universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Self-Consistent GW Calculations for Solids

verfasst von / submitted by

Manuel Grumet, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2017 / Vienna, 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 876
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Physik
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Georg Kresse

Abstract

The GW approximation is a method for *ab initio* simulations of many-electron systems, based on the formal framework of the Green's function formalism and on Hedin's equations. It goes beyond mean field methods such as the Hartree-Fock approximation (HF) and density functional theory (DFT) and constitutes an improvement for calculating excitation spectra and band structures. GW calculations can be done with different levels of self-consistency, leading to the non-self-consistent G_0W_0 method, the partially self-consistent GW_0 method and the fully self-consistent GW method.

In this thesis, fully self-consistent GW calculations were carried out for a number of solid-state systems using the Vienna *Ab Initio* Simulation Package (VASP). All of the systems used were gapped systems, i.e. semiconductors or insultators. The results are compared to G_0W_0 and quasiparticle GW (QPGW) calculations.

Furthermore, this thesis also deals with two specific problems in the implementation of GW calculations. The first of these is the problem of the calculation of the so-called head of the dielectric matrix, which arises because of the singularity of the Coulomb kernel in reciprocal space. This was solved using a fit based on data close to the singularity. The second was the problem of analytic continuation of the the self-energy Σ from the imaginary axis to the real axis. This was solved by using Padé fitting.

Zusammenfassung

Die GW-Näherung ist eine Methode für *ab-initio*-Simulationen von Vielelektronensystemen, die auf dem formalen Gerüst des Green-Funktionen-Formalismus und auf den Hedin-Gleichungen beruht. Sie geht über Molekularfeld-Methoden wie die Hartree-Fock-Näherung (HF) und die Dichtefunktionaltheorie (DFT) hinaus und stellt eine Verbesserung für die Berechnung von Anregungsspektren und Bandstrukturen dar. GW-Berechnungen können mit verschiedenen Stufen von Selbstkonsistenz durchgeführt werden, was zur nicht selbstkonsistenten G_0W_0-Methode, zur teilweise selbstkonsistenten GW_0-Methode und zur vollständig selbstkonsistenten GW-Methode führt.

In dieser Arbeit wurden vollständig selbstkonsistente GW-Berechnungen für einige Festkörper-Systeme mit dem Vienna *Ab Initio* Simulation Package (VASP) durchgeführt. Alle verwendeten Systeme weisen Bandlücken auf, d.h. es sind Halbleiter oder Isolatoren. Die Ergebnisse werden mit G_0W_0-Berechnungen und Quasiteilchen-GW-Berechnungen (QPGW) verglichen.

Des Weiteren behandelt diese Arbeit zwei spezifische Probleme der Implementierung von GW-Berechnungen. Das erste dieser Probleme ist das Problem der Berechnung des sog. Kopfs der dielektrischen Matrix, das auf die Singularität des Coulomb-Kerns im reziproken Raum zurückzuführen ist. Dieses Problem wurde durch Fitting auf Basis von Daten in der Nähe der Singularität gelöst. Das zweite Problem ist das Problem der analytischen Fortsetzung der Selbstenergie Σ von der imaginären Achse zur reellen Achse. Dieses Problem wurde durch Padé-Fitting gelöst.

Acknowledgements

I would like to express my sincere thanks to my supervisor Georg Kresse, who has done a great job guiding me through this thesis over the past months. I have really learned a lot from him. I am also truly thankful to my co-supervisor Peitao Liu, who has taken so much time for me and helped me in so many ways.

I would also like to thank the rest of the Computational Material Physics research group for many fruitful discussions, both scientific and private. In particular, I want to mention Stefan Riemelmoser and Benjamin Ramberger, who have both been my colleagues and friends for many years while studying physics. Furthermore, I want to thank Michael Pöltl and Doris Hecht-Aichholzer for everything they have done for me.

Contents

1 Introduction

Ab initio simulations of solid-state systems can be performed using a variety of different numerical methods. These methods differ primarily in how well many-body effects resulting from the interaction of electrons with each other are captured. Mean-field methods, such as the Hartree-Fock approximation (HF) and density functional theory (DFT) use an independent-particle picture and assume that the full many-electron wave function can be written in terms of single-electron orbitals. These methods can successfully predict ground-state properties of solids, but they suffer from systematic errors when calculating excited state energies and the resulting band structures. This is especially true for the band gaps between valence and conduction bands in semiconductors and insulators, which are overestimated by HF and underestimated by DFT.

The GW approximation goes beyond mean field methods and can yield better results for the band structures of solids. The central quantity in the GW approximation is the GW self-energy $\Sigma = GW$, where G is the single-particle Green's function and W is the screened electron-electron interaction. G describes the propagation of an added electron or an added electron hole (i.e., a removed electron) in a many-electron system. These added electrons or holes travel through the system surrounded by polarization clouds of opposite charge, forming so-called quasiparticles. This concept also has a direct connection to experiments, as the quasiparticle spectral function, which is directly related to the Green's function, can be measured by photoemission spectroscopy (PES) and inverse photoemission spectroscopy (IPES).

There are several variations of GW calculations which differ in their degree of self-consistency. This includes the G_0W_0 method, which is done completely without self-consistency, the GW_0 method, which is partially self-consistent, and the GW method (sometimes also called scGW), which is fully self-consistent (see section 2.6 for details). Furthermore, there is also the so-called quasiparticle GW method (QPGW), which is fully self-consistent, but makes additional simplifications [1]. Of these variations, G_0W_0 is computationally the cheapest and also the most widely used. Fully self-consistent GW

has in the past been studied for atoms and molecules [2], for the uniform electron gas [3], and for K and Si [4], and has recently been applied to more solids [5,6].

In this thesis, quasiparticle energies for 15 solid-state materials are computed with the fully self-consistent GW method and compared to results obtained from G_0W_0 and QPGW.

Chapter 2 discusses the theoretical background for the methods used in this thesis. A brief overview is given for the Hartree-Fock approximation and density functional theory. After that, the theory of Green's functions, Hedin's equations, the GW approximation, and self-consistency are discussed in more detail.

Chapters 3 and 4 deal with two particular problems that occur in the implementation of self-consistent GW calculations. Specifically, chapter 3 deals with the problem of the calculation of the head of the dielectric matrix ϵ, which arises in the long-wavelength limit $\mathbf{q} \to 0$ due to the singularity of the reciprocal-space Coulomb potential. This problem is solved by fitting and extrapolating from finite values for $\mathbf{q}$. Chapter 4 deals with the problem of analytic continuation of Σ, i.e. obtaining values along the real axis from values along the imaginary axis. This is solved by using Padé fitting methods.

Finally, chapter 5 describes the details of the calculations that were performed and presents the results obtained from them, comparing the different methods to each other. All the materials considered here are gapped materials, i.e. either semiconductors or insulators. The list of materials is a subset of the materials considered in a previous paper by Klimeš, Kaltak and Kresse, which used the G_0W_0 method [7].

2 Theoretical Background

2.1 The Many-Electron Problem

A system of N non-relativistic electrons in an external potential V_{ext} can be described by a wave function $\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ which fulfils a Schrödinger equation. In atomic units, where the physical constants $\hbar$ and e and the electron mass m_e are all set to 1, the Hamiltonian of such a system can be written as

$$\hat{H} = \sum_{i=1}^{N} \left[-\frac{1}{2} \nabla^2 \mathbf{r}_i \right] + \sum_{i=1}^{N} V_{\text{ext}}(\mathbf{r}_i) + \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{2.1}$$

Here, the first term being summed over is the kinetic energy of the electrons, the second term is the external potential, and the third term arises from the pair-wise Coulomb interaction of the electrons with each other.

In a solid, the electrons also interact with the atomic nuclei, so in principle the Hamiltonian would also have to contain terms for the nuclei. However, due to the fact that the nuclei have much larger masses than the electrons, their motion occurs on larger time scales. Therefore, it is possible to view the nuclei as static and only consider the motion of the electrons. This is known as the Born-Oppenheimer approximation. In this approximation, the effect of the nuclei can be treated as an external potential in which the electrons move, given as a sum over Coulomb potentials

$$V_{\text{ext}}(\mathbf{r}) = -\frac{1}{2} \sum_{i=1}^{M} \frac{Z_i}{|\mathbf{r} - \mathbf{R}_i|}, \tag{2.2}$$

where M is the total number of nuclei, Z_i are the nuclear charges and $\mathbf{R}_i$ are the nuclear positions. Therefore, the many-electron Hamiltonian within the Born-Oppenheimer approximation can be written in the form of Eq. (2.1). Further simplification is often

possible by also assuming the electrons in the inner atomic shells to be fixed, so that their effect can be absorbed in the definition of $V_{\text{ext}}(\mathbf{r})$ as well, leaving only valence electrons to be considered in the many-electron problem. This is called the frozen core approximation.

In principle, the many-electron wave function $|\Psi\rangle$ is an eigenfunction of the Hamiltonian in Eq. (2.1), i.e. a solution to the many-body Schrödinger equation

$$\hat{H}\,|\Psi\rangle = E\,|\Psi\rangle\,. \tag{2.3}$$

However, this equation can be solved exactly only for very simple systems, such as the hydrogen atom. For more complex systems, the presence of the Coulomb interaction between the electrons makes analytic solutions impossible. Direct numerical solutions are possible, but impractical for almost all cases due to memory and computation time requirements of $\mathcal{O}(e^N)$. Therefore, a number of numerical methods have been developed to find approximate solutions. One category of such methods are mean field methods, which work with independent-particle equations. These include the Hartree-Fock approximation (HF) and density functional theory (DFT). Another category are methods based on Green's functions, which includes the GW approximation.

2.2 The Hartree-Fock Approximation

In the Hartree-Fock approximation, the effect that electrons in a many-electron system have on each other is approximated as an effective potential acting on each electron. Thus, the many-electron Hamiltonian in the Hartree-Fock approximation can be written as a sum of single-electron Hamiltonians. This implies that the wave function $|\Psi\rangle$ can be written in terms of single-electron wave functions $|\chi_i\rangle$. The simplest possible form for such a wave function is a so-called Hartree product

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \chi_1(\mathbf{x}_1) \cdots \chi_N(\mathbf{x}_N). \tag{2.4}$$

Here, the χ_i are spin orbitals, i.e. single-electron wave functions that depend on three spatial coordinates $\mathbf{r}$ and a spin coordinate σ, and $\mathbf{x_i}$ is an abbreviation for $(\mathbf{r}_i, \sigma_i)$. Spin orbitals are complete descriptions of the state of a single electron. They can always be

written as a product of a spatial orbital $\psi(\mathbf{r})$ and one of two orthonormal spin eigenfunctions $\alpha(\sigma)$ and $\beta(\sigma)$ that correspond to the two possible spin states,

$$\chi(\mathbf{x}) = \begin{cases} \psi(\mathbf{r})\alpha(\sigma) \\ \text{ or} \\ \psi(\mathbf{r})\beta(\sigma). \end{cases} \tag{2.5}$$

The Hartree product is a completely uncorrelated wave function, in the sense that for any two electrons, the probability $|\Psi(\mathbf{x}_1, \mathbf{x}_2)|^2 \, \mathrm{d}\mathbf{x}_1 \, \mathrm{d}\mathbf{x}_2$ of finding electron one in a volume element $\mathrm{d}\mathbf{x}_1$ and electron two in a volume element $\mathrm{d}\mathbf{x}_2$ is a simple product $|\chi_1(\mathbf{x}_1)|^2 \, \mathrm{d}\mathbf{x}_1 \, |\chi_2(\mathbf{x}_1)|^2 \, \mathrm{d}\mathbf{x}_2$. However, this does not accurately describe the behavior of electrons. Due to their fermionic nature, electrons are subject to the Pauli exclusion principle, which states that two electrons cannot occupy the same quantum state. A more mathematical formulation of the Pauli principle is the anti-symmetry principle, which states that the many-electron wavefunction must be anti-symmetric under the exchange of any two electrons. This means that

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_n) = -\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_n) \tag{2.6}$$

for any $\mathbf{x}_i$ and $\mathbf{x}_j$ (from this formulation, it immediately follows that Ψ is zero if $\mathbf{x}_i$ and $\mathbf{x}_j$ are the same, which is the physical content of the Pauli principle). Many-electron wave functions which have this property cannot be written as simple Hartree products. Instead, they can be written in the form of Slater determinants

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix}. \tag{2.7}$$

A Slater determinant is an antisymmetric wave function, but still composed of independent single-electron spin orbitals. For two electrons of opposite spin, it is uncorrelated, like the Hartree product. But for two electrons of the same spin, the combined probability of finding electron one in $\mathrm{d}\mathbf{x}_1$ and electron two in $\mathrm{d}\mathbf{x}_2$ is

$$\left|\Psi(\mathbf{x}_1, \mathbf{x}_2)\right|^2 \mathrm{d}\mathbf{x}_1 \, \mathrm{d}\mathbf{x}_2 = \frac{1}{2}\left|\chi_1(\mathbf{x}_1)\chi_2(\mathbf{x}_2) - \chi_1(\mathbf{x}_2)\chi_2(\mathbf{x}_1)\right|^2 \mathrm{d}\mathbf{x}_1 \, \mathrm{d}\mathbf{x}_2, \tag{2.8}$$

which is not a simple product of single-particle probabilities. Therefore, the Slater determinant includes some correlation between electrons of the same spin.

The idea of the Hartree-Fock approximation is to find the ground state $|\Psi_0\rangle$ of a many electron system by minimizing the energy

$$E = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle \,, \tag{2.9}$$

under the assumption that $|\Psi_0\rangle$ can be written as a Slater determinant. From this minimization requirement, a Schrödinger-like equation for the single-electron orbitals can be derived, using calculus of variations (see [8] for a full derivation). The result is

$$\left[-\frac{1}{2}\nabla^2 \mathbf{r} + V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) - V_{\text{x}}(\mathbf{r}) \right] \chi_i(\mathbf{x}) = \epsilon_i \chi_i(\mathbf{x}). \tag{2.10}$$

These equations are called the Hartree-Fock equations. Here, the operator acting on $\chi_i(\mathbf{x})$, i.e. the sum of the terms in square brackets, is called the Fock operator. The first and the second terms in the Fock operator are the kinetic energy and the external potential. The last two terms together constitute an additional effective potential for each electron. The third term, V_{H}, is called the Hartree potential and is defined as

$$V_{\text{H}}(\mathbf{r}) = \sum_j \int \mathrm{d}\mathbf{r}' \, \frac{|\chi_j(\mathbf{r}')|^2}{|\mathbf{r}' - \mathbf{r}|}. \tag{2.11}$$

The Hartree potential is the effective Coulomb potential of all the electrons in the system. It can also be written as

$$V_{\text{H}}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \, \frac{n(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|}, \tag{2.12}$$

where $n(\mathbf{r})$ is the electron density, defined as the sum of the the spatial probability densities of all occupied orbitals,

$$n(\mathbf{r}) = \sum_j^{\text{occ.}} |\chi_j(\mathbf{r})|^2 = 2 \sum_j^{\text{occ.}} |\psi_j(\mathbf{r})|^2, \tag{2.13}$$

where the factor two arises arises from the two possible spin states. The fourth term in Eq. (2.10), $V_{\mathrm{x}}(\mathbf{r})$, is called the exchange potential. It is a consequence of the Pauli principle and expresses the fact that, even within the Slater determinant ansatz, there are still some additional exchange effects between electrons of the same spin. It is given as

$$V_{\mathrm{x}}(\mathbf{r})\,|\chi_i\rangle = -\sum_j \int \mathrm{d}\mathbf{r}'\,\frac{\chi_j^*(\mathbf{r}')\chi_i(\mathbf{r}')}{|\mathbf{r}'-\mathbf{r}|}\,|\chi_j\rangle\,. \tag{2.14}$$

Note that the definition of the Hartree potential in Eq. (2.12) contains a term where $i = j$, i.e. a Coulomb interaction term of an electron with itself. This self-interaction is unphysical. However, the definition of the exchange potential in Eq. (2.14) contains a term of the exact same form, but with opposite sign. Thus, the two unphysical terms cancel each other out and the summation can indeed run over all j in both definitions.

Since the Fock operator depends on its own eigenfunctions, the Hartree-Fock equations have to be solved iteratively. In each step, the Fock operator is calculated from the orbitals that have been obtained in the previous step, until self consistency is reached. In practice, the orbitals are represented as linear combinations of a finite set of basis functions $\phi_i(\mathbf{r})$. The Fock operator can then be written in matrix form by calculating the matrix elements $F_{ij} = \langle\phi_i|\hat{F}|\phi_j\rangle$. The iterative solution of the Hartree-Fock equations can therefore be implemented using matrix operations. This leads to $2K$ spin orbitals as a solution, where K is the number of basis functions used. These orbitals are orthogonal to each other and can themselves be used as a basis set for further calculations, e.g. for GW calculations. This is called the Hartree-Fock basis.

Of the $2K$ orbitals, the N with the lowest eigenvalues are occupied in the ground state, whereas the rest is unoccupied. The many-electron wave function of the Hartree-Fock ground state is the Slater determinant of the occupied orbitals. It is also possible to form other Slater determinants, involving orbitals that are not occupied in the ground state. These can be seen as approximations to excited states of the N-electron system.

Ionization potentials and electron affinities can be obtained from Hartree-Fock calculations using Koopmans' theorem [9]. This theorem states that the ionization potential, i.e. the energy required to produce a system with $N-1$ instead of N electrons is equal (up to a sign) to the energy of the highest occupied orbital, and the electron affinity, i.e. the energy required to produce a system with $N+1$ electrons, is equal (up to a sign) to the energy of the lowest unoccupied orbital. The derivation assumes that the

orbitals stay the same under addition or removal of an electron (this is known as the frozen orbitals approximation). However, the accuracy of these values obtained using the Hartree-Fock approximation is often bad, especially for electron affinities.

2.3 Density Functional Theory

The Hartree-Fock approximation neglects all correlations between electrons except for the exchange effects required by the Pauli principle. One possible way to go beyond the Hartree-Fock approximation is the Kohn-Sham ansatz within density functional theory (DFT), which includes both exchange and further correlation effects in the form of an effective potential V_{xc}, which is a semi-local functional of the electron density $n(\mathbf{r})$. In DFT, the electron density is thus the central quantity. This is physically appealing since the electron density, unlike the full wave function, is in principle a directly measurable quantity and can be measured in X-ray scattering experiments.

The formal justification for this central role of the density is provided by the following two foundational theorems by Hohenberg and Kohn [10].

Theorem: Hohenberg-Kohn I. *For any system of interacting particles in an external potential $V_{\mathrm{ext}}(\mathbf{r})$, there exists exactly one functional $\mathcal{F} : \mathcal{C}^\infty(\mathbb{R}^3) \to \mathbb{R}$ such that $V_{\mathrm{ext}}(\mathbf{r}) = \mathcal{F}[n_0(\mathbf{r})] + c$, where $n_0(\mathbf{r})$ is the ground state density and c is an arbitrary real number. Thus, up to a constant, $V_{\mathrm{ext}}(\mathbf{r})$ is uniquely determined by the ground state density.*

As a corollary, it follows that the entire Hamiltonian of the system is uniquely determined up to a constant by the ground state density. Therefore, the ground state wave function and all observable properties of the system are in principle unique functionals of the density. However, the theorem only asserts the existence of these functionals, it does not provide an explicit formula for any of them.

Theorem: Hohenberg-Kohn II. *There exists exactly one functional $\mathcal{E} : \mathcal{C}^\infty(\mathbb{R}^3) \to \mathbb{R}$ such that $\mathcal{E}[n_0(\mathbf{r})]$ is the ground state energy and the variational property*

$$\frac{\delta \mathcal{E}[n]}{\delta n} = 0. \tag{2.15}$$

is satisfied at the ground state density $n_0(\mathbf{r})$.

This functional is called the total energy functional. The variational property means that the total energy takes on an extremal value at the exact ground state density (in

practice, since energies are positive and unbounded, this is always a minimum). Again, the theorem only guarantees the existence of a functional, without showing an explicit way to compute or derive it. An exact explicit formula for this functional is not known.

In Kohn-Sham DFT [11], it is assumed that the true ground state density of the many-electron system can be approximated as the density of a Slater determinant formed from one-electron orbitals. This is similar to the ansatz made in the Hartree-Fock approximation. Minimizing the total energy under this assumption (see [12] for a full derivation) leads to the Kohn-Sham equations

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r})\right]\chi_i(\mathbf{x}) = \epsilon_i\chi_i(\mathbf{x}). \tag{2.16}$$

Like in the Hartree-Fock case, there exists one such equation for each orbital. The terms in the brackets are, from left to right, the kinetic energy, the external potential, the Hartree potential and the exchange-correlation potential. The first three of these terms can be calculated exactly in terms of noninteracting one-electron orbitals. The exchange-correlation potential is the most difficult term. It describes all many-electron effects beyond the Hartree term. Physically, these effects create a so-called exchange-correlation hole around each electron, i.e. a region of space where the probability of finding another electron is very low (note that this also includes electrons of opposite spin). Unlike the exchange potential in the Hartree-Fock approximation (Eq. (2.14)), V_{xc} is a local potential, since the ground state wave function depends only on the electron density. An exact expression for V_{xc} as a functional of $n_0(\mathbf{r})$ is not known. Therefore, approximations have to be used in practical calculations.

The simplest approximation for V_{xc} is the local density approximation (LDA), which assumes that the system can be approximated locally by a homogeneous electron gas (HEG) with the same density, for which the exact exchange-correlation energy can be calculated. Approximations which go beyond the LDA often also incorporate the gradient of the density in addition to its local value at every point. These are collectively known as generalized gradient approximations (GGA).

Within any chosen approximation for V_{xc}, the Kohn-Sham equations can be solved self-consistently, like the Hartree-Fock equations. This yields a set of orbitals with different energy eigenvalues. These eigenvalues can be interpreted as excitation energies. However, there is no formal justification for this, with the only exception being the energy of the highest occupied orbital, which is equal (up to a sign) to the ionization potential.

For solids, where $V_{\text{ext}}(\mathbf{r})$ is periodic, the spatial orbitals obtained in DFT or Hartree-Fock calculations can always be written in the form

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}), \tag{2.17}$$

where $u_{n,\mathbf{k}}(\mathbf{r})$ are functions with the same periodicity as $V_{\text{ext}}(\mathbf{r})$, i.e. with the same periodicity as the crystal lattice. This is known as Bloch's Theorem. The orbitals can therefore be indexed with a band index n and a wave vector $\mathbf{k}$. It is sufficient to consider only wave vectors that lie in the first Brillouin zone of the crystal's reciprocal lattice, as wave vectors outside the first Brillouin zone do not yield any additional solutions. In practice, the Brillouin zone is sampled at finitely many $\mathbf{k}$-points and the Kohn-Sham equations can be solved for each point separately. Furthermore, the periodicity of the functions $u_{n,\mathbf{k}}(\mathbf{r})$ implies that they can be expanded in terms of plane waves with wave vectors $\mathbf{G}$ that lie on the lattice points of the reciprocal lattice. These plane waves are orthogonal to each other and can therefore be used as a basis set. The size of the basis set in any particular calculation can be chosen by selecting an energy cutoff E_{cut} and discarding waves with energies higher than E_{cut}. This also determines the maximum number of bands that can be obtained. E_{cut} is therefore an important parameter determining both the accuracy and the computational cost of a calculation.

In practical calculations with both Hartree-Fock and DFT, some simplifications can be achieved by replacing the potential V_{ext} from the nuclei with so-called pseudopotentials. These are effective potentials which yield wave functions that are correct outside of a certain cutoff radius r_{cut}, but which do not have the singularity of the true Coulomb potential. This can lead to a significant reduction in the number of plane waves needed in a calculation and therefore a reduction in computational cost.

The band structure of a solid is the set of band energies $\epsilon_n(\mathbf{k})$ as functions of the wave vector $\mathbf{k}$. For semiconductors and insulators, there is always a gap between the energies of the highest occupied and the lowest unoccupied band. These band gaps are typically underestimated by DFT calculations. This is related to the fact that the true exchange-correlation energy of such a system changes discontinuously if the number of electrons is increased continuously from below N to above N (note that the number of electrons can be changed continuously by scaling $n(\mathbf{r})$ with a factor). The existing approximations for the exchange-correlation functional, however, have no such discontinuity, and therefore lack this "jump" in energy, leading to the underestimation of the band gaps in DFT.

It is also possible to combine the HF and DFT approaches into a hybrid approach by weighted averaging of the terms V_{x} from HF and V_{xc} from LDA and/or GGA. This is known as the hybrid functional approach. This approach improves the band gaps, since the results lie in between the HF results (which tend to overestimate band gaps) and the DFT results. See [13] for a detailed description of the hybrid functional approach.

2.4 Green's Function Formalism

The basic quantity in the GW method is the single-particle Green's function G. It describes the propagation of an added electron or a hole, i.e. a removed electron, in a many-electron system. These situations are experimentally realizable in photoemission spectroscopy (PES) and inverse photoemission spectroscopy (IPES) [14]. In PES, a stream of photons with energy ω hits a material and knocks electrons out of occupied states, creating holes in the process. In IPES, the material is hit by a stream of electrons. The incoming electrons can occupy previously unoccupied states, releasing photons in the process. Therefore, PES yields information about occupied states, whereas IPES yields information about unoccupied states. Because of Coulomb repulsion, both holes and added electrons will accumulate a polarization cloud of the opposite charge around them, thereby forming so-called quasiparticles. Because of this cloud, quasiparticles interact only weakly with each other. Furthermore, they differ from bare (i.e. non-clouded) particles by having different energy levels and finite lifetimes.

The single-particle Green's function G is defined such that $iG(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2)$ is the probability amplitude for the propagation of an electron added to a system from $(\mathbf{r}_1, t_1)$ to $(\mathbf{r}_2, t_2)$ if $t_2 > t_1$, and the probability amplitude for the propagation of a hole created in the system at $(\mathbf{r}_1, t_1)$ to $(\mathbf{r}_2, t_2)$ if $t_2 < t_1$. More generally, the definition can be extended to allow states specified by an arbitrary set of arbitrary quantum numbers k. This is just a more abstract notation. In practice, k will represent either a position vector $\mathbf{r}$ or a wave vector $\mathbf{k}$. Furthermore, in a system whose Hamiltonian does not explicitly depend on time, only the time difference $t_2 - t_1$ is relevant. The single-particle Green's function can therefore be written as a function $G(k_1, k_2, t_2 - t_1)$.

The probability amplitudes for the propagation of electrons and holes can also be considered separately, using the notation G^+ for electrons and G^- for holes. $G^+(k_1, k_2, t_2 - t_1)$, which is also called the retarded Green's function, is defined to be zero for $t_2 < t_1$. $G^-(k_1, k_2, t_2 - t_1)$, which is also called the advanced Green's function, is defined to be

zero for $t_2 > t_1$. Physically, this means that electrons propagate forward in time, while holes can be thought of as electrons propagating backward in time.

First, consider the case of a single non-interacting particle. An understanding for the Green's function of this simple system can be be obtained from the time-dependent Schrödinger equation

$$\left(i \frac{\partial}{\partial t} - \hat{H} \right) \psi(\mathbf{r}, t) = 0. \tag{2.18}$$

For states that fulfil the time-dependent Schrödinger equation, time evolution introduces an exponential factor, i.e. if $\psi(\mathbf{r}, t_1) = \phi_{k_1}(\mathbf{r})$ is a solution at time t_1, then $\psi(\mathbf{r}, t_2) = \phi_{k_1}(\mathbf{r}) e^{-i\epsilon_{k_1}(t_2 - t_1)}$ is a solution at time t_2, where ϵ_{k_1} is the energy eigenvalue associated with ϕ_{k_1}. The probability amplitude for the non-interacting particle propagating from the state $\phi_{k_1}(\mathbf{r})$ at time t_1 to the new state $\phi_{k_2}(\mathbf{r})$ at time t_2, under the assumption that $t_2 > t_1$, can therefore be written as

$$\begin{aligned} & iG^+(k_1, k_2, t_2 - t_1) \\ & \quad = \Theta(t_2 - t_1) \left\langle \psi(\mathbf{r}, t_1) \,\middle|\, \psi(\mathbf{r}, t_2) \right\rangle \\ & \quad = \Theta(t_2 - t_1) e^{-i\epsilon_{k_1}(t_2 - t_1)} \delta(k_2 - k_1) \end{aligned} \tag{2.19}$$

where Θ is the Heaviside step function, which arises from the definition of G^+, and δ is the Dirac delta distribution. The last equality above is true under the assumption that the states ϕ_{k_1} and ϕ_{k_2} are orthogonal, which is always the case for position and momentum eigenstates. The result is a so-called fundamental solution of the time-dependent Schrödinger equation, which means that it has the property

$$\left(i \frac{\partial}{\partial t} - \hat{H} \right) G^+(k_1, k_2, t_2 - t_1) = \delta(k_2 - k_1) \delta(t_2 - t_1). \tag{2.20}$$

This can be proven immediately by inserting Eq. (2.19) into the left hand side of Eq. (2.20), using the facts that $(\epsilon_k - \hat{H}) e^{-i\epsilon_k (t_2 - t_1)} = 0$ and that $\delta(t_2 - t_1)$ is the derivative of $\Theta(t_2 - t_1)$. This property coincides precisely with the mathematical definition of Green's functions as fundamental solutions of differential equations, thus justifying the use of this terminology.

G^+ can be Fourier transformed to change its domain from the time difference $t_2 - t_1$ to a frequency parameter ω. However, the Fourier integral

$$\int_{-\infty}^{+\infty} \mathrm{d}(t_2 - t_1)\, \Theta(t_2 - t_1) e^{i\omega(t_2 - t_1)} e^{-i\epsilon_k (t_2 - t_1)} \tag{2.21}$$

is divergent because the integration yields a term of the form $e^{i(\omega - \epsilon_k)(t_2 - t_1)}$, which oscillates for $(t_2 - t_1) \to \infty$. To make the Fourier transform well defined, $G^+(k, t_2 - t_1)$ can be multiplied with an additional factor $e^{-\eta(t_2 - t_1)}$, where η is an infinitesimally small positive real number with the property that $e^{-\eta(t_2 - t_1)}$ goes to zero for $(t_2 - t_1) \to \infty$. The frequency-domain Green's function $G^+(k_1, k_2, \omega)$ is therefore defined as

$$iG^+(k_1, k_2, \omega) = \int_{-\infty}^{+\infty} \mathrm{d}(t_2 - t_1)\, \Theta(t_2 - t_1) e^{i\omega(t_2 - t_1)} e^{-i(\epsilon_{k_1} - \eta)(t_2 - t_1)}. \tag{2.22}$$

This can be evaluated to

$$G^+(k_1, k_2, \omega) = \frac{1}{\omega - \epsilon_{k_1} + i\eta} \delta(k_2 - k_1). \tag{2.23}$$

Note that this last expression is for G^+, not iG^+, since the factor i cancels with another i obtained in the integration.

Now, consider the case of a many-electron system with known single-electron orbitals. In this case, G^+ is the probability amplitude that, in the initial N-particle ground state of the system, an additional electron is created in state $\phi_{k_1}(\mathbf{r})$ at time t_1, propagates to $\phi_{k_2}(\mathbf{r})$ at time t_2, and is then annihilated, leaving the system again in the N-particle ground state. This can be expressed using the formalism of second quantization. In this formalism, the operators $\hat{c}^\dagger$ and $\hat{c}$ are introduced. The creation operator $\hat{c}^\dagger(k)$ creates an electron in the state ϕ_k, turning an N particle system into an $N + 1$ particle system. The annihilation operator $\hat{c}(k)$ removes an electron in the state ϕ_k, turning an N particle system into an $N - 1$ particle system. These operators fulfil the anticommutator relations

$$\{\hat{c}^\dagger(k_1), \hat{c}^\dagger(k_2)\} = \{\hat{c}(k_1), \hat{c}(k_2)\} = 0, \tag{2.24}$$

$$\{\hat{c}(k_1), \hat{c}^\dagger(k_2)\} = \delta(k_2 - k_1), \tag{2.25}$$

where the anticommutator $\{\hat{A}, \hat{B}\}$ is defined as $\hat{A}\hat{B} + \hat{B}\hat{A}$. Using this notation, the expression for G^+ in the many-electron system is

$$iG^+(k_1, k_2, t_2 - t_1) = \sum_n \left\langle \Psi_N \left| \hat{c}_n(k_2) e^{-i\hat{H}(t_2-t_1)} \hat{c}_n^\dagger(k_1) \right| \Psi_N \right\rangle, \tag{2.26}$$

where $|\Psi_N\rangle$ is the N particle ground state, but the summation index n goes over the orbitals of the $N+1$ particle system, since these are the orbitals that the added electron can occupy. Note that this is written in the Schrödinger picture (in the Heisenberg picture, the time dependence would be carried by the operators $\hat{c}^\dagger$ and $\hat{c}$). Eq. (2.28) can be Fourier transformed to frequency domain, again after introducing a factor $e^{-\eta(t_2-t_1)}$ with an infinitesimal η. This yields

$$G^+(k_1, k_2, \omega) = \sum_n \frac{\left\langle \Psi_N \left| \hat{c}_n(k_2) \hat{c}_n^\dagger(k_1) \right| \Psi_N \right\rangle}{\omega - \epsilon_{k_1} + i\eta}. \tag{2.27}$$

The hole Green's function G^- can be treated in exactly the same way as the electron Green's function G^+, with the two exceptions that annihilation happens before creation, which means that the order of the operators is reversed, and that the sign of the infinitesimal η is changed. This leads to

$$iG^-(k_1, k_2, t_2 - t_1) = \sum_n \left\langle \Psi_N \left| \hat{c}_n^\dagger(k_2) e^{-i\hat{H}(t_2-t_1)} \hat{c}_n(k_1) \right| \Psi_N \right\rangle \tag{2.28}$$

and

$$G^-(k_1, k_2, \omega) = \sum_n \frac{\left\langle \Psi_N \left| \hat{c}_n^\dagger(k_2) \hat{c}_n(k_1) \right| \Psi_N \right\rangle}{\omega - \epsilon_{k_1} - i\eta}, \tag{2.29}$$

where the summation index n runs over the orbitals of the $N-1$ particle system. The full Green's function G is a sum of G^+ and G^-. It can be written as

$$iG(k_1, k_2, t_2 - t_1) = \sum_n \left\langle \Psi_N \left| \hat{T}\{\hat{c}(k_1) e^{-i\hat{H}(t_2-t_1)} \hat{c}^\dagger(k_2)\} \right| \Psi_N \right\rangle \tag{2.30}$$

and

$$G(k_1, k_2, \omega) = \sum_n \frac{\left\langle \Psi_N \middle| \hat{T}\{\hat{c}(k_1)\hat{c}^\dagger(k_2)\} \middle| \Psi_N \right\rangle}{\omega - \epsilon_n - i\eta \,\mathrm{sgn}(\epsilon_n - \mu)}. \tag{2.31}$$

The last equation is called the Lehman representation of the Green's function. In both equations, n runs over the orbitals of both the $N+1$ particle system and the $N-1$ particle system. $\hat{T}$ is the time-ordering operator, defined as

$$\begin{aligned} \hat{T}\left\{\hat{c}(k_1, t_1)\hat{c}^\dagger(k_2, t_2)\right\} = \\ \Theta(t_2 - t_1)\,\hat{c}(k_1, t_1)\hat{c}^\dagger(k_2, t_2) + \\ \Theta(t_1 - t_2)\,\hat{c}^\dagger(k_2, t_2)\hat{c}(k_1, t_1), \end{aligned} \tag{2.32}$$

and μ is the chemical potential, defined such that it is larger than the ionization potential, but smaller than the electron affinity, thus always yielding the correct sign for $(\epsilon_n - \mu)$. The energies ϵ_n in Eq. (2.30) and Eq. (2.31) are

$$\epsilon_n = \begin{cases} E_n^{N+1} - E_0^N & \text{for electrons,} \\ E_0^N - E_n^{N-1} & \text{for holes.} \end{cases} \tag{2.33}$$

If the orbitals for an N electron system are known from a mean field calculation, such as an HF or DFT calculation, the Green's function corresponding to this mean field system can be obtained easily, under the approximation that an added electron or hole leaves the orbitals unchanged. This so-called non-interacting Green's G_0 is given as

$$G_0(k_1, k_2, \omega) = \sum_n^{\text{occ.}} \frac{\chi_n(k_1)\chi_n^*(k_2)}{\omega - \epsilon_n - i\eta} + \sum_n^{\text{unocc.}} \frac{\chi_n(k_1)\chi_n^*(k_2)}{\omega - \epsilon_n + i\eta}, \tag{2.34}$$

where the sums only involve orbitals of the original N electron system. Interestingly, the full Green's function G for the interacting system can be obtained from G_0. This is possible via Dyson's equation

$$G(\mathbf{r}_1, \mathbf{r}_2, \omega) = G_0(\mathbf{r}_1, \mathbf{r}_2, \omega) + \int \mathrm{d}\mathbf{r}_3\, \mathrm{d}\mathbf{r}_4\, G_0(\mathbf{r}_1, \mathbf{r}_3, \omega)\Sigma(\mathbf{r}_3, \mathbf{r}_4, \omega)G(\mathbf{r}_4, \mathbf{r}_2, \omega), \tag{2.35}$$

where Σ is the so-called self-energy. The self-energy describes the energy difference between a non-interacting bare particle and an interacting clouded particle, i.e. a quasiparticle. In fact, the Dyson equation can be derived by requiring that G must be a fundamental solution of the Schrödinger equation for the Hamiltonian $\hat{H}_0 + \Sigma$, where $\hat{H}_0$ is the mean field Hamiltonian (see [8]). Note that there are multiple possible definitions for Σ, depending on which effects are already included in $\hat{H}_0$.

The above version of Dyson's equation is written in position space. After Fourier transforming from $\mathbf{r}$ to $\mathbf{k}$, the convolution integral becomes a simple product, leading to

$$G(\mathbf{k}_1, \mathbf{k}_2, \omega) = G_0(\mathbf{k}_1, \mathbf{k}_2, \omega) + G_0(\mathbf{k}_1, \mathbf{k}_2, \omega)\Sigma(\mathbf{k}_1, \mathbf{k}_2, \omega)G(\mathbf{k}_1, \mathbf{k}_2, \omega). \tag{2.36}$$

This can be expanded by iteratively inserting the expression for G into the right hand side. This yields (omitting the arguments $\mathbf{k}_1$, $\mathbf{k}_2$ and ω for brevity)

$$G = G_0 + G_0\Sigma G_0 + G_0\Sigma G_0\Sigma G_0 + \ldots, \tag{2.37}$$

which is a geometric series and can therefore also be written as

$$G = G_0(1 - \Sigma G_0)^{-1} = (G_0^{-1} - \Sigma)^{-1}. \tag{2.38}$$

Starting from a non-interacting system whose Hamiltonian $\hat{H}_0$ includes no many-body effects beyond the Hartree potential and inserting its Green's function G_0 into Dyson's equation leads to the so-called quasiparticle equations

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{ext}}(\mathbf{r}) + V_H(\mathbf{r})\right]\chi_i(\mathbf{x}) + \int \mathrm{d}\mathbf{r}'\, \Sigma(\mathbf{r}, \mathbf{r}', \epsilon_i)\chi_i(\mathbf{x}') = \epsilon_i\chi_i(\mathbf{x}). \tag{2.39}$$

These are separate equations for each orbital and look similar to the Hartree-Fock or Kohn-Sham equations, but they do contain the additional energy-dependent term Σ. Therefore, Σ can be thought of as an energy-dependent potential. These equations can be solved iteratively to self-consistency, if an appropriate approximation for Σ is chosen (such as the GW approximation). This yields a set of quasiparticle orbitals χ_i and their corresponding energy eigenvalues.

Another way to obtain the quasiparticle energies from the Green's function is via the spectral function A, which is defined as

$$A(k_1, k_2, \omega) = \left| \frac{1}{\pi} \operatorname{Im} G(k_1, k_2, \omega) \right|. \tag{2.40}$$

Due to the denominators of the form $(\omega - \epsilon_n)$, the Green's function, and therefore also the spectral function, has poles at the quasiparticle energies ϵ_n. In the non-interacting case, these poles are sharp δ-like peaks at precise energies along the real axis. In the interacting quasiparticle case, the energies ϵ_n become complex, which means that the poles are slightly shifted away from the real axis. This creates spread-out peaks along the real axis. The real part of the ϵ_i corresponds to the position of the peak on the real axis, whereas the imaginary part corresponds to the width of the peak, which is also related to the quasiparticle lifetime. In addition to the main quasiparticle peak for each orbital, the spectral function can also contain smaller satellites. These arise from collective excitations of the electrons in the system (so-called plasmons).

Aside from quasiparticle energies, many other properties can in principle also be obtained from the single-particle Green's function. For example, the electron density $n(\mathbf{r}, t)$ can be determined as

$$n(\mathbf{r}, t) = \left\langle \Psi_N \middle| \hat{c}(\mathbf{r}, t)\hat{c}^\dagger(\mathbf{r}, t) \middle| \Psi_N \right\rangle = -iG(\mathbf{r}, \mathbf{r}, t, t^+). \tag{2.41}$$

Here, the notation t^+ means $t + \eta$, where η is infinitesimal and positive (this is necessary to ensure a well-defined time ordering in the Green's function). It is also possible to calculate the total electronic energy of a system from the single-particle Green's function via the Galitskii-Migdal total energy formula [15]

$$E = \frac{1}{2} \int \mathrm{d}\mathbf{r} \lim_{\mathbf{r}' \to \mathbf{r}} \lim_{t' \to t} \left[\frac{\partial}{\partial t} - ih(\mathbf{r}) \right] G(\mathbf{r}, \mathbf{r}', t, t'), \tag{2.42}$$

where $h(x)$ is the density-independent part of the system's Hamiltonian. Note that this is a direct formula for the energy as a functional $E[G]$, as opposed to the DFT case, where only the existence of a functional $E[n]$ is known.

2.5 Hedin's Equations

The single-particle Green's function G and the self-energy Σ can in principle be obtained from a closed set of integral equations, which are called Hedin's equations. They can be derived from the Heisenberg equations of motion of the creation and annihilation operators (for a full derivation see [16] or [17]).

For the formulation of Hedin's equations, several additional quantities must be introduced. First, the irreducible polarizability P is defined in terms of the reaction of a many-electron system to a small change in the external potential V_{ext}, denoted as δv_{ext}. Such a change in the external potential will also change the mean-field potential of the system itself, leading to a total change δv_{tot}. P is then defined as

$$P(1,2) = \frac{\delta n(1)}{\delta v_{\text{tot}}(2)}, \tag{2.43}$$

whereas the reducible polarizability χ is defined as

$$\chi(1,2) = \frac{\delta n(1)}{\delta v_{\text{ext}}(2)}. \tag{2.44}$$

Here, the numbers 1 and 2 are shorthands for sets of space-time coordinates $(\mathbf{r}_1, t_1)$ and $(\mathbf{r}_2, t_1)$. This notation will be used throughout the rest of this chapter (depending on the context, the numbers can also stand for reciprocal-space coordinates $(\mathbf{k}_1, t_1)$).

Secondly, the vertex function Γ is also defined in terms of δv_{ext},

$$\Gamma(1,2,3) = -\frac{\delta G(1,2)}{\delta v_{\text{tot}}(3)} = \delta(1,2)\delta(1,3) + \frac{\delta \Sigma(1,2)}{\delta v_{\text{tot}}(3)}. \tag{2.45}$$

Thirdly, the screened Coulomb interaction W is defined as

$$W(1,2) = \int d3\, \epsilon^{-1}(1,3)\nu(3-2), \tag{2.46}$$

where ν is the bare Coulomb interaction and ϵ^{-1} is the inverse of the microscopic dielectric function ϵ (see section 3.1 for an exact definition). Using these definitions, the full set of Hedin's equations reads

$$P(12) = -i \int \mathrm{d}(34)\, G(13) G(41^+) \Gamma(342), \tag{2.47}$$

$$W(12) = \nu(12) + \int \mathrm{d}(34)\, \nu(13) P(34) W(42), \tag{2.48}$$

$$\Sigma(12) = i \int \mathrm{d}(34)\, G(14^+) W(13) \Gamma(423), \tag{2.49}$$

$$\Gamma(123) = \delta(12)\delta(13) + \int \mathrm{d}(4567)\, \frac{\delta \Sigma(12)}{\delta G(45)} G(46) G(75) \Gamma(673), \tag{2.50}$$

$$G(12) = G_0(12) + \int \mathrm{d}(34)\, G_0(13) \Sigma(34) G(42). \tag{2.51}$$

Here, the notation 1^+ means $(\mathbf{r}_1, t_1 + \eta)$, where η is again an infinitesimally small positive number. The last of these equations is Dyson's equation for G. Also note that the equation for W is another Dyson-like equation. Σ, as it is used in Hedin's equations, includes both exchange and correlation effects. It can be written as $\Sigma = V_\mathrm{x} + \Sigma^\mathrm{c}$, where V_x is the exchange potential and Σ^c contains the correlation effects.

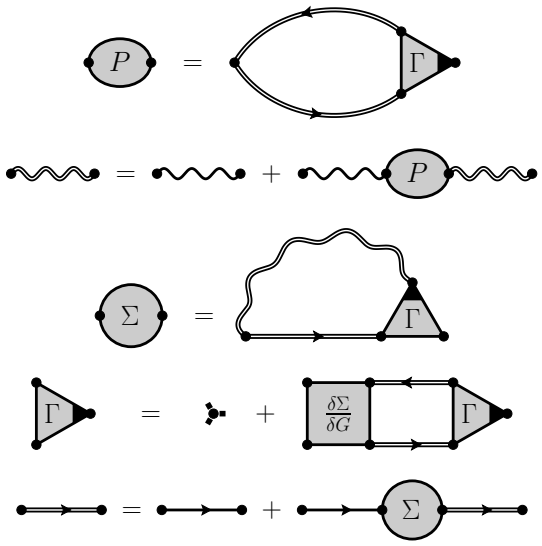

Figure 2.1: Diagrammatic representation of Hedin's equations.

Hedin's equations can also be written in the form of Feynman diagrams. This is shown in Fig. 2.1. Here, the single straight lines represent the non-interacting Green's function G_0, the double straight lines represent the full Green's function G, the single curly

lines represent the bare Coulomb interaction ν, and the double curly lines represent the screened interaction W. The other quantities are labeled in the diagrams. Note that Γ is a three-point quantity and $\delta\Sigma/\delta G$ is a four-point quantity. $\delta\Sigma/\delta G$ is sometimes also called the scattering amplitude and denoted as I.

The equations can in principle be solved iteratively to self-consistency. This can be done by starting with the non-interacting Green's function G_0 as a starting point for G and then computing Γ, P, W and Σ, in this order. Finally, from Σ, a new G can be obtained for the next iteration. However, this has very high computational cost. Furthermore, no simple way to numerically calculate $\delta\Sigma/\delta G$ exists, so that one is forced to make some approximations. One possible simplification is the GW approximation.

2.6 The GW Approximation

The GW approximation [18,19] corresponds to setting $\Gamma(123) = \delta(12)\delta(13)$ in Hedin's equations. This leads to a simplified set of equations,

$$P(12) = -iG(12)G(21^+), \tag{2.52}$$

$$W(12) = \nu(12) + \int \mathrm{d}(34)\, \nu(13)P(34)W(42), \tag{2.53}$$

$$\Sigma(12) = iG(12^+)W(12), \tag{2.54}$$

$$G(12) = G_0(12) + \int \mathrm{d}(34)\, G_0(13)\Sigma(34)G(42). \tag{2.55}$$

Again, these equations can be represented diagrammatically, as shown in Fig. 2.2. Note that the diagrams for P and Σ are simplified compared to Fig. 2.1, whereas the diagrams for G and W are the same. Also note that the scattering amplitude $\delta\Sigma/\delta G$ does not occur anymore. Therefore, these diagrams contain no three- or four-point quantities.

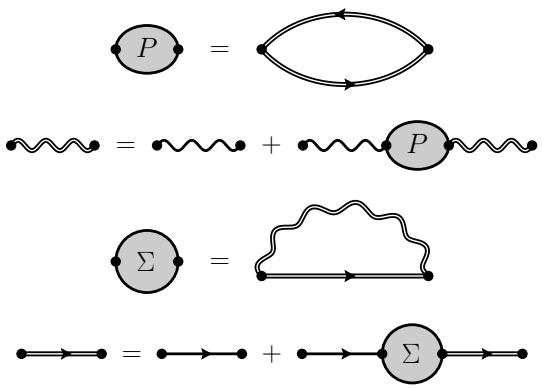

Figure 2.2: Diagrammatic representation of the GW approximation.

In order to solve the equations Eq. (2.52) to Eq. (2.55), G_0 must be provided as a starting point. This is usually done by running a DFT calculation as a preparatory step and obtaining G_0 from the Kohn-Sham orbitals. Using this, the polarizability P can be calculated from G_0, the screened interaction W can be calculated from P and the self-energy Σ can be calculated from W, leading to the quasiparticle energies via the quasiparticle equation Eq. (2.39). These steps constitute the so-called G_0W_0 method (also called single-shot GW). G_0W_0 does not provide any self-consistency, since both G and W are calculated only once, at the level of DFT. Therefore, the results obtained form G_0W_0 calculations depend strongly on the DFT results that are used as a starting point.

Self-consistency can be achieved by iterating the process, using the self-energy obtained in the first step to determine a new estimation for G, which is then used as input for the second step, and so on. This leads to the self-consistent GW method. The iteration process usually converges after a few steps.

Note that in this process, G does not necessarily have to be calculated using Dyson's equation. Instead, it is also possible to obtain G from Σ via the quasiparticle equation, under the approximation that G can be expressed in terms of single-electron DFT orbitals and some optimized quasiparticle energies, in analogy to Eq. (2.34), as

$$G(k_1, k_2, \omega) = \sum_n^{\text{occ.}} \frac{\chi_n(k_1)\chi_n^*(k_2)}{\omega - \epsilon_n - i\eta} + \sum_n^{\text{unocc.}} \frac{\chi_n(k_1)\chi_n^*(k_2)}{\omega - \epsilon_n + i\eta}. \tag{2.56}$$

This is called eigenvalue self-consistent GW.

As a third option, it is also possible to use a partially self-consistent method "in between" G_0W_0 and GW, where W is kept at the DFT level W_0 and only G is updated in every step. This idea is based on the observation that, in many cases, W_0 is already a good approximation to the observable W (physically, this means that DFT can already describe the screening effects represented by W reasonably well). The resulting method is called GW_0. It is computationally cheaper than GW, since there is no need to re-calculate polarizabilites or the dielectric function.

Another possibility is the so-called quasiparticle GW method (QPGW) [1]. This method works by optimizing a Hamiltonian $\hat{H}_0$ such that the perturbation Σ becomes minimized. This means that an approximation for the exchange-correlation potential V_{xc} is obtained from Σ. This V_{xc} accounts for the many-body effects represented by Σ, but has the form of a non-energy-dependent effective potential. It is then included in $\hat{H}_0$. With the updated $\hat{H}_0$, G_0 can be re-calculated, leading to a new Σ. This process can be iterated to self-consistency. QPGW is therefore a self-consistent method, but not the same as GW, since it works with non-interacting effective Hamiltonians $\hat{H}_0$. Convergence is usually also achieved after only a few steps.

2.7 Self-Consistent GW

Even though the GW approximation constitutes a significant simplification compared to the full Hedin equations, fully self-consistent GW calculations are still more computationally demanding than mean field calculations such as HF and DFT, in terms of both time and memory requirements.

Implementations of the self-consistent GW approach generally fall into one of two categories. The first and more widely used category consists of implementations that operate with reciprocal space and real frequencies, i.e. they calculate all quantities as functions of the form $f(\mathbf{k}_1, \mathbf{k}_2, \omega)$. The computation time for these implementations scales as $\mathcal{O}(N^4)$ with the number of electrons in the system. The second category consists of implementations that operate in real space and imaginary times, i.e. they calculate all quantities as functions of the form $f(\mathbf{r}_1, \mathbf{r}_2, it_2 - it_1)$. This approach is called space-time GW (however, in reality, Fourier transformations or sine/cosine transformations can be used at several points in the implementation; therefore, some quantities are also obtained along the imaginary frequency axis, as functions of $i\omega$). Implementations using space-time

GW are possible in $\mathcal{O}(N^3)$. A $\mathcal{O}(N^3)$ implementation of space-time GW has recently been included in the Vienna *Ab-Initio* Simulation Package (VASP) [20]. For both kinds of implementations, the discretization of ω is of great importance for both the accuracy and the cost of the calculations. The number of ω grid points is therefore a parameter that can be tuned for each calculation. For conventional reciprocal-space implementations, usually more than 100 points need to be used, whereas for the recently developed $\mathcal{O}(N^3)$ space-time GW, smaller grids with 16 or 24 points are sufficient. [20]

For the calculations in this thesis, the space-time GW implementation in VASP was used. Therefore, the problems discussed in chapters 3 and 4 are specific to the space-time GW approach. For a fully detailed description of the VASP implementation, see [20].

In general, all implementations follow in principle the path outlined by the equations Eq. (2.52) to Eq. (2.55). However, some simplifications are possible compared to the direct evaluation of these equations. Most importantly, the screened interaction W does not have to be calculated from equation Eq. (2.53). Instead, it can be obtained from its definition in Eq. (2.46). In reciprocal space, this reads

$$W_{\mathbf{G},\mathbf{G}'}(\mathbf{k}, i\omega) = \nu_{\mathbf{G},\mathbf{G}'}(\mathbf{k})\, \epsilon^{-1}_{\mathbf{G},\mathbf{G}'}(\mathbf{k}, i\omega). \tag{2.57}$$

Here, $\mathbf{k}$ represents a wave vector, $\mathbf{G}$ and $\mathbf{G}'$ are reciprocal lattice vectors, and ν is the symmetric reciprocal-space Coulomb kernel, given in atomic units as

$$\nu_{\mathbf{G},\mathbf{G}'}(\mathbf{k}) = \frac{4\pi}{|\mathbf{k}+\mathbf{G}||\mathbf{k}+\mathbf{G}'|}. \tag{2.58}$$

Thus, for the evaluation of W from this formula, the inverse dielectric function ϵ^{-1} is needed (see section 3.1 for the definition of ϵ^{-1}). The evaluation of this quantity is difficult in the case of the so-called head, where $\mathbf{G} = \mathbf{G}'$. This will be discussed in more detail in chapter 3 of this thesis. Note that in the case G_0W_0 and GW_0, it is possible to obtain the head of ϵ^{-1} using $\mathbf{k} \cdot \mathbf{p}$ perturbation theory.

Furthermore, space-time GW has the problem that Σ is obtained along the imaginary time axis, i.e. as $\Sigma(it_2 - it_1)$. It can be transformed to $\Sigma(i\omega)$ using sine and cosine transformations. However, $\Sigma(\omega)$ is needed along the real axis, in order to obtain $G(\omega)$, and ultimately the quasiparticle energies. Therefore, an analytic continuation needs to be performed from the imaginary axis to the real axis, which poses numerical difficulties,

especially if dense grids are used for the discretization of ω. Once the $\Sigma(\omega)$ is known, G can be obtained as

$$G(\omega) = \frac{1}{\omega - T - V_{\text{ext}} - V_{\text{H}} - \Sigma(\omega)}, \tag{2.59}$$

and the quasiparticle energies correspond to the poles of G. The analytic continuation problem is further discussed in chapter 4.

3 The Head of the Dielectric Function

3.1 The Head Calculation Problem

The macroscopic dielectric function ϵ_{mac} is defined such that its inverse $\epsilon_{\text{mac}}^{-1}$ is the proportionality factor between an external electric field $\mathbf{E}_{\text{ext}}$ and the resulting total electric field $\mathbf{E}$ in a material. Thus,

$$\mathbf{E}\left(\mathbf{r},\omega\right) = \int \mathrm{d}\mathbf{r}'\, \epsilon_{\text{mac}}^{-1}\left(\mathbf{r}-\mathbf{r}',\omega\right) \mathbf{E}_{\text{ext}}\left(\mathbf{r}',\omega\right). \tag{3.1}$$

In general, ϵ_{mac} is a Cartesian 3×3 tensor. However, if $\mathbf{E}$ is collinear to $\mathbf{E}_{\text{ext}}$, it reduces to a scalar. It depends only on $\mathbf{r}-\mathbf{r}'$ if the material being considered is homogeneous at a macroscopic scale. Eq. (3.1) can be written in momentum space as

$$\mathbf{E}\left(\mathbf{q},\omega\right) = \epsilon_{\text{mac}}^{-1}\left(\mathbf{q},\omega\right) \mathbf{E}_{\text{ext}}\left(\mathbf{q},\omega\right). \tag{3.2}$$

Similarly, the microscopic dielectric function ϵ is defined via the equation

$$\phi_{\text{tot}}\left(\mathbf{r},\omega\right) = \int \mathrm{d}\mathbf{r}'\, \epsilon^{-1}\left(\mathbf{r},\mathbf{r}',\omega\right) \phi_{\text{ext}}\left(\mathbf{r}',\omega\right), \tag{3.3}$$

where ϕ_{tot} is the total microscopic scalar electric potential. ϵ does depend on both $\mathbf{r}$ and $\mathbf{r}'$, because the microscopic electric field can vary at the length scale of a unit cell. However, ϵ has the periodicity of the crystal, i.e. $\epsilon(\mathbf{r},\mathbf{r}',\omega) = \epsilon(\mathbf{r}+\mathbf{R},\mathbf{r}'+\mathbf{R},\omega)$. Therefore, in the momentum space quantity $\epsilon(\mathbf{q},\mathbf{q}',\omega)$, the difference between $\mathbf{q}$ and $\mathbf{q}'$ must be a reciprocal lattice vector. ϕ_{tot} can therefore be written as a sum

$$\phi_{\text{tot}}(\mathbf{q}+\mathbf{G},\omega) = \sum_{\mathbf{G}'} \epsilon^{-1}_{\mathbf{G},\mathbf{G}'}(\mathbf{q},\omega)\, \phi_{\text{ext}}(\mathbf{q}+\mathbf{G}',\omega), \tag{3.4}$$

where $\mathbf{G}$ and $\mathbf{G}'$ are reciprocal lattice vectors and $\epsilon^{-1}_{\mathbf{G},\mathbf{G}'}(\mathbf{q},\omega)$ is defined as

$$\epsilon^{-1}_{\mathbf{G},\mathbf{G}'}(\mathbf{q},\omega) = \epsilon^{-1}(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}',\omega). \tag{3.5}$$

The macroscopic and microscopic quantities are connected via

$$\mathbf{E} = \lim_{\mathbf{q}\to 0} i\mathbf{q}\,\phi_{\mathrm{tot}}(\mathbf{q},\omega) = \lim_{\mathbf{q}\to 0}\int \mathrm{d}^3\mathbf{r}\,\nabla e^{i\mathbf{q}\mathbf{r}}\phi_{\mathrm{tot}}(\mathbf{r},\omega), \tag{3.6}$$

since $\mathbf{E}$ corresponds to the long-wavelength limit of $\nabla\phi_{\mathrm{tot}}$. Under the assumption that the external field varies on a length scale that is much larger than a unit cell, it can also be shown that ϵ and ϵ_{mac} are related via

$$\epsilon^{-1}_{\mathrm{mac}}(\hat{\mathbf{q}},\omega) = \lim_{\mathbf{q}\to 0}\epsilon^{-1}(\mathbf{q},\mathbf{q},\omega) = \lim_{\mathbf{q}\to 0}\epsilon^{-1}_{00}(\mathbf{q},\omega). \tag{3.7}$$

Here, $\epsilon^{-1}_{\mathrm{mac}}$ only depends on the direction $\hat{\mathbf{q}} = \mathbf{q}/|\mathbf{q}|$. The quantity $\epsilon_{00}(\mathbf{q},\omega)$ in the equation is the so-called head of the matrix $\epsilon_{\mathbf{G}\mathbf{G}'}$, i.e. the component for which $\mathbf{G} = \mathbf{G}' = 0$. Thus, under the assumption that the external field varies on a scale larger than a unit cell, the head of the microscopic dielectric matrix determines the value of the macroscopic dielectric function.

For the purposes of implementations of self-consistent GW calculations, ϵ can be obtained from P within the so-called random phase approximation (RPA) via the formula

$$\epsilon_{\mathbf{G},\mathbf{G}'}(\mathbf{q},i\omega) = \delta_{\mathbf{G}\mathbf{G}'} - \nu_{\mathbf{G},\mathbf{G}'}(\mathbf{q})\chi_{\mathbf{G},\mathbf{G}'}(\mathbf{q},i\omega). \tag{3.8}$$

However, this formula fails for the head in the limit $\mathbf{q}\to 0$, due to the singularity of the Coulomb potential ν, as given in Eq. (2.58) (note that this is only a problem for Eq. (3.8), but not for the similar looking Eq. (2.57), since W has the same singularity as ν, whereas ϵ^{-1} takes on a finite value at $\mathbf{q} = 0$). For G_0W_0 and GW_0 calculations, the $\mathbf{q}\to 0$ limit of the head can instead be calculated via $\mathbf{k}\cdot\mathbf{p}$ perturbation theory using DFT eigenstates and their energies [21]. However, this cannot be done for fully self-consistent GW. Here, other methods have to be used.

3.2 Fitting from finite q

For the purposes of this thesis, the head of ϵ^{-1} at $\mathbf{q} = 0$ was approximated using an extrapolation from a quadratic fit based on data points obtained at $\mathbf{q} \neq 0$ for each ω (note that the ω points used here lie on the imaginary axis, but are written in this section as ω instead of $i\omega$ for notational simplicity). This means that for each ω, $\epsilon(\mathbf{q}, \omega)$ was approximated as a function of the form

$$f(|\mathbf{q} + \mathbf{G}|) = a + b|\mathbf{q} + \mathbf{G}|^2 \,, \tag{3.9}$$

where a and b are fitting parameters. The estimate for $\lim_{\mathbf{q}\to 0} \epsilon_{00}^{-}1(\mathbf{q}, \omega)$ obtained from the fit is then equal to a. Note that it is not obvious how many data points should be used as input for the fit. Therefore, in practice, the fit is performed several times and the least square errors are compared. The value of a is then taken from the best fit. Such a fitting approach can be tested with G_0W_0 calculations, since in this case, the head can be calculated from the DFT orbitals.

Fig. 3.1 shows data points for $\epsilon_{00}^{-1}(\mathbf{q}, \omega)$ at different values of ω, obtained from a G_0W_0 calculation for C, with a DFT calculation as a preparatory step (for further details on the calculations, see section 5.1 and especially Tab. 5.1). From this data, it can be seen qualitatively that $\epsilon_{00}^{-1}(\mathbf{q}, \omega)$ has positive curvature for small $|\mathbf{q}|$, then has a turning point and then goes to 1 for $\mathbf{q} \to \infty$. Therefore, the positively curved quadratic fitting function in Eq. (3.9) is only appropriate in the range before the turning point. This implies that only a limited number of data points can be used for the fit. Furthermore, it can be seen that for large ω, $\epsilon_{00}^{-1}(\mathbf{q}, \omega)$ starts ar relatively high values and goes to 1 fast, whereas for small ω, it starts at lower values and goes to 1 much more slowly. Therefore, it is to be expected that the quality of the fit will be different for large ω than for small ω.

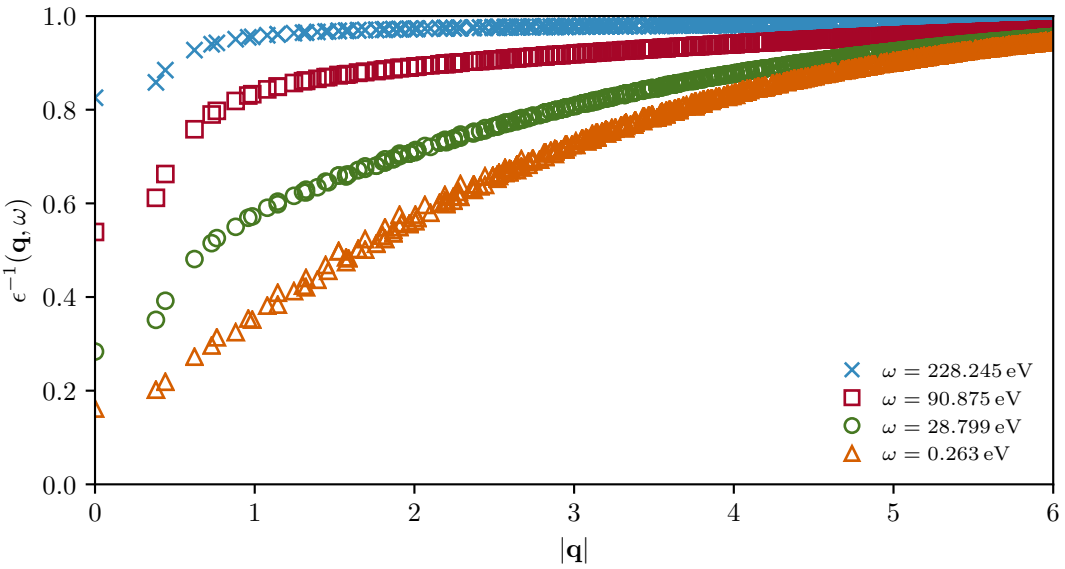

Figure 3.1: Data points for $\epsilon^{-1}(\mathbf{q},\omega)$ at four different values of ω. The data for this plot was obtained from a G_0W_0 calculation for C using $8\times8\times8$ k-points, where W_0 is calculated using DFT orbitals.

Fig. 3.2 and Fig. 3.3 show data points and fitting curves at small and at large ω, respectively (but note that the axis scaling in these figures is different than in Fig. 3.1; it is zoomed in to small values of $|\mathbf{q}|$). Again, the data is obtained from G_0W_0 calculations for C. The fits were performed using standard linear least-square fitting. The number of available data points depends on the number of k-points used in the calculations. Fig. 3.2 and Fig. 3.3 therefore show the results for four different k-point sets. All of the k-point sets used here were Γ-centered Monkhorst-Pack grids [22]. In each of the fits, the first three data points with $|\mathbf{q}| \neq 0$ were used for the fit. The data point at $|\mathbf{q}| = 0$ serves as a reference for comparison (which is only available for W_0). Note that the fits always overestimate the true value at $|\mathbf{q}| = 0$. Also note that an increase of the number of k-points used always leads to an improvement in the estimation of $\epsilon^{-1}_{00}(\mathbf{q}=0,\omega)$, since there are more data points available close to $\mathbf{q}=0$ if many k-points are used.

Fig. 3.4 shows the convergence of the results of this fitting approach with the number of k-points, over the full range of values for ω (24 points were used for the discretization of the ω axis). It can be seen that the choice of the k-point set matters the most for small and medium values of ω, where $\epsilon^{-1}_{00}(\mathbf{q}=0,\omega)$ is small.

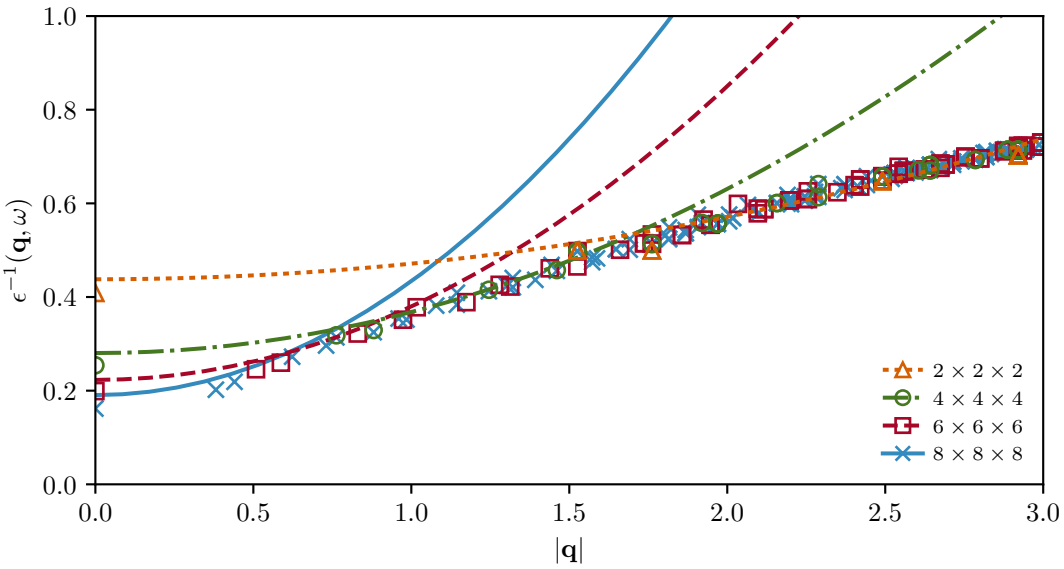

Figure 3.2: Quadratic fits for $\epsilon^{-1}(\mathbf{q},\omega)$ at $\omega = 0.264\,\text{eV}$, with four different k-point sets. The data for this plot was obtained from G_0W_0 calculations for C, where W_0 is calculated using DFT orbitals.

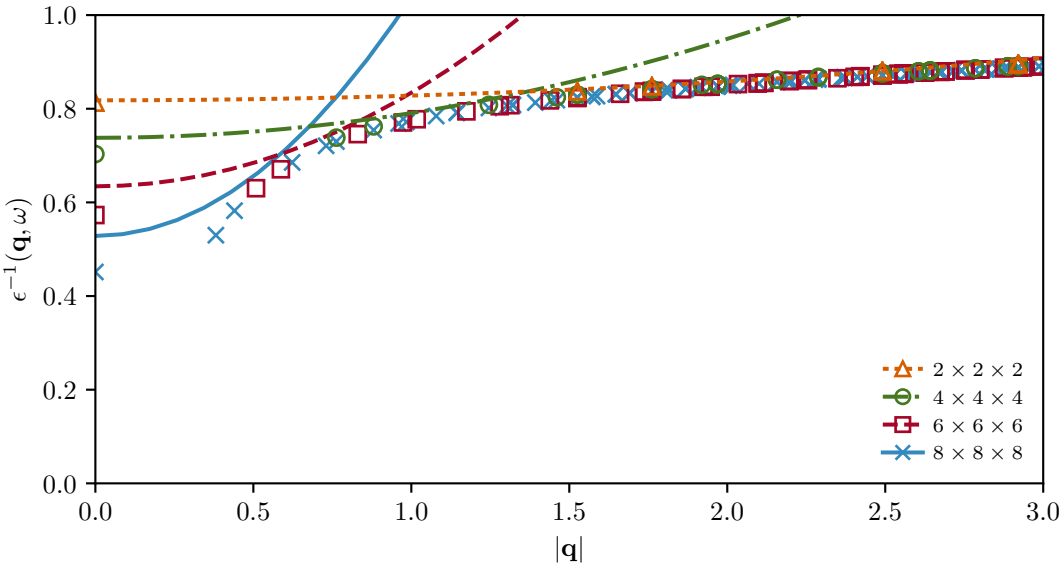

Figure 3.3: Quadratic fits for $\epsilon^{-1}(\mathbf{q},\omega)$ at $\omega = 67.837\,\text{eV}$, with four different k-point sets. The data for this plot was obtained from G_0W_0 calculations for C, where W_0 is calculated using DFT orbitals.

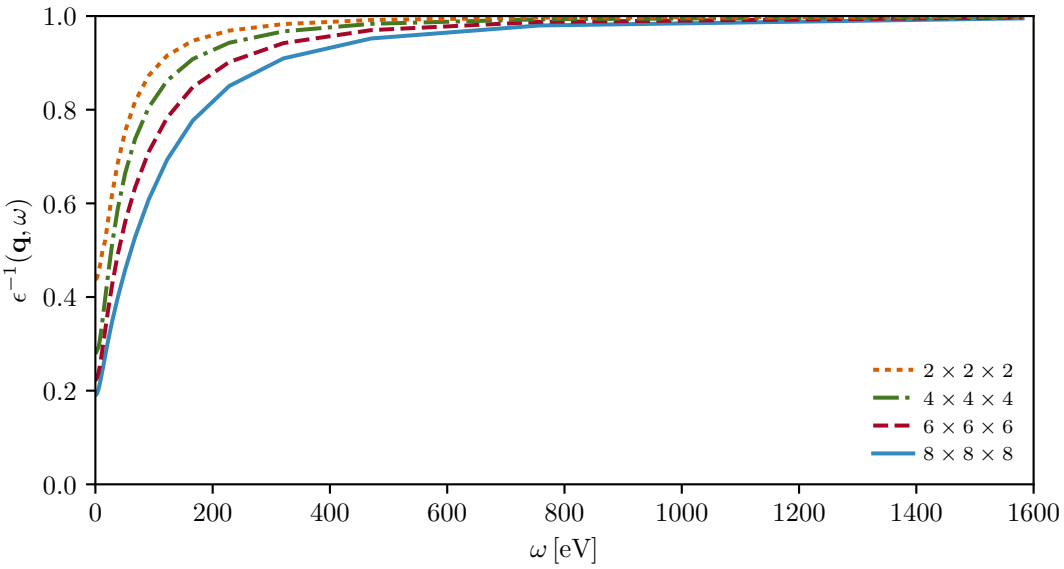

Figure 3.4: Convergence of the quadratic fitting approach with number of k-points. The data for this plot was obtained from a G_0W_0 calculation for C.

Table 3.1: Comparison of quasiparticle energies obtained from G_0W_0 and GW calculations for C using $6 \times 6 \times 6$ k-points and 24 ω-points.

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c
G_0W_0 (head from DFT)	-1.18	-22.04	7.54	-2.93	10.46	-6.58	6.29
G_0W_0 (quadratic fit)	-1.24	-21.99	7.51	-2.92	10.38	-6.56	6.27
GW (head set to 0)	-7.70	-27.95	14.90	-3.37	18.62	-7.45	14.50
GW (quadratic fit)	-1.59	-23.83	8.83	-3.10	11.89	-7.00	7.44

Table 3.1 shows the effects of the head calculation on the resulting quasiparticle energies, again using C as an example. The table lists results for both G_0W_0 and GW, with the different methods for obtaining the head. The column names are the same as in section 5.3: ΔIP is the change in the ionization potential compared to DFT, Γ_{VBmin} is the valence band minimum, and Γ_c L_v, L_c, X_v, X_c are the highest valence band and lowest conduction band energies at Γ, L, and X, respectively. All values are given in eV and all except for ΔIP are given relative to IP_{GW} (note however that this table does not include basis-set corrections as described in section 5.2). The table shows that the differences between G_0W_0 with the head obtained from DFT via $\mathbf{k}\cdot\mathbf{p}$ perturbation theory

and G_0W_0 with the head obtained from the quadratic fit, at least with the k-point set chosen here ($6 \times 6 \times 6$ k-points). For GW, it can be seen that neglecting the head by setting it to 0 yields significantly different results, even with this k-point set. Thus, the head has to be included in order to obtain realistic results.

Fig. 3.5 shows the k-point convergence of the band gap of C (corresponding to the column X_c in Tab. 3.1). Values from the GW calculations with the head set to 0 are not included in this figure, since they lie outside the energy range depicted here. Note that the band gap obtained with G_0W_0 with $\mathbf{k} \cdot \mathbf{p}$ perturbation theory from DFT is already reasonably converged with $4 \times 4 \times 4$ k-points, whereas the other values still change by a significant amount when increasing the k-point set from $4 \times 4 \times 4$ to $6 \times 6 \times 6$. The gap obtained from G_0W_0 with the quadratic fit changes by 170 meV, the one obtained from GW changes by 350 meV. This is most likely due to the k-point sensitivity of the quadratic fit.

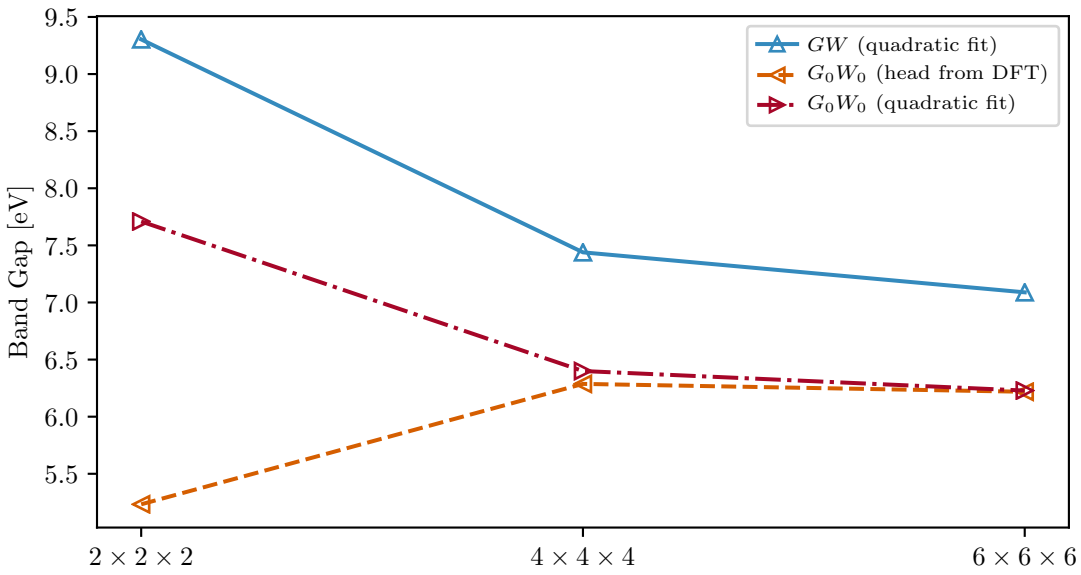

Figure 3.5: k-point convergence of the band gap of C, obtained from G_0W_0 and GW calculations using 24 ω-points.

4 Analytic Continuation of Σ

4.1 The Analytic Continuation Problem

An analytic function is a complex function $f(z) = u(z) + iv(z)$ that is differentiable at every point in its domain. It can be shown that this implies that all derivatives of an analytic function are also differentiable. A necessary and sufficient condition for a function to be analytic is that it fulfils the Cauchy-Riemann differential equations

$$\begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y}, \\ \frac{\partial u}{\partial y} &= -\frac{\partial v}{\partial x} \end{aligned} \tag{4.1}$$

where x and y are the real and the imaginary parts of the complex variable z, respectively.

Analytic continuation is the process of extending the domain of an analytic function, i.e. defining it for regions outside the original domain, in such a way that the resulting function is still analytic. An important special case of this is the continuation of a function that is given on points along the imaginary axis to the real axis. In GW calculations, an analytic continuation problem of this form arises if they are implemented in such a way that Σ is obtained at imaginary frequencies, as $\Sigma(i\omega)$, which is the case in space-time GW. The analytic continuation is needed here in order to obtain $G(\omega)$ using Eq. (2.59), which allows to determine the quasiparticle energies from the poles of $G(\omega)$.

In principle, there exists a unique analytic continuation for any function that is known at infinitely many discrete points along the imaginary axis with infinite precision. In practice however, analytic continuation has to be done from finitely many points on the imaginary axis with finite precision. Therefore, it can be seen as a fitting problem for a complex function with a given set of data points. Several methods exist for numerically solving this problem. In this thesis, least square Padé fitting in monomial basis and in Lagrange basis is used, as well as the continued fraction Padé fitting method. Other

methods include the multipole expansion method [23], which is also based on Padé approximants, and the maximum entropy method (MEM) [24], which is based on Bayesian inference. MEM is robust against statistical noise in the input data than Padé-based methods, which is relevant especially for Monte-Carlo calculations. However, it tends to yield less accurate results for the poles of G, which is why Padé-based methods are preferred for space-time GW implementations.

4.2 Padé Fitting in Monomial Basis

A Padé function of order N_n/N_d is a rational function of the form

$$P(z) = \frac{\sum_{k=0}^{N_\mathrm{n}} p_k z^k}{\sum_{k=0}^{N_\mathrm{d}} q_k z^k} \tag{4.2}$$

where p_k and q_k are complex coefficients and exactly one of the q_k is fixed to 1 (usually q_{N_d}, i.e. the coefficient of the highest power of z, is chosen for this, but in principle any q_k can be used). Padé functions can be seen as finite-order approximants to arbitrary complex functions with N_d poles. In this sense, they are a generalization of Taylor approximants. If $N_d = 1$, i.e. if there are no poles, Eq. (4.2) reduces to a Taylor sum (or, more precisely, a Maclaurin sum) of order N_n.

For the purpose of fitting Σ, N_n is always chosen to be $N_d - 1$. This is necessary to obtain the right asymptotic behavior for $i\omega \to \infty$, where Σ should go to 0 as $1/i\omega$. In the rest of this chapter, $N_n = N_d - 1$ will therefore always be assumed. Also note that, in order for the correct asymptotic behavior to be achieved, only the truly frequency-dependent part of Σ can be fitted. Constants must be subtracted before fitting and added afterwards. For the GW approximation, this means that the exchange potential V_x needs to be subtracted.

In order to perform a fit, a Padé function $P(z)$ can be equated with M given values $\{y_i\}$ at a set of M points $\{z_i\}$,

$$y_i = P(z_i) \quad \text{for } i = 1, \ldots, M. \tag{4.3}$$

By setting $q_{N_d} = 1$ bringing the y_i to the left hand side, this can be brought into the form of a linear system of equations

$$A\mathbf{x} = \mathbf{b}, \tag{4.4}$$

where A is an $M \times 2(N_d - 1)$ matrix of the form

$$A = \left(\begin{array}{cccc|cccc} 1 & z_1 & \cdots & z_1^{N_d-1} & -y_1 & -y_1 z_1 & \cdots & -y_1 z_1^{N_d-1} \\ 1 & z_2 & \cdots & z_2^{N_d-1} & -y_2 & -y_2 z_2 & \cdots & -y_2 z_2^{N_d-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_M & \cdots & z_M^{N_d-1} & -y_M & -y_M z_M & \cdots & -y_M z_M^{N_d-1} \end{array}\right), \tag{4.5}$$

the right hand side $\mathbf{b}$ is a vector of the form

$$\mathbf{b} = \begin{pmatrix} -y_1 z_1^{N_d} \\ \vdots \\ -y_M z_M^{N_d} \end{pmatrix}, \tag{4.6}$$

and $\mathbf{x}$ is a vector containing the unknown Padé coefficients p_k and q_k,

$$\mathbf{x} = \begin{pmatrix} p_1 \\ \vdots \\ p_{N_d-1} \\ q_1 \\ \vdots \\ q_{N_d-1} \end{pmatrix}. \tag{4.7}$$

Eq. (4.4) is an overdetermined system of linear equations. A solution to the fitting problem can be found using a linear least square algorithm. However, the fitting matrix in Eq. (4.5) has the numerical problem that the exponents increase from left to right in each block of the matrix, leading to small values on the left of each block and large values on the right. Furthermore, in the case of fitting $\Sigma(i\omega)$, the input values z_i vary considerably in magnitude, leading also to an increase of the values in each column of the matrix from top to bottom. As a result, the conditioning number of the matrix is large, rendering numerical solutions of the fitting problem difficult, as small errors in the input values can scale up to give very large errors in the output.

Fig. 4.1 shows a logarithmic plot of the real and imaginary parts of the fitting matrix, for the self-energy obtained from a self-consistent GW calculation for C with 16 points in the ω-grid, and thus 16 points as input data for the fit. Note also that some columns in this plot are zero. This is because the input values z_i are, in our case, purely imaginary, which implies that all their even powers are purely real and all their odd powers are purely imaginary. This effect does not occur in the right half of the matrix, because here the powers of z_i are multiplied with the values for y_i, which have both a real and an imaginary part (since in our application, y corresponds to Σ).

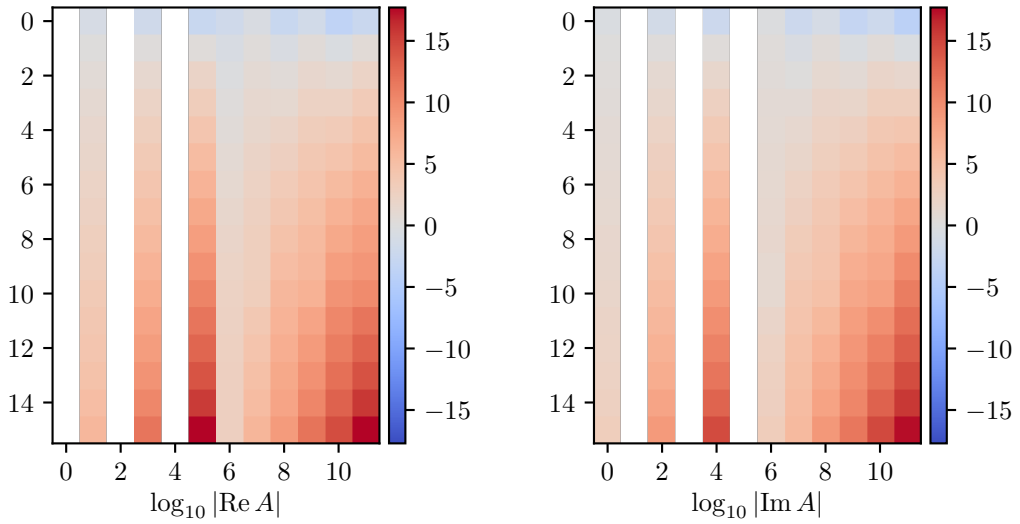

Figure 4.1: Logarithmic plot of the real and imaginary part of the fitting matrix for a Padé fit in monomial basis, with 16 data points and N_d set to 7. The white squares are matrix components which are zero, i.e. for which the logarithm does not exist. The data used for the fit was the self-energy of C in the first band at the Γ point, obtained by a self-consistent GW calculation.

4.3 Padé Fitting in Lagrange Basis

The Lagrange polynomials $L_k^n(x)$ for a set of n control points $t_1 \dots t_n$ are defined as the unique polynomials with the property

$$L_k^n(t_j) = \delta_{kj}. \tag{4.8}$$

For a given n, these are n polynomials of degree n. They can be written in an explicit form as follows:

$$L_k^n(t) = \prod_{\substack{j=0 \\ j \neq k}}^{n} \frac{t - t_j}{t_k - t_j}. \tag{4.9}$$

The Lagrange polynomials $L_k^n(x)$ form a basis for the space of polynomials of degree n. This means that arbitrary polynomials of degree n can be written as a linear combination of Lagrange polynomials. This makes it possible to write the numerator and denominator of a Padé function in terms of Lagrange polynomials

$$P(z) = \frac{\sum_{k=0}^{N_\mathrm{n}} p_k L_k^{N_n}(z)}{\sum_{k=0}^{N_\mathrm{d}} q_k L_k^{N_d}(z)}. \tag{4.10}$$

where, as in Eq. (4.2), exactly one of the coefficients q_i in the denominator has to be set to 1. This form of the Padé function can be used for fitting, as an alternative to Eq. (4.2). This leads to the fitting matrix

$$A = \left(\begin{array}{ccc|ccc} L_1(z_1) & \cdots & L_{N_d-1}(z_1) & -y_1 L_1(z_1) & \cdots & -y_1 L_{N_d-1}(z_1) \\ L_1(z_2) & \cdots & L_{N_d-1}(z_2) & -y_2 L_1(z_2) & \cdots & -y_2 L_{N_d-1}(z_2) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ L_1(z_M) & \cdots & L_{N_d-1}(z_M) & -y_M L_1(z_M) & \cdots & -y_2 L_{N_d-1}(z_M) \end{array} \right) \tag{4.11}$$

and the right hand side

$$\mathbf{b} = \left(\begin{array}{c} -y_1 L_{N_d}(z_1) \\ \vdots \\ -y_M L_{N_d}(z_M) \end{array} \right). \tag{4.12}$$

Compared to the fitting matrix in Eq. (4.5), the Lagrange fitting matrix has the advantage that every component of row i contains z_i to all powers up to z_i^n. This leads to a significantly smaller range of values. Fig. 4.2 contains a plot of such a fitting matrix with $N_d = 7$, generated from the same calculation as Fig. 4.1. The values are still spread out over multiple orders of magintude, but far less so than in Fig. 4.1.

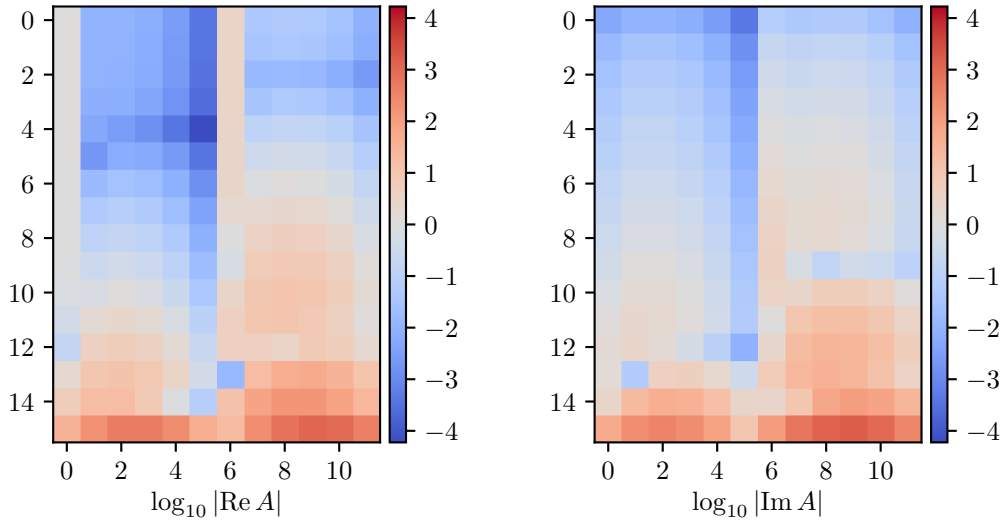

Figure 4.2: Logarithmic plot of the real and imaginary part of the fitting matrix for a Padé fit in Lagrange basis, with 16 data points and N_d set to 7. The data used for the fit is the same as in Fig. 4.1.

For the purposes of this thesis, least-square Padé fitting was implemented using Householder's algorithm for solving the least square problem using a QR decomposition [25]. This was found to work better than an algorithm based on LU factorization. The implementation was done using quadruple precision, i.e. 128 bit floating point numbers, in order to be able to cope with the bad conditioning. The most important parameter for the quality of the fit is N_d, the number of terms in the denominator (or, equivalently, N_n, the number of terms in the numerator). Higher N_d generally lead to more exact results, but this is only true up to a certain threshold, where further increasing can cause numerical instabilities. In general $N_d \in \{5, \ldots, 10\}$ is a safe choice. The stability of the fit also depends largely on the number of ω points given as input. It was found that 16 or 24 data points, which are typical grid sizes in space-time GW calculations, can be handled by the fitting procedure with these values of N_d.

In practice, the fitting of Σ is done per band and per k-point. Fig. 4.3 and Fig. 4.4 show example fits, for the highest occupied band of C and Si. These figures also include plots of the real and imaginary parts of the fitted Σ in the two-dimensional complex plane, allowing to see the pole structure of the Padé function. This reveals that the poles can be relatively far away from the real axis.

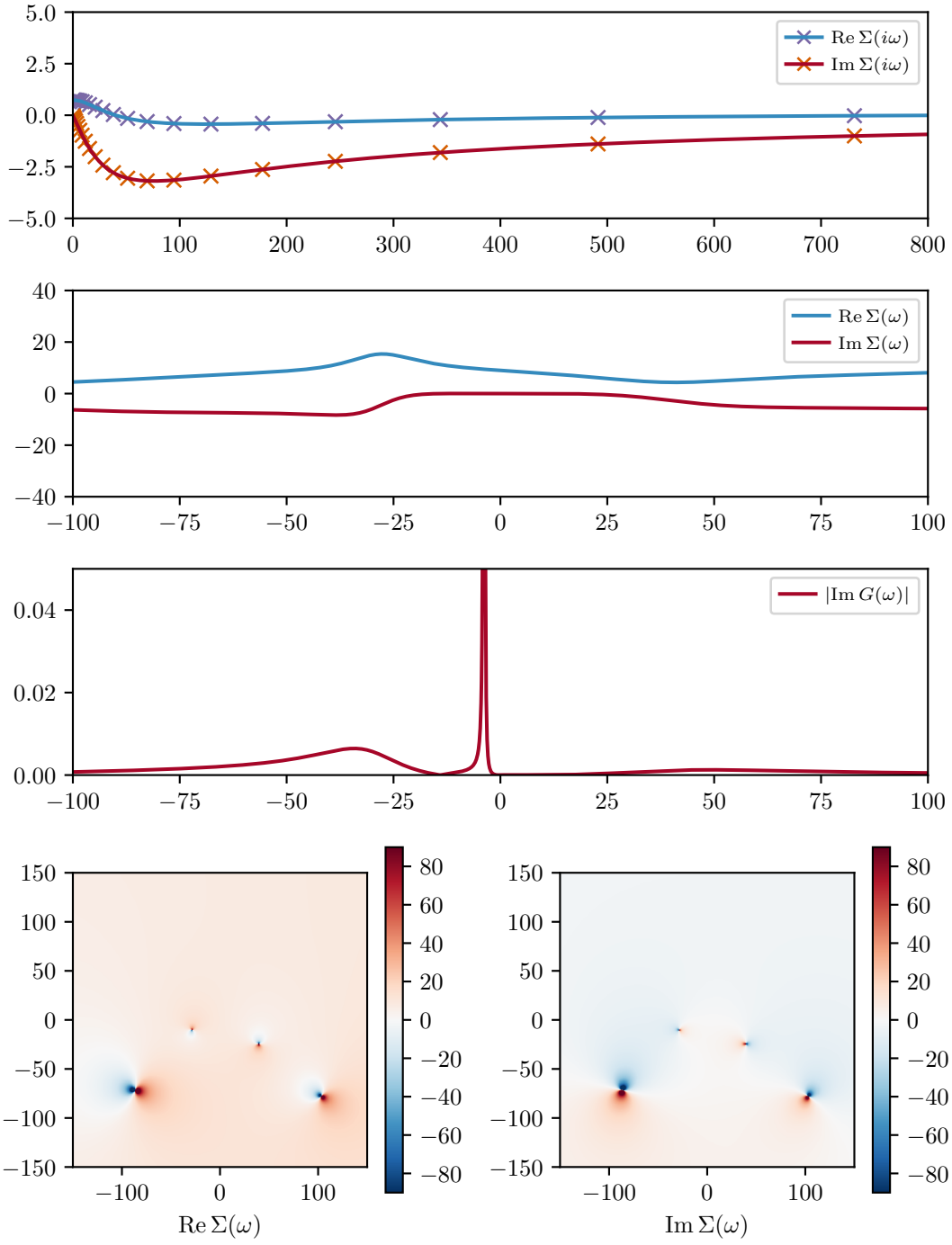

Figure 4.3: QR Padé fitting of Σ for the highest occupied band in C at Γ. From top to bottom: Σ along the imaginary axis, Σ along the real axis, $|\mathrm{Im}\ G|$ along the real axis, and Σ in the complex plane of frequencies. Input data was obtained from a self-consistent GW calculation with $4 \times 4 \times 4$ k-points and 24 ω-points. The parameter N_d was set to 7.

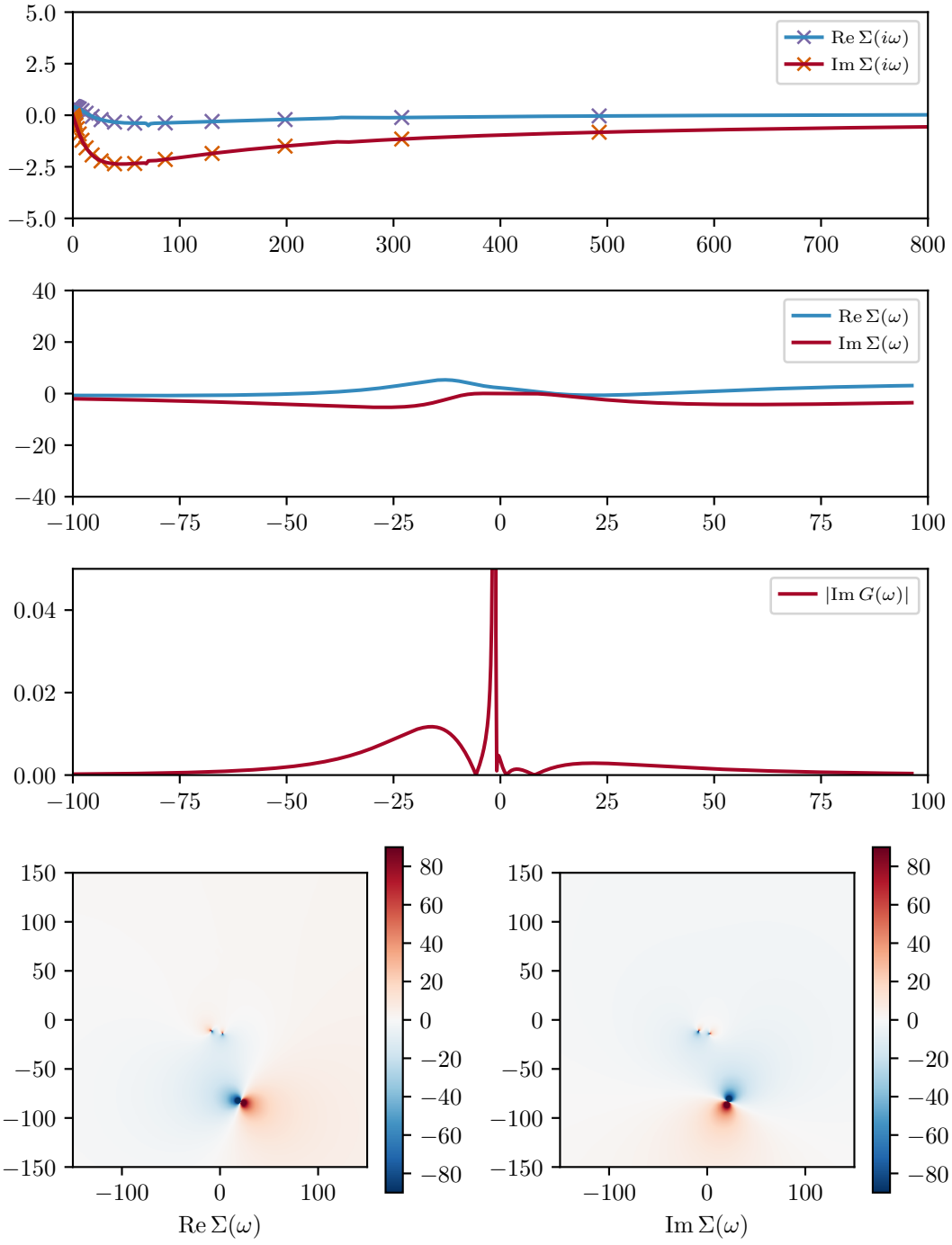

Figure 4.4: QR Padé fitting of Σ for the highest occupied band in Si at Γ. From top to bottom: Σ along the imaginary axis, Σ along the real axis, $|\mathrm{Im}\, G|$ along the real axis, and Σ in the complex plane of frequencies. Input data was obtained from a self-consistent GW calculation with $4 \times 4 \times 4$ k-points and 24 ω-points. The parameter N_d was set to 7.

The fit for C in Fig. 4.3 is an example for a case where Padé fitting works well, yielding a smooth Green's function. The fit for Si in Fig. 4.4, however, shows a potential problem that can arise: the imaginary part of the Green's function changes sign several times to the right of the quasiparticle peak. This is noticeable in the plot of its absolute value as small spikes. Such sign change problems can occur with this fitting approach, depending on the input data and on N_d. However, they have little influence on the position of the quasiparticle peak.

In general, it was found that the results for the quasiparticle energies are numerically stable for bands near the Fermi level in solids, whereas deep bands are more difficult to handle (see the results in section 5.3 for more details).

4.4 Further Numerical Improvements

There are several possible ways to further improve the conditioning number of the fitting matrix, in addition to the change from the monomial basis to the Lagrange basis. These include the choice of nodes for the Lagrange polynomials and the choice of the fixed coefficient in the denominator of Eq. (4.10). For the choice of the Lagrange nodes, several possibilities were tested. The best conditioning numbers were obtained using quadratic (or near-quadratic) spacing, which means that the nodes t_i are calculated as a linear combination

$$t_k = \lambda_k \, t_{\min} + (1 - \lambda_k) \, t_{\max} \tag{4.13}$$

with coefficients that increase quadratically with k:

$$\lambda_k = \left(\frac{k}{n} \right)^2 . \tag{4.14}$$

Fig. 4.5 shows Lagrange polynomials with such a spacing. Qualitatively, it can be seen that the polynomials still tend to vary over several orders of magnitude. This is not ideal for the conditioning, but much better than with the monomial basis.

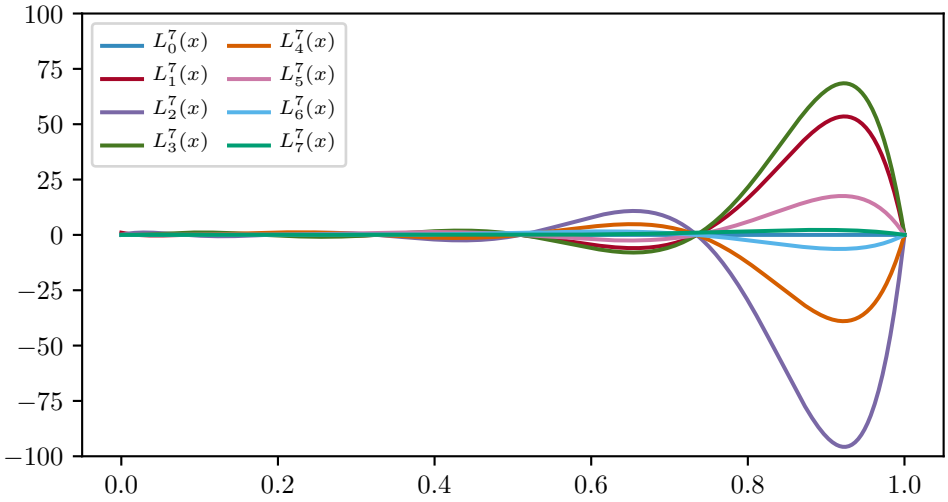

Figure 4.5: Lagrange polynomials with nodes chosen according to the fomula in Eq. (4.13). In this plot, $n = 7$, $t_{\min} = 0$ and $t_{\max} = 1$.

Regarding the choice of the fixed coefficient in the denominator of Eq. (4.10), also several choices were tested. In general, it was found to be best to keep the coefficient corresponding to the Lagrange polynomial for the rightmost node fixed, at least when the nodes are non-uniformly spaced with fewer nodes on the right (which is the case for the spacing in Eq. (4.13)). However, the impact of this choice on the conditioning is generally small.

Another possible way to improve the conditioning is reweighting the rows in the fitting matrix, in order to reduce the range of values. A reweighting that ensures that the row averages are uniform can be obtained by dividing each row by its own average,

$$A_{ij} \to \frac{A_{ij}}{\sum_k A_{ik}/n}. \tag{4.15}$$

The result of such a reweighting is shown in Fig. 4.6. Qualitatively, the matrix now looks more uniform than in Fig. 4.6. However, such a reweighting scheme might distort the results of the fit. While fully-determined systems of equations are invariant under reweighting of rows, over-determined systems (such as this fitting problem) are generally not. Therefore, reweighting is a risky decision and should be avoided if possible. For the calculations presented in chapter 5 of this thesis, reweighting was not used.

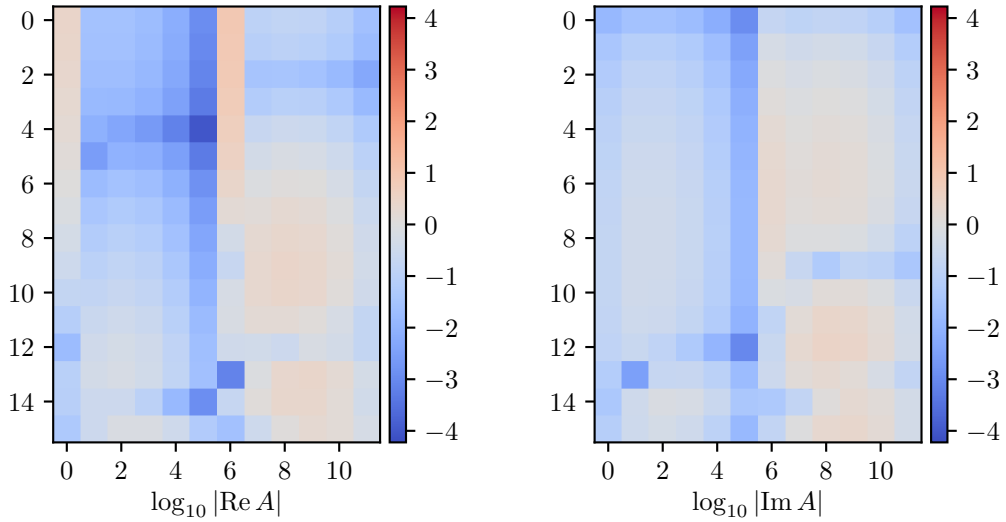

Figure 4.6: Logarithmic plot of the real and imaginary part of the fitting matrix for a Padé fit in Lagrange basis, with 16 data points and N_d set to 7 and with reweighting applied according to Eq. (4.15). The data used for the fit is the same as in Fig. 4.1.

4.5 Continued Fraction Fitting

Another method for generating Padé fits is by expanding the Padé functions as continued fractions of the form

$$P_M(z) = \cfrac{a_1}{1 + \cfrac{a_2(z - z_1)}{1 + \cfrac{a_3(z - z_2)}{\ddots \; 1 + \cfrac{a_M(z - z_{M-1})}{1 + (z - z_M)g_{M+1}(z)}}}}, \tag{4.16}$$

where the coefficients a_n and the functions g_n can be determined via

$$a_n = g_n(z_n), \tag{4.17}$$

$$g_1(z_n) = g_n(y_n), \tag{4.18}$$

$$g_n(z) = \frac{g_{n-1}(z_{n-1}) - g_{n-1}(z)}{(z - z_{n-1})g_{n-1}(z)}. \tag{4.19}$$

This is known as Thiele's reciprocal difference method (TRD). The P_M obtained from this method fulfil $P_M(z_n) = y_n$ exactly at all of the given input points. Therefore, these functions are interpolations rather than fits, and the number of degrees of freedom is always M. This provides less flexibility than the previously discussed methods, where the number of degrees of freedom can be chosen at will. However, this method has the advantage of being generally more numerically stable. This method was already implemented in VASP. The existing implementation was used in the calculations shown in chapter 5.

5 Calculations

5.1 Computational Details

GW calculations were performed with the Vienna *Ab Initio* Simulation Package (VASP) for 15 different semiconductors and insulators, which are listed in Tab. 5.1. Highly accurate norm-conserving PBE [26] PAW [27,28] potentials were used for all materials.

Table 5.1: Structures, lattice constants and energy cutoffs for the materials used. All numbers are given to all digits as used in the calculations.

	structure	a [Å]	E_{cut} [eV]
BN	zinc blende	3.61590	700.000
C	diamond	3.566986	741.689
SiC	zinc blende	4.35810	741.689
MgO	rock salt	4.21	821.518
GaN	zinc blende	4.534993	801.992
ZnO	zinc blende	4.584200	802.268
Si	diamond	5.43092	609.828
AlP	zinc blende	5.46250	616.622
AlAs	zinc blende	5.6610	613.911
InP	zinc blende	5.86870	616.622
AlSb	zinc blende	6.13550	571.798
CdS	zinc blende	5.818	657.510
ZnS	zinc blende	5.40930	802.268
GaP	zinc blende	5.4508	801.992
InSb	zinc blende	6.479370	561.758

The methods used were GW, G_0W_0, and QPGW. In each calculation, a preparatory step was performed using DFT with a plane-wave basis. The resulting Kohn-Sham orbitals

were used as starting point. An exception for this was made for InSb, because DFT yields qualitatively wrong results for InSb, predicting the band gap to be zero. Therefore, for InSb, the preparatory step was done using the HSE hybrid-functional approach [29].

For both GW and G_0W_0, the VASP space-time GW implementation [20] was used. The discretization of ω was done with 24 frequency points for these methods. For QPGW, a reciprocal-space implementation was used and ω was discretized using 128 frequency points. For the calculation of the polarizabilities, the PEAD approach (perturbation expansion after discretization) was used in all cases [30]. For the calculation of the head of the dielectric function in the limit $\mathbf{q} \to 0$ in the fully self-consistent GW calculations, quadratic fitting from finite $\mathbf{q}$ was used (as described in chapter 3). For the analytic continuation of Σ, both the TRD continued fraction method and the quadruple-precision QR Padé fitting in the Lagrange basis (as described in chapter 4) were used, in order to be able to compare the two. For the QR fit, the parameter N_d was set to 7 in all cases. The number of iterations used for the self-consistent GW and QPGW iterations was 5. This was found to be enough to get convergence for the quasiparticle energies.

The GW and G_0W_0 calculations were done once with $4 \times 4 \times 4$ k-points with the default energy cutoff for each material, as listed in Tab. 5.1, as well as three times with $3 \times 3 \times 3$ k-points at higher energy cutoffs, in order to obtain corrections (see the next section for details). The QPGW calculations were done with $6 \times 6 \times 6$ k-points at the default cutoffs and with $3 \times 3 \times 3$ k-points for the corrections. In all cases, the k-points correspond to Γ-centered Monkhorst-Pack grids [22].

5.2 Energy Cutoff Corrections

In any calculation with a finite basis set, there is a small error that results from the fact that only basis functions up to a certain cutoff energy E_{cut} can be included. For any energy E obtained from GW calculations, the difference between its true value E_∞ and its actual value E_{red} in a reduced basis set has the behavior

$$|E_\infty - E_{\text{red}}| \propto \left(\frac{1}{E_{\text{cut}}}\right)^{\frac{3}{2}} \propto \frac{1}{N_{\text{pw}}}. \tag{5.1}$$

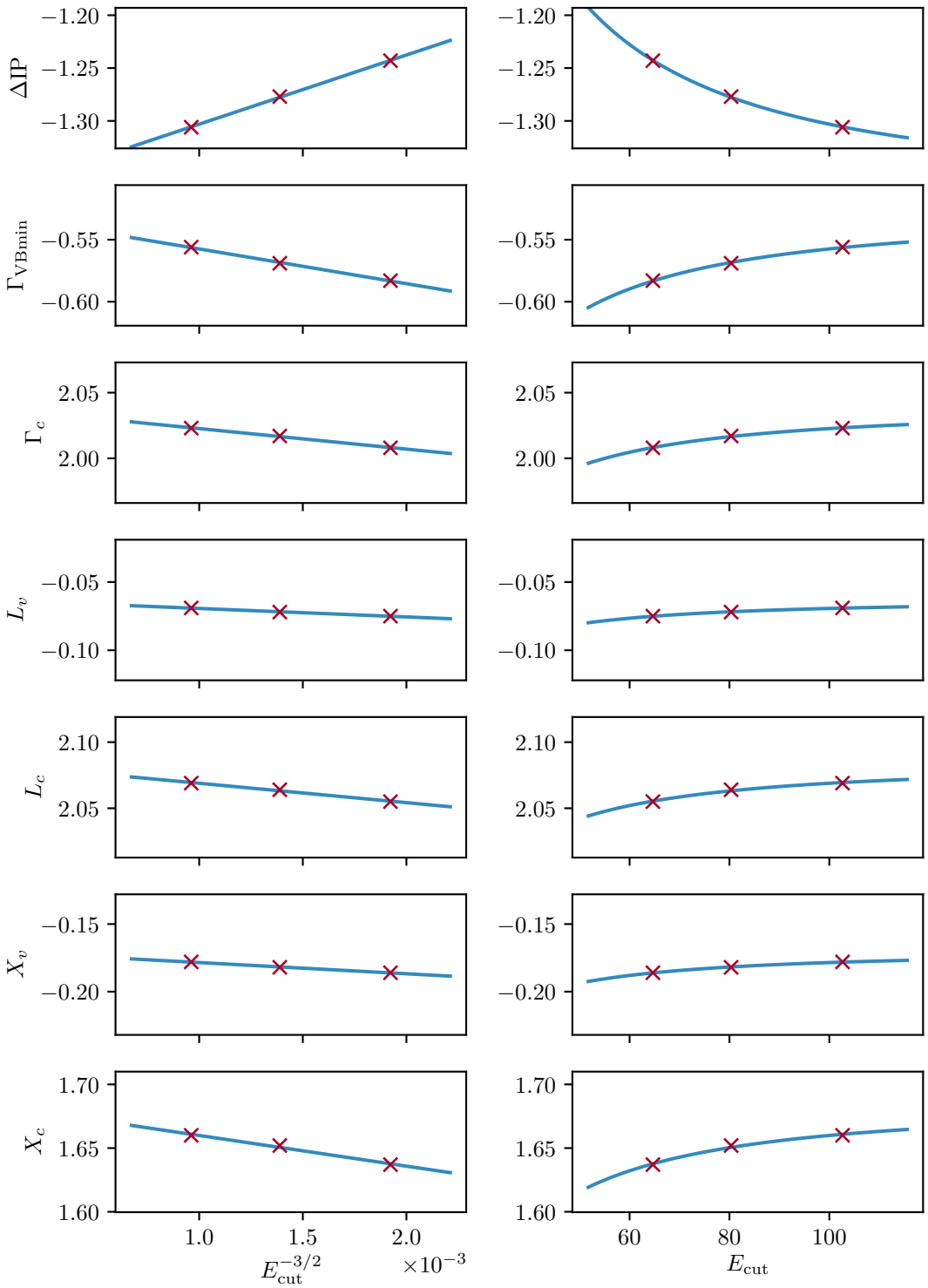

Figure 5.1: Quasiparticle energies for C from G_0W_0 calculations at three different E_{cut} with fitting curves, plotted both as a function of $E_{\mathrm{cut}}^{-3/2}$ (left) and E_{cut} (right).

This makes it possible to fit E as a function of E_{cut} or of $E_{\text{cut}}^{-3/2}$ using a least square fit, if several data points are given at different cutoffs. This function can then be extrapolated to infinity, yielding an estimate for E_∞. In practice, the fitting was done for E as a function of $E_{\text{cut}}^{-3/2}$, since the data is linear in this case, making the fit easier.

All cutoff correction fits were performed with data obtained by running calculations at $3 \times 3 \times 3$ k-points, with three different cutoffs. The values of E_∞ for $4 \times 4 \times 4$ k-points were then approximated as

$$E_\infty(4 \times 4 \times 4) = E_{\text{red}}(4 \times 4 \times 4) - E_{\text{red}}(3 \times 3 \times 3) + E_\infty(3 \times 3 \times 3). \tag{5.2}$$

These corrections were done for all calculations. Fig. 5.1 shows plots with fitting curves for a G_0W_0 calculation for C as an example.

5.3 Results and Discussion

The tables on the following pages contain the quasiparticle energies obtained in the calculations. In these tables, ΔIP is the change of the ionization potential, i.e. the energy of the highest occupied state, compared to its DFT value. This means

$$\Delta\text{IP} = \text{IP}_{GW} - \text{IP}_{\text{DFT}}. \tag{5.3}$$

All other energies are given relative to IP_{GW}. This means that energies below IP_{GW} are negative and energies above IP_{GW} are positive. Γ_{VBmin} is the valence band minimum, i.e. the deepest state, at the Γ point. L_v and X_v are the highest valence band energies at L and X. Γ_c, L_c and X_c are the lowest conduction band energies at Γ, L and X. Because all values are given relative to the ionization potential, the band gaps for all materials can also be read directly from the tables. They are equal to $\min(\Gamma_c, L_c, X_c)$. Furthermore, the direct band gaps at the Γ point are equal to Γ_c. All numbers in the tables are given in eV .

Tables 5.2 and 5.3 contain quasiparticle energies obtained with continued fraction fitting from self-consistent GW calculations, without cutoff corrections and with cutoff corrections, respectively. It can be seen from these tables that the corrections are typically on the order of 100 meV or less. Si stands out as a material for which the corrections are

exceptionally large, indicating that Si can only accurately be described using energy cut-offs larger than 610 eV. It can also be seen that in general, the corrections are relatively small for L_v and X_v.

Tables 5.4 and 5.5 contain data from the same calculations, but obtained using QR Padé fitting instead of continued fraction fitting. In general, the values are close to those obtained from continued fraction fitting. For most states, the differences are below 100 meV, but in some cases, especially for the deep Γ_{VBmin} states, they can be larger.

Tab. 5.6 and Tab. 5.7 contain data from G_0W_0 and QPGW, respectively. Here, only the basis-set corrected values are given. It can be seen that the band gaps in G_0W_0 are smaller than in GW. The band gaps and direct band gaps obtained from the different methods are summarized for all materials in Fig. 5.2 and Fig. 5.3. These figures also include experimental data, taken from the supplemental material of [31]. They are $T = 0$ values for all materials except for ZnS and CdS. For some materials, also data points from a recent paper by Kutepov [6] are included. This paper used a self-consistent GW approach, similar to the present work.

It can be seen from these figures that self-consistent GW calculations overestimate band gaps, compared to the experimental values. Note that the values from [6] that were also obtained with self-consistent GW are lower than the values obtained in this thesis. The difference is most likely due to inaccuracies resulting from the slow k-point convergence of the head calculations for the dielectric matrix in the $\mathbf{q} \to 0$ limit. However, the values from [6] are still larger than the experiental values. Furthermore, self-consistent QPGW also overestimates band gaps, but in general not as strongly as GW. The values from G_0W_0 lie in between those obtained from GW and those obtained from DFT. Of the methods considered here, G_0W_0 comes closest to the experimental values.

Table 5.2: Quasiparticle energies obtained from self-consistent *GW* calculations, with the analytic continuation done using continued-fraction fitting, without energy cutoff corrections.

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c
BN	-2.10	-22.60	13.06	-2.15	14.05	-5.46	7.94
SiC	-1.49	-17.01	9.00	-1.18	8.11	-3.58	3.70
C	-1.60	-23.80	8.84	-3.09	11.90	-6.99	7.46
Si	-0.96	-13.11	4.31	-1.28	3.11	-3.08	2.14
AlP	-1.29	-12.48	5.56	-0.85	4.99	-2.35	3.60
AlAs	-1.31	-12.85	4.32	-0.90	4.30	-2.38	3.42
AlSb	-1.10	-11.61	3.52	-0.98	3.11	-2.39	2.87
InP	-1.08	-12.20	2.34	-1.07	3.18	-2.56	3.30
InSb	-0.86	-11.76	0.75	-1.04	1.69	-2.36	2.43
GaN	-1.92	-18.14	4.58	-1.03	7.79	-2.87	6.20
GaP	-1.16	-13.53	3.88	-1.21	3.60	-2.93	3.37
ZnO	-3.17	-19.11	5.25	-0.80	10.53	-2.19	9.79
ZnS	-1.93	-13.86	5.07	-0.89	6.41	-2.29	6.11
CdS	-1.70	-13.12	3.62	-0.82	5.71	-2.06	5.96
MgO	-3.30	-19.64	9.81	-0.78	13.06	-1.58	14.14

Table 5.3: Quasiparticle energies obtained from self-consistent *GW* calculations, with the analytic continuation done using continued-fraction fitting, with energy cutoff corrections.

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c
BN	-2.22	-22.45	13.18	-2.13	14.08	-5.43	8.03
SiC	-1.53	-17.04	8.97	-1.19	8.11	-3.58	3.68
C	-1.70	-23.88	8.86	-3.09	11.92	-6.97	7.47
Si	-1.19	-13.94	4.71	-1.36	3.46	-3.25	2.44
AlP	-1.39	-12.48	5.60	-0.85	5.07	-2.31	3.71
AlAs	-1.37	-12.81	4.29	-0.89	4.28	-2.37	3.43
AlSb	-1.18	-11.56	3.59	-0.93	3.14	-2.33	2.92
InP	-1.18	-12.44	2.38	-1.04	3.24	-2.50	3.39
InSb	-0.94	-11.68	0.78	-1.31	1.72	-2.35	2.50
GaN	-1.89	-18.50	4.56	-1.02	7.82	-2.87	6.17
GaP	-1.21	-13.49	3.93	-1.16	3.64	-2.93	3.42
ZnO	-3.28	-18.74	5.31	-0.79	10.63	-2.15	9.84
ZnS	-2.01	-13.79	5.22	-0.87	6.55	-2.26	6.21
CdS	-1.74	-13.17	3.75	-0.80	5.84	-2.01	6.06
MgO	-3.40	-19.58	9.80	-0.74	13.07	-1.58	13.92

Table 5.4: Quasiparticle energies obtained from self-consistent *GW* calculations, with the analytic continuation done using quadruple-precision QR Padé fitting in Lagrange basis, without energy cutoff corrections.

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c
BN	-2.10	-22.57	13.06	-2.15	14.05	-5.46	7.94
SiC	-1.50	-16.96	9.00	-1.18	8.10	-3.56	3.71
C	-1.60	-23.86	8.84	-3.09	11.90	-6.98	7.45
Si	-0.99	-12.85	4.34	-1.29	3.14	-3.10	2.16
AlP	-1.30	-12.47	5.56	-0.84	5.00	-2.31	3.61
AlAs	-1.32	-12.84	4.33	-0.90	4.31	-2.38	3.42
AlSb	-1.10	-11.61	3.54	-0.97	3.11	-2.37	2.86
InP	-1.09	-12.38	2.34	-1.07	3.19	-2.55	3.31
InSb	-0.92	-11.78	0.74	-1.05	1.61	-2.34	2.36
GaN	-1.93	-18.13	4.60	-1.03	7.82	-2.86	6.22
GaP	-1.16	-13.30	3.88	-1.23	3.59	-3.02	3.36
ZnO	-3.17	-19.31	5.24	-0.80	10.54	-2.19	9.78
ZnS	-1.93	-13.84	5.07	-0.88	6.41	-2.26	6.11
CdS	-1.70	-13.15	3.62	-0.81	5.71	-2.03	5.96
MgO	-3.30	-19.47	9.82	-0.78	13.05	-1.57	14.13

Table 5.5: Quasiparticle energies obtained from self-consistent *GW* calculations, with the analytic continuation done using quadruple-precision QR Padé fitting in Lagrange basis, with energy cutoff corrections.

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c
BN	-2.22	-22.55	13.18	-2.14	14.09	-5.43	8.04
SiC	-1.55	-16.91	8.88	-1.19	8.11	-3.57	3.72
C	-1.70	-23.86	8.86	-3.08	11.93	-6.95	7.47
Si	-1.17	-13.35	4.67	-1.37	3.41	-3.35	2.44
AlP	-1.41	-12.30	5.63	-0.84	5.10	-2.30	3.73
AlAs	-1.34	-12.91	4.25	-0.89	4.24	-2.35	3.38
AlSb	-1.13	-11.51	3.54	-0.96	3.07	-2.34	2.87
InP	-1.17	-12.51	2.38	-1.04	3.24	-2.49	3.38
InSb	-0.99	-11.69	0.89	-1.02	1.73	-2.29	2.43
GaN	-1.90	-18.48	4.73	-1.02	8.03	-2.84	6.18
GaP	-1.12	-13.36	3.82	-1.21	3.54	-2.99	3.31
ZnO	-3.34	-19.18	5.37	-0.78	10.75	-2.13	9.91
ZnS	-2.02	-13.66	5.24	-0.87	6.57	-2.25	6.28
CdS	-1.76	-13.31	3.77	-0.81	5.84	-2.00	6.09
MgO	-3.47	-19.17	9.89	-0.79	13.07	-1.59	14.11

Table 5.6: Quasiparticle energies obtained from G_0W_0 calculations, with the analytic continuation done using continued-fraction fitting, with energy cutoff corrections.

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c
BN	-1.59	-20.91	11.41	-2.04	12.37	-5.16	6.48
SiC	-1.11	-15.55	7.45	-1.10	6.70	-3.32	2.48
C	-1.31	-21.99	7.57	-2.92	10.49	-6.57	6.33
Si	-0.77	-11.84	3.32	-1.20	2.20	-2.85	1.33
AlP	-1.01	-11.27	4.31	-0.78	3.86	-2.14	2.54
AlAs	-1.07	-11.70	3.08	-0.83	3.17	-2.17	2.38
AlSb	-0.89	-10.61	2.48	-0.91	2.14	-2.21	1.93
InP	-0.83	-11.15	1.35	-0.97	2.15	-2.40	2.35
InSb	-0.59	-10.94	0.67	-1.40	1.31	-2.38	2.00
GaN	-1.26	-15.01	2.98	-0.97	5.99	-2.72	4.57
GaP	-0.91	-12.28	2.72	-1.14	2.53	-2.73	2.36
ZnO	-1.75	-18.80	2.60	-0.79	7.57	-2.31	7.10
ZnS	-1.36	-11.92	3.54	-0.85	4.84	-2.19	4.72
CdS	-1.15	-11.61	2.25	-0.78	4.27	-2.01	4.63
MgO	-2.15	-17.85	7.60	-0.72	10.82	-1.48	11.83

Table 5.7: Quasiparticle energies obtained from QPGW calculations, with the analytic continuation done using continued-fraction fitting, with energy cutoff corrections.

	ΔIP	Γ_{VBmin}	Γ_c	L_v	L_c	X_v	X_c
BN	-2.06	-21.69	12.33	-2.12	13.31	-5.35	7.25
SiC	-1.50	-16.22	8.04	-1.20	7.30	-3.45	3.01
C	-1.61	-22.87	8.15	-3.03	11.12	-7.21	6.81
Si	-1.11	-12.20	3.80	-1.24	2.64	-2.94	1.75
AlP	-1.39	-11.81	4.91	-0.82	4.44	-2.24	3.06
AlAs	-1.43	-12.29	3.67	-0.86	3.69	-2.29	2.85
AlSb	-1.22	-11.07	2.94	-0.94	2.57	-2.29	2.34
InP	-1.21	-11.78	1.80	-1.01	2.65	-2.43	2.79
InSb	-1.11	-11.15	0.66	-1.06	1.44	-2.39	2.20
GaN	-2.22	-17.06	4.16	-0.77	7.39	-2.55	5.71
GaP	-1.23	-12.84	3.23	-1.18	3.01	-2.83	2.78
ZnO	-2.96	-18.99	4.41	-0.81	9.58	-2.28	8.95
ZnS	-1.89	-13.30	4.42	-0.89	5.76	-2.29	5.51
CdS	-1.73	-12.60	3.05	-0.80	5.10	-2.01	5.39
MgO	-3.50	-18.36	9.70	-0.54	12.80	-1.37	13.82

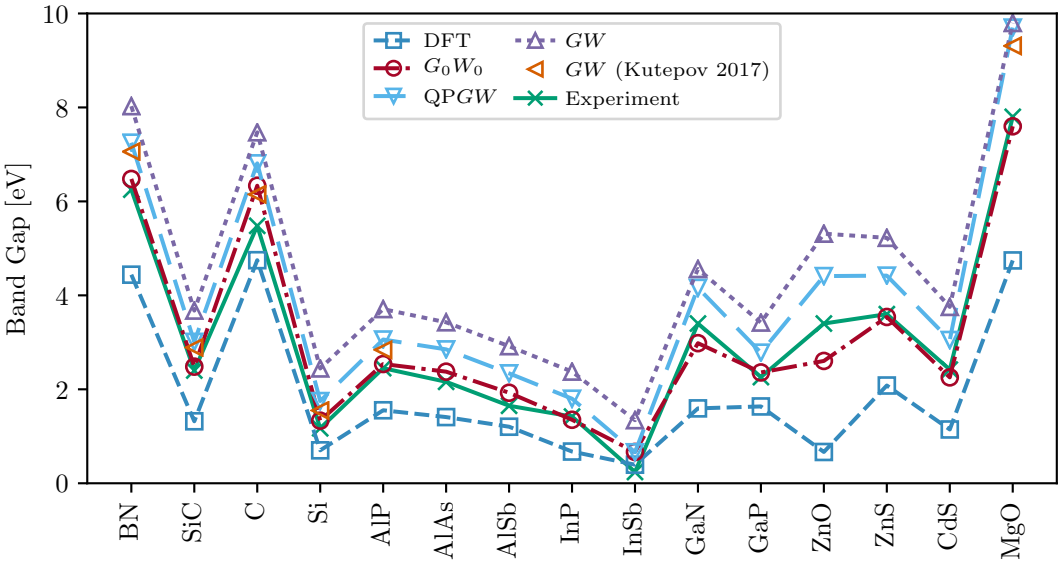

Figure 5.2: Band gaps for all materials considered in this thesis. Note that for InSb, an HSE calculation was used as a starting point.

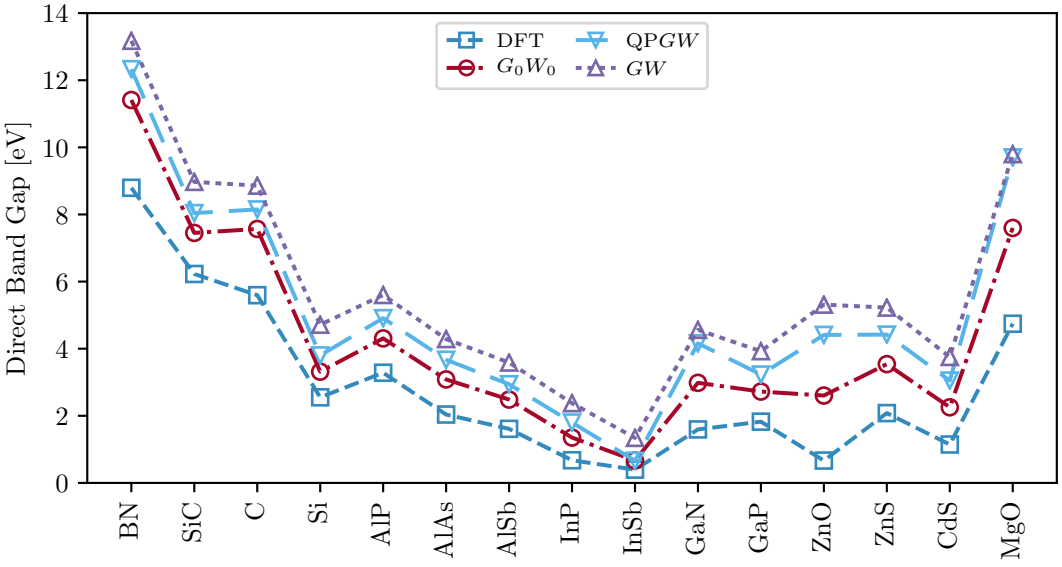

Figure 5.3: Direct band gaps (Γ–Γ) for all materials considered in this thesis. Note that for InSb, an HSE calculation was used as a starting point.

From the analytic continuation methods, it is also possible to obtain Green's functions that are averaged over all k-points and summed over bands. Summing over all occupied states leads to a spectral function that corresponds directly to the spectrum that is measurable in PES. Figures 5.4 and 5.5 show the k-point averaged Green's function summed over all occupied bands for C and for Si, respectively. Note that these plots contain multiple peaks, originating from, different bands and k-points.

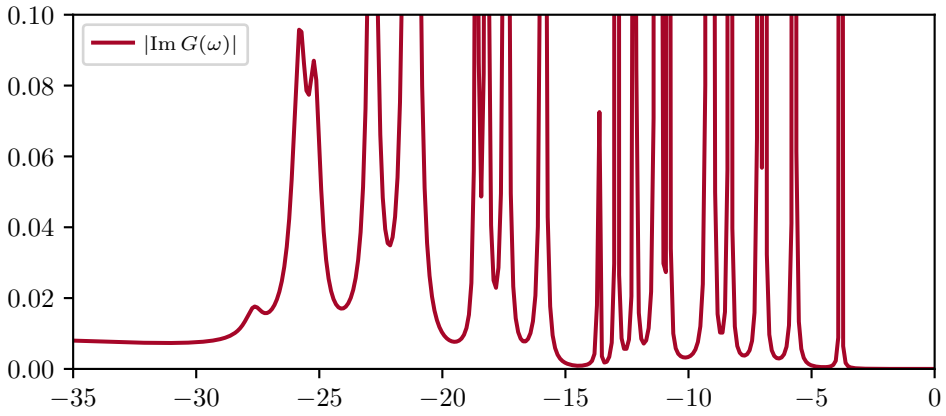

Figure 5.4: k-point averaged Green's function summed over all occupied bands for C, obtained from a self-consistent GW calculation, using quadriple-precision QR Padé fitting with $N_d = 7$.

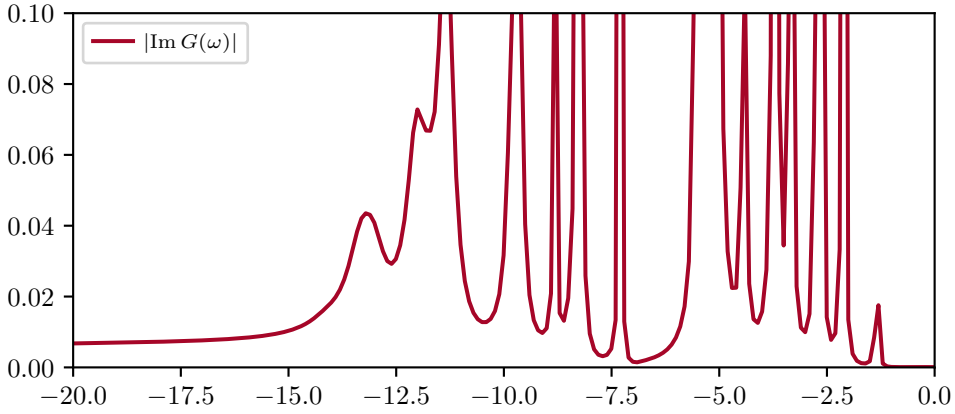

Figure 5.5: k-point averaged Green's function summed over all occupied bands for Si, obtained from a self-consistent GW calculation, using quadruple-precision QR Padé fitting with $N_d = 7$.

In principle it is also possible to get continuous band structure plots from GW calculations. However, this requires many k-points and is therefore computationally demanding. Figures 5.6 to 5.8 therefore contain crude approximations to the continuous band structures along $L - \Gamma - X$ for some of the materials, obtained by spline-fitting from the existing data points. DFT band structures are shown for comparison. Note that these comparisons are only qualitative, firstly because of the small number of k-points used for the fits, and secondly because the values for these plots are not cutoff-corrected. This is because cutoff corrections were done with $3 \times 3 \times 3$ k-points, so using the cutoff correction would mean even fewer data points.

Even though these fits are only qualitative, it can be seen that the band energies are spread out over a larger range in GW, as compared to DFT. The resulting opening of the band gaps is also clearly visible, especially for ZnO, where the DFT gap is very small.

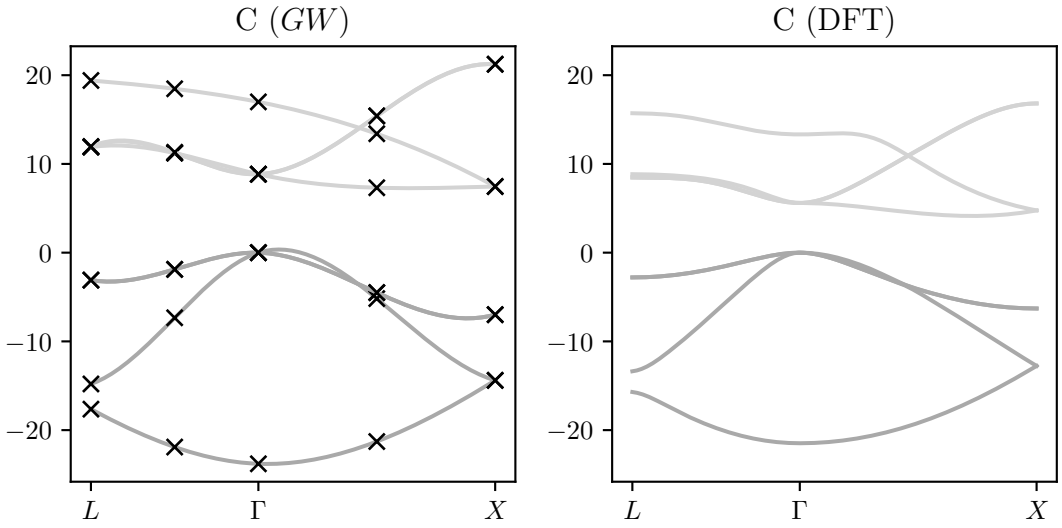

Figure 5.6: Band structure plots for C, with data obtained from a GW calculation without cutoff corrections (left) and from a DFT calculation (right). The lines between the data points were drawn using cubic spline interpolation.

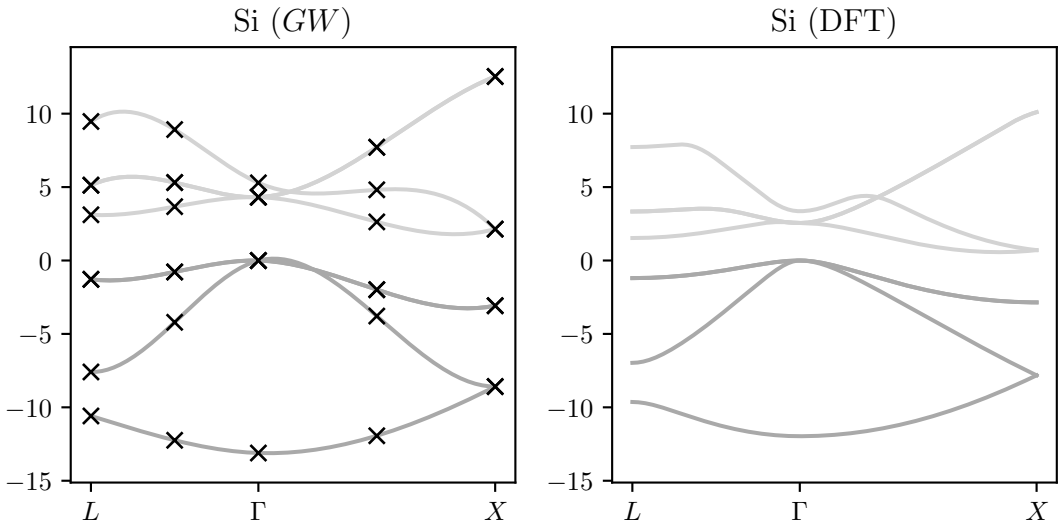

Figure 5.7: Band structure plots for Si, with data obtained from a *GW* calculation without cutoff corrections (left) and from a DFT calculation (right). The lines between the data points were drawn using cubic spline interpolation.

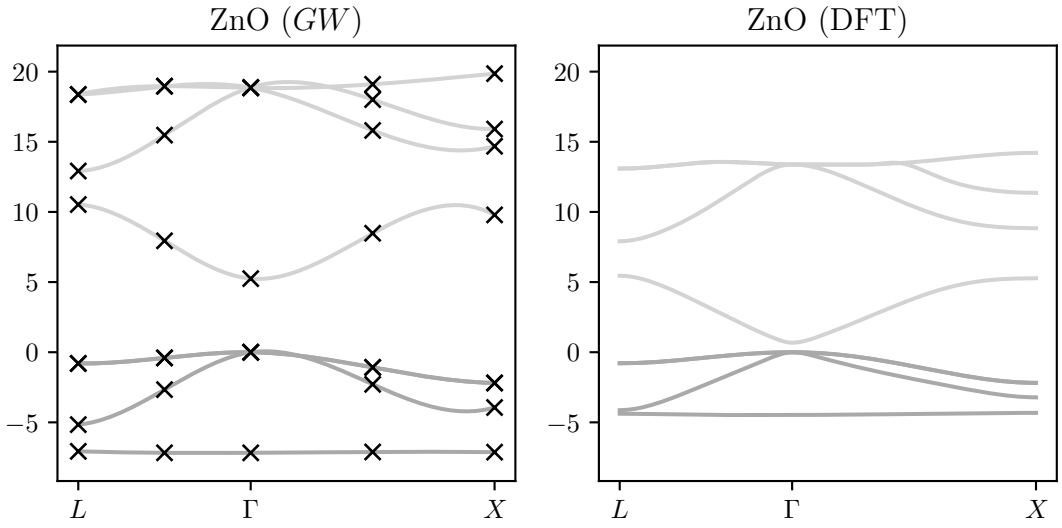

Figure 5.8: Band structure plots for ZnO, with data obtained from a *GW* calculation without cutoff corrections (left) and from a DFT calculation (right). The lines between the data points were drawn using cubic spline interpolation.

6 Conclusion

In this thesis, the problem of the calculation of the head of the dielectric matrix in the limit $\mathbf{q} \to 0$ and the problem of the analytic continuation of Σ from the imaginary frequency axis to the real frequency axis were considered. Both are important components in the implementation of the space-time GW method.

It was found that the inclusion of the head is of great importance for the accuracy of self-consistent calculations. A quadratic fitting method was used to deal with this problem. However, the results obtained with this method were found to be considerably dependent on the size of the k-point mesh.

For the analytic continuation, a Padé fitting routine using a QR algorithm in quadruple precision was developed and tested. It was found that both the use of quadruple precision and the choice of basis for the numerator and denominator polynomials in the Padé function are of utmost importance for the numerical stability of the fit, with the Lagrange basis yielding significantly better stability than the monomial basis. The resulting routine provides more flexibility than continued-fraction fitting with Thiele's reciprocal difference method, since it allows for a free choice of the degree of the polynomials. It yields good results for the quasiparticle energies in most cases, but can disagree with continued-fraction fitting by over 100 meV in some cases.

In chapter 5, results for the quasiparticle energies of 15 semiconductor and insulator materials were presented. The results show that fully self-consistent GW calculations yield significantly higher band gaps than DFT, while the G_0W_0 and QPGW band gaps lie in between. In general, G_0W_0 comes closest to the experimental values, whereas both QPGW and GW overestimate the gaps. For the self-consistent GW calculations, the accuracy of the results was found to be limited due to the slow k-point convergence of the head of the dielectric function.

References

[1] M. van Schilfgaarde, T. Kotani, S. Faleev, Quasiparticle self-consistent *GW* theory, Phys. Rev. Lett. 96 (2006) 226402. doi:10.1103/PhysRevLett.96.226402.

[2] F. Caruso, P. Rinke, X. Ren, M. Scheffler, A. Rubio, Unified description of ground and excited states of finite systems: The self-consistent *GW* approach, Phys. Rev. B. 86 (2012) 081102. doi:10.1103/PhysRevB.86.081102.

[3] B. Holm, U. von Barth, Fully self-consistent *GW* self-energy of the electron gas, Phys. Rev. B. 57 (1998) 2108–2117. doi:10.1103/PhysRevB.57.2108.

[4] W.-D. Schöne, A.G. Eguiluz, Self-consistent calculations of quasiparticle states in metals and semiconductors, Phys. Rev. Lett. 81 (1998) 1662–1665. doi:10.1103/PhysRevLett.81.1662.

[5] A.L. Kutepov, Electronic structure of Na, K, Si, and LiF from self-consistent solution of Hedin's equations including vertex corrections, Phys. Rev. B. 94 (2016) 155101. doi:10.1103/PhysRevB.94.155101.

[6] A.L. Kutepov, Self-consistent solution of Hedin's equations: Semiconductors and insulators, Phys. Rev. B. 95 (2017) 195120. doi:10.1103/PhysRevB.95.195120.

[7] J. Klimeš, M. Kaltak, G. Kresse, Predictive *GW* calculations using plane waves and pseudopotentials, Phys. Rev. B. 90 (2014) 075125. doi:10.1103/PhysRevB.90.075125.

[8] A. Szabo, N.S. Ostlund, Modern quantum chemistry: Introduction to advanced electronic structure theory, 1st ed., revised, Mineola, NY, Dover Publications, 1996 [1982].

[9] T. Koopmans, Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen eines Atoms, Physica. 1 (1934) 104–113. doi:10.1016/S0031-

8914(34)90011-2.

[10] P. Hohenberg, W. Kohn, Inhomogeneous electron gas, Phys. Rev. 136 (1964) B864–B871. doi:10.1103/PhysRev.136.B864.

[11] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects, Phys. Rev. 140 (1965) A1133–A1138. doi:10.1103/PhysRev.140.A1133.

[12] R.M. Martin, Electronic structure: Basic theory and practical methods, 1st ed., Cambridge, Cambridge University Press, 2004.

[13] B.G. Janesko, T.M. Henderson, G.E. Scuseria, Screened hybrid density functionals for solid-state chemistry and physics, Phys. Chem. Chem. Phys. 11 (2009) 443–454. doi:10.1039/b812838c.

[14] F. Reinert, S. Hüfner, Photoemission spectroscopy—from early days to recent applications, New Journal of Physics. 7 (2005) 97. doi:10.1088/1367-2630/7/1/097.

[15] V.M. Galitskii, A.B. Migdal, Application of quantum field theory methods to the many body problem, Sov. Phys. JETP. 7 (1958) 96–104. http://www.jetp.ac.ru/cgi-bin/e/index/e/7/1/p96?a=list.

[16] L. Hedin, New method for calculating the one-particle Green's function with application to the electron-gas problem, Phys. Rev. 139 (1965) 796–823. doi:10.1103/PhysRev.139.A796.

[17] R. Starke, G. Kresse, Self-consistent Green function equations and the hierarchy of approximations for the four-point propagator, Phys. Rev. B. 85 (2012) 075119. doi:10.1103/PhysRevB.85.075119.

[18] F. Aryasetiawan, O. Gunnarsson, The GW method, Reports on Progress in Physics. 61 (1998) 237–312. doi:10.1088/0034-4885/61/3/002.

[19] G. Onida, L. Reining, A. Rubio, Electronic excitations: Density-functional versus many-body Green's-function approaches, Rev. Mod. Phys. 74 (2002) 601–659. doi:10.1103/RevModPhys.74.601.

[20] P. Liu, M. Kaltak, J. Klimeš, G. Kresse, Cubic scaling GW: Towards fast quasiparticle calculations, Phys. Rev. B. 94 (2016) 165109. doi:10.1103/PhysRevB.94.165109.

[21] M. Gajdoš, K. Hummer, G. Kresse, J. Furthmüller, F. Bechstedt, Linear optical properties in the projector-augmented wave methodology, Phys. Rev. B. 73 (2006)

045112. doi:10.1103/PhysRevB.73.045112.

[22] H.J. Monkhorst, J.D. Pack, Special points for Brillouin-zone integrations, Phys. Rev. B. 13 (1976) 5188–5192. doi:10.1103/PhysRevB.13.5188.

[23] H.N. Rojas, R.W. Godby, R.J. Needs, Space-time method for ab initio calculations of self-energies and dielectric response functions of solids, Phys. Rev. Lett. 74 (1995) 1827–1830. doi:10.1103/PhysRevLett.74.1827.

[24] M. Jarrell, J. Gubernatis, Bayesian inference and the analytic continuation of imaginary-time quantum monte carlo data, Physics Reports. 269 (1996) 133–195. doi:10.1016/0370-1573(95)00074-7.

[25] A.S. Householder, Unitary triangularization of a nonsymmetric matrix, J. ACM. 5 (1958) 339–342. doi:10.1145/320941.320947.

[26] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868. doi:10.1103/PhysRevLett.77.3865.

[27] P.E. Blöchl, Projector augmented-wave method, Phys. Rev. B. 50 (1994) 17953–17979. doi:10.1103/PhysRevB.50.17953.

[28] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B. 59 (1999) 1758–1775. doi:10.1103/PhysRevB.59.1758.

[29] J. Heyd, G.E. Scuseria, M. Ernzerhof, Hybrid functionals based on a screened coulomb potential, The Journal of Chemical Physics. 118 (2003) 8207–8215. doi:10.1063/1.1564060.

[30] R.W. Nunes, X. Gonze, Berry-phase treatment of the homogeneous electric field perturbation in insulators, Phys. Rev. B. 63 (2001) 155107. doi:10.1103/PhysRevB.63.155107.

[31] A. Grüneis, G. Kresse, Y. Hinuma, F. Oba, Ionization potentials of solids: The importance of vertex corrections, Phys. Rev. Lett. 112 (2014) 096401. doi:10.1103/PhysRevLett.112.096401.