Imperial College London
Department of Materials

Machine Learning static RPA response properties for accelerating GW calculations

Mario Gernot Zauchner

Submitted in part fulfilment of the requirements
for the degree of Doctor of Philosophy at
Imperial College London, 2023

I hereby declare that all material presented in this thesis is my own work, except where other wise noted.

Mario Gernot Zauchner, 2023

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

In this thesis, I explore the possibility of constructing machine-learning models of the interacting density-density response function (DDRF) and quantities derived from it. Accurate models of the DDRF are a crucial ingredient to enabling GW quasiparticle calculations of more complex systems. Model DDRFs bypass the expensive calculation and inversion of the dielectric matrix, which is the origin of the poor scaling of the GW method with the number of atoms.

The thesis is organized as follows:

- Chapter 2 systematically reviews common descriptors used for machine-learning physical quantities. The key ideas behind the construction of such descriptors are discussed. First, I introduce several descriptors that systematically incorporate symmetry transformations that leave the target quantity invariant. These descriptors can be used for learning quantities such as the ground-state energy, atomization energies and scalar polarizabilities. Next, I discuss several descriptors and models that are equivariant under transformations of the molecular structure. These descriptors are ideal for learning quantities which transform in a defined way under the action of a transformation, such as vectors, tensors and functions, including the DDRF.

- In Chapter 3, I introduce the key electronic structure methods employed throughout the thesis. I start by introducing density functional theory, followed by a detailed introduction to the GW method and the DDRF.

- In Chapter 4, I develop a machine-learning model of an invariant quantity derived from the random phase approximation (RPA) DDRF: the scalar polarizability. In this chapter, I calculate the DDRF of 110 hydrogenated silicon clusters. The results of these calculations are then used to train a model of the scalar polarizability based on the SOAP descriptor [16]. The resulting model is then used to predict the scalar polarizability of clusters with up to 3000 silicon atoms while converging to the correct silicon scalar polarizability bulk limit. The findings of this chapter indicate that the scalar polarizability - even though derived from the non-local DDRF - can be accurately predicted from structural descriptors that only encode the local environment of each atom. These results indicate that the response of a non-metallic system to an external potential described by the DDRF may also be

approximated as a sum of localized atomic contributions, which forms the motivation for the following two chapters.

- In Chapter 5, I develop an approximation to the DDRF of the silicon clusters based on a projection onto atom-centred auxiliary density-fitting basis sets. The results of this chapter indicate that the plane-wave DDRF can be efficiently represented by a small localized basis, thus significantly reducing the size of the DDRF. At the end of this section, I develop a simple neural-network model of the DDRF in this localized basis, highlighting the necessity for using an equivariant descriptor and motivating the next chapter's developments.
- In Chapter 6, I develop a new approximation to the DDRF, which allows a decomposition into atomic contributions. I further introduce the neighbourhood density matrix (NDM), a non-local extension of the SOAP descriptor [16], which transforms under rotations in the same way as the atomic contributions to the DDRF. The developed method is then applied to the silicon clusters from the previous chapters. Using the NDM, I develop a neural-network model capable of accurately predicting the atomic contributions to the DDRF. These atomic contributions are transformed into a plane-wave basis and summed to obtain the DDRF of a silicon cluster. The predicted DDRFs are then used in GW calculations, which show that the model DDRFs accurately reproduce the quasiparticle energy corrections from GW calculations, as obtained within the atomic decomposition of the DDRF. This methodology can be used to construct arbitrarily complex model DDRFs based on purely structural properties of clusters and nanoparticles, paving the way towards GW calculations of complex systems, such as disordered materials, liquids, interfaces and nanoparticles.

Acknowledgements

First, I would like to thank my supervisors, Johannes Lischner and Andrew Horsfield for their patience, their timely feedback and invaluable expertise. You two are the best supervisors a PhD student could ask for. Not only did you help me navigate the completion and publication of my research, but you also kept encouraging me when nothing seemed to work. Even when the pandemic created a significant change in circumstances for me, you supported me throughout my studies and motivated me in times of despair.

I would also like to thank my collaborators Gábor Csányi and Stefano Dal Forno for their valuable feedback and contributions to my second publication, and first publication in my PhD studies.

Of course, I thank my parents, in particular my mother Elisabeth for enabling my undergraduate studies in London, which ultimately led to me pursuing a PhD. Even though you still can't explain to any of your friends what I worked on, you always supported me in my pursuit of further studies, showed me how proud you are and listened to my attempts at explaining my research.

I also want to thank the Theory and Simulation CDT and all the people I met there. Even though we rarely saw each other after the first year, I will never forget the intellectual exchanges, evening study sessions in our Masters's degree and the night we got a warning for "trespassing on the future King's backyard" (walking through Kensington Gardens at night). Maybe, now that Charles is King, we will finally receive that warning letter.

Finally, I would like to thank my Fiancé Xenia. Thank you for your incredible patience and support throughout the last three years. Even after long nights writing this thesis, you somehow always managed to put a smile on my face. I don't think I could have completed this without you and I will always be grateful for that. To many more years together. *I love you.*

To my mother.

The first principle is that you must not fool yourself and you are the easiest person to fool.
- Richard P. Feynman

Publications

The majority of work presented in this thesis has been published, or is currently under review and has been published as a preprint as detailed below:

1. **Chapter 4**: the work presented in this chapter has been published in IOP Machine Learning: Science and Technology in the manuscript entitled 'Predicting polarizabilities of silicon clusters using local chemical environments' [196]. I contributed to all calculations, analysis of all results and the development of a parallelized implementation of the RPA polarizability calculation code based on Stefano Dal Forno's code for computing absorption cross-sections.

2. **Chapter 5 and 6**: Part of the work in Chapter 5 and the majority of the work in Chapter 6 has been submitted for publication in a manuscript entitled 'Accelerating GW calculations through machine learned dielectric matrices' [197]. A preprint of the manuscript is available at https://arxiv.org/abs/2305.02990. I developed the methodology, implemented all the code and contributed to all numerical calculations and the analysis of all the results.

Contents

Abstract **3**

Acknowledgements **5**

1 Introduction **16**

1.1 Excited state calculations 16

1.2 Model dielectric functions 19

2 Navigating the feature forest **22**

2.1 Invariant representations 23

2.1.1 Coulomb matrix 23

2.1.2 Bag of Bonds 25

2.1.3 Introduction to spectral descriptors 26

2.1.4 Bispectrum 28

2.1.5 SOAP 30

2.1.6 Long-distance equivariant descriptor 33

2.1.7 Wavelet scattering representation 34

2.1.8 Atomic cluster expansion 36

2.2 Equivariant representations 40

2.2.1 Symmetry-adapted SOAP 41

2.2.2 Equivariant neural networks 44

2.3 Discussion 47

3 Electronic structure methods **50**

3.1 Density Functional Theory 51

3.1.1 Kohn-Sham equations 53

3.1.2 Plane-wave DFT 55

3.1.3 Pseudopotentials 56

3.2 Electronic excitations and the GW approximation 57

3.2.1 The interacting Green's function 58

3.2.2 The density-density response function 60

3.2.3 The random phase approximation and electronic screening 62

3.2.4 The self energy 63

3.3 GW in practice 67

3.3.1 Extension of the static dielectric matrix to finite frequencies 67

3.3.2 Coulomb truncation 69

4 Predicting polarizabilities of silicon clusters using local chemical environments 71
4.1 Introduction 71
4.2 Methods 74
  4.2.1 Random Phase Approximation polarizabilities 74
  4.2.2 Kernel ridge regression 76
  4.2.3 Environment descriptors 78
  4.2.4 Learning cluster polarizabilities 79
  4.2.5 Physical models 83
  4.2.6 Generation of clusters 84
  4.2.7 Computational details 84
4.3 Results and Discussion 85
  4.3.1 Fitting polarizabilities 85
  4.3.2 Predicting polarizabilities 88
4.4 Conclusions 94

5 Approximating the RPA density-density response function 97
5.1 Introduction 97
5.2 Recap of the role of the DDRF within GW 99
5.3 A compact representation of the DDRF 100
  5.3.1 Introducing a basis of CGTOs 101
  5.3.2 Evaluation of the Fourier transform 102
  5.3.3 Enforcing of the integration sum-rule of the projected density-density response function 104
  5.3.4 Cartesian vs. solid harmonic Gaussians 105
  5.3.5 Symmetry of the projected density-density response function 106
5.4 Application to silicon clusters 106
  5.4.1 Computational Details 106
  5.4.2 Neural network model 108
  5.4.3 Results and Discussion 110
  5.4.4 Machine Learning the expansion coefficients 115
5.5 Conclusion 118

6 Accelerating GW calculations through machine-learned dielectric matrices 120
6.1 Introduction 120
6.2 Theory 122
  6.2.1 Atomic decomposition of the density-density response function 122
  6.2.2 Density based descriptor 130
6.3 Computational Details 133
  6.3.1 Data generation 133
  6.3.2 DFT and GW calculations 133
  6.3.3 Projection onto intermediate basis 133
  6.3.4 Projection onto atomic basis 134
  6.3.5 Descriptors 135
  6.3.6 Neural network 136
6.4 Results and Discussion 138
6.5 Conclusion 148

7 Conclusions 151

A Further details on many-body perturbation theory **176**
A.1 Second quantization 176
A.2 Field operators 177
A.3 Introduction to Feynman Diagrams 178

B Further details on neural networks **181**
B.1 Backpropagation 181

List of Tables

4.1 Regularization paramaters λ_{av} and λ_{sum} determined from k-fold cross validation at different cut-off radii r_{cut}. 89

6.1 Hydrogen basis 134
6.2 Silicon basis 134

List of Figures

3.1 Diagrammatic expansion of the screened Coulomb interaction within the RPA approximation. 62
3.2 Diagrammatic expansion of the interacting Green's function in terms of the non-interacting Green's function and the Coulomb interaction. 64
3.3 Diagrammatic expansion of the interacting Green's function in terms of the non-interacting Green's function and the proper self-energy Σ. 64
3.4 Diagrammatic expansion of the self energy Σ. 65
3.5 Approximate self-consistent expansion of the self-energy Σ. 66
3.6 Self-energy within the self-consistent GW method Σ. 66
3.7 2D visualization of the cell-box truncation scheme. 69

4.1 (a) RPA scalar polarizability α of hydrogenated silicon clusters versus number of silicon atoms N_{Si}. The polarizability increases approximately linearly with N_{Si}. (b) RPA polarizability divided by N_{Si} shows deviations from linearity. 86
4.2 Comparison of α/N_{Si} (cluster polarizability divided by the number of silicon atoms) from RPA calculations (blue dots) with fits from various models. (a) The black line shows the optimal fit for the linear N_{Si} model, see Eq. 4.38. Orange triangles denote results from the bond polarizability model; see Eq. 4.39. (b) Results obtained using SOAP with the sum kernel (green squares), the average kernel (red squares) and the coherent kernel (black crosses). A cut-off of 2.5 Å was employed. (c) Same as (b) but with a cut-off of 20 Å. 87
4.3 Average mean absolute error (MAE) of α/N_{Si} (the cluster polarizability divided by the number of silicon atoms) from various machine learning models and physical-based models versus the cut-off radius r_{cut} that determines the size of the local chemical environment. Optimal regularization parameters were determined using five-fold cross-validation. Error bars indicate the standard error of the average MAE across the five training and validation sets used in the cross-validation procedure. 90
4.4 Matrix elements of the average SOAP kernel K^{av}_{IJ}, see Eq. (4.31), for cut-offs r_{cut} of 5.0 Å (a) and 17.5 Å (b). Note the rapid decay of the matrix elements along rows and columns when a large cut-off radius is used. 91
4.5 Average kernel test set error as a function of SOAP cut-off. The smallest n_t clusters were included in the training set for each curve. The test set consists of the remaining $100 - n_t$ clusters. 92
4.6 Comparison of RPA results for α/N_{Si} (cluster polarizability divided by the number of silicon atoms) and training and test set predictions of the average kernel model. The training set consists of the (a) $n_t = 60$, (b) $n_t = 70$ and (c) $n_t = 80$ smallest clusters and the test set contains the remaining $100 - n_t$ large clusters. 93

4.7 Cluster polarizability divided by the number of silicon atoms for all clusters of Silicon Quantum Dot database [13] from the average SOAP kernel model with $r_{cut} = 5.0$ Å. Bulk polarizability obtained from Hybertsen et al. [91]. 94

4.8 Atomic polarizabilities of the $Si_{2109}H_{604}$ cluster obtained from the SOAP average kernel method. Shown is a cross-section through the centre of the cluster. (a) Atomic polarizabilities when only silicon chemical environments are used. (b) Atomic polarizabilities when both silicon and hydrogen chemical environments are used. For hydrogen environments $r_{cut} = 1.6$ Å was used and for silicon environments $r_{cut} = 5.0$ Å was used. Large dots represent silicon atoms, and small dots represent hydrogen atoms. 96

5.1 Matrix $\chi^{ij}_{\alpha_i\alpha_j}$ for cluster $Si_{10}H_{18}$. The red rectangles highlight the different types of blocks. Si blocks correspond to silicon onsite terms, Si-Si blocks to silicon-silicon crosssite terms, H to hydrogen onsite, H-H to hydrogen-hydrogen crossite terms and Si-H to silicon-hydrogen crosssite terms. 108

5.2 Mean of the magnitude of onsite Silicon blocks against the number of Silicon atoms in the cluster in the admm-1 and admm-2 CGTO basis. The average magnitude appears to drop off until $N_{Si} = 60$, beyond which it remains roughly constant. 111

5.3 Average magnitude of Silicon-Silicon cross-site terms vs. distance between basis function centres using the admm-1 and admm-2 CGTO basis. Clearly, the cross-site terms drop off sharply with the distance between sites, the density response is localised around a site. 113

5.4 HOMO-LUMO gap against number of silicon atoms in the cluster for plane-wave G_0W_0 and G_0W_0 with projected DDRF using different CGTO basis sets. RPA bulk value obtained from Hybertsen et al. [90]. 114

5.5 HOMO-LUMO gap against number of silicon atoms in the cluster for plane-wave G_0W_0 and G_0W_0 with projected DDRF using different CGTO and SGTO basis sets. The two outliers in the CGTO admm-2 basis were removed to highlight the differences between the SGTO and CGTO basis sets. RPA bulk value obtained from Hybertsen et al. [90]. 116

5.6 First 100 rows and columns of the RPA dielectric matrix as used in plane-wave G_0W_0 and the SGTO dielectric matrix with no s-orbitals. The colour bar indicates the value of the real part of elements of the dielectric matrix. 117

5.7 Predicted vs. actual coefficients of χ_{ij}. The coefficients were predicted using the NN models for each block type. The results shown correspond to the relaxed versions of the clusters. 118

5.8 HOMO-LUMO gaps obtained using the plane-wave G_0W_0, G_0W_0 using the admm-1 CGTO DDRF and the predicted DDRFs obtained using the neural network model (ML-G_0W_0). The negative HOMO-LUMO gaps are a consequence of the reordering of energy levels caused by the QP corrections. 119

6.1 Coefficients of the 2C-DDRF in the admm-2 SGTO basis without s-orbitals for the cluster $Si_{10}H_{18}$. The red boxes indicate the coefficients included in the atomic contribution of the second Si atom. 127

6.2 G_0W_0 HOMO-LUMO gaps of unperturbed silicon clusters, obtained using the full plane-wave GW-method (blue), and ML-G_0W_0 with different loss functions. The models were trained on the unperturbed cluster, thus the resulting HOMO-LUMO gaps are indicative of the training-set accuracy of the models. 138
6.3 Coefficients $\chi^{(i)}_{nlmn'l'm'}$ of a silicon atom in cluster $Si_{10}H_{18}$ using the exact 1C-DDRF (top left) and predicted 1C-DDRFs using the squared-loss model (top right), the real-space loss (bottom left) and 2C-loss (bottom right). 139
6.4 HOMO-LUMO gaps of hydrogenated silicon clusters from DFT-PBE Kohn-Sham eigenvalues, plane-wave G_0W_0 and G_0W_0 calculations using the 2C-DDRF, see section “Atomic decomposition of the density-density response function”. 140
6.5 (a) HOMO-LUMO gaps of hydrogenated silicon clusters from plane-wave G_0W_0 and G_0W_0 calculations using the 2C-DDRF and G_0W_0 calculations using the 1C-DDRF, see section ”Atomic decomposition of the density-density response function”. (b) HOMO and LUMO energies of hydrogenated Si clusters. 142
6.6 (a) Quasiparticle corrections from plane-wave G_0W_0 and G_0W_0 with the 1C-DDRF for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. (b) Histogram of difference in quasiparticle corrections from plane-wave G_0W_0 and G_0W_0 calculations with the 1C-DDRF for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. The mean-field energies are referenced to the middle of the mean-field HOMO-LUMO gap. 143
6.7 HOMO-LUMO gaps of hydrogenated silicon clusters from plane-wave G_0W_0 and G_0W_0 calculations using the 1C-DDRF and ML-G_0W_0. 144
6.8 Histogram of difference in quasiparticle corrections from G_0W_0 using the 1C-DDRF and ML-G_0W_0 for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. The mean-field energies are referenced to the middle of the mean-field HOMO-LUMO gap. The energies of the smallest cluster were excluded. 145
6.9 Histogram of difference in quasiparticle corrections from plane-wave G_0W_0 and ML-G_0W_0 DDRF for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. The mean-field energies are referenced to the middle of the mean-field HOMO-LUMO gap. The energies of the smallest cluster were excluded. 146
6.10 HOMO-LUMO gaps (left panels) and errors in quasiparticles shifts (right panels) from explicit G_0W_0 calculations with the 1C-DDRF and from ML-G_0W_0 trained on clusters containing up to $N_{max} = 60$ Si atoms (upper panels), $N_{max} = 50$ Si atoms (middle panels) and $N_{max} = 40$ Si atoms (lower panels). The red vertical line indicates N_{max}. The panels on the right hand side only contain results for clusters with more Si atoms than N_{max}. The mean-field energies are referenced to the middle of the mean-field HOMO-LUMO gap. 147
6.11 Scalar polarizabilities of unperturbed silicon clusters obtained using the ML 1C-DDRF (red), the exact 1C-DDRF (black), and the RPA-DDRF (orange). . . 149

A.1 Diagrammatic expansion of the interacting Green’s function in terms of the non-interacting Green’s function and the Coulomb interaction. 179

Chapter 1

Introduction

1.1 Excited state calculations

Density functional theory [85, 103] has shown tremendous success in the calculation of ground state properties from first principles. However, it is well known that the fundamental band gap is significantly underestimated when computed using Kohn-Sham eigenvalues [169, 166]. The band gap E_G of a system with N electrons is defined as the difference between its ionization potential I and its electron affinity A [30]

$$E_G = I - A. \tag{1.1}$$

Naively, one might assume that the Kohn-Sham band gap, *i.e.* the difference between the valence band maximum (VBM) and conduction band minimum (CBM) [30]

$$E_G^{KS} = \epsilon_{CBM} - \epsilon_{VBM}, \tag{1.2}$$

yields a good approximation to E_G. However, it can be shown [145, 169] that these two quantities differ by the so-called derivative discontinuity of the exchange-correlation potential Δ_{XC}, which corresponds to a discontinuity in the exchange-correlation energy as a function of the total number of electrons. This is commonly used as an explanation for the underestimation of the band gap [169, 30]. A common procedure for calculating the band gap is the Δ-scf method which estimates E_G as [37]

$$E_G = I - A = E(N+1) + E(N-1) - 2E(N), \tag{1.3}$$

by explicitly calculating the ground state total energies of systems with N, $N+1$ and $N-1$ electrons. This approach significantly improves the HOMO-LUMO gap in molecules [37, 126]. For solids, however, this quantity is problematic as there are no localised states in which the occupation can be changed and adding an electron to a completely delocalized state would not have an effect [126]. Moreover, it should be noted that only the band gap is accessible through the Δ-scf method, and higher excitations are inaccessible.

One approach that yields significantly better estimations of the band gap in solids is the GW approximation to the self-energy Σ, which is based on the self-consistent set of equations for the one-electron Green's function first derived by Hedin in 1965 [79]. The earliest applications of the GW approximation were developed by Strinati et al. in 1982 [175] and later by Hybertsen and Louie [90]. A detailed summary of the GW approximation will follow in Chapter 3, but to motivate this thesis, I will briefly introduce the core concepts. The key idea behind the GW approximation is to replace the local or semi-local exchange-correlation potential V_{xc} by the non-local, frequency-dependent self-energy Σ, yielding corrected quasi-particle energies according to [90]

$$E^{QP}_{n\mathbf{k}} = \epsilon^{DFT}_{n\mathbf{k}} - \langle n\mathbf{k}|V_{xc}|n\mathbf{k}\rangle + \langle n\mathbf{k}|\Sigma|n\mathbf{k}\rangle, \tag{1.4}$$

where evaluating only the diagonal elements of the self-energy operator $\langle n\mathbf{k}|\Sigma|n\mathbf{k}\rangle$ already yields excellent estimates of the band gap, with errors of the order of 0.01 eV [90], compared to the result obtained by also including off-diagonal elements. In the case of silicon, Hybertsen et al. found that the GW approximation yields errors in the band gap as small as 0.12 eV.

The name GW stems from the approximation to the self-energy operator Σ in terms of the many-body one-electron Green's function G and the screened Coulomb interaction W [139]

$$\Sigma = iGW. \tag{1.5}$$

In contrast to the bare Coulomb interaction ν, the screened Coulomb interaction is frequency dependent and is evaluated *via* the inverse dielectric function ϵ^{-1} according to

$$W(\mathbf{r}, \mathbf{r}', \omega) = \int d\mathbf{r_2} \epsilon^{-1}(\mathbf{r}, \mathbf{r_2}, \omega) \nu(\mathbf{r_2}, \mathbf{r}'). \tag{1.6}$$

Within the random-phase approximation (RPA), which will be introduced in Chapter 3, the inverse dielectric function is computed by inverting the dielectric function [90]

$$\epsilon(\mathbf{r}, \mathbf{r}', \omega) = \delta(\mathbf{r} - \mathbf{r}') - \int d\mathbf{r}'' \nu(\mathbf{r}, \mathbf{r}'') \chi_0(\mathbf{r}', \mathbf{r}'', \omega). \tag{1.7}$$

Here χ_0 denotes the non-interacting density-density response function (DDRF), which maps an external potential to the resulting density response of a system of non-interacting electrons. χ_0 is commonly calculated as a sum over states within the Adler-Wiser formulation [3, 195], which I will derive in Chapter 3.

In many cases, it is sufficient to only evaluate χ_0 at $\omega = 0$, since the zero frequency inverse dielectric matrix can be extended to finite frequencies using the generalised plasmon-pole approximation (GPP) [90, 171]. The GPP models the frequency dependence of the dielectric matrix in terms of collective modes called plasmons and will be introduced in detail in Chapter 3. The GPP has been shown to yield excitation energies in excellent agreement with experiment while significantly reducing the computational cost of computing the self-energy [119]. However, the DDRF within the Adler-Wiser formulation requires a sum over empty and occupied states [3, 195]. The summation scales roughly as $1/\epsilon_{n_{max}}$, with $\epsilon_{n_{max}}$ being the highest unoccupied state included in the summation. As such, many empty states are required to converge χ_0, making it one of the critical bottlenecks of GW calculations [139, 153]. The poor convergence properties of χ_0 led to the development of a number of model dielectric functions, which are discussed in the next section.

1.2 Model dielectric functions

In order to understand the development of more advanced dielectric functions, it is instructive first to consider a classical treatment of screening via the Drude model. The Drude model results from a classical treatment of the equations of motion in a free electron gas, where it is assumed that atoms lose their valence electrons to become positive ions [11]. The electrons form a non-interacting electron gas and are scattered randomly by the metallic ions [7]. The response of these electrons to an external electric field $\mathbf{E}(t)$ is given by the classical equation of motion

$$m\frac{\partial^2}{\partial t^2}\mathbf{r}(t) + m\gamma\frac{\partial}{\partial t}\mathbf{r}(t) = -e\mathbf{E}(t), \tag{1.8}$$

where m is the effective mass of an electron, and γ is a characteristic collision frequency [7, 11]. By taking the Fourier transform of the above expression, one obtains

$$\mathbf{r}(\omega) = \frac{e\mathbf{E}(\omega)}{m(\omega^2 + i\omega\gamma)}. \tag{1.9}$$

The macroscopic polarization $\mathbf{P}(\omega)$ resulting from the displaced electrons is then given via the density of charge carriers n according to

$$\mathbf{P}(\omega) = -\frac{ne^2\mathbf{E}(\omega)}{m(\omega^2 + i\omega\gamma)}. \tag{1.10}$$

Since the macroscopic polarization is given via the density response function χ, with $\mathbf{P}(\omega) = \chi(\omega)\mathbf{E}(\omega)$ we find that [7]

$$\chi(\omega) = -\frac{ne^2}{m(\omega^2 + i\omega\gamma)}. \tag{1.11}$$

Using the definition of the dielectric function in terms of the density-density response function as $\epsilon(\omega) = 1 + 4\pi\chi(\omega)$, we obtain the Drude dielectric function [7]

$$\epsilon(\omega) = 1 - \frac{4\pi ne^2}{m(\omega^2 + i\omega\gamma)}, \tag{1.12}$$

which is often written as

$$\epsilon(\omega) = \epsilon_\infty - \frac{\omega_p^2}{(\omega^2 + i\omega\gamma)}, \tag{1.13}$$

where ω_p is the plasma frequency and ϵ_∞ is a parameter that accounts for the contribution of positive ion cores [7]. From Eq. (1.13), it is clear that the Drude model only considers the large wavelength limit $q \to 0$.

One of the first rigorous quantum mechanical treatments of screening in the free electron gas was developed by Lindhard [118], who derived the well-known Lindhard dielectric function of the electron gas. The Lindhard dielectric function is derived from the non-interacting DDRF χ_0 of the free electron gas, for which an analytical solution can be obtained at zero temperature [172]. This allows the determination of the RPA dielectric function according to Eq (1.7). Lindhard [118] showed that the non-interacting density-density response function of the free electron gas can be written as

$$\chi_0(\mathbf{q}, \omega) = \frac{2}{V} \sum_{\mathbf{k}} \frac{f_0(\epsilon_{\mathbf{k}}) - f_0(\epsilon_{\mathbf{k}+\mathbf{q}})}{\hbar\omega - \epsilon_{\mathbf{k}+\mathbf{q}} + \epsilon_{\mathbf{k}} + i\delta}, \tag{1.14}$$

where δ is an infinitesimal constant to shift the poles away from the real line, V is the unit-cell volume, and f_0 is the zero temperature Fermi distribution [172]. This sum can be evaluated by replacing it with an integral over k and by assuming parabolic dispersion in k [7]. Several generalisations have been introduced since then to allow the treatment of semiconductors using model functions. These model functions are usually constructed to obey the Kramers-Kronig relations and the sum rules of the exact dielectric function. For example, the models proposed by Hybertsen, Levine and Louie [114, 92] or the one by Capllini et al. [35, 19] have been applied in the past [136, 23, 76].

One of the key limitations of these models is that they rely on a spatially uniform dielectric constant, which may prevent them from being used for complex systems with spatially dependent screening [153]. In order to apply model dielectric functions to more complex systems, Rohlfing [153] proposed to model the dielectric function as a sum of atomic contributions

$$\epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) = \delta_{\mathbf{G}\mathbf{G}'} + \frac{1}{|\mathbf{G}+\mathbf{q}||\mathbf{G}'+\mathbf{q}|} \sum_{n}^{N} \frac{V_n}{V} \chi_{n\mathbf{G}\mathbf{G}'}(\mathbf{q}), \tag{1.15}$$

where V_n is the atomic volume and $\mathbf{G}$ are reciprocal lattice vectors. The atomic DDRF $\chi_{n\mathbf{G}\mathbf{G}'}(\mathbf{q})$ is computed from the Fourier transform of a Gaussian charge density centred on individual atoms. The atomic volume can be used to assign a different weight to the contribution of each atom to the total density response.

The modelling of the dielectric function as a sum over atomic contributions forms the key motivation for this project: to develop flexible model dielectric functions using machine learning. While machine learning usually focuses on the prediction of scalar quantities, several attempts at predicting more complex quantities such as the ground state electron density have been made [31, 5, 72, 38]. The work by Grisafi et al. [72], who predicted ground state densities in a local atom-centred basis using the equivariant SOAP kernel, is of particular relevance in the context of this thesis. In this thesis, I develop an approach similar to that of Grisafi [72] for the electronic density, where the DDRF is first decomposed into atomic contributions, each of which can then be learned by a neural network model based on a purely structural descriptor. To the best of my knowledge, this work constitutes the first application of machine learning to predict non-local quantities based on purely structural descriptors.

Chapter 2

Navigating the feature forest

Modern machine learning algorithms such as Kernel Ridge Regression (KRR), Gaussian Process Regression (GPR) or Artificial Neural Networks (ANNs) are well understood and widely used, with extensive literature describing them in detail such as Barber [12] or Murphy [132]. However, in order to apply these methods in the context of materials property prediction, information about the structure and chemistry of a given material has to be encoded appropriately in a descriptor, also commonly referred to as a feature vector in the machine learning community. The following sections discuss the key ideas involved in the construction of these descriptors and how to classify them into different types of descriptors. The prediction of scalar properties requires descriptors to be invariant under transformations that leave the target property unchanged. As we will see throughout this chapter, this invariance can be achieved in different ways. For tensor properties, as well as functions, invariant descriptors are no longer sufficient to encode the structure of a molecule or cluster, and further considerations are required. The solution to this problem is discussed in Section 2.2, where I discuss equivariant descriptors, as well as what properties a model, such as a neural network, has to obey in order to respect the behaviour of a target property under the action of certain transformations. These equivariant descriptors are particularly interesting in the context of developing an ML model of the DDRF, as the DDRF is not invariant with respect to many transformations of the underlying structure, such as rotation or translation. As such, an understanding of the ideas behind equivariant descriptors is a key requirement for the development of model DDRFs based on structural descriptors of molecules and clusters.

2.1 Invariant representations

In order to apply machine learning methods to the prediction of scalar properties, descriptors should be constructed such that they are invariant under symmetry transformations of a molecule (permutation, rotation, reflection, translation, etc.) and such that they uniquely define the system. One of the simplest descriptors that characterise a system would be a list of chemical species and Cartesian coordinates of a system. However, in such a list, the ordering of atoms is arbitrary. Thus, two systems which are physically the same can have a very different descriptor [15]. Hence, a number of more elaborate descriptors have been proposed in the literature, such as wavelet scattering invariants [83, 56, 82], the bispectrum [104, 100], the smooth overlap of atomic positions (SOAP) descriptor [15, 36], the bag of bonds descriptor [77] and the Coulomb matrix [159].

2.1.1 Coulomb matrix

The Coulomb matrix is a descriptor inspired by the input to electronic structure calculations. It encodes information about pairwise distances $|\mathbf{r_i} - \mathbf{r_j}|$ and nuclear charges Z_i in a matrix according to [158]

$$M_{ij} = \begin{cases} 0.5 Z_i^n, & \text{if } i = j. \\ \frac{Z_i Z_j}{|\mathbf{r_i} - \mathbf{r_j}|}, & \text{if } i \neq j, \end{cases} \tag{2.1}$$

where n is a positive number.

Using this definition, the Coulomb matrix only depends on distances between atoms and is thus invariant under translations, rotations and reflections of the molecule. However, similar to a list of Cartesian coordinates and atomic species, the ordering of the rows of the Coulomb matrix is arbitrary. Moreover, the dimensions of the matrix depend on the number of atoms in a given system. The dimensionality problem is typically solved by padding smaller Coulomb matrices with zeros, such that they all have the same dimensions as the largest Coulomb matrix in the data set [78]. In order to solve the problem of ordering, a number of approaches have been proposed in the literature, namely the eigenspectrum representation, sorted Coulomb matrices, and using a set of randomly sorted Coulomb matrices [78].

In the eigenspectrum representation proposed by Rupp et al. [160], the Coulomb matrix of a system is diagonalised, and the eigenvalues are sorted by magnitude. The list of sorted eigenvalues is now invariant with respect to permutations of atoms and also retains the invariants of the original Coulomb matrix. While the eigenspectrum representation is invariant under the symmetries of a system, the dimensionality of the descriptor is reduced drastically, which may lead to a loss of information as noted by Hansen et al. [78].

The second approach to construct permutationally invariant Coulomb matrices is achieved by sorting the rows of a given matrix in descending order of their norm [78]. While this approach maintains the dimensionality of the original Coulomb matrix, it may lead to drastic changes in the representation if the ordering of rows changes due to small perturbations to atomic positions [78].

The most robust variant of the Coulomb matrix descriptor consists of using a set of randomly permuted Coulomb matrices. Hansen et al. [78, 130] proposed to construct a vector $|\mathbf{C}|$, containing the norms of rows of Coulomb matrices and to add a noise term ϵ drawn from a normal distribution. A randomly sorted Coulomb matrix is then computed by permuting the rows of the original Coulomb matrix such the permutation sorts $|\mathbf{C}| + \epsilon$ in descending order. This procedure is repeated until a sufficiently large sample size from the set of all valid Coulomb matrices of a system is obtained.

The key disadvantage of this approach is that the computational cost of training and prediction grows with the number of randomly sorted coulomb matrices included in the sample [78].

The Coulomb matrix has been successfully applied to several problems in quantum chemistry, such as the prediction of atomization energies [78, 160, 130], and several other molecular properties, such as polarizabilities, ionization potentials and electron affinities [131]. Furthermore, a number of generalisations have been proposed to extend the applicability of the Coulomb matrix to crystalline systems [57, 167].

2.1.2 Bag of Bonds

The bag of bonds descriptor, originally proposed by Hansen et al. [77] for fitting many-body potentials, is inspired by a descriptor called bag of words, frequently used in the field of natural language processing [99]. The basic idea behind the bag of bonds descriptor is similar to the Coulomb matrix in that information about different bonds is collected in "bags", each bag corresponding to a certain type of bond (e.g. a C-C bond) [77]. The entries in each bag are essentially the same as the entries of the Coulomb matrix, i.e. $Z_i Z_j / |\mathbf{r_i} - \mathbf{r_j}|$, with Z_i and Z_j being the nuclear charges and $\mathbf{r_i}$ and $\mathbf{r_j}$ being the positions of the two atoms forming the bond [77]. Once the entries of each bag are computed, they are sorted by their magnitude. Then, bags are padded with zeros such that the length of all bags in the data set is equal. Finally, all bags are concatenated to form a vector [77]. As such, the bag of bonds descriptor is essentially a different ordering scheme for the Coulomb matrix [42].

While this descriptor, like the Coulomb matrix, is invariant to translations and rotations, it suffers from the same problem as the Coulomb matrix, namely permutational invariance. Hansen et al. [77] solved this problem by employing a consistent sorting scheme for bags across the data set. Another disadvantage of the bag of bonds model is that it is not a unique descriptor. Specifically, it cannot distinguish between systems which have distinct geometries but have the same pairwise distances [77]. This flaw has been demonstrated by Bing et al. [88], who showed that two homomeric molecules yield exactly the same energy curves, even though their structures are distinct. As a result of this, Bing et al. [88] developed the so-called BA-representation, which extends the bag of bonds model to include bond angles and torsions. They showed that adding bond angles and torsions to the representation creates a unique descriptor and, as such, improves the learning rate and accuracy of the bag of bonds model [88].

The bag of bonds descriptor has been successfully applied in constructing many-body potentials, achieving chemical accuracy (1 kcal/mol) [77]. The extended bag of bonds representation has been shown to yield predictive accuracy among a large range of chemical properties and is competitive with many other descriptors, such as SOAP, the Coulomb matrix, and other bag-of-bonds variants [88].

2.1.3 Introduction to spectral descriptors

There is a large class of descriptors derived from rigorous mathematical arguments. These methods use spectral representations by expanding an atomic neighbourhood density in either spherical harmonics (to enforce rotational invariance) or Fourier series (to enforce translational invariance). In order to understand this class of methods, this section starts by introducing the rotationally invariant power spectrum [15] and the 1-D Fourier representation [186]. The starting point for the rotationally invariant power spectrum is the projection of an atomic neighbourhood density $\rho(\mathbf{r}) = \sum_i w_{Zi}\delta(\mathbf{r} - \mathbf{r_i})$ formed by delta functions centred on the position $\mathbf{r_i}$ of the neighbour of a central atom with a charge dependent weighting w_{Z_i} onto the unit sphere according to [15]

$$\rho(\hat{\mathbf{r}}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} c_{lm} Y_{lm}(\hat{\mathbf{r}}). \tag{2.2}$$

The expansion coefficients c_{lm} can be collected in a vector $\mathbf{c_l}$ for each value of l. This vector then transforms under rotation by multiplication with the Wigner $\mathbf{D^l}$ matrix (the rotation matrix for spherical harmonics) [15, 104]

$$\mathbf{c_l} \to \mathbf{D^l c_l}. \tag{2.3}$$

Analogous to the Fourier power spectrum, the coefficients of the spherical harmonic power spectrum p_l are then obtained by calculating the squared norm for each value of l according to

$$p_l = \mathbf{c}_\mathbf{l}^\dagger \mathbf{c}_\mathbf{l}. \tag{2.4}$$

Clearly, this representation is rotationally invariant since $\mathbf{D}^\mathbf{l}$ is a unitary matrix and under rotation we find

$$p_l \rightarrow (\mathbf{D}^\mathbf{l}\mathbf{c}_\mathbf{l})^\dagger \mathbf{D}^\mathbf{l}\mathbf{c}_\mathbf{l} = \mathbf{c}_\mathbf{l}^\dagger \mathbf{c}_\mathbf{l} = p_l.$$

Moreover, the power spectrum is invariant to reflection [15]. Translational invariance is enforced by using local neighbourhood densities centred on each atom rather than constructing a density for the full cluster or molecule. While Bartók et al. [15] proved that the power spectrum is not a complete descriptor, they showed in numerical experiments that representation can be improved by introducing a truncation distance for the number of neighbours to be included in the neighbourhood density. They also found that increasing the cut-off requires an increase in the maximum value of l to maintain an accurate representation of the system. The incompleteness of descriptors based on n-body correlations, such as the power spectrum, has since been investigated in detail by Pozdnyakov et al. [147] and will be discussed later in this chapter.

An idea related to the power spectrum was used by v. Lilienfeld et al. [186], who expanded the global density of a molecule in a sum of atom-centred Gaussian functions, followed by taking its Fourier transform. While the Fourier transform is naturally invariant to translations, it is not invariant to rotations. Rotational invariance is achieved by multiplication with its conjugate and projecting the 3-dimensional Fourier transform onto one dimension yielding a descriptor [186]

$$FD(\omega) = \frac{1}{(2a)^3} e^{-\frac{\omega^2}{2a}} \sum_{ij=0}^{N_{atoms}} Z_i Z_j \cos(\omega r_{ij}), \tag{2.5}$$

where a is a hyperparameter, Z_i are the charges of the respective atoms and ω is a frequency. It should be noted that the sum runs over all atoms in the molecule. The model was further tuned by using a radial distribution function instead of the interatomic distance r_{ij}. The descriptor was tested on atomization enthalpies [186], but was shown to be less accurate than the Coulomb matrix, which may be a consequence of the descriptor not being complete.

In the following section, I will focus on the spectral representation in the spherical harmonic domain and present improvements that can be made to the simple power spectrum presented above.

2.1.4 Bispectrum

The Bispectrum is a descriptor originally introduced by Kondor [104] in the context of image processing and was further developed for use in the context of quantum chemistry by Bartók et al. [15]. In contrast to the previous descriptors, which are mainly constructed ad-hoc, the bispectrum has a rigorous mathematical foundation and symmetries are incorporated by group theoretical arguments.

Construction of the bispectrum descriptor starts from projecting the atomic neighbourhood density $\rho(\mathbf{r}) = \sum_i w_{Zi}\delta(\mathbf{r} - \mathbf{r_i})$, with the summation running over atomic positions $\mathbf{r_i}$ and charge dependent weight factors w_{Zi}, onto the unit sphere and expanding it in a basis of spherical harmonics according to [15]

$$\rho(\hat{\mathbf{r}}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} c_{lm} Y_{lm}(\hat{\mathbf{r}}). \tag{2.6}$$

The starting point for the bispectrum invariants is the tensor product between different angular momentum channels $\mathbf{c_l}$, where $\mathbf{c_l}$ contains all c_{lm} coefficients in Eq. (2.6). The tensor product of two of these vectors transform according to [15]

$$\mathbf{c_{l_1}} \otimes \mathbf{c_{l_2}} \to (\mathbf{D^{l_1}} \otimes \mathbf{D^{l_2}})(\mathbf{c_{l_1}} \otimes \mathbf{c_{l_2}}), \tag{2.7}$$

where $\mathbf{D^{l_i}}$ denote the Wigner matrices, which are the rotation operators for spherical harmonics. The tensor product of the two Wigner-Matrices is reducible [127, 98]. Thus, the tensor product Eq. (2.7) can be rewritten in terms of the matrices of Clebsch-Gordan coefficients $\mathbf{C}^{l_1 l_2}$ and a direct sum of Wigner matrices [15, 104]

$$\mathbf{c_{l_1}} \otimes \mathbf{c_{l_2}} \rightarrow (\mathbf{C^{l_1 l_2}})^\dagger \Bigg(\bigoplus_{l=|l_1-l_1|}^{l_1+l_2} \mathbf{D^l} \Bigg) \mathbf{C^{l_1 l_2}} (\mathbf{c_{l_1}} \otimes \mathbf{c_{l_2}}). \tag{2.8}$$

Since the matrices of Clebsch-Gordan coefficients are unitary [15], we find (by multiplication of Eq. (2.8) with the Clebsch-Gordan matrix) that the vector $\mathbf{C^{l_1 l_2}} \mathbf{c_{l_1}} \otimes \mathbf{c_{l_2}}$ transforms as

$$\mathbf{C^{l_1 l_2}} \mathbf{c_{l_1}} \otimes \mathbf{c_{l_2}} \rightarrow \Bigg(\bigoplus_{l=|l_1-l_1|}^{l_1+l_2} \mathbf{D^l} \Bigg) \mathbf{C^{l_1 l_2}} (\mathbf{c_{l_1}} \otimes \mathbf{c_{l_2}}). \tag{2.9}$$

Finally by rewriting

$$\mathbf{C^{l_1 l_2}} \mathbf{c_{l_1}} \otimes \mathbf{c_{l_2}} = \bigoplus_{l=|l_1-l_2|}^{l_1+2} \mathbf{g_{l,l_1,l_2}}$$

we can define the coefficients of the bispectrum descriptor by multiplication of each $\mathbf{g_{l,l_1,l_2}}$ by the spherical harmonic expansion coefficients $\mathbf{c_l}^\dagger$ as [104, 15]

$$b_{l,l_1,l_2} = \mathbf{c_l}^\dagger \cdot \mathbf{g_{l,l_1,l_2}}. \tag{2.10}$$

Each $\mathbf{g_{l,l_1,l_2}}$ transforms in the same way as $\mathbf{c_l}$. This can be used to trivially show that the bispectrum coefficients are invariant under rotations [15]

$$b_{l,l_1,l_2} = (\mathbf{D^l} \mathbf{c_l})^\dagger \mathbf{D^l} \mathbf{g_{l,l_1,l_2}} = \mathbf{c_l}^\dagger \cdot \mathbf{g_{l,l_1,l_2}}. \tag{2.11}$$

It should be noted that the projection onto the unit sphere does not allow for encoding distances between the central atom and its neighbours [15]. However, the formalism can be extended to encode distances to the neighbours by either introducing radial basis functions [15, 178] or by projecting the neighbourhood density onto the 4D unit sphere [15, 16]. Bartók et al. [15] systematically reviewed the completeness of the bispectrum using a series of reconstruction experiments and showed that in order to obtain an accurate representation for a larger number of neighbours included in the environment, the length of the descriptor vectors has to be increased. As we will see in Section 2.1.5; however, the completeness of the bispectrum has since been disproven. The bispectrum descriptor has been used in the context of fitting potential energy surfaces using Gaussian approximation potentials [16]. Despite the incompleteness, they managed to predict surface energies of different semiconductors with close to DFT accuracy, with the resulting potential energy surfaces yielding elastic constants within 25% of the DFT value [16].

2.1.5 SOAP

The smooth overlap of atomic positions (SOAP) representation is another descriptor from the class of spectral descriptors and was first introduced by Bartók et al. [15]. The key insight that inspired the SOAP representation is that in kernel methods, such as kernel ridge regression or Gaussian process regression, the kernel should define a smooth similarity measure between different atomic environments [36]. Moreover, the kernel measure needs to be invariant with respect to the symmetries of a system, such as translation, permutation, reflection and rotation [36, 15]. In order to fulfil these requirements, Bartók et al. derived the SOAP representation directly from a kernel that adheres to all these conditions.

Specifically, the starting point for the SOAP representation is an expansion of the local neighbourhood density as a sum of atom-centred Gaussians up to a certain cut-off radius

$$\rho_\eta(\mathbf{r}) = \sum_{i=1}^{N_\eta} e^{-\gamma_\eta(\mathbf{r}-\mathbf{r_i})^2}, \tag{2.12}$$

where η denotes a specific element that is present in the environment of an atom and N_η is the number of atoms of that species present in the environment. Furthermore, γ_η is a parameter that can be thought of as the atomic size and can be used to control the sensitivity of the SOAP kernel to differences between atomic environments. The kernel or similarity measure is then constructed as an exponentiated inner product between two chemical environments $\rho = \{\rho_\eta\}$ and $\tilde{\rho_\eta} = \{\tilde{\rho_\eta}\}$ according to [36]

$$k(\rho, \tilde{\rho}) = \sum_\eta \int d\hat{R} \left| \int d\mathbf{r} \rho_\eta(\mathbf{r}) \tilde{\rho}_\eta(\hat{R}\mathbf{r}) \right|^k, \tag{2.13}$$

where $\hat{R}$ is the rotation operator and the exponent k is another hyperparameter, typically set to

$k = 2$ [36]. In a multi-species system, Eq. (2.13) can be generalised by moving the sum over η into the norm, thus adding correlations between different species in the environment. However, for the sake of conciseness, I will avoid this generalisation here. The kernel integral in Eq. (2.13) has the appealing property that it can be evaluated analytically in a basis of spherical harmonics Y_{lm} and orthogonal radial basis functions $g_n(r)$

$$\rho_\eta(\mathbf{r}) = \sum_{nlm} c^{(\eta)}_{nlm} g_n(r) Y_{lm}(\hat{\mathbf{r}}). \tag{2.14}$$

Substituting into Eq. (2.13) for the case $k = 2$, we obtain

$$k(\rho, \tilde{\rho}) = \sum_{\eta} \int d\hat{R} \left| \int d\mathbf{r} \sum_{nlm} \sum_{n'l'm_1} \sum_{m'} c^{(\eta)}_{nlm} c^{(\eta)*}_{n'l'm'} g_n(r) Y_{lm}(\hat{\mathbf{r}}) g_{n'}(r) D^{*}_{m_1 m'}(\hat{R}) Y^{*}_{l'm_1}(\hat{\mathbf{r}}) \right|^2, \tag{2.15}$$

where $D_{m_1 m'}$ are the coefficients of the Wigner-D matrix [155]. Evaluating the integral inside the norm, we get

$$\begin{aligned} k(\rho, \tilde{\rho}) &= \sum_{\eta} \int d\hat{R} \left| \sum_{nlm} \sum_{n'l'm_1} \sum_{m'} c^{(\eta)}_{nlm} \tilde{c}^{(\eta)*}_{n'l'm'}(r) D^{*l}_{m_1 m'}(\hat{R}) \delta_{nn'} \delta_{ll'} \delta_{mm_1} \right|^2 \\ &= \sum_{\eta} \int d\hat{R} \left| \sum_{nlm} \sum_{m'} c^{(\eta)}_{nlm} D^{*l}_{mm'}(\hat{R}) \tilde{c}^{(\eta)*}_{nlm'} \right|^2 \\ &= \sum_{\eta} \int d\hat{R} \sum_{nlm} \sum_{m'} \sum_{n'l'm_1} \sum_{m''} c^{(\eta)}_{nlm} D^{*l}_{mm'}(\hat{R}) \tilde{c}^{(\eta)*}_{nlm'} c^{*(\eta)}_{n'l'm_1} D^{l'}_{m_1 m''}(\hat{R}) \tilde{c}^{(\eta)}_{n'l'm''}. \end{aligned} \tag{2.16}$$

The Wigner-D matrices are orthogonal [155], yielding

$$\begin{aligned} k(\rho, \tilde{\rho}) &= \sum_{\eta} \sum_{nlm} \sum_{m'} \sum_{n'l'm_1} \sum_{m''} \frac{8\pi^2}{2l'+1} c^{(\eta)}_{nlm} \tilde{c}^{(\eta)*}_{nlm'} c^{*(\eta)}_{n'l'm_1} \tilde{c}^{(\eta)}_{n'l'm''} \delta_{mm_1} \delta_{m'm''} \delta_{ll'} \\ &= \sum_{\eta} \sum_{nn'l} \sum_{mm'} \frac{8\pi^2}{2l+1} c^{(\eta)}_{nlm} \tilde{c}^{(\eta)*}_{nlm'} c^{*(\eta)}_{n'lm} \tilde{c}^{(\eta)}_{n'lm'}. \end{aligned} \tag{2.17}$$

As noted by Bartók et al. [15, 193], the summations over m and m' can now be carried out individually for each atomic environment, where the coefficients of the SOAP descriptor are defined as

$$d^{(\eta)}_{nn'l} = \frac{\sqrt{8}\pi}{\sqrt{2l+1}} \sum_{m} c^{(\eta)}_{nlm} (c^{(\eta)}_{n'lm})^*, \tag{2.18}$$

yielding

$$k(\rho, \tilde{\rho}) = \sum_{\eta} \sum_{nn'l} d^{(\eta)}_{nn'l} \tilde{d}^{(\eta)}_{nn'l}, \tag{2.19}$$

where $d^{(\eta)}_{nn'l}$ and $\tilde{d}^{(\eta)}_{nn'l}$ are the coefficients of the descriptor vectors corresponding to the neighbourhood densities ρ and $\tilde{\rho}$, respectively. The factor of $\frac{\sqrt{8}\pi}{\sqrt{2l+1}}$ was added in an Erratum to the original paper by Bartók et al.[15] and slightly improved interatomic potential fits. This result indicates that using SOAP with $k = 2$ is equivalent to computing the power spectrum in Eq. (2.4) of a smooth Gaussian neighbourhood density [15]. Furthermore, it can be shown [36, 15] that for $k = 3$, the SOAP kernel corresponds to computing the bispectrum of the Gaussian neighbourhood densities.

While the power spectrum (or bispectrum) and SOAP are formally equivalent, SOAP has a few advantages. Specifically, expanding the neighbourhood densities in terms of Gaussians instead of delta functions leads to a smoother similarity measure [15]. In practice, this has the benefit of being more resistant to artifacts in the predicted quantity due to highly oscillatory basis functions corresponding to high angular momentum channels [15].

The SOAP representation and kernel have been applied to a large array of problems such as the prediction of atomization energies [45], electron densities [72], scalar- [192] and tensor-polarizabilities [73, 74] and the prediction of grain boundary structures [148]. Despite the success of SOAP in various applications, Pozdnyakov et al. [147] showed that SOAP and other descriptors based on n-body correlation functions are, in fact, not complete and may be degenerate. Specific examples of degenerate pairs of environments in the case of 3-body correlations (power spectrum or $\nu = 2$ SOAP) and four-body correlation functions, which correspond to the bispectrum or $\nu = 3$ SOAP, were provided by Pozdnyakov [147] to demonstrate the incompleteness of the SOAP descriptor. As further pointed out by Pozdnyakov and co-workers, the success of n-body correlation functions like SOAP, despite this incompleteness, is largely attributable to the decomposition of properties such as potential energy surfaces into atomic contributions [147]. Very recently, Nigam et al. [135] developed an extension to the usual framework of constructing descriptors from n-body correlations (such as SOAP), which is based on building descriptors by taking tensor products between the smooth neighbourhood density ρ and one or more localized functions ϕ centred on the distance vectors between neighbouring atoms. This procedure yields descriptors of the form

$$d(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) = \phi(\mathbf{r}_{ij}) \otimes \phi(\mathbf{r}_{ik}) \otimes \rho(\mathbf{r}_i). \tag{2.20}$$

A rotationally invariant descriptor is then obtained by expanding $\phi(\mathbf{r}_{ij})$, $\phi(\mathbf{r}_{ik})$ and $\rho(\mathbf{r}_i)$ in a basis of spherical harmonics and radial basis functions and symmetrizing apprOrpiately [135], yielding

$$d^{ijk}_{n_1n_2n_3l_1l_2l_3} = \sum_{m_1m_2m_3} C^{l_3m_3}_{l_1m_1l_2m_2} \phi^{ij}_{n_1l_1m_1} \phi^{ik}_{n_2l_2m_2} \rho^{i}_{n_3l_3m_3}, \tag{2.21}$$

where $C^{l_3m_3}_{l_1m_1l_2m_2}$ is a Clebsch-Gordan coefficient. Summing over all neighbours j and k would yield the bispectrum, which, as described previously, is incomplete. Thus, the authors in [135] instead pass the features in Eq. (2.21) through a non-linear function before summing over j, k. This procedure helped in eliminating the degeneracies discovered by Pozdnyakov [147].

2.1.6 Long-distance equivariant descriptor

In many cases, it is necessary to incorporate long-range information into a structural descriptor. In order to remedy this issue, Grisafi and Ceriotti developed the Long-distance equivariant (LODE) descriptor [71]. The starting point for the LODE descriptor is very similar to the SOAP descriptor, but the smooth neighbourhood density is replaced by a neighbourhood potential function for each species η

$$V^{(\eta)}_j(\mathbf{r}, p) = \sum_i \int d\mathbf{s} \frac{1}{|\mathbf{s} - \mathbf{r}|^p} \int d\mathbf{t} e^{-\gamma_\eta(\mathbf{t}-\mathbf{r_j})^2} e^{-\gamma_\eta(\mathbf{s}+\mathbf{t}-\mathbf{r_i})^2}. \tag{2.22}$$

Similar to SOAP, the resulting potential function can be expanded in a basis of spherical harmonics and radial basis functions and then symmetrized for rotational invariance, resulting in

$$V^{(\eta,\eta')}_{pnn'l} \propto \sum_m c^{(\eta)}_{pnlm} c^{(\eta')*}_{pn'lm}, \tag{2.23}$$

where $c^{(\eta)}_{pnlm}$ are the expansion coefficients of the neighbourhood potential in the spherical harmonic basis. The LODE descriptor was used in conjunction with the SOAP descriptor in the prediction of the dielectric response of water in the liquid phase and binding curves of charged organic dimers [71] and was found to yield significantly better predictions than using SOAP alone.

2.1.7 Wavelet scattering representation

Wavelet scattering is a technique first introduced by Mallat [122] in the context of image classification. Mallat created a set of translationally and rotationally invariant representations of an image by successive convolutions followed by a non-linear function [122]. This procedure is related to convolutional neural networks in the sense that they apply a sequence of wavelet filters followed by a non-linear "pooling" or averaging operation. In contrast to convolutional neural networks, however, the filters are predefined and not learned from data, which can reduce the number of samples required to achieve accurate results [122]. A variant of this technique has been used by Hirn et al. to generate rotationally and translationally invariant descriptors for 2D molecules [83, 82] by performing convolutions with oriented wavelets. However, it has been noted by Eickenberg et al. [56] that for 3D molecules too many different orientations would be required to cover all possible rotations in 3D. As such, Eickenberg et al. [56] used convolutions with solid harmonic wavelets where the invariance to rotations is derived from the properties of solid harmonic wavelets. While a full account of the theory behind wavelet scattering invariants is beyond the scope of this work, I will provide a brief summary of the basic ideas behind this type of descriptor.

The starting point for the 3D wavelet scattering representation is very similar to the starting point of the SOAP descriptor, where a "naive" electronic density constructed from Gaussians is associated with each molecule i with $N_{at}^{(i)}$ atoms and z_k nuclear charge [56]

$$\rho_i(\mathbf{r}) = \sum_{k}^{N_{at}^{(i)}} c(z_k) g(\mathbf{r} - \mathbf{r_k}), \tag{2.24}$$

where the sum over k runs over all atomic positions $\mathbf{r_k}$ and charges z_k in molecule i and $c(z_k)$ is used to discriminate between different species [56]. In contrast to SOAP, which assigns a local descriptor to each atom in a molecule [15], the scattering transform descriptor encodes an entire molecule. The next step in computing the scattering transform descriptor is to perform a wavelet transform with respect to a solid harmonic wavelet which is formed from a product of solid harmonics and a Gaussian function [56]

$$\psi_{l,m}(r,\phi,\gamma) = \frac{1}{\sqrt{(2\pi)^2}} e^{-\frac{1}{2}r^2} r^l Y_{lm}(\phi,\gamma). \tag{2.25}$$

We then define the wavelet dilated by 2^j (i.e. r is scaled by a factor of 2^{-j}) as $\psi_{l,m,j}$. Next, the wavelet modulus operator $U[l,j_1]$ is defined as a convolution ($*$) of ρ_i with $\psi_{l,m,j}$ [56]

$$U[l,j_1]\rho_i(\mathbf{r}) = \left(\sum_{m=-l}^{l} |\rho_i * \psi_{l,m,j_1}|^2 \right)^{1/2}, \tag{2.26}$$

where the sum over m results in rotational covariance [56]. Application of the operator $U[l,j_1]$ is equivalent to the filtering operators in convolutional neural networks [56]. Integration (or summation in a discrete setting) over all spatial variables then yields a set of rotationally invariant first-order scattering coefficients

$$S[l,j_1,q]\rho_i = \int d^3\mathbf{r} \left| U[l,j_1]\rho_i(\mathbf{r}) \right|^q, \tag{2.27}$$

where different powers of q can be used to obtain a larger number of coefficients. The second-order scattering coefficients are then obtained by applying the operator $U[l,j_2]$ onto Eq. (2.26) and integrating over all space according to

$$S[l,j_1,j_2,q]\rho_i = \int d^3\mathbf{r} \left| U[l,j_2]U[l,j_1]\rho_i(\mathbf{r}) \right|^q. \tag{2.28}$$

The third order scattering coefficients can then be obtained by applying the operator $U[l,j_3]$ analogously to Eq (2.28); however, it is usually not necessary to go beyond second order [56]. The scattering transform descriptor is then formed by collecting all computed coefficients in a vector.

Mallat [122] showed that in the case of the SOAP descriptor, the changes in power spectrum coefficients resulting from a small deformation are proportional to the cut-off radius. Thus, it can be expected that the utility of the SOAP descriptor deteriorates as the cut-off radius is increased. While the wavelet scattering representation is slightly more involved than many of the descriptors discussed here, it has a number of appealing properties. Most notably, apart from translational and rotational invariance, the scattering coefficients are Lipschitz continuous with respect to deformations [122, 56]. This means that if ρ_i is slightly deformed, the scattering coefficients change by a small amount which is proportional to the deformation of ρ_i [122]. This non-trivial property is particularly useful in the context of learning physical properties, as small changes in structure usually lead to minor changes in a given property. It is well known from the fields of signal- [8] and image-processing [32] that the Fourier transform is not stable to deformation. In particular, high frequencies undergo large shifts when small deformations to the signal occur. This can lead to the representations of a system changing significantly when atoms are moved slightly. Similar arguments can be made for methods based on autocorrelation functions [81], which include the spectral methods described above. When transformed into real space, the power spectrum corresponds to an autocorrelation and the bispectrum to a triple autocorrelation [104]. Since SOAP is, in essence, equivalent to the bispectrum or the power spectrum [15], depending on the exponent of the kernel integral, it can also be said that SOAP is not stable to deformations.

Scattering invariants have been used to predict DFT ground state energies of 2D [82, 81] and 3D [56] molecules, achieving close to state-of-the-art accuracy even when used with simple linear models. Scattering invariants have also been applied to the prediction of grain boundary structures [86].

2.1.8 Atomic cluster expansion

The atomic cluster expansion (ACE) was first introduced by Drautz [51] and can be thought of as a natural generalization of SOAP and related methods to arbitrary body-orders [55]. The fundamental idea of the ACE descriptor is the definition of a scalar property of a certain atom i as a function of all distance vectors $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$

$$E_i(\sigma) \equiv E_i(\mathbf{r}_{1i}, \mathbf{r}_{2i}, \mathbf{r}_{3i}, \ldots, \mathbf{r}_{Ni}). \tag{2.29}$$

Next, a complete and orthogonal single-bond basis $\phi_\nu(\mathbf{r})$ is defined, with $\phi_0 = 1$ [51]. Using the single-bond basis functions, a cluster basis is defined using all single-bond basis functions with $\nu > 0$

$$\Phi_{\alpha\nu} = \phi_{\nu_1}(\mathbf{r}_{j_1 i})\phi_{\nu_2}(\mathbf{r}_{j_2 i})\phi_{\nu_3}(\mathbf{r}_{j_3 i}) \ldots \phi_{\nu_K}(\mathbf{r}_{j_K i}), \tag{2.30}$$

where the cluster index α labels the tuple of all bonds $(j_1 i, j_2 i, \ldots, j_K i)$, in a cluster with K elements, with ν labelling the tuple of all single-bond basis functions $(\nu_1, \nu_2, \ldots, \nu_k)$ [51]. The orthogonality of the cluster basis is inherited from the single-bond basis, which can be seen from

$$\begin{aligned}\langle \Phi_{\alpha\nu} | \Phi_{\beta\mu} \rangle = \delta_{\alpha\beta} \int d\mathbf{r}_{j_1 i} \phi^*_{\nu_1}(\mathbf{r}_{j_1 i})\phi_{\mu_1}(\mathbf{r}_{j_1 i}) \int d\mathbf{r}_{j_2 i} \phi^*_{\nu_2}(\mathbf{r}_{j_2 i})\phi_{\mu_2}(\mathbf{r}_{j_2 i}) \cdots \int d\mathbf{r}_{j_K i} \phi^*_{\nu_K}(\mathbf{r}_{j_K i})\phi_{\mu_K}(\mathbf{r}_{j_K i}) \\ = \delta_{\alpha\beta}\delta_{\nu\mu}, \end{aligned} \tag{2.31}$$

where the term $\delta_{\alpha\beta}$ results from the fact that $\phi_0(\mathbf{r}_{ji}) = 1$ is orthogonal to the other single-bond basis functions [50]: where there are no matching bonds, the integrals will be of the form of

$$\int d\mathbf{r}_{ij} 1 \cdot \phi_\nu(\mathbf{r}_{ij}) = 0 \quad \forall \quad \nu \neq 0. \tag{2.32}$$

Using these results in conjunction with the completeness relation [51] with $\sigma \equiv (\mathbf{r}_{1j}, \mathbf{r}_{2j}, \mathbf{r}_{3j}, \ldots, \mathbf{r}_{N_{ij}})$

$$1 + \sum_{\gamma \subseteq \alpha} \sum_{\nu} \Phi^*_{\gamma\nu}(\sigma)\Phi_{\gamma\nu}(\sigma') = \delta(\sigma - \sigma'), \tag{2.33}$$

the atomic property may be expanded as

$$E_i(\sigma) = J_0 + \sum_{\alpha\nu} J_{\alpha\nu}\Phi_{\alpha\nu}(\sigma). \tag{2.34}$$

In the case of single-species materials, the order of bonds can be permuted arbitrarily, which allows us to replace the coefficients $J_{\alpha\nu}$, by another set of coefficients which is only characterized by the body-order K of the expansion term $J_{\nu}^{(K)}$

$$\begin{aligned} E_i(\sigma) = \sum_{j\nu} J_{\nu}^{(1)} \phi_{\nu}(\mathbf{r}_{ji}) &+ \frac{1}{2!} \sum_{j_1 j_2}^{j_1 \neq j_2} \sum_{\nu_1 \nu_2} J_{\nu_1 \nu_2}^{(2)} \phi_{\nu_1}(\mathbf{r}_{j_1 i}) \phi_{\nu_2}(\mathbf{r}_{j_2 i}) \\ &+ \frac{1}{3!} \sum_{j_1 j_2 j_3}^{j_1 \neq j_2 \neq j_3} \sum_{\nu_1 \nu_2 \nu_3} J_{\nu_1 \nu_2 \nu_3}^{(3)} \phi_{\nu_1}(\mathbf{r}_{j_1 i}) \phi_{\nu_2}(\mathbf{r}_{j_2 i}) \phi_{\nu_3}(\mathbf{r}_{j_3 i}) + \dots . \end{aligned} \tag{2.35}$$

As noted by Drautz [51], the expansion can also be written as an unrestricted sum using a different set of coefficients $c_{\nu}^{(K)}$, where the completeness of the basis is invoked by expanding "self-interaction" terms, i.e. where $j_1 = j_2$, etc., at a lower body order

$$\begin{aligned} E_i(\sigma) = \sum_{j\nu} c_{\nu}^{(1)} \phi_{\nu}(\mathbf{r}_{ji}) &+ \frac{1}{2!} \sum_{j_1 j_2} \sum_{\nu_1 \nu_2} c_{\nu_1 \nu_2}^{(2)} \phi_{\nu_1}(\mathbf{r}_{j_1 i}) \phi_{\nu_2}(\mathbf{r}_{j_2 i}) \\ &+ \frac{1}{3!} \sum_{j_1 j_2 j_3} \sum_{\nu_1 \nu_2 \nu_3} c_{\nu_1 \nu_2 \nu_3}^{(3)} \phi_{\nu_1}(\mathbf{r}_{j_1 i}) \phi_{\nu_2}(\mathbf{r}_{j_2 i}) \phi_{\nu_3}(\mathbf{r}_{j_3 i}) + \dots . \end{aligned} \tag{2.36}$$

One of the challenges with the expansion in Eq. (2.36) is the unfavourable scaling with increasing body-order terms. In order to avoid these issues, Drautz [51] introduces an atomic neighbourhood density $\rho_i(\mathbf{r}) = \sum_j \delta(\mathbf{r} - \mathbf{r}_{ji})$ and a set of coefficients

$$A_{i\nu} = \int d\mathbf{r} \rho_i(\mathbf{r}) \phi_{\nu}(\mathbf{r}) = \sum_j \phi_{\nu}(\mathbf{r}_{ji}). \tag{2.37}$$

Using these coefficients, Eq. (2.36) may be rewritten as

$$E_i(\sigma) = \sum_{\nu} c_{\nu}^{(1)} A_{i\nu} + \frac{1}{2!} \sum_{\nu_1 \nu_2}^{\nu_1 \geq \nu_2} c_{\nu_1 \nu_2}^{(2)} A_{i\nu_1} A_{i\nu_2} + \frac{1}{3!} \sum_{\nu_1 \nu_2 \nu_3}^{\nu_1 \geq \nu_2 \geq \nu_3} c_{\nu_1 \nu_2 \nu_3}^{(3)} A_{i\nu_1} A_{i\nu_2} A_{i\nu_3} + \dots, \tag{2.38}$$

which now scales linearly with the number of atoms. Finally, a basis of spherical harmonics and radial basis functions is introduced

$$\phi_{\nu}(\mathbf{r}) = R_{nl}(\mathbf{r}) Y_{lm}(\hat{\mathbf{r}}). \tag{2.39}$$

Rotational invariance is then imposed by only including coefficients in the expansion that do not change under rotation, thus changing the expansion in Eq. (2.38) to

$$E_i(\sigma) = \sum_{\nu} c_n^{(1)} B_{in}^{(1)} + \frac{1}{2!} \sum_{n_1 n_2 l} c_{n_1 n_2 l}^{(2)} B_{i n_1 n_2 l}^{(2)} + \frac{1}{3!} \sum_{n_1 n_2 n_3 l_1 l_2 l_3} c_{n_1 n_2 n_3 l_1 l_2 l_3}^{(3)} B_{i n_1 n_2 n_3 l_1 l_2 l_3}^{(3)} + \dots, \tag{2.40}$$

where the summation is ordered as in Eq. (2.38) to avoid double counting of equivalent terms [51]. Expressions for the first few expansion coefficients were by provided Drautz [51], the first three are given by

$$B_{in}^{(1)} = A_{in00}, \tag{2.41}$$

$$B_{i n_1 n_2 l}^{(2)} = \sum_{m} (-1)^m A_{i n_1 l m} A_{i n_2 l -m}, \tag{2.42}$$

$$B_{i n_1 n_2 n_3 l_1 l_2 l_3}^{(3)} = \sum_{m_1 m_2 m_3} \frac{(-1)^{l_1 - l_2 - l_3}}{\sqrt{2 l_3 + 1}} C_{l_1 m_1 l_2 m_2}^{l_3 (-m_3)} A_{i n_1 l_1 m_1} A_{i n_2 l_2 m_2} A_{i n_2 l_2 m_2}, \tag{2.43}$$

where $C_{l_1 m_1 l_2 m_2}^{lm}$ are the Clebsch-Gordan coeffcients. The significance of these expressions is that only terms where $m_1 + m_2 = 0$, $m_1 + m_2 + m_3 = 0$, etc., are included by the selection rules of the Clebsch-Gordan coefficients. Drautz [51] demonstrated the power of the atomic cluster expansion by parametrizing an interatomic potential for copper and small clusters, achieving accuracies in energies on par with DFT. Efficient implementations of ACE were developed in the `PACE` [121] software package, and efficient parametrization strategies for ACE were implemented in the `PACEMAKER` [27] software. ACE has been applied to the construction of interatomic potentials [28, 150], on-the-fly learning schemes used in molecular dynamics simulations [120], and the representation of wave functions in terms of ACE [198, 53]. The success of ACE can largely be attributed to the generality and completeness of the expansion [55], which is demonstrated by the fact that many previously developed descriptors such as the bispectrum [16], SOAP [15], Behler-Parinello symmetry functions [22, 10] and moment tensor potentials [170] can be recast in terms of ACE [51, 55], and even wavelet-transforms [162] can be related to the atomic cluster expansion. As such, it can be argued that ACE encompasses many previously developed descriptors. For the SOAP descriptor specifically, we can express SOAP(2) and SOAP(3) in terms of ACE coefficients (obtained from a smooth neighbourhood density instead of the delta-function density) using [51]

$$k^{(2)}(\rho, \rho') = \sum_{nn'l} B^{(2)}_{inn'l}(\rho) B^{(2)}_{in_1n_2l}(\rho'), \tag{2.44}$$

and

$$k^{(3)}(\rho, \rho') = \sum_{n_1n_2n_3l_1l_2l_3} B^{(3)}_{in_1n_2n_3l_1l_2l_3}(\rho) B^{(3)}_{in_1n_2n_3l_1l_2l_3}(\rho'). \tag{2.45}$$

2.2 Equivariant representations

The descriptors discussed so far are all invariant with respect to rotation and translation. This is sufficient in the case of modelling scalar properties, such as potential energy surfaces, scalar polarizabilities or atomization energies. In order to apply ML methods to tensor- and function outputs, such as polarizability tensors, the electron density, or the density-density response function, descriptors that transform in a predictable way are required. In order to understand the construction of descriptors and models that are equivariant with respect to a certain transformation, it is worth formalizing the definition of equivariance, which is also commonly referred to as covariance. Covariance, when it comes to a regression model, is a property that requires the outputs of a model to transform under the action of a group element $g \in G$ in exactly the same way as the inputs. To make this more concrete, we require a bit of jargon from representation theory. A more thorough treatment can be found, for example, in Ref. [98].

A representation D of a group G is defined as a function from G to square matrices, with the following property [179]

$$D(g)D(h) = D(gh) \quad \forall g, h \in G. \tag{2.46}$$

A function $f : V \longrightarrow W$, where V and W are vector spaces, is then said to be covariant with respect to a group G, if it has the following property [179, 105]

$$f(D^V(g)\mathbf{x}) = D^W(g)f(\mathbf{x}) \quad \forall g \in G, \mathbf{x} \in V, f(\mathbf{x}) \in W, \tag{2.47}$$

where we have denoted $D^V(g)$ as the the group transformation matrix of g acting on the vector space V. In simple terms, this means that the action of the function f needs to commute with the group action of g. This means that an ML model can only be covariant if the transformation of the descriptor or model input transforms if the underlying atomic coordinates are transformed. In the case of neural networks, for example, it has been shown by Kondor et al. that a Neural Network layer can only be covariant with respect to a group if it implements a generalized form of convolution [106]

$$(f * g)(u) = \int_G f(uv^{-1})g(v)d\mu(v), \tag{2.48}$$

where f and g are functions from a compact group $G \longrightarrow \mathbb{C}$, and $\mu(v)$ is the Haar Measure.

2.2.1 Symmetry-adapted SOAP

As we have seen before, the SOAP descriptor achieves rotational invariance by integrating over the space of rotation matrices $\hat{R}$, an integral which can be computed analytically when the chemical environment is expanded in a basis of spherical harmonics and radial basis functions [15]. Due to their well-defined transformation properties under rotation, the spherical harmonics form a natural basis for constructing descriptors which are equivariant under the action of elements of SO(3), the rotation group in three dimensions. As such, the SOAP descriptor lends itself to the construction of a descriptor SO(3) covariant descriptors. The construction of such a descriptor was first achieved by Grisafi et al. [73] in the context of machine-learning tensorial properties. The starting point for the so-called λ-SOAP descriptor introduced by Grisafi and co-workers is the scalar SOAP kernel

$$k(\rho, \tilde{\rho}) = \left| \int d\mathbf{r} \rho(\mathbf{r}) \tilde{\rho}(\hat{R}\mathbf{r}) \right|^2. \tag{2.49}$$

For simplicity, I have also dropped the coefficient η, which labels different kinds of atoms in the chemical environments. The neighbourhood densities ρ, are again a sum of Gaussians centred on atoms in the neighbourhood of the central atom. Next, the above integral is augmented by introducing an integral over rotation matrices $\hat{R}$ and the Wigner-D matrix corresponding to the rotation $\mathbf{D}^\lambda(\hat{R})$. This yields a series of matrix-valued kernels of dimension $2\lambda+1 \times 2\lambda+1$

$$\mathbf{k}^\lambda(\rho,\tilde{\rho}) = \int d\hat{R}\mathbf{D}^\lambda(\hat{R})\left|\int d\mathbf{r}\rho(\mathbf{r})\tilde{\rho}(\hat{R}\mathbf{r})\right|^2. \tag{2.50}$$

As noted by Grisafi [73], the original scalar SOAP kernel in Eq. (2.13) is recovered in the case $\lambda = 0$. The coefficients of the descriptor can be computed in a similar way to the standard SOAP descriptor [73], for arbitrary λ. The starting point of the derivation is the evaluation of the norm which was already computed as part of Eq. (2.16)

$$k(\rho,\hat{R}\tilde{\rho}) = \left|\int d\mathbf{r}\rho(\mathbf{r})\tilde{\rho}(\hat{R}\mathbf{r})\right|^2 = \sum_{nlm}\sum_{m'}\sum_{n'l'm_1}\sum_{m''} c_{nlm}D^{*l}_{mm'}(\hat{R})\tilde{c}^*_{nlm'}c^*_{n'l'm_1}D^{l'}_{m_1m''}(\hat{R})\tilde{c}_{n'l'm''}, \tag{2.51}$$

where I removed the species labels for simplicity. The coefficients of the tensor kernel $k^\lambda_{\mu\nu}(\rho,\tilde{\rho})$ are then obtained by computing the Haar integral

$$k^\lambda_{\mu\nu}(\rho,\tilde{\rho}) = \int d\hat{R}D^\lambda_{\mu\nu}(\hat{R})\sum_{nlm}\sum_{m'}\sum_{n'l'm_1}\sum_{m''} c_{nlm}D^{*l}_{mm'}(\hat{R})\tilde{c}^*_{nlm'}c^*_{n'l'm_1}D^{l'}_{m_1m''}(\hat{R})\tilde{c}_{n'l'm''}. \tag{2.52}$$

The Haar integral of three Wigner-D matrices can be evaluated in terms of the Clebsch-Gordan coefficients $C^{LM}_{l_1m_1l_2m_2}$ [155]

$$\int d\hat{R}D^{*l}_{mm'}(\hat{R})D^\lambda_{\mu\nu}(\hat{R})D^{l'}_{m_1m''}(\hat{R}) = \frac{8\pi^2}{2l+1}C^{lm}_{\lambda\mu l'm_1}C^{*lm'}_{\lambda\nu l'm''}. \tag{2.53}$$

Substituting into Eq. (2.52), we obtain,

$$k^\lambda_{\mu\nu}(\rho,\tilde{\rho}) = \sum_{nlm}\sum_{m'}\sum_{n'l'm_1}\sum_{m''}\left(\frac{8\pi^2}{2l+1}\right)C^{lm}_{\lambda\mu l'm_1}C^{lm'*}_{\lambda\nu l'm''}c_{nlm}\tilde{c}^*_{nlm'}c^*_{n'l'm_1}\tilde{c}_{n'l'm''}. \tag{2.54}$$

We can now define the coefficients of the λ-SOAP descriptor as

$$d^{\lambda\mu}_{nn'll'} = \sum_{m_1m}\sqrt{\frac{8\pi^2}{2l+1}}C^{lm}_{\lambda\mu l'm_1}c_{nlm}c^*_{n'l'm_1}, \tag{2.55}$$

and

$$\tilde{d}^{\lambda\nu*}_{nn'll'} = \sum_{m'm''} \sqrt{\frac{8\pi^2}{2l+1}} C^{lm'*}_{\lambda\nu l'm''} \tilde{c}^*_{nlm'} \tilde{c}_{n'l'm''}. \tag{2.56}$$

Similar to the original paper by Grisafi et al. [73], we arrive at an expression for the elements of the kernel in Eq. (2.50) $k^{\lambda}_{\mu\nu}(\rho, \tilde{\rho})$, as a summation over outer products between the descriptors

$$k^{\lambda}_{\mu\nu}(\rho, \tilde{\rho}) = \sum_{nn'll'} d^{\lambda\mu}_{nn'll'} \tilde{d}^{\lambda\nu*}_{nn'll'}, \tag{2.57}$$

or

$$\mathbf{k}^{\lambda}(\rho, \tilde{\rho}) = \sum_{nn'll'} \mathbf{d}^{\lambda}_{nn'll'} \left(\tilde{\mathbf{d}}^{\lambda}_{nn'll'}\right)^{\dagger}. \tag{2.58}$$

It should be noted that each of the vectors now contains $2\lambda + 1$ rows. Now consider how the kernel in Eq. (2.50) changes if the atoms in the environment $\rho(\mathbf{r})$ are rotated by the action of a rotation $\hat{R}_1$. In this case, we obtain

$$\begin{aligned}\mathbf{k}^{\lambda}(\hat{R}_1\rho, \tilde{\rho}) = \int d\hat{R}\mathbf{D}^{\lambda}(\hat{R})k(\hat{R}_1\rho, \hat{R}\tilde{\rho}) = \int d\hat{R}\mathbf{D}^{\lambda}(\hat{R}_1)\mathbf{D}^{\lambda}(\hat{R}_1^{-1})\mathbf{D}^{\lambda}(\hat{R})k(\rho, \hat{R}_1^{-1}\hat{R}\tilde{\rho}) \\ = \mathbf{D}^{\lambda}(\hat{R}_1)\int d\hat{R}'\mathbf{D}^{\lambda}(\hat{R}')k(\rho, \hat{R}'\tilde{\rho}) = \mathbf{D}^{\lambda}(\hat{R}_1)\mathbf{k}'^{\lambda}(\rho, \tilde{\rho})\end{aligned} \tag{2.59}$$

where the invariance of the kernel under simultaneous rotation of environments ρ and $\tilde{\rho}$ was exploited in the second step [73] and $\hat{R}' = \hat{R}_1^{-1}\hat{R}$. The renaming of $d\hat{R}$ to $d\hat{R}'$ is possible due to the left and right invariance of the Haar measure in compact Lie groups [33]. Now consider a Cartesian tensor property α, which can be written as a spherical tensor [98] with components α^{λ}. For a given atomic environment i, the components of the spherical tensor are obtained using

$$\alpha^{\lambda}_i = \sum_j \mathbf{k}^{\lambda}(\rho_i, \rho_j) w_j, \tag{2.60}$$

where w_j are learned weights. Using Eq. (2.59), rotation of the chemical environment i transforms the components of the spherical tensor according to

$$\alpha'^{\lambda}_i = \sum_j \mathbf{k}^{\lambda}(\hat{R}\rho_i, \rho_j) w_j = \mathbf{D}^{\lambda}(\hat{R}) \sum_j \mathbf{k}'^{\lambda}(\rho_i, \rho_j) w_j, \tag{2.61}$$

where $\mathbf{k'}^{\lambda}(\rho_i, \rho_j)$ is the kernel corresponding to the unrotated environment. This proves the covariance of the λ-SOAP kernel according to Eq. (2.47). The λ-SOAP kernel has been used in the prediction of dipole moments, polarizabilities and hyper-polarizabilities of water [73], dipole moments of small organic molecules [184] in the QM9 dataset [26, 157] and the prediction of electronic densities [72, 115, 60].

2.2.2 Equivariant neural networks

As mentioned, the requirements for neural network layers to be equivariant were formalized by Kondor [106] and need to be of the form of a generalized convolution in Eq. (2.48). Many different architectures are discussed in the literature, such as Clebsch-Gordan networks [105, 106, 6], tensor-field networks [179] and spherical CNNs [41, 40]. Furthermore, the `E3NN` library [112, 67] was developed to allow for the construction of many proposed layer architectures. For the purposes of this discussion, I will focus on the Clebsch-Gordan network architecture [105]. This type of network works by forming the tensor product between all angular momentum channels of the layer inputs, followed by the reduction of the tensor product by Clebsch-Gordan summation. The input of the first layer has to be the spherical harmonic transform of the real-space input. The input should be a real-space function encoding information about the molecule or cluster. For example, Kondor [105] and Cohen [41] used a generalized version of the Coulomb matrix as their input function. Here, I will use the smooth electron density as employed in the SOAP descriptor [15]

$$\rho^{(k)}(\mathbf{r}) = \sum_i e^{-\alpha_i (r - r_{ik})^2}. \tag{2.62}$$

As we have seen before, the neighbourhood density can then be expanded in a basis of spherical harmonics and radial basis function, yielding a set of expansion coefficients

$$\rho^{(k)}_{nlm} = \int d\mathbf{r} \rho^{(k)}(\mathbf{r}) Y_{lm}(\hat{\mathbf{r}}) g_n(r). \tag{2.63}$$

These coefficients are then placed into matrices $\rho^{(k)}_l$ according to their angular momentum l. The layer operation of the Clebsch-Gordan network starts by building the tensor product between all pairs of angular momentum channels, which is effectively a quadratic non-linearity [105]. For the first layer, the layer operation is written as

$$\mathbf{g}_{l,l'}^{(k,1)} = \rho_l^{(k)} \otimes \rho_{l'}^{(k)}. \tag{2.64}$$

After building all possible tensor products, a Clebsch-Gordan summation is computed [155]. The result of this summation is given by Kondor [105] as

$$g_{n_1 n_2 l m}^{(k,1)} = \sum_{m_1,m_2} C_{l_1 m_1 l_2 m_2}^{lm} \rho_{n_1 l_1 m_1}^{(k)} \rho_{n_2 l_2 m_2}^{(k)} \tag{2.65}$$

where $C_{l_1 m_1 l_2 m_2}^{lm}$ are the Clebsch-Gordan Coefficients. An equivalent Clebsch-Gordan summation is also employed to generalise the ACE descriptor to tensorial properties [52]. One can show that the Clebsch-Gordan sum commutes with the Wigner-D matrix by assuming that the environment is rotated - i.e. $\rho_{nlm}^{(k)} \to \sum_{m'} D_{mm'}^{l} \rho_{nlm'}^{(k)}$. Substituting this expression into Eq. (2.65), one obtains

$$g_{n_1 n_2 l m}^{(k,1)} = \sum_{m_1,m_2} \sum_{m_3,m_4} C_{l_1 m_1 l_2 m_2}^{lm} D_{m_1 m_3}^{l_1} D_{m_2 m_4}^{l_2} \rho_{n_1 l_1 m_3}^{(k)} \rho_{n_2 l_2 m_4}^{(k)}. \tag{2.66}$$

Now using [4]

$$D_{m_1 m_3}^{l_1} D_{m_2 m_4}^{l_2} = \sum_{l_3 m_5 m_6} D_{m_5 m_6}^{l_3} C_{l_1 m_1 l_2 m_2}^{l_3 m_5} C_{l_1 m_3 l_2 m_4}^{l_3 m_6}, \tag{2.67}$$

and substituting into Eq. (2.66), we obtain

$$g_{n_1 n_2 l m}^{(k,1)} = \sum_{m_1,m_2} \sum_{m_3,m_4} \sum_{l_3 m_5 m_6} D_{m_5 m_6}^{l_3} C_{l_1 m_1 l_2 m_2}^{l_3 m_5} C_{l_1 m_3 l_2 m_4}^{l_3 m_6} C_{l_1 m_1 l_2 m_2}^{lm} \rho_{n_1 l_1 m_3}^{(k)} \rho_{n_2 l_2 m_4}^{(k)}. \tag{2.68}$$

We can now sum over m_1 and m_2 and use the orthogonality relation for the Clebsch Gordan coefficients [4]

$$\delta_{l_3 l} \delta_{m_5 m} = \sum_{m_1,m_2} C_{l_1 m_1 l_2 m_2}^{l_3 m_5} C_{l_1 m_1 l_2 m_2}^{lm}. \tag{2.69}$$

Hence,

$$g_{n_1 n_2 l m}^{(k,1)} = \sum_{m_3,m_4} \sum_{l_3 m_5 m_6} D_{m_5 m_6}^{l_3} C_{l_1 m_3 l_2 m_4}^{l_3 m_6} \delta_{l_3 l} \delta_{m_5 m} \rho_{n_1 l_1 m_3}^{(k)} \rho_{n_2 l_2 m_4}^{(k)}. \tag{2.70}$$

Again summing over m_5 and l_3 yields

$$g^{(k,1)}_{n_1n_2lm} = \sum_{m_3,m_4}\sum_{m_6} D^l_{mm_6} C^{lm_6}_{l_1l_2m_3m_4} \rho^{(k)}_{n_1l_1m_3}\rho^{(k)}_{n_2l_2m_4}, \tag{2.71}$$

which is the same as

$$g^{(k,1)}_{n_1n_2lm} = \sum_{m_6} D^l_{mm_6} g^{(k,1)}_{n_1n_2lm_6}, \tag{2.72}$$

which shows that transforming the inputs $\rho^{(k)}_{n_1l_1m_1}$ of a layer is equivalent to transforming the outputs of a layer. The number of radial channels in Eq. (2.65) grows quadratically in each layer. Moreover, several pairs of tensor products will produce the same output angular momentum l. The fragments with the same angular momentum are collected in a larger matrix. To avoid the exponential growth of n channels, the Clebsch-Gordan network includes a multiplication with a learnable weight matrix [105]. Assuming concatenation of equivalent l channels has been performed, the multiplication with the weights looks as follows

$$o^{(k,S)}_{nlm} = \sum_{n_1n_1} g^{(k,S)}_{n_1n_2lm} W_{n,n_1n_2}, \tag{2.73}$$

where $o^{(k,S)}_{nlm}$ forms the output of layer S, which is then passed to the next layer $(S+1)$, where the process is repeated. In contrast to traditional neural networks, the non-linearity of the Clebsch-Gordan layer is performed before reduction with the weight matrix. In general (ignoring the atomic index k), the output of layer S is given by

$$o^{(S)}_{nlm} = \sum_{n_1n_2} C^{lm}_{l_1m_1l_2m_2} o^{(S-1)}_{n_1lm} o^{(S-1)}_{n_2l'm'} W_{n,n_1n_2}, \tag{2.74}$$

In the final layer of the network, only the required angular momentum channels are output [105]: for scalar properties, only the $l = 0$ outputs are needed. For vectors, outputs with $l = 1$ are used, and for rank-2 tensors, outputs up to $l = 2$ are used, depending on the symmetry of the rank-2 tensor.

2.3 Discussion

The construction of representations for molecules and clusters has seen significant developments over the last few decades. In this chapter, we explored several invariant descriptors, such as the Coulomb matrix [158] and the bag-of-bonds descriptor [77]. While these descriptors are trivially invariant under rotations and translations, their main drawback is that they are not invariant under the permutation of atoms in a system. To achieve permutational invariance, several heuristics have to be employed, such as using consistent orderings of descriptor entries [77], eigenspectrum decompositions [78] or summing over randomly permuted descriptors [130]. Moreover, zero padding is usually required in order to ensure all descriptors in the dataset have the same dimension. Next, we looked at a class of descriptors which are systematically constructed from an atomic neighbourhood density, such as the bispectrum [104, 16], SOAP [15], ACE [51] and wavelet-scattering descriptors [122]. These descriptors impose the invariances of the system by starting from a spectral decomposition and integrating out rotational degrees of freedom. In our exploration of these different spectral descriptors, we saw that there is a close relationship between most spectral descriptors. For example, the power spectrum from Section 2.1.3, as well as the bispectrum from Section 2.1.4, are equivalent to SOAP(2) and SOAP(3) descriptors under the assumption that a smooth atomic neighbourhood density is used. The coefficients of the SOAP descriptor can, in turn, be recast within the ACE framework [51], illustrating that the hierarchy of SOAP descriptors SOAP(n), is equivalent to the n-body cluster coefficients of ACE. Thus, it can be said that ACE is a generalization of previous descriptors based on expansions of the neighbourhood density. While ACE arguably provides the most complete description of a system if the expansion is taken to the maximum order, as stated by Pozdnyakov [147], the common practice of introducing a cutoff sphere beyond which neighbours are ignored introduces a further consideration with regards to the choice of descriptors. As argued by Mallat [81], simply increasing the radius of the cutoff sphere may negatively impact the stability of the descriptor to deformations of the atomic environment. To still include long-range physics typically ignored in density-based descriptors, methods such as the LODE descriptor [71] may be employed in conjunction with descriptors based on n-body correlations. All of these considerations make the selection of appropriate descriptors a challenging problem, and one can find advantages and disadvantages for each specific choice. A reasonable procedure for predicting scalar properties, such as the scalar polarizability, would be to start with a simple descriptor such as SOAP(2) and increase the body order of the SOAP kernel if the incompleteness is found to cause issues. If this is still insufficient, another descriptor, such as ACE, could be employed and long-range information could be incorporated using a long-range descriptor such as LODE. Another consideration in this procedure would be the extrapolation capability of the model. If the model is trained on small molecules or clusters and subsequently used to predict a property of a larger cluster, it may be desirable only to include information about the immediate neighbourhood of an atom to facilitate the recognition of chemical environments across different system sizes [196], as we will see in Chapter 4.

In the case of machine learning functions such as the DDRF, the use of a descriptor that transforms under rotation is required. For example, the symmetry adapted λ-SOAP descriptor was employed by Grisafi [72] who exploited the fact that the spherical harmonics of angular momentum l transform in the same way as a spherical tensor of rank l. While this is a viable strategy for developing models of local functions such as the electron density, it cannot be easily extended to non-local functions. Looking at Eq. (2.58), it is clear that an extension to non-local functions would result in a hierarchy of λ-SOAP kernels indexed by two angular momentum indexes λ_1 and λ_2, each with dimension $(2\lambda_1+1)\times(2\lambda_1+1)\times(2\lambda_2+1)\times(2\lambda_2+1)$. Such a kernel would take the form

$$\mathbf{k}^{\lambda_1,\lambda_2}(\rho,\tilde{\rho}) = \int d\hat{R}\mathbf{D}^{\lambda_1}(\hat{R})\otimes\mathbf{D}^{\lambda_2}(\hat{R})\left|\int d\mathbf{r}d\mathbf{r}'\rho(\mathbf{r},\mathbf{r}')\tilde{\rho}(\hat{R}\mathbf{r},\hat{R}\mathbf{r}')\right|^2, \tag{2.75}$$

for a non-local descriptor $\rho(\mathbf{r},\mathbf{r}')$. Following similar to steps to Eq. (2.59), it can be shown that this kernel transforms as

$$\begin{aligned}\mathbf{k}^{\lambda_1,\lambda_2}(\hat{R}_1\rho,\tilde{\rho}) &= \int d\hat{R}\mathbf{D}^{\lambda_1}(\hat{R})\otimes\mathbf{D}^{\lambda_2}(\hat{R})\left|\int d\mathbf{r}d\mathbf{r}'\rho(\hat{R}_1\mathbf{r},\hat{R}_1\mathbf{r}')\tilde{\rho}(\hat{R}\mathbf{r},\hat{R}\mathbf{r}')\right|^2\\ &= \mathbf{D}^{\lambda_1}(\hat{R}_1)\otimes\mathbf{D}^{\lambda_2}(\hat{R}_1)\mathbf{k}^{\lambda_1,\lambda_2}(\rho,\tilde{\rho}),\end{aligned} \tag{2.76}$$

which would be the desired transformation property for a non-local quantity, such as the DDRF. Thus, while formally possible, an extension of the λ-SOAP framework to non-local functions such as the DDRF may not be practical. Another option would be the use of the Clebsch-Gordan network architecture [105], or general equivariant neural networks constructed using `E3NN` [67, 112]. In order to use this network architecture to learn the density-density response function, the final layer (S) has to be slightly modified. In this layer, the Clebsch-Gordan sum is omitted, and instead, the following expression is evaluated

$$o^{(S)}_{nlmn'l'm'} = \sum_{n_1 n_2} o^{(S-1)}_{n_1 lm} o^{(S-1)}_{n_2 l'm'} W_{nn',n_1 n_2}, \tag{2.77}$$

where W_{nn',n_1n_2} is again a learnable weight matrix and $o^{(N)}_{nlmn'l'm'}$ is a prediction for the expansion coefficients of a non-local function $\chi^{(i)}_{nlmn'l'm'}$, expanded in a basis of spherical harmonics and radial basis functions. This final layer also trivially commutes with the Wigner-D matrix $D^l_{mm_1}(\hat{R})$, as no summation over m is involved. Depending on the number of radial channels in the final layer and the number of radial channels in the output, the weight matrix W_{nn',n_1n_2} may still become prohibitively large. Thus, it is questionable if an equivariant model for non-local functions is practical. This issue is revisited in Chapters 5 and 6, where I develop a machine-learning model of the DDRF.

Chapter 3

Electronic structure methods

The development of the famous Schrödinger equation by the Austrian physicist Erwin Schrödinger [164] formed one of the most impactful discoveries of the 20th century and the starting point for the rigorous mathematical treatment of quantum mechanics. Unfortunately, an analytical solution to the Schrödinger equation is only possible for the most trivial systems, which led to the development of many approximate methods. In this chapter, I will start by introducing density functional theory, one of the most successful approximate methods for finding the ground-state energy of a system. In the second part of this chapter, I introduce the GW method, which is a solution to the famous band-gap problem of density-functional theory [169, 166]. At the heart of these theories lies the many-body Hamiltonian [126] $\hat{H}$ given by

$$\hat{H} = -\frac{1}{2}\sum_{i}\nabla_i^2 + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_J|} - \frac{1}{2}\sum_{I}\frac{\nabla_I^2}{M_I} + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}, \tag{3.1}$$

where Z_I is the number of protons in an atom, and M_I is the mass of the nucleus. Hartree atomic units are used throughout this chapter. The first term in Eq. (3.1) corresponds to the kinetic energy of electrons, the second to electron-electron interactions, the third to the attraction between electrons and nuclei and the fourth and fifth to the kinetic energy of nuclei and nucleus-nucleus repulsion, respectively, where electron positions are given by coordinates $\mathbf{r}_i$ and nuclear coordinates are given by positions $\mathbf{R}_I$. The first step in simplifying Eq. (3.1) is the so-called Born-Oppenheimer approximation [126], in which electron and nuclear wavefunctions are assumed to be separable and electrons respond to nuclear motion adiabatically. This approximation is justified since nuclei are significantly heavier than electrons. Under this approximation, the fourth term can be neglected when solving the electronic Schrödinger equation. If nuclei are further assumed to be completely static, the fifth term results in a constant energy shift to the total energy.

3.1 Density Functional Theory

Density functional theory was first introduced in 1964 by Hohenberg and Kohn [85], where instead of finding an explicit solution to the many-body Schrödinger equation, the problem of finding the ground state of a many-body system was recast by expressing properties of the many-body system as a functional of the ground-state electron density $\rho_0(\mathbf{r})$ [126]. The foundation of density-functional theory is formed by two fundamental theorems [85]:

- **Theorem I**: In a system of N interacting particles in an external potential $V_{\text{ext}}(\mathbf{r})$, the external potential $V_{\text{ext}}(\mathbf{r})$ is uniquely determined (up to a constant) by the ground state density $\rho_0(\mathbf{r})$ of the system.
- **Theorem II**: There exists a universal energy functional of the electron density $E[\rho]$, valid for any external potential $V_{\text{ext}}(\mathbf{r})$. This universal functional is minimized by the ground state density of the N-particle system $\rho_0(\mathbf{r})$.

The first theorem follows from a proof-by-contradiction [126]: Assume there exist two different external potentials $V_{\text{ext}}^{(1)}(\mathbf{r})$ and $V_{\text{ext}}^{(2)}(\mathbf{r})$, which have the same ground state density $\rho_0(\mathbf{r})$. These two external potentials yield differing Hamiltonians and, thus, differing ground state wave functions. It follows that

$$E^{(1)} = \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle < \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} \middle| \Psi^{(2)} \right\rangle. \tag{3.2}$$

This inequality can be rewritten as

$$E^{(1)} < E^{(2)} + \int d\mathbf{r} [V_{\text{ext}}^{(1)}(\mathbf{r}) - V_{\text{ext}}^{(2)}(\mathbf{r})] \rho_0(\mathbf{r}). \tag{3.3}$$

An analogous inequality can be derived by starting at $E^{(2)}$, yielding

$$E^{(2)} < E^{(1)} + \int d\mathbf{r}[V_{\text{ext}}^{(2)}(\mathbf{r}) - V_{\text{ext}}^{(1)}(\mathbf{r})]\rho_0(\mathbf{r}). \tag{3.4}$$

Adding Eq. (3.3) and Eq. (3.4), we arrive at the contradicting inequality

$$E^{(1)} + E^{(2)} < E^{(2)} + E^{(1)}, \tag{3.5}$$

which proves that there cannot be two external potentials yielding the same ground state density. Given that the ground state density uniquely determines the external potential, it follows that each property of the system, including the total energy can be written as a functional of the electronic density $\rho(\mathbf{r})$

$$E[\rho] = T[\rho] + E_{int}[\rho] + \int d\mathbf{r}V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}), \tag{3.6}$$

where $T[\rho]$ is the kinetic energy and $E_{\text{int}}[\rho]$ is the electron-electron interaction energy. These two terms are universal functionals and independent of the external potential. Suppose the ground state density associated with an external potential $V_{\text{ext}}^{(1)}(\mathbf{r})$ is given by $\rho^{(1)}(\mathbf{r})$ and has a corresponding ground state wave function $\Psi^{(1)}$. To prove the second theorem, we define the energy $E^{(1)}[\rho]$ as

$$E^{(1)}[\rho] = T[\rho] + E_{int}[\rho] + \int d\mathbf{r}V_{\text{ext}}^{(1)}(\mathbf{r})\rho(\mathbf{r}). \tag{3.7}$$

A density $\rho^{(2)}(\mathbf{r})$ associated with a different external potential $V_{\text{ext}}^{(2)}(\mathbf{r})$ and wave function $\Psi^{(2)}$ yields a larger energy $E^{(1)}[\rho^{(2)}]$ [85], when evaluated against the Hamiltonian $H^{(1)}$

$$E^{(1)}[\rho^{(1)}] = \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle < \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} \middle| \Psi^{(2)} \right\rangle = E^{(1)}[\rho^{(2)}]. \tag{3.8}$$

Since no assumption was made about $\rho^{(2)}(\mathbf{r})$ (other than it being the ground state density associated with some external potential), it follows that the ground state density yields a lower total energy than any other density [126]. While these two theorems justify the electron density as a central property of a many-body system, they don't provide explicit expressions for the form of the universal density functional. A solution to finding the ground state energy was proposed by Kohn and Sham in 1965 [103], which introduces an auxiliary system of non-interacting electrons.

3.1.1 Kohn-Sham equations

The Ansatz for solving the many-body problem proposed by Kohn and Sham [103] relies on two fundamental assumptions: (1) An auxiliary system of non-interacting electrons representing the exact ground state density exists. (2) The Hamiltonian of this auxiliary system consists of the single-particle kinetic energy operator and an effective potential acting on the electron in question. For the purposes of this presentation, I will ignore the spin index of the wavefunctions, which is a valid assumption in the case of non-magnetic systems. In atomic units, the Hamiltonian of individual electrons in the auxiliary system is given by [126]

$$\hat{H}_{\text{aux}} = -\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r}), \tag{3.9}$$

yielding solutions $\{\psi_i(\mathbf{r}), \epsilon_i\}$. The electron density of the auxiliary system is then defined as

$$\rho(\mathbf{r}) = 2\sum_{i=1}^{N/2} |\psi_i(\mathbf{r})|^2, \tag{3.10}$$

where the factor of 2 is included to account for spin degeneracy. It should be noted that the restricted sum in the above equation is only valid at zero temperature. This auxiliary system gives rise to an expression for the total energy of a system

$$E_{\text{KS}} = T_s[\rho] + \int d\mathbf{r} V_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) + E_{\text{Hartree}}[\rho] + E_{\text{xc}}[\rho], \tag{3.11}$$

where $T_s[\rho]$ is the independent particle kinetic energy [126]

$$T_s = -\sum_{i}^{N/2} \langle \psi_i | \nabla^2 | \psi_i \rangle, \tag{3.12}$$

$E_{\text{Hartree}}[\rho]$ is the Hartree contribution to the total energy,

$$E_{\text{Hartree}}[\rho] = \frac{1}{2}\int d\mathbf{r} \int d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{3.13}$$

$V_{\text{ext}}(\mathbf{r})$ is the external potential, which, in the absence of any external fields, is simply the potential caused by the presence of the nuclei. Finally, $E_{\text{xc}}[\rho]$ combines together all many-body effects and the difference between the single-particle kinetic energy and the true kinetic energy of the many-particle system [126]. A system of linear equations can be derived by taking the functional derivative

$$\frac{\delta E_{\text{KS}}}{\delta \psi_i^*(\mathbf{r})} = \frac{\delta T_s}{\delta \psi_i^*(\mathbf{r})} + \left[\frac{\delta E_{\text{ext}}}{\delta \rho(\mathbf{r})} + \frac{\delta E_{\text{Hartree}}}{\delta \rho(\mathbf{r})} + \frac{\delta E_{\text{xc}}}{\delta \rho(\mathbf{r})}\right] \frac{\delta \rho(\mathbf{r})}{\delta \psi_i^*(\mathbf{r})} - \frac{\delta \left[2 \sum_i^{N/2} \epsilon_i \int d\mathbf{r} \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r})\right]}{\delta \psi_i^*(\mathbf{r})} = 0, \tag{3.14}$$

under the constraint that wavefunctions are normalized. This constraint is given by the last term in the above equation, where the energy ϵ_i plays the role of the Lagrange multiplier. This results in the equation

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{ext}}(\mathbf{r}) + V_{\text{Hartree}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \tag{3.15}$$

where we identify the effective potential V_{eff} with

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + V_{\text{Hartree}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}). \tag{3.16}$$

Eq. (3.15) is usually solved self-consistently: An initial guess for the starting density is generated, and the effective potential is calculated from which wavefunctions are obtained. Then the new density is calculated, and the process repeats. This procedure is iterated until convergence is reached. The only unknown term in Eq. (3.16) is the so-called exchange-correlation potential $V_{\text{xc}}(\mathbf{r})$. The exchange-correlation energy $E_{\text{xc}}[\rho]$ can be written as the Coulomb repulsion of the density of an electron and the so-called exchange-correlation hole charge density $\rho_{\text{xc}}(\mathbf{r}, \mathbf{r}')$ [75]

$$E_{\text{XC}}[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) \int d\mathbf{r}' \frac{\rho_{\text{xc}}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{3.17}$$

The simplest approximation to the exchange-correlation energy is given by the local density approximation (LDA), in which it is assumed that $\epsilon_{\text{xc}}(\rho, \mathbf{r})$ at a given point $\mathbf{r}$ is assumed to be the same as that of a uniform electron gas of the same density as $\rho(\mathbf{r})$, evaluated at point $\mathbf{r}$ [126]. In this case, the exchange-correlation energy can be written as

$$E_{\text{XC}}^{\text{LDA}}[\rho] = \int d\mathbf{r}\rho(\mathbf{r})\epsilon_{\text{xc}}^{\text{LDA}}(\rho(\mathbf{r})). \tag{3.18}$$

In practice, $\epsilon_{\text{xc}}^{\text{LDA}}(\rho(\mathbf{r}))$ is parametrized based on known low- and high-density limits and fitted to the result of quantum Monte Carlo (QMC) calculations, leading to parameterizations such as those obtained by Perdew and Zunger (PZ) [143] or by Vosko, Wilk and Nusair [187]. More sophisticated approximations, such as the generalized gradient approximation (GGA), where the exchange-correlation density depends on the magnitude of the gradient $|\nabla\rho(\mathbf{r})|$ and density at point $\mathbf{r}$. The most famous approximation of this kind is the parametrization proposed by Perdew, Burke and Ernzerhof (PBE) [144]. Approximations beyond the GGA exist in so-called meta-GGA functionals such as the recently popularized SCAN functional [176, 17], or hybrid functionals which include a fraction of the exact Hartree-Fock exchange energy [20, 21].

3.1.2 Plane-wave DFT

Many DFT codes rely on the use of plane-wave basis sets, including Quantum Espresso (QE) [70, 69], which is used throughout this thesis. In plane-wave DFT, single-electron wavefunctions are expanded as [126]

$$\psi_n(\mathbf{r}) = \frac{1}{\sqrt{V}}\sum_{\mathbf{q}} c_{n,\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} \equiv \sum_{\mathbf{q}} c_{n,\mathbf{q}}\langle\mathbf{r}|\mathbf{q}\rangle, \tag{3.19}$$

where V is the crystal volume, $\mathbf{q}$ the crystal momentum and $c_{n,\mathbf{q}}$ a set of expansion coefficients. Eq. (3.9) then becomes

$$\sum_{\mathbf{q}}\langle\mathbf{q}'|\hat{H}_{aux}|\mathbf{q}\rangle c_{n,\mathbf{q}} = \epsilon_n c_{n,\mathbf{q}'}. \tag{3.20}$$

The kinetic energy contribution to the Hamiltonian trivially evaluates to

$$\langle\mathbf{q}'| - \frac{1}{2}\nabla^2|\mathbf{q}\rangle = \frac{1}{2}|q|^2\delta_{\mathbf{q},\mathbf{q}'}. \tag{3.21}$$

The effective potential V_{eff} is unit-cell periodic and thus can be written as a Fourier series

$$V_{eff}(\mathbf{r}) = \sum_{m} V_{eff}(\mathbf{G}_m) e^{i\mathbf{G}_m \cdot \mathbf{r}}, \tag{3.22}$$

where $\mathbf{G}$ are reciprocal lattice vectors. The matrix elements of the effective potential are given by

$$\langle \mathbf{q}' | V_{eff}(\mathbf{r}) | \mathbf{q} \rangle = \frac{1}{V} \sum_{m} \int d\mathbf{r} V_{eff}(\mathbf{G}_m) e^{i[\mathbf{G}_m - (\mathbf{q}' - \mathbf{q})] \cdot \mathbf{r}} = \sum_{m} V_{eff}(\mathbf{G}_m) \delta_{\mathbf{q}-\mathbf{q}', \mathbf{G}_m}. \tag{3.23}$$

Now using the substitutions $\mathbf{q} = \mathbf{k} + \mathbf{G}_m$ and $\mathbf{q}' = \mathbf{k} + \mathbf{G}_{m'}$ [126], where $\mathbf{k}$ is the crystal momentum constrained to the first Brillouin-zone, we arrive at a set of linear equations for each $\mathbf{k}$-point

$$\sum_{m'} H_{mm'}(\mathbf{k}) c_{i,m'}(\mathbf{k}) = \epsilon_i(\mathbf{k}) c_{i,m}(\mathbf{k}), \tag{3.24}$$

where

$$H_{mm'}(\mathbf{k}) = \frac{1}{2} |\mathbf{k} + \mathbf{G}_m|^2 \delta_{mm'} + V_{eff}(\mathbf{G}_m - \mathbf{G}_{m'}). \tag{3.25}$$

In the case where pseudopotentials are used, Eq. (3.25) has to be generalized to non-local effective potentials [126].

3.1.3 Pseudopotentials

In order to avoid all-electron calculations, where core electrons are included explicitly, pseudopotentials are often employed in plane-wave DFT calculations. The first step towards the pseudopotential approximation is the frozen core approximation [168], in which core electrons are assumed to remain in the states occupied in an isolated atom. However, core states are localized around the nucleus and thus rapidly oscillate. In order to maintain orthogonality with the core states, valence states also rapidly oscillate inside the core region [168]; therefore, many plane waves are required to correctly describe the valence states in the vicinity of the nucleus. Even if one replaces the effect of the core electrons with an effective potential, as is done in the effective core approximation [168], valence states still oscillate rapidly in the vicinity of the nucleus. The fundamental idea of the pseudopotential is to replace the valence wave functions with pseudo wavefunctions, which are equal to the valence wave functions in an all-electron calculation of an isolated atom outside a core region with cut-off R_c but are smooth and nodeless inside the core region [168]. As noted by Kresse et al. [107], the pseudo wavefunctions must also be continuously differentiable at least twice at R_c. Another important consideration is that BerkeleyGW, the GW code employed throughout this thesis, requires norm-conserving pseudopotentials [90, 49]. In norm-conserving pseudopotentials, the pseudo wavefunctions still differ from the all-electron wavefunctions; however, the charge enclosed inside the core region is the same [107]. These criteria are enforced when fitting the pseudo wavefunctions on an isolated atom. Once the pseudo wavefunctions are obtained by smoothing the all-electron valence wavefunctions, the radial Schrödinger equation can be inverted to obtain a non-local pseudopotential [107] which replaces the bare Coulomb potential of the nucleus. In this thesis, norm-conserving pseudopotentials fitted using the method by Rappe et al. [151] (RRKJ) and the method by Von Barth and Car [185] (VBC) are employed.

3.2 Electronic excitations and the GW approximation

As mentioned in the introduction, the DFT band gap is significantly underestimated when calculated using the Kohn-Sham energies [145, 169]. In order to tackle this problem, the GW method is often employed. The system of self-consistent equations that form the basis for the GW method was first derived by Hedin [79]; however, these equations are rarely solved self-consistently due to the difficulty of accounting for so-called vertex-corrections [139]. The vertex correction effectively accounts for higher-order perturbative corrections, as will become clear in this section. Thus, in this section, I will develop the central equations of the so-called one-shot $\mathrm{G}_0\mathrm{W}_0$ approach, which is used in all quasiparticle calculations in this thesis. The GW method is based on many-body perturbation theory, where the central quantity is the single-particle Green's function $G(\mathbf{r}, \mathbf{r}', t, t')$. Roughly speaking, the Green's function describes the propagation of a particle added to the system at time t' and point $\mathbf{r}'$ and removed from the system at $\mathbf{r}, t$. To fully understand the significance of the Green's function, familiarity with second quantization and diagrammatic perturbation theory are required. A brief overview of these concepts is provided in Appendix A. For a more comprehensive introduction, I refer the interested reader to Fetter and Walecka [63] Chapter 3.

3.2.1 The interacting Green's function

The Green's function is a fundamental property of many-particle systems, and its utility has been demonstrated in great detail, for example, by Fetter and Walecka [63], Galitskii et al. [66], Klein et al. [102] or Martin et al. [125]. The Green's function is defined as [63]

$$iG_{\alpha\beta}(\mathbf{r},\mathbf{r}',t,t') = \frac{\langle\Psi_0|T[\hat{\psi}_{H\alpha}(\mathbf{r},t)\hat{\psi}^{\dagger}_{H\beta}(\mathbf{r}',t')]|\Psi_0\rangle}{\langle\Psi_0|\Psi_0\rangle}, \tag{3.26}$$

where $|\Psi_0\rangle$ is the interacting ground state in the Heisenberg picture, and $\hat{\psi}_{H\alpha}(\mathbf{r},t)$ is a field operator in the Heisenberg picture (see Appendix A). Moreover, the operator $T[\dots]$ orders operators inside the T-product from left to right in order of increasing time. Additionally, a factor of $(-1)^k$ has to be added, where k is the number of interchanges necessary to achieve the correct time order [63]. We can express Eq. (3.26) more explicitly by using the expressions for operators in the Heisenberg picture and by expressing the time-ordering with Heaviside step functions $\theta(t-t')$

$$\begin{aligned} iG_{\alpha\beta}(\mathbf{r},\mathbf{r}',t,t') = \theta(t-t')\langle\Psi_0|e^{i\hat{H}t}\hat{\psi}_{\alpha}(\mathbf{r})e^{-i\hat{H}t}e^{i\hat{H}t'}\hat{\psi}^{\dagger}_{\beta}(\mathbf{r}')e^{-i\hat{H}t'}|\Psi_0\rangle \\ - \theta(t'-t)\langle\Psi_0|e^{i\hat{H}t'}\hat{\psi}^{\dagger}_{\beta}(\mathbf{r}')e^{-i\hat{H}t'}e^{i\hat{H}t}\hat{\psi}_{\alpha}(\mathbf{r})e^{-i\hat{H}t}|\Psi_0\rangle, \end{aligned} \tag{3.27}$$

where it was further assumed that the ground state is normalized to unity. Applying the operators to the ground states, we obtain

$$\begin{aligned} iG_{\alpha\beta}(\mathbf{r},\mathbf{r}',t,t') = \theta(t-t')e^{iE_0(t-t')}\langle\Psi_0|\hat{\psi}_{\alpha}(\mathbf{r})e^{-i\hat{H}t}\hat{\psi}^{\dagger}_{\beta}(\mathbf{r}')e^{-i\hat{H}t'}|\Psi_0\rangle \\ - \theta(t'-t)e^{-iE_0(t-t')}\langle\Psi_0|\hat{\psi}^{\dagger}_{\beta}(\mathbf{r}')e^{-i\hat{H}t'}e^{i\hat{H}t}\hat{\psi}_{\alpha}(\mathbf{r})|\Psi_0\rangle, \end{aligned} \tag{3.28}$$

where E_0 is the ground-state of the N-electrons system.

Inserting a complete set of states $|\Psi_n\rangle$, including all possible numbers of electrons, in between the exponential operators, we find that

$$\begin{aligned} iG_{\alpha\beta}(\mathbf{r},\mathbf{r}',t,t') = \sum_n \theta(t-t')e^{-i(E_n-E_0)(t-t')}\langle\Psi_0|\hat{\psi}_\alpha(\mathbf{r})|\Psi_n\rangle\langle\Psi_n|\hat{\psi}^\dagger_\beta(\mathbf{r}')|\Psi_0\rangle \\ - \theta(t'-t)e^{i(E_n-E_0)(t-t')}\langle\Psi_0|\hat{\psi}^\dagger_\beta(\mathbf{r}')|\Psi_n\rangle\langle\Psi_n|\hat{\psi}_\alpha(\mathbf{r})|\Psi_0\rangle. \quad (3.29) \end{aligned}$$

Now inserting the integral expression for the step function [63]

$$\theta(t-t') = -\int d\omega \frac{1}{2\pi i}\frac{e^{-i\omega(t-t')}}{\omega+i\eta}, \quad (3.30)$$

and substituting $\omega' = \omega + (E_n - E_0)$ and $\omega' = \omega - (E_n - E_0)$ in the first and second term respectively, we obtain

$$\begin{aligned} iG_{\alpha\beta}(\mathbf{r},\mathbf{r}',t,t') = \sum_n \Bigg\{ -\int d\omega' \frac{1}{2\pi i}\frac{e^{-i(\omega')(t-t')}}{\omega'-(E_n-E_0)+i\eta}\langle\Psi_0|\hat{\psi}_\alpha(\mathbf{r})|\Psi_n\rangle\langle\Psi_n|\hat{\psi}^\dagger_\beta(\mathbf{r}')|\Psi_0\rangle \\ -\int d\omega' \frac{1}{2\pi i}\frac{e^{-i(\omega')(t-t')}}{\omega'+(E_n-E_0)-i\eta}\langle\Psi_0|\hat{\psi}^\dagger_\beta(\mathbf{r}')|\Psi_n\rangle\langle\Psi_n|\hat{\psi}_\alpha(\mathbf{r})|\Psi_0\rangle\Bigg\}. \quad (3.31) \end{aligned}$$

Now computing the Fourier transform, we finally obtain the so-called Lehmann representation [113] of the Green's function

$$G_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega) = \sum_n \left(\frac{\langle\Psi_0|\hat{\psi}_\alpha(\mathbf{r})|\Psi_n\rangle\langle\Psi_n|\hat{\psi}^\dagger_\beta(\mathbf{r}')|\Psi_0\rangle}{\omega-(E_n-E_0)+i\eta} + \frac{\langle\Psi_0|\hat{\psi}^\dagger_\beta(\mathbf{r}')|\Psi_n\rangle\langle\Psi_n|\hat{\psi}_\alpha(\mathbf{r})|\Psi_0\rangle}{\omega+(E_n-E_0)-i\eta} \right). \quad (3.32)$$

Since the quantities in the numerator involve fully-interacting states, they are not practical to compute. However, since the field operators add or remove a particle from the ground state, we can already conclude that only states containing $N \pm 1$ electrons will contribute, as overlaps with the ground state will otherwise vanish. It is also instructive to consider the case where the states $|\Psi_n\rangle$ and the ground state are given by a single Slater determinant, as is the case in Hartree-Fock and Kohn-Sham DFT methods. We will denote the states with an additional or missing electron in state s compared to the ground state $|N \pm 1, s\rangle$. We then obtain

$$G_{\alpha\beta}(\mathbf{r}, \mathbf{r}', \omega) = \sum_s \Bigg\{ \frac{\psi_{\alpha s}(\mathbf{r})\psi^*_{\beta s}(\mathbf{r}')\langle \Psi_0 | \hat{c}_{\alpha s} | N+1, s\rangle \langle N+1, s | \hat{c}^\dagger_{\beta s} | \Psi_0 \rangle}{\omega - (E_s - E) + i\eta} + \frac{\psi^*_{\beta s}(\mathbf{r}')\psi_{\alpha s}(\mathbf{r})\langle \Psi_0 | \hat{c}^\dagger_{\beta s} | N-1, s\rangle \langle N-1, s | \hat{c}_{\alpha s}(\mathbf{r}) | \Psi_0 \rangle}{\omega + (E_s - E) - i\eta} \Bigg\}. \quad (3.33)$$

which is the same as

$$G_{\alpha\beta}(\mathbf{r}, \mathbf{r}', \omega) = \sum_s \Bigg\{ \frac{\psi_{\alpha s}(\mathbf{r})\psi^*_{\beta s}(\mathbf{r}')(1 - f_{\alpha s})(1 - f_{\beta s})}{\omega - \epsilon_s + i\eta} + \frac{\psi^*_{\beta s}(\mathbf{r}')\psi_{\alpha s}(\mathbf{r}) f_{\alpha s} f_{\beta s}}{\omega - \epsilon_s - i\eta} \Bigg\}, \quad (3.34)$$

where $f_{\alpha s}$ is the occupancy of state s with spin-index α. Additionally, the total energy differences will correspond to the Kohn-Sham/Hartree-Fock single-particle energies ϵ_s. The first term in Eq. (3.34) runs over all unoccupied states, while the second term runs over all occupied states. It is important to emphasize that this form of the Green's function is equivalent to the Green's function of a non-interacting system, which is why this form is typically used as an initial guess for self-consistent GW calculations, or as the Green's function G_0 in one-shot G_0W_0 calculations [139]. Additionally, if the Hamiltonian is not spin-dependent, we can further simplify the expression for the Green's function to obtain

$$G_{\alpha\beta}(\mathbf{r}, \mathbf{r}', \omega) = \delta_{\alpha\beta} \sum_s \frac{\psi_s(\mathbf{r})\psi^*_s(\mathbf{r}')(1 - f_s)}{\omega - \epsilon_s + i\eta} + \frac{\psi^*_s(\mathbf{r}')\psi_s(\mathbf{r}) f_s}{\omega - \epsilon_s - i\eta}. \quad (3.35)$$

3.2.2 The density-density response function

As described in the Introduction, the density-density response function maps a perturbing potential to the change in charge density resulting from the perturbation. Formally, the DDRF is defined as [63]

$$\chi(\mathbf{r}, \mathbf{r}', t, t') = \frac{\langle \Psi_0 | T[\hat{n}_{H\alpha}(\mathbf{r}, t)\hat{n}_{H\beta}(\mathbf{r}', t')] | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}, \quad (3.36)$$

where $\hat{n}_{H\alpha}(\mathbf{r}, t)$ is the number-density operator in the Heisenberg picture. Starting from this definition, it can be shown (see, for example, Fetter and Walecka Chapter 4 [63]) that for a non-interacting system, the DDRF in the frequency domain can be written as

$$\chi_0(\mathbf{r},\mathbf{r}',\omega) = -\frac{i}{\pi}\int d\omega' G_0(\mathbf{r},\mathbf{r}',\omega')G_0(\mathbf{r}',\mathbf{r},\omega+\omega'), \tag{3.37}$$

where $G_0(\mathbf{r},\mathbf{r}',\omega')$ is the Green's function of a non-interacting system, defined in Eq. (3.35). The single-particle states $\psi_s(\mathbf{r})$, and corresponding energies are obtained from a non-interacting calculation or a mean-field calculation like Kohn-Sham DFT. In this product, two of the terms will have poles on the same side of the real axis; thus, their contributions vanish. The remaining terms are given by

$$\chi_0(\mathbf{r},\mathbf{r}',\omega) = \sum_{ss'} -\frac{i}{\pi}\int d\omega' \left[\frac{\psi_s(\mathbf{r})\psi_s^*(\mathbf{r}')\psi_{s'}(\mathbf{r}')\psi_{s'}^*(\mathbf{r})(1-f_s)f_{s'}}{(\omega'-\epsilon_s+i\eta)(\omega+\omega'-\epsilon_{s'}-i\eta)} + \frac{\psi_s^*(\mathbf{r}')\psi_s(\mathbf{r})\psi_{s'}^*(\mathbf{r})\psi_{s'}(\mathbf{r}')f_s(1-f_{s'})}{(\omega'-\epsilon_s-i\eta)(\omega+\omega'-\epsilon_{s'}+i\eta)}\right]. \tag{3.38}$$

We can evaluate this integral by contour integration, closing the contour in the upper half-plane in both terms. The two residues are given by

$$\begin{aligned}\mathrm{Res}\left[\frac{1}{(\omega'-\epsilon_s+i\eta)(\omega+\omega'-\epsilon_{s'}-i\eta)}\right] &= \lim_{\omega'\to-\omega+\epsilon_s'+i\eta}\frac{\omega'+\omega-\epsilon_s'-i\eta}{(\omega'-\epsilon_s+i\eta)(\omega+\omega'-\epsilon_{s'}-i\eta)}\\ &= \frac{1}{(-\omega+\epsilon_{s'}-\epsilon_s+2i\eta)},\end{aligned} \tag{3.39}$$

and

$$\begin{aligned}\mathrm{Res}\left[\frac{1}{(\omega'-\epsilon_s-i\eta)(\omega+\omega'-\epsilon_{s'}+i\eta)}\right] &= \lim_{\omega'\to\epsilon_s+i\eta}\frac{\omega'-\epsilon_s-i\eta}{(\omega'-\epsilon_s-i\eta)(\omega+\omega'-\epsilon_{s'}+i\eta)}\\ &= \frac{1}{(\omega-\epsilon_{s'}+\epsilon_s+2i\eta)}.\end{aligned} \tag{3.40}$$

Thus (by relabelling $2\eta \to \eta'$, we find that the non-interacting DDRF is given by

$$\chi_0(\mathbf{r},\mathbf{r}',\omega) = \sum_{ss'} 2\left(\frac{\psi_s^*(\mathbf{r}')\psi_s(\mathbf{r})\psi_{s'}^*(\mathbf{r})\psi_{s'}(\mathbf{r}')f_s(1-f_{s'})}{\omega+\epsilon_s-\epsilon_{s'}+i\eta'} - \frac{\psi_s(\mathbf{r})\psi_s^*(\mathbf{r}')\psi_{s'}(\mathbf{r}')\psi_{s'}^*(\mathbf{r})(1-f_s)f_{s'}}{\omega+\epsilon_s-\epsilon_{s'}-i\eta'}\right). \tag{3.41}$$

We can further simplify this expression (including absorption of the factor of 2 into the sum and summing over both spin states), yielding the Adler-Wiser formulation [3, 195] of the (time-ordered) non-interacting DDRF

$$\chi_0(\mathbf{r},\mathbf{r}',\omega) = \sum_{s}^{occ}\sum_{s'}^{empty} \psi_s^*(\mathbf{r}')\psi_s(\mathbf{r})\psi_{s'}^*(\mathbf{r})\psi_{s'}(\mathbf{r}')\left(\frac{1}{\omega+\epsilon_s-\epsilon_{s'}+i\eta'} - \frac{1}{\omega+\epsilon_{s'}-\epsilon_s-i\eta'}\right). \quad (3.42)$$

3.2.3 The random phase approximation and electronic screening

The random phase approximation (RPA) is an approximation that introduces the concept of a screened Coulomb interaction [29]. This screened Coulomb interaction is dynamic and frequency or time-dependent and can be drawn diagrammatically in terms of the Feynman diagrams in Fig. 3.1. For a brief introduction to Feynman diagrams, please refer to Appendix A. The polarization bubbles, consisting of two Green's functions in opposite directions, correspond to the non-interacting DDRF, and the interaction lines correspond to the bare Coulomb interaction.

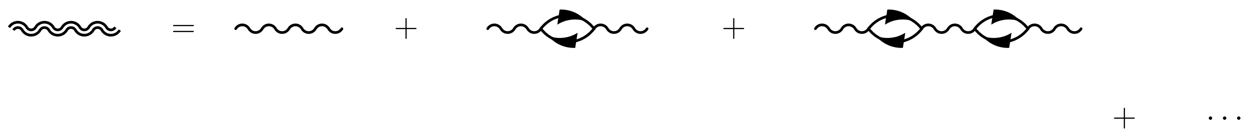


Figure 3.1: Diagrammatic expansion of the screened Coulomb interaction within the RPA approximation.

We can write the diagrammatic expansion in Fig. 3.1 as a Dyson equation by using the Feynman rules. This results in

$$\begin{aligned} W(\mathbf{r},\mathbf{r}',t,t') &= V(\mathbf{r},\mathbf{r}')\delta(t-t') + \int d\mathbf{r}_1 \int d\mathbf{r}_2 V(\mathbf{r},\mathbf{r}_1)\chi_0(\mathbf{r}_1,\mathbf{r}_2,t,t')V(\mathbf{r}_2,\mathbf{r}')+ \\ &\int d\mathbf{r}_1 \int d\mathbf{r}_2 \int d\mathbf{r}_3 \int d\mathbf{r}_4 \int dt_1 V(\mathbf{r},\mathbf{r}_1)\chi_0(\mathbf{r}_1,\mathbf{r}_2,t,t_1)V(\mathbf{r}_2,\mathbf{r}_3)\chi_0(\mathbf{r}_3,\mathbf{r}_4,t_1,t')V(\mathbf{r}_4,\mathbf{r}') + \ldots \\ &= V(\mathbf{r},\mathbf{r}')\delta(t-t') + \int d\mathbf{r}_1 \int d\mathbf{r}_2 V(\mathbf{r},\mathbf{r}_1)\chi(\mathbf{r}_1,\mathbf{r}_2,t,t')V(\mathbf{r}_2,\mathbf{r}'), \quad (3.43) \end{aligned}$$

where the screened Coulomb interaction is given by $W(\mathbf{r},\mathbf{r}',t,t')$ and I have defined the interacting

DDRF as $\chi(\mathbf{r},\mathbf{r}',t,t')$, which is given by

$$\begin{aligned}\chi(\mathbf{r},\mathbf{r}',t,t') &= \chi_0(\mathbf{r},\mathbf{r}',t,t') + \int d\mathbf{r}_1 \int d\mathbf{r}_2 \int dt_1 \chi_0(\mathbf{r},\mathbf{r}_1,t,t_1) V(\mathbf{r}_1,\mathbf{r}_2) \chi_0(\mathbf{r}_2,\mathbf{r}',t_1,t') + \ldots \\ &= \chi_0(\mathbf{r},\mathbf{r}',t,t') + \int d\mathbf{r}_1 \int d\mathbf{r}_2 \int dt_1 \chi_0(\mathbf{r},\mathbf{r}_1,t,t_1) V(\mathbf{r}_1,\mathbf{r}_2) \chi(\mathbf{r}_2,\mathbf{r}',t_1,t'). \quad (3.44)\end{aligned}$$

We can further simplify Eq. (3.43) by introducing the inverse dielectric function $\epsilon^{-1}(\mathbf{r},\mathbf{r}',t,t')$ defined as

$$\epsilon^{-1}(\mathbf{r},\mathbf{r}',t,t') = \delta(\mathbf{r}-\mathbf{r}')\delta(t-t') + \int d\mathbf{r}_1 V(\mathbf{r},\mathbf{r}_1)\chi(\mathbf{r}_1,\mathbf{r}',t,t'), \quad (3.45)$$

yielding the equation for the RPA-screened Coulomb interaction

$$W(\mathbf{r},\mathbf{r}',t,t') = \int d\mathbf{r}_1 \epsilon^{-1}(\mathbf{r},\mathbf{r}_1,t,t') V(\mathbf{r}_1,\mathbf{r}'). \quad (3.46)$$

The inverse dielectric function describes the screening of an electron's charge by the surrounding electrons' response, yielding an effective quasiparticle interaction. The repulsive interaction between electrons results in electron density depletion around each electron. The electrons with the depleted charge region surrounding them are modelled as quasiparticles within the RPA. We will see in the next section how this weak interaction is exploited.

3.2.4 The self energy

When introducing the Green's function, we have defined it in terms of the interacting ground state $|\Psi_0\rangle$. However, the interacting ground state is inaccessible for all intents and purposes due to the complexity of solving the many-particle Schödinger equation. Fortunately, a diagrammatic expansion of the many-body Green's function is made possible by the Gell-Mann and Low theorem [68]. The Gell-Mann and Low theorem connects the interacting ground state to the non-interacting ground state, which yields a diagrammatic expansion in terms of the non-interacting Green's function G_0 and the Coulomb interaction V. In terms of Feynman diagrams, the interacting Green's function (denoted with a directed double line) is given by the diagrams in Fig. 3.2. The first term in Fig. 3.2 corresponds to the non-interacting Green's function G_0, and the other two terms to the first order perturbation terms. Only the zeroth and first-order terms

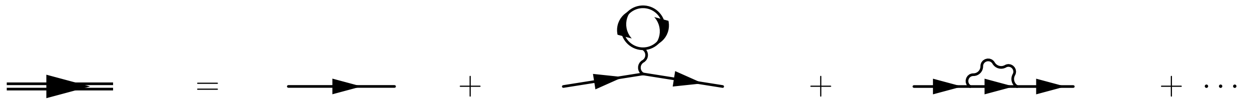


Figure 3.2: Diagrammatic expansion of the interacting Green's function in terms of the non-interacting Green's function and the Coulomb interaction.

are included explicitly. The higher order terms correspond to all possible Feynman diagrams with $2n+1$ Green's function lines and n interaction lines, where n is the perturbation order [63]. We can replace the series of non-repeating diagrams between the incoming and outgoing lines by the proper self-energy insertion Σ, resulting in Fig. 3.3. We can write down an equation for

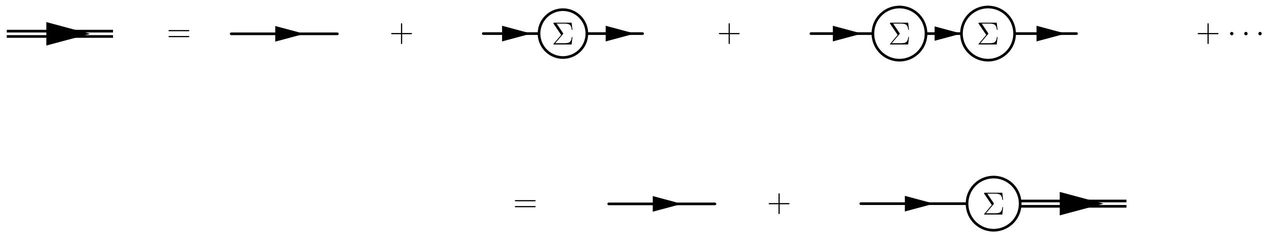


Figure 3.3: Diagrammatic expansion of the interacting Green's function in terms of the non-interacting Green's function and the proper self-energy Σ.

the diagram in Fig. 3.3, resulting in a Dyson equation for the interacting Green's function

$$G(\mathbf{r},\mathbf{r}',t,t') = G_0(\mathbf{r},\mathbf{r}',t,t') + \int d\mathbf{r}_1 dt_1 \int d\mathbf{r}_2 dt_2 G_0(\mathbf{r},\mathbf{r}_1,t,t_1)\Sigma(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2)G(\mathbf{r}_2,\mathbf{r}',t_2,t'). \tag{3.47}$$

By introducing a basis set (as in Eq. (3.35)), the above equation can be turned into an algebraic equation, which allows us to define the interacting Green's function as

$$G = \frac{1}{G_0^{-1} - \Sigma}. \tag{3.48}$$

The non-interacting Green's function is diagonal in the basis of single-particle eigenstates. Assuming that the self-energy is also diagonal, we can get a simple approximation for the poles of the Green's function as

$$0 = G_{0ss}^{-1} - \Sigma_{ss} = \omega - \epsilon_s - \Sigma_{ss}. \tag{3.49}$$

The above equation highlights the significance of the self-energy: it is equivalent to a perturbative correction to the energy levels of the non-interacting system and incorporates all many-body interactions. Thus, it plays a similar role to the exchange-correlation functional in DFT. The self-energy is, however, again given by an infinite series of Feynman diagrams, the first few of which are shown in Fig. 3.4.

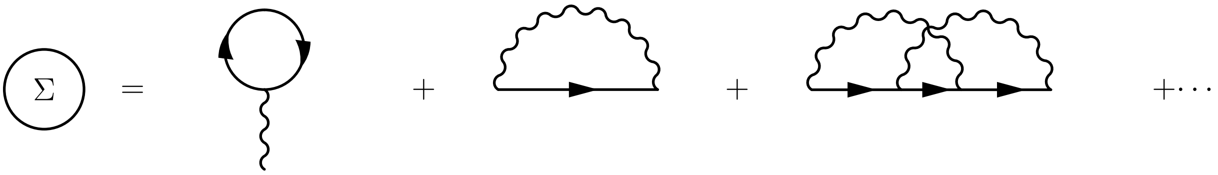


Figure 3.4: Diagrammatic expansion of the self energy Σ.

To make progress, one has to approximate the self-energy by only retaining certain diagrams. One potential approximation is only to retain the first two terms. Evaluating the first term results in

$$-i\int d\mathbf{r}' V(\mathbf{r},\mathbf{r}')G_0(\mathbf{r}',\mathbf{r}',t,t^+) = \int d\mathbf{r}' V(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}'),$$

where $\rho(\mathbf{r}')$ is the electron density. The second term is given by

$$iV(\mathbf{r},\mathbf{r}')G_0(\mathbf{r},\mathbf{r}',t,t')\delta(t-t'), \tag{3.50}$$

The first term corresponds to the Hartree potential, and the second term to the Hartree-Fock exchange potential. One can further replace the non-interacting Green's function with the interacting Green's function, as shown in Fig 3.5. However, it can be shown [63] that this corresponds precisely to the self-consistent Hartree-Fock method, which is known to overestimate properties like the band gap [139] significantly. Going beyond the first order terms can be problematic. In fact, the series in Fig. 3.4 may not have a limit at all [61] and some diagrams,

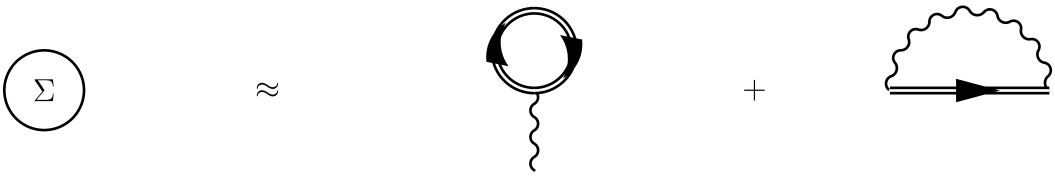


Figure 3.5: Approximate self-consistent expansion of the self-energy Σ.

in particular, the ones containing repetitions of polarization bubbles (e.g. the ones in Fig. 3.1) can be unbounded [61] due to the long range of the Coulomb interaction. One solution to these problems is to select the problematic diagrams containing polarization bubbles and sum them to infinite order. It turns out that this is equivalent to replacing the bare Coulomb interaction with the screened Coulomb interaction in Fig. 3.1. The screened Coulomb interaction W is much shorter ranged than the bare Coulomb interaction V [63], and a diagrammatic expansion in terms of W instead of V is thus more well-behaved and should converge rapidly. Within the RPA, the self-consistent GW method corresponds to a truncation after the first order in W and is thus equivalent to replacing the bare Coulomb interaction in the exchange term with the dynamically screened Coulomb interaction. The distinction between the classical RPA method and the self-consistent GW method is that the DDRF is recalculated at each iteration step, thus replacing the non-interacting Green's function lines of the polarization insertion with interacting Green's function lines. The self-consistent GW method is thus equivalent to the Hartree-Fock method with a dynamically screened exchange energy. Graphically the resulting self-energy is therefore given by the diagram in Fig. 3.6 (excluding the Hartree potential). However, when

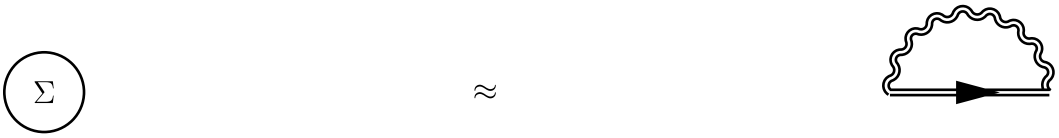


Figure 3.6: Self-energy within the self-consistent GW method Σ.

using this self-consistent approach, introducing a suitable vertex correction is crucial [139]. The vertex correction effectively accounts for higher-order contributions to the self-energy, such as the third term in Fig. 3.4. However, including vertex corrections is far from trivial [142] and often does not outperform non-selfconsistent approaches [139]. In order to avoid these complexities, in this thesis, I will focus on the so-called one-shot $\mathrm{G_0W_0}$ method. Within this approximation, the self energy Σ is given by [119]

$$\Sigma(\mathbf{r}, \mathbf{r}', t, t') = iG_0(\mathbf{r}, \mathbf{r}', t, t')W_0(\mathbf{r}, \mathbf{r}', t, t'), \tag{3.51}$$

with G_0 being an initial guess for the Green's function obtained from a suitable reference calculation (typically Kohn-Sham DFT or Hartree-Fock) and W_0 being evaluated using the interacting DDRF as defined in Eq. (3.44). Once the self-energy is computed, a first-order correction to the Kohn-Sham/Hartree-Fock (denoted "MF") energies is computed, obtaining quasiparticle energies [173]

$$\epsilon_s^{QP} = \epsilon_s^{MF} + (1 - \langle\psi_s^{MF}|\frac{\partial\Sigma}{\partial\omega}|_{\epsilon_s^{MF}}|\psi_s^{MF}\rangle)^{-1}\langle\psi_s^{MF}|\Sigma(\mathbf{r}, \mathbf{r}', \omega = \epsilon_s^{MF}) - V_{xc}(\mathbf{r})|\psi^{MF}\rangle, \tag{3.52}$$

where it was further assumed that the off-diagonal elements of Σ are vanishingly small [173].

3.3 GW in practice

3.3.1 Extension of the static dielectric matrix to finite frequencies

So far, we have assumed that the dielectric function is computed at all frequencies. In practice, this is challenging, as it often has to be evaluated at many discrete frequencies. Therefore the generalized plasmon-pole model [90] (GPP) is often employed. Within the GPP, the frequency dependence of the dielectric matrix is described by collective modes called plasmons [43]. Specifically, the GPP relies on the observation that ϵ^{-1} is dominated by energies close to the plasmon frequency [173]. As the BerkeleyGW [90, 49] code employed in this thesis uses a plane wave basis, it is convenient to express all quantities in terms of plane waves. In this basis, the dielectric matrix is written as [173]

$$\begin{aligned}\epsilon_{\mathbf{GG}'}(\mathbf{q},\omega) = \delta_{\mathbf{G},\mathbf{G}'} - V(\mathbf{q}+\mathbf{G}) \sum_{s}^{occ} \sum_{s'}^{empty} \sum_{\mathbf{k}} M_{s,s'}^{\mathbf{G}}(\mathbf{k},\mathbf{q}) M_{s,s'}^{*\mathbf{G}'}(\mathbf{k},\mathbf{q}) \\ \times \left\{ \frac{1}{E_{s,\mathbf{k}-\mathbf{q}} - E_{s,\mathbf{k}} - \omega + i\eta} + \frac{1}{E_{s,\mathbf{k}-\mathbf{q}} - E_{s,\mathbf{k}} + \omega - i\eta} \right\}, \quad (3.53)\end{aligned}$$

where $\mathbf{G}, \mathbf{G}'$ are reciprocal lattice vectors and $M_{s,s'}^{\mathbf{G}}(\mathbf{k},\mathbf{q}) = \langle s, \mathbf{k} | e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} | s', \mathbf{k}-\mathbf{q} \rangle$. Within the GPP, the imaginary part of the inverse dielectric function is given by [90]

$$\mathrm{Im}(\epsilon_{\mathbf{GG}'}^{-1})(\mathbf{q},\omega) = -\frac{\pi}{2} \frac{\Omega_{\mathbf{GG}'}^2}{\tilde{\omega}_{\mathbf{GG}'}} \left[\delta(\omega - \tilde{\omega}_{\mathbf{GG}'}) - \delta(\omega + \tilde{\omega}_{\mathbf{GG}'}) \right]. \quad (3.54)$$

The real part can be obtained using the Kramers-Kronig relations, yielding [90]

$$\begin{aligned}\mathrm{Re}(\epsilon_{\mathbf{GG}'}^{-1})(\mathbf{q},\omega) &= \delta_{\mathbf{GG}'} - \mathtt{P} \int_0^\infty \frac{\Omega_{\mathbf{GG}'}^2}{\tilde{\omega}_{\mathbf{GG}'}} \left[\delta(\omega' - \tilde{\omega}_{\mathbf{GG}'}) - \delta(\omega' + \tilde{\omega}_{\mathbf{GG}'}) \right] \frac{\omega'}{\omega'^2 - \omega^2} d\omega' \\ &= \delta_{\mathbf{GG}'} + \frac{\Omega_{\mathbf{GG}'}^2}{\omega^2 - \omega_{\mathbf{GG}'}^2}. \quad (3.55)\end{aligned}$$

The coefficients $\tilde{\omega}_{\mathbf{GG}'}$ are obtained using the the Kramers-Kronig relation evaluated at the static limit of the dielectric function ($\omega = 0$), which is explicitly calculated, yielding [177]

$$\tilde{\omega}_{\mathbf{GG}'}^2 = \frac{\Omega_{\mathbf{GG}'}^2}{\delta_{\mathbf{GG}'} - \epsilon_{\mathbf{GG}'}^{-1}(\mathbf{q}, \omega = 0)}. \quad (3.56)$$

The coefficients $\Omega_{\mathbf{GG}'}^2$ are calculated by using the generalized f-sum rule [91]

$$\int_0^\infty \omega \, \mathrm{Im}(\epsilon_{\mathbf{GG}'}^{-1})(\mathbf{q},\omega) = -\frac{\pi}{2} \frac{\omega_p(\mathbf{q}+\mathbf{G}) \cdot (\mathbf{q}+\mathbf{G}')\rho(\mathbf{G}-\mathbf{G}')}{|\mathbf{q}+\mathbf{G}|^2 \rho(\mathbf{0})}, \quad (3.57)$$

where ω_p is the plasma frequency and $\rho(\mathbf{G})$ is the reciprocal space representation of the electron density [173]. Thus, the coefficients evaluate to

$$\Omega_{\mathbf{GG}'}^2 = \frac{\omega_p(\mathbf{q}+\mathbf{G}) \cdot (\mathbf{q}+\mathbf{G}')\rho(\mathbf{G}-\mathbf{G}')}{|\mathbf{q}+\mathbf{G}|^2 \rho(\mathbf{0})}. \quad (3.58)$$

Using the GPP, a full frequency calculation can often be avoided while obtaining excellent agreement with experimental excitation energies [119].

3.3.2 Coulomb truncation

For finite systems, truncation of the Coulomb interaction can significantly improve the speed of convergence with respect to the unit cell size. The Coulomb truncation replaces the Coulomb potential with a truncated version

$$V_t(\mathbf{r}-\mathbf{r}') = \frac{\theta(f(|\mathbf{r}-\mathbf{r}'|))}{|\mathbf{r}-\mathbf{r}'|}, \tag{3.59}$$

where the function f defines the geometry of the truncation [49]. Within the cell-box truncation scheme used in this thesis, the Coulomb interaction is truncated at the boundaries of a supercell-shaped region around each point. The cell-box truncation is achieved through a numerical truncation; however, to illustrate the behaviour, it is instructive to look at the definition of the spherical truncation, which has an analytic expression [93]

$$V_t(\mathbf{r}-\mathbf{r}') = \frac{\theta(r_c-|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}, \tag{3.60}$$

where r_c is a cut-off radius. The cell-box (or supercell) truncation behaves analogously and is visualized in Fig. 3.7: the black circles indicate a spherical cluster in the centre of the supercell. Now suppose the truncated Coulomb interaction is evaluated at point $\mathbf{r}$ indicated in Fig. 3.7. The red box centred at $\mathbf{r}$ has the same size as the supercell and indicates the region where the truncated Coulomb interaction is non-zero: For every point $\mathbf{r}'$ outside the red box, the Coulomb interaction is truncated. This truncation scheme helps to limit interactions between periodic

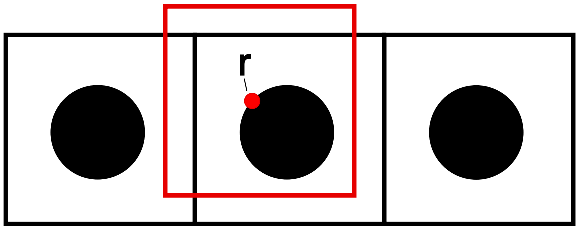


Figure 3.7: 2D visualization of the cell-box truncation scheme.

images and thus leads to faster convergence with respect to supercell size [49]. To fully avoid interactions with periodic images, the supercell has to be large enough, such that for every point $\mathbf{r}$ inside the cluster, the red truncation box does not intersect with the clusters in the neighbouring supercells. While spherical truncation can be used as well, supercell truncation is recommended for the treatment of 0D systems within BerkeleyGW [49].

Chapter 4

Predicting polarizabilities of silicon clusters using local chemical environments

4.1 Introduction

Clusters and nanoparticles are used in a variety of scientific and industrial applications, including optoelectronics [189], photocatalysis [44], medical imaging [141, 138] or single electron transistors [65]. Electronic excitations often play a key role in these applications, but theoretical techniques for calculating excited-state properties of materials, such as the first-principles GW/Bethe-Salpeter method, are typically limited to very small systems. As discussed in the introduction, the main bottleneck of the GW method is the calculation of the interacting DDRF.

As an efficient alternative to first-principles techniques, machine learning (ML) based techniques have been explored in recent years. For example, ML has efficiently been applied to potential energy surfaces [16, 77, 116, 59] or to predict electronic ground state densities [31, 5, 72, 38]. Additionally, projects such as the Materials Project [96] and the Open Quantum Materials database [101, 161] have made an effort to make first-principles data of a wide range of materials publicly available. As discussed in Chapter 2, descriptors or fingerprints encoding the structure of a molecule or cluster are crucial ingredients in the application of ML in materials physics.

For example, the smooth-overlap of atomic positions (SOAP) kernel [15] has been widely used as a similarity measure between different chemical environments. Recently, several groups have also started to explore the applicability of ML approaches to calculate molecular polarizabilities and dipole moments [192, 73, 181, 184]. For example, Grisafi et al. [73] introduced a symmetry-adapted variant of the SOAP descriptor [15] to predict polarizability tensors of molecules. Similarly, Wilkins and coworkers [192] used the symmetry-adapted SOAP kernel to predict polarizabilities and first hyperpolarizabilities of small organic molecules with high accuracy. Recently, Ceriotti et al. [184] used a combination of the symmetry-adapted SOAP kernel and the scalar SOAP kernel to predict dipole moments of small molecules with close to DFT accuracy. However, the applicability of machine learning models to the prediction of non-local response functions such as the DDRF remains largely unexplored.

Before approaching the development of a model of the DDRF, this chapter focuses on a simpler problem: the RPA scalar polarizability, which is closely related to the DDRF, and thus may provide valuable insights into the ability to predict response functions of clusters based on purely structural descriptors. In order to test this approach, we create a dataset of RPA scalar polarizabilities of hydrogenated silicon clusters. These systems are well suited for this purpose because their polarizabilities have been studied in detail with a variety of modelling techniques. For example, simple empirical models, such as bond polarizability models, have been used to predict Raman spectra of silicon clusters in good agreement with experiment [146]. Empirical models can be extended beyond the assumption of additivity of atomic polarizabilities. A class of models that captures interactions between polarization centres are dipole interaction models [188], which have been successfully applied to the construction of polarizable force fields [190]. Highly accurate cluster polarizabilities can be obtained using ab initio approaches such as density functional theory (DFT) [183, 47, 18, 95, 94, 124], Møller-Plesset perturbation theory [124], coupled-cluster theory [123, 140, 124] or the random phase approximation (RPA) [63, 97, 129]. For example, Mochizuki and Agren [129] used the RPA and the second-order polarization propagator approximation to calculate the polarizabilities of spherical hydrogenated silicon clusters with up to 35 Si atoms and found that the polarizability per silicon atom approaches the bulk limit from below. In contrast, for unhydrogenated silicon clusters, Jackson and coworkers found that the bulk value is approached from above as the size of the cluster increases [95, 94]. This behaviour was attributed to the presence of dangling bonds on the surface. Furthermore, it was observed that the polarizability depends sensitively on the shape of the cluster [97, 94]. Jansik et al. [97] compared polarizabilities of three-dimensional (3D), two-dimensional (2D) and one-dimensional (1D) hydrogenated silicon structures and found that the presence of π-bonds in 2D systems leads to a much stronger increase in the polarizability as a function of cluster size when compared to 1D and 3D clusters [97]. A similar trend was observed when comparing prolate and compact clusters, with prolate structures showing a significantly larger polarizability per silicon atom than compact structures [94].

Here, we explore the ability of machine learning models based on the SOAP [15] descriptor to describe and predict static polarizabilities of hydrogenated silicon clusters calculated from RPA DDRFs. Despite the incompleteness of the SOAP descriptor discussed in Chapter 2, the SOAP descriptor has proven to reliably represent chemical environments while being invariant to rigid translations and rotations. Previous work [192, 73] has already demonstrated the ability to predict isotropic scalar polarizabilities and also the full polarizability tensor using SOAP and generalizations thereof. The symmetry-adapted SOAP descriptor has also been used successfully in conjunction with physical insights [184] to predict molecular dipole moments. As SOAP is a generic 3-body descriptor of the neighbourhood density [133], many simpler descriptors such as RDFs and ADFs [87, 39, 154] (which are particular projections of the neighbour density) can be represented in the SOAP basis, and in the limit of no basis truncation, SOAP is equivalent to other 3-body descriptors such as Behler-Parrinello Atom Centered Symmetry Functions [22] and the FCHL [58] descriptors. To generate a data set, we start by calculating scalar isotropic polarizabilities of a set of clusters containing between 10 and 110 silicon atoms using the RPA DDRF. Next, we investigate the ability of the ML approach to reproduce the calculated polarizabilities and find that almost perfect agreement can be obtained when the size of the local chemical environments is sufficiently large to contain the whole cluster. Importantly, the ML models already describe the qualitative behaviour of the average polarizability per atom if the local environment only contains nearest neighbour atoms. These findings establish the fittability of RPA scalar polarizabilities using local SOAP descriptors, which has not been explored to date in contrast to mean-field DFT data. Next, we study the ability of ML to predict the polarizabilities of clusters not included in the training set. Interestingly, the predictive power of ML is strongest when the size of the chemical environment is relatively small. These insights enable the reliable prediction of polarizabilities of large clusters, which are difficult to calculate with standard first-principles techniques and constitute a first step towards efficient ML approaches for excited-state properties of materials, such as the DDRF which will be explored in Chapters 5 and 6.

4.2 Methods

4.2.1 Random Phase Approximation polarizabilities

Scalar polarizabilities of molecules and clusters were calculated within the RPA in a linear response framework. The RPA was chosen because it is known to give an accurate description of the dielectric properties of bulk silicon [91]. The polarizability tensor α_{ij} relates the induced dipole moment with Cartesian components μ_i to the applied static electric field E_j according to

$$\mu_i = \sum_j \alpha_{ij} E_j. \tag{4.1}$$

To obtain an expression for α_{ij}, we express μ_i in terms of the induced electronic charge density $\Delta\rho(\mathbf{r})$ via

$$\mu_i = -e \int d\mathbf{r} \Delta\rho(\mathbf{r}) r_i, \tag{4.2}$$

where e denotes the proton charge, and r_i is the Cartesian component of the position vector. The induced charge density is determined by the interacting density-density response function $\chi(\mathbf{r}, \mathbf{r}')$ according to

$$\Delta\rho(\mathbf{r}) = e \sum_j E_j \int d\mathbf{r}' \chi(\mathbf{r}, \mathbf{r}') r'_j, \tag{4.3}$$

where the potential associated with the applied electric field $V(\mathbf{r}) = e\sum_j E_j r_j$ was used. Combining these equations yields

$$\alpha_{ij} = -e^2 \int d\mathbf{r}d\mathbf{r}'\chi(\mathbf{r},\mathbf{r}')r_i r_j'. \tag{4.4}$$

The scalar polarizability α is obtained by dividing the trace of α_{ij} by three, which shows the close relationship between the scalar polarizabilitiy and the DDRF. To evaluate Eq. (4.4), the interacting density-density response function must be determined. As described in Chapter 3, the RPA χ obeys the Dyson equation

$$\chi(\mathbf{r},\mathbf{r}') = \chi_0(\mathbf{r},\mathbf{r}') + \int d\mathbf{r}_1 d\mathbf{r}_2 \chi_0(\mathbf{r},\mathbf{r}_1)V(\mathbf{r}_1,\mathbf{r}_2)\chi(\mathbf{r}_2,\mathbf{r}'), \tag{4.5}$$

where $V(\mathbf{r}_1,\mathbf{r}_2)$ denotes the Coulomb interaction and χ_0 is the non-interacting density-density response function within the Adler-Wiser formulation [3, 195] in Eq. (3.42). To numerically calculate scalar polarizabilities, the plane-wave/pseudopotential approach is employed. Specifically, the BerkeleyGW program [90, 49] is used to calculate $\chi_{\mathbf{G}\mathbf{G}'}$ where $\mathbf{G}$ and $\mathbf{G}'$ denote reciprocal lattice vectors of the periodically repeated supercell. Note that interactions between images are avoided by using a truncated Coulomb interaction. The interacting density-density response function in real space is then given by

$$\chi(\mathbf{r},\mathbf{r}') = \frac{1}{V}\sum_{\mathbf{G},\mathbf{G}'} e^{i\mathbf{G}\cdot\mathbf{r}}\chi_{\mathbf{G},\mathbf{G}'}e^{-i\mathbf{G}'\cdot\mathbf{r}'}, \tag{4.6}$$

where $V = L^3$ denotes the volume of the cubic supercell, with L being the side length. Finally, the scalar polarizability is found to be

$$\alpha = \frac{e^2}{3V}\sum_i \sum_{\mathbf{G},\mathbf{G}'} \chi_{\mathbf{G},\mathbf{G}'}\Delta_{\mathbf{G},i}\Delta^*_{\mathbf{G}',i}, \tag{4.7}$$

with

$$\Delta_{\mathbf{G},x} = \begin{cases} \frac{L^4}{2}\delta_{G_x,0}\delta_{G_y,0}\delta_{G_z,0} & \text{if } G_x = 0, \\ \frac{L^3}{iG_x}\delta_{G_y,0}\delta_{G_z,0} & \text{otherwise,} \end{cases} \tag{4.8}$$

and similar expressions for $\Delta_{\mathbf{G},y}$ and $\Delta_{\mathbf{G},z}$.

It should be noted that other, more efficient approaches than the one described above exist for the calculation of the static scalar polarizability, such as the finite field method [34]. However, the ultimate interest of this work is in applying ML techniques to accelerate excited-state calculations, and these methods require the full interacting density-density response function, which cannot easily be obtained with other methods.

4.2.2 Kernel ridge regression

In order to model the polarizabilities obtained using the method described above, we use kernel ridge regression (KRR). The relevant expressions can be derived starting from ordinary linear regression, where the target quantity y_i is modelled as a linear combination of individual inputs in the feature vector $\mathbf{x}_i$

$$y_i \approx f(\mathbf{x}_i) = \mathbf{x}_i^T \mathbf{w}, \tag{4.9}$$

where $\mathbf{w}$ is a learnable vector of weight parameters. First, the input vector is projected into a (usually higher dimensional) feature space through a function $\phi(\mathbf{x}_i)$ [152]. An example would be the projection of a scalar x into the space of powers $\phi(x) = (1, x, x^2, \dots)$ [152]. After applying the projections, the linear model takes the form of

$$f(\mathbf{x}_i) = \phi(\mathbf{x}_i)^T \mathbf{w}. \tag{4.10}$$

Next, the weight vector is expanded as a linear combination of feature vectors contained in the training set $\mathbf{w} = \sum_j^{N_\text{train}} \zeta_j \phi(\mathbf{x}_j)$ [132], yielding

$$f(\mathbf{x}_i) = \sum_j^{N_\text{train}} \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \zeta_j. \tag{4.11}$$

The inner product between feature vectors corresponds to elements of the symmetric kernel matrix K_{ij}, which allows us to rewrite Eq. (4.11)

$$f(\mathbf{x}_i) = \sum_j^{N_\text{train}} K_{ij} \zeta_j. \tag{4.12}$$

Thus, elements of the kernel matrix effectively measure the similarity between different feature vectors. When training a KRR model, we want to minimize the squared difference between the target quantity y and our model function $f(\mathbf{x})$

$$\epsilon = (\mathbf{y} - \mathbf{K}\zeta)^T(\mathbf{y} - \mathbf{K}\zeta) + \lambda\zeta^T\mathbf{K}\zeta, \tag{4.13}$$

where the kernel matrix is computed between all feature vectors in the training set, and the vector $\mathbf{y}$ is the vector containing the corresponding target quantities. The second term in Eq. (4.13) penalizes for larger magnitudes in the weight vector, and λ is a positive parameter determining the strength of the penalty. λ is usually referred to as a regularization parameter [132]. Taking derivatives with respect to ζ and equating to zero, we obtain a solution for the coefficients ζ in terms of the kernel matrix $\mathbf{K}$ and the vector of target quantities $\mathbf{y}$

$$\zeta = (\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{y}. \tag{4.14}$$

If the training set is large, Eq. (4.14) can become prohibitively expensive to solve. One possible solution is the subset of regressors approach [152] in which a subset of M representative points is selected from the training set. In this case, the use of an alternative equation for obtaining the coefficients ζ is required [36]

$$\zeta = \left[\mathbf{K}_{MM} + \mathbf{K}_{MN}\mathbf{\Lambda}^{-1}\mathbf{K}_{NM}\right]^{-1}\mathbf{K}_{MN}\mathbf{\Lambda}^{-1}\mathbf{y}, \tag{4.15}$$

where K_{MM} is the kernel matrix computed between representative points and $\mathbf{K}_{NM}$ is the kernel matrix between the N training points and the M representative points. $\Lambda = \lambda\mathbf{I}$ is again a regularization term. Equation (4.15) is the result of minimizing [36]

$$\epsilon = (\mathbf{y} - \mathbf{K}_{NM}\zeta)^T\mathbf{\Lambda}^{-1}(\mathbf{y} - \mathbf{K}_{NM}\zeta) + \zeta^T\mathbf{K}_{MM}\zeta. \tag{4.16}$$

4.2.3 Environment descriptors

As the kernel matrix measures the similarity between features, the ability to assess the similarity of different chemical environments plays a key role in machine learning of material properties. Here, we use the SOAP approach [15] where the environment of atom i is described by the set of neighbourhood densities

$$\rho_i^\nu(\mathbf{r}) = \sum_{i=j}^{N_\nu} e^{-\gamma_\nu(\mathbf{r}-\mathbf{r}_{ij})^2}, \tag{4.17}$$

where ν denotes a specific element that is present in the atom's environment with N_ν being the number of such atoms up to a given cut-off radius r_{cut}. In addition, γ_ν is a hyperparameter describing the size of the neighbour atom.

The similarity of two chemical environments described by the neighbourhood densities $\rho_i = \{\rho_i^\nu\}_\nu$ and $\rho_j = \{{\rho_j}^\nu\}_\nu$ can be measured by the kernel [36]

$$k(\rho_i, \rho_j) = \int d\hat{R} \left| \sum_\nu \int d\mathbf{r} \rho_i^\nu(\mathbf{r}) \rho_j^\nu(\hat{R}\mathbf{r}) \right|^2, \tag{4.18}$$

where $\hat{R}$ denotes a rotation matrix. Note that in Eq. (4.18), the sum is inside the norm in contrast to the simpler SOAP kernel introduced in Chapter 2. In this version of the SOAP kernel, correlations between the environments corresponding to different species ν and ν' are included in the descriptor. To evaluate the kernel integral, the angular dependence of the neighbourhood densities is expanded in a basis of spherical harmonics Y_{lm} and the radial part in a set of orthogonal radial basis functions $g_n(r)$ according to

$$\rho_i^\nu(\mathbf{r}) = \sum_{nlm} c_{i,nlm}^\nu g_n(r) Y_{lm}(\hat{\mathbf{r}}), \tag{4.19}$$

where $c_{i,nlm}^\nu$ is an expansion coefficient. Here, l ranges from zero to a cut-off value $l_{\max}$ and m ranges from $-l$ to l and the radial index n has a cut-off of $n_{\max}$.

The similarity kernel Eq. (4.18) has the appealing property that the integrals can be carried out analytically yielding [15]

$$k(\rho_i, \rho_j) = \sum_{\nu \leq \nu'} \sum_{nn'l} d^{\nu,\nu'}_{i,nn'l} d^{\nu,\nu'}_{j,nn'l} \tag{4.20}$$

$$d^{\nu,\nu'}_{i,nn'l} = \sum_m c^{\nu}_{i,nlm} (c^{\nu'}_{i,n'lm})^*. \tag{4.21}$$

From the above expressions, it can be seen that the set of coefficients $\{d^{\nu,\nu'}_{i,nn'l}\}$ with $\nu \leq \nu'$ plays the role of a descriptor vector $\mathbf{d}_i$ for the environment of atom i. In practice, the descriptor vectors are calculated using the `QUIPPY` software package [15]. The kernel matrix $k(\rho_i, \rho_j)$ is then calculated according to

$$k(\rho_i, \rho_j) = \mathbf{d}_i \cdot \mathbf{d}_j. \tag{4.22}$$

The sensitivity of the kernel to differences between atomic environments can be increased by defining the effective SOAP kernel [15, 36]

$$K(\rho_i, \rho_j) = \left(\frac{k(\rho_i, \rho_j)}{\sqrt{k(\rho_i, \rho_i) k(\rho_j, \rho_j)}} \right)^{\epsilon}. \tag{4.23}$$

In this work, we use $\epsilon = 2$.

4.2.4 Learning cluster polarizabilities

The SOAP descriptor allows the comparison of different environments of atoms in a cluster. However, the polarizability is calculated for an entire molecule or cluster consisting of many atoms. To harness the SOAP approach for the prediction of cluster polarizabilities, it is, therefore, necessary to relate atomic properties to cluster properties. One way to achieve this is by expressing the polarizability α_I of cluster I as the sum of atomic contributions α_i [36] according to

$$\alpha_I = \sum_{i=1}^{N_I} \alpha_i, \tag{4.24}$$

where N_I denotes the total number of atoms in the cluster. While the atomic contributions can provide some valuable intuition about the dielectric response of complex clusters, it is important to stress that these quantities are not directly measurable and should be interpreted with care [14].

Using standard kernel ridge regression, the atomic polarizabilities can be expressed as

$$\alpha_i = \sum_j^{N_{\mathrm{train}}} K_{ij}\zeta_j, \tag{4.25}$$

where N_{train} denotes the total number of atoms in the training set (i.e. the total number of atoms contained in all training set clusters), ζ_j is a coefficient obtained from training the SOAP model, and $K_{ij} \equiv K(\rho_i, \rho_j)$. Inserting Eq. (4.25) into Eq. (4.24) yields

$$\alpha_I = \sum_i^{N_I} \sum_j^{N_{\mathrm{train}}} K_{ij}\zeta_j = \sum_j^{N_{\mathrm{train}}} K_{I,j}^{\mathrm{sum}}\zeta_j, \tag{4.26}$$

where we defined the sum kernel $K_{I,j}^{\mathrm{sum}} = \sum_i^{N_I} K_{ij}$.

Determining the coefficients ζ_j is difficult as the fit to the calculated cluster polarizabilities is strongly underdetermined (as the number of coefficients is the total number of atoms of all clusters in the training set). To make progress, the number of coefficients must be reduced. Intuitively, this should be possible as the atomic environments of many atoms in the training set are very similar. While it is also possible to select a subset of data points from the training set [152], we achieve this sparsification by means of a singular value decomposition (SVD) of the descriptor matrix $\mathbf{D}$ whose rows contain the descriptor vectors from Eq. (4.21). Specifically, $\mathbf{D}$ is expressed as

$$\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{4.27}$$

where $\mathbf{U}$ and $\mathbf{V}^T$ contain the right and left singular vectors, respectively, and $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values. If many environments in $\mathbf{D}$ are similar, only a few singular values will have large magnitudes. Only those singular values which are larger than a given threshold are retained, and the corresponding left singular vectors (which form a matrix $\tilde{\mathbf{V}}$) are used as a new basis to represent $\mathbf{D}$.

The elements of the SOAP kernel $\tilde{\mathbf{K}}$ corresponding to this new set of effective descriptors are obtained by projecting the descriptors $\mathbf{d}_i$ onto the rows $\tilde{\mathbf{v}}_j$ of the truncated matrix of singular vectors $\tilde{\mathbf{V}}$ according to

$$\tilde{K}_{ij} = \mathbf{d}_i \cdot \tilde{\mathbf{v}}_j. \tag{4.28}$$

Next, the effective sum kernel $\tilde{\mathbf{K}}^{\text{sum}}$ can be calculated using Eq. (4.26), but now the number of coefficients ζ_i is equal to the number of singular vectors whose singular values exceed the threshold. Finally, the vector of coefficients ζ is obtained from Eq. (4.15)

$$\zeta = \left[\tilde{\mathbf{V}}^T\tilde{\mathbf{V}} + (\tilde{\mathbf{K}}^{\text{sum}})^T\mathbf{\Lambda}^{-1}\tilde{\mathbf{K}}^{\text{sum}}\right]^{-1}(\tilde{\mathbf{K}}^{\text{sum}})^T\mathbf{\Lambda}^{-1}\boldsymbol{\alpha}, \tag{4.29}$$

where $\mathbf{\Lambda} = \lambda\mathbf{I}$ with λ being a regularization parameter and $\boldsymbol{\alpha}$ denotes the vector of calculated cluster polarizabilities.

Alternatively, the cluster polarizability can be expressed as the number of silicon atoms multiplied by their average polarizability α^{av} (note that in this definition, α^{av} also contains the smaller contribution from the hydrogen atoms)

$$\alpha = N_{Si}\alpha^{\text{av}}_{Si}. \tag{4.30}$$

To calculate the average polarizability, the SOAP kernel matrix is averaged over environments belonging to pairs of clusters [14]

$$K^{\text{av}}_{IJ} = \frac{1}{N_I N_J}\sum_i^{N_I}\sum_j^{N_J} K_{ij}. \tag{4.31}$$

Using kernel ridge regression, the average polarizability of the silicon atoms in a given cluster is expressed as

$$\alpha^{\text{av}}_{Si} = \sum_J^{n_{\text{train}}} K^{\text{av}}_J \zeta_J, \tag{4.32}$$

where n_{train} denotes the number of clusters in the training set, and the vector of coefficients ζ_J

is determined by

$$\zeta = \left(\mathbf{K}^{\mathrm{av}} + \boldsymbol{\Lambda}\right)^{-1} \boldsymbol{\alpha}^{\mathrm{av}}, \tag{4.33}$$

where $\boldsymbol{\alpha}^{\mathrm{av}}$ is the vector containing the average polarizabilities per atom of the training set clusters. As a consequence of the averaging, no additional sparsification procedure is required, as in the case of the sum kernel.

This method has the advantage that the average polarizability can be written as a sum of atomic contributions, which allows one to assign polarizabilities to individual atoms. This can be achieved by omitting the average over the index i in Eq. 4.31, which yields a prediction for each silicon atom in a cluster

It is interesting to note that the polarizability obtained from the average kernel can also be expressed as a sum of atomic contributions given by

$$\alpha_i = \frac{1}{N_J} \sum_J^{n_{\mathrm{train}}} \sum_j^{N_J} K_{ij} \zeta_J. \tag{4.34}$$

Apart from the scaling factor $1/N_J$, the last equation is very similar to Eq. (4.25) of the sum kernel approach, with the additional constraint that the coefficients ζ_j on atoms in a cluster J are all equal, $\zeta_j = \zeta_J \ \forall j \in J$. The effect of the scaling factor is that while the sum kernel is extensive (its magnitude scales with the number of atoms in the cluster), the average kernel is intensive, independent of system size. Consequently, large clusters get more heavily weighted in the solution of the least squares problem, Eq. (4.29), compared with that for the average kernel.

Finally, we also use the "coherent average" kernel (denoted "coh"), which is obtained as follows. Rather than computing a SOAP descriptor for each atomic environment, as in Eq. (4.20), we take the spherical harmonic coefficients c^{ν}_{nlm} and average them first to obtain, for cluster I

$$\bar{c}^{\nu}_{I,nlm} = \frac{1}{N_I} \sum_{i=1}^{N_I} c^{\nu}_{i,nlm}, \tag{4.35}$$

and then square these to form the averaged descriptor vector $\bar{\mathbf{d}}_I$ with components,

$$\bar{d}^{\nu,\nu'}_{I,nn'l} = \sum_m \bar{c}^{\nu}_{I,nlm} (\bar{c}^{\nu'}_{I,n'lm})^*, \tag{4.36}$$

and the coherent (unnormalized) kernel between clusters I and J as

$$k^{\mathrm{coh}}(I, J) = \bar{\mathbf{d}}_I \cdot \bar{\mathbf{d}}_J. \tag{4.37}$$

4.2.5 Physical models

Apart from the SOAP models, two simpler, physical-based models are used to fit the calculated RPA polarizabilities. In the first approach, the cluster polarizability is assumed to be proportional to the number of silicon atoms N_{Si} in the cluster, i.e.

$$\alpha = \alpha_{Si}^{\mathrm{av}} N_{Si}, \tag{4.38}$$

with α_{Si}^{av} denoting the average polarizability per silicon atom (again, any contributions from hydrogen atoms are included in α_{Si}^{av} in this definition). In contrast to the SOAP fitting with the average kernel, the average polarizability is assumed to be the same for all clusters.

The second model is a bond polarizability approach where the cluster polarizability is expressed as a sum of contributions from Si-Si bonds and Si-H bonds according to

$$\alpha = \alpha_{Si-Si} N_{Si-Si} + \alpha_{Si-H} N_{Si-H}, \tag{4.39}$$

where α_{Si-H} and α_{Si-Si} are the polarizabilities of Si-H and Si-Si bonds, respectively, and N_{Si-Si} and N_{Si-H} are the number of Si-Si and Si-H bonds, respectively. While this model explicitly includes the contribution of the hydrogen atoms, it is also assumed that the bond polarizabilities are independent of the cluster size and shape.

In a hydrogenated silicon cluster with only sp^3 bonding, the number of Si-Si bonds and Si-H bonds can be expressed in terms of the number of silicon and hydrogen atoms. In particular, N_{Si-Si} is given by $(4N_{Si} - N_H)/2$, and N_{Si-H} is equal to N_H. Substituting these expressions into Eq. 4.39 yields

$$\alpha = \frac{4N_{Si} - N_H}{2} \alpha_{Si-Si} + N_H \alpha_{Si-H}. \tag{4.40}$$

Dividing both sides by N_{Si} yields the polarizability per silicon atom

$$\frac{\alpha}{N_{Si}} = 2\alpha_{Si-Si} + \left(\alpha_{Si-H} - \frac{\alpha_{Si-Si}}{2}\right)\frac{N_H}{N_{Si}}. \tag{4.41}$$

Interestingly, this shows that the polarizability per silicon atom is a function of the ratio of hydrogen and silicon atoms only.

4.2.6 Generation of clusters

To generate atomic structures of hydrogenated silicon clusters we initially followed a similar procedure as Barnard et al. [13] who carve spherical clusters from a perfect silicon crystal, terminate the dangling bonds on the surface with hydrogen atoms and then relax the atomic positions using DFT. Unfortunately, this approach only yields very few clusters with 100 or fewer silicon atoms. Because of the relatively large computational cost associated with the RPA polarizability calculations, we instead use the following approach to generate clusters: starting from the spherical $\mathrm{Si}_{123}\mathrm{H}_{100}$ cluster, we remove silicon atoms from the surface, terminating any dangling bonds with hydrogen atoms and relax the structure with DFT. In this way, a set of 100 hydrogenated silicon clusters containing between 10 and 110 Si atoms is obtained for which RPA polarizabilities are calculated. In addition, we include the spherical clusters with fewer than 123 Si atoms from Barnard et al. [13].

4.2.7 Computational details

The plane-wave/pseudopotential DFT code Quantum Espresso [70, 69] was used to obtain Kohn-Sham energies ϵ_n and wavefunctions $\phi_n(\mathbf{r})$. We employed the PBE exchange-correlation functional, norm-conserving pseudopotentials from the original Quantum Espresso Pseudopotential library [70, 69] and a plane-wave cut-off of 65 Ry. The clusters were placed in a cubic unit cell with sufficient vacuum to avoid interactions between periodically repeated images. Next, cluster polarizabilities were calculated with BerkeleyGW [90, 49] using a plane-wave cutoff of 6 Ry and a truncated Coulomb interaction. A total of 600 Kohn-Sham states were included in the summation for χ which was found to be sufficient to converge the scalar polarizabilities.

SOAP descriptors were constructed with $l_{max} = 9$ and $n_{max} = 20$ and $\gamma_\nu = 2.0$ for $r_{cut} \leq 10.0$Å and $\gamma_\nu = 0.5$ for $r_{cut} > 10.0$Å. In all calculations, we only study local environments of silicon atoms. As all hydrogen atoms are bonded to silicon atoms, their contribution to the cluster polarizabilities can be captured indirectly through their influence on the silicon atoms.

4.3 Results and Discussion

4.3.1 Fitting polarizabilities

Fig. 4.1(a) shows the RPA polarizabilities of the hydrogenated silicon clusters as function of the number of silicon atoms in the cluster. We observe that the polarizability exhibits a linear behaviour which suggests that the Si atoms provide the dominant contribution.

Deviations from the linear behaviour become explicit when the cluster polarizability is divided by the number of silicon atoms, see Fig. 4.1(b). For clusters containing more than 80 Si atoms, the polarizability per silicon atom decreases. Interestingly, α/N_{Si} increases for cluster between 70 and 80 silicon atoms, but decreases again for clusters between 40 and 70 silicon atoms. For clusters with fewer than 40 Si atoms, there is a significant amount of scatter in the polarizabilities but overall α/N_{Si} tends to increase with increasing number of Si atoms. Overall, the polarizability per silicon atom has an M-like shape as function of the number of silicon atoms.

For very large clusters, α/N_{Si} should converge to the atomic RPA polarizability of bulk silicon which is 3.77 Å^3 (determined using the Clausius-Mossotti relation and the bulk dielectric constant of 12.2 [91]). This explains the observed decrease of α/N_{Si} for $N_{Si} > 80$. Note that in our results the bulk value is not approached from below because we have not removed the hydrogen contributions from the cluster polarizabilities [129, 97].

To understand these findings, we first compare the results to two physical-based models: a model in which the cluster polarizability is assumed to be proportional to the number of Si atoms (denoted the linear N_{Si} model) and a bond polarizability model (see Methods). The parameters of both models were fitted to the calculated RPA data using a least squares optimization. The results are shown in Fig. 4.2 (a). While the linear N_{Si} model cannot capture any dependence of

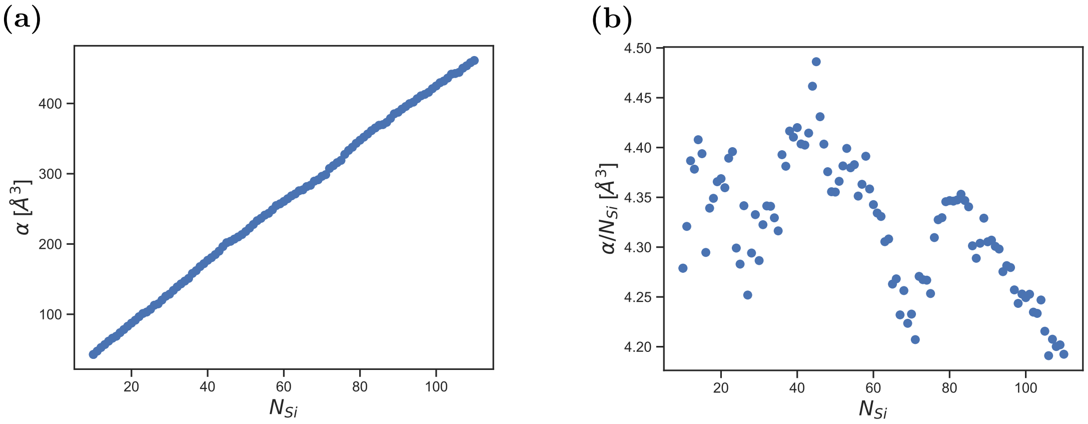


Figure 4.1: (a) RPA scalar polarizability α of hydrogenated silicon clusters versus number of silicon atoms N_{Si}. The polarizability increases approximately linearly with N_{Si}. (b) RPA polarizability divided by N_{Si} shows deviations from linearity.

α/N_{Si} on the number of silicon atoms, the bond polarizability model correctly describes several key features. In particular, it shows a decreasing trend for large clusters and a minimum near $N_{Si} = 70$. For small clusters, the bond polarizability model predicts an increase in polarizability as the number of Si atoms is reduced in disagreement with the RPA data. Interestingly, the bond polarizability model also features a significant scatter for small clusters. As discussed in the methods section, α/N_{Si} in the bond polarizability model only depends on the ratio of hydrogen and silicon atoms N_H/N_{Si} suggesting that the fraction of surface hydrogen atoms captures a large part of the variation the polarizability of the clusters.

While the bond polarizability model captures several features, we note that neither of the two physical models can capture the full M-shape of the polarizability per Si atom in Fig. 4.2 (a). Furthermore, from the least square fits of the linear N_{Si} model to the RPA data, we find $\alpha_{Si}^{\text{av}} = 4.29$ Å^3. This is significantly larger than the RPA value in bulk Si of 3.77 Å^3. The parameters of the bond polarizability model are found to be $\alpha_{Si-Si} = 1.98$ Å^3 and $\alpha_{Si-H} = 1.32$ Å^3. As the polarizability per Si atom is $2\alpha_{Si-Si}$, the predicted bulk value is 3.96 Å^3, which is in better agreement with RPA results.

The above analysis demonstrates that both physical-based models have several shortcomings. This is a consequence of two factors: (i) their parameters do not depend on the properties of the local chemical environment, i.e. bond lengths or bond angles. In particular, significant atomic relaxations occur for small clusters, resulting in changes to the bond polarizabilities compared

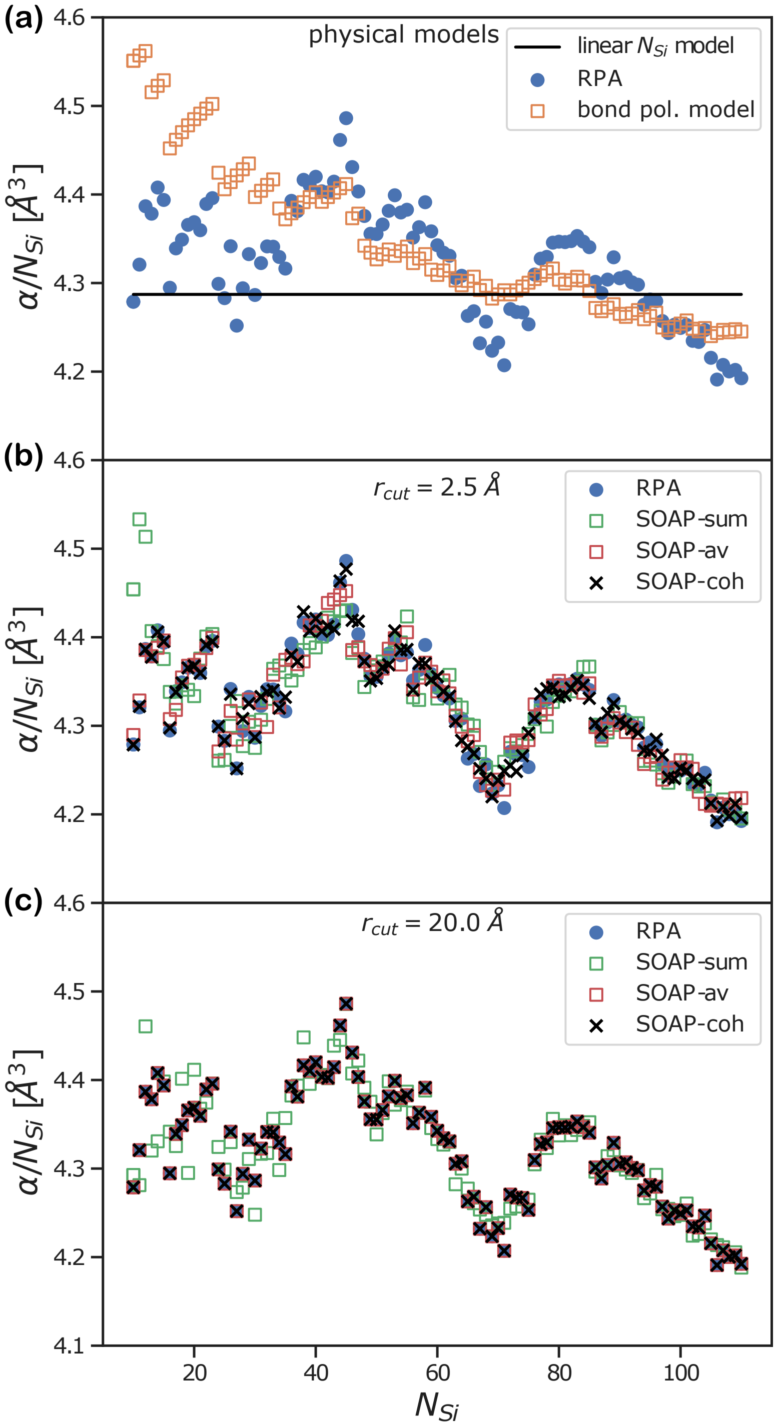


Figure 4.2: Comparison of α/N_{Si} (cluster polarizability divided by the number of silicon atoms) from RPA calculations (blue dots) with fits from various models. (a) The black line shows the optimal fit for the linear N_{Si} model, see Eq. 4.38. Orange triangles denote results from the bond polarizability model; see Eq. 4.39. (b) Results obtained using SOAP with the sum kernel (green squares), the average kernel (red squares) and the coherent kernel (black crosses). A cut-off of 2.5 Å was employed. (c) Same as (b) but with a cut-off of 20 Å.

to the larger clusters that the bond polarizability model does not capture. (ii) The models do not capture the effects of interactions between the polarizable units. Consequently, they cannot distinguish between clusters containing the same numbers of Si and H atoms and do not capture the dependence of the polarizability on the cluster shape. To overcome these problems, we now explore the ability of machine learning models to describe the polarizabilities of Si clusters.

Figures 4.2 (b) and 4.2 (c) show the results from the machine learning model using the sum kernel, the average kernel and the coherent kernel (see Methods). The real-space cut-off that determines the size of the chemical environment of each atom is $r_c = 2.5$ Å in Fig. 4.2(b) and $r_c = 20$ Å in Fig. 4.2(c). In the fit, the regularization parameter λ was kept small (10^{-15} for the sum kernel model and 10^{-12} for the average and coherent kernels) in order to allow as much flexibility in the parameters as possible. Furthermore, no train-test split was performed. For the smaller cut-off (where only the nearest neighbour atoms are included in the local environment), all three kernels provide an improved description compared to the physical-based models. Specifically, they capture the M-shape of α/N_{Si} as a function of N_{Si} and also reproduce the scatter for smaller clusters. The coherent kernel is slightly better than the averaged kernel. Significant deviations from the calculated polarizabilities are only observed for the smallest cluster sizes when the sum kernel is used. When r_c is increased to 20 Å, the agreement between the ML models and the calculated polarizabilities significantly improves. In particular, the average and the coherent kernel results almost perfectly agree with the data, while the sum kernel results show minor deviations for smaller clusters. The good results obtained for the short cut-off indicate that local chemical effects dominate polarizabilities. However, long-range interactions also influence polarizabilities. These long-range interactions are captured when the cut-off radius is increased.

4.3.2 Predicting polarizabilities

Up to this point, we only considered the ability of the SOAP approach to fit the calculated cluster polarizabilities. To investigate SOAP's capacity to predict polarizabilities of clusters that it was not trained on, k-fold cross validation [152] is employed. In this procedure, the clusters in the data set are randomly assigned to five sub-sets. Next, four sub-sets are used to train the ML approach, and the fifth sub-set is used as the test set. This is done five times, with each sub-set acting as a test set once. We optimize the regularization parameter λ to minimize the mean average error (MAE). The optimal parameters are listed in Table 4.1 below.

Table 4.1: Regularization paramaters λ_{av} and λ_{sum} determined from k-fold cross validation at different cut-off radii r_{cut}.

r_{cut} [Å]	λ_{av}	λ_{sum}	λ_{coh}
2.5	10^{-8}	10^{-8}	10^{-8}
5.0	10^{-5}	0.0001	10^{-6}
7.5	10^{-5}	0.01	10^{-5}
10.0	0.0001	0.01	0.0001
12.5	10^{-5}	10^{-6}	10^{-5}
15.0	10^{-5}	10^{-5}	10^{-5}
17.5	0.0001	0.001	0.0001
20.0	0.0001	0.001	0.0001

The resulting MAE and its standard deviation as a function of r_{cut} are shown in Fig. 4.3 (a). The average kernel and the coherent kernel yield very similar results and are compared in Fig. 4.3 (b). Strikingly, the sum kernel model produces the largest MAE for the test set among all methods. In particular, the test set MAE is significantly larger than the training set MAE indicating poor capacity to predict polarizabilities. In contrast, the average kernel model yields the smallest test set MAE, which is only slightly worse than the training set error. The coherent kernel model yields slightly worse predictions than the average kernel, with the most significant difference between the two occurring at $r_{cut} = 5.0$ Å. The MAE of the two physical-based models lies between those of the sum kernel and the average kernel. The different performances of the sum kernel and the average kernels originate from the different training procedures: the sum kernel model is trained on total cluster polarizabilities. In contrast, the average kernel is trained on the average polarizability per silicon atom, see Eq. (4.33). Consequently, the sum kernel model is biased towards more accurate predictions for large clusters and is less accurate for small clusters. This can also be seen in Fig. 4.2(c), which shows that the quality of the sum kernel fit improves for larger clusters. This has been observed before by Stocker et al. [174], who argued that the intensive average kernel has the advantage of equally weighting small and large molecules, which is beneficial when learning quantities over an extensive range of cluster sizes. Interestingly, the average kernel performs somewhat better than the coherent kernel suggesting

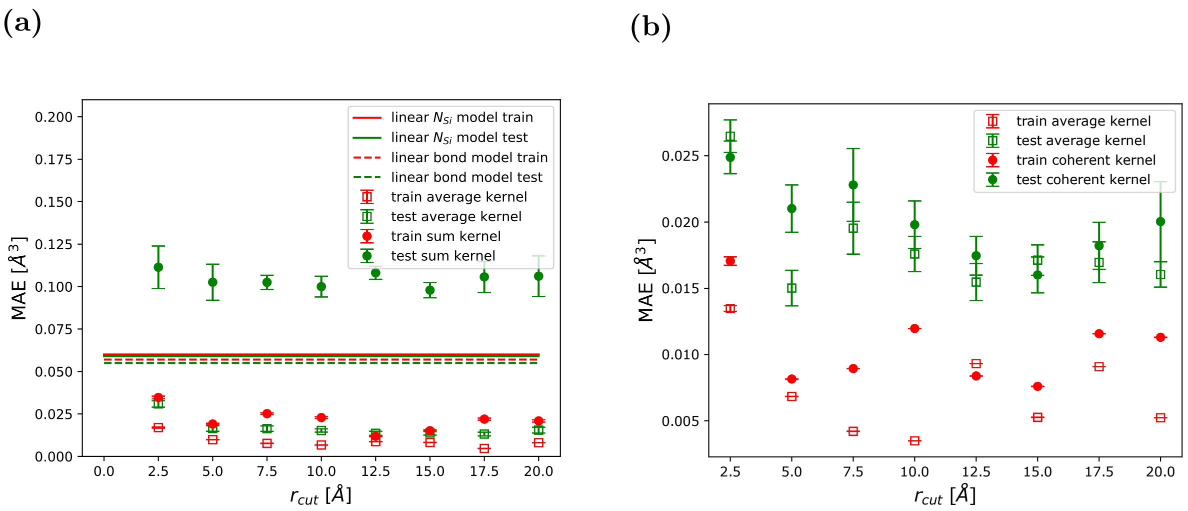


Figure 4.3: Average mean absolute error (MAE) of α/N_{Si} (the cluster polarizability divided by the number of silicon atoms) from various machine learning models and physical-based models versus the cut-off radius r_{cut} that determines the size of the local chemical environment. Optimal regularization parameters were determined using five-fold cross-validation. Error bars indicate the standard error of the average MAE across the five training and validation sets used in the cross-validation procedure.

that a model of the cluster polarizability that can be expressed as a sum of atomic contributions constitutes a better representation of the system's dielectric response.

Fig. 4.3 also shows that the minimum test set MAE for the average kernel and the coherent kernel is obtained for cut-offs around $r_{cut} = 12.5$ Å. In contrast, for the sum kernel, the minimum is achieved for $r_{cut} = 15.0$ Å. Interestingly, neither kernel benefits significantly from increasing r_{cut} beyond 5 Å. To understand this finding, we compare the elements of the average kernel matrix for $r_{cut} = 5.0$ Å and $r_{cut} = 17.5$ Å, see Fig. 4.4. When a smaller cut-off is used, the elements of the kernel matrix decay slowly along the rows and columns. In contrast, the decay is significantly more pronounced for the larger cut-off, suggesting that a smaller cut-off facilitates the recognition of similar chemical environments in clusters of different sizes. This is not surprising because, for large cut-offs, the chemical environment contains a significant amount of vacuum for small clusters but not for large clusters.

Next, the ability of the ML approach to predict polarizabilities of large clusters based on a training set of small clusters is explored. For this, we train the average kernel on the 60, 70 or 80 smallest clusters and then predict the polarizabilities of the remaining large clusters in the data set. Fig. 4.5 shows the resulting test set MAE as a function of the cutoff radius. All curves

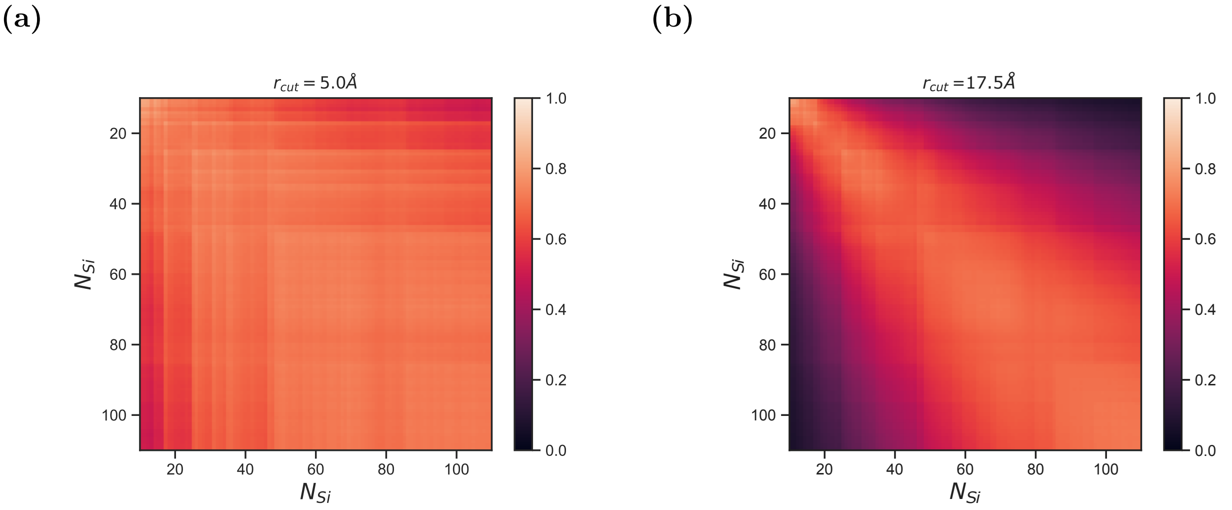


Figure 4.4: Matrix elements of the average SOAP kernel K^{av}_{IJ}, see Eq. (4.31), for cut-offs r_{cut} of 5.0 Å (a) and 17.5 Å (b). Note the rapid decay of the matrix elements along rows and columns when a large cut-off radius is used.

exhibit a minimum at small cut-offs near $r_{cut} = 5$ Å and the smallest MAE is obtained for the largest training set. For the smaller training sets ($n_t = 60$ or 70), the MAE increases rapidly as the cutoff is increased, while for the largest training set, the increase is mild (and another minimum is found at $r_{cut} = 15.0$ Å). Similar to our findings in the k-fold cross-validation, this shows that increasing the cut-off radius beyond a certain value is not beneficial.

Figures 4.6 (a)-(c) compare the predictions of the average kernel with $r_{\rm cut} = 5$ Å with the calculated RPA polarizabilities per silicon atom. The ML model captures the qualitative trends for all three training set sizes. For $n_t = 60$, the average kernel correctly predicts the increase of α/N_{Si} at $N_{Si} = 70$ and also the decrease starting at $N_{Si} = 80$. While the ML models underestimate the polarizabilities per Si atom for large clusters when $n_t = 60$ and $n_t = 70$, good quantitative agreement is achieved for $n_t = 80$.

Finally, we train the average kernel model on the entire data set (using $r_{cut} = 5$ Å) and predict the average polarizabilities of the entire Silicon Quantum Dot data set containing clusters with up to 3000 silicon atoms [13]. The results are shown in Fig. 4.7. It can be observed that the polarizability per Si atom converges slowly to its bulk limit as N_{Si} increases, and there is a significant scatter in the results. The scatter in α/N_{Si} reflects the different N_H/N_{Si} ratios and different environments present in the clusters. To understand the slow convergence to the bulk value, note that the number of silicon atoms scales with the cluster volume, while the number

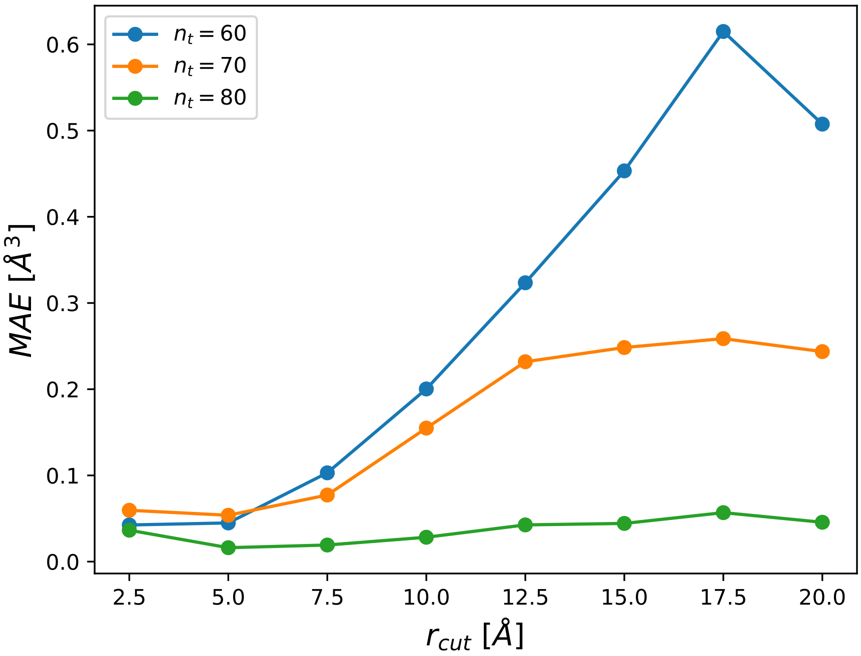


Figure 4.5: Average kernel test set error as a function of SOAP cut-off. The smallest n_t clusters were included in the training set for each curve. The test set consists of the remaining $100 - n_t$ clusters.

of hydrogen atoms is roughly proportional to the surface area. This suggests that α/N_{Si} should be proportional to the inverse radius of the cluster or, equivalently, to $1/N_{Si}^{1/3}$. Indeed, Fig. 4.7 shows that the ML predictions are well described by the function $a + b/N_{Si}^{1/3}$ with $a = 3.89$ Å^3 and $b = 1.55$ obtained from a least-squares fit. The value of a agrees well with the RPA atomic polarizability of bulk silicon of 3.77 Å^3[91].

Additional insights can be obtained by analyzing the atomic polarizabilities obtained from the SOAP average kernel method; see Eq. (4.34). Fig. 4.8 shows the atomic polarizabilities of a $Si_{2109}H_{604}$ cluster. In Fig. 4.8 (a) only local chemical environments of silicon atoms are considered (and the effect of the hydrogen atoms is captured indirectly through their influence on the silicon chemical environments). Silicon atoms in the centre of the cluster have a polarizability of 3.76 Å^3, in excellent agreement with the value extracted from bulk calculations of 3.77 Å^3 [91]. The polarizability of the silicon atoms in the two surface layers is larger, sometimes as large as 5 Å^3. The reason for this increase is that the surface silicon atoms are bonded to hydrogen atoms, and their atomic polarizability is effectively the sum of the silicon and hydrogen contributions.

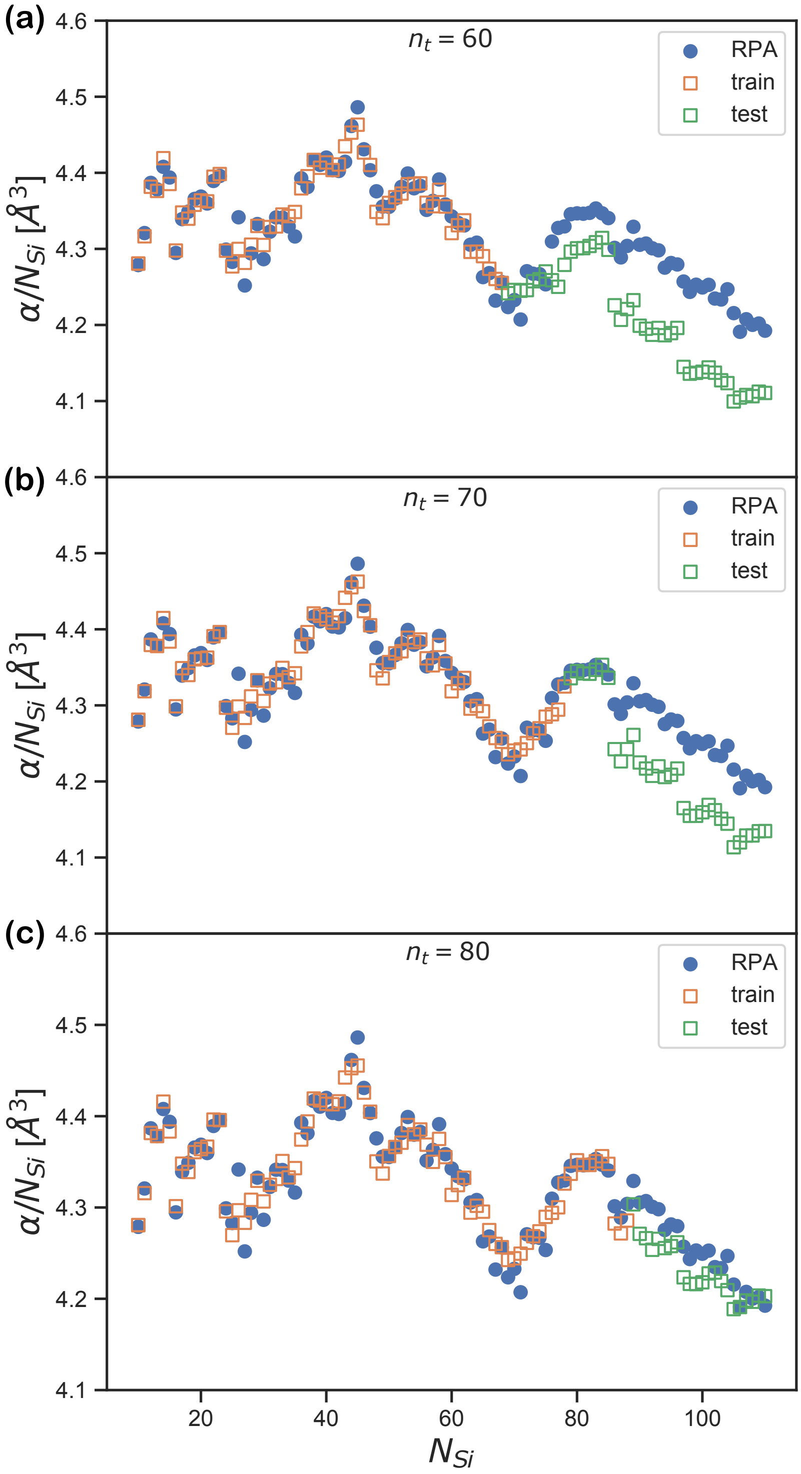


Figure 4.6: Comparison of RPA results for α/N_{Si} (cluster polarizability divided by the number of silicon atoms) and training and test set predictions of the average kernel model. The training set consists of the (a) $n_t = 60$, (b) $n_t = 70$ and (c) $n_t = 80$ smallest clusters and the test set contains the remaining $100 - n_t$ large clusters.

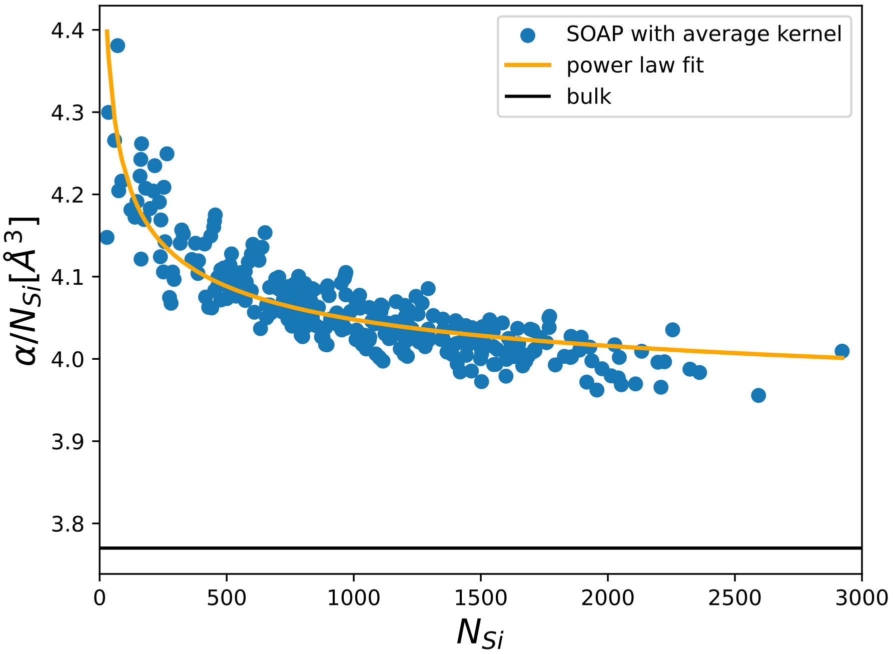


Figure 4.7: Cluster polarizability divided by the number of silicon atoms for all clusters of Silicon Quantum Dot database [13] from the average SOAP kernel model with $r_{cut} = 5.0$ Å. Bulk polarizability obtained from Hybertsen et al. [91].

To disentangle contributions from silicon and hydrogen atoms to the cluster polarizability, Fig. 4.8 shows the atomic polarizabilities from a calculation that explicitly takes the chemical environments of hydrogen atoms into account. Interestingly, the results suggest that the atomic polarizability of subsurface silicon atoms is larger than the bulk value, but the polarizability of surface silicon atoms (which are bonded to hydrogens) is smaller. The average atomic polarizability of the silicon atoms is found to be 3.63 Å^3. This is in agreement with the results of Mochizuki et al. [129], who predicted that the bulk limit of the silicon atomic polarizability is approached from below.

4.4 Conclusions

In this chapter, it was demonstrated that machine learning models based on the smooth overlap of atomic positions (SOAP) descriptor can be used to accurately and efficiently predict polarizabilities of large hydrogenated silicon clusters. Using the random phase approximation, the polarizabilities of a set of hydrogenated silicon clusters containing between 10 and 110 silicon atoms were calculated. We then assessed the ability of three machine learning models (one using the sum kernel, one using the average kernel and one the coherent kernel) to fit the calculated polarizabilities and find that all three models perform well when the local environment includes nearest neighbour atoms only. Increasing the size of the environment improves the quality of the fit. Next, we investigated the ability of the machine learning models to predict polarizabilities of clusters that are not in the training set. Using k-fold cross-validation, we find that the average kernel performs significantly better than the sum kernel and that the predictions only weakly depend on the size of the chemical environment. We also tested the predictive power of the average kernel when it is trained on small clusters only and found that quantitative accuracy can be achieved if the training set is sufficiently large. Finally, we use the average kernel approach to predict the polarizabilities of hydrogenated silicon atoms with up to 3000 silicon atoms and find that the results approach the correct bulk limit. The ability to efficiently calculate polarizabilities of large clusters paves the way towards using machine learning for excited-state properties of these systems. These results further suggest that quantities derived from the DDRF are well described by descriptors based on the local chemical environment of an atom. In the following two chapters, the locality of the DDRF in semi-conductor clusters is exploited to achieve an atomic decomposition of the DDRF and develop flexible model DDRFs which can be used in GW calculations.

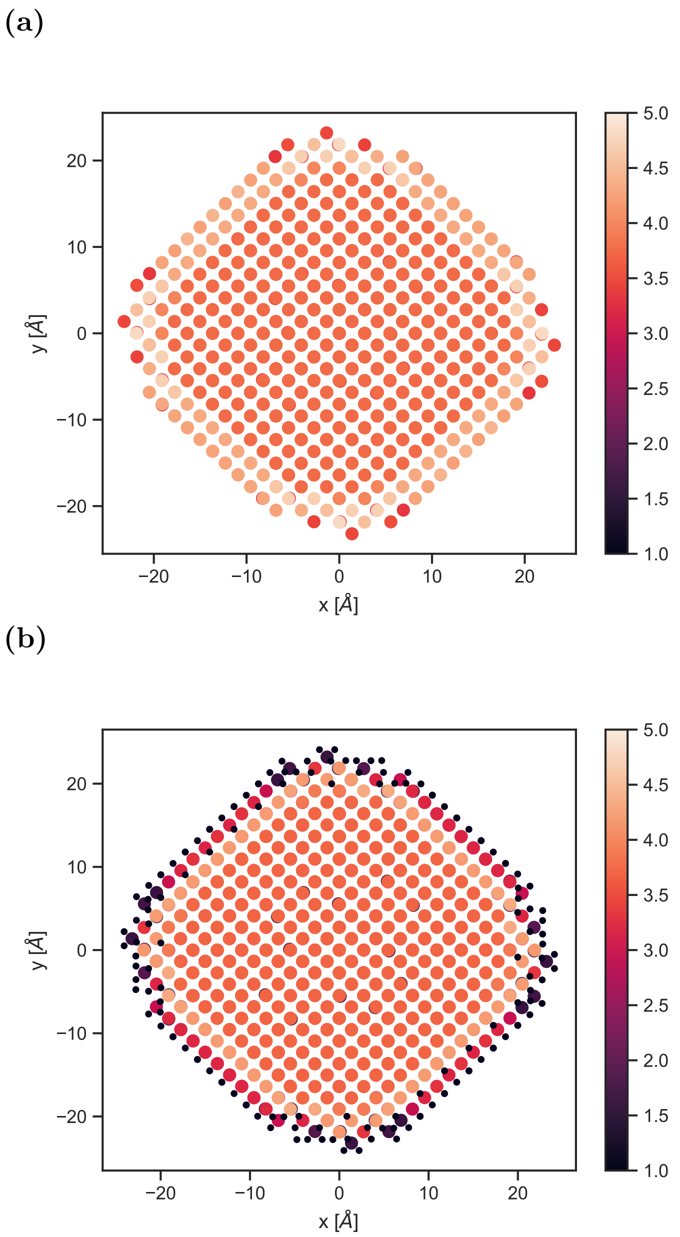


Figure 4.8: Atomic polarizabilities of the $Si_{2109}H_{604}$ cluster obtained from the SOAP average kernel method. Shown is a cross-section through the centre of the cluster. (a) Atomic polarizabilities when only silicon chemical environments are used. (b) Atomic polarizabilities when both silicon and hydrogen chemical environments are used. For hydrogen environments $r_{cut} = 1.6$ Å was used and for silicon environments $r_{cut} = 5.0$ Å was used. Large dots represent silicon atoms, and small dots represent hydrogen atoms.

Chapter 5

Approximating the RPA density-density response function

5.1 Introduction

As mentioned in Chapter 1, the GW method [90, 139] is often employed in the calculation of quasiparticle energies, in which corrections to the DFT energies are computed perturbatively. While yielding highly accurate excitation energies, the GW method is largely restricted to small systems due to its poor scaling with system size [139, 153]; the most expensive step is the computation of the density-density response function (DDRF), which is typically computed as a sum over states using Eq. (3.42). In order to converge the sum, however, a large number of empty states have to be included in the summation, which is the main cause of the poor scaling with system size [139].

The poor convergence of the computation of the DDRF led to the development of several model dielectric functions, the simplest of which are based on the Lindhard dielectric function [118] of the free electron gas. More general model functions were proposed later by Hybertsen, Levine and Louie [114, 92] and Cappellini et al. [35, 19]. However, as noted by Rohlfing [153], these model functions cannot be used for systems with non-uniform screening, as they rely on a spatially uniform dielectric constant. This restriction makes these model functions especially unsuitable in the extreme case of isolated clusters or molecules, where one cannot start from the assumption of uniform screening. Rohlfing [153] proposed to model the dielectric function as a sum of atomic contributions, attributing a density response resulting from a Gaussian-shaped charge density to each atom. The various parameters in Rohlfing's model are obtained by fitting to existing data for RPA dielectric functions. The resulting dielectric function yields reliable estimates of the band gaps in extended systems [54, 9]. However, the errors for individual quasiparticle (QP) energies, in particular energies below the conduction band and above the valence band, are less reliable [153]. Furthermore, the errors in QP energies for finite systems are substantially larger, approximately 0.5 eV for HOMO-LUMO gaps and 0.7 eV for HOMO energies [153]. This failure is likely a consequence of Rohlfing's approach using a baseline model, which assumes metallic screening [153], to which corrections are calculated.

In Chapter 4, I showed that the RPA scalar polarizability, which is a scalar quantity derived from the DDRF, can be accurately predicted within an atom-centred framework. The results from the previous chapter and the limitations of existing model dielectric functions are the motivation for the next two chapters, where I build the foundation for accurate machine learning models of the DDRF. Predicting such quantities is a formidable challenge: for example, the DDRF of a small silicon cluster can be tens of gigabytes in size when represented in a plane-wave basis, even when a modest plane-wave cutoff is used. As such, in order to make progress, it is necessary to express the DDRF in a more compact way. In this chapter, I develop an approximation to the DDRF by projecting it onto an auxiliary basis of Gaussian-type orbitals (GTOs). This approach drastically reduces the size of the DDRF while retaining most of the accuracy in quasi-particle (QP) energies.

I then apply the method of projecting the DDRF onto an atom-centred basis of GTOs to the silicon clusters used in the polarizability calculations in Chapter 4. Finally, a simple machine-learning model of the approximated DDRF is developed, and its shortcomings are discussed in order to motivate the development of a more systematic machine-learning approach in the next chapter.

5.2 Recap of the role of the DDRF within GW

The GW method relies on applying a correction to mean-field energy levels that is based on the self-energy $\Sigma(\mathbf{r},\mathbf{r}',\omega)$, which is in turn calculated from the many-body Green's function $G(\mathbf{r},\mathbf{r}')$ and the screened Coulomb interaction $W(\mathbf{r},\mathbf{r}',\omega)$: [139, 90, 89]

$$\Sigma(\mathbf{r},\mathbf{r}',\omega) = \frac{i}{2\pi}\int e^{-i\delta\omega'} G(\mathbf{r},\mathbf{r}',\omega+\omega')W(\mathbf{r},\mathbf{r}',\omega')d\omega' \tag{5.1}$$

The screened Coulomb interaction is, in turn, computed from the bare Coulomb interaction $\nu(\mathbf{r},\mathbf{r}')$ and the dielectric function $\epsilon(\mathbf{r},\mathbf{r}',\omega)$:

$$W(\mathbf{r},\mathbf{r}',\omega) = \int \epsilon^{-1}(\mathbf{r},\mathbf{r_2},\omega)\nu(\mathbf{r_2},\mathbf{r}')d\mathbf{r_2}. \tag{5.2}$$

As such, the dielectric function forms one of the key ingredients of GW calculations, where $\epsilon(\mathbf{r},\mathbf{r_2})$ is computed from the interacting DDRF

$$\epsilon^{-1}(\mathbf{r},\mathbf{r}',\omega) = \delta(\mathbf{r},\mathbf{r}') + \int \nu(\mathbf{r},\mathbf{r_2})\chi(\mathbf{r}',\mathbf{r_2},\omega)d\mathbf{r_2}. \tag{5.3}$$

The non-interacting DDRF χ_0 is calculated as a sum over empty and occupied states within the Adler-Wiser [3, 195] formulation in Eq. (3.42). Then, the interacting DDRF χ is computed using the Dyson equation

$$\chi(\mathbf{r},\mathbf{r}') = \chi_0(\mathbf{r},\mathbf{r}') + \int d\mathbf{r}_1 d\mathbf{r}_2 \chi_0(\mathbf{r},\mathbf{r}_1)v(\mathbf{r}_1,\mathbf{r}_2)\chi(\mathbf{r}_2,\mathbf{r}'). \tag{5.4}$$

The resulting Dyson-equation can be solved by either a truncated summation or, more accurately, by inversion of the dielectric function given by

$$\epsilon(\mathbf{r},\mathbf{r}',\omega) = \delta(\mathbf{r},\mathbf{r}') - \int \nu(\mathbf{r},\mathbf{r_2})\chi_0(\mathbf{r}',\mathbf{r_2},\omega)d\mathbf{r_2}. \tag{5.5}$$

Computing the inverse dielectric function is one of the critical bottlenecks of a GW calculation, the reason for which is twofold: (1) Due to the energy dependence of the non-interacting DDRF in Eq. (3.42), a large number of empty states is required for convergence. (2) To invert the dielectric function in Eq. (5.5), it has to be represented on either a real-space grid or in a basis set, both of which require the inversion of a large matrix. Since the density-density response function can be several gigabytes in size when represented in a plane-wave basis, a more efficient representation that makes matrix elements readily learnable has to be found. In the remainder of this chapter, I will focus on the static DDRF $\chi(\mathbf{r},\mathbf{r}',0) \equiv \chi(\mathbf{r},\mathbf{r}')$, since the frequency dependence can be approximated using the Generalized Plasmon-Pole model (GPP) [90, 171].

5.3 A compact representation of the DDRF

The Adler-Wiser formulation in Eq. (3.42) described in Chapter 3 consists of a product of four mean-field Eigenfunctions. When expanding Eq. (3.42) in a real, atom-centred basis set, one obtains a four-point function for χ_0

$$\chi_0(\mathbf{r},\mathbf{r}') = \sum_{ijkl} d_{ijkl}\phi_i(\mathbf{r})\phi_j(\mathbf{r})\phi_k(\mathbf{r}')\phi_l(\mathbf{r}'). \tag{5.6}$$

By expanding Eq. (5.4), one can show that the interacting DDRF χ can also be written as a four-point function using a different set of coefficients

$$\chi(\mathbf{r},\mathbf{r}') = \sum_{ijkl} \chi_{ijkl}\phi_i(\mathbf{r})\phi_j(\mathbf{r})\phi_k(\mathbf{r}')\phi_l(\mathbf{r}'). \tag{5.7}$$

However, as commonly done in quantum chemistry for fitting electron densities, one can use an auxiliary density-fitting basis set to represent products of basis functions [110] efficiently. Thus, using an appropriate basis set, one can approximate χ as a two-point function:

$$\chi(\mathbf{r},\mathbf{r}') = \sum_{ij} \chi_{ij}\phi_i(\mathbf{r})\phi_j(\mathbf{r}'). \tag{5.8}$$

Within BerkeleyGW, the DDRF is represented in a plane wave basis [90, 49]. Thus, the real-space DDRF is given by

$$\chi(\mathbf{r},\mathbf{r}') = \frac{1}{V}\sum_{\mathbf{G},\mathbf{G}'}\chi_{\mathbf{G},\mathbf{G}'}e^{i\mathbf{G}\cdot\mathbf{r}}e^{-i\mathbf{G}'\cdot\mathbf{r}'}. \tag{5.9}$$

Assuming the atom-centred basis set is orthogonal, the projection of the BerkeleyGW DDRF onto the atom-centred basis set is computed as

$$\chi_{ij} = \frac{1}{V}\sum_{\mathbf{G},\mathbf{G}'}\chi_{\mathbf{G},\mathbf{G}'}\int_V \phi_i(\mathbf{r})e^{i\mathbf{G}\cdot\mathbf{r}}d\mathbf{r}\int_V \phi_j(\mathbf{r}')e^{-i\mathbf{G}'\cdot\mathbf{r}'}d\mathbf{r}' = \sum_{\mathbf{G},\mathbf{G}'}\tilde{\phi}_i^*(\mathbf{G})\tilde{\phi}_j(\mathbf{G}')\chi_{\mathbf{G},\mathbf{G}'}. \tag{5.10}$$

The functions $\tilde{\phi}_i(\mathbf{G})$ denote the Fourier transforms of the orthogonal basis functions, defined as

$$\tilde{\phi}_i(\mathbf{G}) = \frac{1}{\sqrt{V}}\int_V \phi_i(\mathbf{r})e^{-i\mathbf{G}\cdot\mathbf{r}}d\mathbf{r}, \tag{5.11}$$

where the integral runs over the unit-cell volume V.

5.3.1 Introducing a basis of CGTOs

The first step in achieving the previously outlined projection for an explicit basis of Cartesian Gaussian-type orbitals (CGTOs) is constructing an orthogonal set of basis functions. The starting point are a set of CGTOs $\{\phi^a_{\alpha_a}(\mathbf{r})\}$, with a labelling the atom on which the CGTO is centred and α_a labelling the basis functions on atom a. Building the orthogonal basis set can be achieved in several ways. However, the simplest method is to choose the orthogonal set $\{\hat{\phi}^a_{\alpha_a}(\mathbf{r})\}$ from the Eigenvectors of the overlap matrix

$$S^{ab}_{\alpha_a\alpha_b} = \int d\mathbf{r}\phi^a_{\alpha_a}(\mathbf{r})\phi^b_{\alpha_b}(\mathbf{r}). \tag{5.12}$$

The orthogonal basis functions are then defined as

$$\hat{\phi}^i_{\alpha_i}(\mathbf{r}) = \sum_k\sum_{\alpha_k}A^{\alpha_i\alpha_k}_{ik}\phi^k_{\alpha_k}(\mathbf{r}), \tag{5.13}$$

where $A_{ik}^{\alpha_i\alpha_k}$ is the matrix of eigenvectors of the overlap matrix. The coefficients of the DDRF, when expanded in the orthogonalized basis are

$$\hat{\chi}_{\alpha_i\alpha_j}^{ij} = \frac{1}{V}\sum_{\mathbf{G},\mathbf{G}'}\chi_{\mathbf{G},\mathbf{G}'} \times \int_{-\infty}^{\infty}\hat{\phi}_{\alpha_i}^{i}(\mathbf{r})e^{i\mathbf{G}\cdot\mathbf{r}}d\mathbf{r}\int_{-\infty}^{\infty}e^{-i\mathbf{G}'\cdot\mathbf{r}'}\hat{\phi}_{\alpha_j}^{j}(\mathbf{r}')d\mathbf{r}', \tag{5.14}$$

where, due to the localised nature of the basis functions, I extended the integral from an integral over the supercell to an integral over all space. It should be noted that orthogonalizing the CGTO basis generally yields extended functions. However, the orthogonalized functions are a linear combination of GTOs. Thus, extending the integral over all space is a valid approximation if the decay coefficients of the GTOs and the distances to the supercell edge are sufficiently large. These integrals are proportional to the Fourier transforms of the basis functions (or their complex conjugates). $\hat{\chi}$ can then be transformed from the orthogonal basis to the atom-centred basis using the coefficients $A_{ik}^{\alpha_i\alpha_k}$ from Eq. (5.13). This transformation can be expressed in terms of matrix multiplication with the transformation matrix $\mathbf{A}$, containing these coefficients.

$$\begin{aligned}\chi(\mathbf{r},\mathbf{r}') = \sum_{\alpha_i\alpha_j}\sum_{ij}\hat{\chi}_{\alpha_i\alpha_j}^{ij}\hat{\phi}_{\alpha_i}^{i}(\mathbf{r})\hat{\phi}_{\alpha_j}^{j}(\mathbf{r}') &= \sum_{\alpha_k\alpha_l}\sum_{kl}\sum_{\alpha_i\alpha_j}\sum_{ij}A_{ik}^{\alpha_i\alpha_k}A_{jl}^{\alpha_j\alpha_l}\hat{\chi}_{\alpha_i\alpha_j}^{ij}\phi_{\alpha_k}^{k}(\mathbf{r})\phi_{\alpha_l}^{l}(\mathbf{r}') \\ &= \sum_{\alpha_k\alpha_l}\sum_{kl}\chi_{\alpha_k\alpha_l}^{kl}\phi_{\alpha_k}^{k}(\mathbf{r})\phi_{\alpha_l}^{l}(\mathbf{r}'), \end{aligned}\tag{5.15}$$

where I defined,

$$\chi_{\alpha_k\alpha_l}^{kl} = \sum_{\alpha_i\alpha_j}\sum_{ij}\hat{\chi}_{\alpha_i\alpha_j}^{ij}A_{ik}^{\alpha_i\alpha_k}A_{jl}^{\alpha_j\alpha_l}. \tag{5.16}$$

In matrix notation, the above equation can be written as

$$\chi = \mathbf{A^T}\hat{\chi}\mathbf{A}. \tag{5.17}$$

5.3.2 Evaluation of the Fourier transform

A primitive CGTO is defined as

$$\phi_{abc}(\mathbf{r}) = N_{abc}(x-x_0)^a(y-y_0)^b(z-z_0)^c e^{-\alpha|\mathbf{r}-\mathbf{r_0}|^2}, \tag{5.18}$$

where N_{abc} is a normalization constant and $\mathbf{r_0}$ is the position of the atom the orbital is centred on. The angular momentum l of a CGTO is given by the sum of the exponents, i.e. $l = a+b+c$. The 3-D Fourier transform of a CGTO decomposes into a product of three 1-D Fourier transforms, which simplifies the problem to three 1-D Fourier transforms of functions of the form (ignoring the normalization constant)

$$f_k(x) = (x - x_0)^k e^{-\alpha(x-x_0)^2}. \tag{5.19}$$

Moreover, using the shift property of the Fourier transform, the Fourier transform only has to be computed for $x_0 = 0$. The resulting Fourier transform is multiplied by a phase factor of $e^{-iG_x x_0}$. Thus, the problem is simplified to evaluating Fourier transforms of the form:

$$\tilde{f}_k(G_x) = \int x^k e^{-\alpha x^2} e^{-iG_x x} dx \tag{5.20}$$

Conveniently, the Fourier transform has the property that $FT(xf(x))(G_x) = i\frac{\partial}{\partial G_x} FT(f(x))(G_x)$. This allows the derivation of a recursive expression for the Fourier transform of a Gaussian of order k, where I will denote

$$F_k = FT(f_k)(G_x). \tag{5.21}$$

F_0 can be computed by simply using the Fourier transform of a 1-D Gaussian, centred at 0. Extending the Fourier integral over the unit cell to an integral over all space, the Fourier transform can be evaluated analytically and is given by

$$\tilde{f}_0(G_x) = \sqrt{\frac{\pi}{\alpha}} e^{-\frac{G_x^2}{4\alpha}}. \tag{5.22}$$

The recursive expression can then be constructed by repeatedly applying the operator $i\frac{\partial}{\partial G_x}$ to F_0. Starting with F_1

$$F_1 = i\frac{\partial}{\partial G_x} F_0 = i\frac{\partial}{\partial G_x} \sqrt{\frac{\pi}{\alpha}} e^{-\frac{G_x^2}{4\alpha}} = -\frac{iG_x}{2\alpha} F_0 \tag{5.23}$$

$$F_2 = i\frac{\partial}{\partial G_x}F_1 = \frac{1}{2\alpha}F_0 - \frac{iG_x}{2\alpha}F_1 \tag{5.24}$$

$$F_3 = i\frac{\partial}{\partial G_x}F_1 = \frac{2}{2\alpha}F_1 - \frac{iG_x}{2\alpha}F_2 \tag{5.25}$$

Repeated application of the operator results in the following pattern

$$F_n = \frac{n-1}{2\alpha}F_{n-2} - \frac{iG_x}{2\alpha}F_{n-1}. \tag{5.26}$$

The Fourier transform of the CGTO is then given by

$$\tilde{\phi}_{klm}(\mathbf{G}) = F_k(G_x)F_l(G_y)F_m(G_z)e^{-i\mathbf{G}\cdot\mathbf{r_0}}. \tag{5.27}$$

5.3.3 Enforcing of the integration sum-rule of the projected density-density response function

Since the DDRF is used to calculate the density response to a perturbing potential $V(\mathbf{r})$, it is clear that for the charge to be conserved, the following sum rule has to apply

$$\int\int d\mathbf{r}d\mathbf{r}'\chi(\mathbf{r},\mathbf{r}')V(\mathbf{r}') = 0 \quad \forall V(\mathbf{r}'). \tag{5.28}$$

When χ is expanded in a plane wave basis, this results in the following constraint on the plane wave coefficients

$$\int\int d\mathbf{r}d\mathbf{r}' \sum_{\mathbf{G},\mathbf{G}'} \chi_{\mathbf{G}\mathbf{G}'}e^{i\mathbf{G}\cdot\mathbf{r}}e^{-i\mathbf{G}'\cdot\mathbf{r}'}V(\mathbf{r}') = \sum_{\mathbf{G},\mathbf{G}'}\chi_{\mathbf{G}\mathbf{G}'}\delta_{\mathbf{G}\mathbf{0}}V(\mathbf{G}') = \sum_{\mathbf{G}'}\chi_{\mathbf{0}\mathbf{G}'}V(\mathbf{G}') = 0. \tag{5.29}$$

For this condition to be true for all $V(\mathbf{r})$, it is necessary that all $\chi_{\mathbf{0}\mathbf{G}'}$ are 0. Since the static density-response is real, $\chi_{\mathbf{G}\mathbf{G}'}$ is a hermitian matrix. Thus it is also required that all $\chi_{\mathbf{G}\mathbf{0}}$ are 0. For the CGTO basis, I enforced this constraint as a post-processing step, i.e. by setting the first row and column of the Fourier transform of the projected matrix to zero. This constraint helped in improving the quality of the quasi-particle energy levels when performing a GW calculation with the projected matrix.

5.3.4 Cartesian vs. solid harmonic Gaussians

An alternative to the CGTO basis are the solid harmonic Gaussian basis functions (SGTOs), defined as

$$\phi_{nlm}(r,\theta,\phi) = N_l(\beta_{nl}) r^l e^{-\beta_{nl} r^2} Y_{lm}(\theta,\phi), \tag{5.30}$$

where $Y_{lm}(\theta,\phi)$ is a spherical harmonic following the Condon-Shortey phase convention. The CGTOs and SGTOs are connected via a transformation matrix, the expressions for which are provided, for example, in Schlegel et al. [163]. For s- and p-type orbitals, the CGTOs can be expressed exactly in terms of the SGTOs of the same angular momentum l. For d-orbitals and higher, this is no longer the case. There are a total of 6 CGTOs with angular momentum $l = 2$, but only 5 SGTOs of the same angular momentum. In order to express CGTOs in terms of SGTOs, an additional basis function with s-type symmetry has to be added to the $l = 2$ SGTOs

$$\phi_{n26}(r,\theta,\phi) = N_l(\beta_{nl}) r^2 e^{-\beta_{nl} r^2} Y_{00}(\theta,\phi). \tag{5.31}$$

For $l = 3$, there are 7 SGTOs and 10 CGTOs, and three p-symmetry orbitals need to be added to the SGTOs. In practice, these additional orbitals are not included in most quantum chemistry codes. This is also the case in LibInt [182], which is used for the calculation of overlap integrals in this chapter. Thus, the SGTOs can be expressed in terms of CGTOs, but the reverse is not the case for $l = 2$ and higher. However, SGTOs have a few advantages: 1) There are fewer SGTOs than CGTOs for the same angular momentum l, thus providing a more compact basis set. 2) As shown by Kuang et al. [108], the Fourier Transform of an SGTO is also an SGTO. Provided the SGTOs follow the Condon-Shortley phase convention, the Fourier Transform of the SGTO in Eq. (5.30) is given by

$$\frac{1}{(2\pi)^{3/2}} \int d\mathbf{r} e^{-i\mathbf{G}\cdot\mathbf{r}} N_l(\beta_n) r^l e^{-\beta r^2} Y_{lm}(\hat{\mathbf{r}}) = (-i)^l \tilde{N}_l(\beta_n) G^l e^{-G^2/(4\beta_n)} Y_{lm}(\hat{\mathbf{G}}), \tag{5.32}$$

with $\tilde{N}_l(\beta_n) = N_l(\beta_n)/(2\beta_n)^{3/2}$. Equation (5.32) shows that only the $l = 0$ SGTOs are non-zero at $|\mathbf{G}| = 0$. Thus, for SGTOs, the integration sum rule for the DDRF could also be enforced by removing s-type orbitals from the basis set. SGTOs can also be expressed in terms of real basis functions. The relationship between complex SGTOs and real SGTOs is given by the relationship between complex spherical harmonics Y_{lm} and their real counterparts R_{lm}. The real and complex spherical harmonics are related by the unitary transformation

$$R_{lm}(\theta,\phi) = \begin{cases} \frac{i}{\sqrt{2}}\left(Y_{l-|m|}(\theta,\phi) - (-1)^m Y_{l|m|}(\theta,\phi)\right) \text{ if } m < 0 \\ Y_{lm}(\theta,\phi) \text{ if } m = 0 \\ \frac{1}{\sqrt{2}}\left(Y_{l-|m|}(\theta,\phi) + (-1)^m Y_{l|m|}(\theta,\phi)\right) \text{ if } m > 0. \end{cases} \tag{5.33}$$

In this chapter, the real SGTOs (RSGTOs) are used exclusively. The Fourier transform of the RSGTOs is computed by first evaluating Eq. (5.32), followed by applying the transformation in Eq. (5.33).

5.3.5 Symmetry of the projected density-density response function

The density-density response function is symmetric under the exchange of $\mathbf{r}$ and $\mathbf{r}'$. Looking at Eq. (5.15), we thus require that

$$\chi^{kl}_{\alpha_k\alpha_l} = \chi^{lk}_{\alpha_l\alpha_k}. \tag{5.34}$$

Since $\chi_{\mathbf{GG}'}$ is hermitian, this constraint is automatically satisfied for RSGTOs and CGTOs.

5.4 Application to silicon clusters

This section applies the previously outlined method to the silicon clusters from Chapter 4. First, the properties of the approximate DDRFs are investigated, followed by an assessment of the accuracy in HOMO-LUMO gaps when using the approximate DDRFs in GW calculations. Then, a neural network model is trained to predict the resulting DDRFs.

5.4.1 Computational Details

In order to increase the size of the data set, I perturbed the atomic coordinates of the Si-H clusters used to train the polarizability model in Chapter 4 by a random vector with between 0 and 0.1 Å magnitude. Only clusters with fewer than 60 silicon atoms are included in the data set. Using this approach, I generated 6 randomly perturbed clusters out of each relaxed cluster in the original data set. These perturbed clusters are then used as a training set for a simple neural network model of the DDRF. The DDRF and QP corrections were calculated using the BerkeleyGW software package [90, 49]. Mean-field DFT calculations were performed using the Quantum Espresso code [69, 70]. Norm-conserving pseudopotentials from the Quantum Espresso Pseudopotential Library were used. The parameters of the DFT calculations were the same as those used by Zauchner et al. [196] (see Chapter 4): a plane-wave cut-off of 65 Ry, and a supercell with sufficient vacuum to avoid interactions between periodic images. Also, a plane-wave cut-off of 6 Ry and a truncated Coulomb interaction were used for the calculation of the DDRF. The QP corrections were calculated using the generalized plasmon-pole approximation (GPP) [90], an explicit sum over 1000 Kohn-Sham states and also a static remainder correction [48]. To calculate the HOMO and LUMO energies, the vacuum level was determined by averaging the electrostatic potential over the faces of the supercell. To perform the projections onto the localised basis, auxiliary density fitting basis sets from the admm-series [109] were used. The basis set was obtained from Basis Set Exchange (BSE) [62, 149, 165]. The smallest of these basis sets (admm-1) uses 3 contracted s-type orbitals and 2 contracted p-type orbitals for silicon atoms, and 2 contracted s-type orbitals for hydrogen atoms. Overlap matrices were computed using LibInt [182]. A parallel implementation of the projection of the DDRF described above was developed. The smallest CGTO basis set used results in a total of 2 basis functions per hydrogen atom and 9 basis functions per silicon atom. Since the basis of the DDRF is formed by products of the basis functions in $\mathbf{r}$ and $\mathbf{r}'$, the matrix $\chi^{ij}_{\alpha_i\alpha_j}$ can be constructed out for five different "types" of blocks. A 9x9 block for silicon onsite terms, i.e. where the basis functions in $\mathbf{r}$ and $\mathbf{r}'$ reside on the same atom. A 9x9 block for silicon-silicon cross-site terms, where the basis functions in $\mathbf{r}$ and $\mathbf{r}'$ reside on different atoms. Equivalent 2x2 blocks for hydrogen atoms. And finally, a 9x2 block for silicon-hydrogen cross-site terms. The different types of blocks are shown in Figure 5.1. These five types of blocks form the output of the neural network models developed at the end of this chapter.

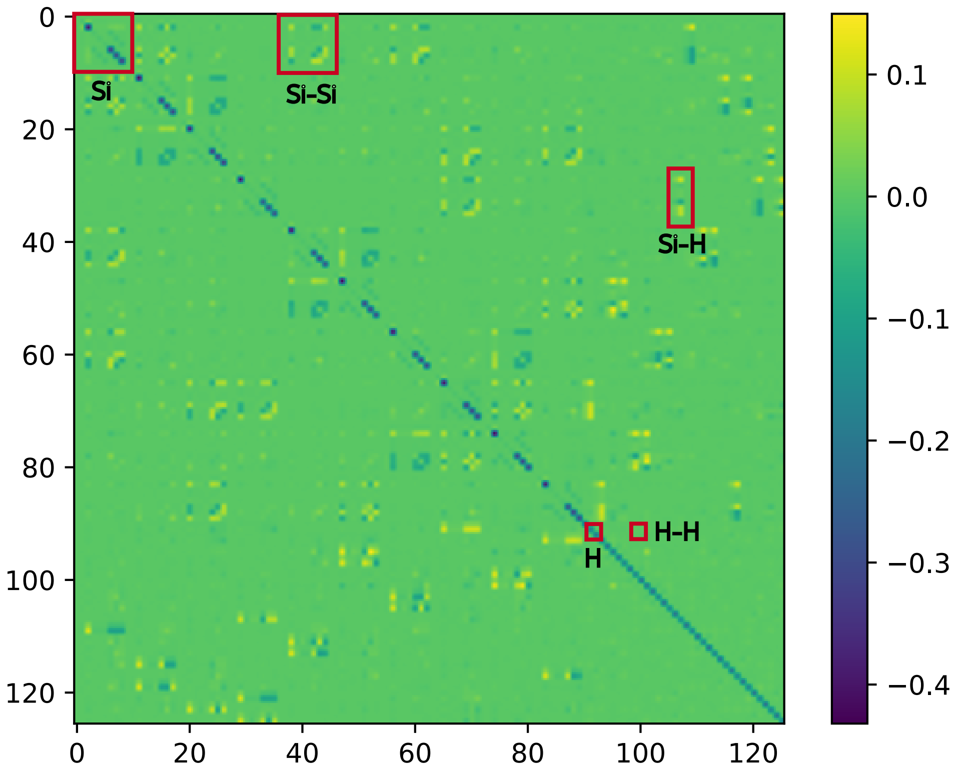


Figure 5.1: Matrix $\chi^{ij}_{\alpha_i\alpha_j}$ for cluster $Si_{10}H_{18}$. The red rectangles highlight the different types of blocks. Si blocks correspond to silicon onsite terms, Si-Si blocks to silicon-silicon crosssite terms, H to hydrogen onsite, H-H to hydrogen-hydrogen crossite terms and Si-H to silicon-hydrogen crosssite terms.

5.4.2 Neural network model

In order to predict the expansion coefficients of the smallest CGTO basis, five dense neural network models - one for each type of block - were constructed and trained using the Tensorflow Python-API [2]. A dense neural network is a non-linear model consisting of several layers which apply a non-linear transformation to the output of the previous layer. All layers between the input and output layer are usually referred to as hidden layers [25]. The output of an N-layer neural network model is given by the composition of the non-linear functions of all layers, i.e.

$$\mathbf{f}(\mathbf{x}) = \mathbf{o}^N(\mathbf{o}^{N-1}(\cdots \mathbf{o}^1(\mathbf{x}))), \tag{5.35}$$

where $\mathbf{x}$ is the input of the model and $\mathbf{o}(\cdot)$ is some non-linear function on the input to the layer. In a dense neural network, the non-linear function consists of a linear transformation

$$z_i^N = \sum_j w_{ij}^N o_j^{N-1} + b_i, \tag{5.36}$$

where w_{ij}^N and b_i^N are learnable parameters, followed a non-linear function h applied to each intermediate output

$$o_i^N = h(z_i^N). \tag{5.37}$$

Neural networks are trained using the backpropagation algorithm [156], which is explained in Appendix B. The five neural network models used to represent each block type use the same neural network structure, with three hidden layers (with 400, 300 and 100 nodes, respectively). The non-linearity used in these models is the Leaky-ReLU function, which has a gradient of 1 for positive values and a gradient which is specified as a hyperparameter for negative values. It should be noted that the Leaky-ReLU activation function has a discontinuity in the first derivative. However, since no derivatives of the DDRF are required in the GW method, the discontinuity should not cause any problems. The models for onsite blocks have a single input - the first 50 principle components (PCs) of the SOAP descriptor of the corresponding atom, as obtained from principal component analysis (PCA). While this is a natural choice for onsite terms, it is not clear how to construct a descriptor for the cross-site terms. Since SOAP is rotationally invariant by design, there is no information about the relative orientation of the two atoms encoded in the descriptor. This led to really poor regression performance for certain types of cross-site terms. In particular, I found that s-p cross-terms (s orbital on one atom and p orbital on the other) and p-p cross-terms are highly dependent on the relative orientation of the two atoms. To provide additional orientational information, the cross-site term models have an additional input: the overlap matrix block corresponding to the two atoms. The overlap matrix provides two additional pieces of information: 1) It contains information about the distance between the two atoms; and 2) Due to the presence of p-orbitals in the basis set, it also contains information about the relative orientation of the two atoms, as the sign of the corresponding overlap matrix element has a strong dependence on the orientation of the p-orbitals. Cross-site blocks thus have three inputs: (1) 50 PCs of the SOAP descriptor of the first atom, (2) 50 PCs of the SOAP descriptor of the second atom and (3) the overlap matrix block corresponding to the two atoms. SOAP descriptors were constructed with $l_{max} = 8$ and $n_{max} = 12$ and $\gamma_\nu = 0.5$ for $r_{cut} = 8.0$ Å using the `QUIPPY` library [15]. In order to enforce the symmetry of the DDRF in the neural network (NN) model, only $\chi^{ij}_{\alpha_i\alpha_j}$ for $j \geq i$ are predicted and the remaining coefficients are obtained by using the symmetry property of the DDRF. Additionally, the coefficients $\chi^{ij}_{\alpha_i\alpha_j}$ were scaled to be between -1 and 1 before training. Both the scaling and the addition of the of overlap matrix blocks improved regression performance, in particular, a decrease in terminal training error was observed. The models were trained for 250 epochs on the perturbed versions of the clusters. The relaxed clusters were retained as a test set. A learning rate of 0.001 was used. During training dropout of 0.05 was used in each layer to prevent overfitting. The Leaky-ReLu activation function has an additional parameter defining the slope for negative outputs, which was set to 0.13. These parameters were optimized using hyperopt [24].

5.4.3 Results and Discussion

Analysis of matrix elements

Before evaluating the accuracy of the QP energies in the projected basis, it is worth analysing some of the properties of the expansion coefficients. The following analysis shows some of the properties of the expansion coefficients in the admm-1 and admm-2 CGTO basis sets. A plot of the average magnitude of matrix elements in the admm-1 and admm-2 CGTO basis sets vs the number of Si atoms in a cluster is shown in Fig. 5.2. Specifically, the coefficients $\chi^{ab}_{\alpha_a\alpha_b}$ were averaged over the onsite-blocks of silicon atoms using

$$\chi_{N_{Si}} = \sum_{a=0}^{N_{Si}-1} \sum_{\alpha_a\beta_a} \left|\chi^{aa}_{\alpha_a\beta_a}\right| / N_{Si}, \tag{5.38}$$

for each cluster. The resulting average values $\chi_{N_{Si}}$ were then divided by the maximum value of $\chi_{N_{Si}}$ for a given basis set to bring the averages of the two basis sets to the same scale. The average magnitude of the onsite terms in the admm-1 basis set appears to drop off by around 20% as the number of silicon atoms in a cluster increases. The steepest drop-off in magnitude occurs up to 40 silicon atoms. In the admm-2 basis set, there is no systematic drop-off but large variations in $\chi_{N_{Si}}$. Significant scatter in the average magnitude can be seen, followed by a distinctive drop-off between 35 and 40 Si atoms and another increase between 45 and 60 Si atoms. The variation in the average magnitude of the expansion coefficients indicates that there is a global dependence on the entire cluster. Thus, it is expected that a descriptor that only captures the first neighbours is not sufficient for learning the coefficients $\chi^{ab}_{\alpha_a\alpha_b}$.

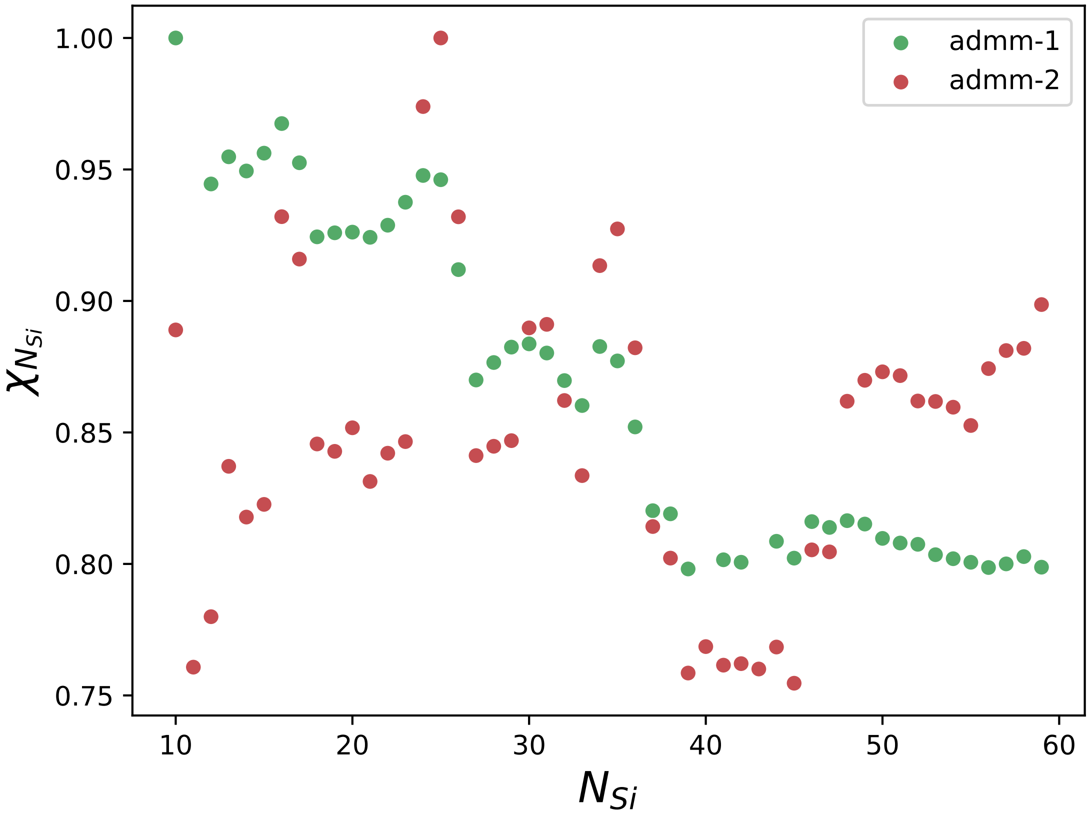


Figure 5.2: Mean of the magnitude of onsite Silicon blocks against the number of Silicon atoms in the cluster in the admm-1 and admm-2 CGTO basis. The average magnitude appears to drop off until $N_{Si} = 60$, beyond which it remains roughly constant.

Decay of cross terms as a function of distance

The cross-site terms are the expansion coefficients $\chi^{ab}_{\alpha_a\alpha_b}$ corresponding to basis functions $\phi^a_{\alpha_a}(\mathbf{r}-\mathbf{r_a})$ and $\phi^b_{\alpha_b}(\mathbf{r}'-\mathbf{r_b})$, where $\mathbf{r_a} \neq \mathbf{r_b}$. In order to investigate the non-locality of the DDRF, we average the magnitude of the cross-site terms by sorting them into bins depending on the distance $|\mathbf{r_a} - \mathbf{r_b}|$: for example, atoms with $|\mathbf{r_a} - \mathbf{r_b}| \geq 2.5$ Å and $|\mathbf{r_a} - \mathbf{r_b}| < 3.5$ Å are placed into the bin at 3 Å. The average of the absolute value of the coefficients in each bin is then calculated, and the results are again divided by the maximum in order to bring the admm-1 and admm-2 results to the same scale. The result of this average χ_r is plotted in Fig. 5.3. Fig. 5.3 only includes silicon-silicon cross-site terms, as these are the most numerous and largest in magnitude; however, a similar trend can be seen in the hydrogen-hydrogen and silicon-hydrogen terms. Clearly, the magnitude of the matrix elements drops off sharply and is negligible beyond a distance of around $8 - 10$ Å. The decrease in χ_r is slower when the admm-2 basis set is used; however, a drop-off around 90% at a distance of 8 Å can still be observed. The decay of the cross-site terms indicates electronic localisation around a site - the density response to a perturbing potential at point $\mathbf{r_a}$ only extends to the atoms surrounding the point at which the perturbation is applied. This effect has been studied in detail by Mussard et al. [134], where it was found that the charge density response is non-zero where the exchange-correlation hole is non-zero. The consequence of this observation is that the charge density response to a perturbing potential typically only extends a few atoms away from the perturbation in semiconductors and insulators, where charge is localised [134].

Accuracy of QP energy levels

The QP HOMO-LUMO gaps (E_g) were calculated at the mean-field (PBE-DFT) level, at the GW level and using the projected DDRF in the admm-series CGTO basis sets. Fig. 5.4 shows how the results with the projected matrix compare to the DDRF and the mean-field (DFT) result. The HOMO-LUMO gap was only calculated for the unperturbed versions of the clusters. We can see that the bulk band gap of silicon (obtained from GW-RPA calculations by Hybertsen et al. [90]) is significantly smaller than the HOMO-LUMO gaps of the hydrogenated silicon clusters. An interesting feature is the drop-off in HOMO-LUMO gap over the range of cluster sizes. At the mean-field level, the HOMO-LUMO gap decreases by ∼1.8 eV from 10 to 60 silicon atoms, while at the GW+RPA level, this decrease is more pronounced ∼3.1 eV. The decrease in HOMO-LUMO gap with cluster size is well documented and has also been found by Degoli et al. [46] or Wippermann et al. [194]. The decrease in HOMO-LUMO gap is largely

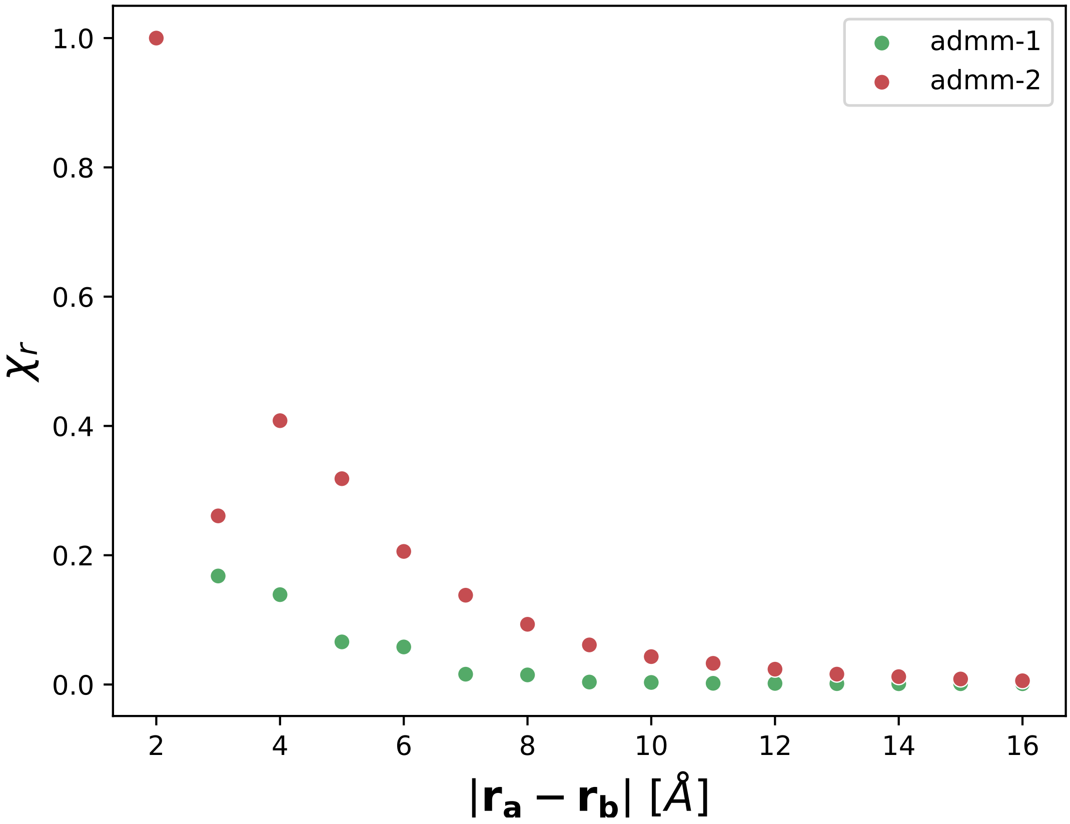


Figure 5.3: Average magnitude of Silicon-Silicon cross-site terms vs. distance between basis function centres using the admm-1 and admm-2 CGTO basis. Clearly, the cross-site terms drop off sharply with the distance between sites, the density response is localised around a site.

attributed to quantum-confinement effects, which become less significant as the cluster size increases. Moreover, Wippermann [194] found that the GW QP-corrections to the mean-field HOMO-LUMO gap were comparable to the corrections shown here (between 3 eV and 4 eV). For the smallest basis set (admm-1), the projected matrix yields errors in the HOMO-LUMO gap of up to 1 eV. This is still an improvement over DFT-PBE, even though a minimal basis set is used to compute the projections. The QP corrections are significantly improved when the larger admm-2 basis set is used for the projected DDRF. An exception to this improvement is observed for the clusters with 15 and 16 silicon atoms. Here, the projection fails to yield accurate quasiparticle energies and significantly underestimates the HOMO-LUMO gap. Further, minor improvements are achieved with the even larger aug-admm-2 basis set. Moreover, the large error for the two aforementioned clusters is not observed with this basis set. Interestingly for the admm-1 basis set, the difference to the plane-wave G_0W_0 increases as the cluster size increases. In contrast, for the two larger basis sets, the error remains constant over the entire range of clusters. Several calculations with the approximate CGTO basis sets also failed due to numerical instabilities, as can be seen in the missing points in Fig. 5.4. In Fig. 5.5, different variants of

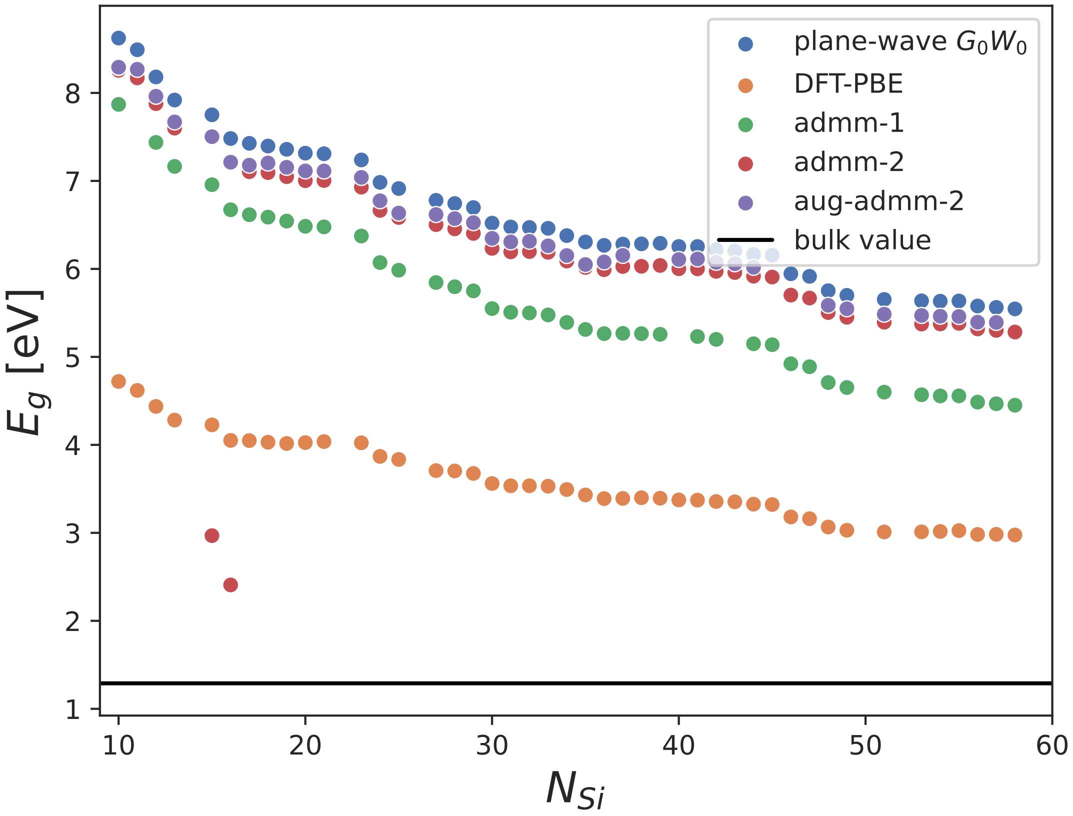


Figure 5.4: HOMO-LUMO gap against number of silicon atoms in the cluster for plane-wave G_0W_0 and G_0W_0 with projected DDRF using different CGTO basis sets. RPA bulk value obtained from Hybertsen et al. [90].

the admm-2 basis set are compared. In orange, the CGTO admm-2 basis is shown. The other two admm-2 variants use an SGTO basis. For the SGTO basis, including s-orbitals, one crucial modification was made: the contracted GTOs, which are basis functions consisting of a linear combination of GTOs, were separated into individual basis functions. This additional flexibility in the basis yields slightly more accurate HOMO-LUMO gaps than the CGTO basis, even though SGTOs generally contain fewer basis functions for the same angular momentum l. In this SGTO basis, the integration sum rule of the DDRF was enforced as a post-processing step in the same way as for the CGTOs. In red, the SGTO admm-2 basis with removed s-orbitals is shown. This modification eliminates the need for the post-processing step, as discussed in Section 5.3.4. The accuracy in HOMO-LUMO gaps is slightly lower than for the CGTO basis. The removal of the s-orbitals will play an essential role in the next chapter, where I will discuss why despite the slight decrease in accuracy, the removal of s-orbitals is a worthwhile trade-off. Even for the SGTO basis without s-orbitals, the difference in HOMO-LUMO gap to the plane-wave G_0W_0 result is consistently within ~ 0.4 eV. This result indicates that the plane-wave DDRF can indeed be compactly represented using auxiliary basis sets, thereby reducing the size of the DDRF from several Gigabytes to a few Megabytes, without losing too much accuracy. It should also be noted that the numerical instabilities encountered with the CGTO basis sets were not observed with the SGTO basis sets, making the SGTO basis the choice for the calculations carried out in the next chapter. To gain more insight into the effect of the projection on the dielectric matrix, Fig. 5.6 shows the real part of the first 100 rows and columns of the dielectric matrix of the cluster $Si_{10}H_{18}$. The LHS heatmap shows the RPA dielectric matrix as used in the plane-wave G_0W_0 calculations and the RHS shows the dielectric matrix obtained from the Fourier transform of the DDRF in the approximate SGTO basis (with no s-orbitals). On both heatmaps, similar patterns can be observed. Most notably positive values close to 1.0 along the diagonal. Moreover, the sign of individual elements is mostly equal, which can be seen clearly in between row 20 and row 80 within the first 20 columns. While the signs of the matrix elements are mostly correct, it can be seen that the magnitude is consistently overestimated when using the SGTO basis. The dielectric matrix is directly related to the DDRF, the elements of which are a rough indicator of the strength of the density response. It appears that the stronger density response causes the system to become slightly more metallic, decreasing the HOMO-LUMO gap compared to the plane-wave G_0W_0 HOMO-LUMO gap.

5.4.4 Machine Learning the expansion coefficients

One of the challenges in developing a model to predict these coefficients is the choice of descriptor. Ideally, such a descriptor should transform in the same way as the outputs. In the case of the neural network models representing the different types of blocks of the DDRF, this is not the

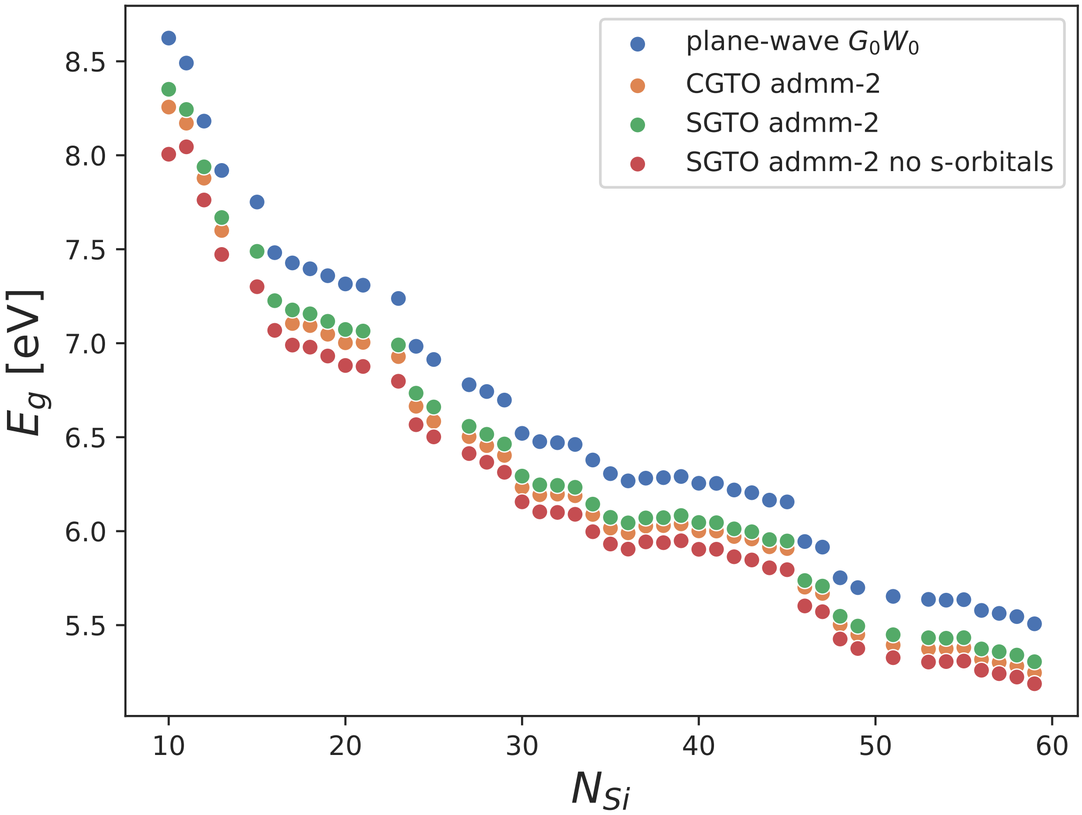


Figure 5.5: HOMO-LUMO gap against number of silicon atoms in the cluster for plane-wave G_0W_0 and G_0W_0 with projected DDRF using different CGTO and SGTO basis sets. The two outliers in the CGTO admm-2 basis were removed to highlight the differences between the SGTO and CGTO basis sets. RPA bulk value obtained from Hybertsen et al. [90].

case. The SOAP descriptor is rotationally invariant [16] and thus does not provide a complete descriptor for the target quantity $\chi^{ab}_{\alpha_a\alpha_b}$. As a consequence, it was found that both the training and validation error stopped decreasing after around 250 epochs. The models still appear to reliably predict the larger coefficients of the DDRF, as illustrated in Fig. 5.7.

Fig. 5.7 (a) and 5.7 (b) shows the predicted coefficients against the actual coefficients for a small ($Si_{10}H_{18}$) and a large ($Si_{69}H_{68}$) cluster. We can see that for smaller coefficients, there is significantly more scatter than for the larger coefficients. Without pre-scaling the coefficients to be within the range [-1.0, 1,0], this problem becomes worse, as the mean squared error loss function is biased towards accurate predictions on larger coefficients. These smaller coefficients correspond to cross-site terms that have strong orientational dependence discussed before. While using the overlap matrix as an additional input to the model significantly improves the accuracy

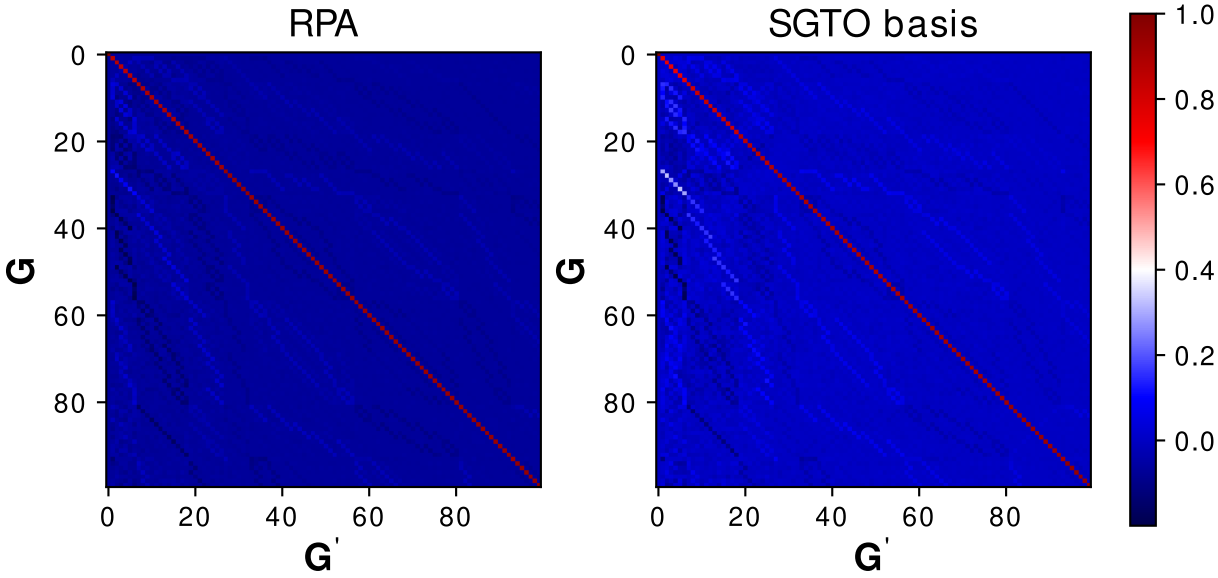
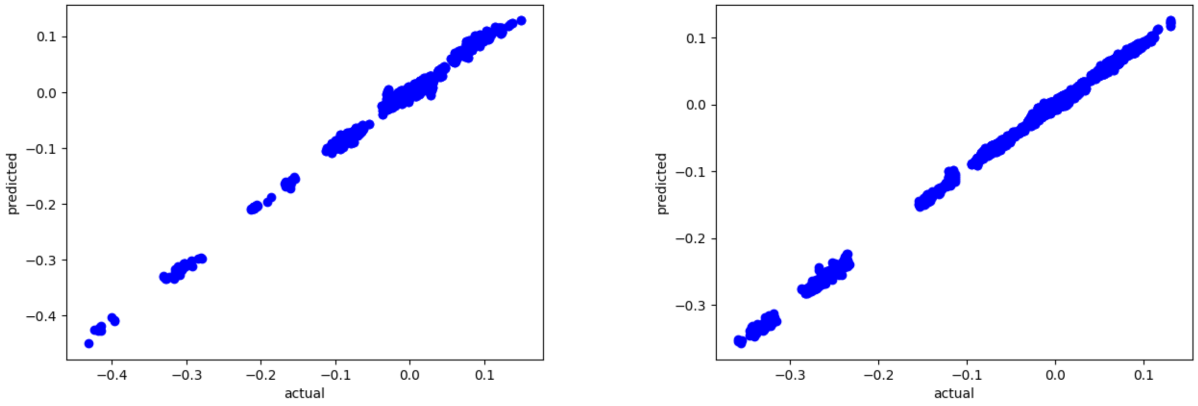


Figure 5.6: First 100 rows and columns of the RPA dielectric matrix as used in plane-wave G_0W_0 and the SGTO dielectric matrix with no s-orbitals. The colour bar indicates the value of the real part of elements of the dielectric matrix.

in predicting these coefficients, it does not entirely mitigate the issue. The HOMO-LUMO gap appears to be very sensitive to these coefficients, so the predicted DDRFs do not yield satisfactory results for the HOMO-LUMO gap. This can be seen in Fig. 5.8 (ML-G_0W_0), where the predicted DDRFs in GW calculations led to extremely unreliable QP corrections and changes in the ordering of the QP energies. In particular, it can be seen that for most silicon clusters, the QP corrections shift the DFT-PBE LUMO energy below the HOMO energy, leading to a negative HOMO-LUMO gap when energies are ordered according to their DFT-PBE energies. Reordering the energies would necessitate the calculation of all states, which is not practical. The high sensitivity of the QP-energies to small differences in the coefficients $\chi^{ab}_{\alpha_a\alpha_b}$ necessitates a more systematic approach for machine-learning the DDRF. This is the subject of the next chapter, where I introduce an atomic decomposition of the DDRF and a corresponding descriptor which encodes the transformation properties of these atomic contributions under rotation of the chemical environment.



(a) $\mathbf{Si_{10}H_{18}}$ **(b)** $\mathbf{Si_{69}H_{68}}$

Figure 5.7: Predicted vs. actual coefficients of χ_{ij}. The coefficients were predicted using the NN models for each block type. The results shown correspond to the relaxed versions of the clusters.

5.5 Conclusion

In this chapter, I have demonstrated how the DDRF can be efficiently approximated using a set of auxiliary basis functions of GTOs. The resulting DDRFs yield HOMO-LUMO gaps within ~ 0.4 eV of a full plane-wave G_0W_0 calculation. While this constitutes significant progress towards the development of an ML-DDRF, further work is needed.

The failure of the ML model introduced in this chapter is largely due to the fact that $\phi^i_{\alpha_i}(\mathbf{r})$ and $\phi^j_{\alpha_j}(\mathbf{r}')$ can be centred on different atoms and thus a model that predicts the corresponding expansion coefficient $\chi^{ij}_{\alpha_i\alpha_j}$ would require a descriptor that depends on the chemical environment of both atoms, as well as their relative distance and orientation. Ideally, such a descriptor should also transform in the same way as the target quantity $\chi^{ij}_{\alpha_i\alpha_j}$ under a rotation of the cluster. Combining these properties into one descriptor is far from trivial, and the ‘naive’ machine learning model and its failure to predict the DDRF with sufficient accuracy prove that a more systematic approach is needed. In the next chapter, I describe how an atomic decomposition of the DDRF is achieved, followed by deriving the transformation properties of these atomic contributions under rotation. The predictable transformation properties allow for constructing a neighbourhood density-based descriptor that encodes these transformation properties and can be used to train a neural network, predicting the expansion coefficients of the atomic DDRFs.

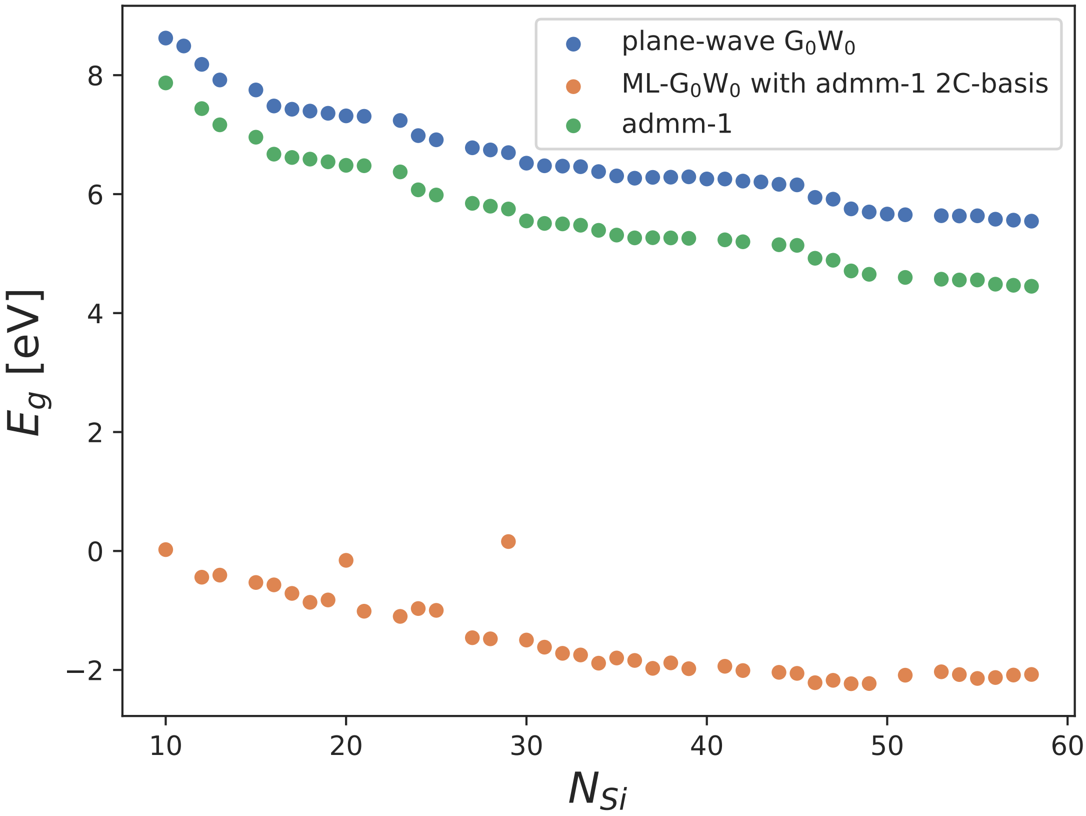


Figure 5.8: HOMO-LUMO gaps obtained using the plane-wave G_0W_0, G_0W_0 using the admm-1 CGTO DDRF and the predicted DDRFs obtained using the neural network model (ML-G_0W_0). The negative HOMO-LUMO gaps are a consequence of the reordering of energy levels caused by the QP corrections.

Chapter 6

Accelerating GW calculations through machine-learned dielectric matrices

6.1 Introduction

Recently machine learning has shown tremendous success in reproducing various scalar properties and may provide a promising path to more flexible model dielectric functions. As discussed in Chapter 5, to make progress on developing a machine-learned model DDRF, a systematic approach is required. The starting point of every machine learning model is the descriptor. Many descriptors used in the fields of computational chemistry are explicitly constructed to be invariant under rotation and translation. For example, ACE [51], SOAP [16], the Coulomb matrix [78, 160], Bag-of-bonds [77] or fingerprint-based descriptors have been shown to be reliable descriptors for the prediction of scalar quantities. As I have shown in Chapter 5, when predicting tensors or functions, however, it is no longer sufficient to rely on a rotationally invariant descriptor. Several approaches aimed at addressing this problem can be found in the literature. For example, Grisafi [73] developed a symmetry-adapted version of the SOAP kernel, which is equivariant under rotations and was successfully used in the prediction of polarizability tensors and first hyperpolarizabilities [192, 73], dipole moments [184] and electronic densities [72].

The prediction of electronic densities is particularly interesting in the context of this work, as, similar to learning the DDRF, the model's output is a function instead of a scalar or tensor. As shown by Grisafi and co-workers [72], the electronic density can be expanded in a basis of spherical harmonics and radial basis functions, yielding a set of coefficients which transform as spherical tensors. Several other groups have also explored machine learning approaches of the electronic density in the past, for example, Brockherde et al. [31], Alred et al. [5] and Chandrasekaran and co-workers [38]. Moreover, the construction of group-equivariant neural networks, such as Clebsch-Gordan networks [105, 106, 6], tensor-field networks [179] and spherical CNNs [41, 40] have seen significant developments in recent years and implementation of these methods has been significantly simplified by frameworks such as `e3NN` [112] developed by Geiger et al., thus providing promising alternatives to symmetry adapted SOAP for the learning of functions. To the best of my knowledge, however, there has been no attempt at developing machine learning models for the prediction of non-local functions, such as the DDRF. In this chapter, I aim to bridge this gap and present a scheme to decompose the DDRF into atomic contributions, which can be learned within an atom-centred framework. The decomposition into atomic contributions is almost a necessity in the context of learning non-local functions: as discussed in Chapter 5, the DDRF can quickly grow to several tens of gigabytes in size when represented on a real-space or plane-wave grid. The immense size of these matrices for larger systems presents a significant computational and storage bottleneck in GW calculations and a challenging machine-learning problem that an atomic decomposition can significantly simplify. In the last chapter, I have shown that the size of the DDRF can be drastically reduced by approximating it in a basis of GTOs. In this chapter, I build on the previous results and show that by decomposing the DDRF into atomic contributions and constructing a new atom-centred descriptor, an accurate machine-learning model of the DDRF can be developed. The resulting DDRFs are then used in GW calculations to obtain QP corrections. This hybrid ML/GW approach is called the ML-GW method. I would further like to point out that while a fully equivariant model would be desirable, extending the previously mentioned equivariant models to non-local functions quickly becomes intractable due to the large number of parameters required and the computationally expensive operations needed for each layer. Alternatively, one could use a purely linear model, which can be easily constructed to be equivariant. However, in doing so, one would lose the immense flexibility of neural networks, primarily attributed to the non-linearities in each layer. As such, I employ a dense neural network to predict the DDRF while ensuring that the descriptor can discern between different orientations of the atomic environment.

6.2 Theory

6.2.1 Atomic decomposition of the density-density response function

To bypass the expensive computation of the DDRF and pave the way towards a machine-learning approach, it is desirable to decompose the DDRF $\chi(\mathbf{r},\mathbf{r}')$ as a sum of atomic contributions $\chi_i(\mathbf{r},\mathbf{r}')$ according to

$$\chi(\mathbf{r},\mathbf{r}') = \sum_{i=1}^{N} \chi_i(\mathbf{r},\mathbf{r}'), \tag{6.1}$$

where i labels atoms and N is the total number of atoms.

How this partitioning is achieved is not immediately obvious. However, one would expect certain properties from these atomic contributions: (1) the function has small contributions far away from the atom; (2) the function retains the global symmetry of χ, e.g.

$$\chi(\mathbf{r},\mathbf{r}') = \chi(\mathbf{r}',\mathbf{r}). \tag{6.2}$$

The (intractable) direct approach

The most obvious solution is to define a purely atom-centred basis. The basis functions for atom i are given by the solid harmonic Gaussians of the form

$$\psi^i_{nlm}(\mathbf{r}) = N(\beta_{nl}) r^l e^{-\beta_{nl} r^2} Y_{lm}(\theta,\phi), \tag{6.3}$$

where $N(\beta_{nl})$ is a normalization constant. Using this basis, we can expand the atomic density-density response function as

$$\chi_i(\mathbf{r},\mathbf{r}') = \sum_{n=1}^{N_{max}}\sum_{l=0}^{l_{max}}\sum_{m=-l}^{l}\sum_{n'=1}^{N_{max}}\sum_{l'=0}^{l_{max}}\sum_{m'=-l'}^{l'} \chi^{(i)}_{nlmn'l'm'}\psi^i_{nlm}(\mathbf{r})\psi^{i*}_{n'l'm'}(\mathbf{r}'). \tag{6.4}$$

Recall that the plane-wave density-density response function is given by

$$\chi(\mathbf{r},\mathbf{r}') = \frac{1}{V}\sum_{\mathbf{G},\mathbf{G}'}\chi_{\mathbf{G},\mathbf{G}'}e^{i\mathbf{G}\cdot\mathbf{r}}e^{-i\mathbf{G}'\cdot\mathbf{r}'}. \tag{6.5}$$

The projection of $\chi(\mathbf{r},\mathbf{r}')$ onto the basis $\psi^i_{nlm}(\mathbf{r})\psi^i_{n'l'm'}(\mathbf{r}')$ can then be computed using the following integral

$$w^{(i)}_{nlmn'l'm'} = \frac{1}{V}\sum_{\mathbf{G},\mathbf{G}'}\chi_{\mathbf{G},\mathbf{G}'}\int d\mathbf{r}\psi^i_{nlm}(\mathbf{r})e^{i\mathbf{G}\cdot\mathbf{r}}\int d\mathbf{r}'\psi^{i*}_{n'l'm'}(\mathbf{r}')e^{-i\mathbf{G}'\cdot\mathbf{r}'} = \sum_{\mathbf{G},\mathbf{G}'}\chi_{\mathbf{G},\mathbf{G}'}\tilde{\psi}^{*i}_{nlm}(\mathbf{G})\tilde{\psi}^i_{n'l'm'}(\mathbf{G}'), \tag{6.6}$$

where $\tilde{\psi}^i_{nlm}(\mathbf{G})$ denotes the Fourier transform of a basis function evaluated at $\mathbf{G}$. However, since the basis functions on different atoms are not orthogonal, the coefficients $w^i_{nlmn'l'm'}$ are different from $\chi^{(i)}_{nlmn'l'm'}$ in Eq. (6.4). To find the relationship between the two coefficients, we first expand the density in the previously described basis set and then multiply by $\psi^j_{nlm}(\mathbf{r})\psi^j_{n'l'm'}(\mathbf{r}')^*$ and integrate over all space

$$\begin{aligned} w^{(j)}_{nlmn'l'm'} = &\sum_i\sum_{n_1l_1m_1n_2l_2m_2}\chi^{(i)}_{n_1l_1m_1n_2l_2m_2}\int\psi^i_{n_1l_1m_1}(\mathbf{r})\psi^{j*}_{nlm}(\mathbf{r})d\mathbf{r}\int\psi^{i*}_{n_2l_2m_2}(\mathbf{r}')\psi^j_{n'l'm'}(\mathbf{r}')d\mathbf{r}' = \\ &\sum_i\sum_{n_1l_1m_1n_2l_2m_2}\chi^{(i)}_{n_1l_1m_1n_2l_2m_2}S^{(i,j)}_{n_1l_1m_1nlm}S^{(j,i)}_{n'l'm'n_2l_2m_2}. \end{aligned} \tag{6.7}$$

By collecting indices nlm together into a combined index k, we arrive at the following expression

$$w^{(j)}_{kk'} = \sum_i\sum_{k_1k_2}S^{(i,j)}_{k_1k}\chi^{(i)}_{k_1k_2}S^{(j,i)}_{k'k_2}. \tag{6.8}$$

Next, we can use the following property

$$S^{(i,j)*}_{nlmn'l'm'} = S^{(j,i)}_{n'l'm'nlm} \longrightarrow S^{(i,j)*}_{kk'} = S^{(j,i)}_{k'k}, \tag{6.9}$$

to obtain,

$$w^{(j)}_{kk'} = \sum_i \sum_{k_1 k_2} S^{(j,i)*}_{kk_1} \chi^{(i)}_{k_1 k_2} S^{(j,i)}_{k'k_2}. \tag{6.10}$$

Now we define

$$S^{(j,i)*}_{kk_1} S^{(j,i)}_{k'k_2} = \tilde{S}^{(i,j)}_{kk'k_1k_2}. \tag{6.11}$$

While this large matrix $\tilde{\mathbf{S}}$ may, in principle, be inverted naively, it makes sense to identify it with the Khatri–Rao product [128] of two matrices denoted $\mathbf{S}^\dagger \circledast \mathbf{S}$,

$$\tilde{\mathbf{S}} = \mathbf{S}^\dagger \circledast \mathbf{S} = \begin{pmatrix} \mathbf{S}^{\dagger(1,1)} \otimes \mathbf{S}^{(1,1)} & \cdots & \mathbf{S}^{\dagger(1,N)} \otimes \mathbf{S}^{(1,N)} \\ \vdots & \ddots & \vdots \\ \mathbf{S}^{\dagger(N,1)} \otimes \mathbf{S}^{(N,1)} & \cdots & \mathbf{S}^{\dagger(N,N)} \otimes \mathbf{S}^{(N,N)} \end{pmatrix}. \tag{6.12}$$

Now using the fact that the inverse of a Kronecker Product is the Kronecker Product of the inverses

$$(A^{-1} \otimes B^{-1})(A \otimes B) = (A^{-1}A) \otimes (B^{-1}B) = I \otimes I = I, \tag{6.13}$$

the inversion can be recursively solved via inversion by partitioning [64], where the matrix is first divided into four sub-matrices. This process recursively repeats on each sub-matrix until the sub-matrices only consist of a single tensor product. Then the inverse tensor product can be built from the tensor product of the inverses.

Thus, to compute the expansion coefficients in Eq. (6.4), one must first calculate the projections in Eq. (6.7), followed by multiplication with the inverse of the above matrix.

While this approach of direct inversion of the overlap matrix is appealing, it is unfortunately intractable for even the most trivial systems. This is because the tensor products in Eq. (6.12) increase drastically in size as the number of basis functions is increased. For example, if a total of 100 basis functions are used in $\mathbf{r}$ and $\mathbf{r}'$ respectively, each of the tensor-products in Eq. (6.12) corresponds to a 10000×10000 matrix. For a 10-atom system, the total overlap matrix would be around 160GB if double-precision complex floating point numbers are used. As such, an alternative approach through partitioning the DDRF into atomic contributions before the projection is required.

Divide and conquer

This can be achieved, for example, without approximation, by defining a symmetric partitioning function $p_i(\mathbf{r}, \mathbf{r}')$ in a similar fashion to Hirshfeld's Stockholder partitioning [84], which decays as the distance from the atom increases. Thus, we can define the atomic contributions as

$$\chi_i(\mathbf{r}, \mathbf{r}') = p_i(\mathbf{r}, \mathbf{r}')\chi(\mathbf{r}, \mathbf{r}'), \tag{6.14}$$

requiring that

$$\sum_{i=1}^{N} p_i(\mathbf{r}, \mathbf{r}')\chi(\mathbf{r}, \mathbf{r}') = \chi(\mathbf{r}, \mathbf{r}'). \tag{6.15}$$

Such a partitioning function can be constructed from any symmetric function $\tilde{p}_i(\mathbf{r}, \mathbf{r}')$, by defining

$$p_i(\mathbf{r}, \mathbf{r}') = \frac{\tilde{p}_i(\mathbf{r}, \mathbf{r}')}{\sum_j \tilde{p}_j(\mathbf{r}, \mathbf{r}')}. \tag{6.16}$$

While this approach is appealing, it has two significant disadvantages: 1) integrals and Fourier transforms can no longer be evaluated analytically. 2) It is not apparent how to enforce the integration sum rule of the DDRF. As such, I propose a different approach, leveraging the results from the previous chapter. To extract atomic contributions from the total DDRF, it is convenient to switch to Bra-Ket notation, where we define

$$\chi(\mathbf{r}, \mathbf{r}') = \langle \mathbf{r} | \hat{\chi} | \mathbf{r}' \rangle. \tag{6.17}$$

and a projection operator $\hat{P}_n$ to extract atomic contributions from the DDRF. The projection operator is defined in the real atom-centred basis from the previous chapter $|\phi^a_{\alpha_a}\rangle \otimes \langle \phi^b_{\alpha_b}|$, where a, b index atoms and α_a, α_b index a specific basis function on the respective atoms

$$\hat{P}_n = \frac{1}{2} \sum_{v,\alpha_v} \sum_{w,\alpha_w} (\delta_{nv} + \delta_{wn}) \sum_{k,\alpha_k} \sum_{l,\alpha_l} S^{-1,wl}_{\alpha_w \alpha_l} S^{-1,vk}_{\alpha_v \alpha_k} |\phi^w_{\alpha_w}\rangle \otimes |\phi^k_{\alpha_k}\rangle \otimes \langle \phi^v_{\alpha_v}| \otimes \langle \phi^l_{\alpha_l}|, \tag{6.18}$$

where $S^{-1,vk}_{\alpha_v\alpha_k}$ are the elements of the inverse overlap matrix of the basis set, and the factor $\frac{1}{2}$ is introduced to avoid double counting. The overlap matrix of the basis set is defined as

$$S^{ij}_{\alpha_i\alpha_j} = \int d\mathbf{r}\, \phi^{i*}_{\alpha_i}(\mathbf{r})\phi^{j}_{\alpha_j}(\mathbf{r}) = \langle \phi^{i}_{\alpha_i}|\phi^{j}_{\alpha_j}\rangle. \tag{6.19}$$

Starting from an expansion of $\hat{\chi}$ in the intermediate real atom-centred basis as

$$\hat{\chi} = \sum_{a,\alpha_a}\sum_{b,\alpha_b} \chi^{ab}_{\alpha_a\alpha_b} |\phi^{a}_{\alpha_a}\rangle \otimes \langle \phi^{b}_{\alpha_b}|, \tag{6.20}$$

where $\chi^{ab}_{\alpha_a\alpha_b}$ is a symmetric matrix to ensure the symmetry of $\chi(\mathbf{r},\mathbf{r}')$ under exchange of $\mathbf{r}$ and $\mathbf{r}'$, the action of the operator on $\hat{\chi}$ is given by the contraction

$$\hat{P}_n\hat{\chi} = \frac{1}{2}\sum_{v,\alpha_v}\sum_{w,\alpha_w}(\delta_{nv}+\delta_{wn})\sum_{k,\alpha_k}\sum_{l,\alpha_l} S^{-1,wl}_{\alpha_w\alpha_l} S^{-1,vk}_{\alpha_v\alpha_k}\sum_{a,\alpha_a}\sum_{b,\alpha_b}\chi^{\alpha_a\alpha_b}_{ab}\langle\phi^{b}_{\alpha_b}|\phi^{k}_{\alpha_k}\rangle\langle\phi^{l}_{\alpha_l}|\phi^{a}_{\alpha_a}\rangle|\phi^{w}_{\alpha_w}\rangle\otimes\langle\phi^{v}_{\alpha_v}|. \tag{6.21}$$

After some algebra, we arrive at the definition of the atomic contribution to the DDRF given by (notice the renaming of the remaining summation index over atoms)

$$\hat{\chi}_n = \frac{1}{2}\sum_{\alpha_n}\sum_{w,\alpha_w}\left[\left(\chi^{nw}_{\alpha_n\alpha_w}|\phi^{n}_{\alpha_n}\rangle\otimes\langle\phi^{w}_{\alpha_w}|\right) + \left(\chi^{wn}_{\alpha_w\alpha_n}|\phi^{w}_{\alpha_w}\rangle\otimes\langle\phi^{n}_{\alpha_n}|\right)\right]. \tag{6.22}$$

In real space, these atomic contributions can be written as

$$\begin{aligned}\chi_n(\mathbf{r},\mathbf{r}') &= \langle\mathbf{r}|\hat{\chi}_n|\mathbf{r}'\rangle \\ &= \frac{1}{2}\sum_{\alpha_n}\sum_{w,\alpha_w}\left[\left(\chi^{nw}_{\alpha_n\alpha_w}\langle\mathbf{r}|\phi^{n}_{\alpha_n}\rangle\otimes\langle\phi^{w}_{\alpha_w}|\mathbf{r}'\rangle\right) + \left(\chi^{wn}_{\alpha_w\alpha_n}\langle\mathbf{r}|\phi^{w}_{\alpha_w}\rangle\otimes\langle\phi^{n}_{\alpha_n}|\mathbf{r}'\rangle\right)\right] \\ &= \frac{1}{2}\sum_{\alpha_n}\sum_{w,\alpha_w}\left[\left(\chi^{nw}_{\alpha_n\alpha_w}\phi^{n}_{\alpha_n}(\mathbf{r})\phi^{w*}_{\alpha_w}(\mathbf{r}')\right) + \left(\chi^{wn}_{\alpha_w\alpha_n}\phi^{w}_{\alpha_w}(\mathbf{r})\phi^{n*}_{\alpha_n}(\mathbf{r}')\right)\right]. \end{aligned} \tag{6.23}$$

Since the basis functions are real, i.e. $\phi^{w*}_{\alpha_w}(\mathbf{r}) = \phi^{w}_{\alpha_w}(\mathbf{r})$, it can be easily verified using the symmetry of $\chi^{nw}_{\alpha_n\alpha_w}$, that $\chi_n(\mathbf{r},\mathbf{r}')$ inherits the symmetry of the total DDRF. The locality of $\chi_n(\mathbf{r},\mathbf{r}')$ is less trivial and depends on the decay of the expansion coefficients $\chi^{nw}_{\alpha_n\alpha_w}$ as the distance between atom n and atom w increases. In practice, this property is also inherited from the total DDRF of the silicon clusters, which I discussed in Chapter 5. I will refer to the DDRF in this basis as the two-centre (2C) DDRF. Fig. 6.1 shows the coefficients $\chi^{ab}_{\alpha_a\alpha_b}$ for the silicon cluster $Si_{10}H_{18}$, with the red boxes indicating the coefficients that are included in the atomic contribution of the second silicon atom. We can clearly see the origin of the factor of 1/2 in the atomic contributions: due to the overlap of the red boxes, each coefficient is used exactly twice, thus it is necessary to divide the atomic contributions by a factor of 2 in order to avoid double counting. I want to stress that this atomic representation of the DDRF is exact, i.e. $\sum_i \chi_i(\mathbf{r}, \mathbf{r}')$ reproduces the full interacting DDRF when the local basis set is complete. However, the atomic contributions to the DDRF contain contributions from pairs of basis functions which are centred on different atoms: see Eq. (6.23). These contributions are difficult to learn using atom-centred descriptors. To make progress, I exploit the localization of $\chi_i(\mathbf{r}, \mathbf{r}')$ and expand it

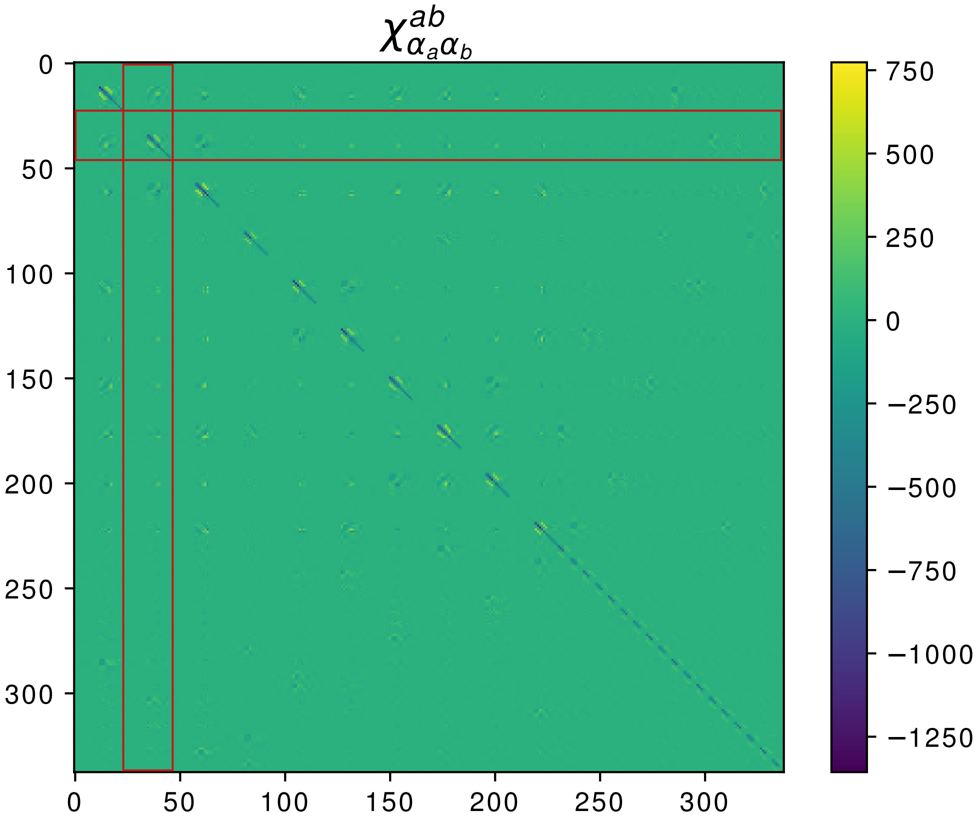


Figure 6.1: Coefficients of the 2C-DDRF in the admm-2 SGTO basis without s-orbitals for the cluster $Si_{10}H_{18}$. The red boxes indicate the coefficients included in the atomic contribution of the second Si atom.

in terms of a set of basis functions $\psi^i_{nlm}(\mathbf{r}) = Y_{lm}(\hat{\mathbf{r}})R_n(|\mathbf{r}|)$ (with Y_{lm} denoting the spherical harmonics and R_n a set of radial functions) which are all centered on atom i according to

$$\chi_i(\mathbf{r},\mathbf{r}') = \sum_{nlm}\sum_{n'l'm'} \hat{\chi}^{(i)}_{nlmn'l'm'}\hat{\psi}^i_{nlm}(\mathbf{r})\hat{\psi}^i_{n'l'm'}(\mathbf{r}') \tag{6.24}$$

with $\chi^{(i)}_{nlmn'l'm'}$ denoting the expansion coefficients given by

$$\hat{\chi}^{(i)}_{nlmn'l'm'} = \int\int d\mathbf{r}d\mathbf{r}'\chi_i(\mathbf{r},\mathbf{r}')\hat{\psi}^i_{nlm}(\mathbf{r})\hat{\psi}^i_{n'l'm'}(\mathbf{r}'), \tag{6.25}$$

where $\hat{\psi}^i_{nlm}(\mathbf{r})$ are again an orthogonal basis built from the overlap matrix of the orbitals $\psi^i_{nlm}(\mathbf{r})$. These coefficients can be learned using a neural network based on atom-centred descriptors. I refer to the representation of the DDRF in the basis $\{\psi^i_{nlm}(\mathbf{r})\}$ as 1-centre DDRF (1C-DDRF) because it only contains pairs of basis functions centred on the same atom.

Integral evaluation

We start from the DDRF in the atom-centred 2C-basis

$$\chi(\mathbf{r},\mathbf{r}') = \frac{1}{2}\sum_{\alpha_n}\sum_{w,\alpha_w}\left[\left(\chi^{nw}_{\alpha_n\alpha_w}\phi^n_{\alpha_n}(\mathbf{r})\phi^{w*}_{\alpha_w}(\mathbf{r}')\right) + \left(\chi^{wn}_{\alpha_w\alpha_n}\phi^w_{\alpha_w}(\mathbf{r})\phi^{n*}_{\alpha_n}(\mathbf{r}')\right)\right]. \tag{6.26}$$

Next, we can exploit the rapid decay of the cross terms in $\chi^{nw}_{\alpha_n\alpha_w}$, with respect to the distance between the centres of basis functions by evaluating Eq. (6.25). These basis functions are also given by the solid harmonic Gaussians, however, a larger number of radial basis functions and larger angular momentum cut-off will be used in order to accurately approximate the cross-terms. From this point onwards, I will drop the complex conjugates, since we will be using real solid harmonic Gaussians. If the onsite basis functions are orthogonalised appropriately (e.g. by choosing the eigenfunctions of the onsite overlap matrix, denoted $\hat{\psi}^i_{nlm}(\mathbf{r})$), we can compute the projection of $\chi^{(i)}(\mathbf{r},\mathbf{r}')$ for each atom individually

$$\hat{\chi}^{(k)}_{n_1l_1m_1n_2l_2m_2} = \frac{1}{2}\sum_j \sum_{\alpha_i\alpha_j} \Bigg(\chi^{ij}_{\alpha_i\alpha_j} \int d\mathbf{r} \phi^i_{\alpha_i}(\mathbf{r}) \hat{\psi}^k_{n_1l_1m_1}(\mathbf{r}) \int d\mathbf{r}' \phi^j_{\alpha_j}(\mathbf{r}') \hat{\psi}^k_{n_2l_2m_2}(\mathbf{r}') + \chi^{ji}_{\alpha_j\alpha_i} \int d\mathbf{r} \hat{\psi}^k_{n_1l_1m_1}(\mathbf{r}) \phi^j_{\alpha_j}(\mathbf{r}) \int d\mathbf{r}' \hat{\psi}^k_{n_2l_2m_2}(\mathbf{r}') \phi^i_{\alpha_i}(\mathbf{r}') \Bigg),$$

which we can simplify to

$$\hat{\chi}^{(k)}_{n_1l_1m_1n_2l_2m_2} = \frac{1}{2}\sum_j \sum_{\alpha_i\alpha_j} \left(\chi^{ij}_{\alpha_i\alpha_j} O^{k,i}_{n_1l_1m_1\alpha_i} O^{k,j}_{n_2l_2m_2\alpha_j} + \chi^{ji}_{\alpha_j\alpha_i} O^{k,j}_{n_1l_1m_1\alpha_j} O^{k,i}_{n_2l_2m_2\alpha_i} \right), \tag{6.27}$$

where $O^{k,i}_{n_2l_2m_2\alpha_i}$ denotes the overlap between basis functions of the 2C and 1C-basis:

$$O^{k,i}_{nl_2m_2\alpha_i} = \int \hat{\psi}^k_{n_2l_2m_2}(\mathbf{r}) \phi^i_{\alpha_i}(\mathbf{r}) d\mathbf{r}. \tag{6.28}$$

The coefficients $\hat{\chi}^{(k)}_{n_1l_1m_1n_2l_2m_2}$ can then be transformed from the orthogonalized basis to the atom-centred basis using a unitary transformation, similar to the 2C-basis:

$$\chi^{(k)}_{nlmn'l'm'} = \sum_{n_1l_1m_1n_2l_2m_2} A_{nlmn_1l_1m_1} A_{n_2l_2m_2n'l'm'} \hat{\chi}^{(k)}_{n_1l_1m_1n_2l_2m_2}. \tag{6.29}$$

One drawback of the above partitioning, followed by a projection onto atom-centred basis functions, is that the number of coefficients included in the summation has to be truncated. This leads to a basis set error that will be introduced in both the intermediate basis, as well as the atom-centred basis. However, doing so allows the evaluation of all involved integrals in closed form, e.g. through the Obara-Saika scheme [137] or the method described by Kuang et al. [108]. These integrals are again evaluated using LibInt [182].

As mentioned in Chapter 5, the total DDRF has to integrate to zero. This property is a formal requirement since the density-density response function can be used to compute the density response from a perturbing potential [139]

$$\Delta\rho(\mathbf{r}) = \int \chi(\mathbf{r}, \mathbf{r}') V_{pert}(\mathbf{r}') d\mathbf{r}', \tag{6.30}$$

which has to integrate to zero for any perturbing potential in order to be a true density response. In practice, this property is enforced by removing s-orbitals from the basis set, in the same way as discussed for the 2C-basis in Chapter 5. Specifically, we need to ensure that $\chi_{\mathbf{G0}} = \chi_{\mathbf{0G'}} = 0$, which requires the contribution of basis functions at $\mathbf{G} = 0$ to vanish. The removal of s-orbitals is a sufficient condition for enforcing this property since only the Fourier transforms of s-orbitals has a $\mathbf{G} = 0$ contribution.

6.2.2 Density based descriptor

As previously mentioned, it is no longer sufficient to use a scalar descriptor that is invariant under the action of the rotation group. To alleviate this issue, it is important to use a descriptor which changes under rotation in order to discern between different orientations of the atomic neighbourhood. The starting point for such a descriptor is the neighbourhood density matrix (NDM), a non-local extension of the smooth neighbourhood density employed in the SOAP descriptor [16], which is analogous to a naive density matrix, defined as

$$\rho_i^\eta(\mathbf{r}, \mathbf{r}') = \sum_{k\in\eta} \sum_{l\in\eta} e^{-\alpha r_{ik}^2} e^{-\alpha {r'_{il}}^2}, \tag{6.31}$$

for each species η, where k, l run over atoms in the neighbourhood of atom i, within a certain cut-off radius R_{cut}, α is a hyperparameter, which loosely defines the size of an atom, and r_{ik} is given by $|\mathbf{r} - (\mathbf{r}_k - \mathbf{r}_i)|$. In our scheme, this corresponds to simply creating a different neighbourhood density for each species in the environment. This neighbourhood density matrix is then expanded in a basis of spherical harmonics and radial basis functions,

$$\rho_i^\eta(\mathbf{r}, \mathbf{r}') = \sum_{nlm} \sum_{n'l'm'} \rho_{nlmn'l'm'}^{(i,\eta)} Y_{lm}(\theta, \phi) Y_{l'm'}^*(\theta', \phi') R_n(r) R_{n'}^*(r'), \tag{6.32}$$

where the expansion coefficients $\rho_{nlmn'l'm'}^{(i,\eta)}$ form the descriptor. Next, we explore the transformation properties of the atomic contributions to the density-density response function under the action of a representation of the rotation group $D(\hat{R})$ [98, 155]. The same steps can be followed to show that the coefficients $\rho^{(i,\eta)}_{nlmn'l'm'}$ transform in the same way.

$$D(\hat{R}) \otimes D^{\dagger}(\hat{R})\chi_i(\mathbf{r}, \mathbf{r}') = \chi_i(\hat{R}^{-1}\mathbf{r}, \hat{R}^{-1}\mathbf{r}'). \tag{6.33}$$

The transformation is defined by the tensor product representation $D(\hat{R}) \otimes D^{\dagger}(\hat{R})$ of two SO(3) representations. Given that the descriptor defines a physical system, both $\mathbf{r}$ and $\mathbf{r}'$ rotate simultaneously. Eq. (6.33) yields

$$\begin{aligned}
&\chi_i(\hat{R}^{-1}\mathbf{r}, \hat{R}^{-1}\mathbf{r}') = \\
&\sum_{nlm}\sum_{n'l'm'} \chi^{(i)}_{nlmn'l'm'} R_n(r) R^*_{n'}(r') Y_{lm}(\hat{R}^{-1}\hat{\mathbf{r}}) Y^*_{l'm'}(\hat{R}^{-1}\hat{\mathbf{r}}') = \sum_{\eta}\sum_{nlm}\sum_{n'l'm'} \chi^{(i)}_{nlmn'l'm'} R_n(r) R^*_{n'}(r') \\
&\qquad \sum_{m_1,m_2} D^{l*}_{mm_1}(\hat{R}^{-1}) D^{l'}_{m'm_2}(\hat{R}^{-1}) Y_{l'm_1}(\hat{\mathbf{r}}) Y^*_{l'm_2}(\hat{\mathbf{r}}').
\end{aligned} \tag{6.34}$$

where $\chi^{(i)}_{nlmn'l'm'}$ is an expansion coefficient.

Now, exchanging the summations of m, m' and m_1, m_2 and using the unitarity property of the Wigner-D matrices $D^l(\hat{R})$ [155]

$$D^l(\hat{R}^{-1}) = D^{l\dagger}(\hat{R}), \tag{6.35}$$

we arrive at

$$\chi_i(\hat{R}^{-1}\mathbf{r}, \hat{R}^{-1}\mathbf{r}') = \sum_{nlm}\sum_{n'l'm_1} R_n(r) R^*_{n'}(r') Y_{lm_1}(\hat{\mathbf{r}}) Y^*_{l'm_2}(\hat{\mathbf{r}}') \sum_{m,m'} D^l_{m_1 m}(\hat{R}) D^{l'*}_{m_2 m'}(\hat{R}) \chi^{(i)}_{nlmn'l'm'}. \tag{6.36}$$

Thus, the atomic DDRF transforms under rotation as

$$\chi^{(i)}_{nlm_1n'l'm_2} = \sum_{m,m'} D^l_{m_1 m}(\hat{R}) D^{l'*}_{m_2 m'}(\hat{R}) \chi^{(i)}_{nlmn'l'm'}. \tag{6.37}$$

or in matrix notation

$$\tilde{\chi}^i_{nn'll'} = \mathbf{D}^l(\hat{R}) \chi^i_{nn'll'} \mathbf{D}^{l'\dagger}(\hat{R}). \tag{6.38}$$

where $\tilde{\chi}^{i}_{nn'll'}$ is the expansion coefficient for the rotated system. Equation (6.38) shows that, when expanded in terms of spherical harmonics and radial basis functions, the atomic contributions to the DDRF transform in a well-defined way under rotation of the atomic environment.

Given that the NDM descriptor transforms in exactly the same way, in an equivariant scheme, it may be necessary to use Eq. (6.31) as a descriptor. It is, however, instructive to notice that the expression in Eq. (6.31), can be written as a product of two sums

$$\rho_i^{\eta}(\mathbf{r}, \mathbf{r}') = (\sum_{k\in\eta} e^{-\alpha r_{ik}^2})(\sum_{l\in\eta} e^{-\alpha {r'_{il}}^2}), \tag{6.39}$$

which shows that the non-local NDM is simply the η-block-wise tensor product of two equal local neighbourhood densities. This means that all information about the transformation properties of the non-local descriptor is already present in the local version, which can be seen clearly by writing

$$\rho_i^{\eta}(\mathbf{r}, \mathbf{r}') = \rho_i^{\eta}(\mathbf{r})\rho_i^{\eta}(\mathbf{r}'). \tag{6.40}$$

Similar to the neighbourhood density matrix, $\rho_i^{\eta}(\mathbf{r})$ can be expanded in a basis of spherical harmonics and radial basis functions $R_n(|\mathbf{r}|)$ with coefficients $\rho_{nlm}^{(i,\eta)}$. It follows that

$$\rho_{nlmn'l'm'}^{(i,\eta)} = \rho_{nlm}^{(i,\eta)}\rho_{n'l'm'}^{(i,\eta)}. \tag{6.41}$$

In practice, I did not find any significant benefit to building the tensor product explicitly and found that it is sufficient to use the local version of the descriptor as input to our model, i.e.

$$\rho_i^{\eta}(\mathbf{r}) = \sum_{k\in\eta} e^{-\alpha r_{ik}^2} = \sum_{nlm} \rho_{nlm}^{(i,\eta)} Y_{lm}(\theta, \phi) R_n(r). \tag{6.42}$$

Relationship to SOAP

It is worth noting that while the starting point of the descriptor used here is the same as in SOAP, they are, in fact, distinct. I specifically avoid making the descriptor rotationally invariant, which is a crucial requirement for this method to work. They are, however, closely related. Specifically, the rotationally invariant SOAP descriptor can be computed from the spherical harmonic expansion of the NDM, using

$$d^{\eta,\eta'}_{i,nn'l} = \sqrt{\frac{8\pi^2}{2l+1}} \sum_m \rho^{(i,\eta)}_{nlm} (\rho^{(i,\eta')}_{n'lm})^*. \tag{6.43}$$

6.3 Computational Details

6.3.1 Data generation

The atomic structures of the hydrogenated silicon clusters were obtained in the same way as described by Zauchner et al. [196] (see Chapter 4): starting from the $\mathrm{Si}_{123}\mathrm{H}_{100}$ cluster of the silicon Quantum Dot data set [13], we remove the silicon atom furthest from the centre of the cluster, terminate the dangling bonds with hydrogen atoms and relax the resulting structure using DFT. The process is repeated until only 10 silicon atoms remain. From this set of silicon clusters, only clusters with fewer than 60 silicon atoms were used in the training set for DDRF prediction. From each cluster with less than 60 silicon atoms, we created six additional clusters in which random displacements were added to the atomic positions. The magnitudes of the displacements were drawn from a uniform distribution with a width of 0.1 Å. Finally, calculations were also carried out for clusters with between 60 and 70 silicon atoms. These clusters are not part of the training set but are used to test the extrapolation capacity of the ML approach.

6.3.2 DFT and GW calculations

The parameters for the DFT and GW calculations are the same as those given in Section 5.4.1.

6.3.3 Projection onto intermediate basis

We first use BerkeleyGW to calculate the inverse dielectric matrix $\epsilon^{-1}_{\mathbf{GG}'}$ in a plane-wave basis [49]. From this, we determine the interacting DDRF via

$$\chi_{\mathbf{GG}'} = (\epsilon_{\mathbf{GG}'} - \delta_{\mathbf{GG}'})/v_{\mathbf{G}} \tag{6.44}$$

with $v_{\mathbf{G}}$ being the Fourier transform of the truncated Coulomb interaction.

Next, the DDRF in real space is obtained as

$$\chi(\mathbf{r},\mathbf{r}') = \frac{1}{V}\sum_{\mathbf{G},\mathbf{G}'} e^{i\mathbf{G}\cdot\mathbf{r}}\chi_{\mathbf{G}\mathbf{G}'}e^{-i\mathbf{G}'\cdot\mathbf{r}'}, \tag{6.45}$$

where V is the volume of the supercell. To compute the projection onto the 2C-basis, we follow the steps outlined in the previous chapter, using a basis of RSGTOs, as defined in LibInt [182]

$$\phi_{lm}(r,\theta,\phi) = N_l(\beta)r^l e^{-\beta r^2}R_{lm}(\theta,\phi), \tag{6.46}$$

with the Fourier transform in Eq. (5.32). The basis set used in this work is a modified version of the admm-2 basis set [109] (see Tables 6.2 and 6.1), in which the s-orbitals were removed and contracted Gaussians were uncontracted into individual basis functions.

Table 6.1: Hydrogen basis

l	β [$1/a_0^2$]
1	1.0
1	0.457639

Table 6.2: Silicon basis

l	β [$1/a_0^2$]
1	13.8028
1	59.9261
1	4.34446
1	0.267360
1	0.0765250
2	0.45

6.3.4 Projection onto atomic basis

The fully atom-centred basis set also consists of RSGTOs. The basis set was constructed following the same procedure as in the DScribe library [80], where individual basis functions are given by

$$\psi_{nlm}(r,\theta,\phi) = N_l(\beta_{nl})r^l e^{-\beta_{nl}r^2}R_{lm}(\theta,\phi), \tag{6.47}$$

with the basis set being truncated at a maximum angular momentum l_{max} and a maximum principal quantum number n_{max}. For silicon atoms we use $l_{max} = n_{max} = 4$. For hydrogen atoms we use $l_{max} = n_{max} = 3$.

The exponents β_{nl} are constructed such that the corresponding basis functions decay to zero at a cutoff radius R_n, i.e. $\beta_{nl} = -\ln\left(\frac{T}{R_n^l}\right)/R_n^2$ with $T = 10^{-3}$ Å^l being a threshold parameter. The cutoff radius $R_n = R_i + (R_o - R_i)/n$ lies between an inner radius R_i and an outer radius R_o. For hydrogen atoms, we used $R_i = 0.1$ Å and $R_o = 3.0$ Å and for silicon atoms, we used $R_i = 1.0$ Å and $R_o = 8.0$ Å. Both R_i and R_o were optimized to minimize linear dependencies in the basis set, as such dependencies significantly deteriorate the accuracy of the neural network predictions. A similar observation was made by Grisafi et al. [72] when learning electron densities, although a different approach was taken to remedy this issue in their work. For silicon atoms, I also added the basis functions included in the modified admm-2 basis, which allows for an exact representation of the onsite terms in the 2C-basis.

In order to compute the coefficients of the atomic contributions to the DDRF in the fully atom-centered basis the same procedure as in the intermediate basis was used: the basis was first orthogonalized by computing the eigenvectors of the overlap matrix. Then the atomic DDRFs in the intermediate basis were projected onto the orthogonalized fully-atom centred basis with overlaps between the different basis functions being computed using LibInt [182]. Finally the atomic DDRFs were transformed back to the non-orthogonal basis, producing the desired coefficients $\chi^{(i)}_{nlmn'l'm'}$.

6.3.5 Descriptors

The basis set for the NDM was generated using the same procedure as for the fully atom-centred basis for the DDRF. However, s-orbitals were not removed and the basis functions of the admm-2 basis set were not included. We used $R_i = 1.0$ Å for both hydrogen and silicon atoms and $R_o = 4.0$ Å for hydrogen atoms and $R_o = 9.0$ Å for silicon atoms. The exponents of the Gaussians in Eq. (6.31) were set such that the standard deviation of the Gaussians is 0.5 Å. LibInt [182] was again used to compute the required integrals for the projection.

6.3.6 Neural network

A dense neural network with four hidden layers with 2000, 1500, 1000 and 2000 nodes, respectively, was constructed for both silicon and hydrogen atoms. Each layer uses a Leaky-ReLu activation function with a leak parameter of 0.1. The output layer was further symmetrized by adding its transpose. The loss used was the mean-squared error between the predicted and true expansion coefficients $\chi^{(i)}_{nlmn'l'm'}$ (see section "Loss function selection"). The neural network was trained on the perturbed clusters for 20,000 epochs.

Loss function selection

Three loss functions were explored in order to select the optimal optimization method for our ML-G_0W_0 method. The first loss function, which I call the squared loss, is simply the mean-squared error between the predicted $\overline{\chi}^{(i)}_{nlmn'l'm'}$ and actual coefficients $\chi^{(i)}_{nlmn'l'm'}$, defined as

$$\epsilon_i = \frac{1}{N} \sum_{nlmn'l'm'} \left| \overline{\chi}^{(i)}_{nlmn'l'm'} - \chi^{(i)}_{nlmn'l'm'} \right|^2, \tag{6.48}$$

where N is the total number of coefficients in the 1C-DDRF. The second loss function, which I call the real-space loss is defined as the squared real-space difference between the predicted and actual DDRFs

$$\epsilon_i = \int\int \left| \sum_{nlmn'l'm'} \overline{\chi}^{(i)}_{nlmn'l'm'} \psi^i_{nlm}(\mathbf{r}) \psi^i_{n'l'm'}(\mathbf{r}') - \chi^{(i)}_{nlmn'l'm'} \psi^i_{nlm}(\mathbf{r}) \psi^i_{n'l'm'}(\mathbf{r}') \right|^2 d\mathbf{r} d\mathbf{r}'. \tag{6.49}$$

The third loss function, which I refer to as the 2C-loss, bypasses the projection step of the 2C basis onto the 1C basis and instead directly computes the real-space squared error between the predicted 1C-DDRF and the actual 2C-DDRF. This approach is motivated by the observation of Grisafi et al. [72], who found that direct optimization can help in avoiding numerical instabilities associated with near-linear dependencies in the basis set. The 2C-loss is defined as

$$
\begin{aligned}
\epsilon_i = \int\int \Bigg| \Bigg(\sum_{nlmn'l'm'} \overline{\chi}^{(i)}_{nlmn'l'm'} \psi^{i}_{nlm}(\mathbf{r}) \psi^{i}_{n'l'm'}(\mathbf{r}') \Bigg) - \\
\frac{1}{2}\Bigg[\sum_{\alpha_i} \sum_{j,\alpha_j} \Bigg(\chi^{ij}_{\alpha_i\alpha_j} \phi^{i}_{\alpha_i}(\mathbf{r}) \phi^{j}_{\alpha_j}(\mathbf{r}') \Bigg) + \Bigg(\chi^{ji}_{\alpha_j\alpha_i} \phi^{j}_{\alpha_j}(\mathbf{r}) \phi^{i}_{\alpha_i}(\mathbf{r}') \Bigg) \Bigg] \Bigg|^2 d\mathbf{r} d\mathbf{r}'. \quad (6.50)
\end{aligned}
$$

In contrast to Grisafi and co-workers [72], we evaluate this loss function analytically by computing all resulting SGTO overlaps using LinInt [182]. In order to assess the fidelity of each loss function, we train three equivalent models as described above with the only distinction between the models being the loss function used for optimizing the NN parameters. Each of the models is trained on the unperturbed clusters for which the QP energies are calculated: we examine the training accuracy of the models. The resulting HOMO-LUMO gaps obtained from using each of the model DDRFs in G_0W_0 calculations are shown in Fig. 6.2. In blue, the G_0W_0 QP-energies using the exact projection onto the 1C-basis is shown. We can see that the model using the squared loss (orange) closely follows the HOMO-LUMO gaps and is able to nearly exactly reproduce the HOMO-LUMO gaps. The second-best results are obtained using the real-space loss (green), where HOMO-LUMO gaps are accurately reproduced up to around clusters with 40 Si atoms, beyond which the model DDRF underestimates the HOMO-LUMO gaps. Surprisingly, the 2C-loss (red) results in the worst HOMO-LUMO gaps, initially overestimating the gap by a large margin, then underestimating between clusters with 20 and 40 Si atoms. Finally, in the region beyond 40 silicon atoms, the 2C-loss model again predicts larger HOMO-LUMO gaps than obtained with the exact 1C-DDRF. This result is in stark contrast to the original hypothesis of the direct loss calculation providing improved numerical stability. I attribute this result to the optimization of the NN parameters being stuck in a local minimum. This hypothesis is supported by a closer examination of the DDRFs predicted by each model. In Fig. 6.3 the coefficients $\chi^{(i)}_{nlmn'l'm'}$ of a silicon atom in the cluster $Si_{10}H_{18}$ are shown. We can clearly see that the squared loss closely resembles the exact 1C-DDRF, whereas both the real-space loss and the 2C-loss DDRFs are visually clearly distinct. This is likely a consequence of the gradient descent optimizer getting stuck in a local minimum. Unfortunately, this could not be resolved, even by choosing different optimizers and optimizer parameters. As such, the squared loss was selected for all of the models presented in the following results.

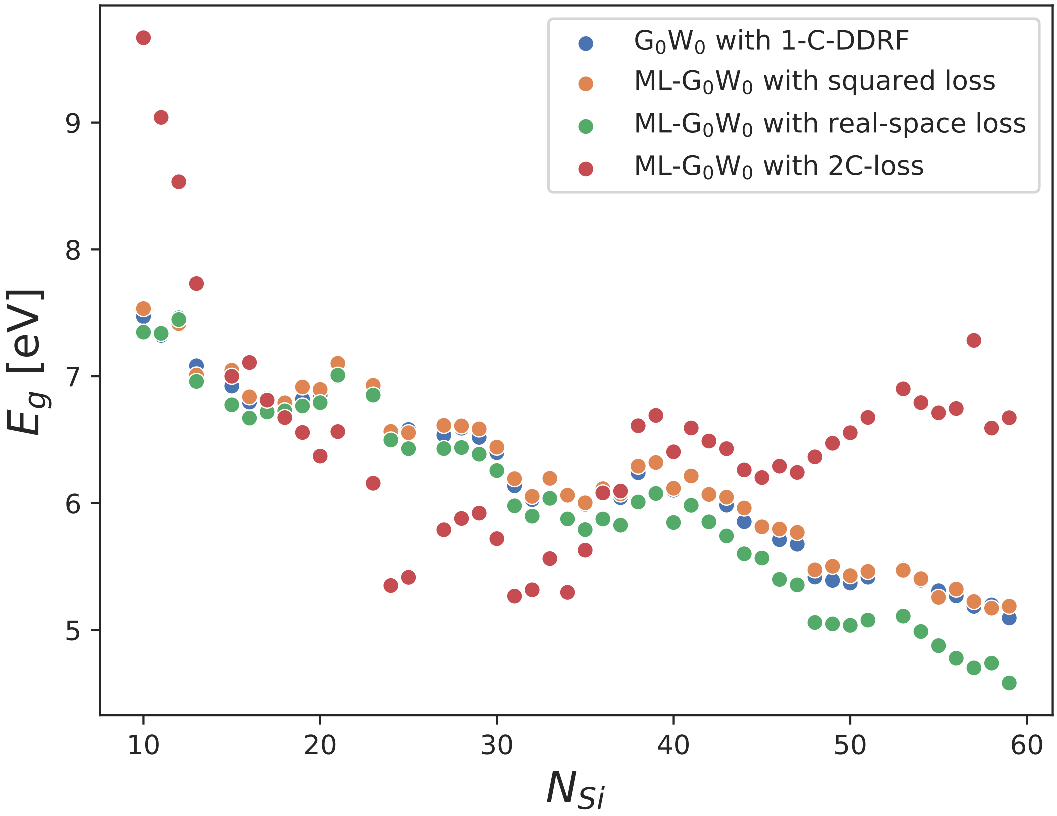


Figure 6.2: G_0W_0 HOMO-LUMO gaps of unperturbed silicon clusters, obtained using the full plane-wave GW-method (blue), and ML-G_0W_0 with different loss functions. The models were trained on the unperturbed cluster, thus the resulting HOMO-LUMO gaps are indicative of the training-set accuracy of the models.

6.4 Results and Discussion

We apply our ML approach for predicting DDRFs to the hydrogenated silicon clusters from the previous chapters and then use the DDRFs to calculate GW quasiparticle energies for these systems. I refer to this technique as the ML-GW approach. The atomic positions of the clusters were constructed as described in the computational methods section and then relaxed using DFT.

To establish the accuracy of this approach, we first investigate the error in the GW quasipar-

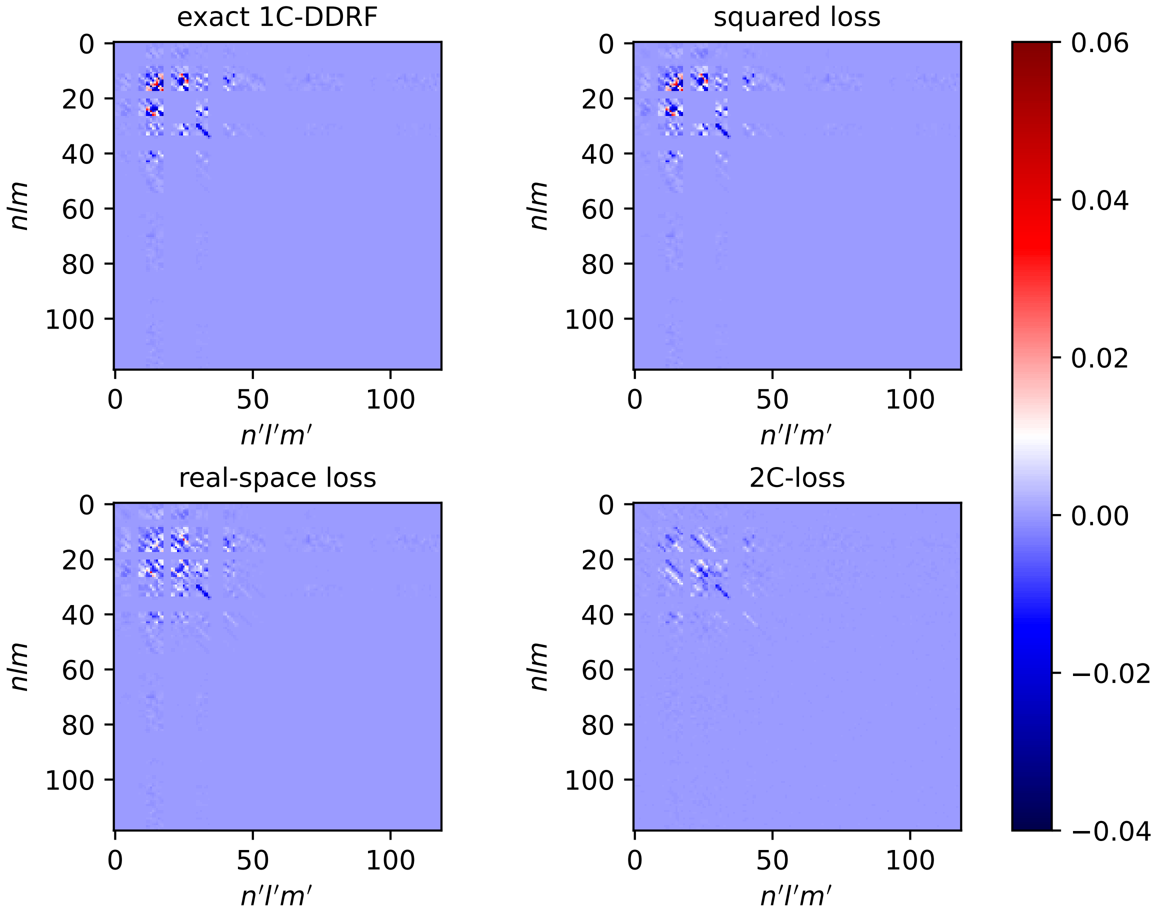


Figure 6.3: Coefficients $\chi^{(i)}_{nlmn'l'm'}$ of a silicon atom in cluster $\mathrm{Si}_{10}\mathrm{H}_{18}$ using the exact 1C-DDRF (top left) and predicted 1C-DDRFs using the squared-loss model (top right), the real-space loss (bottom left) and 2C-loss (bottom right).

ticle energies resulting from the expansion of the DDRF in terms of the intermediate local basis $\{\phi^a_{\alpha_a}(\mathbf{r})\}$: see Eq. (6.22). Fig. 6.4 compares the HOMO-LUMO gaps obtained from mean-field DFT-PBE calculations, a standard plane-wave $\mathrm{G}_0\mathrm{W}_0$ calculation using a generalized plasmon-pole approximation [90, 171] and a $\mathrm{G}_0\mathrm{W}_0$ calculation using the 2C-DDRF, where the DDRF is expanded in terms of a modified version of the admm-2 basis set [109]: see computational methods section. The DFT-PBE results show that the HOMO-LUMO gap decreases with increasing cluster size from $E_g \approx 4.8$ eV for the smallest cluster containing 10 Si atoms to $E_g \approx 3$ eV for the biggest cluster with almost 60 Si atoms. This decrease is a consequence of quantum confinement effects which are less pronounced for bigger clusters. The plane-wave GW HOMO-LUMO gaps show a similar trend as a function of cluster size, but the gaps are larger than the DFT-PBE gaps by several electron volts. Interestingly, the GW corrections are larger for smaller clusters than for larger clusters. As a consequence, the reduction in the GW HOMO-LUMO gaps as a function of cluster size is larger compared to the DFT-PBE result: in particular, the gap is as large as 8.6 eV for the smallest clusters and shrinks to 5.5 eV for the largest clusters corresponding to a decrease of 3.1 eV (compared to a decrease of 1.8 eV in the DFT-PBE HOMO-LUMO gap energies). Similar results were obtained by Chelikowsky et al. [180] who also carried out GW calculations on hydrogenated Si clusters. In particular, they found that the HOMO-LUMO gap shrinks from $\sim$ 9 eV for a 10 Si atom cluster to $\sim$ 6.5 eV for a 47 Si atom cluster. The GW results obtained with the 2C-DDRF are qualitatively similar to the plane-wave GW results. However, the HOMO-LUMO gaps that are obtained with this approach are consistently $\sim$ 0.4 eV smaller than the plane-wave results. This is a consequence of the incompleteness of the local basis set.

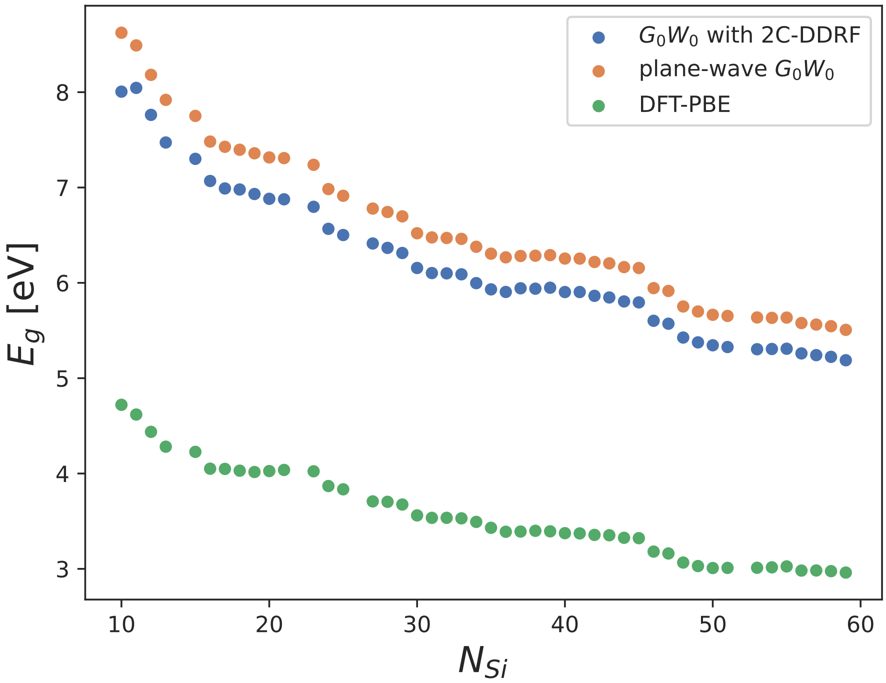


Figure 6.4: HOMO-LUMO gaps of hydrogenated silicon clusters from DFT-PBE Kohn-Sham eigenvalues, plane-wave G_0W_0 and G_0W_0 calculations using the 2C-DDRF, see section "Atomic decomposition of the density-density response function".

Next, we determine the 1C-DDRF. For the basis set we use solid harmonic Gaussians with optimized decay coefficients: see the methods section. Fig. 6.5 (a) compares the HOMO-LUMO gaps from G_0W_0 calculations with the 1C-DDRF to those obtained with the 2C-DDRF and also to plane-wave G_0W_0 results. For small clusters, the HOMO-LUMO gaps obtained with the 1C-DDRF are smaller than those obtained with the 2C-DDRF, while the opposite behaviour is observed for larger clusters. The largest difference between the two methods is obtained for clusters containing ~ 40 Si atoms. The root-mean-square error (RMSE) of the 1C-basis results relative to the 2C-basis results is 0.22 eV and the RMSE relative to the plane-wave results is 0.45 eV for all clusters. Fig. 6.5 (b) shows the HOMO and LUMO quasiparticle energies. It can be seen that better agreement with the plane-wave result is obtained for the LUMO than for the HOMO.

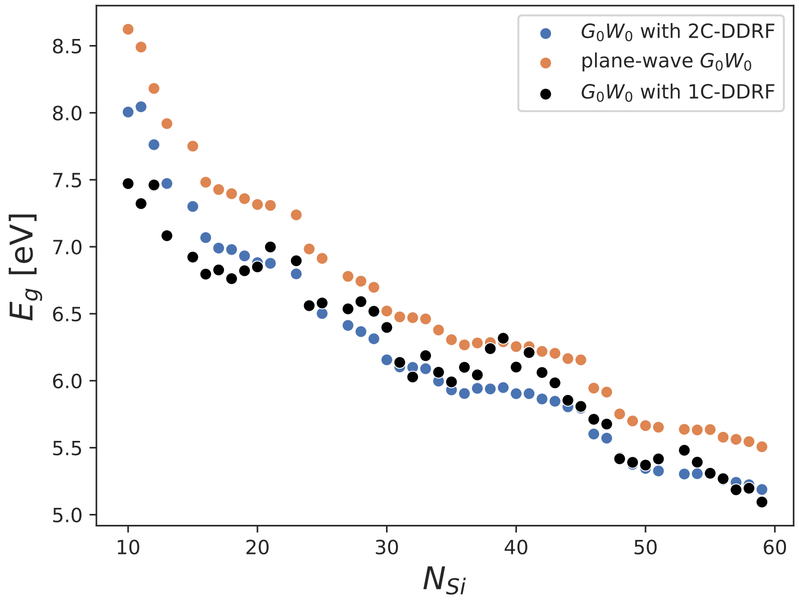
Fig. 6.6 (a) shows the quasiparticle energy corrections of the ten lowest conduction orbitals and the ten highest valence orbitals from plane-wave G_0W_0 and G_0W_0 with the 1C-DDRF. The corrections obtained with the 1C-DDRF follow a similar trend as those obtained from the plane-wave calculation. For the unoccupied states, the quantitative agreement is better than for the occupied states, but the 1C-DDRF results for the unoccupied states are scattered over a larger energy range than the plane-wave results. To analyze the errors that arise from the use of the 1C-DDRF in more detail, Fig. 6.6 (b) shows a two-dimensional histogram of the difference in QP corrections between plane-wave G_0W_0 and G_0W_0 with the 1C-DDRF. For the occupied states the differences are mostly smaller than 0.4 eV, while they are somewhat smaller for the unoccupied states. The RMSE over all energy levels is 0.32 eV.

Now that we have established the accuracy of the method used to generate the training set, we use a dense neural network (NN) in conjunction with NDM descriptor to generate the coefficients of the 1C-DDRF according to

$$\chi^{(i)}_{nlmn'l'm'} = f(\rho^{(i,Si)}_{nlm}, \rho^{(i,H)}_{nlm}), \tag{6.51}$$

where f is the neural network function. The hydrogen and silicon environment descriptors are concatenated into a single vector before being fed into the neural network. A separate network is trained for Si and H contributions to the DDRF. The exact architecture of the network as well as the practical computation of the atomic decomposition and the descriptors, are described in the computational methods section. To generate the training data for the neural network,

(a)



(b)

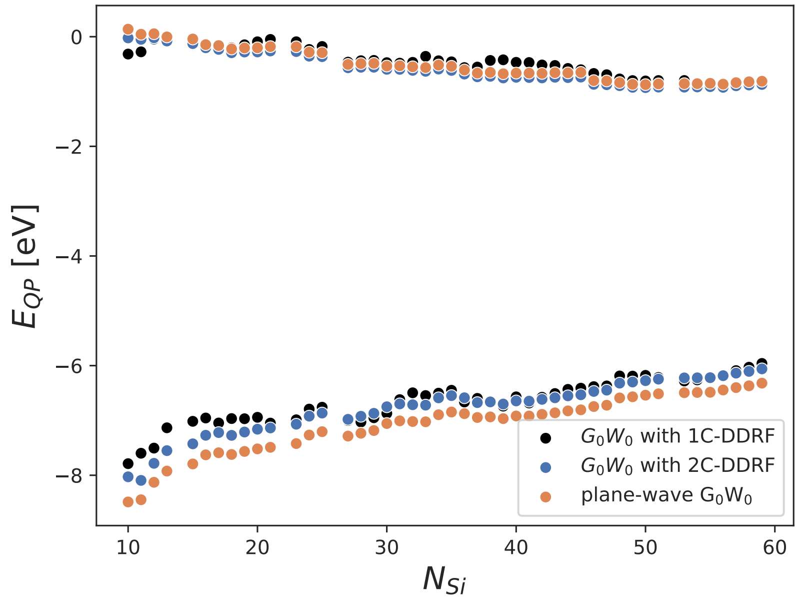


Figure 6.5: (a) HOMO-LUMO gaps of hydrogenated silicon clusters from plane-wave G_0W_0 and G_0W_0 calculations using the 2C-DDRF and G_0W_0 calculations using the 1C-DDRF, see section "Atomic decomposition of the density-density response function". (b) HOMO and LUMO energies of hydrogenated Si clusters.

we start from the set of relaxed hydrogenated Si clusters that were studied above. From each relaxed cluster, we generate six new configurations by randomly displacing the atoms with the magnitude of the displacements being drawn from a uniform distribution with a maximum of

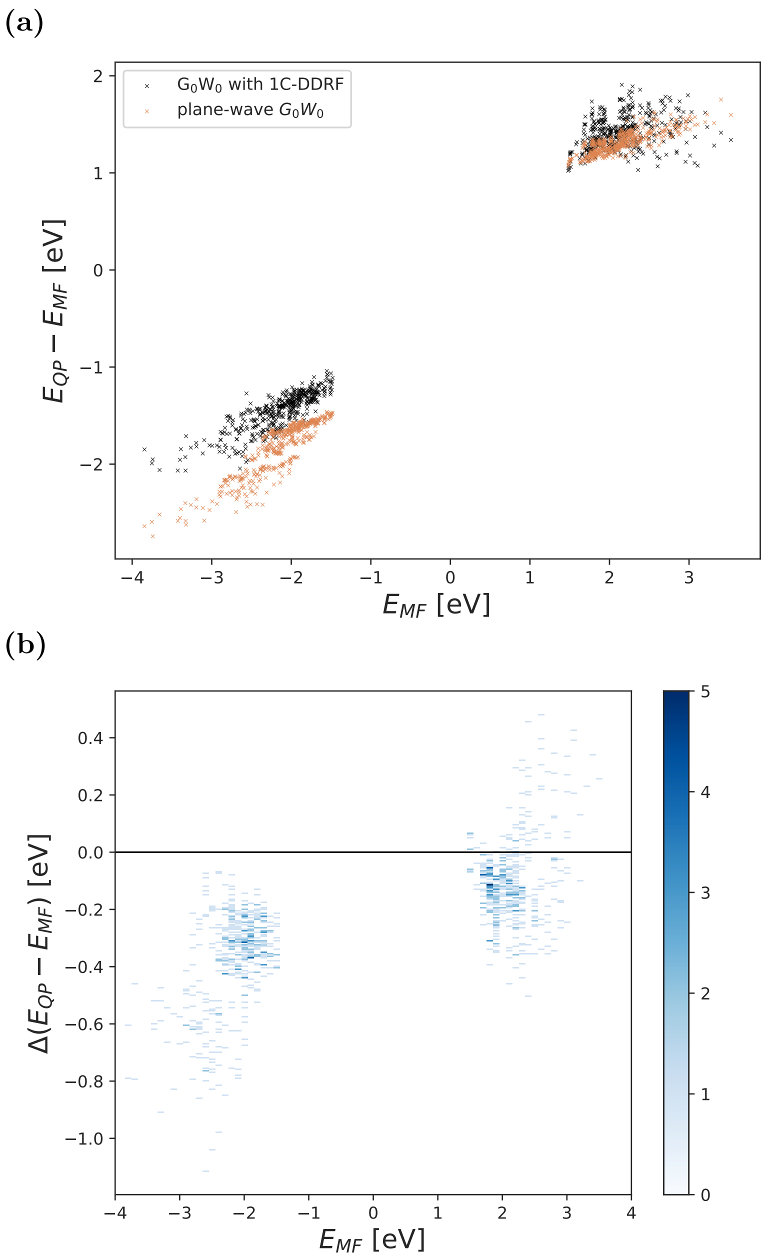


Figure 6.6: (a) Quasiparticle corrections from plane-wave G_0W_0 and G_0W_0 with the 1C-DDRF for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. (b) Histogram of difference in quasiparticle corrections from plane-wave G_0W_0 and G_0W_0 calculations with the 1C-DDRF for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. The mean-field energies are referenced to the middle of the mean-field HOMO-LUMO gap.

0.1 Å. For these clusters, we then calculate the 1C-DDRF.

Once the neural network is trained on the 1C-DDRF of the randomly displaced clusters, we use it to calculate the 1C-DDRFs of the relaxed clusters and then determine quasiparticle energies via the ML-GW approach. Fig. 6.7 compares the HOMO-LUMO gaps from ML-GW and GW with explicitly calculated 1C-DDRFs. Except for the smallest cluster, the ML-GW method accurately reproduces the HOMO-LUMO gaps of the explicit GW calculations. The worse performance for the smallest cluster is a consequence of the training set which contains a large number of bigger clusters containing atomic environments that differ from those found in the smallest clusters. The overall RMSE of the ML-GW method relative to the explicit GW with the 1C-basis is only 0.15 eV but reduces to 0.06 eV when the smallest cluster is excluded.

Fig. 6.8 shows the difference in QP corrections between ML-GW and GW with the 1C-DDRF for the 10 highest valence states and 10 lowest conduction states. The energies of the smallest cluster were excluded from the plot. ML-GW produces QP shifts for both valence and conduction states within 0.1 eV from the explicit G_0W_0 with the 1C-DDRF. The majority of valence states exhibit a positive error, while for conduction states, the error is largely negative.

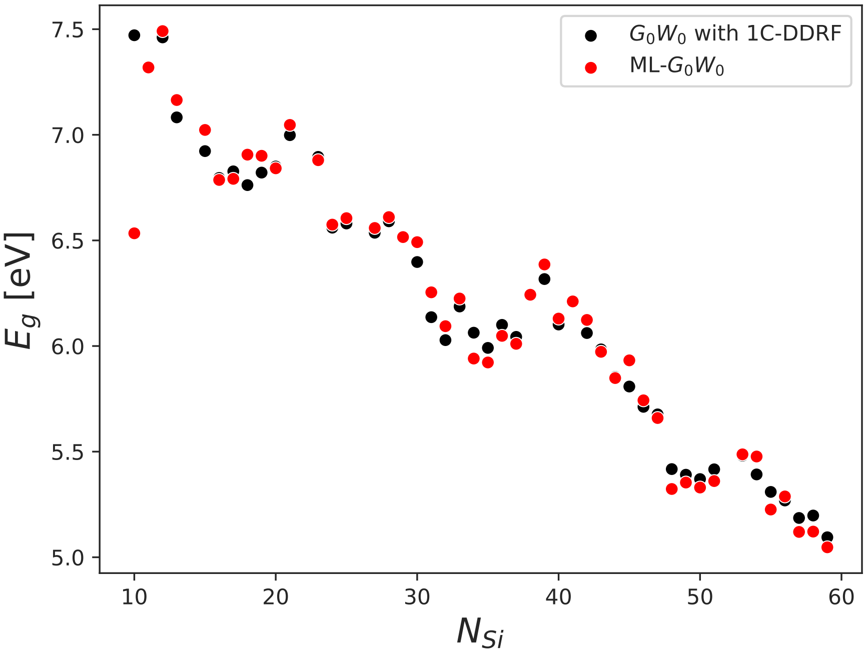


Figure 6.7: HOMO-LUMO gaps of hydrogenated silicon clusters from plane-wave G_0W_0 and G_0W_0 calculations using the 1C-DDRF and ML-G_0W_0.

Fig. 6.9 compares the ML-G_0W_0 QP corrections to plane-wave G_0W_0 results. The energies of the smallest cluster were again excluded from the plot. As expected, the differences are very similar to those between plane-wave G_0W_0 and the explicit G_0W_0 with the 1C-basis. In particular, the RMSE is 0.34 eV for all clusters and reduces to 0.30 eV when the smallest cluster

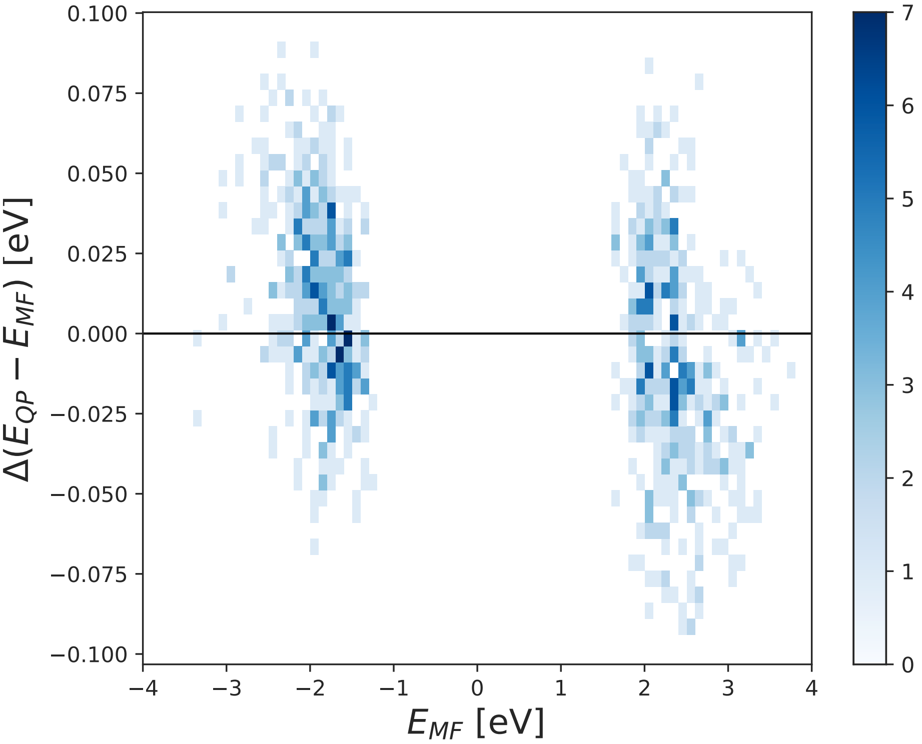


Figure 6.8: Histogram of difference in quasiparticle corrections from G_0W_0 using the 1C-DDRF and ML-G_0W_0 for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. The mean-field energies are referenced to the middle of the mean-field HOMO-LUMO gap. The energies of the smallest cluster were excluded.

is excluded. This result demonstrates that the key obstacle to improving the ML-GW approach is the development of a better basis set.

Next, we test the ability of the ML-GW approach to predict the quasiparticle energies of clusters which are larger than those included in the training data. For this, we only include clusters with up to N_{max} Si atoms in the training set with N_{max} being 60, 50 and 40. Again, the training set only includes clusters with randomly displaced atoms and the test set consists of the relaxed clusters. The predicted ML-GW for the whole set of relaxed clusters is shown in Fig. 6.10. From this graph, it is clear that the accuracy of the prediction for the largest clusters deteriorates as N_{max} is reduced: while for $N_{max} = 60$, the gaps and QP corrections for clusters with more than 60 Si atoms are still highly accurate, larger differences are observed for $N_{max} = 50$. For $N_{max} = 40$, errors as larger as 1 eV are obtained for the gaps of clusters with around 50 Si atoms. Fig. 6.10(f) shows that the large error in the gaps is a consequence of having a negative error in the QP shifts for occupied states and a positive error in the shift for

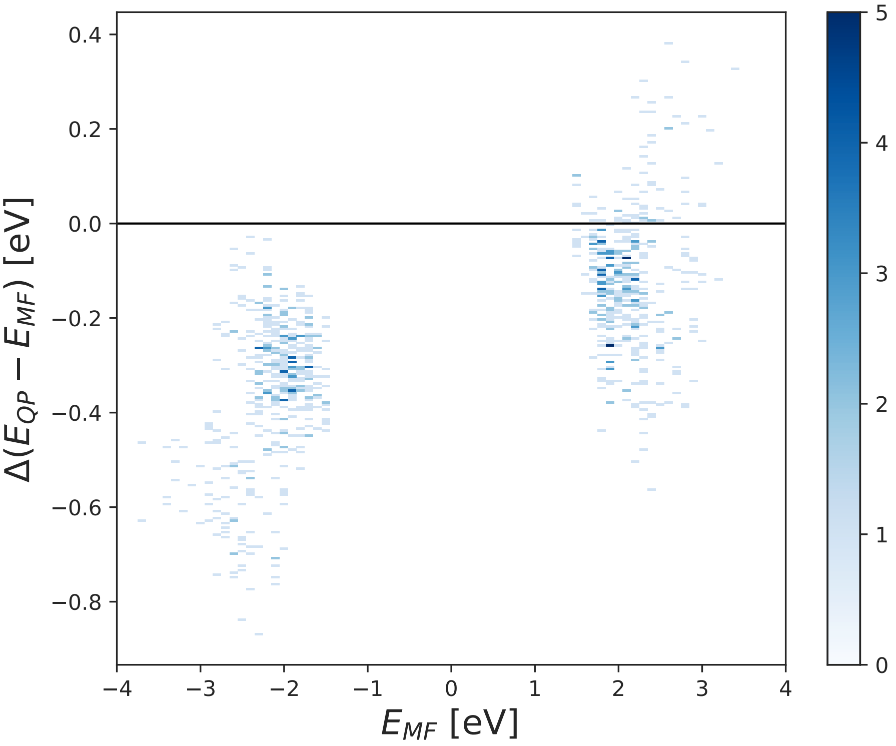


Figure 6.9: Histogram of difference in quasiparticle corrections from plane-wave G_0W_0 and ML-G_0W_0 DDRF for the 10 highest valence orbitals and the 10 lowest conduction orbitals of hydrogenated silicon clusters. The mean-field energies are referenced to the middle of the mean-field HOMO-LUMO gap. The energies of the smallest cluster were excluded.

unoccupied states. In other words: instead of a cancellation, we get an accumulation of errors when computing HOMO-LUMO gaps.

Finally, we tie together everything discussed in the three main chapters and investigate the polarizabilities obtained from the ML-DDRFs. To achieve this, we transform the predicted 1C-DDRFs into **G**-space and use the method outlined in Chapter 4 to compute the scalar polarizability. The resulting polarizabilities are shown in Fig. 6.11. The plot shows the polarizabilities obtained in Chapter 4 (RPA-DDRF), the polarizabilities obtained from the predicted DDRFs (ML 1C-DDRF), and the polarizabilities obtained using the exact 1C DDRFs. We can see that the polarizabilities obtained from the 1C-DDRF are initially overestimated, and underestimated for clusters containing more than around 20 Si atoms. The largest deviations occur between 35 and 50 Si atoms, where the error increases to almost 50 Å^3. Similar to the HOMO-LUMO gaps, however, the polarizabilities computed using the predicted DDRFs closely follow those obtained from the exact 1C-DDRFs. We also note that for clusters between

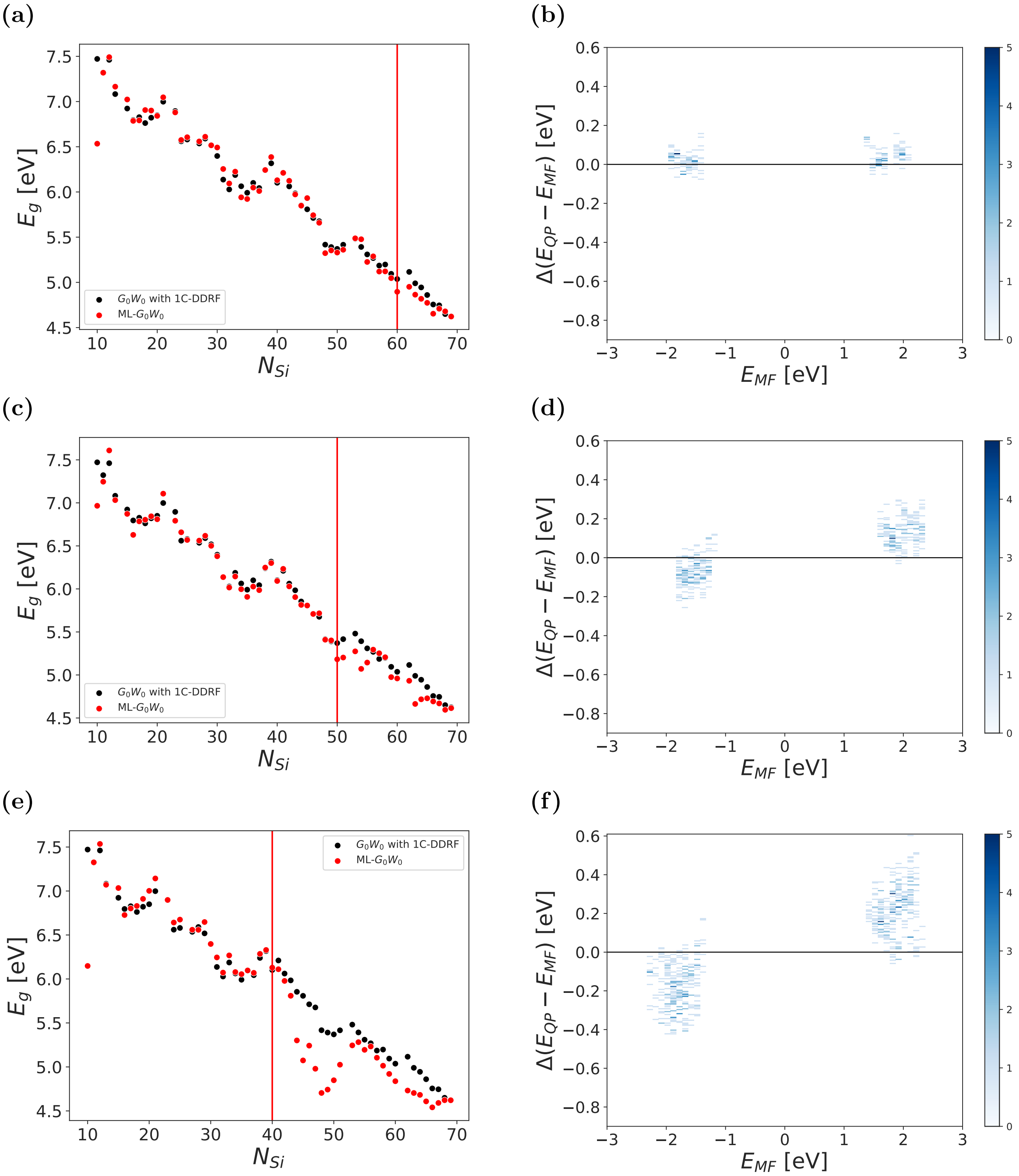


Figure 6.10: HOMO-LUMO gaps (left panels) and errors in quasiparticles shifts (right panels) from explicit G_0W_0 calculations with the 1C-DDRF and from ML-G_0W_0 trained on clusters containing up to $N_{max} = 60$ Si atoms (upper panels), $N_{max} = 50$ Si atoms (middle panels) and $N_{max} = 40$ Si atoms (lower panels). The red vertical line indicates N_{max}. The panels on the right hand side only contain results for clusters with more Si atoms than N_{max}. The mean-field energies are referenced to the middle of the mean-field HOMO-LUMO gap.

15 and 25 Si atoms both the ML-DDRF and 1C-DDRF polarizabilities closely follow the RPA-DDRF result. Moreover, beyond 50 Si atoms, the 1C-DDRF polarizabilities approach

the RPA-DDRF polarizabilities. Another interesting observation is that the linear trend in the RPA polarizability is rather poorly reproduced by the polarizabilities obtained from the 1C-DDRF. This is surprising, as the 1C-DDRF explicitly decomposes the RPA density-density response function into localised atomic contributions. However, it should be noted that the basis functions used in the 1C-DDRF are relatively long-ranged. In light of the long-rangedness of the basis functions, one would expect to obtain a linear trend in the polarizability once the clusters become significantly larger than the extent of the basis functions. However, in order to confirm this, further research is required. Furthermore, it should be stressed that the 1C basis was tuned to yield HOMO-LUMO gaps as close to the G_0W_0 gaps as possible and was not optimized with the polarizability in mind. In light of these results, it is even more clear that the largest potential for improvement in this method lies in the basis set. To this end, the major obstacle is the improvement of the radial basis functions, which are currently non-orthogonal. Chapter 7 briefly discusses a few potential alternatives to the non-orthogonal basis sets used here.

6.5 Conclusion

In this chapter, we have developed a machine-learning approach to predict the interacting density-density response function (DDRF) of materials. To achieve this, we introduce a decomposition of the DDRF into atomic contributions which form the output of a neural network. We also introduce the neighbourhood density-matrix descriptor which is a generalization of the widely used SOAP descriptor [16]: instead of symmetrizing the descriptor using a Haar integral over a symmetry group [111], we construct the tensor product of the expansion coefficients of the neighbourhood density which transforms under rotation in the same way as the atomic contributions to the DDRF. Thus, while not fully covariant, this approach is able to distinguish between different orientations of a chemical environment, which is a key requirement for predicting functions, such as the DDRF.

The machine learning technique for DDRFs is then combined with the GW approach. The resulting approach is called the ML-GW approach. We apply this method to hydrogenated

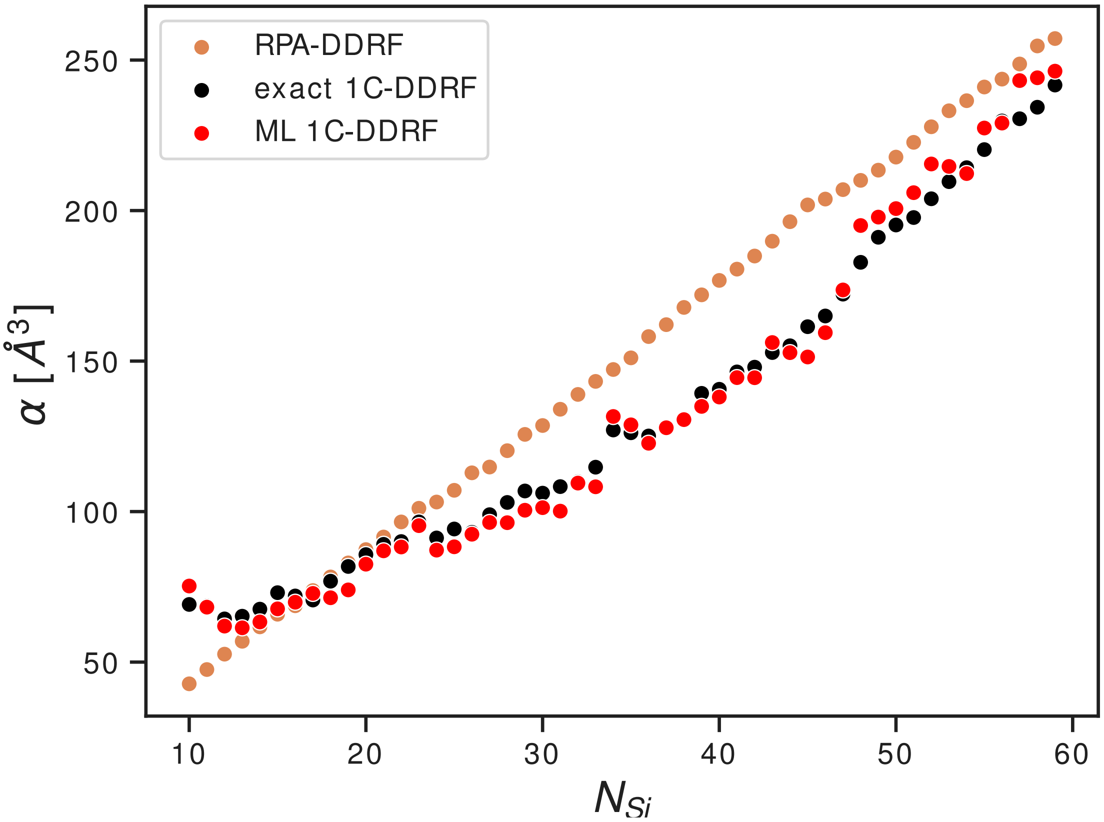


Figure 6.11: Scalar polarizabilities of unperturbed silicon clusters obtained using the ML 1C-DDRF (red), the exact 1C-DDRF (black), and the RPA-DDRF (orange).

silicon clusters. The ML-GW approach reproduces HOMO-LUMO gaps and quasiparticle energies of GW calculations using the explicitly calculated 1C-DDRF, i.e. the DDRF in a pair basis where the basis functions of each pair are centred on the same atom, with an accuracy of about 0.1 eV. The accuracy of the results deteriorates when it is applied to clusters which are larger than those included in the training set.

However, the error of ML-GW is significantly larger when compared to standard plane-wave GW results: HOMO-LUMO gaps are reproduced to within 0.5 eV, but the error reduces to 0.4 eV when the smallest cluster is excluded from the test set. These errors are comparable to those obtained by Rohlfing in his GW calculations for silane using a model dielectric function [153].

These findings demonstrate that the main challenge towards improving the ML-GW method is the construction of better local basis sets for the DDRF. The basis used for the 2C-DDRF can be improved straightforwardly by using larger basis sets, such as aug-admm-2, admm-3 or aug-admm-3 [109]. However, it is more difficult to increase the basis used for the 1C-DDRF as this leads to linear dependencies which deteriorate the predictive accuracy of the neural network. This was also observed by Grisafi et al. [72] when predicting the expansion coefficients of the electronic density using the symmetry-adapted SOAP kernel [73].

I expect that the ML-GW method can be applied to calculate quasiparticle energies in systems that have so far been out of reach for standard implementations. Examples include disordered materials, liquids, interfaces or nanoparticles. It could also be combined with on-the-fly machine learning methods [117] to perform GW calculations on molecular-dynamics snapshots to determine finite-temperature quasiparticle energies.

Chapter 7

Conclusions

In this thesis, I explored the possibility of developing a Machine Learning model of the Density-Density response function (DDRF). We started with a review of commonly used descriptors in the field of materials property prediction. The SOAP descriptor [16] was then used in Chapter 4 to construct an ML model of the scalar polarizability, a quantity which is directly derived from the DDRF. In this chapter, we found that the SOAP descriptor can indeed be used to predict scalar polarizabilities of large hydrogenated silicon clusters accurately. This result was achieved by constructing a kernel-ridge regression model using different variants of the SOAP kernel. One of the key observations from Chapter 4 was that using a small cut-off radius, which only includes nearest neighbour and next-nearest neighbour information in the descriptor, facilitates the recognition of chemical environments across different cluster sizes. Based on this observation, we used the resulting model to predict the polarizabilities of hydrogenated silicon clusters with up to 3000 silicon atoms. We found that the results approach the correct bulk limit. These results gave the first indication that the DDRF may be decomposed into atomic contributions.

In Chapter 5, we tackled the problem of reducing the size of the DDRF by projecting it onto an auxiliary basis of GTOs, thereby reducing the size of the DDRF from several gigabytes to a few megabytes, depending on the size of the auxiliary basis, while retaining a high degree of accuracy in the QP-corrections resulting from using the approximate the DDRF in a GW calculation. In doing so, we also found that the density response resulting from the DDRF decays exponentially with the distance from the atom at which the perturbation is applied. While the approximate DDRFs yield acceptable QP corrections, we failed to construct a machine-learning model to accurately predict the expansion coefficients in this basis. The failure to accurately predict the DDRF in this basis was attributed to the fact that a descriptor encoding the chemical environments of two atoms and their relative distance and orientation would be required. It is not apparent how such a descriptor would be constructed. Thus an alternative approach was developed in Chapter 6. In this chapter, we used the approximate DDRF from Chapter 5, where basis functions in $\mathbf{r}$ and $\mathbf{r}'$ were allowed to be centred on different atoms as an intermediate result, which allowed the partitioning of the DDRF into atomic contributions. These atomic contributions were then projected onto a second basis set of GTOs, where each atomic contribution is centred on a single atom. We then derived the transformation properties of these atomic contributions under rotation and constructed a descriptor called the neighbourhood density matrix, which obeys the same transformation rule. This descriptor was then used to train a dense neural network capable of predicting the atomic contributions and used the resulting DDRF in GW calculations in a hybrid ML/GW approach called the ML-GW method. The resulting method is capable of reproducing the HOMO-LUMO gaps and quasiparticle energies obtained with the approximate DDRF with remarkable accuracy. However, compared to a full G_0W_0 calculation, the average error in HOMO-LUMO gaps increases to 0.5 eV, similar to the errors that Rohlfing observed in his GW calculations for silane using a model dielectric function [153]. These results indicate that the key obstacle to achieving chemical accuracy is the use of more accurate basis sets in both the intermediate basis introduced in Chapter 5 and in the fully atom-centred basis introduced in Chapter 6. While the intermediate basis can be improved by choosing a larger auxiliary basis set [109], the size of the atom-centred basis is limited due to near-linear dependencies caused by the GTOs being non-orthogonal. One potential alternative to the SGTO basis sets used in this thesis are hydrogen wavefunctions, which have an analytical Fourier transform but require numerical evaluation of the overlap integrals with the intermediate basis of SGTOs. It should be noted, however, that the numerical evaluation of the aforementioned overlap integrals would only be required in the generation of training data for the ML-GW method. Furthermore, when using Fourier-space methods for evaluating the overlap integrals, such as those proposed by Kuang et al. [108], only the radial part of the overlap integrals has to be evaluated numerically.

Another candidate for an orthogonal radial basis is given by functions of the form

$$R_{nl}(r) = N_{nl} r^l L^n_{l+1/2}(r^2) e^{-r^2/2}, \tag{7.1}$$

where $L^n_{l+1/2}(r^2)$ are the associated Laguerre polynomials. These functions are orthogonal, which can be seen by evaluating

$$\begin{aligned} \delta_{nn'} = \int dr N_{nl} N_{n'l} r^{2l+2} L^n_{l+1/2}(r^2) L^{n'}_{l+1/2}(r^2) e^{-r^2} \\ = \frac{1}{2} \int dr' N_{nl} N_{n'l} r'^{l+1/2} L^{n'}_{l+1/2}(r') L^n_{l+1/2}(r') e^{-r'}, \end{aligned} \tag{7.2}$$

where the second equality was achieved by substitution $r' = r^2$. Replacing $l + 1/2$ with ν, we obtain

$$\delta_{nn'} = \frac{1}{2} \int dr' r'^{\nu} N_{nl} N_{n'l} L^n_\nu(r') L^{n'}_\nu(r') e^{-r'}, \tag{7.3}$$

which is simply the orthogonality relation for Laguerre polynomials [1]. The radial part of the 3-D Fourier can also be evaluated analytically [1]

$$\int_0^\infty r^{\nu+1} e^{-\beta r^2} L^n_\nu(\alpha r^2) J_\nu(kr) = 2^{-\nu-1} \beta^{-\nu-n-1} (\beta - \alpha)^n k^\nu e^{-\frac{k^2}{4\beta}} L^n_\nu\left(\frac{\alpha k^2}{4\beta(\alpha - \beta)}\right), \tag{7.4}$$

where $J_\nu(kr)$ is a Bessel function and $\nu = l + 1/2$. Moreover, the overlap of these basis functions with SGTOs can also be evaluated analytically using the aforementioned method by Kuang et al. [108]. Hence, these radial functions possess all the properties required for efficient implementation of the ML-GW method and do not require any integrals to be evaluated numerically.

While further refinements are required, several potential paths for improvements to the ML-GW method are available. Thus, the ML-GW method has the potential to enable several applications of the GW method that are currently outside the reach of traditional GW implementations.

Bibliography

[1] Chapter 7 - definite integrals of special functions. In Daniel Zwillinger, Victor Moll, I.S. Gradshteyn, and I.M. Ryzhik, editors, *Table of Integrals, Series, and Products (Eighth Edition)*, pages 776–865. Academic Press, Boston, eighth edition edition, 2014.

[2] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[3] Stephen L. Adler. Quantum theory of the dielectric constant in real solids. *Phys. Rev.*, 126:413–420, Apr 1962.

[4] S. Aja-Fernández, R. de Luis Garcia, D. Tao, and X. Li. *Tensors in Image Processing and Computer Vision*. Advances in Computer Vision and Pattern Recognition. Springer London, 2009.

[5] John M. Alred, Ksenia V. Bets, Yu Xie, and Boris I. Yakobson. Machine learning electron density in sulfur crosslinked carbon nanotubes. *Composites Science and Technology*, 166:3–9, 2018.

[6] Brandon Anderson, Truong Son Hy, and Risi Kondor. Cormorant: Covariant molecular neural networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[7] AV Andrade-Neto. Dielectric function for free electron gas: comparison between drude and lindhard models. *Revista Brasileira de Ensino de Física*, 39, 2016.

[8] J. Andén and S. Mallat. Deep scattering spectrum. *IEEE Transactions on Signal Processing*, 62(16):4114–4128, 2014.

[9] Ashish Arora, Jonathan Noky, Matthias Drüppel, Bhakti Jariwala, Thorsten Deilmann, Robert Schneider, Robert Schmidt, Osvaldo Del Pozo-Zamudio, Torsten Stiehm, Arnab Bhattacharya, Peter Krüger, Steffen Michaelis de Vasconcellos, Michael Rohlfing, and Rudolf Bratschitsch. Highly anisotropic in-plane excitons in atomically thin and bulklike 1t'-rese2. *Nano Letters*, 17(5):3202–3207, May 2017.

[10] Nongnuch Artrith and Jörg Behler. High-dimensional neural network potentials for metal surfaces: A prototype study for copper. *Phys. Rev. B*, 85:045439, Jan 2012.

[11] N. W. Ashcroft and N. D. Mermin. *Solid State Physics.* Holt-Saunders, 1976.

[12] D. Barber. *Bayesian Reasoning and Machine Learning.* Cambridge University Press, 2012.

[13] Amanda; Wilson Barnard and Hugh. Silicon quantum dot data set. *CSIRO*, v2., 2015.

[14] Albert P. Bartók, Sandip De, Carl Poelking, Noam Bernstein, James R. Kermode, Gábor Csányi, and Michele Ceriotti. Machine learning unifies the modeling of materials and molecules. *Science Advances*, 3(12), 2017.

[15] Albert P. Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Phys. Rev. B*, 87:184115, May 2013.

[16] Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104:136403, Apr 2010.

[17] Albert P. Bartók and Jonathan R. Yates. Regularized SCAN functional. *The Journal of Chemical Physics*, 150(16), 04 2019. 161101.

[18] Víctor E. Bazterra, María C. Caputo, Marta B. Ferraro, and Patricio Fuentealba. On the theoretical determination of the static dipole polarizability of intermediate size silicon clusters. *The Journal of Chemical Physics*, 117(24):11158–11165, 2002.

[19] F. Bechstedt, R. Del Sole, G. Cappellini, and Lucia Reining. An efficient method for calculating quasiparticle energies in semiconductors. *Solid State Communications*, 84(7):765 – 770, 1992.

[20] Axel D. Becke. A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics*, 98(2):1372–1377, 01 1993.

[21] Axel D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 04 1993.

[22] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401, Apr 2007.

[23] Lorin X. Benedict and Eric L. Shirley. Ab initio calculation of $\epsilon_2(\omega)$ including the electron-hole interaction: Application to gan and caf$_2$. *Phys. Rev. B*, 59:5441–5451, Feb 1999.

[24] James Bergstra, Daniel Yamins, and David Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 115–123, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

[25] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006. Chapter 5.

[26] L. C. Blum and J.-L. Reymond. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.*, 131:8732, 2009.

[27] Anton Bochkarev, Yury Lysogorskiy, Sarath Menon, Minaam Qamar, Matous Mrovec, and Ralf Drautz. Efficient parametrization of the atomic cluster expansion. *Phys. Rev. Mater.*, 6:013804, Jan 2022.

[28] Anton Bochkarev, Yury Lysogorskiy, Christoph Ortner, Gábor Csányi, and Ralf Drautz. Multilayer atomic cluster expansion for semilocal interactions. *Phys. Rev. Res.*, 4:L042019, Nov 2022.

[29] David Bohm and David Pines. A collective description of electron interactions: III. coulomb interactions in a degenerate electron gas. *Phys. Rev.*, 92:609–625, Nov 1953.

[30] Pedro Borlido, Jonathan Schmidt, Ahmad W. Huran, Fabien Tran, Miguel A. L. Marques, and Silvana Botti. Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning. *npj Computational Materials*, 6(1):96, Jul 2020.

[31] Felix Brockherde, Leslie Vogt, Li Li, Mark E. Tuckerman, Kieron Burke, and Klaus-Robert Müller. Bypassing the Kohn-Sham equations with machine learning. *Nature Communications*, 8(1), 2017.

[32] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1872–1886, 2013.

[33] D. Bump. *Lie Groups*. Graduate Texts in Mathematics. Springer New York, 2013.

[34] P. Calaminici, K. Jug, and A. M. Köster. Density functional calculations of molecular polarizabilities and hyperpolarizabilities. *The Journal of Chemical Physics*, 109(18):7756–7763, 1998.

[35] G Cappellini, Del Sole R, Lucia Reining, and F Bechstedt. Model dielectric function for semiconductors. *Physical review. B, Condensed matter*, 47:9892–9895, 05 1993.

[36] Michele Ceriotti, Michael J. Willatt, and Gábor Csányi. *Machine Learning of Atomic-Scale Properties Based on Physical Principles*, pages 1–27. Springer International Publishing, Cham, 2018.

[37] M. K. Y. Chan and G. Ceder. Efficient band gap prediction for solids. *Phys. Rev. Lett.*, 105:196403, Nov 2010.

[38] Anand Chandrasekaran, Deepak Kamal, Rohit Batra, Chiho Kim, Lihua Chen, and Rampi Ramprasad. Solving the electronic structure problem with machine learning. *npj Computational Materials*, 5(1), 2019.

[39] Kamal Choudhary, Brian DeCost, and Francesca Tavazza. Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape. *Phys. Rev. Materials*, 2:083801, Aug 2018.

[40] Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR.

[41] Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. *CoRR*, abs/1801.10130, 2018.

[42] Christopher R. Collins, Geoffrey J. Gordon, O. Anatole von Lilienfeld, and David J. Yaron. Constant size descriptors for accurate machine learning models of molecular properties. *The Journal of Chemical Physics*, 148(24):241718, 2018.

[43] Basil J B Crowley. Generalized plasmon-pole approximation for the dielectric function of a multicomponent plasma. *arXiv: Plasma Physics*, 2015.

[44] Isabel S. Curtis, Ryan J. Wills, and Mita Dasog. Photocatalytic hydrogen generation using mesoporous silicon nanoparticles: influence of magnesiothermic reduction conditions and nanoparticle aging on the catalytic activity. *Nanoscale*, 13:2685–2692, 2021.

[45] Sandip De, Albert P. Bartók, Gábor Csányi, and Michele Ceriotti. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.*, 18:13754–13769, 2016.

[46] Elena Degoli, G. Cantele, Eleonora Luppi, Rita Magri, D. Ninno, O. Bisi, and Stefano Ossicini. Ab initio structural and electronic properties of hydrogenated silicon nanoclusters in the ground and excited state. *Phys. Rev. B*, 69:155411, Apr 2004.

[47] Ke Deng, Jinlong Yang, and C. T. Chan. Calculated polarizabilities of small si clusters. *Phys. Rev. A*, 61:025201, Jan 2000.

[48] Jack Deslippe, Georgy Samsonidze, Manish Jain, Marvin L Cohen, and Steven G Louie. Coulomb-hole summations and energies for g w calculations with limited number of empty orbitals: A modified static remainder approach. *Physical Review B*, 87(16):165124, 2013.

[49] Jack Deslippe, Georgy Samsonidze, David A. Strubbe, Manish Jain, Marvin L. Cohen, and Steven G. Louie. Berkeleygw: A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures. *Computer Physics Communications*, 183(6):1269 – 1289, 2012.

[50] R. Drautz and M. Fähnle. Spin-cluster expansion: Parametrization of the general adiabatic magnetic energy surface with ab initio accuracy. *Phys. Rev. B*, 69:104404, Mar 2004.

[51] Ralf Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B*, 99:014104, Jan 2019.

[52] Ralf Drautz. Atomic cluster expansion of scalar, vectorial, and tensorial properties including magnetism and charge transfer. *Phys. Rev. B*, 102:024104, Jul 2020.

[53] Ralf Drautz and Christoph Ortner. Atomic cluster expansion and wave function representations. *arXiv preprint arXiv:2206.11375*, 2022.

[54] Matthias Drüppel, Thorsten Deilmann, Jonathan Noky, Philipp Marauhn, Peter Krüger, and Michael Rohlfing. Electronic excitations in transition metal dichalcogenide monolayers from an lda+GdW approach. *Phys. Rev. B*, 98:155433, Oct 2018.

[55] Geneviève Dusson, Markus Bachmayr, Gábor Csányi, Ralf Drautz, Simon Etter, Cas van der Oord, and Christoph Ortner. Atomic cluster expansion: Completeness, efficiency and stability. *Journal of Computational Physics*, 454:110946, 2022.

[56] Michael Eickenberg, Georgios Exarchakis, Matthew Hirn, and Stephane Mallat. Solid harmonic wavelet scattering: Predicting quantum molecular energy from invariant descriptors of 3d electronic densities. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6540–6549. Curran Associates, Inc., 2017.

[57] Felix Faber, Alexander Lindmaa, O. Anatole von Lilienfeld, and Rickard Armiento. Crystal structure representations for machine learning models of formation energies. *International Journal of Quantum Chemistry*, 115(16):1094–1101, 2015.

[58] Felix A. Faber, Anders S. Christensen, Bing Huang, and O. Anatole von Lilienfeld. Alchemical and structural distribution based representation for universal quantum machine learning. *The Journal of Chemical Physics*, 148(24):241717, 2018.

[59] Felix A. Faber, Alexander Lindmaa, O. Anatole Von Lilienfeld, and Rickard Armiento. Machine Learning Energies of 2 Million Elpasolite(ABC2d6)Crystals. *Physical Review Letters*, 117(13), 2016.

[60] Alberto Fabrizio, Andrea Grisafi, Benjamin Meyer, Michele Ceriotti, and Clemence Corminboeuf. Electron density learning of non-covalent systems. *Chemical science*, 10(41):9424–9432, 2019.

[61] Behnam Farid. A note on the many-body perturbation theory. *Philosophical Magazine Letters*, 79(8):581–593, 1999.

[62] David Feller. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.*, 17:1571–1586, 1996.

[63] A.L. Fetter and J.D. Walecka. *Quantum Theory of Many-particle Systems*. Dover Books on Physics. Dover Publications, 2003.

[64] In Fortran, William Press, Saul Teukolsky, William Vetterling, and Brian Flannery. Numerical recipes. *Cambridge, UK, Cambridge University Press*, 01 1992.

[65] Y. Fu, A. Dutta, M. Willander, and S. Oda. Carrier conduction in a Si-nanocrystal-based single-electron transistor-I. effect of gate bias. *Superlattices and Microstructures*, 28(3):177 – 187, 2000.

[66] Viktor Mikhailovich Galitskii and Arkadii Beinusovich Migdal. Application of quantum field theory methods to the many body problem. *Sov. Phys. JETP*, 7(96):18, 1958.

[67] Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

[68] Murray Gell-Mann and Francis Low. Bound states in quantum field theory. *Phys. Rev.*, 84:350–354, Oct 1951.

[69] P Giannozzi, O Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, and et al. Advanced capabilities for materials modelling with quantum espresso. *Journal of Physics: Condensed Matter*, 29(46):465901, Oct 2017.

[70] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, Andrea Dal Corso, Stefano de Gironcoli, Stefano Fabris, Guido Fratesi, Ralph Gebauer, Uwe Gerstmann, Christos Gougoussis, Anton Kokalj, Michele Lazzeri, Layla Martin-Samos, Nicola Marzari, Francesco Mauri, Riccardo Mazzarello, Stefano Paolini, Alfredo Pasquarello, Lorenzo Paulatto, Carlo Sbraccia, Sandro Scandolo, Gabriele Sclauzero, Ari P Seitsonen, Alexander Smogunov, Paolo Umari, and Renata M Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, sep 2009.

[71] Andrea Grisafi and Michele Ceriotti. Incorporating long-range physics in atomic-scale machine learning. *The Journal of Chemical Physics*, 151(20):204105, 2019.

[72] Andrea Grisafi, Alberto Fabrizio, Benjamin Meyer, David M. Wilkins, Clemence Corminboeuf, and Michele Ceriotti. Transferable machine-learning model of the electron density. *ACS Central Science*, 5(1):57–64, 2019.

[73] Andrea Grisafi, David M. Wilkins, Gábor Csányi, and Michele Ceriotti. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.*, 120:036002, Jan 2018.

[74] Andrea Grisafi, David M. Wilkins, Michael J. Willatt, and Michele Ceriotti. *Atomic-Scale Representation and Statistical Learning of Tensorial Properties*, chapter 1, pages 1–21.

[75] O. Gunnarsson and B. I. Lundqvist. Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism. *Phys. Rev. B*, 13:4274–4298, May 1976.

[76] P. H. Hahn, W. G. Schmidt, and F. Bechstedt. Bulk excitonic effects in surface optical spectra. *Phys. Rev. Lett.*, 88:016402, Dec 2001.

[77] Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O. Anatole Von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *The Journal of Physical Chemistry Letters*, 6(12):2326–2331, 2015.

[78] Katja Hansen, Grégoire Montavon, Franziska Biegler, Siamac Fazli, Matthias Rupp, Matthias Scheffler, O. Anatole von Lilienfeld, Alexandre Tkatchenko, and Klaus-Robert Müller. Assessment and validation of machine learning methods for predicting molecular atomization energies. *Journal of Chemical Theory and Computation*, 9(8):3404–3419, 2013. PMID: 26584096.

[79] Lars Hedin. New method for calculating the one-particle green's function with application to the electron-gas problem. *Phys. Rev.*, 139:A796–A823, Aug 1965.

[80] Lauri Himanen, Marc O. J. Jäger, Eiaki V. Morooka, Filippo Federici Canova, Yashasvi S. Ranawat, David Z. Gao, Patrick Rinke, and Adam S. Foster. DScribe: Library of descriptors for machine learning in materials science. *Computer Physics Communications*, 247:106949, 2020.

[81] Matthew Hirn, Stéphane Mallat, and Nicolas Poilvert. Wavelet scattering regression of quantum chemical energies. *Multiscale Modeling & Simulation*, 15(2):827–863, 2017.

[82] Matthew Hirn, Stéphane Mallat, and Nicolas Poilvert. Wavelet scattering regression of quantum chemical energies. *Multiscale Modeling & Simulation*, 15(2):827–863, 2017.

[83] Matthew Hirn, Nicolas Poilvert, and Stéphane Mallat. Quantum energy regression using scattering transforms. *arXiv preprint arXiv:1502.02077*, 2015.

[84] F. L. Hirshfeld. Bonded-atom fragments for describing molecular charge densities. *Theoretica chimica acta*, 44(2):129–138, Jun 1977.

[85] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, Nov 1964.

[86] Eric R. Homer, Derek M. Hensley, Conrad W. Rosenbrock, Andrew H. Nguyen, and Gus L. W. Hart. Machine-learning informed representations for grain boundary structures. *Frontiers in Materials*, 6:168, 2019.

[87] Shreyas J. Honrao, Stephen R. Xie, and Richard G. Hennig. Augmenting machine learning of energy landscapes with local structural information. *Journal of Applied Physics*, 128(8):085101, 2020.

[88] Bing Huang and O. Anatole von Lilienfeld. Communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity. *The Journal of Chemical Physics*, 145(16):161102, 2016.

[89] Mark S. Hybertsen and Steven G. Louie. First-principles theory of quasiparticles: Calculation of band gaps in semiconductors and insulators. *Phys. Rev. Lett.*, 55:1418–1421, Sep 1985.

[90] Mark S. Hybertsen and Steven G. Louie. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Phys. Rev. B*, 34:5390–5413, Oct 1986.

[91] Mark S. Hybertsen and Steven G. Louie. Ab initio static dielectric matrices from the density-functional approach. i. formulation and application to semiconductors and insulators. *Phys. Rev. B*, 35:5585–5601, Apr 1987.

[92] Mark S. Hybertsen and Steven G. Louie. Model dielectric matrices for quasiparticle self-energy calculations. *Phys. Rev. B*, 37:2733–2736, Feb 1988.

[93] Sohrab Ismail-Beigi. Truncation of periodic image interactions for confined systems. *Phys. Rev. B*, 73:233103, Jun 2006.

[94] K. A. Jackson, M. Yang, I. Chaudhuri, and Th. Frauenheim. Shape, polarizability, and metallicity in silicon clusters. *Phys. Rev. A*, 71:033205, Mar 2005.

[95] Koblar Jackson, Mark Pederson, Cai-Zhuang Wang, and Kai-Ming Ho. Calculated polarizabilities of intermediate-size si clusters. *Phys. Rev. A*, 59:3685–3689, May 1999.

[96] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin a. Persson. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.

[97] B. Jansik, B. Schimmelpfennig, P. Norman, Y. Mochizuki, Y. Luo, and H. Ågren. Size, order, and dimensional relations for silicon cluster polarizabilities. *The Journal of Physical Chemistry A*, 106(2):395–399, 2002.

[98] N. Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists*. An Introduction to Tensors and Group Theory for Physicists. Birkhäuser Boston, 2011.

[99] Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In Claire Nédellec and Céline Rouveirol, editors, *Machine Learning: ECML-98*, pages 137–142, Berlin, Heidelberg, 1998. Springer Berlin Heidelberg.

[100] R. Kakarala. A group-theoretic approach to the triple correlation. In *[1993 Proceedings] IEEE Signal Processing Workshop on Higher-Order Statistics*, pages 28–32, June 1993.

[101] Scott Kirklin, James E. Saal, Bryce Meredig, Alex Thompson, Jeff W. Doak, Muratahan Aykol, Stephan Rühl, and Chris Wolverton. The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *npj Computational Materials*, 1(1):15010, Dec 2015.

[102] Abraham Klein and Richard Prange. Perturbation theory for an infinite medium of fermions. *Phys. Rev.*, 112:994–1007, Nov 1958.

[103] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[104] Risi Kondor. A novel set of rotationally and translationally invariant features for images based on the non-commutative bispectrum. *arXiv preprint cs/0701127*, 2007.

[105] Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch–gordan nets: a fully fourier space spherical convolutional neural network. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[106] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2747–2755. PMLR, 10–15 Jul 2018.

[107] G Kresse and J Hafner. Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements. *Journal of Physics: Condensed Matter*, 6(40):8245, oct 1994.

[108] Jiyun Kuang and C D Lin. Molecular integrals over spherical gaussian-type orbitals: I. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 30(11):2529–2548, jun 1997.

[109] Chandan Kumar, Heike Fliegl, Frank Jensen, Andrew M. Teale, Simen Reine, and Thomas Kjærgaard. Accelerating kohn-sham response theory using density fitting and the auxiliary-density-matrix method. *International Journal of Quantum Chemistry*, 118(16):e25639, 2018.

[110] Dmitri N. Laikov. Fast evaluation of density functional exchange-correlation terms using the expansion of the electron density in auxiliary basis sets. *Chemical Physics Letters*, 281(1):151–156, 1997.

[111] Marcel F. Langer, Alex Goeßmann, and Matthias Rupp. Representations of molecules and materials for interpolation of quantum-mechanical simulations via machine learning. *npj Computational Materials*, 8(1):41, Mar 2022.

[112] Kostiantyn Lapchevskyi, Benjamin Miller, Mario Geiger, Tess Smidt, Swiss National Science Foundation, USDOE, and Free University of Berlin. Euclidean neural networks (e3nn) v1.0, version v1.0, 10 2020.

[113] H Lehmann. On the properties of propagation functions and renormalization contants of quantized fields. *Nuovo Cim*, 11(342):10, 1954.

[114] Zachary H. Levine and Steven G. Louie. New model dielectric function and exchange-correlation potential for semiconductors and insulators. *Phys. Rev. B*, 25:6310–6316, May 1982.

[115] Alan M. Lewis, Andrea Grisafi, Michele Ceriotti, and Mariana Rossi. Learning electron densities in the condensed phase. *Journal of Chemical Theory and Computation*, 17(11):7203–7214, 2021. PMID: 34669406.

[116] Zhenwei Li, James R. Kermode, and Alessandro De Vita. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.*, 114:096405, Mar 2015.

[117] Zhenwei Li, James R. Kermode, and Alessandro De Vita. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.*, 114:096405, Mar 2015.

[118] Jens Lindhard. Dielectric function for free electron gas: comparison between Drude and Lindhard models. *Danske Matematisk-fysiske Meddeleiser*, 28(8):1–57, 1954.

[119] Johannes Lischner, Sahar Sharifzadeh, Jack Deslippe, Jeffrey B. Neaton, and Steven G. Louie. Effects of self-consistency and plasmon-pole models on *gw* calculations for closed-shell molecules. *Phys. Rev. B*, 90:115130, Sep 2014.

[120] Yury Lysogorskiy, Anton Bochkarev, Matous Mrovec, and Ralf Drautz. Active learning strategies for atomic cluster expansion models. *Phys. Rev. Mater.*, 7:043801, Apr 2023.

[121] Yury Lysogorskiy, Cas van der Oord, Anton Bochkarev, Sarath Menon, Matteo Rinaldi, Thomas Hammerschmidt, Matous Mrovec, Aidan Thompson, Gábor Csányi, Christoph Ortner, and Ralf Drautz. Performant implementation of the atomic cluster expansion (pace) and application to copper and silicon. *npj Computational Materials*, 7(1):97, Jun 2021.

[122] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[123] George Maroulis, Didier Begué, and Claude Pouchan. Accurate dipole polarizabilities of small silicon clusters from ab initio and density functional theory calculations. *The Journal of Chemical Physics*, 119(2):794–797, 2003.

[124] George Maroulis and Claude Pouchan. Assessing the performance of ab initio methods on static (hyper)polarizability predictions for silicon clusters. si4 as a test case. *Phys. Chem. Chem. Phys.*, 5:1992–1995, 2003.

[125] Paul C. Martin and Julian Schwinger. Theory of many-particle systems. I. *Phys. Rev.*, 115:1342–1373, Sep 1959.

[126] Richard M. Martin. *Electronic Structure: Basic Theory and Practical Methods.* Cambridge University Press, 2004.

[127] Heinrich Maschke. Ueber den arithmetischen charakter der coefficienten der substitutionen endlicher linearer substitutionsgruppen. *Mathematische Annalen*, 50(4):492–498, Dec 1898.

[128] Jadranka Micic. Inequalities involving the khatri-rao product of matrices. *JOURNAL OF MATHEMATICAL INEQUALITIES*, 3(4):617–630, 2009.

[129] Yuji Mochizuki and Hans Ågren. Polarizability of silicon clusters. *Chemical Physics Letters*, 336(5):451 – 456, 2001.

[130] Grégoire Montavon, Katja Hansen, Siamac Fazli, Matthias Rupp, Franziska Biegler, Andreas Ziehe, Alexandre Tkatchenko, Anatole V. Lilienfeld, and Klaus-Robert Müller. Learning invariant representations of molecules for atomization energy prediction. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 440–448. Curran Associates, Inc., 2012.

[131] Grégoire Montavon, Matthias Rupp, Vivekanand Gobre, Alvaro Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New Journal of Physics*, 15(9):095003, sep 2013.

[132] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[133] Felix Musil, Andrea Grisafi, Albert P. Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Physics-inspired structural representations for molecules and materials. *Chemical Reviews*, 2021. PMID: 34310133.

[134] Bastien Mussard and János G. Ángyán. Relationships between charge density response functions, exchange holes and localized orbitals. *Computational and Theoretical Chemistry*, 1053:44 – 52, 2015. Special Issue: Understanding structure and reactivity from topology and beyond.

[135] Jigyasa Nigam, Sergey N Pozdnyakov, Kevin K Huguenin-Dumittan, and Michele Ceriotti. Completeness of atomic structure representations. *arXiv preprint arXiv:2302.14770*, 2023.

[136] John E. Northrup. Electronic structure of si(100)c(4×2) calculated within the GW approximation. *Phys. Rev. B*, 47:10032–10035, Apr 1993.

[137] S. Obara and A. Saika. Efficient recursive computation of molecular integrals over cartesian gaussian functions. *The Journal of Chemical Physics*, 84(7):3963–3974, 1986.

[138] Norah O'Farrell, Andrew Houlton, and Benjamin R. Horrocks. Silicon nanoparticles: applications in cell biology and medicine. *International journal of nanomedicine*, 1(4):451–472, 2006.

[139] Giovanni Onida, Lucia Reining, and Angel Rubio. Electronic excitations: density-functional versus many-body green's-function approaches. *Rev. Mod. Phys.*, 74:601–659, Jun 2002.

[140] M. G. Papadopoulos, H. Reis, A. Avramopoulos, Ş Erkoç, and L. Amirouche. Polarizabilities and second hyperpolarizabilities of zn m cd n clusters. *Molecular Physics*, 104(13-14):2027–2036, 2006.

[141] Ji-Ho Park, Luo Gu, Geoffrey von Maltzahn, Erkki Ruoslahti, Sangeeta N. Bhatia, and Michael J. Sailor. Biodegradable luminescent porous silicon nanoparticles for in vivo applications. *Nature Materials*, 8(4):331–336, Apr 2009.

[142] Yaroslav Pavlyukh, Jamal Berakdar, and Angel Rubio. Taming singularities of the diagrammatic many-body perturbation theory. *arXiv preprint arXiv:1601.04285*, 2016.

[143] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048–5079, May 1981.

[144] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[145] John P. Perdew and Mel Levy. Physical content of the exact kohn-sham orbital energies: Band gaps and derivative discontinuities. *Phys. Rev. Lett.*, 51:1884–1887, Nov 1983.

[146] M.E. Povarnitsyn, N.S. Shcheblanov, D.S. Ivanov, V. Yu. Timoshenko, and S.M. Klimentov. Vibrational analysis of silicon nanoparticles using simulation and decomposition of raman spectra. *Phys. Rev. Applied*, 14:014067, 2020.

[147] Sergey N. Pozdnyakov, Michael J. Willatt, Albert P. Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Incompleteness of atomic structure representations. *Phys. Rev. Lett.*, 125:166001, Oct 2020.

[148] Jonathan L. Priedeman, Conrad W. Rosenbrock, Oliver K. Johnson, and Eric R. Homer. Quantifying and connecting atomic and crystallographic grain boundary structure using local environment representation and dimensionality reduction techniques. *Acta Materialia*, 161:431 – 443, 2018.

[149] Benjamin P. Pritchard, Doaa Altarawy, Brett Didier, Tara D. Gibsom, and Theresa L. Windus. A new basis set exchange: An open, up-to-date resource for the molecular sciences community. *J. Chem. Inf. Model.*, 59:4814–4820, 2019.

[150] Minaam Qamar, Matous Mrovec, Yury Lysogorskiy, Anton Bochkarev, and Ralf Drautz. Atomic cluster expansion for quantum-accurate large-scale simulations of carbon. *Journal of Chemical Theory and Computation*, 2022. PMID: 37347981.

[151] Andrew M. Rappe, Karin M. Rabe, Efthimios Kaxiras, and J. D. Joannopoulos. Optimized pseudopotentials. *Phys. Rev. B*, 41:1227–1230, Jan 1990.

[152] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning).* The MIT Press, 2005.

[153] Michael Rohlfing. Electronic excitations from a perturbative LDA + gdw approach. *Phys. Rev. B*, 82:205127, Nov 2010.

[154] Franz M Rohrhofer, Santanu Saha, Simone Di Cataldo, Bernhard C Geiger, Wolfgang von der Linden, and Lilia Boeri. Importance of feature engineering and database selection in a machine learning model: a case study on carbon crystal structures. *arXiv preprint arXiv:2102.00191*, 2021.

[155] M.E. Rose. *Elementary Theory of Angular Momentum.* Dover Books on Physics Series. Dover Publications, Incorporated, 2013.

[156] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[157] 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.

[158] Matthias Rupp. Machine learning for quantum mechanics in a nutshell. *International Journal of Quantum Chemistry*, 115(16):1058–1073, 8 2015.

[159] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.*, 108:058301, Jan 2012.

[160] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.*, 108:058301, Jan 2012.

[161] James E. Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and C. Wolverton. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (oqmd). *JOM*, 65(11):1501–1509, Nov 2013.

[162] J. M. Sanchez. Cluster expansion and the configurational theory of alloys. *Phys. Rev. B*, 81:224202, Jun 2010.

[163] H. Bernhard Schlegel and Michael J. Frisch. Transformation between cartesian and pure spherical harmonic gaussians. *International Journal of Quantum Chemistry*, 54:83–87, 1995.

[164] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.*, 28:1049–1070, Dec 1926.

[165] Karen L. Schuchardt, Brett T. Didier, Todd Elsethagen, Lisong Sun, Vidhya Gurumoorthi, Jared Chase, Jun Li, and Theresa L. Windus. Basis set exchange: A community database for computational sciences. *J. Chem. Inf. Model.*, 47:1045–1052, 2007.

[166] Peter A. Schultz. Theory of defect levels and the “band gap problem” in silicon. *Phys. Rev. Lett.*, 96:246401, Jun 2006.

[167] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Phys. Rev. B*, 89:205118, May 2014.

[168] Peter Schwerdtfeger. The pseudopotential approximation in electronic structure theory. *ChemPhysChem*, 12(17):3143–3155, 2011.

[169] L. J. Sham and M. Schlüter. Density-functional theory of the energy gap. *Phys. Rev. Lett.*, 51:1888–1891, Nov 1983.

[170] Alexander V. Shapeev. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling & Simulation*, 14(3):1153–1173, 2016.

[171] S. Sharifzadeh, I. Tamblyn, P. Doak, P. T. Darancet, and J. B. Neaton. Quantitative molecular orbital energies within a G0W0 approximation. *The European Physical Journal B*, 85(9):323, Sep 2012.

[172] J. Sólyom. *Fundamentals of the Physics of Solids: Volume 3 - Normal, Broken-Symmetry, and Correlated Systems.* Theoretical Solid State Physics: Interaction Among Electrons. Springer Berlin Heidelberg, 2010.

[173] Catalin-Dan Spataru. *Electron excitations in solids and novel materials.* PhD thesis, 2004.

[174] Sina Stocker, Gábor Csányi, Karsten Reuter, and Johannes T. Margraf. Machine learning in chemical reaction space. *Nature Communications*, 11(1):5505, Oct 2020.

[175] G. Strinati, H. J. Mattausch, and W. Hanke. Dynamical aspects of correlation corrections in a covalent crystal. *Phys. Rev. B*, 25:2867–2888, Feb 1982.

[176] Jianwei Sun, Adrienn Ruzsinszky, and John P. Perdew. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.*, 115:036402, Jul 2015.

[177] M Taut. Frequency moments of the dielectric function for an inhomogeneous electron gas. *Journal of Physics C: Solid State Physics*, 18(13):2677, 1985.

[178] Christopher D. Taylor. Connections between the energy functional and interaction potentials for materials simulations. *Phys. Rev. B*, 80:024104, Jul 2009.

[179] Nathaniel Thomas, Tess Smidt, Steven M. Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *CoRR*, abs/1802.08219, 2018.

[180] Murilo L. Tiago and James R. Chelikowsky. Optical excitations in organic molecules, clusters, and defects studied by first-principles green's function methods. *Phys. Rev. B*, 73:205334, May 2006.

[181] Tran Tuan-Anh and Robert Zaleśny. Predictions of high-order electric properties of molecules: Can we benefit from machine learning? *ACS Omega*, 5(10):5318–5325, 2020.

[182] E. F. Valeev. Libint: A library for the evaluation of molecular integrals of many-body operators over gaussian functions. http://libint.valeyev.net/, 2022. version 2.8.0.

[183] Igor Vasiliev, Serdar Öğüt, and James R. Chelikowsky. Ab initio calculations for the polarizabilities of small semiconductor clusters. *Phys. Rev. Lett.*, 78:4805–4808, Jun 1997.

[184] Max Veit, David M. Wilkins, Yang Yang, Robert A. DiStasio, and Michele Ceriotti. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *The Journal of Chemical Physics*, 153(2):024113, 2020.

[185] U von Barth and A C Pedroza. The cohesive energy and charge-density form factors of beryllium as a test on the langreth–perdew–mehl approximation. *Physica Scripta*, 32(4):353, oct 1985.

[186] O. Anatole von Lilienfeld, Raghunathan Ramakrishnan, Matthias Rupp, and Aaron Knoll. Fourier series of atomic radial distribution functions: A molecular fingerprint for machine learning models of quantum chemical properties. *International Journal of Quantum Chemistry*, 115(16):1084–1093, 2015.

[187] Seymour H Vosko, Leslie Wilk, and Marwan Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of physics*, 58(8):1200–1211, 1980.

[188] Junmei Wang, Piotr Cieplak, Jie Li, Tingjun Hou, Ray Luo, and Yong Duan. Development of polarizable models for molecular mechanical calculations i: Parameterization of atomic polarizability. *The Journal of Physical Chemistry B*, 115(12):3091–3099, 2011. PMID: 21391553.

[189] Y. Q. Wang, Y. G. Wang, L. Cao, and Z. X. Cao. High-efficiency visible photoluminescence from amorphous silicon nanoparticles embedded in silicon nitride. *Applied Physics Letters*, 83(17):3474–3476, 2003.

[190] Zhi-Xiang Wang, Wei Zhang, Chun Wu, Hongxing Lei, Piotr Cieplak, and Yong Duan. Strike a balance: Optimization of backbone torsion parameters of amber polarizable force field for simulations of proteins and peptides. *Journal of Computational Chemistry*, 27(6):781–790, 2006.

[191] G. C. Wick. The evaluation of the collision matrix. *Phys. Rev.*, 80:268–272, Oct 1950.

[192] David M. Wilkins, Andrea Grisafi, Yang Yang, Ka Un Lao, Robert A. DiStasio, and Michele Ceriotti. Accurate molecular polarizabilities with coupled cluster theory and machine learning. *Proceedings of the National Academy of Sciences*, 116(9):3401–3406, 2019.

[193] Michael J. Willatt, Félix Musil, and Michele Ceriotti. Atom-density representations for machine learning. *The Journal of Chemical Physics*, 150(15):154110, Apr 2019.

[194] S. Wippermann, M. Vörös, D. Rocca, A. Gali, G. Zimanyi, and G. Galli. High-pressure core structures of si nanoparticles for solar energy conversion. *Phys. Rev. Lett.*, 110:046804, Jan 2013.

[195] Nathan Wiser. Dielectric constant with local field effects included. *Phys. Rev.*, 129:62–69, Jan 1963.

[196] Mario G Zauchner, Stefano Dal Forno, Gábor Csányi, Andrew Horsfield, and Johannes Lischner. Predicting polarizabilities of silicon clusters using local chemical environments. *Machine Learning: Science and Technology*, 2(4):045029, oct 2021.

[197] Mario G Zauchner, Andrew Horsfield, and Johannes Lischner. Accelerating gw calculations through machine learned dielectric matrices. *arXiv preprint arXiv:2305.02990*, 2023.

[198] Dexuan Zhou, Huajie Chen, Cheuk Hin Ho, and Christoph Ortner. A multilevel method for many-electron schrödinger equations based on the atomic cluster expansion. *arXiv preprint arXiv:2304.04260*, 2023.

Appendix A

Further details on many-body perturbation theory

A.1 Second quantization

In second quantization, the many-body wave function is written in terms of populations in certain single-particle states. For example the wave function $|0, 0, n_3, \cdots 0\rangle$ contains n_3 particles in state 3.

We can now define the creation operator $\hat{c}_n^\dagger$, which creates a particle in state n and the destruction operator, $\hat{c}_n$, which destroys a particle in state n. Assuming the vacuum state is given by $|0\rangle$, we get

$$\hat{c}_n^\dagger|0\rangle = |0, 0, \cdots 1, \cdots 0\rangle, \tag{A.1}$$

where the 1 indicates a particle in state n and

$$\hat{c}_n|0, 0, \cdots 1, \cdots 0\rangle = |0\rangle. \tag{A.2}$$

Since only a single fermion can ooccupy a state, the creation operator yields zero if state n is already occupied, and the destruction operator yields zero if n is not occupied. For fermions, creation and destruction operators have the following anti-commutation relations [63]

$$[\hat{c}_r, \hat{c}_s] = [\hat{c}_r^\dagger, \hat{c}_s^\dagger] = 0, \tag{A.3}$$

and

$$[\hat{c}_r, \hat{c}_s^\dagger] = \delta_{rs}, \tag{A.4}$$

where the anti-commutator is defined as

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} + \hat{B}\hat{A}. \tag{A.5}$$

A.2 Field operators

Given a complete set of single-particle wave functions $\{\psi_\alpha(\mathbf{r})\}$, where α is a spin index, we can define field operators as [63]

$$\hat{\psi}_\alpha(\mathbf{r}) = \sum_n \psi_{\alpha n}(\mathbf{r})\hat{c}_{\alpha n}, \tag{A.6}$$

which follow the anti-commutation relations

$$[\hat{\psi}_\alpha(\mathbf{r}), \hat{\psi}_\beta^\dagger(\mathbf{r}')] = \delta_{\alpha\beta}\delta(\mathbf{r} - \mathbf{r}'), \tag{A.7}$$

and

$$[\hat{\psi}_\alpha(\mathbf{r}), \hat{\psi}_\beta(\mathbf{r}')] = 0 \tag{A.8}$$

$$[\hat{\psi}_\alpha^\dagger(\mathbf{r}), \hat{\psi}_\beta^\dagger(\mathbf{r}')] = 0. \tag{A.9}$$

We can now further define the operator $\hat{\psi}(\mathbf{r})$ as

$$\hat{\psi}(\mathbf{r}) = \begin{bmatrix} \hat{\psi}_1(\mathbf{r}) \\ \hat{\psi}_2(\mathbf{r}) \end{bmatrix}, \tag{A.10}$$

where the indices correspond to up and down spins, respectively. With these definitions, a general single-particle operator in second quantization is defined as [63]

$$\hat{O} = \int d\mathbf{r} \hat{\psi}^{\dagger}(\mathbf{r}) O(\mathbf{r}) \hat{\psi}(\mathbf{r}), \tag{A.11}$$

where $O(\mathbf{r})$ is the corresponding first quantization operator, while two-body operators are written as

$$\hat{V} = \int d\mathbf{r} \int d\mathbf{r}' \hat{\psi}^{\dagger}(\mathbf{r}) \hat{\psi}^{\dagger}(\mathbf{r}') V(\mathbf{r}, \mathbf{r}') \hat{\psi}(\mathbf{r}') \hat{\psi}(\mathbf{r}). \tag{A.12}$$

A.3 Introduction to Feynman Diagrams

Feynman diagrams are a method to pictorially represent terms occurring in many-body perturbation theory. The terms represented by Feynman diagrams are a consequence of the Gell-Mann and Low theorem [68], which connects the interacting ground state to the non-interacting ground state. The result is an expression for the time-ordered Green's function as [63]

$$iG_{\alpha\beta}(\mathbf{r}, \mathbf{r}', t, t') = \sum_{n=0}^{\infty} (-i)^n \frac{1}{n!} \int dt_1 \cdots \int dt_n \frac{\left\langle \Phi_0 | T[\hat{V}(t_1) \ldots \hat{V}(t_n) \hat{\psi}_{\alpha}(\mathbf{r}, t) \hat{\psi}_{\beta}(\mathbf{r}', t')] | \Phi_0 \right\rangle}{\langle \Psi_0 | \Psi_0 \rangle}, \tag{A.13}$$

where $\hat{V}(t_n)$ are Coulomb operators, $\hat{\psi}_{\alpha}(\mathbf{r}, t)$ are the field operators defined in Section A.2, $|\Phi_0\rangle$ are non-interacting ground states and $|\Psi_0\rangle$ are interacting ground states. As shown in Eq. (A.12), the Coulomb operators involve a product of four field operators. Thus, each term in Eq. (A.13) contains a total of $2 + 4n$ field operators. Writing the Coulomb potential as $U_{ijkl}(\mathbf{r}, \mathbf{r}', t, t') = \delta_{ij} \delta_{kl} V(\mathbf{r}, \mathbf{r}') \delta(t - t')$, we obtain (ignoring the denominator)

$$\begin{aligned} iG_{\alpha\beta}(\mathbf{r}, \mathbf{r}', t, t') = {} & iG^0_{\alpha\beta}(\mathbf{r}, \mathbf{r}', t, t') \\ & + (-i) \sum_{ijkl} \int dt_1 dt_2 d\mathbf{r}_1 d\mathbf{r}_2 U_{ijkl}(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2) \\ & \times \left\langle \Phi_0 | T[\hat{\psi}_i(\mathbf{r}_1, t_1) \hat{\psi}_j(\mathbf{r}_2, t_2) \hat{\psi}_k(\mathbf{r}_2, t_2) \hat{\psi}_l(\mathbf{r}_1, t_1) \hat{\psi}_{\alpha}(\mathbf{r}, t) \hat{\psi}_{\beta}(\mathbf{r}', t')] | \Phi_0 \right\rangle \\ & + \ldots, \end{aligned} \tag{A.14}$$

where $G^0_{\alpha\beta}(\mathbf{r},\mathbf{r}',t,t')$ is the Green's function of the non-interacting system. Wick [191] showed that each of the terms can be expressed as a sum over terms involving $2n+1$ non-interacting Green's functions [63]. Furthermore, specific terms in Eq. (A.14) can be factored to cancel the denominator in Eq (A.13). The remaining terms can be represented as Feynman diagrams, where each directed line corresponds to a non-interacting Green's function, and each wavy line corresponds to the Coulomb interaction. For example, Eq. (A.14) is given by the Feynman diagrams in Fig. A.1, where the second and third terms correspond to the surviving first-order terms in Eq. (A.14).

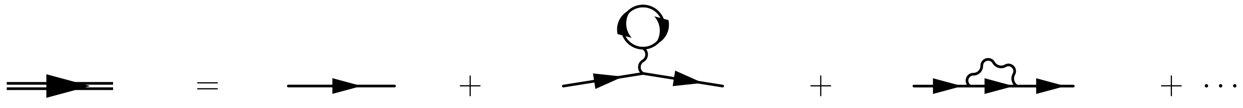


Figure A.1: Diagrammatic expansion of the interacting Green's function in terms of the non-interacting Green's function and the Coulomb interaction.

The terms in Feynman diagrams are evaluated using the following set of rules [63]:

1. Each interaction vertex is labelled with a space-time point $\mathbf{r}, t$ and incoming and outgoing spin indices i and j.
2. For each interaction vertex, a sum over incoming and outgoing spin-indices is required.
3. A factor of $(-1)^P$ has to be added to each term, where P is the number of Green's function lines closing on themselves.
4. Another factor of $(i)^n$ is required for each n-th order term.
5. A Green's function with two equal space-time points is interpreted as $G^0_{\alpha\beta}(\mathbf{r},\mathbf{r},t,t^+)$.
6. Finally, an integration over all internal space-time variables has to be performed.

Using these rules, the second diagram in Fig. A.1 evaluates to

$$(-i)\sum_{ijkl}\int dt_1 dt_2 d\mathbf{r}_1 d\mathbf{r}_2 G^0_{\alpha i}(\mathbf{r},\mathbf{r}_1,t,t_1)U_{ijkl}(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2)G^0_{kl}(\mathbf{r}_2,\mathbf{r}_2,t_2,t_2^+)G^0_{j\beta}(\mathbf{r}_1,\mathbf{r}',t_2,t'). \tag{A.15}$$

This expression can be simplified by noting that

$$iG^0_{\alpha\beta}(\mathbf{r},\mathbf{r},t,t^+) = -\frac{i}{2}\delta_{\alpha\beta}\rho^0(\mathbf{r}), \tag{A.16}$$

where $\rho^0(\mathbf{r})$ is the electron density of the non-interacting system. Furthermore, the Coulomb potential is given by

$$U_{ijkl}(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2) = \delta_{ij}\delta_{kl}V(\mathbf{r}_1,\mathbf{r}_2)\delta(t_1-t_2), \tag{A.17}$$

and the Green's functions are usually also diagonal in the spin indices. We thus obtain

$$\begin{aligned}\frac{1}{2}\sum_{ijkl}\int dt_2 d\mathbf{r}_1 d\mathbf{r}_2 \delta_{\alpha i}\delta_{ij}\delta_{kl}\delta_{kl}\delta_{j\beta}G^0(\mathbf{r},\mathbf{r}_2,t,t_2)V(\mathbf{r}_1,\mathbf{r}_2)\rho^0(\mathbf{r}_2)G^0(\mathbf{r}_1,\mathbf{r}',t_2,t')\\ = \int dt_2 d\mathbf{r}_1 d\mathbf{r}_2 \delta_{\alpha\beta}G^0(\mathbf{r},\mathbf{r}_1,t,t_2)V(\mathbf{r}_1,\mathbf{r}_2)\rho^0(\mathbf{r}_2)G^0(\mathbf{r}_1,\mathbf{r}',t_2,t').\end{aligned} \tag{A.18}$$

The integral involving the electron density can be identified with the Hartree potential [63]

$$V_{\text{Hartree}}(\mathbf{r}_1) = \int d\mathbf{r}_2 V(\mathbf{r}_1,\mathbf{r}_2)\rho^0(\mathbf{r}_2). \tag{A.19}$$

Similarly, we can evaluate the third term in Fig. (A.1) as

$$-i\delta_{\alpha\beta}\int dt_1 dt_2 d\mathbf{r}_1 d\mathbf{r}_2 G^0(\mathbf{r},\mathbf{r}_1,t,t_1)V(\mathbf{r}_1,\mathbf{r}_2)\delta(t_1-t_2)G^0(\mathbf{r}_1,\mathbf{r}_2,t_1,t_2)G^0(\mathbf{r}_2,\mathbf{r},t_2,t'). \tag{A.20}$$

Appendix B

Further details on neural networks

B.1 Backpropagation

The backpropagation algorithm, initially introduced by Rumelhart et al. [156], is a method for adjusting the weights of a neural network, such that they approach the minimum of some loss function E, which computes an error metric between the output of the neural network model $\mathbf{f}$ and the target quantity $\mathbf{y}$. Since neural networks are highly non-linear functions, an analytical solution cannot be found for this optimization problem. Thus other methods, such as gradient descent, where the weights of the neural network are adjusted in the direction of the negative gradient of the loss function with respect to the weights $-\frac{\partial E}{\partial w_{ij}}$, are required. Here, I will consider the simplest case, where the loss function is computed for each data point individually, and the weights are adjusted accordingly. This procedure can be generalized for batches of data points [25], but for simplicity, I will avoid this generalization here.

To simplify the notation, we can absorb the biases of each layer into the weight matrix by adding a dummy output which is always equal to 1 to each layer [25]. The linear part of the layer operation in then given by

$$z_i^N = \sum_j w_{ij}^N o_j^{N-1}, \tag{B.1}$$

and the output of the model is given by

$$f_i = o_i^N = h(z_i^N), \tag{B.2}$$

for some non-linear function h. The loss function we use here is simply the squared error between the model output $\mathbf{f}_n$ for input $\mathbf{x}_n$ and the true target quantity $\mathbf{y}_n$

$$E_n = \frac{1}{2}\sum_i (y_{ni} - f_{ni})^2. \tag{B.3}$$

We now need to evaluate the required gradients using the chain rule

$$\frac{\partial E_n}{\partial w_{ij}^N} = \frac{\partial E_n}{\partial z_i^N}\frac{\partial z_i^N}{\partial w_{ij}^N}. \tag{B.4}$$

The second part of the above equation is simply given by the output of the previous layer

$$\frac{\partial z_i^N}{\partial w_{ij}^N} = o_i^{N-1}. \tag{B.5}$$

Furthermore, assuming the output layer is linear, the first partial derivative is

$$\frac{\partial E_n}{\partial z_i^N} = -(y_{ni} - f_{ni}) \equiv \delta_i^N. \tag{B.6}$$

The crucial step in the backpropagation algorithm is the evaluation of the quantity $\frac{\partial E_n}{\partial z_i^M}$ for hidden layers. For the layer $N-1$ we use the chain rule for partial derivatives [25], to obtain

$$\frac{\partial E_n}{\partial z_i^{N-1}} = \sum_k \frac{\partial E_n}{\partial z_k^N}\frac{\partial z_k^N}{\partial z_i^{N-1}} = \sum_k \delta_k^N \frac{\partial z_k^N}{\partial z_i^{N-1}} \equiv \delta_i^{N-1}. \tag{B.7}$$

Finally the derivative $\frac{\partial z_k^N}{\partial z_i^{N-1}}$ is given by

$$\frac{\partial z_k^N}{\partial z_i^{N-1}} = h'(z_i^{N-1}) w_{ki}^N. \tag{B.8}$$

Thus, we have an expression for the quantity δ_i^{N-1} in terms of δ_i^N

$$\delta_i^{N-1} = h'(z_i^{N-1}) \sum_k w_{ki}^N \delta_k^N, \tag{B.9}$$

where the sum runs over all k that are connected to the hidden unit j [25]. This suggests the following procedure for updating the weights [25]:

1. Propagate the input $\mathbf{x}_n$ forward through the network.
2. Compute the quantity δ_i^N and the gradient $\frac{\partial E_n}{\partial w_{ij}^N}$ for the final layer.
3. Use the computed δ_i^N's to compute the δ_j^{N-1}'s and gradients $\frac{\partial E_n}{\partial w_{ij}^{N-1}}$ for the previous layers.
4. Once all gradients are computed, update the weights with $-\lambda\frac{\partial E_n}{\partial w_{ij}^N}$, where λ is a positive constant.
5. Repeat steps 1-4 for the next input $\mathbf{x}_n$.