic molecules. In a comparative study, Wasserstein based kernels are set side by side against conventional kernels in quantum machine learning, underscoring the former's data efficiency through empirical evidence.

In a subsequent exploration, we introduce the Δ -Machine Learning paradigm, tailored for the prediction of GW quasiparticle energies and photoelectron spectra of molecules and clusters. By leveraging orbital-sensitive graph-based representations, this technique underscores the feasibility of predicting the full photoelectron spectrum of molecules within the dataset through a singular model. Additionally, this methodology's adaptability is accentuated when integrated into multiscale simulation frameworks. This adaptability is demonstrably evident in our study on solvatochromic shifts of quasiparticle and electron-hole excitation energies of solvated acetone in a setup combining Molecular Dynamics, DFT, the GW approximation and the Bethe–Salpeter Equation (GW -BSE).

Moreover, we introduce a general framework for the construction of a deep feedforward neural network to predict distance and orientation dependent electronic coupling elements in disordered molecular materials. An evolutionary algorithm automatizes the selection of an optimal architecture of the artificial neural network within a predefined search space. As a prototypical system, we consider hole transport in amorphous tris(8-hydroxyquinolinato)aluminum. The final optimized deep feedforward neural network is tested for transport models without and

with energetic disorder. It predicts electronic coupling elements and mobilities in excellent agreement with the reference data.

In our concluding section, we draw attention to the profound interplay between the foundations of quantum mechanics and kernel-based machine learning. Given the collection of foundational principles between quantum mechanics and kernel-based machine learning—principally rooted in linear algebra constructs like Hilbert spaces and high-dimensional linear operators—we discuss the idea of interrelation between quantum-mechanical operators and high-dimensional nonlinear regressions, including instructions for systematic derivation of machine-learning equivalents of quantum-mechanical calculations. We illustrate this relationship with analytic and numerical examples.

Acknowledgement

I am immensely grateful for the support and guidance I've received throughout my PhD journey. A special thanks goes to my supervisor, Dr. B. Baumeier, whose expertise and patience have been fundamental to my research. I also deeply appreciate the support from my co-promotor, Prof. Dr. O.A. von Lilienfeld, for all invaluable insights and encouragements. My sincere appreciation extends to the committee members: Prof. Dr. K. P. L. Veroy-Grep, Prof. Dr. S. Calero, Dr. M. Rupp, and Prof. Dr. T. Bereau. Your constructive feedback have greatly enriched my academic experience.

To my parents, Hulya and Ali Çaylak, and my sister, Tugce Çaylak – your constant support and belief in me have been a source of strength and motivation throughout this journey, and for that, I am forever thankful.

I would also like to thank my colleagues and friends in our group and department. Gianluca, Vivek, Jens, Pranav, Zhong, Wouter, Anil, and Yuriy – our discussions and the feedback you've provided have been crucial in shaping my research. Your friendship and shared passion for our field have made this journey enjoyable and rewarding.

Additionally, I acknowledge the broader academic community and everyone who has directly or indirectly contributed to my research. Your collective wisdom and knowledge have been immensely inspiring.

About the author

Born on October 23, 1989, in Venray, the Netherlands, Mr. Onur Caylak pursued his initial academic interests at VU University Amsterdam, where he completed a Bachelor of Science in Mathematics in 2015. Under the guidance of Prof. Dr. Andre Ran.

Moving to Eindhoven for further studies, he enrolled in a Master's program in Industrial and Applied Mathematics. In 2017, as a part of his master's programm, he wrote a thesis supervised by Dr. Georg Prokert.

Later in 2017, he initiated his doctoral research, sponsored by the NWO, at the Centre for Analysis, Scientific Computing, and Applications in Eindhoven. Under the guidance and support of Dr. Bjoern Baumeier, his research primarily revolved around integrating machine learning techniques into electronic structure calculations, a core theme of this dissertation.

Upon completing the research component of his PhD, he took on a new challenge in the beginning of 2021. He joined IMC Trading, a high-frequency trading company based in Amsterdam.

Bibliography

- [1] V. Helms, *Principles of Computational Cell Biology: From Protein Complexes to Cellular Networks* (Wiley, 2018).
- [2] T. Förster, “Zwischenmolekulare Energiewanderung und Fluoreszenz”, Annalen der Physik **437**, 55 (1948).
- [3] D. L. Dexter, “A theory of sensitized luminescence in solids”, The Journal of Chemical Physics **21**, 836 (1953).
- [4] B. Valeur and M. N. Berberan-Santos, *Molecular Fluorescence* (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, Apr. 2012).
- [5] R. Kancherla, K. Muralirajan, A. Sagadevan, and M. Rueping, “Visible Light-Induced Excited-State Transition-Metal Catalysis”, Trends in Chemistry **1**, 510 (2019).
- [6] P. Kirkpatrick and C. Ellis, “Chemical space”, Nature **432**, 823 (2004).
- [7] E. Schrödinger, “An Undulatory Theory of the Mechanics of Atoms and Molecules”, en, Physical Review **28**, 1049 (1926).
- [8] D. R. Hartree, “The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion”, Mathematical Proceedings of the Cambridge Philosophical Society **24**, 111 (1928).
- [9] P. Hohenberg and W. Kohn, “Inhomogeneous electron gas”, Physical Review **136**, B864 (1964).

- [10] W. Kohn and L. J. Sham, “Self-Consistent Equations Including Exchange and Correlation Effects”, *Phys. Rev.* **140**, A1133 (1965).
- [11] A. D. Becke, “Density functionals for static, dynamical, and strong correlation”, *The Journal of Chemical Physics* **138**, 074109 (2013).
- [12] A. Dreuw and M. Head-Gordon, “Failure of time-dependent density functional theory for long-range charge-transfer excited states: The zincbacteriochlorin-bacteriochlorin and bacteriochlorophyll-spheroidene complexes”, *J. Am. Chem. Soc.* **126**, 4007 (2004).
- [13] S. Kristyán and P. Pulay, “Can (semi)local density functional theory account for the london dispersion forces?”, *Chemical Physics Letters* **229**, 175 (1994).
- [14] A. J. Cohen, P. Mori-Sánchez, and W. Yang, “Insights into current limitations of density functional theory”, *Science* **321**, 792 (2008).
- [15] J. Behler, “Constructing high-dimensional neural network potentials: a tutorial review”, *Int. J. Quantum Chem.* **115**, 1032 (2015).
- [16] M. Rupp, “Special issue on machine learning and quantum mechanics”, *International Journal of Quantum Chemistry* **115**, 1003 (2015).
- [17] J. Behler, “Perspective: machine learning potentials for atomistic simulations”, *J. Chem. Phys.* **145**, 170901 (2016).
- [18] M. Rupp, O. A. von Lilienfeld, and K. Burke, “Guest editorial: special topic on data-enabled theoretical chemistry”, *J. Chem. Phys.* **148**, 241401 (2018).
- [19] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, “Machine learning for molecular simulation”, arXiv preprint arXiv:1911.02792 (2019).
- [20] F. Musil and M. Ceriotti, “Machine learning at the atomic scale”, *CHIMIA International Journal for Chemistry* **73**, 972 (2019).

- [21] M. Ceriotti, “Unsupervised machine learning in atomistic simulations, between predictions and understanding”, *The Journal of chemical physics* **150**, 150901 (2019).
- [22] M. Ceriotti, M. J. Willatt, and G. Csányi, “Machine-learning of atomic-scale properties based on physical principles”, arXiv preprint arXiv:1901.10971 (2019).
- [23] O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller, “Exploring chemical compound space with quantum-based machine learning”, arXiv preprint arXiv:1911.10084 (2019).
- [24] B. Huang and O. A. von Lilienfeld, “Communication: understanding molecular representations in machine learning: the role of uniqueness and target similarity”, *J. Chem. Phys.* **145**, 161102, <http://dx.doi.org/10.1063/1.4964627> (2016).
- [25] R. Ramakrishnan, P. Dral, M. Rupp, and O. A. von Lilienfeld, “Quantum chemistry structures and properties of 134 kilo molecules”, *Scientific Data* **1**, 140022 (2014).
- [26] M. Born and R. Oppenheimer, “Zur Quantentheorie der Moleküle”, *Annalen der Physik* **389**, 457 (1927).
- [27] R. Martin, R. Martin, and C. U. Press, *Electronic structure: basic theory and practical methods* (Cambridge University Press, 2004).
- [28] C. Cramer, *Essentials of computational chemistry: theories and models* (Wiley, 2013).
- [29] S. Palm, P. D., and M. D., “On certain correspondences among various coupled-cluster theories for closed-shell systems”, *Pramana – Journal of Physics | Indian Academy of Sciences*, 261 (1982).
- [30] T. Van Voorhis and M. Head-Gordon, “Benchmark variational coupled cluster doubles results”, *Journal of Chemical Physics* **113**, 8873 (2000).

-
- [31] B. Cooper and P. J. Knowles, “Benchmark studies of variational, unitary and extended coupled cluster methods”, *Journal of Chemical Physics* **133**, 234102 (2010).
 - [32] F. A. Evangelista, “Alternative single-reference coupled cluster approaches for multireference problems: The simpler, the better”, *Journal of Chemical Physics* **134**, 224102 (2011).
 - [33] B. Jeziorski and H. J. Monkhorst, “Coupled-cluster method for multiterminantal reference states”, *Physical Review A* **24**, 1668 (1981).
 - [34] I. Lindgren and D. Mukherjee, “On the connectivity criteria in the open-shell coupled-cluster theory for general model spaces”, *Physics Reports* **151**, 93 (1987).
 - [35] K. Sneskov and O. Christiansen, “Excited state coupled cluster methods”, *WIREs Computational Molecular Science* **2**, 566 (2012).
 - [36] L. Zhao and E. Neuscamman, “Variational Excitations in Real Solids: Optical Gaps and Insights into Many-Body Perturbation Theory”, *Physical Review Letters* **123**, 036402 (2019).
 - [37] W. Kohn and L. J. Sham, “Self-consistent equations including exchange and correlation effects”, *Physical Review* **140**, A1133 (1965).
 - [38] K. Burke, “Perspective on density functional theory”, *Journal of Chemical Physics* **136**, 150901 (2012).
 - [39] T. Tsuneda, in *Density functional theory in quantum chemistry* (Springer Japan, 2014), pp. 101–124.
 - [40] E. Runge and E. K. Gross, “Density-functional theory for time-dependent systems”, *Physical Review Letters* **52**, 997 (1984).
 - [41] E. K. Gross and W. Kohn, “Local density-functional theory of frequency-dependent linear response”, *Physical Review Letters* **55**, 2850 (1985).

- [42] M. Casida, A. Ipatov, and F. Cordova, in *Time-Dependent Density Functional Theory*, Vol. 706, edited by R. Beig, W. Beiglböck, W. Domcke, B.-G. Englert, U. Frisch, P. Hänggi, G. Hasinger, K. Hepp, W. Hillebrandt, D. Imboden, R. L. Jaffe, R. Lipowsky, H. v. Löhneysen, I. Ojima, D. Sornette, S. Theisen, W. Weise, J. Wess, J. Zittartz, M. A. Marques, C. A. Ullrich, F. Nogueira, A. Rubio, K. Burke, and E. K. U. Gross, Series Title: Lecture Notes in Physics (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006), pp. 243–257.
- [43] L. Hedin, “New Method for Calculating the One-Particle Green’s Function with Application to the Electron-Gas Problem”, Phys. Rev. **139**, A796 (1965).
- [44] L. Hedin and S. Lundqvist, en, in *Solid State Physics*, Vol. 23, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic Press, Jan. 1970), pp. 1–181.
- [45] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, en, in *Solid State Physics*, Vol. 54, edited by H. Ehrenreich and F. Spaepen (Academic Press, Jan. 2000), pp. 1–218.
- [46] M. Rohlfing, “Excited states of molecules from Green’s function perturbation techniques”, en, Int. J. Quantum Chem. **80**, 807 (2000).
- [47] M. S. Hybertsen and S. G. Louie, “First-Principles Theory of Quasiparticles: Calculation of Band Gaps in Semiconductors and Insulators”, Phys. Rev. Lett. **55**, 1418 (1985).
- [48] D. Golze, J. Wilhelm, M. J. van Setten, and P. Rinke, “Core-Level Binding Energies from GW: An Efficient Full-Frequency Approach within a Localized Basis”, J. Chem. Theory Comput. **14**, 4856 (2018).
- [49] D. Golze, M. Dvorak, and P. Rinke, “The GW Compendium: A Practical Guide to Theoretical Photoemission Spectroscopy”, English, Front. Chem. **7** (2019).

-
- [50] G. Tirimbò, V. Sundaram, O. Çaylak, W. Scharpach, J. Sijen, C. Jung-hans, J. Brown, F. Z. Ruiz, N. Renaud, J. Wehner, and B. Baumeier, “Excited-state electronic structure of molecules using many-body Green’s functions: Quasiparticles and electron–hole excitations with VOTCA-XTP”, *J. Chem. Phys.* **152**, 114103 (2020).
 - [51] O. Caylak and B. Baumeier, “Excited-state geometry optimization of small molecules with many-body green’s functions theory”, *J. Chem. Theory Comput.* **17**, 879 (2021).
 - [52] M. S. Hybertsen and S. G. Louie, “Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies”, *Phys. Rev. B* **34**, 5390 (1986).
 - [53] R. W. Godby and R. J. Needs, “Metal-insulator transition in Kohn-Sham theory and quasiparticle theory”, *Phys. Rev. Lett.* **62**, 1169 (1989).
 - [54] M. Rohlfing and S. G. Louie, “Electron-hole excitations and optical spectra from first principles”, *Phys. Rev. B* **62**, 4927 (2000).
 - [55] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (Courier Corporation, June 2003).
 - [56] T. Rangel, S. M. Hamed, F. Bruneval, and J. B. Neaton, “An assessment of low-lying excitation energies and triplet instabilities of organic molecules with an ab initio Bethe-Salpeter equation approach and the Tamm-Dancoff approximation”, *Journal of Chemical Physics* **146**, 10.1063/1.4983126 (2017).
 - [57] D. Jacquemin, I. Duchemin, A. Blondel, and X. Blase, “Benchmark of Bethe-Salpeter for Triplet Excited-States”, *Journal of Chemical Theory and Computation* **13**, 767 (2017).
 - [58] Y. Ma, M. Rohlfing, and C. Molteni, “Excited states of biological chromophores studied using many-body perturbation theory: Effects

- of resonant-antiresonant coupling and dynamical screening”, Physical Review B - Condensed Matter and Materials Physics **80**, 241405 (2009).
- [59] S. Obara and A. Saika, “Efficient recursive computation of molecular integrals over Cartesian Gaussian functions”, en, The Journal of Chemical Physics **84**, 3963 (1986).
- [60] S. Reine, T. Helgaker, and R. Lindh, “Multi-electron integrals”, en, Wiley Interdisciplinary Reviews: Computational Molecular Science **2**, 290 (2012).
- [61] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, “Auxiliary basis sets to approximate Coulomb potentials (Chem. Phys. Letters 240 (1995) 283) (PII:0009-2614(95)00621-4)”, Chemical Physics Letters **242**, 652 (1995).
- [62] X. Blase, C. Attaccalite, and V. Olevano, “First-principles GW calculations for fullerenes, porphyrins, phtalocyanine, and other molecules of interest for organic photovoltaic applications”, Physical Review B - Condensed Matter and Materials Physics **83**, 115103 (2011).
- [63] D. Golze, J. Wilhelm, M. J. van Setten, and P. Rinke, “Core-Level Binding Energies from $\langle i \rangle$ GW $\langle /i \rangle$: An Efficient Full-Frequency Approach within a Localized Basis”, Journal of Chemical Theory and Computation **14**, 4856 (2018).
- [64] D. Golze, M. Dvorak, and P. Rinke, “The GW Compendium: A Practical Guide to Theoretical Photoemission Spectroscopy”, Frontiers in Chemistry **7**, 10.3389/fchem.2019.00377 (2019).
- [65] M. S. Hybertsen and S. G. Louie, “Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies”, Physical Review B, 10.1103/PhysRevB.34.5390 (1986).

- [66] R. W. Godby and R. J. Needs, “Metal-insulator transition in Kohn-Sham theory and quasiparticle theory”, Physical Review Letters, 10.1103/PhysRevLett.62.1169 (1989).
- [67] M. Rohlfing, P. Krüger, and J. Pollmann, “Efficient scheme for GW quasiparticle band-structure calculations with applications to bulk Si and to the Si(001)-(2×1) surface”, Physical Review B **52**, 1905 (1995).
- [68] C. Adachi, “Third-generation organic electroluminescence materials”, Jpn. J. Appl. Phys. **53**, 060101 (2014).
- [69] C. W. Tang and S. A. Vanslyke, “Organic electroluminescent diodes”, Appl. Phys. Lett. **51**, 913 (1987).
- [70] J. J. Halls, C. A. Walsh, N. C. Greenham, E. A. Marseglia, R. H. Friend, S. C. Moratti, and A. B. Holmes, “Efficient photodiodes from interpenetrating polymer networks”, Nature **376**, 498 (1995).
- [71] S. Baral, M. Phillips, H. Yan, J. Aveno, L. Gundlach, B. Baumeier, and E. Lyman, “Ultrafast Formation of the Charge Transfer State of Prodan Reveals Unique Aspects of the Chromophore Environment”, J. Phys. Chem. B **124**, 2643 (2020).
- [72] T. Stein, L. Kronik, and R. Baer, “Reliable prediction of charge transfer excitations in molecular complexes using time-dependent density functional theory”, J. Am. Chem. Soc. **131**, 2818 (2009).
- [73] L. Hedin and S. Lundqvist, “Effects of Electron-Electron and Electron-Phonon Interactions on the One-Electron States of Solids”, Solid State Physics - Advances in Research and Applications **23**, edited by F. Seitz, D. Turnbull, and H. Ehrenreich, 1 (1970).
- [74] M. S. Hybertsen and S. G. Louie, “First-principles theory of quasi-particles: Calculation of band gaps in semiconductors and insulators”, Physical Review Letters **55**, 1418 (1985).

- [75] G. Onida, L. Reining, and A. Rubio, “Electronic excitations: Density-functional versus many-body Green’s-function approaches”, *Reviews of Modern Physics* **74**, 601 (2002).
- [76] M. Rohlfing, “Excited states of molecules from Green’s function perturbation techniques”, en, *International Journal of Quantum Chemistry* **80**, 807 (2000).
- [77] B. Baumeier, D. Andrienko, Y. Ma, and M. Rohlfing, “Excited states of dicyanovinyl-substituted oligothiophenes from many-body Green’s functions theory”, *Journal of Chemical Theory and Computation* **8**, 10.1021/ct2008999 (2012).
- [78] B. Baumeier, D. Andrienko, and M. Rohlfing, “Frenkel and charge-transfer excitations in donor-acceptor complexes from many-body green’s functions theory”, *Journal of Chemical Theory and Computation* **8**, 2790 (2012).
- [79] B. Baumeier, M. Rohlfing, and D. Andrienko, “Electronic excitations in push-pull oligomers and their complexes with fullerene from many-body Green’s functions theory with polarizable embedding”, *Journal of Chemical Theory and Computation* **10**, 3104 (2014).
- [80] M. J. Van Setten, F. Weigend, and F. Evers, “The GW-method for quantum chemistry applications: Theory and implementation”, *Journal of Chemical Theory and Computation* **9**, 232 (2013).
- [81] M. J. Van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers, and P. Rinke, “GW100: Benchmarking G0W0 for Molecular Systems”, *Journal of Chemical Theory and Computation* **11**, 5665 (2015).
- [82] D. Varsano, E. Coccia, O. Pulci, A. M. Conte, and L. Guidoni, “Ground state structures and electronic excitations of biological chromophores

- at Quantum Monte Carlo/Many Body Green's Function Theory level”, Computational and Theoretical Chemistry, Excited states: From isolated molecules to complex environments **1040-1041**, 338 (2014).
- [83] F. Kaplan, M. E. Harding, C. Seiler, F. Weigend, F. Evers, and M. J. Van Setten, “Quasi-Particle Self-Consistent GW for Molecules”, Journal of Chemical Theory and Computation **12**, 2528 (2016).
 - [84] P. Boulanger, D. Jacquemin, I. Duchemin, and X. Blase, “Fast and accurate electronic excitations in cyanines with the many-body bethe-salpeter approach”, Journal of Chemical Theory and Computation **10**, 1212 (2014).
 - [85] D. Jacquemin, I. Duchemin, and X. Blase, “Benchmarking the Bethe-Salpeter Formalism on a Standard Organic Molecular Set”, Journal of Chemical Theory and Computation **11**, 3290 (2015).
 - [86] F. Bruneval, T. Rangel, S. M. Hamed, M. Shao, C. Yang, and J. B. Neaton, “MOLGW 1: Many-body perturbation theory software for atoms, molecules, and clusters”, Computer Physics Communications **208**, 149 (2016).
 - [87] G. Strinati, “Application of the Green's functions method to the study of the optical properties of semiconductors”, en, La Rivista Del Nuovo Cimento Series 3 **11**, 1 (1988).
 - [88] M. E. CASIDA, in (Nov. 1995), pp. 155–192.
 - [89] S. Ismail-Beigi and S. G. Louie, “Excited-State Forces within a First-Principles Green's Function Formalism”, Physical Review Letters **90**, 4 (2003).
 - [90] R. W. Godby, M. Schlüter, and L. J. Sham, “Self-energy operators and exchange-correlation potentials in semiconductors”, Physical Review B **37**, 10159 (1988).

- [91] J. Wehner, L. Brombacher, J. Brown, C. Junghans, O. Çaylak, Y. Khalak, P. Madhikar, G. Tirimbò, and B. Baumeier, “Electronic Excitations in Complex Molecular Environments: Many-Body Green’s Functions Theory in VOTCA-XTP”, *Journal of Chemical Theory and Computation* **14**, 6253 (2018).
- [92] C. Adamo and V. Barone, “Toward reliable density functional methods without adjustable parameters: The PBE0 model”, *J. Chem. Phys.* **110**, 6158 (1999).
- [93] C. A. Guido, D. Jacquemin, C. Adamo, and B. Mennucci, “On the TD-DFT accuracy in determining single and double bonds in excited-state structures of organic molecules”, *J. Phys. Chem. A* **114**, 13402 (2010).
- [94] C. A. Guido, S. Knecht, J. Kongsted, and B. Mennucci, “Benchmarking time-dependent density functional theory for excited state geometries of organic molecules in gas-phase and in solution”, *J. Chem. Theory Comput.* **9**, 2209 (2013).
- [95] C. S. Page and M. Olivucci, “Ground and excited state CASPT2 geometry optimizations of small organic molecules”, *J. Comput. Chem.* **24**, 298 (2003).
- [96] T. Yanai, D. P. Tew, and N. C. Handy, “A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP)”, *Chem. Phys. Lett.* **393**, 51 (2004).
- [97] Y. Zhao and D. G. Truhlar, “The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals”, *Theor. Chem. Acc.* **120**, 215 (2008).
- [98] T. H. Dunning, “Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen”, *J. Chem. Phys.* **90**, 1007 (1989).

-
- [99] F. Weigend, A. Köhn, and C. Hättig, “Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations”, *J. Chem. Phys.* **116**, 3175 (2002).
 - [100] J. F. Stanton, J. Gauss, N. Ishikawa, and M. Head-Gordon, “A comparison of single reference methods for characterizing stationary points of excited state potential energy surfaces”, *The Journal of Chemical Physics* **103**, 4160 (1995).
 - [101] K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure* (Springer US, 1979).
 - [102] R. Guareschi and C. Filippi, “Ground- and excited-state geometry optimization of small organic molecules with quantum Monte Carlo”, *Journal of Chemical Theory and Computation* **9**, 5513 (2013).
 - [103] Š. Budzák, G. Scalmani, and D. Jacquemin, “Accurate excited-state geometries: a caspt2 and coupled-cluster reference database for small molecules”, *J. Chem. Theory Comput.* **13**, 6237 (2017).
 - [104] H. Eshuis, J. Yarkony, and F. Furche, “Fast computation of molecular random phase approximation correlation energies using resolution of the identity and imaginary frequency integration”, *J. Chem. Phys.* **132**, 234114 (2010).
 - [105] C. Holzer, X. Gui, M. E. Harding, G. Kresse, T. Helgaker, and W. Klopper, “Bethe–salpeter correlation energies of atoms and molecules”, *J. Chem. Phys.* **149**, 144106 (2018).
 - [106] P.-F. Loos, A. Scemama, I. Duchemin, D. Jacquemin, and X. Blase, “Pros and cons of the bethe–salpeter formalism for ground-state energies”, *J. Phys. Chem. Lett.* **11**, 3536 (2020).
 - [107] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, “Fast and accurate modeling of molecular atomization energies with machine learning”, *Phys. Rev. Lett.* **108**, 058301 (2012).

- [108] G. Montavon, K. Hansen, S. Fazli, M. Rupp, F. Biegler, A. Ziehe, A. Tkatchenko, O. A. von Lilienfeld, and K.-R. Müller, “Learning invariant representations of molecules for atomization energy prediction”, NIPS proceedings **25**, 449 (2012).
- [109] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K. R. Müller, and O. Anatole Von Lilienfeld, “Machine learning of molecular electronic properties in chemical compound space”, New Journal of Physics **15**, 10 . 1088/1367 - 2630/15/9/095003 (2013).
- [110] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. Von Lilienfeld, A. Tkatchenko, and K. R. Müller, “Assessment and validation of machine learning methods for predicting molecular atomization energies”, Journal of Chemical Theory and Computation **9**, 3404 (2013).
- [111] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, and R. Ramprasad, “Accelerating materials property predictions using machine learning”, Scientific reports **3**, 2810 (2013).
- [112] S. De, A. P. Bartok, G. Csanyi, and M. Ceriotti, “Comparing molecules and solids across structural and alchemical space”, Phys. Chem. Chem. Phys. **18**, 13754 (2016).
- [113] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, “Machine learning unifies the modeling of materials and molecules”, Science Advances **3**, 10 . 1126/sciadv.1701816 (2017).
- [114] O. A. von Lilienfeld, “Quantum machine learning in chemical compound space”, Angewandte Chemie International Edition **57**, doi:201709686, 4164 (2018).
- [115] O. A. von Lilienfeld, *Towards the computational design of compounds from first principles*, edited by V. Bach and L. D. Site, Vol. IX, Mathematical Physics Studies (Springer, 2014), p. 417.

-
- [116] N. Marzari, “Materials modelling: the frontiers and the challenges”, *Nature materials* **15**, 381 (2016).
 - [117] F. A. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, “Machine learning energies of 2 million elpasolite (ABC_2D_6) crystals”, *Phys. Rev. Lett.* **117**, 135502 (2016).
 - [118] O. A. von Lilienfeld, R. Ramakrishnan, M. Rupp, and A. Knoll, “Fourier series of atomic radial distribution functions: a molecular fingerprint for machine learning models of quantum chemical properties”, *Int. J. Quantum Chem.* **115**, <http://arxiv.org/abs/1307.2918>, 1084 (2015).
 - [119] O. A. von Lilienfeld, “First principles view on chemical compound space: gaining rigorous atomistic control of molecular properties”, *International Journal of Quantum Chemistry* **113**, 1676 (2013).
 - [120] K. Hansen, F. Biegler, O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, “Interaction potentials in molecules and non-local information in chemical space”, *J. Phys. Chem. Lett.* **6**, 2326 (2015).
 - [121] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, “Prediction errors of molecular machine learning models lower than hybrid DFT error”, *J. Chem. Theory Comput.* **13**, 5255 (2017).
 - [122] C. R. Collins, G. J. Gordon, O. A. von Lilienfeld, and D. J. Yaron, “Constant size descriptors for accurate machine learning models of molecular properties”, *The Journal of Chemical Physics* **148**, 241718 (2018).
 - [123] B. Huang and O. A. von Lilienfeld, “The “DNA” of chemistry: Scalable quantum machine learning with “amons””, arXiv preprint arXiv:1707.04146, submitted (2017).

- [124] F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, “Alchemical and structural distribution based representation for universal quantum machine learning”, *The Journal of Chemical Physics* **148**, 241717 (2018).
- [125] W. Pronobis, A. Tkatchenko, and K.-R. Müller, “Many-body descriptors for predicting molecular properties with machine learning: analysis of pairwise and three-body interactions in molecules”, *Journal of chemical theory and computation* (2018).
- [126] M. Eickenberg, G. Exarchakis, M. Hirn, S. Mallat, and L. Thiry, “Solid harmonic wavelet scattering for predictions of molecule properties”, *The Journal of Chemical Physics* **148**, 241732 (2018).
- [127] T.-S. Ho and H. Rabitz, “A general method for constructing multidimensional molecular potential energy surfaces from ab initio calculations”, *J. Chem. Phys.* **104**, 2584 (1996).
- [128] M. Soloviov and M. Meuwly, “Reproducing kernel potential energy surfaces in biomolecular simulations: nitric oxide binding to myoglobin”, *The Journal of chemical physics* **143**, 09B607_1 (2015).
- [129] O. T. Unke and M. Meuwly, “Toolkit for the construction of reproducing kernel-based representations of data: application to multidimensional potential energy surfaces”, *Journal of chemical information and modeling* **57**, 1923 (2017).
- [130] L. N. Vaserstein, “Markov processes over denumerable products of spaces, describing large systems of automata”, *Problemy Peredachi Informatsii* **5**, 64 (1969).
- [131] R. Ramakrishnan, M. Hartmann, E. Tapavicza, and O. A. V. Lilienfeld, “Electronic spectra from TDDFT and machine learning in chemical space”, *J. Chem. Phys.* **143**, 84111 (2015).

-
- [132] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. V. Lilienfeld, “Big data meets quantum chemistry approximations: The Δ -machine learning approach”, *J. Chem. Theory Comput.* **11**, 2087 (2015).
 - [133] O. Çaylak, O. A. von Lilienfeld, and B. Baumeier, “Wasserstein metric for improved quantum machine learning with adjacency matrix representations”, *en, Mach. Learn.: Sci. Technol.* **1**, 03LT01 (2020).
 - [134] G. Tirimbó, O. Çaylak, and B. Baumeier, “A kernel-based machine learning approach to computing quasiparticle energies within many-body green’s functions theory”, arXiv (2020).
 - [135] D. Lemm, G. F. von Rudorff, and O. A. von Lilienfeld, “Energy-free machine learning predictions of *{ab initio} structures*”, (2021).
 - [136] M. Rupp, A. Tkatchenko, K. R. Müller, and O. A. V. Lilienfeld, “Fast and accurate modeling of molecular atomization energies with machine learning”, *Phys. Rev. Lett.* **108**, 058301 (2012).
 - [137] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. V. Lilienfeld, K. R. Müller, and A. Tkatchenko, “Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space”, *J. Phys. Chem. Lett.* **6**, 2326 (2015).
 - [138] O. Çaylak, A. Yaman, and B. Baumeier, “Evolutionary approach to constructing a deep feedforward neural network for prediction of electronic coupling elements in molecular materials”, *J. Chem. Theory Comput.* **15**, 1777 (2019).
 - [139] F. Faber, A. Lindmaa, O. A. von Lilienfeld, and R. Armiento, “Crystal structure representations for machine learning models of formation energies”, *Int. J. Quantum Chem.* **115**, <http://arxiv.org/abs/1503.07406>, 1094 (2015).

- [140] O. Çaylak, A. Yaman, and B. Baumeier, “Evolutionary Approach to Constructing a Deep Feedforward Neural Network for Prediction of Electronic Coupling Elements in Molecular Materials”, *Journal of Chemical Theory and Computation* **15**, 1777 (2019).
- [141] K. Ghosh, A. Stuke, M. Todorović, P. B. Jørgensen, M. N. Schmidt, A. Vehtari, and P. Rinke, “Deep learning spectroscopy: neural networks for molecular excitation spectra”, *Advanced science* **6**, 1801367 (2019).
- [142] J. Behler, “Atom-centered symmetry functions for constructing high-dimensional neural networks potentials”, *J. Chem. Phys.* **134**, 074106 (2011).
- [143] A. P. Bartók, R. Kondor, and G. Csányi, “On representing chemical environments”, *Phys. Rev. B* **87**, 184115 (2013).
- [144] C. Frogner, C. Zhang, H. Mobahi, M. Araya-Polo, and T. A. Poggio, “Learning with a Wasserstein loss”, *Advances in Neural Information Processing Systems (NIPS)* 28 (2015).
- [145] G. Montavon, K.-R. Müller, and M. Cuturi, “Wasserstein training of boltzmann machines”, *arXiv preprint arXiv:1507.01972* (2015).
- [146] J. Rabin, S. Ferradans, and N. Papadakis, “Adaptive color transfer with relaxed optimal transport”, in *2014 ieee international conference on image processing (icip)* (IEEE, 2014), pp. 4852–4856.
- [147] A. Ramdas, N. G. Trillos, and M. Cuturi, “On wasserstein two-sample testing and related families of nonparametric tests”, *Entropy* **19**, 10 . 3390/e19020047 (2017).
- [148] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, “Towards exact molecular dynamics simulations with machine-learned force fields”, *Nature communications* **9**, 3887 (2018).

-
- [149] S. Chmiela, H. E. Sauceda, I. Poltavsky, K.-R. Müller, and A. Tkatchenko, “Sgdml: constructing accurate and data efficient molecular force fields using machine learning”, Computer Physics Communications (2019).
 - [150] J. Westermayr, F. A. Faber, A. S. Christensen, O. A. von Lilienfeld, and P. Marquetand, “Neural networks and kernel ridge regression for excited states dynamics of ch₂nh₂⁺: from single-state to multi-state representations and multi-property machine learning models”, arXiv preprint arXiv:1912.08484 (2019).
 - [151] J. Westermayr, M. Gastegger, M. F. Menger, S. Mai, L. González, and P. Marquetand, “Machine learning enables long time scale molecular photodynamics simulations”, Chemical Science **10**, 8100 (2019).
 - [152] A. S. Christensen, F. A. Faber, B. Huang, L. A. Bratholm, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, “Qml: a python toolkit for quantum machine learning”, <http://www.qmlcode.org>, 10.5281/zenodo.817332 (2017).
 - [153] O. A. von Lilienfeld, en, in *Many-Electron Approaches in Physics, Chemistry and Mathematics: A Multidisciplinary View*, edited by V. Bach and L. Delle Site, Mathematical Physics Studies (Springer International Publishing, Cham, 2014), pp. 169–189.
 - [154] B. Huang and O. A. V. Lilienfeld, “Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity”, J. Chem. Phys. **145**, 241730 (2016).
 - [155] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, “Machine Learning a General-Purpose Interatomic Potential for Silicon”, Phys. Rev. X **8**, 041048 (2018).
 - [156] C. R. Collins, G. J. Gordon, O. A. von Lilienfeld, and D. J. Yaron, “Constant size descriptors for accurate machine learning models of molecular properties”, J. Chem. Phys. **148**, 241718 (2018).

- [157] K. T. Schütt, M. Gastegger, A. Tkatchenko, K. R. Müller, and R. J. Maurer, “Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions”, *Nat. Commun.* **10**, 1 (2019).
- [158] O. A. von Lilienfeld and K. Burke, “Retrospective on a decade of machine learning for chemical discovery”, *Nat. Commun.* **11**, 1 (2020).
- [159] A. Y.-T. Wang, R. J. Murdock, S. K. Kauwe, A. O. Oliynyk, A. Gurlo, J. Brgoch, K. A. Persson, and T. D. Sparks, “Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices”, *Chem. Mater.* **32**, 4954 (2020).
- [160] Z. Song, X. Chen, F. Meng, G. Cheng, C. Wang, Z. Sun, and W.-J. Yin, “Machine learning in materials design: Algorithm and application”, en, *Chinese Phys. B* **29**, 116103 (2020).
- [161] A. Tkatchenko, “Machine learning for chemical discovery”, en, *Nat. Commun.* **11**, 4125 (2020).
- [162] O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, “Exploring chemical compound space with quantum-based machine learning”, en, *Nat. Rev. Chem.* **4**, 347 (2020).
- [163] J. Westermayr and P. Marquetand, “Machine Learning for Electronically Excited States of Molecules”, *Chem. Rev.* (2020).
- [164] J. Westermayr and R. J. Maurer, “Physically inspired deep learning of molecular excitations and photoemission spectra”, *Chem. Sci.* **12**, 10755 (2021).
- [165] Y. Juan, Y. Dai, Y. Yang, and J. Zhang, “Accelerating materials discovery using machine learning”, en, *Journal of Materials Science & Technology* **79**, 178 (2021).

-
- [166] C. Rauer and T. Bereau, “Hydration free energies from kernel-based machine learning: Compound-database bias”, *J. Chem. Phys.* **153**, 014101 (2020).
 - [167] S. S. Dong, M. Govoni, and G. Galli, “Machine learning dielectric screening for the simulation of excited state properties of molecules and materials”, *en, Chem. Sci.* **12**, 4970 (2021).
 - [168] P. O. Dral and M. Barbatti, “Molecular excited states through a machine learning lens”, *en, Nat. Rev. Chem.*, 1 (2021).
 - [169] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. V. Lilienfeld, “Quantum chemistry structures and properties of 134 kilo molecules”, *Sci. Data* **1**, 1 (2014).
 - [170] N. Meftahi, M. Klymenko, A. J. Christofferson, U. Bach, D. A. Winkler, and S. P. Russo, “Machine learning property prediction for organic photovoltaic devices”, *en, Npj Comput. Mater.* **6**, 1 (2020).
 - [171] Z. Liu, L. Lin, Q. Jia, Z. Cheng, Y. Jiang, Y. Guo, and J. Ma, “Transferable Multilevel Attention Neural Network for Accurate Prediction of Quantum Chemistry Properties via Multitask Learning”, *J. Chem. Inf. Model.* **61**, 1066 (2021).
 - [172] K. Ghosh, A. Stuke, M. Todorović, P. B. Jørgensen, M. N. Schmidt, A. Vehtari, and P. Rinke, “Deep learning spectroscopy: Neural networks for molecular excitation spectra”, *Adv. Sci.* **6**, 1801367 (2019).
 - [173] R. S. Mulliken, “Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I”, *J. Chem. Phys.* **23**, 1833 (1955).
 - [174] A. J. Stone, “Distributed Multipole Analysis: Stability for Large Basis Sets”, *J. Chem. Theory Comput.* **1**, 1128 (2005).

- [175] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, “Machine-learning approach for one- and two-body corrections to density functional theory: Applications to molecular and condensed water”, *Phys. Rev. B* **88**, 054104 (2013).
- [176] M. Rupp, “Machine learning for quantum mechanics in a nutshell”, *Int. J. Quantum Chem.* **115**, 1058 (2015).
- [177] P. Zaspel, B. Huang, H. Harbrecht, and O. A. V. Lilienfeld, “Boosting quantum machine learning models with a multilevel combination technique: Pople diagrams revisited”, *J. Chem. Theory Comput.* **15**, 1546 (2019).
- [178] J. Wehner, L. Brombacher, J. Brown, C. Junghans, O. Çaylak, Y. Khalak, P. Madhikar, G. Tirimbò, and B. Baumeier, “Electronic Excitations in Complex Molecular Environments: Many-Body Green’s Functions Theory in VOTCA-XTP”, *J. Chem. Theory Comput.* **14**, 6253 (2018).
- [179] C. Adamo and V. Barone, “Toward reliable density functional methods without adjustable parameters: The PBE0 model”, *The Journal of Chemical Physics* **110**, 6158 (1999).
- [180] F. Weigend and R. Ahlrichs, “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy”, en, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
- [181] F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, “RI-MP2: optimized auxiliary basis sets and demonstration of efficiency”, en, *Chemical Physics Letters* **294**, 143 (1998).
- [182] N. S. Bayliss and E. G. McRae, “Solvent Effects in the Spectra of Acetone, Crotonaldehyde, Nitromethane and Nitrobenzene”, *J. Phys. Chem.* **58**, 1006 (1954).

-
- [183] M. Merchán, B. O. Roos, R. McDiarmid, and X. Xing, “A combined theoretical and experimental determination of the electronic spectrum of acetone”, *J. Chem. Phys.* **104**, 1791 (1996).
 - [184] O. Crescenzi, M. Pavone, F. De Angelis, and V. Barone, “Solvent Effects on the UV ($n \rightarrow \Pi^*$) and NMR (^{13}C and ^{17}O) Spectra of Acetone in Aqueous Solution. An Integrated Car-Parrinello and DFT/PCM Approach”, *J. Phys. Chem. B* **109**, 445 (2005).
 - [185] L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, and W. L. Jorgensen, “LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands”, *Nucleic Acids Research* **45**, W331 (2017).
 - [186] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water”, *J. Chem. Phys.* **79**, 926 (1983).
 - [187] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, “Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids”, *J. Am. Chem. Soc.* **118**, 11225 (1996).
 - [188] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”, en, *SoftwareX* **1-2**, 19 (2015).
 - [189] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, “A smooth particle mesh Ewald method”, *J. Chem. Phys.* **103**, 8577 (1995).
 - [190] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling”, *J. Chem. Phys.* **126**, 014101 (2007).

- [191] L. Verlet, "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules", *Phys. Rev.* **159**, 98 (1967).
- [192] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath", *J. Chem. Phys.* **81**, 3684 (1984).
- [193] M. Rohlfing, P. Krüger, and J. Pollmann, "Efficient scheme for GW quasiparticle band-structure calculations with applications to bulk Si and to the Si(001)-(2x1) surface", *Phys. Rev. B* **52**, 1905 (1995).
- [194] J.-L. Brédas, D. Beljonne, V. Coropceanu, and J. Cornil, "Charge-Transfer and Energy-Transfer Processes in π -Conjugated Oligomers and Polymers: A Molecular Picture", *Chemical Reviews* **104**, 4971 (2004).
- [195] A. A. Bakulin, A. Rao, V. G. Pavelyev, P. H. M. van Loosdrecht, M. S. Pshenichnikov, D. Niedzialek, J. Cornil, D. Beljonne, and R. H. Friend, "The Role of Driving Energy and Delocalized States for Charge Separation in Organic Semiconductors", *English, Science* **335**, 1340 (2012).
- [196] C. Poelking, M. Tietze, C. Elschner, S. Olthof, D. Hertel, B. Baumeier, F. Würthner, K. Meerholz, K. Leo, and D. Andrienko, "Impact of mesoscale order on open-circuit voltage in organic solar cells", *en, Nature Materials* **14**, 434 (2015).
- [197] C. Duan, R. E. M. Willem, J. J. van Franeker, B. J. Bruijnaers, M. M. Wienk, and R. A. J. Janssen, "Effect of side chain length on the charge transport, morphology, and photovoltaic performance of conjugated polymers in bulk heterojunction solar cells", *en, Journal of Materials Chemistry A* **4**, 1855 (2016).
- [198] P. M. Borsenberger, L. Pautmeier, and H. Bässler, "Charge transport in disordered molecular solids", *The Journal of Chemical Physics* **94**, 5447 (1991).

-
- [199] H. Bässler, “Charge Transport in Disordered Organic Photoconductors a Monte Carlo Simulation Study”, en, *physica status solidi (b)* **175**, 15 (1993).
 - [200] S. V. Novikov, D. H. Dunlap, V. M. Kenkre, P. E. Parris, and A. V. Vannikov, “Essential Role of Correlations in Governing Charge Transport in Disordered Organic Materials”, *Physical Review Letters* **81**, 4472 (1998).
 - [201] W. F. Pasveer, J. Cottaar, C. Tanase, R. Coehoorn, P. A. Bobbert, P. W. M. Blom, D. M. de Leeuw, and M. A. J. Michels, “Unified Description of Charge-Carrier Mobilities in Disordered Semiconducting Polymers”, *Physical Review Letters* **94**, 206601 (2005).
 - [202] R. Coehoorn, W. F. Pasveer, P. A. Bobbert, and M. A. J. Michels, “Charge-carrier concentration dependence of the hopping mobility in organic materials with Gaussian disorder”, *Physical Review B* **72**, 155206 (2005).
 - [203] J. Cottaar and P. A. Bobbert, “Calculating charge-carrier mobilities in disordered semiconducting polymers: Mean field and beyond”, *Physical Review B* **74**, 115204 (2006).
 - [204] V. Rühle, A. Lukyanov, F. May, M. Schrader, T. Vehoff, J. Kirkpatrick, B. Baumeier, and D. Andrienko, “Microscopic Simulations of Charge Transport in Disordered Organic Semiconductors”, *Journal of Chemical Theory and Computation* **7**, 3335 (2011).
 - [205] D. Beljonne, J. Cornil, L. Muccioli, C. Zannoni, J.-L. Brédas, and F. Castet, “Electronic Processes at Organic-Organic Interfaces: Insight from Modeling and Implications for Opto-electronic Devices†”, *Chem. Mater.* **23**, 591 (2010).
 - [206] X. de Vries, P. Friederich, W. Wenzel, R. Coehoorn, and P. A. Bobbert, “Full quantum treatment of charge dynamics in amorphous molecular semiconductors”, *Physical Review B* **97**, 075203 (2018).

- [207] C. Risko, M. D. McGehee, and J.-L. Brédas, “A quantum-chemical perspective into low optical-gap polymers for highly-efficient organic solar cells”, en, *Chemical Science* **2**, 1200 (2011).
- [208] M. Schrader, R. Fitzner, M. Hein, C. Elschner, B. Baumeier, K. Leo, M. Riede, P. Baeuerle, and D. Andrienko, “Comparative Study of Microscopic Charge Dynamics in Crystalline Acceptor-Substituted Oligothiophenes”, *Journal of the American Chemical Society* **134**, 6052 (2012).
- [209] F. May, B. Baumeier, C. Lennartz, and D. Andrienko, “Can Lattice Models Predict the Density of States of Amorphous Organic Semiconductors?”, *Physical Review Letters* **109**, 136401 (2012).
- [210] E. F. Valeev, V. Coropceanu, D. A. da Silva Filho, S. Salman, and J.-L. Brédas, “Effect of Electronic Polarization on Charge-Transport Parameters in Molecular Organic Semiconductors”, *Journal of the American Chemical Society* **128**, 9882 (2006).
- [211] B. Baumeier, J. Kirkpatrick, and D. Andrienko, “Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies”, en, *Physical Chemistry Chemical Physics* **12**, 11103 (2010).
- [212] B. Baumeier, O. Stenzel, C. Poelking, D. Andrienko, and V. Schmidt, “Stochastic modeling of molecular charge transport networks”, *Physical Review B* **86**, 184202 (2012).
- [213] P. Kordt, O. Stenzel, B. Baumeier, V. Schmidt, and D. Andrienko, “Parametrization of Extended Gaussian Disorder Models from Microscopic Charge Transport Simulations”, *Journal of Chemical Theory and Computation* **10**, 2508 (2014).
- [214] O. Schütt and J. VandeVondele, “Machine Learning Adaptive Basis Sets for Efficient Large Scale Density Functional Theory Simulation”, *Journal of Chemical Theory and Computation* **14**, 4168 (2018).

- [215] M. Misra, D. Andrienko, B. Baumeier, J.-L. Faulon, and O. A. von Lilienfeld, “Toward Quantitative Structure–Property Relationships for Charge Transfer Rates of Polycyclic Aromatic Hydrocarbons”, *Journal of Chemical Theory and Computation* **7**, 2549 (2011).
- [216] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller, “Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies”, *Journal of Chemical Theory and Computation* **9**, 3404 (2013).
- [217] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, “Quantum chemistry structures and properties of 134 kilo molecules”, en, *Scientific Data* **1**, 140022 (2014).
- [218] R. Ramakrishnan, M. Hartmann, E. Tapavicza, and O. A. von Lilienfeld, “Electronic spectra from TDDFT and machine learning in chemical space”, *The Journal of Chemical Physics* **143**, 084111 (2015).
- [219] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, “Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning”, *Physical Review Letters* **108**, 058301 (2012).
- [220] S. O. Haykin, *Neural Networks and Learning Machines*, en (Pearson Education, Nov. 2011).
- [221] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st (Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989).
- [222] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, “The missing term in effective pair potentials”, *The Journal of Physical Chemistry* **91**, 6269 (1987).
- [223] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling”, *The Journal of Chemical Physics* **126**, 014101 (2007).

- [224] A. Lukyanov, C. Lennartz, and D. Andrienko, “Amorphous films of tris(8-hydroxyquinolinato)aluminium: Force-field, morphology, and charge transport”, en, *physica status solidi (a)* **206**, 2737 (2009).
- [225] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, “GROMACS: Fast, flexible, and free”, en, *Journal of Computational Chemistry* **26**, 1701 (2005).
- [226] R. A. Marcus, “Electron transfer reactions in chemistry. Theory and experiment”, *Reviews of Modern Physics* **65**, 599 (1993).
- [227] A. D. Becke, “Density-functional thermochemistry. III. The role of exact exchange”, *The Journal of Chemical Physics* **98**, 5648 (1993).
- [228] F. Weigend and R. Ahlrichs, “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy”, en, *Physical Chemistry Chemical Physics* **7**, 3297 (2005).
- [229] B. Thole, “Molecular polarizabilities calculated with a modified dipole interaction”, *Chemical Physics* **59**, 341 (1981).
- [230] P. T. van Duijnen and M. Swart, “Molecular and Atomic Polarizabilities: Thole’s Model Revisited”, *J. Phys. Chem. A* **102**, 2399 (1998).
- [231] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized Gradient Approximation Made Simple”, *Physical Review Letters* **77**, 3865 (1996).
- [232] F. Neese, “The ORCA program system”, en, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2**, 73 (2012).
- [233] J. Wehner, L. Brombacher, J. Brown, C. Junghans, O. Çaylak, Y. Khalak, P. Madhikar, G. Tirimbò, and B. Baumeier, “Electronic Excitations in Complex Molecular Environments: Many-Body Green’s Functions Theory in VOTCA-XTP”, *Journal of Chemical Theory and Computation* **14**, 6253 (2018).

-
- [234] D. C. Elton, Z. Boukouvalas, M. S. Butrico, M. D. Fuge, and P. W. Chung, “Applying machine learning techniques to predict the properties of energetic materials”, en, *Scientific Reports* **8**, 9059 (2018).
 - [235] D. Floreano, P. Dürr, and C. Mattiussi, “Neuroevolution: from architectures to learning”, en, *Evolutionary Intelligence* **1**, 47 (2008).
 - [236] A. Yaman, D. C. Mocanu, G. Iacca, G. Fletcher, and M. Pechenizkiy, “Limited Evaluation Cooperative Co-evolutionary Differential Evolution for Large-scale Neuroevolution”, in *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO ’18* (2018), pp. 569–576.
 - [237] R. Miikkulainen, J. Liang, E. Meyerson, A. Rawal, D. Fink, O. Francon, B. Raju, H. Shahrzad, A. Navruzyan, N. Duffy, and B. Hodjat, “Evolving Deep Neural Networks”, arXiv:1703.00548 [cs] (2017).
 - [238] T. Vehoff, B. Baumeier, A. Troisi, and D. Andrienko, “Charge Transport in Organic Crystals: Role of Disorder and Topological Connectivity”, *Journal of the American Chemical Society* **132**, 11702 (2010).
 - [239] F. J. Ayala and J. C. Avise, *Essential Readings in Evolutionary Biology*, en (JHU Press, Mar. 2014).
 - [240] P. Borsenberger, L. Pautmeier, and H. Bässler, “Nondispersive-to-dispersive charge-transport transition in disordered molecular solids”, *Physical Review B* **46**, 12145 (1992).
 - [241] J. Hermann, Z. Schätzle, and F. Noé, “Deep-neural-network solution of the electronic schrödinger equation”, *Nature Chemistry* **12**, 891 (2020).
 - [242] Y. Wang, Y. Liao, and H. Xie, “Solving schrödinger equation using tensor neural network”, (2022).
 - [243] A. Radu and C. A. Duque, “Neural network approaches for solving schrödinger equation in arbitrary quantum wells”, *Scientific Reports* **12**, 2535 (2022).

-
- [244] J. Han, L. Zhang, and W. E, “Solving many-electron schrödinger equation using deep neural networks”, *Journal of Computational Physics* **399**, 108929 (2019).
 - [245] K.-F. Pu, H.-L. Li, H.-L. Lü, and L.-G. Pang, “Solving schrodinger equations using a physically constrained neural network*”, *Chinese Physics C* **47**, 054104 (2023).