



UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Faculté des Sciences Appliquées
CERMIN
Unité PCPM

CORE-ELECTRONS AND SELF-CONSISTENCY IN THE *GW* APPROXIMATION FROM A PAW PERSPECTIVE

Dissertation présentée en vue de l'obtention du grade de
Docteur en Sciences de l'Ingénieur

par

Matteo Giantomassi

Promoteurs:

Prof. Xavier Gonze

Prof. Gian-Marco Rignanese

May 2009



UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Faculté des Sciences de l'Ingénieur

MAPR

Unité PCPM

IMPLEMENTATION OF THE GW FORMALISM
WITHIN THE PROJECTOR AUGMENTED WAVE
METHOD

Dissertation présentée en vue de l'obtention du grade de

Docteur en Sciences de l'Ingénieur

par

Matteo Giantomassi

Membres du jury:

Prof. Xavier Gonze, promoteur

Prof. Gian-Marco Rignanese, promoteur

Prof. Jean-Christophe Charlier, président

Prof. Giovanni Battista Bachelet

Prof. Rex Godby

Prof. Valerio Olevano

May 2009

To my parents

Contents

0.1	Acronyms	iv
1	Resume	vii
2	Preface	xi
3	Density Functional Theory	1
3.1	The many-body problem	1
3.2	Density Functional Theory	2
3.2.1	The Hohenberg-Kohn theorem (1964)	2
3.2.2	The Kohn-Sham equations (1965)	4
3.3	Approximations for the Exchange-Correlation functional	6
3.4	Successes and failures of DFT: the band gap problem	7
4	An Overview of Many-Body Theory	11
4.1	Introduction	11
4.2	The Green's Function	12
4.3	The Lehmann representation	13
4.4	The spectral function	14
4.5	Equation of motion of G and the concept of self-energy	16
4.6	Hedin's Equations	18
4.7	The GW approximation	21
4.8	Perturbative Approach	24
4.9	Hartree-Fock as a limiting case of GW	25
4.10	Quasiparticle self-consistent GW	26
4.11	Core-valence decoupling	27
5	GW Implementation	33
5.1	Evaluation of the Screened Interaction	33
5.2	RPA for the Irreducible Polarizability	35
5.3	Scaling of the RPA computation in Fourier space	41
5.4	Symmetries of the irreducible polarizability	43
5.4.1	Symmetries in $\mathbf{q}$ -space	43
5.4.2	Magnetic Symmetries	44

5.4.3	Symmetries in $\mathbf{k}$ -space	46
5.5	Plasmon-Pole Models	48
5.5.1	Godby-Needs and Hibertsen-Louie Plasmon-Pole Models	49
5.5.2	Von der Linden-Horsch and Engel-Farid Plasmon-Pole Models	51
5.5.3	Pros and Cons of the Plasmon-Pole Technique	54
5.6	Spectral Method for the Polarizability	55
5.7	Evaluation of the GW self-energy	58
5.7.1	Exchange Part of the Self-Energy	59
5.7.1.1	Symmetrization of $\hat{\Sigma}_x$	62
5.7.2	The Correlation Part of the Self-Energy	63
5.7.2.1	Plasmon-Pole	66
5.7.2.2	Contour Deformation	68
5.7.2.3	Analytic Continuation	70
5.8	Contour Deformation vs Analytic Continuation	72
5.9	QPSCGW using Kohn-Sham states as basis set	76
6	GW Calculations in Silicon Dioxide	79
6.1	On the Reliability of the Different Plasmon-Pole Models	79
6.2	GW Results for α -SiO ₂	83
7	The GW method within the PAW formalism	93
7.1	Introduction	93
7.2	The PAW formalism	93
7.2.1	Remarks on the notation and terminology	97
7.3	Expectation values	98
7.4	Charge Densities in the PAW formalism	99
7.5	PAW Hamiltonian	102
7.5.1	Treatment of the Different Contributions to D_{ij}	104
7.6	Treatment of v_{xc} and v_H in PAW	105
7.6.1	Onsite Exchange-Correlation Potentials	105
7.6.2	Onsite Hartree Potentials	109
7.7	Oscillator matrix elements within PAW	110
7.8	On the Symmetrization of the PAW projections	111
7.9	Testing the GW + PAW Implementation	112
7.10	QPSCGW with PAW	117
8	Wannier Interpolation	121
8.1	Introduction	121
8.2	The Wannier Representation	121
8.3	Wannier Interpolation of Matrix Elements	123
8.4	Wannier interpolation of GW results	124
9	Conclusions and Perspectives	129

A	Fourier Transform	133
A.1	Conventions	133
A.2	Properties	133
A.3	Fourier series for periodic lattice quantities	133
B	Symmetry Properties	135
B.1	Effect of space group symmetries on electron energies and wavefunctions .	135
B.2	Magnetic Space Groups	137
B.3	Symmetries of the oscillator matrix elements	138
B.3.1	Symmetrization in $\mathbf{k}$ -space	138
B.3.2	Symmetrization in $\mathbf{q}$ -space	139
C	Optical properties within the pseudopotential formalism	141
C.1	Norm-conserving Pseudopotentials in the Kleyman Bylander form	141
C.2	Treatment of the long-wavelength limit	142
D	Complex and Real Spherical Harmonics	149
D.1	Complex Spherical Harmonics	149
D.2	Real spherical harmonics	150
D.3	Symmetry transformation of RSH	151
D.4	Useful expansions in terms of RSH	152
E	Optical properties within PAW	153
E.1	Matrix elements of the gradient operator between PAW partial waves . . .	153
E.2	Optical limit within LDA+U	155
F	List of Publications	157

0.1 Acronyms

HF	Hartree-Fock
DFT	Density Functional Theory
HK	Hohenberg and Kohn
KS	Kohn and Sham
XC	Exchange-Correlation
LDA	Local Density Approximation
GGA	Generalized Gradient Approximation
TDDFT	Time-Dependent Density Functional Theory
MBPT	Many-Body Perturbation Theory
QP	Quasiparticle
RPA	Random Phase Approximation
PPM	Plasmon-Pole Model
CD	Contour Deformation
AC	Analytic Continuation
KB	Kleynman and Bylander
NC	Norm-Conserving
RSH	Real Spherical Harmonic
AE	All-Electron
PAW	Projector Augmented Wave
BZ	first Brillouin zone
IBZ	irreducible Brillouin zone
FFT	Fast Fourier Transform
QPSCGW	Quasiparticle Self-Consistent GW
GW	GW approximation
WFs	Wannier Functions
MLWFs	Maximally Localized Wannier Functions
PPs	pseudopotentials

Nomenclature

$\chi^0(12)$ Independent particle polarizability

$\epsilon(12)$ Dielectrix matrix

$\epsilon^{-1}(12)$ Inverse dielectrix matrix

$\Gamma(12;3)$ Vertex function

$G_0(12)$ Independent particle Green's function

$\hat{\mathcal{R}}_{\mathbf{t}}$ Symmetry operation of the space group

$\hat{H}$ Hamiltonian operator

$\mathbf{K}$ $\mathbf{k} + \mathbf{G}$

$\mathcal{G}$ Gaunt coefficient

$\mathcal{R}$ Real orthonormal matrix representing a proper or improper rotation

Ω Volume of the unit cell in real space

$\phi_i(\mathbf{r})$ AE PAW partial wave

$\mathcal{Y}_m^l(\hat{\mathbf{r}})$ Real spherical harmonic

$\Sigma(12)$ Self-energy

$\tilde{\chi}(12)$ Irreducible polarizability

$\tilde{\epsilon}(12)$ Symmetrized dielectrix matrix

$\tilde{\epsilon}^{-1}(12)$ Symmetrized inverse dielectrix matrix

$\tilde{\phi}_i(\mathbf{r})$ Pseudo PAW partial wave

$\tilde{p}_i(\mathbf{r})$ PAW projector

$\mathbf{t}$ Fractional translation

$v(\mathbf{r})$ Bare Coulomb interaction

$v_{\text{xc}}(\mathbf{r})$	Exchange-correlation potential
$Y_m^l(\hat{\mathbf{r}})$	Complex spherical harmonic
$G(12)$	One-particle many-body Green's function
$G(1234)$	Two-particle many-body Green's function
V	Volume of the Born-von Kármán supercell
$W(12)$	Screened Coulomb interaction

Chapter 1

Resume

Density functional theory (DFT) performs reasonably well for the determination of structural properties of many materials, but fails to predict electronic band gap values accurately. Such a failure of DFT is not unexpected, since there exists no formal justification for interpreting the DFT eigenvalues as addition or removal energies of the many-body system (quasiparticle energies). An alternative approach to the study of exchange-correlation effects in many-particle systems is provided by many-body perturbation theory (MBPT), which defines a rigorous approach to description of excited-state properties based on the Green's function formalism and the concept of quasiparticle electrons. Within MBPT, one can calculate the quasiparticle (QP) energies and the QP amplitudes by approximately solving, within the so-called *GW* approximation, the set of coupled integro-differential equations proposed by Hedin in 1965. The *GW* method generally yields significantly better values for QP energies with respect to DFT as it accounts for the dynamic many-body effects in the electron-electron interaction going beyond the mean-field approximation of independent Kohn-Sham particles. The drawback of such an involved approach, however, is its large computational cost, which is mainly due to the evaluation of dielectric matrices, their inversion, and the solution of a non-Hermitian and nonlinear eigenvalue problem.

The first part of this PhD research is devoted to the implementation of an efficient and scalable coarse-grained parallel algorithms and the development of state-of-the-art methods to solve the *GW* equations. All these techniques are then applied to the study of the quasiparticle band structures of SiO₂ in the alpha quartz crystalline structure. The effects of the different approximations involved in the theory, the influence of self-consistency, and a systematic analysis of the reliability of the different plasmon-pole models are presented and discussed. The second part of this PhD work is dedicated to the implementation of a methodological approach to the solution of the *GW* equations based on the so-called projector augmented wave method (PAW) proposed by Blochl in 1994. This new approach permits to remove many of the limitations intrinsic to the use of the pseudopotential technique widely used for *ab-initio* calculations, allowing one to reach *GW* results closer to an all-electron implementation while still maintaining computational efficiency, and flexibility. This allows us to achieve a coherent implementation of many

different algorithms which enables a detailed comparison of GW calculations not yet performed until now. Results for the quasiparticle band structure and optical spectra of prototype s - and p -compounds are discussed and compared to recent studies reported in the scientific literature (when available).

Chapter 2

Preface

In spectroscopic experiments, a sample is excited by means of an external perturbation, for instance by incoming photons or by an external electron source, and the response of the system is measured. The results of such measurements are key for our understanding of the electronic structure of matter, as they give direct access to crucial information about the properties of the sample. They also shed light on the underlying physical laws governing the process. A well known example comes from early studies of the optical excitations in the hydrogen atom, whose interpretation led to the development of the quantum theory of electronic states. This mode of discovery has continued over the last century up to now thanks to important improvements in the experimental techniques employed to measure electronic excitations. Many contributions come from the progress made at synchrotron-radiation sources and with ultrafast lasers. Thanks to these new technological progresses, high-brilliance sources are now commonplace and photon beams in the extreme ultraviolet region are increasingly used, allowing for high quality measurements characterized by high spatial resolution and short measurement times.

Nowadays, most of our current understanding of the electronic structure of materials of atoms, molecules and solids comes from examining their optical and dielectric properties. The rapid and astonishing improvement of experimental techniques continuously stimulates a more precise theoretical understanding, as well as the development of new techniques to tackle the complicated many-body problem involved in the description of excitations. For this reason, theoretical spectroscopy has become one of the most active research fields in condensed matter physics. In particular, parameter-free approaches – also called *ab initio* methods – are commonly employed as a powerful tool to improve our understanding of the physical properties of real materials. The joint use of theory and computer simulations provides a direct interpretation of experimental data, and it allows us to formulate predictions about the electronic and optical properties of new compounds when experimental results are missing.

Unfortunately, studying optical excitations using contemporary standard quantum mechanical methods represents a challenging task due to the level of accuracy required by present scientific and technological needs. The widely used density functional theory (DFT) – one of the most successful approaches to the study of quantum mechanical

properties of many-body systems – performs reasonably well for the determination of structural properties of many materials, but fails to predict band gap values accurately. Such a failure of DFT is not unexpected, since there exists no formal justification for interpreting the DFT eigenvalues as addition or removal energies of the many-body system (quasiparticle energies). Reliable theoretical simulations require a much more involved approach to the many-body problem. The time-dependent extension to static density functional theory, Time-Dependent Density Functional Theory (TDDFT) [1] has become a promising and appealing approach to the study of optical properties in finite systems [2], but unfortunately, it does not provide results of the same quality when applied to extended systems [2].

An alternative, more traditional, approach to the study of exchange-correlation effects in many-body systems is provided by Many-Body Perturbation Theory (MBPT) [3, 4] which defines a rigorous framework for the description of excited-state properties based on the Green’s function formalism. Within MBPT, one can calculate charged excitations with good precision (*i.e.* electron addition and removal energies), using *e.g.* Hedin’s *GW* approximation [5, 6] for the electron self-energy. In the same framework, neutral excitations (experimentally related to optical and energy-loss spectra) are also well described through the solution of the Bethe-Salpeter equation [7, 8]. The *GW* approximation generally yields significantly better values for quasiparticle energies with respect to DFT as it accounts for many-body effects in the electron-electron interaction going beyond the mean-field approximation of independent particles. This improved description is achieved by screening the nonlocal exchange interaction with the inverse frequency-dependent dielectric matrix, which is a much more sophisticated approach when compared to the effective local exchange-correlation potentials of DFT. The drawback of such an involved approach, however, is its large computational cost, which is mainly due to the evaluation of dielectric matrices, their inversion, and the solution of a non-Hermitian and nonlinear eigenvalue problem. The increase of computing power has, only in recent years, made numerical calculations within MBPT feasible for real systems. Today, the *GW* method has been extensively used to study physical properties of different types of materials ranging from the electronic properties of atoms, clusters, bulk materials, super-lattices, up to such properties as the valence off-set between different bulk semiconductors [2]. All these different studies have established the *GW* approximation as an accurate and reliable first-principles method for the computation of quasiparticle properties.

However, several issues related to the particular *ab initio* implementations, and the approximations involved in the solution of the *GW* equations still remain to be investigated. Most of the original *GW* implementations were based on a plane-wave expansion of the Bloch wave function, and employed norm-conserving pseudopotentials within the frozen core approximation to reduce the computing effort. For several decades, thanks to the excellent agreement obtained with respect to experiments, pseudopotential-based methods have represented a *de facto* standard for many-body calculations. In recent years, however, elaborate full-potential linear muffin-tin-orbital FP-LMTO or full-potential linear augmented plane-wave FLAPW methods, have been used in combination with *GW* for precise and unbiased MBPT calculations [9, 10, 11]. *GW* calculations per-

formed with All-Electron (AE) approaches have revealed that a fully consistent treatment of the electronic degrees of freedom produces band gaps that are systematically smaller than the pseudopotential results, thus worsening the agreement between *GW* and experiments. These findings have led to an intense debate in the scientific literature concerning the reliability of the pseudopotential approach for quasiparticle calculations [9, 11]. According to the point of view of the all-electron community, using pseudopotentials for many-body applications leads to a fortuitous improvement that can be traced back both to the neglect of core relaxation effects as well as to spurious effects introduced by the pseudization of the wavefunctions [12]. Systematic studies performed within the norm-conserving pseudopotential approach support these criticisms since it has been shown that results in reasonable agreement with AE calculations can be obtained only by employing highly transferable pseudo-atoms at the price of a dramatic increase in computing effort [13]. Although further investigations are needed in order to assess the reliability of the pseudopotential approach for MBPT applications, one can safely say that, in the case of *GW* calculations, many of the advantages obtained from the pseudization of the wave functions are lost. A *GW* study based on pseudopotentials has to rely on several assumptions which are difficult to control as well as on physical approximations whose impact is hard to estimate.

A very promising alternative to the use of pseudopotentials for the solution of the *GW* equations, first adopted by Arnaud and co-workers in 2000 [14], is based on the so-called Projector Augmented Wave (PAW) formalism proposed by Blöchl in 1994 [15]. The PAW method allows the removal of many of the limitations intrinsic to the use of pseudopotentials, thus enabling, in principle, results of equivalent quality as all-electron approaches. From the FLAPW method, it inherits a fast convergence with respect to the basis set size and, contrary to the norm-conserving formalism, can be efficiently applied to *d*- and *f*-elements or semi-core states. Moreover, the PAW approach constitutes a full potential method which gives direct access to the true single-particle wave function with the correct nodal shape in the atomic regions. This eliminates shape approximations on the charge distribution, thus allowing for an accurate description of physical quantities sensitive to variations in the nuclear region such as, for instance, magnetic properties and optical matrix elements. Compared to other full potential methods, PAW retains most of the simplicity of a pseudopotential approach (fast convergence with respect to the basis set truncation, computational efficiency, and flexibility), allowing one to obtain *GW* results that are closer to an all-electron implementation, with a considerable saving in computational cost. It is the aim of the present manuscript to provide a detailed account of the implementation of the *GW* method in the frequency- and Fourier-space domain, discussing the subtle technical details related to the use of the PAW method.

The thesis is organized as follows. The theoretical framework used throughout this work is described in the first part of the manuscript. Chapter 3 provides a brief introduction to density functional theory. The Green's function formalism and the approach proposed by Hedin to find the Green's function and the self-energy of a many-body interacting system are reviewed in Chapter 4. Chapter 5 provides a detailed description of the methodological approach used to solve the *GW* equations in the frequency-

and reciprocal-space. The explicit expression for the relevant quantities used in the *GW* formalism (polarizabilities, dielectric matrices, self-energies) are derived, and the state-of-the-art algorithms that can be used to solve the *GW* equations are analyzed. Particular emphasis is given to the use of symmetry properties to reduce both the computing effort and the memory requirements. In Chapter 6, we then present a systematic analysis of the reliability of the plasmon-pole approximation for the calculation of quasiparticle corrections in a variety of simple *s* and *p* prototype compounds. The quasiparticle band structure of silicon dioxide in the hexagonal alpha quartz phase is discussed, and the effect of *GW* self-consistency is analyzed. In Chapter 7, we then proceed to discuss how the *GW* formalism can be implemented in the framework of the PAW formalism. The changes needed with respect to the standard norm-conserving approach are outlined, and selected cases are presented in order to validate our implementation against previously published works. The changes needed to implement the constrained self-consistent *GW* method proposed by Faleev and co-workers [16] are also briefly described. Finally, Chapter 8 discusses an efficient and reliable interpolation scheme for the quasiparticle energies based on the use of Maximally Localized Wannier Functions (MLWFs) introduced by Marzari and Vanderbilt [17]. The conclusions, with a brief summary of the main topics addressed in this work, end the thesis.

Chapter 3

Density Functional Theory

3.1 The many-body problem

The description of the chemical and physical properties of materials requires a quantum mechanical treatment of the many-particle system of electrons and nuclei with their basic electrostatic Coulomb interaction. Even if we assume that the motion of the fast moving electrons can be decoupled from that of the more massive nuclei (Born-Oppenheimer approximation [18]), the description of the physical properties of a non relativistic time-independent system of N electrons requires the solution of the Schrödinger equation for a system of $3N$ spatial variables and N spins¹,

$$\hat{H}\Psi_\nu = E_\nu\Psi_\nu, \quad \hat{H} = \sum_{i=1}^N \left(-\frac{1}{2}\Delta_i + v_{\text{ext}}(\mathbf{r}_i) \right) + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (3.1)$$

where $\mathbf{r}_i$ is the coordinate of the i -th electron, and $v_{\text{ext}}(\mathbf{r})$ describes the interaction with the nuclei. The last term in equation 3.1 accounts for the Coulomb electron-electron interaction and its presence dramatically complicates the solution of the problem. It correlates the motion of the particles leading to a quantum mechanical behavior that cannot be easily described in terms of mean-field theories.

The second-order differential equation in Eq. 3.1 can be solved exactly only for a few high-symmetry systems with a small number of electrons. For systems with a macroscopic number of particles, the mathematical solution of the problem is out-of-reach even with the computing facilities available nowadays, and we have to resort to approximate methods to obtain reasonably accurate results for the physical properties of interest.

In the past century several approximate approaches have been proposed and applied to the solution of the problem. Shortly after the basic work of Schrödinger and Heisenberg on the foundations of quantum-mechanics, Thomas [19] and Fermi [20] proposed (in 1927) a heuristic equation for the ground-state density of the many-electron system which uses the electronic density as the key ingredient, under the assumption that the

¹Unless otherwise specified, Hartree atomic units will be used throughout.

inhomogeneous interacting electron gas can be locally described as if it were homogeneous and non-interacting. Unfortunately, neither the kinetic energy, nor the electron-electron interaction beyond the classical electrostatic term, can be explicitly formulated in terms of the density and, indeed, the theory predicts ions to be unstable, and does not account for the electronic shell structure of atoms.

The failure of this first approach based on the electronic density as fundamental variable lead the physics community to concentrate on theories explicitly based on electronic wavefunctions. For example, in the so-called Hartree-Fock (HF) method [21, 22], the Schrödinger problem is reformulated in an approximate way in terms of a minimization problem for the total energy in the Hilbert space spanned by (automatically antisymmetric) determinants of single-particle wave functions. For several years, the Hartree-Fock approach and its different variants and improvements were the leading method in performing electronic structure calculations, especially in the chemistry community.

Approaches based on the density as the basic quantity were put on firm mathematical grounds only in the second half of the past century, thus receiving a formal justification. In 1964, Hohenberg and Kohn (HK) [23] showed that the ground-state electronic density – a simple scalar function of the position – theoretically contains all the information needed to describe the properties of the many-electron system by virtue of the existence of a one-to-one mapping between the ground-state density and the external potential up to a trivial additive constant. Density Functional Theory (DFT) formally reduces the many-electron problem to a single-electron problem, and uses the electron density distribution $n(\mathbf{r})$ as the basic variable instead of the more complicated many-body wavefunction.

The theorem, despite its formal simplicity, represents a milestone in the modern approach to the description of matter, since it reveals that all properties of the system can be formally expressed in terms of functionals of the density itself, thus opening the way towards the study of macroscopic systems. It is the aim of this chapter to introduce the basic concepts of density-functional theory and to discuss its successes and failures in the description of the physical properties of materials from first principles.

This chapter is organized as follows: The Hohenberg-Kohn theorem and its fundamental consequences are discussed in section 3.2.1. Section 3.2.2 introduces the Kohn-Sham method of mapping the complicated many-electron system to a set of single-particle equations. In section 3.3, some of the most famous and currently used approximations for the exchange-correlation potential are introduced. In the last part of the chapter, we then proceed to discuss how DFT performs in practice: we show that density functional theory gives a good description of structural and vibrational properties of materials, whereas it fails to give a reliable quantitative description of excitation energies and band gaps.

3.2 Density Functional Theory

3.2.1 The Hohenberg-Kohn theorem (1964)

Since the electron-electron interaction is known, it is clear from equation 3.1 that the external potential completely determines the Hamiltonian of the many-electron system.

Thus $v_{\text{ext}}(\mathbf{r})$ also completely determines the ground-state wavefunction as well as the ground-state density of the many-electron system. This implies that a unique mapping exists from the external potential to the ground-state density, $n_0(\mathbf{r})$, and that the ground-state density is a uniquely determined functional of $v_{\text{ext}}(\mathbf{r})$.

The first part of the Hohenberg-Kohn theorem states that the mapping from the external potential to the ground-state density is invertible up to a trivial additive constant in the potential. As a consequence, $v_{\text{ext}}(\mathbf{r})$ is uniquely determined by $n_0(\mathbf{r})$. Since every wavefunction $\Psi(\mathbf{r})$ depends implicitly on the external potential, it immediately follows that also $\Psi(\mathbf{r})$ is a functional of the ground-state density. The theorem leads to the conclusions that every quantum mechanical observable, *i.e.* every expectation value $\langle \Psi | \hat{O} | \Psi \rangle$, can be expressed in the form of a functional of the ground-state density.

The second statement of the Hohenberg-Kohn theorem proves that a unique functional $E[n]$ of the density exists which, under the constraint $\int n(\mathbf{r}) d\mathbf{r} = N$, reaches its minimum for $n_0(\mathbf{r})$, and gives the ground-state energy $E_0 = E[n_0]$. According to the re-formulation of DFT proposed by Levy [24, 25] in terms of a “constrained search”, the energy functional $E[n]$ can be defined as the minimum over all antisymmetric N -particle wavefunctions yielding the same density $n(\mathbf{r})$, namely that:

$$E[n] = \min_{\Psi \rightarrow n} \langle \Psi | \hat{H} | \Psi \rangle = \min_{\Psi \rightarrow n} \langle \Psi | \hat{T} + \hat{V}_{e-e} + V_{\text{ext}} | \Psi \rangle = F[n] + \int n(\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r}. \quad (3.2)$$

$F[n]$ is universal in the sense that it does not depend on the external potential and is the same for all systems having the same electron-electron interaction. From equation 3.2, minimizing over all N -electron densities arising from an antisymmetric wavefunction, one obtains

$$E_0 = \min_n E[n]. \quad (3.3)$$

This variational principle can therefore be used to determine the ground-state density $n_0(\mathbf{r})$ and the ground-state energy E_0 provided that the functional $F[n]$ is explicitly defined. The constraint of fixed number of particles can be explicitly included through the introduction of the Lagrange multiplier μ :

$$\delta \left\{ F[n] + \int v_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} - \mu \int n(\mathbf{r}) d\mathbf{r} \right\} = 0, \quad (3.4)$$

which is equivalent to the Euler equation:

$$\frac{\delta F}{\delta n(\mathbf{r})} + v_{\text{ext}}(\mathbf{r}) = \mu. \quad (3.5)$$

Equation 3.5 makes it clear that the external potential is uniquely determined by the ground-state density.

3.2.2 The Kohn-Sham equations (1965)

The theory discussed above has transformed the formidable problem of finding the minimum of $\langle \Psi | \hat{H} | \Psi \rangle$ with respect to the $3N$ -dimensional trial function Ψ into the seemingly trivial problem of locating the minimum of $E[n]$ with respect to the three-dimensional trial function $n(\mathbf{r})$. However, the explicit functional form of $F[n]$ is not known: the HK theorem assures that such a functional exists, but it does not provide any algorithmic procedure to construct $F[n]$. The re-formulation of Levy defines $F[n]$ in terms of an impractical minimization which is even more complicated than the original problem. The theory, up to this point, might seem rather abstract –useful for formal discussions, but of little practical use. In 1965, Kohn and Sham (KS) showed how a practical computational method can be constructed starting from the HK theorem [26]. In the formulation of DFT given by the two authors, the problem of finding the ground-state density of the interacting system is replaced by a much more affordable set of equations for an *auxiliary* ensemble of non-interacting electrons in an effective potential, $v_{\text{eff}}(\mathbf{r})$, whose ground-state density is supposed to be equal to that of the fully interacting system (the so-called *v*-representability ansatz).

For a system of non-interacting electrons the electron-electron interaction, V_{e-e} , in equation 3.2 vanishes, and $F[n]$ reduces to the kinetic energy functional $T_s[n]$ of the non-interacting electrons:

$$T_s[n] = \min_{\Psi \rightarrow n} \langle \Psi | \hat{T} | \Psi \rangle = \langle \Phi | \hat{T} | \Phi \rangle, \quad (3.6)$$

where Φ is a single Slater determinant or a linear combination of several of them. The total energy functional can thus be rewritten as

$$E_s[n] = T_s[n] + \int n(\mathbf{r}) v_{\text{eff}}(\mathbf{r}) \, d\mathbf{r}, \quad (3.7)$$

and the Hohenberg-Kohn variational principle leads to the Euler-Lagrange equation:

$$\frac{\delta \left\{ E_s[n] + \mu \left[N - \int n(\mathbf{r}) \, d\mathbf{r} \right] \right\}}{\delta n(\mathbf{r})} = \frac{\delta T_s[n]}{\delta n(\mathbf{r})} + v_{\text{eff}}(\mathbf{r}) - \mu = 0, \quad (3.8)$$

where the Lagrange parameter μ has been introduced to guarantee charge conservation. In the above equation, both the functional $T_s[n]$ for the kinetic energy of the non-interacting system and the non-interacting density can be explicitly expressed in terms of single-particle orbitals:

$$n(\mathbf{r}) = \sum_i^{\text{occ}} |\psi_i(\mathbf{r})|^2, \quad (3.9)$$

$$T_s[n] = -\frac{1}{2} \sum_i^{\text{occ}} \int \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) \, d\mathbf{r}. \quad (3.10)$$

The variation of $E[n]$ with respect to the orbitals leads to the famous Kohn-Sham equations:

$$\left[-\frac{1}{2}\Delta + v_{\text{eff}}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}), \quad (3.11)$$

where the Lagrange multipliers, ϵ_i , have been introduced in order to fulfill the orthonormality constraint $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. Kohn and Sham suggested the decomposition of the energy contribution associated to $v_{\text{eff}}(\mathbf{r})$ according to:

$$\int n(\mathbf{r})v_{\text{eff}}(\mathbf{r}) \, d\mathbf{r} = \int n(\mathbf{r})v_{\text{ext}}(\mathbf{r}) \, d\mathbf{r} + \frac{1}{2} \iint \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \, d\mathbf{r}_1 \, d\mathbf{r}_2 + E_{\text{xc}}[n]. \quad (3.12)$$

The first term on the right hand side of equation 3.12 represents the contribution to the total energy given by the external potential while the second integral accounts for the classical electrostatic repulsions among charged particles. The last term, also called the Exchange-Correlation (XC) energy, is formally defined as

$$E_{\text{xc}}[n] = F[n] - T_s[n] - \frac{1}{2} \iint \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \, d\mathbf{r}_1 \, d\mathbf{r}_2. \quad (3.13)$$

The usefulness of the decomposition 3.12 is that $T_s[n]$ can be exactly described in terms of the orbitals of the auxiliary system, leaving only $E_{\text{xc}}[n]$ to be approximate. From a physical point of view the approximation is reasonable and expected to be accurate enough for practical applications since the kinetic energy of the auxiliary system and the long-range Hartree term constitute a significant part of the total energy. The remaining part, $E_{\text{xc}}[n]$, is usually much smaller, and can be reasonably approximate as a local (or nearly local) functional of $n(\mathbf{r})$. Using the decomposition given in Eq. 3.12, equation 3.11 can be rewritten as:

$$\left[-\frac{1}{2}\Delta + v_{\text{ext}}(\mathbf{r}) + v_H[n](\mathbf{r}) + v_{\text{xc}}[n](\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}) \quad (3.14)$$

where the Hartree term, $v_H[n](\mathbf{r})$, describes the classical Coulomb repulsion:

$$v_H[n](\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, d\mathbf{r}', \quad (3.15)$$

and $v_{\text{xc}}[n](\mathbf{r})$, the so-called exchange-correlation potential, is formally defined by

$$v_{\text{xc}}[n](\mathbf{r}) = \frac{\delta E_{\text{xc}}[n]}{\delta n(\mathbf{r})}. \quad (3.16)$$

The obtained equations have the form of independent-particle equations, but – as the effective potential $v_{\text{eff}}(\mathbf{r}) = v_H[n](\mathbf{r}) + v_{\text{xc}}[n](\mathbf{r}) + v_{\text{ext}}(\mathbf{r})$ depends on the density – the solution must be found self-consistently. This can be achieved by an iterative process: starting from a trial density, the effective potential is determined for which the set of equations in 3.11 are solved to determine a new $n(\mathbf{r})$. The process is iterated until the

new density or the total energy equals the previous one within the required accuracy. It is important to stress, however, that the self-consistent KS orbitals have no direct link with the full many-body wavefunction; this is even less true for the KS eigenvalues which, except for some exceptions, cannot be interpreted as the energies of the electrons in the interacting system. The KS eigenvectors and eigenvalues are just mathematical tools used to solve the problem for a fictitious non interacting system whose ground-state density is assumed to be equal to the original many-body system. Strictly speaking, only the density given by equation 3.9 has a direct physical meaning.

3.3 Approximations for the Exchange-Correlation functional

The density-functional theory presented above is exact in principle, however, the functionals $E_{xc}[n]$ and $v_{xc}[n]$ in which the complexity of the many-particle system is somehow hidden, are not exactly known and must be approximated. The widespread use of DFT for first principles calculations is mainly due to the availability of approximations which are simple and accurate enough for practical applications.

A rather simple expression, the so-called Local Density Approximation (LDA), was already proposed in the original work of Kohn and Sham [26]. In the LDA, $E_{xc}[n](\mathbf{r})$ is simply given by the exchange-correlation energy of the homogeneous electron gas with density $n(\mathbf{r})$:

$$E_{xc}^{\text{LDA}}[n] = \int n(\mathbf{r}) \epsilon_{xc}^{\text{HEG}}(n(\mathbf{r})) d\mathbf{r}, \quad (3.17)$$

where $\epsilon_{xc}^{\text{HEG}}([n], \mathbf{r})$ is the energy per particle of an electron gas with uniform density, thus a function (not a functional) of the density. The LDA is exact only for densities that vary slowly over space [26] and was therefore expected to perform well mainly for solids that are similar to the homogeneous electron gas. Despite this limitation, the LDA proved to be remarkably accurate and hard to improve upon. It is indeed able to give a reasonable description of the structural and vibrational properties of a wide variety of crystalline systems, and it still remains the most popular method for electronic-structure calculations in the solid state community. By contrast, the LDA is rarely used in the chemistry community which is mainly concerned with strongly inhomogeneous systems such as atoms or molecules, where energetics are much more important than in solid state physics.

Several approaches have been proposed to go beyond the efficient but crude LDA approximation. For systems with more inhomogeneous densities, an attempt to include the spatial variations of the density was done in the early 1990's by means of the Generalized Gradient Approximation (GGA) [27]. Following the GGA approach, the integrand in equation 3.17 is generalized by taking into account the gradient of the density:

$$E_{xc}^{\text{GGA}}[n] = \int f(n(\mathbf{r}), \nabla n(\mathbf{r})) d\mathbf{r}. \quad (3.18)$$

While the input $\epsilon_{xc}^{\text{HEG}}$ in equation 3.17 is unique, the function $f(n, \nabla n)$ in equation 3.18 is

not uniquely determined and many different forms have been proposed in order to incorporate a number of known properties of the exact XC functional. Functionals belonging to the GGA family usually result in a better description of the structural properties of solids and improve the cohesive energies and lattice constants of $3d$ transition metals. The GGA is also quite successful in improving binding energies, total energies, bond-lengths and bond-angles of atoms as compared to the LDA results, although its accuracy is still far from the chemical accuracy required for chemistry applications.²

3.4 Successes and failures of DFT: the band gap problem

By 1998, DFT had become so successful and popular for large-scale applications in quantum chemistry that half of the Nobel prize in chemistry went to Walter Kohn “for his development of density functional theory” [28]. Nowadays, DFT has become a leading method for electronic structure calculations in both the solid-state and the quantum-chemistry community, for several reasons. First of all, the computational cost is relatively low when compared to traditional methods, such as Hartree-Fock and its descendants, which are based on complicated many-electron wavefunctions. Furthermore, DFT gives direct and easy access to forces and stresses, thus allowing structural relaxations in order to predict the minimum energy configuration of the system. Density functional perturbation theory and the $2n + 1$ theorem [29, 30] can be used to calculate derivatives of the energy up to any order, which opens the way towards the study of vibrational and elastic properties as well as the response of the system beyond the harmonic regime.

Presently, a wide variety of physical properties are within the realm of applicability of DFT: lattice-dynamical behavior, infra-red-, Raman-, and neutron-diffraction spectra, specific heat, thermal expansion, heat conduction, electron-phonon interaction, resistivity of metals and conventional superconductivity. Lattice parameters and atomic positions are usually predicted to within 1-5%, and the calculated phonon dispersion seldom departs more than a few percents from the experimental results.

On the other hand, the efficiency of DFT in describing and predicting excitation energies and band gaps is not as good as that for structural and vibrational properties. In standard *ab initio* calculations, the energy spectrum of the KS Hamiltonian is routinely employed to describe and predict band structures, optical properties and effects due to the presence of the Fermi surface in metals. In this case the agreement with experiments is quite unsatisfactory; standard functionals (LDA and GGA), as shown in Figure 3.1, systematically underestimate band gaps, giving values that are about 30-40% smaller than experimental data.

The inability of standard KS theory to give band gaps close to experiment is often referred to as the “band-gap problem”. From a theoretical point of view this is not surprising since KS eigenvalues are not supposed to give the correct band energies. The band structure of a crystal is rigorously defined as the energies needed to add or subtract electrons from the many-body system which, in turn, are related to the difference between

²Typically of the order of 0.05 eV.

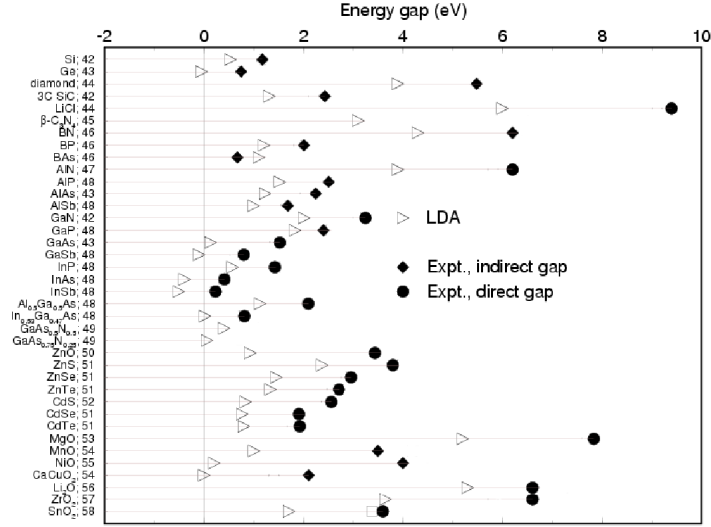


Figure 3.1: Comparison of LDA band gaps with respective experimental values for a variety of materials. From [31].

total energies of many-body states differing by one electron. Within the Kohn-Sham formalism, the eigenvalues can be related to the derivative of the total energy with respect to the occupation of a state but, from the perspective of optical properties, they are just Lagrangian multipliers included in the KS equations to satisfy the orthonormality constraint. The only exception is the energy of the highest occupied state which, in exact DFT, equals the ionization potential (the chemical potential for metals) [32].

As explained by Perdew and Levy [33] and Sham and Schlüter [34, 35], the true fundamental gap in semiconductors and insulators, E_g , differs from the exact KS fundamental gap, E_g^{KS} , by a term, Δ_{xc} , originating from the discontinuity in the exchange-correlation potential upon the addition of an electron:

$$E_g = E_g^{\text{KS}} + \Delta_{\text{xc}}. \quad (3.19)$$

The magnitude of the Δ_{xc} discontinuity has not been established yet, and there is active research in the field to clarify its importance. Nevertheless it is evident from figure 3.1 that fundamental gaps of materials are not correctly reproduced by standard functionals, and that the error can be of the same order of magnitude as E_g^{KS} .

In principle, since the ground-state density determines the entire Hamiltonian and the Kohn-Sham density is supposed to be exact by construction, there should be some way to use the Kohn-Sham potential and eigenvalues to determine all excitations exactly. In practice, however, the simplest and naive interpretation of the KS eigenvalues as addition or removal energies leads to a poor agreement with experiments, and it is still not so clear how to construct simple density functionals able to describe the many-body excitation spectrum.

The capability of performing reliable first-principles estimates of band gaps in semiconductors is of paramount importance for the design and synthesis of new materials and optical devices. The search for materials with the required properties would be greatly facilitated if we were able to predict band gaps and excitation energies with a sufficient degree of accuracy and without adjustable external parameters. Unfortunately, in striking contrast to the present scientific and technological needs, the actual progress in developing techniques for describing interaction effects within DFT has been relatively slow, and we are still far from the required accuracy.

The past decades have seen the rise of new appealing and promising exchange-correlation functionals based on KS orbitals, but several open questions on the nature and importance of the Δ_{xc} discontinuity remain. The time-dependent extension to static density functional theory (TDDFT [1]) has become a promising and appealing approach to the study of optical properties of finite systems [2]. Unfortunately, the presently available approximations for the nonlocal and frequency-dependent kernel,³ once applied to extended systems, do not provide results of the same quality [2].

An alternative, more traditional, approach to the study of exchange-correlation effects in many-body systems is provided by MBPT [3, 4, 36, 37], which defines a rigorous approach to the description of excited-state properties, based on the Green's function formalism. MBPT also provides a proper framework for accurately computing excited state properties from KS results. The mathematical aspects of many-body theory and the link between DFT and MBPT are the subjects of the next chapter.

³ The exchange-correlation kernel, f_{xc} , is defined as the functional derivative of the time-dependent KS potential with respect to the time-dependent density.

Chapter 4

An Overview of Many-Body Theory

4.1 Introduction

The aim of this chapter is to introduce the Green's function formalism [3, 4, 36, 37]. This is the basic tool of many-body theory which has been successfully applied, during the last three decades, to the description of one-particle and two-particle excitations. The chapter is organized as follows: The concept of time-ordered Green's function is introduced and discussed in section 4.2. Section 4.3 introduces the so-called Lehmann representation, an alternative formulation of the Green's function concept which allows one to extract information about the excitation energies of the many-body system. The equation of motion for the Green's function is derived in section 4.5, while section 4.6 describes the approach proposed by Hedin in 1965 to find the Green's function and the self-energy of a many-body system.

Most of the expressions presented hereafter are obtained within the second-quantization formulation of quantum mechanics, in the Heisenberg representation [4, 36]. Unless explicitly noted, it is also understood that all the operators are given in this representation. $\hat{\psi}^\dagger(t\mathbf{r})$ and $\hat{\psi}(t\mathbf{r})$ indicate the field operators describing the creation and the annihilation of an electron at the position $\mathbf{r}$ at time t . To simplify the notation, space-time coordinates $(\mathbf{r}_1, t_1)$ are denoted with a natural number (1). Moreover, the following shorthand notations are employed:

$$\delta(12) = \delta(\mathbf{r}_1 - \mathbf{r}_2) \delta(t_1 - t_2) \quad (4.1)$$

$$\int d1 = \int d\mathbf{r}_1 \int_{-\infty}^{+\infty} dt_1 \quad (4.2)$$

$$v(12) = v(\mathbf{r}_1, \mathbf{r}_2) \delta(t_1 - t_2) \quad (4.3)$$

$$1^+ = (\mathbf{r}_1, t_1 + \eta)_{\eta \rightarrow 0^+} \quad (4.4)$$

where $v(\mathbf{r}_1, \mathbf{r}_2)$ represents the bare Coulomb interaction, and η is a positive infinitesimal. In all equations, the spin quantum number is not explicitly considered as it can be easily added to the formulas by considering it as part of the spatial coordinate $\mathbf{r}$. Unless otherwise specified, Hartree atomic units will be used throughout.

4.2 The Green's Function

The time-ordered Green's function $G(12)$, also called the propagator, defines the probability amplitude for the propagation of an added or removed electron in a many-body system. Since the probability amplitude is simply given by the overlap between the final and the initial state, $G(12)$ can be expressed as

$$G(12) = -i\langle\Theta_0^N|\hat{T}[\hat{\psi}(1)\hat{\psi}^\dagger(2)]|\Theta_0^N\rangle, \quad (4.5)$$

where $|\Theta_0^N\rangle$ represents the ground state of the N -electron many-body Hamiltonian. The time-ordering operator $\hat{T}$ rearranges a series of field operators for fermionic systems in order of ascending time arguments from right to left with a factor (-1) for each pair permutation, for example:

$$\hat{T}[\hat{\psi}(1)\hat{\psi}^\dagger(2)] := \begin{cases} \hat{\psi}(1)\hat{\psi}^\dagger(2) & \text{if } t_1 > t_2, \\ -\hat{\psi}^\dagger(2)\hat{\psi}(1) & \text{if } t_2 > t_1. \end{cases} \quad (4.6)$$

Depending on the time-ordering, equation 4.5 describes either the propagation of an electron added to the system ($t_1 > t_2$) or the propagation of a hole ($t_1 < t_2$), *i.e.* the time evolution of a system with $N \pm 1$ particles. The Green's function contains only part of the full information carried by the many-body wave function, but it includes the relevant information for the study of charged excitations, which is the problem we are dealing with. As shown in [4], the ground-state expectation value of any single particle operator is determined by the knowledge of $G(12)$. For example the ground-state electron density can be expressed as

$$n(\mathbf{r}) = \langle\Theta_0^N|\hat{\psi}^\dagger(\mathbf{r})\hat{\psi}(\mathbf{r})|\Theta_0^N\rangle = -iG(11^+) \quad (4.7)$$

where 1^+ serves to enforce the correct ordering of the field operators. Also the ground-state energy of the system is accessible, once the one-particle Green's function is available, by means of the Galitskii-Migdal formula [38]:

$$E = \frac{1}{2} \int d\mathbf{r}_2 \lim_{t_2 \rightarrow t_1^+} \left[\frac{\partial}{\partial t} - ig(\mathbf{r}) \right] G(12)_{\mathbf{r}_2 \rightarrow \mathbf{r}_1}, \quad (4.8)$$

where $g(\mathbf{r})$ is the density-independent part of the effective Hamiltonian [38].

4.3 The Lehmann representation

Useful physical information about the charged excitation energies of the many-body system can be obtained by expressing $G(12)$ in the so-called Lehmann representation [39]. To this purpose it is useful to introduce the following notation to denote the charged excitation energies of the N -electron system [40]:

$$\epsilon_i = \begin{cases} E_i^{(N+1)} - E_0^N & \epsilon_i > \mu \\ E_0^N - E_i^{(N-1)} & \epsilon_i < \mu \end{cases} \quad (4.9)$$

where E_0^N is the ground state energy of the electron system with N electrons, and i is the set of quantum numbers labeling the excited states with $N \pm 1$ electrons. Finally μ is the chemical potential of the system. Other important quantities that will be used in the following are the so-called Lehmann amplitudes defined, in the Schrödinger representation, by

$$\Psi_i(\mathbf{r}) \equiv \begin{cases} \langle \Theta_0^N | \hat{\psi}(\mathbf{r}) | \Theta_i^{N+1} \rangle & \epsilon_i > \mu \\ \langle \Theta_i^{N-1} | \hat{\psi}(\mathbf{r}) | \Theta_0^N \rangle & \epsilon_i < \mu \end{cases} \quad (4.10)$$

Inserting the closure relation $\sum_i |\Theta_i^{N\pm 1}\rangle \langle \Theta_i^{N\pm 1}| = \hat{1}$ between the two field operators in equation 4.5, transforming the result in the frequency domain, we arrive at the final result [40]:

$$G(\mathbf{r}_1, \mathbf{r}_2; \omega) = \sum_i \frac{\Psi_i(\mathbf{r}_1) \Psi_i^\dagger(\mathbf{r}_2)}{\omega - \epsilon_i + i\eta \text{sign}(\epsilon_i - \mu)} \quad \eta \rightarrow 0^+, \quad (4.11)$$

To obtain the expression in 4.11, we used the property that the Green's function of a stationary system is invariant under an arbitrary translation in time, and therefore depends on t_1 and t_2 only through the difference $t_1 - t_2$.

Equation 4.11 with equation 4.9 makes it clear that, in the frequency domain, the time-ordered Green's function contains the complete excitation spectrum corresponding to excitations of an $(N - 1)$ -particle and an $(N + 1)$ -particle system. Hence locating the poles of the Green's function in the complex plane provides the information needed to interpret those processes measured in experiments in which a single electron is inserted to or removed from the system. Figure 4.1 gives a schematical representation of the location of the poles of the time-ordered Green's function.

In principle there is only one way to measure the single-particle properties, which is to insert or remove single electrons into or out of the system. This can be achieved by the so-called tunnel junction device or by subjecting the sample to a beam of electrons. However, even optical experiments measure the single particle density of states approximately (see [2], for a discussion of the different spectroscopic experiments).

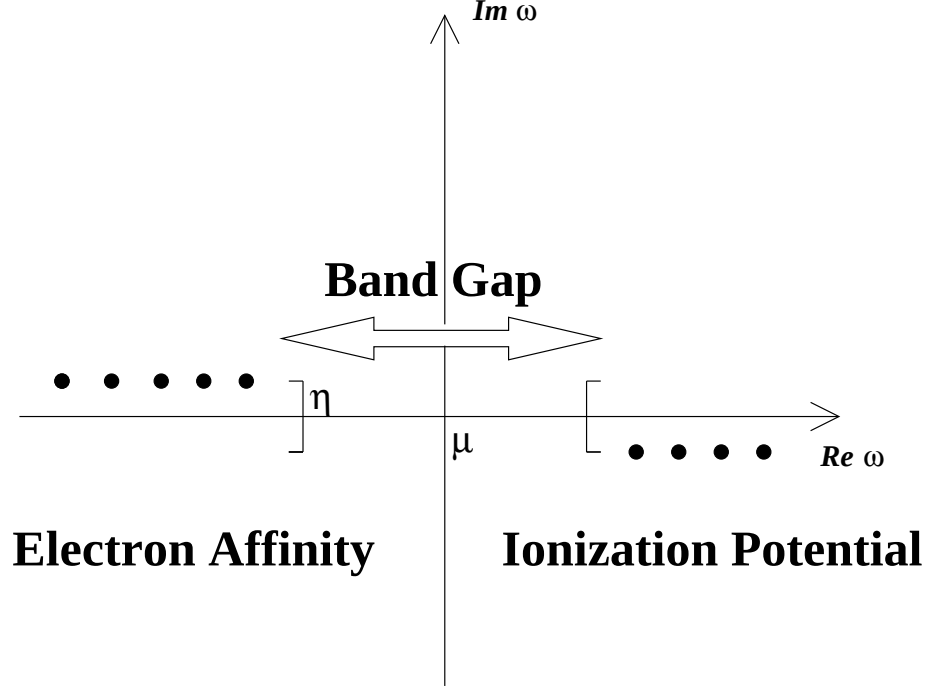


Figure 4.1: Schematic representation of the analytic structure of the time-ordered Green's function, the chemical potential defines the zero of the energy. The branch cuts in the second (fourth) quadrant of the complex plane correspond to electron removal (addition) energies. In semiconductors, no poles are present within the fundamental gap *i.e.* within the minimum gap in the excitation spectrum of the $N \pm 1$ -electron states with respect to the ground-state energy of the N -particle state.

4.4 The spectral function

An important quantity closely connected to spectroscopic measurements is the spectral function $A(\mathbf{r}_1, \mathbf{r}_2; \omega)$ given by:

$$A(\mathbf{r}_1, \mathbf{r}_2; \omega) \equiv \left| \frac{1}{\pi} \text{Im} G(\mathbf{r}_1, \mathbf{r}_2; \omega) \right| = \sum_i \Psi_i(\mathbf{r}_1) \Psi_i^\dagger(\mathbf{r}_2) \delta(\omega - \epsilon_i), \quad (4.12)$$

which defines the density of the excited states contributing to the electron or hole propagation. The spectral function can be thought as either the quantum state resolution of a particle with given energy ω or as the energy resolution for a particle in a given quantum number i . It gives an indication of how well the excitation created by adding a particle in state i can be described by a free non-interacting particle. For the idealized case of non-interacting electrons described by the Hamiltonian $\hat{h}_0$, the spectral function

$A(\mathbf{r}_1, \mathbf{r}_2; \omega)$ reduces to the simple expression:

$$A_0(\mathbf{r}_1, \mathbf{r}_2; \omega) = \sum_i \psi_i(\mathbf{r}_1) \psi_i^\dagger(\mathbf{r}_2) \delta(\omega - \epsilon_i), \quad (4.13)$$

where $\psi(\mathbf{r})$ and ϵ_i are the eigenvectors and eigenvalues of $\hat{h}_0$. Since $\hat{h}_0$ is Hermitian, its eigenfunctions can always be chosen so that they form an orthonormal basis set in the Hilbert space, *i.e.* $\langle \psi_i | \psi_j \rangle = \delta_{ij}$. As a consequence, the spectral function operator in the single-particle representation is given by a diagonal matrix whose elements are delta functions centered at the single-particle energy, ϵ_i :

$$\langle \psi_i | \hat{A}_0(\omega) | \psi_j \rangle = \delta_{ij} \delta(\omega - \epsilon_i). \quad (4.14)$$

When many-body interactions are switched on, the Lehmann amplitudes are not orthogonal anymore, although they still satisfy the closure relation. In the interacting case, the diagonal matrix elements of $\hat{A}$ are given by an envelope of delta functions according to:

$$\langle \Psi_i | \hat{A}(\omega) | \Psi_i \rangle = \sum_k \langle \Psi_i | \Psi_k \rangle \langle \Psi_k | \Psi_i \rangle \delta(\omega - \epsilon_k). \quad (4.15)$$

Many-body interactions thus lead to a redistribution of the spectral weight and the appearance of satellites which can dramatically change the profile of the spectral function with respect to the ideal delta profile on the non-interacting system. This is schematically depicted in figure 4.2.

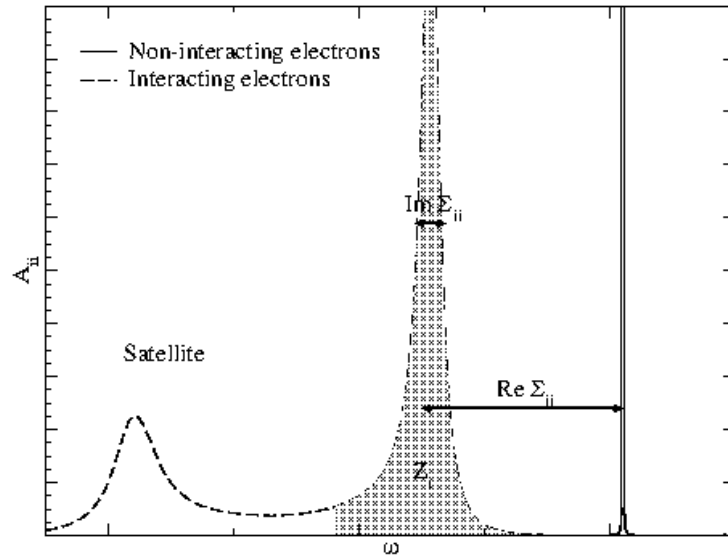


Figure 4.2: Matrix elements of the spectral function of interacting (dashed line) and non-interacting (solid line) electron systems. Taken from [41].

The Quasiparticle (QP) picture relies on the assumption that equation 4.15 gives rise to something that can be seen as a function with a broadened peak, so that it is still possible to isolate an excitation that can be described as a renormalized single-electron. There is, of course, no guarantee that such single-particle will exist, the spectral weight may indeed turn out to be a completely smooth function with no quasiparticle states at all. The success of the single-particle theory of solids suggests, however, that well-defined quasiparticles *do* exist and that the non-interacting approximation is not too far from the truth.

4.5 Equation of motion of G and the concept of self-energy

In the previous sections, the concept of Green's functions was introduced and some extremely useful properties were presented. However, at this stage, the question of how it can be calculated remains. There are several ways to tackle the problem, one of which is the equation of motion technique. In this approach, a series of coupled differential equations are generated by differentiating the Green's function a number of times. With the help of the second-quantization formalism, the many-particle Hamiltonian:

$$\hat{H} = \sum_{i=1}^N \left(-\frac{1}{2} \Delta_i + v_{\text{ext}}(\mathbf{r}_i) \right) + \frac{1}{2} \sum_{i \neq j} v(\mathbf{r}_i, \mathbf{r}_j) = \sum_{i=1}^N \hat{h}_0(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} v(\mathbf{r}_i, \mathbf{r}_j), \quad (4.16)$$

can be rewritten in terms of a field operator involving the annihilation and creation operators:

$$\hat{H} = \int \hat{\psi}^\dagger(\mathbf{r}) \hat{h}_0(\mathbf{r}) \hat{\psi}(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \iint \hat{\psi}^\dagger(\mathbf{r}_1, t) \hat{\psi}^\dagger(\mathbf{r}_2, t) v(\mathbf{r}_1, \mathbf{r}_2) \hat{\psi}(\mathbf{r}_2) \hat{\psi}(\mathbf{r}_1) d\mathbf{r}_1 d\mathbf{r}_2, \quad (4.17)$$

where we have explicitly separated the one-electron part of the Hamiltonian

$$\hat{h}_0(\mathbf{r}) \equiv -\frac{1}{2} \Delta + v_{\text{ext}}(\mathbf{r}), \quad (4.18)$$

from the Coulomb interaction. The expression 4.17 is just a mathematical formulation of the initial Hamiltonian, introduced in order to automatically satisfy the requirement of antisymmetry of the total fermionic wave function. From the equation of motion for the field operator in the Heisenberg representation [4]:

$$i \frac{\partial \hat{\psi}}{\partial t} = [\hat{\psi}, \hat{H}] \quad (4.19)$$

it is possible to derive the following equation, defining the time-evolution of the Green's function:

$$i \frac{\partial G(12)}{\partial t_1} = \delta(12) + \hat{h}_0(\mathbf{r}) G(12) - i \int v(1^+ 3) G(1323^+) d3, \quad (4.20)$$

which involves the two-particle Green's function:

$$G(1234) = -\langle \Theta_0^N | \hat{T} [\hat{\psi}(1)\hat{\psi}(2)\hat{\psi}^\dagger(4)\hat{\psi}^\dagger(3)] | \Theta_0^N \rangle. \quad (4.21)$$

The presence of the two-particle Green's functions in equation 4.20 complicates the solution of the problem, since the knowledge of the time-evolution of a four-point quantity is needed. One might think that it would be possible to obtain a closed expression for the derivative of $G(12)$ by analyzing the time evolution of $G(1234)$. Unfortunately, the evolution of the two-particle Green function depends on a three-particle Green's function, the equation for the three-particle Green function requires the knowledge of a four-particle Green's function and so on. Hence the solution of the equation of motion for $G(12)$ would require the solution of an infinite series of coupled differential equations involving functions with a progressively increasing number of electrons. The presence of an infinite series is somehow expected since the complete solution for the propagation of an added particle requires the inclusion of progressively more complicated interactions among the electrons under study. The solution of such equations presents the same complexity as the original Schrödinger problem, and it might seem that the formulation of the problem in terms of Green's functions has not lead to any kind of advantage. What is required is a way of either summing or terminating the series such that the most important physics is correctly described. This can be formally achieved by employing basic notions from the theory of integro-differential equations.

Let us now assume for the moment that we have solved the infinite series of equations for the equation of motion of $G(12)$. Under this assumption, we can introduce a function, $\Sigma(12)$, called the self-energy, which formally satisfies the equation:

$$\left[i \frac{\partial}{\partial t_1} - \hat{h}_0(\mathbf{r}_1) \right] G(12) - \int \Sigma(13) G(32) d3 = \delta(12). \quad (4.22)$$

The case $\Sigma = 0$ leads to the following equation for the Green's function $G_0(12)$ of the non-interacting system described by the single-particle Hamiltonian $\hat{h}_0$:

$$\left[i \frac{\partial}{\partial t_1} - \hat{h}_0(1) \right] G_0(12) = \delta(12). \quad (4.23)$$

Since $\hat{h}_0(\mathbf{r})$ does not have any time dependence, it follows that:

$$G_0^{-1}(\omega) = \left[\omega - \hat{h}_0(\mathbf{r}_1) \right]. \quad (4.24)$$

Introducing a function to help solve the integro-differential equation is a well-known method used in mathematical theory which, in this particular case, turns out to have a clear physical meaning.¹ To see why $\Sigma(12)$ is termed the self-energy, we have to Fourier transform equation 4.22 in the energy domain:

$$\left[\omega - \hat{h}_0(\mathbf{r}_1) \right] G(\mathbf{r}_1, \mathbf{r}_2; \omega) - \int \Sigma(\mathbf{r}_1, \mathbf{r}_3; \omega) G(\mathbf{r}_3, \mathbf{r}_2, \omega) d\mathbf{r}_3 = \delta(\mathbf{r}_1 - \mathbf{r}_2). \quad (4.25)$$

¹ The self-energy can be shown to exist by Wick's theorem. A much more consistent and rigorous introduction to the physical concept of self-energy can be found in [4].

Multiplying with G_0 both sides of equation 4.25 and integrating, one obtains:

$$G(12) = G_0(12) + \int G_0(13) \Sigma(34) G(42) d34. \quad (4.26)$$

From the above equation, it follows that:

$$G^{-1}(12) = G_0^{-1}(12) - \Sigma(12), \quad (4.27)$$

from which it can be seen that the poles of the many-body Green's function $G(12)$ (zeros of the inverse $G^{-1}(12)$) are moved in energy by $\Sigma(12)$ from those of the non-interacting Green's function $G_0(12)$. Thus the self-energy can be considered to be the contribution of the interaction effects to the excitation energies of the system.

By virtue of equation 4.27, the calculation of $G(12)$ has now been reduced to the calculation of the kernel of the integral equation 4.22. Furthermore, the knowledge of the self-energy operator allows one to describe the quantum-mechanical state of a renormalized electron in the many-body system. The quasiparticle amplitudes and energies can indeed be found by solving the following nonlinear system of equations [40]:

$$\hat{h}_0(\mathbf{r}_1)\Psi_j(\mathbf{r}_1) + \int \Sigma(\mathbf{r}_1, \mathbf{r}_2; \epsilon_j)\Psi_j(\mathbf{r}_2) d\mathbf{r}_2 = \epsilon_j\Psi_j(\mathbf{r}_1). \quad (4.28)$$

4.6 Hedin's Equations

In the previous section, we have seen that the calculation of the fully interacting Green's function can be reduced to the problem of finding the self-energy of the system. Historically, the first attempt to tackle the problem of determining $\Sigma(12)$ was done in the 50's [4]. The original approach used the exact non-interacting Green's functions $G_0(12)$ and the instantaneous bare Coulomb interaction $v(12) = \delta(t_1 - t_2) v(\mathbf{r}_1 - \mathbf{r}_2)$ as its basic quantities. The final equations are exact but hard to use if a perturbative expansion is employed: developments in terms of $v(12)$ are known to fail, for instance, the second order term in $v(12)$ for the correlation energy of an homogeneous electron gas is infinite. Moreover since the Coulomb interaction is strong and long-ranged, high order terms in the expansion are expected to decay extremely slowly. For all these reasons, it is thus reasonable to expect that a perturbation expansion based on the bare Coulomb interaction can give meaningful physical results only if high order contributions are included.

On the other hand, each particle in a many-body system does not react to the applied external potential, but to the total potential. This includes the response of the other particles, which significantly reduce the effective potential. Thus, a more realistic answer valid even for low orders should be obtained if we consider the change of the Green's function with respect to the total potential, rather than the external potential. This is the basic idea behind the formulation of the many-body problem proposed by Hedin in 1965 [5]. He showed how to derive a set of coupled integro-differential equations whose self-consistent solution, in principle, solves the many-electron problem exactly. For his derivation, Hedin employed the functional-derivative method using the exact "dressed"

propagator $G(12)$ and the dynamical screened interaction $W(12)$ (to be defined later on) as fundamental building blocks.

The derivation starts by introducing an external time-dependent perturbation $U(1)$ which is taken to the zero limit at the end of the derivation.² The total effective potential:

$$U_{\text{eff}}(1) \equiv U(1) + \int v(1, 2)n(2) d2 = U(1) - i \int v(1, 2)G(22^+) d2, \quad (4.29)$$

is given by the superposition of the external perturbation $U(1)$ and the internal classical Hartree potential.

To facilitate the subsequent discussion it is useful at this point to introduce some important quantities that will enter the final equations. In analogy with the classical polarization, we can introduce the irreducible polarizability:

$$\tilde{\chi}(12) \equiv \frac{\delta n(1)}{\delta U_{\text{eff}}(2)} = -i \frac{\delta G(11^+)}{\delta U_{\text{eff}}(2)}, \quad (4.30)$$

which describes the linear response of the density to changes in the total effective potential, and the reducible polarizability:

$$\chi(12) \equiv \frac{\delta n(1)}{\delta U(2)} = -i \frac{\delta G(11^+)}{\delta U(2)}, \quad (4.31)$$

which reflects the linear change of the density with respect to the external perturbation. The time-ordered inverse dielectric matrix is defined as

$$\epsilon^{-1}(12) \equiv \frac{\delta U_{\text{eff}}(1)}{\delta U(2)} = \delta(12) + \int v(13)\chi(32) d3, \quad (4.32)$$

and its inverse can be expressed in terms of the irreducible polarizability according to:

$$\epsilon(12) \equiv \frac{\delta U(1)}{\delta U_{\text{eff}}(2)} = \delta(12) - \int v(13)\tilde{\chi}(32) d3. \quad (4.33)$$

The dynamically screened interaction generated at position (1) by an electron at (2) is expressed in terms of the bare Coulomb term and the inverse dielectric matrix as

$$W(12) \equiv \int \epsilon^{-1}(13) v(32) d3. \quad (4.34)$$

In matrix notation, Eq. 4.33 and Eq. 4.34 are equivalent to $\epsilon = \mathbf{1} - \mathbf{v}\tilde{\chi}$ and $\mathbf{W} = \epsilon^{-1}\mathbf{v}$, respectively. Therefore $\mathbf{v} = \epsilon\mathbf{W} = (\mathbf{1} - \mathbf{v}\tilde{\chi})\mathbf{W}$ from which, after some manipulation, it follows that:

$$W(12) = v(12) + \int v(13)\tilde{\chi}(34) W(42) d34. \quad (4.35)$$

²The perturbation $U(1)$ is a mathematical tool introduced to employ the functional derivative method.

Equation 4.35 and the Dyson equation in 4.26 constitute the first two of Hedin's equations. At this point most of the ingredients needed for the derivation have been defined, and we can proceed with the derivation of the full set of Hedin's equations.

For the functional derivative of the Green's function with respect to the perturbation $U(1)$ one finds [5]:

$$\frac{\delta G(12)}{\delta U(3)} = G(12)G(33^+) - G(1323^+). \quad (4.36)$$

This important relation can be used to eliminate the two-particle Green's function in the equation of motion in 4.20, leading to:

$$\begin{aligned} i \frac{\partial G(12)}{\partial t_1} &= \delta(12) + \hat{h}_0(\mathbf{r})G(12) - i \underbrace{\left(\int v(13)G(33^+) d3 \right)}_{v_H(1)} G(12) + i \int v(1^+3) \frac{\delta G(12)}{\delta U(3)} d3 = \\ &= \delta(12) + \left[\hat{h}_0(\mathbf{r}) + v_H(1) \right] G(12) + \int \Sigma(13)G(32) d3, \end{aligned} \quad (4.37)$$

where the self-energy operator is now given by:

$$\Sigma(12) = i \iint v(1^+3) \frac{\delta G(14)}{\delta U(3)} G^{-1}(42) d(34) = \quad (4.38)$$

$$= -i \iint v(1^+3)G(14) \frac{\delta G^{-1}(42)}{\delta U(3)} d(34) = \quad (4.39)$$

$$= i \iint W(1^+3)G(14)\Gamma(42;3) d(34). \quad (4.40)$$

For the second identity, we have used partial integration and for the third equality the chain rule for functional derivatives. The so-called vertex function, $\Gamma(12;3)$, is formally defined by:

$$\Gamma(12;3) = -\frac{\delta G^{-1}(12)}{\delta U_{\text{eff}}(3)}. \quad (4.41)$$

To obtain a tractable expression for the vertex, we use the Dyson equation (Eq. 4.27), from which:

$$\Gamma(12;3) = \delta(12)\delta(13) + \frac{\delta \Sigma(12)}{\delta U_{\text{eff}}(3)}. \quad (4.42)$$

Employing the easily verifiable identity:

$$\begin{aligned} \frac{\delta G(12)}{\delta U_{\text{eff}}(3)} &= \frac{\delta}{\delta U_{\text{eff}}(3)} \int G(14) G^{-1}(45) G(52) d(45) = \\ &= 2 \frac{\delta G(12)}{\delta U_{\text{eff}}(3)} + \int G(14) \frac{\delta G^{-1}(45)}{\delta U_{\text{eff}}(3)} G(52) d(45), \end{aligned} \quad (4.43)$$

one finally obtains:

$$\Gamma(12; 3) = \delta(12)\delta(13) - \int \frac{\delta\Sigma(12)}{\delta G(45)} G(56) \Gamma(67; 3) G(74) d(4567). \quad (4.44)$$

Equation 4.44 defines an integral equation for the vertex function involving only $\Sigma(12)$ and $G(12)$. At this stage, the irreducible polarizability (Eq. 4.30) still needs to be expressed in terms of the other quantities. Using equation 4.43 again, one obtains:

$$\tilde{\chi}(12) = -i \frac{\delta G(11^+)}{\delta U_{\text{eff}}(2)} = -i \int G(13) \Gamma(34; 2) G(41) d(34). \quad (4.45)$$

At this point, any formal reference to the perturbing potential $U(1)$ has been removed, and the limit $U \rightarrow 0$ can be safely performed. This completes the derivation of Hedin's equations. For convenience, the final results are gathered in the table below.

$$\tilde{\chi}(12) = -i \iint G(13) \Gamma(34; 2) G(41) d(34) \quad (4.46a)$$

$$\Sigma(12) = i \iint G(14) W(1^+3) \Gamma(42; 3) d(34) \quad (4.46b)$$

$$W(12) = v(12) + \iint v(13) \tilde{\chi}(34) W(42) d(34) \quad (4.46c)$$

$$G(12) = G_0(12) + \iint G_0(13) \Sigma(34) G(42) d(34) \quad (4.46d)$$

$$\Gamma(12; 3) = \delta(12)\delta(13) + \iiint \frac{\delta\Sigma(12)}{\delta G(45)} G(56) \Gamma(67; 3) G(74) d(4567) \quad (4.46e)$$

This set of coupled equations constitutes Hedin's widely celebrated set of integro-differential equations whose self-consistent solution, in principle, solves the many-electron problem exactly. The pentagon sketched in figure 4.3 shows how the various physical quantities are interrelated: the polarization function re-normalizes the bare interaction resulting in the screened interaction $W(12)$. The screened interaction, $W(12)$, the many-body propagator $G(12)$, and the vertex function, $\Gamma(12; 3)$, which describe the interactions between virtual holes and electrons [40], are the essential ingredients for the determination of $\Sigma(12)$.

The practical solution of Hedin's equations is extremely complicated as they are not just numerical relations but contain a functional derivative (Eq. 4.46e) whose direct evaluation is challenging. The set of equations can, however, be iterated in order to derive useful approximations. Hedin himself stressed that his approach would be interesting only if few iterations were actually needed to obtain meaningful physical results.

4.7 The GW approximation

A widely used approach to the approximated solution of Hedin's equation is the so-called *GW* approximation (GW) approximation [5], which consists in neglecting all orders

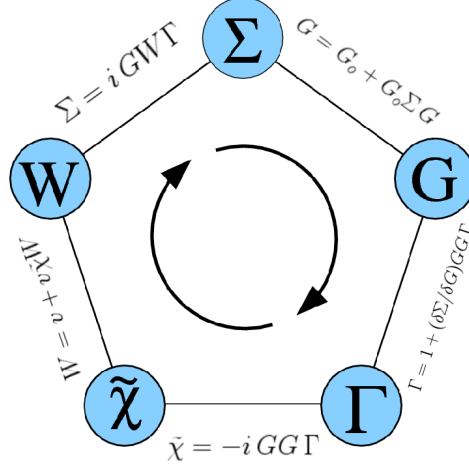


Figure 4.3: Graphical representation of Hedin's equations.

beyond the 0-th in the perturbative expansion of the vertex. In this case the vertex function is approximated as a local and instantaneous function:

$$\Gamma(12; 3) \approx \delta(1, 2)\delta(1, 3) \equiv \Gamma^{GW}(12; 3). \quad (4.47)$$

In real space, the self-energy reduces to a simple direct product of the dressed electron propagator, $G(12)$, and the dynamically screened interaction, $W(12)$:

$$\Sigma(12) = iG(12)W(1^+2). \quad (4.48)$$

Thanks to the neglect of vertex corrections, the irreducible polarizability $\tilde{\chi}(\omega)$ can also be easily evaluated starting from the many-body propagator, giving:

$$\tilde{\chi} = -iG(12)G(21^+). \quad (4.49)$$

Equation 4.49, once rewritten in terms of orbitals and energies, reduces to the Random Phase Approximation (RPA) for the polarizability proposed by Adler and Wiser [42, 43], as we will see in the next chapter. For convenience the complete set of equations to be solved within the GW formalism is reported in the table below.

$$\tilde{\chi}(12) = -iG(12)G(21^+) \quad (4.50a)$$

$$\Sigma(12) = iG(12)W(1^+2) \quad (4.50b)$$

$$W(12) = v(12) + \iint v(13)\tilde{\chi}(34)W(42)d(34) \quad (4.50c)$$

$$G(12) = G_0(12) + \iint G_0(13)\Sigma(34)G(42)d(34) \quad (4.50d)$$

$$\Gamma(12; 3) \approx \delta(12)\delta(13) \quad (4.50e)$$

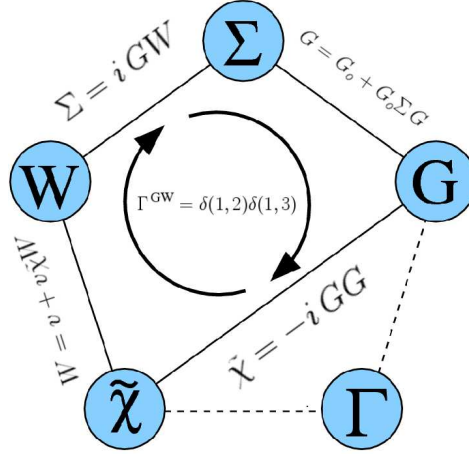


Figure 4.4: Graphical representation of the GW approximation. Usually only one "loop" is performed.

In practice, the dressed electron propagator $G(12)$ is never explicitly known and has to be approximated with the Green's function, $G_0(12)$, of an appropriate non-interacting system. For the sake of consistency, $G_0(12)$ should be obtained from the Hartree Hamiltonian, but since the Hartree theory completely neglects exchange and correlation effects, its propagator does not represent the best approximation. For this reason, the *GW* approximation is usually formulated using the KS Green's function as the starting point and the KS exchange-correlation potential has to be subtracted later from the resulting self-energy.

Ideally, the set of *GW* equations schematically depicted in Figure 4.4 should be iterated until self-consistency in all terms is reached. Hereafter this approach will be denoted with the name *SCGW*. *SCGW* is the most satisfactory implementation from a theoretical point of view since it preserves energy, electron number and total momentum [44, 45, 46]. However *SCGW* calculations for real systems are challenging, and very few have been reported in the literature (see for example [47, 48, 49] for applications to the homogeneous electron gas and [50, 51] for calculations on simple semiconductors). Fully self-consistent *GW* results are still under debate within the scientific community, and many authors agree that a proper self-consistent approach would need the inclusion of some kind of vertex correction during the solution of the equations.

For this reason, the most common approach employed in the *ab initio* community consists of using the best available approximation for G and W as a starting point and performing only a single-iteration of the parallelogram shown in Figure 4.4 (the so-called G_0W_0 , or one-shot *GW* method). In this case the self-energy is simply given by:

$$\Sigma(12) = iG_0^{\text{KS}}(12)W_0(1^+2) \quad (4.51)$$

where $G_0^{\text{KS}}(12)$ is the independent-particle propagator of the Kohn-Sham Hamiltonian, and the screened interaction is approximated with the random-phase approximation

(RPA) with KS energies and wave functions:

$$\chi^0(12) = -iG_0^{\text{KS}}(12)G_0^{\text{KS}}(12). \quad (4.52)$$

G_0W_0 gives results that are in much better agreement with experiments compared to $SCGW$ and with a considerably reduced computational effort. When available $SCGW$ calculations are compared to G_0W_0 , it turns out that self-consistency leads to a worsening of the results with respect to the one-shot approach in particular for spectroscopic properties. A systematic deterioration in the description of the bandwidth, quasiparticle excitations, lifetimes, and plasmon satellites is found, and gaps are overestimated. The better performance of G_0W_0 has been attributed to cancellation effects between the GW self-consistency and the vertex correction which are both neglected in G_0W_0 .

Improvements due the self-consistency can instead be found for the description of total energies [52, 53, 54]. These improvements can be traced back to the fulfillment of the above mentioned conservation laws, although further investigations are required to clarify the real effectiveness of the self-consistent approach for total energy calculations.

4.8 Perturbative Approach

For band structure calculations, it is more efficient to obtain the quasiparticle energies directly from equation 4.28 instead of solving the Dyson integral equation given in 4.50d, and searching for the poles of the Green's function. Following this approach, the equation to be solved is:

$$\left[\hat{h}_0(\mathbf{r}_1) + v_H(\mathbf{r}_1) \right] \Psi(\mathbf{r}_1) + \int \Sigma(\mathbf{r}_1, \mathbf{r}_2; \epsilon^{\text{QP}}) \Psi(\mathbf{r}_2) d\mathbf{r}_2 = \epsilon^{\text{QP}} \Psi(\mathbf{r}_1) \quad (4.53)$$

where

$$\Sigma(\mathbf{r}_1, \mathbf{r}_2; \omega) = \frac{i}{2\pi} \int G(\mathbf{r}_1, \mathbf{r}_2; \omega + \omega') W(\mathbf{r}_1, \mathbf{r}_2; \omega') e^{i\omega' \delta^+} d\omega'. \quad (4.54)$$

Although equation 4.53 looks similar to a mean-field approach, it does not constitute a mean-field formulation since the self-energy takes all dynamic many-electron processes into account. Moreover equation 4.53 departs from the KS equation since the QP eigenvectors and eigenvalues do have a direct physical meaning: they can be used to obtain both the charge density of the interacting system and to describe charged excitation properties. Unlike the KS approach, the problem in equation 4.53 is nonlinear and therefore the quasiparticle amplitudes are not orthogonal. Furthermore, as $\Sigma(12)$ is not Hermitian, the eigenvalues are complex-valued.

Despite all these fundamental differences, the Kohn-Sham exchange-correlation potential can be seen as a static, local and Hermitian approximation to $\Sigma(12)$. In many cases, the Kohn-Sham energies already provide a reasonable estimate of the band structure and are usually in qualitative agreement with experiment. For simple systems where the quasiparticle amplitudes, $\Psi(\mathbf{r})$, have been calculated, one also finds that the KS wave functions, $\Psi^{\text{KS}}(\mathbf{r})$, are usually very close to the QP results [55, 56, 57]. In silicon, for

example, the overlap between QP and LDA wave functions have been reported to be close to 99.9% (see [58], however, for comments and criticisms). This indicates that in the KS representation, the self-energy is generally a diagonally dominant matrix with negligible off-diagonal contributions.

These observations suggest that a simple, albeit accurate, solution for the QP energies can be obtained using first-order perturbation theory, treating $v_{xc}(\mathbf{r})$ as a zero-order approximation to the nonlocal and energy dependent $\Sigma(12)$ [55, 56]. Following this approach, the quasiparticle energies can be obtained as the root of the nonlinear equation:³

$$\epsilon^{\text{QP}} = \epsilon^{\text{KS}} + \langle \Psi^{\text{KS}} | \Sigma(\epsilon^{\text{QP}}) - v_{xc} | \Psi^{\text{KS}} \rangle. \quad (4.55)$$

The solution of equation 4.55 requires the knowledge of the frequency dependence of $\Sigma(\epsilon)$, which is usually not available. Under the initial assumption that the QP energies are only a correction to the Kohn-Sham eigenvalues, we can expand the self-energy operator around ϵ^{KS} obtaining a closed expression for ϵ^{QP} :

$$\Sigma(\epsilon^{\text{QP}}) \approx \Sigma(\epsilon^{\text{KS}}) + (\epsilon^{\text{QP}} - \epsilon^{\text{KS}}) \frac{\partial \Sigma(\epsilon)}{\partial \epsilon^{\text{KS}}}, \quad (4.56)$$

where a compact notation for the diagonal elements of Σ has been used: $\Sigma(\epsilon) \equiv \langle \Psi^{\text{KS}} | \Sigma(\epsilon) | \Psi^{\text{KS}} \rangle$. The final result is:

$$\epsilon^{\text{QP}} = \epsilon^{\text{KS}} + Z \langle \Psi^{\text{KS}} | \Sigma(\epsilon^{\text{KS}}) - v_{xc} | \Psi^{\text{KS}} \rangle, \quad (4.57)$$

where

$$Z \equiv \left[1 - \langle \Psi^{\text{KS}} | \frac{\partial \Sigma(\epsilon)}{\partial \epsilon^{\text{KS}}} | \Psi^{\text{KS}} \rangle \right]^{-1} \quad (4.58)$$

is the so-called renormalization factor.

4.9 Hartree-Fock as a limiting case of GW

At this point, it is interesting to relate the *GW* approximation to the Hartree-Fock (HF) method. In Hartree-Fock theory [21, 22, 2], an upper bound for the total energy is obtained by minimizing the energy in the Hilbert subspace made up of all possible antisymmetric Slater determinants constructed from single-particle orbitals. The Hamiltonian whose eigenstates define the optimal Slater determinant contains the Fock exchange operator:

$$\Sigma_x^{\text{HF}}(\mathbf{r}_1, \mathbf{r}_2) = - \sum_i \theta(\mu - \epsilon_i) \psi_i(\mathbf{r}_1) \psi_i^\dagger(\mathbf{r}_2) v(\mathbf{r}_1, \mathbf{r}_2), \quad (4.59)$$

³ In principle, only the real part of the self-energy should be considered when calculating the QP energy. To keep the notation as simple as possible, this has been notationally omitted.

a nonlocal and static operator which can be re-expressed in terms of $G(12)$ and $v(12)$ as:

$$\Sigma^{\text{HF}} = iG(1, 2)v(1, 2^+). \quad (4.60)$$

Equation 4.60 shows that the GW expression given in Eq. 4.50b reduces to the Hartree-Fock value in the limit $W \rightarrow v$, *i.e.* when the dynamical screening due to the surrounding electronic charge can be neglected. As a consequence, the GW approximation should give results very similar to the HF ones in case of atoms or molecules in which Hartree-Fock is known to work well. On the other hand, GW improves upon HF as it can formally be regarded as a Hartree-Fock approach with a dynamically screened interaction instead of the static Coulomb term. Indeed, GW improves the description of crystalline solids, where screening effects play a fundamental role, whereas HF leads to a severe overestimation of band gaps.

4.10 Quasiparticle self-consistent GW

As already discussed in section 4.7, $SCGW$ requires the iteration of Hedin's equation with the vertex function approximated as local and instantaneous. A few $SCGW$ calculations have been reported in the literature and the results are not encouraging. $SCGW$ has indeed shown to worsen the results obtained with the standard one-shot G_0W_0 method. Holm and von Barth found that $SCGW$ overestimates the bandwidth in the homogeneous electron gas [48]. In potassium, the valence band width is overestimated by more than 1 eV while both the direct and indirect band gap in silicon are dramatically overestimated [51]. It is thus not clear that a fully self-consistent approach to GW consistency represents a clear improvement upon G_0W_0 and, moreover, $SCGW$ is much more demanding than standard G_0W_0 .

A different approach to the self-consistency issue is the Quasiparticle Self-Consistent GW (QPSCGW) method proposed by Faleev and co-workers [16, 59, 11]. QPSCGW employs a perturbation theory constructed around some initial single-particle Hamiltonian $\hat{h}$, and uses a self-consistent procedure to determine an optimal non-interacting Hamiltonian

$$h_{QPSCGW}(\mathbf{r}_1, \mathbf{r}_2) = \hat{h}(\mathbf{r}_1) + \tilde{\Sigma}(\mathbf{r}_1, \mathbf{r}_2), \quad (4.61)$$

with the true self-energy approximated via a static and Hermitian operator $\tilde{\Sigma}(\mathbf{r}_1, \mathbf{r}_2)$ constructed according to:

$$\tilde{\Sigma}_{ij}^{\mathbf{k}} \equiv \frac{1}{2} \text{Herm} [\Sigma_{ij}^{\mathbf{k}}(\epsilon_{i\mathbf{k}}) + \Sigma_{ij}^{\mathbf{k}}(\epsilon_{j\mathbf{k}})], \quad (4.62)$$

where Herm means that only the Hermitian part of the matrix has to be considered. Then the approximated self-energy, $\tilde{\Sigma}$, is diagonalized providing a new set of quasiparticle amplitudes and energies. From this new set of orbitals, a new density and the corresponding Hartree potential is generated, a new $\tilde{\Sigma}$ is constructed and the procedure is iterated to self-consistency. Ideally, the final result should not depend on the

initial Hamiltonian even though no firm mathematical proof for this has been reported so far. The $QPSCGW$ approach improves the G_0W_0 results, giving band gaps very close to experiments albeit systematically slightly overestimated. Figure 4.5 compares the values for the fundamental gap of a variety of compounds as calculated using the standard one-shot G_0W_0 approach and the results coming from the self-consistent solution of equation 4.61. The $QPSCGW$ data reported in right panel of the picture clearly shows that self-consistent approach of Faleev and co-workers leads to better agreement with experiment, errors are small and highly systematic.

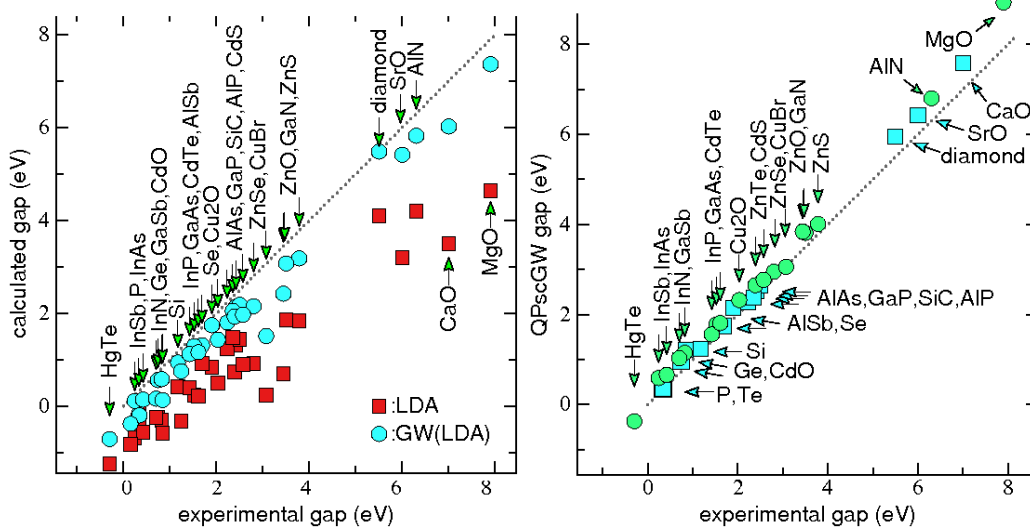


Figure 4.5: Left panel: LDA and G_0W_0 results for the band gap. The G_0W_0 gaps improve on the LDA, but are still systematically underestimated. Right panel: Results obtained with the $QPSCGW$ method. Direct $\Gamma-\Gamma$ transitions are shown as green circles, all other gaps are shown as blue squares. Taken from [59], also available at cond-mat/0510408

4.11 Core-valence decoupling

A point deserving clarification is the appropriate treatment of the interactions between core and valence electrons, when the frozen core approximation is employed to separate electrons into core and valence states. The separation of the electronic degrees of freedom into two different groups is the standard approach followed in many *ab initio* methods in order to avoid having to deal with the strongly localized and quickly oscillating core wave functions. The proper description of these orbitals would require either an exceedingly dense FFT mesh or an appropriate basis set localized within the core region (*e.g.* a radial support mesh or a set of atomic-like orbitals). On the other hand, considering the core frozen in the atomic reference configuration gives results for the total energy that are close to fully self-consistent all-electron results since, as shown in [60], the correction to the frozen-core approximation vanishes to first order in the charge density differences.

There is, however, no analogous valid cancellation theorem for many-body calculations, and the description of the core contribution within a formalism based on the frozen core approximation is quite a delicate issue. All the quantities introduced so far – the self-energy $\hat{\Sigma}(\omega)$, the irreducible polarization $\tilde{\chi}(\omega)$, and the screened interaction $W(\omega)$ – should include many-body contributions coming from all the electrons of the system, both core and valence states. To clarify this important point, following the discussion in [56], it is illuminating to separate the Green’s function into two different parts:

$$G(\mathbf{r}_1, \mathbf{r}_2; \omega) = G_c(\mathbf{r}_1, \mathbf{r}_2; \omega) + G_v(\mathbf{r}_1, \mathbf{r}_2, \omega), \quad (4.63)$$

where G_c includes only the core electrons and G_v the remaining states. A similar decomposition can be done for the polarizability:

$$\tilde{\chi}(\mathbf{r}_1, \mathbf{r}_2; \omega) = \tilde{\chi}_c(\mathbf{r}_1, \mathbf{r}_2; \omega) + \tilde{\chi}_v(\mathbf{r}_1, \mathbf{r}_2, \omega). \quad (4.64)$$

The self-energy operator can thus be decomposed into three different contributions [61]

$$\Sigma(\mathbf{r}_1, \mathbf{r}_2; \omega) = iG_c W + iG_v W_v \tilde{\chi}_c W_v + iG_v W_v. \quad (4.65)$$

The first term in the equation 4.65 is a core-valence screened exchange contribution. Since the screening is ineffective in the core region [62], $iG_c W$ can be accurately described in terms of a bare core-valence exchange interaction which reduces to the Fock operator, $iG_c v$, generated by core states. The second term in equation 4.65 is a screened polarization potential arising from the polarization of core electrons due to many-body effects. Finally, the last term, $iG_v W_v$, describes many-body interactions among valence electrons. In a GW implementation based on the frozen-core approximation, only the last term $iG_v W_v$ is explicitly accounted for, whereas the first two contributions in equation 4.65 have to be approximated. If pseudopotentials (PPs) are used, the missing two terms are replaced by the core-valence exchange-correlation term which is frozen inside the pseudopotential operator after the unscreening procedure in the pseudopotential generation [63]. In addition, the exact many-body contribution is described at the Kohn-Sham level because this term is directly imported from the self-consistent DFT calculation performed in the atomic environment. As a consequence, many-body effects are only partly accounted for and, in addition, possible core-relaxation effects due to the change in the chemical environment are completely neglected. To justify the above mentioned approximations, Hybertsen and Louie argued that, although the net contribution of the first two terms in equation 4.65 might even be of the order of some eV, the error introduced by replacing the proper core-valence terms by the frozen Kohn-Sham contribution should be much smaller [56]. For simple systems, the two authors gave a quantitative analysis of the error which, in the worst case, was estimated to be at most 0.3 eV.

For several decades – thanks to the excellent results obtained with pseudopotential-based GW applications – the argumentation provided by Hybertsen and Louie represented a commonly accepted justification of the validity of the core-valence decoupling

within the frozen-core approximation. In recent years, however, GW calculations performed with AE approaches [9, 10, 64, 11] have revealed that a fully consistent treatment of the electronic degrees of freedom produces band gaps that are systematically smaller than pseudopotential results. This worsens the agreement between GW and experiments. These findings have led to quite an intense debate in the scientific literature concerning the reliability of the pseudopotential approach for many-body calculations. According to the point of view of the all-electron community, using pseudopotentials for many-body calculations leads to a fortuitous improvement of the agreement between GW results and experiments. The fortuitous improvement can be traced back both to the neglect of core relaxation effects as well as to spurious effects introduced by replacing the true wave function with a smooth and nodeless pseudized image [12]. An accurate analysis of the relative importance of the two approximations has been recently reported in [12] whose most important results are summarized in figure 4.6.

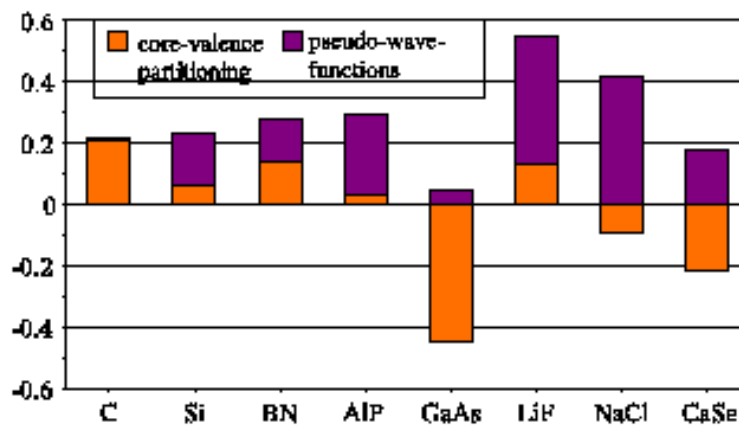


Figure 4.6: Contributions to the discrepancies between AE and PP G_0W_0 gaps due to core-valence partitioning and the replacement of the true all-electron wave function with the pseudized image. The dominant contribution to the discrepancy between AE and PP results is system dependent. For systems with shallow cores, such as GaAs, most of the discrepancy comes from the neglecting of many-body effects due to the core. In other compounds, such as LiF and NaCl, the dominant contribution comes from the pseudization of the wave functions. Taken from [12].

In the pseudopotential community, on the other hand, there is a common agreement that the pseudopotential theory still represents a reliable tool for many-body calculations, although the pseudopotentials have to be generated with much more care than that needed for standard ground-state applications. For instance, since the convergence of the quasiparticle correction is exceedingly slow with respect to the number of empty states [65], pseudopotentials for accurate GW calculations should have a high transferability which can be achieved only if multiple projectors per angular channel are

used [66, 67].⁴ In systems with shallow cores, a proper treatment of many-body core-valence effects usually requires the inclusion of semi-core states in the valence together with the correct description of the nodal shape of the functions in the core region. This is especially true for the case of localized *d*- or *f*-electrons that, although well separated in energy from inner core states, present a significant spatial overlap with these [68]. Shallow cores are commonly included also to perform total energy calculations, but – since the total energy is mainly sensitive to the rearrangement of the tails of the wave functions in the interstitial region – a relatively smooth pseudized wave function (requiring a reasonable cutoff energy) can usually guarantee an accurate description of structural and vibrational properties. Unfortunately, the same degree of accuracy is not sufficient for *GW* studies as QP energies are sensitive both to the localization of the wave function and to the correct description of the nodal shape of the functions in the core regions. Within the pseudopotential formalism, the localization of the wave function can be reasonably described by choosing a core radius close the maximum of the AE wave function, at the price of a drastic increase of the plane wave basis set [13, 69]. Figure 4.7 reports the case of nickel as a typical example. In Ni the spatial overlap between the *3d* electrons and the inner *3s* and *3p* shells plays a crucial role for an accurate evaluation of the self-energy, in particular for the estimate of the Fock-like contribution to $\hat{\Sigma}$. A small core radius for the *3s*, *3p* and *3d* channels is needed to properly describe the orbital overlap, but this choice produces quite a hard pseudopotential, requiring around 120 Hartree to converge the atomic energy below 1 meV. Easier convergence with respect to cut-off energies may be obtained by relaxing the norm-conservation condition following the ultrasoft pseudopotential formalism [70], but even this approach does not help improve the description of the spatial localization. Ultrasoft wave functions are not even correctly normalized, and additional compensation charges have to be added to describe the correct electrostatics outside the core region [70].

Another drawback of the PP formalism is that having the correct nodal shape is a task virtually impossible unless an appropriate mapping between the pseudized wave function and the corresponding AE wave function is established as proposed in [72, 73, 74]. The proper description of many-body core-polarization effects with PPs does not represent an unsurmountable problem since a partial description of the contribution given by $iG_v W_v \tilde{\chi}_c W_v$ can be achieved following the method proposed by Shirley and Martin [75, 76, 77] in which core-polarization effects are included from first principles in the pseudopotential operator.

It is, however, clear from the above discussion that, in the case of many-body calculations, many of the advantages obtained from the pseudization of the wave functions are lost. A *GW* calculation based on the pseudopotentials has to rely on several (difficult to control) assumptions, as well as on physical approximations that are hard to estimate.

⁴Pseudopotential projectors are radial functions entering the nonlocal part of the Hamiltonian. They are constructed such that the scattering properties of the true atom are correctly reproduced over the energy range of interest.

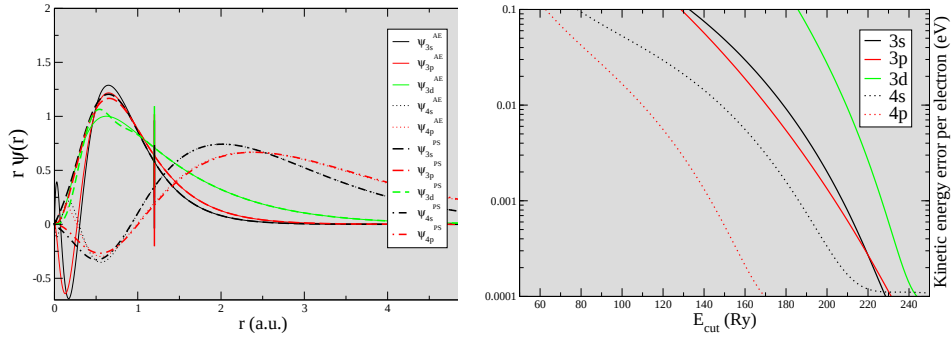


Figure 4.7: Left panel: All-electron wave functions (solid lines) and corresponding pseudized images (dashed lines) corresponding to the third and fourth shell of Ni. A norm-conserving Troullier-Martins [71] scheme is used. The vertical bar indicates the position of the core radius for the pseudization of the AE orbitals. Only one projector per l -channel is used, and an ionized atomic configuration is considered as reference state. Right panel: Convergence of the kinetic energy as a function of the plane wave cutoff on a logarithmic scale. The curves give an initial estimate of the convergence rate of the total energy in the crystalline environment. The slow convergence of the plane wave expansion is principally due to the small core radius used for the $3d$ electrons.

Results in reasonable agreement with AE calculations can be obtained only at the price of a dramatic increase of the computational effort [13]. One of the goals of the present work consists in implementing a different methodological approach to the solution of Hedin's equations based on the so-called PAW formalism proposed by Blöchl in 1994 [15]. This new approach will permit to remove many of the limitations intrinsic to the use of pseudopotentials allowing us to reach *GW* results closer to an all-electron implementation while still maintaining many of the advantages of the pseudopotential theory: fast convergence with respect to the basis set truncation, computational efficiency, and flexibility. In the next chapter, we shall describe the basic equations needed to implement the *GW* equations in the frequency- and reciprocal-space domain. In chapter 7, we then proceed to discuss how the implementation has to be generalized in order to benefit from the improvements introduced by the PAW method.

Chapter 5

GW Implementation

The aim of this chapter is to introduce the basic equations required to implement the *GW* formalism in frequency and reciprocal space as well as the different algorithms available in the ABINIT code [78, 79] to evaluate the self-energy and the quasiparticle energies. The material is organized as follows. Section 5.1 summarizes the problem and introduces the basic objects and the notations used in the following of the discussion. The RPA expression for the irreducible polarizability is presented in section 5.2. An analysis of the scaling of the formalism in Fourier space is done in section 5.3. In section 5.4, we then discuss how to take advantage of symmetries to reduce both the computational effort and the memory requirements needed for the evaluation of the RPA polarizability. The plasmon-pole technique, commonly used to model the frequency dependence of $W(\omega)$ in an efficient way, is presented in section 5.5, and the four different approaches presently implemented in the code are briefly described. In section 5.6, we present an algorithm for the RPA calculation which outperforms the standard implementation based on the Adler-Wiser expression, when a wide range of frequencies is required. Finally, in section 5.7, the different available methods for the evaluation of the quasiparticle energies are described.

Henceforth the spin quantum degree of freedom is explicitly suppressed as spin degeneracy in spin-compensated systems can usually be taken into account by means of a multiplicative factor. The changes needed to treat systems with collinear magnetism are explicitly described when needed.

5.1 Evaluation of the Screened Interaction

The *GW* self-energy, a simple product in the time domain according to equation 4.50b, becomes a computationally expensive convolution in frequency-space:

$$\Sigma(\mathbf{r}_1, \mathbf{r}_2; \omega) = \lim_{\delta \rightarrow 0^+} \frac{i}{2\pi} \int G(\mathbf{r}_1, \mathbf{r}_2; \omega + \omega') W(\mathbf{r}_1, \mathbf{r}_2; \omega') e^{i\omega'\delta} d\omega', \quad (5.1)$$

where the complex exponential has been introduced to enforce the correct time-ordering, and the limit $\delta \rightarrow 0^+$ is taken after the integration is performed. The first step of any *GW* implementation therefore consists of calculating the screened interaction, $W(\mathbf{r}_1, \mathbf{r}_2; \omega)$,

using Kohn-Sham wavefunctions and eigenvalues as a zeroth order approximation to the quasiparticle amplitudes and energies. The two-point function $W(\mathbf{r}_1, \mathbf{r}_2; \omega)$ can then be obtained from the knowledge of the inverse dielectric matrix according to equation 4.34. Once $W(\omega)$ is available, several techniques can be used to perform the convolution given in Eq. 5.1, finally obtaining the quasiparticle corrections via equation 4.53.

As already discussed in section 4.6, the dielectric matrix $\epsilon(12)$ is related to the irreducible polarizability $\tilde{\chi}(12)$ by the following relation:

$$\begin{aligned} \epsilon(1, 2) &= \frac{\delta v_{\text{ext}}(1)}{\delta v_{\text{tot}}(2)} = \\ &= \frac{\delta}{\delta v_{\text{tot}}(2)} \left[v_{\text{tot}}(1) - \int v(1, 3) n(3) d3 \right] = \delta(1, 2) - \int v(1, 3) \tilde{\chi}(3, 2) d3, \end{aligned} \quad (5.2)$$

from which, Fourier transforming to the frequency domain, one obtains:

$$\epsilon(\mathbf{r}_1, \mathbf{r}_2; \omega) = \delta(\mathbf{r}_1, \mathbf{r}_2) - \int v(\mathbf{r}_1, \mathbf{r}_3) \tilde{\chi}(\mathbf{r}_3, \mathbf{r}_2; \omega) d\mathbf{r}_3. \quad (5.3)$$

Since we are assuming an ideal crystal environment, the two-point function $\epsilon(\mathbf{r}_1, \mathbf{r}_2; \omega)$ must be invariant under an arbitrary crystalline translation $\mathbf{R}$ acting on $\mathbf{r}_1$ and $\mathbf{r}_2$, *i.e.*

$$\epsilon(\mathbf{r}_1 + \mathbf{R}, \mathbf{r}_2 + \mathbf{R}; \omega) = \epsilon(\mathbf{r}_1, \mathbf{r}_2; \omega), \quad (5.4)$$

and any two-point function having the same translational symmetry as the underlying Bravais lattice can be expanded in a Fourier series according to the conventions reported in section A.1. In frequency and reciprocal space, equation 5.3 can hence be rewritten as:

$$\epsilon_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}; \omega) = \delta_{\mathbf{G}_1 \mathbf{G}_2} - v(\mathbf{q}, \mathbf{G}_1) \tilde{\chi}_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}; \omega), \quad (5.5)$$

where $\mathbf{q}$ is a point lying within the first Brillouin zone and

$$v(\mathbf{q}, \mathbf{G}) = \frac{4\pi}{|\mathbf{q} + \mathbf{G}|^2} \quad (5.6)$$

is the Fourier representation of the bare Coulomb potential, singular at the origin due to the long range of the interaction in real space. Due to the presence of the Coulomb singularity, special care has to be used in equation 5.5 when considering the long-wavelength limit $\mathbf{q} \rightarrow 0$ as the value will in general depend on the direction in which the limit is taken [80, 81, 82]. A detailed discussion of this rather technical point is reported in the two appendices C and E.

To facilitate the use of symmetry properties, it is preferable to replace $\epsilon_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}, \omega)$ with the so-called symmetrized dielectric matrix [83]:

$$\tilde{\epsilon}_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}, \omega) \equiv \frac{|\mathbf{q} + \mathbf{G}_1|}{|\mathbf{q} + \mathbf{G}_2|} \epsilon_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}, \omega), \quad (5.7)$$

and $\epsilon_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega)$ with the corresponding symmetrized inverse dielectric matrix:

$$\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega) \equiv \frac{|\mathbf{q} + \mathbf{G}_1|}{|\mathbf{q} + \mathbf{G}_2|} \epsilon_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega). \quad (5.8)$$

Inserting the definition 5.8 in equation 4.50c yields the following expression for the screened Coulomb interaction in frequency- and momentum-space

$$W_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega) = 4\pi \frac{\epsilon_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega)}{|\mathbf{q} + \mathbf{G}_2|^2} = 4\pi \frac{\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega)}{|\mathbf{q} + \mathbf{G}_1||\mathbf{q} + \mathbf{G}_2|}. \quad (5.9)$$

Since $\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega)$ can be obtained from equation 5.5 by a simple matrix inversion, the knowledge of the irreducible polarizability $\tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega)$ is sufficient to evaluate the screened interaction.

5.2 RPA for the Irreducible Polarizability

Within the *GW* approximation, the irreducible polarizability in the space-time domain is simply given by the product of two Green's functions:

$$\tilde{\chi}(12) \approx -i G(12) G(21) \equiv \tilde{\chi}^{\text{RPA}}(12), \quad (5.10)$$

where the notation $\tilde{\chi}^{\text{RPA}}(12)$ has been introduced to stress that the true many-body polarizability is now replaced by an approximated expression where vertex corrections are completely neglected. According to Eq. A.4, the product in the time domain becomes a convolution in the frequency domain. In $\mathbf{r}$ - and ω -space, $\tilde{\chi}^{\text{RPA}}(12)$ thus reads:

$$\tilde{\chi}^{\text{RPA}}(\mathbf{r}_1, \mathbf{r}_2; \omega) = -\frac{2i}{2\pi} \int G(\mathbf{r}_1, \mathbf{r}_2, \omega + \omega') G(\mathbf{r}_2, \mathbf{r}_1, \omega') d\omega', \quad (5.11)$$

where the factor two accounts for spin degeneracy. Substituting the Lehmann representation of the time-ordered Green's function into equation 5.11 leads to:

$$\begin{aligned} \tilde{\chi}^{\text{RPA}}(\mathbf{r}_1, \mathbf{r}_2; \omega) = & -\frac{i}{\pi} \sum_{\alpha\beta} \Psi_{\alpha}(\mathbf{r}_1) \Psi_{\beta}^{\dagger}(\mathbf{r}_1) \Psi_{\alpha}^{\dagger}(\mathbf{r}_2) \Psi_{\beta}(\mathbf{r}_2) \times \\ & \times \underbrace{\int \frac{1}{\omega + \omega' - \epsilon_{\alpha} + i\eta \text{sign}(\epsilon_{\alpha} - \mu)} \frac{1}{\omega' - \epsilon_{\beta} + i\eta \text{sign}(\epsilon_{\beta} - \mu)} d\omega'}_{\Theta_{\alpha\beta}(\omega)}, \end{aligned} \quad (5.12)$$

where $\Psi_{\alpha}(\mathbf{r})$ is either a single-particle wave function of an appropriate mean-field Hamiltonian or a quasiparticle amplitude, depending on the approximation used for the propagator $G(12)$. The index α is a contracted notation to indicate an appropriate set of quantum numbers. The frequency dependence of $\tilde{\chi}^{\text{RPA}}(\mathbf{r}_1, \mathbf{r}_2; \omega)$ is entirely contained in the integral $\Theta_{\alpha\beta}(\omega)$, which can be analytically evaluated via the residue theorem by

closing the integration path either in the upper or the lower half of the complex plane. The complex poles of the integrand are located at

$$\omega' = \begin{cases} -\omega + \epsilon_\alpha - i\eta \operatorname{sign}(\epsilon_\alpha - \mu) \\ \epsilon_\beta - i\eta \operatorname{sign}(\epsilon_\beta - \mu). \end{cases} \quad (5.13)$$

From equation 5.13 it follows that occupied states lead to poles positioned in the upper half of the complex plane, while empty states produce poles in the lower. Consequently $\Theta_{\alpha\beta}(\omega)$ vanishes when the two poles of the Green's functions are located in the same half plane as the integration path can be closed in the complex region where no pole occurs. $\Theta_{\alpha\beta}(\omega)$ will differ from zero if and only if one state, say α , is occupied and the other, say β , is empty. Only the transitions from occupied to empty states contribute to $\tilde{\chi}^{\text{RPA}}(\omega)$, and this constraint can be accounted for by introducing two Heaviside step-functions $\theta(x)$, one for valence states, the other for conduction electrons. Then the integration can be easily carried out yielding

$$\Theta_{\alpha\beta}(\omega) \equiv \left[\frac{\theta(\epsilon_\alpha - \mu)\theta(\mu - \epsilon_\beta)}{\omega - \epsilon_\alpha + \epsilon_\beta + i\eta} - \frac{\theta(\mu - \epsilon_\alpha)\theta(\epsilon_\beta - \mu)}{\omega - \epsilon_\alpha + \epsilon_\beta - i\eta} \right] = \frac{f(\epsilon_\beta) - f(\epsilon_\alpha)}{\omega - \epsilon_\alpha + \epsilon_\beta + i\eta \operatorname{sign}(\epsilon_\alpha - \epsilon_\beta)}, \quad (5.14)$$

where $f(\epsilon)$ are the occupation factors defined in the interval $[0, 1]$ and η is an infinitesimally small positive number that shifts the poles such that the resulting function is time-ordered.

Substituting Eq. 5.14 in Eq. 5.12, one realizes that the expression obtained for $\tilde{\chi}^{\text{RPA}}(12)$ is equivalent to the RPA expression given by Adler and Wiser in [42, 43]. in which, however, single-particle orbitals are replaced by quasiparticle amplitudes. If the many-body propagator $G(12)$ is approximated with the DFT independent-particle Green's function, G_0^{KS} , one obtains that:

$$\tilde{\chi}^{\text{RPA}}(12) = -i G_0^{\text{KS}}(12) G_0^{\text{KS}}(21) \equiv \chi^0(12), \quad (5.15)$$

where $\chi^0(12)$ describes the linear response of the auxiliary non-interacting Kohn-Sham system. Note, however, that the response of the KS system differs from that of the true many-body system, $\tilde{\chi}(\omega)$, because excitonic or other many-body effects are not accounted for. The difference between $\chi^0(\omega)$ and $\tilde{\chi}(\omega)$ is made clear in figure 5.1 where the analytic structure of $\chi^0(\omega)$ and $\tilde{\chi}(\omega)$ is depicted schematically. According to equation 5.14, $\chi^0(\omega)$ has poles in correspondence with the neutral energy transitions of the KS system, namely transitions between empty and occupied states of the N -electron KS Hamiltonian. Since the KS eigenvalues have no direct physical meaning, the poles of $\chi^0(\omega)$ are not expected to coincide with the poles of $\tilde{\chi}(\omega)$ although a link between the $\chi^0(\omega)$ and $\tilde{\chi}(\omega)$ can be established within TDDFT [1, 85]. The fact that $\chi^0(\omega)$ constitutes an approximation to the true many-body response function is an important point worth considering when performing *GW* calculations, since the particular approximation used for $G(12)$ in equation 5.11 has a direct influence on the physical accuracy of

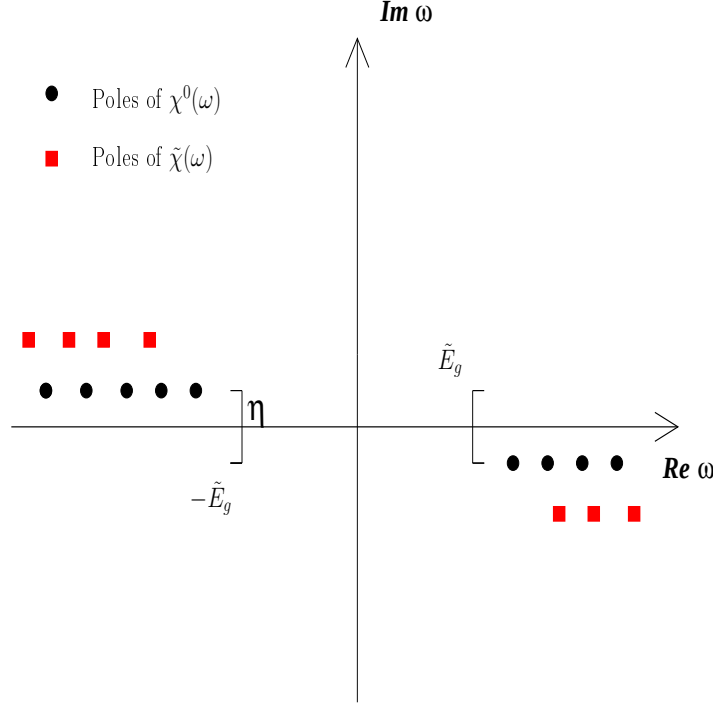


Figure 5.1: Schematic analytic structure of the time-ordered density-density correlation function $\chi^0(\omega)$ (and $\tilde{\chi}(\omega)$) of the KS system. In an infinite system, the poles form two distinct branch cuts: a branch cut located in the lower half of the complex plane, from $\tilde{E}_g$ to $+\infty$, which corresponds to positive (resonant) transitions, and a second branch cut in the complex upper plane, from $-\infty$ to $-\tilde{E}_g$, associated to negative (anti-resonant) transitions. $\tilde{E}_g$ is the minimum gap in the excitation spectrum of the N -electron KS Hamiltonian (neutral excitations). No poles are present in the energy range from $-\tilde{E}_g$ to $+\tilde{E}_g$. The exact many-body polarizability, $\tilde{\chi}(\omega)$, also has a similar analytic structure, although its poles are located at the *true* neutral excitation energies of the many-body system [84].

the polarizability and, consequently, on the resulting screened interaction. For example, if the dressed propagator is approximated with the single-particle Kohn-Sham Green's function, the poles of $\chi^0(\omega)$ in the fourth quadrant of the complex plane will be shifted towards lower energy, due to the underestimation of the band gap in LDA or GGA, and the resulting $W(\omega)$ will be overestimated. This is especially true if the underestimation of the gap in DFT is severe or if the Kohn-Sham theory predicts the system to be metallic.

At this point it is useful to discuss the symmetry properties of $\chi^0(\omega)$ for real and purely imaginary frequencies since these symmetry relations will play an important role in reducing the computational effort of calculations. From the definition of $\Theta_{\alpha\beta}(\omega)$ given

in equation 5.14, it immediately follows that:

$$\Theta_{\alpha\beta}(\omega) = \Theta_{\beta\alpha}(-\omega), \quad (5.16)$$

from which it can be shown that $\chi^0(\omega)$ is an even function of ω :

$$\chi^0(\mathbf{r}_1, \mathbf{r}_2; \omega) = \chi^0(\mathbf{r}_1, \mathbf{r}_2; -\omega). \quad (5.17)$$

In the particular case of a crystalline system, the eigenstates of the Hamiltonian can be indexed by the crystalline momentum $\mathbf{k}$, assumed to be in the first Brillouin zone, and the band index n . Equation 5.12 can therefore be rewritten as

$$\chi^0(\mathbf{r}_1, \mathbf{r}_2; \omega) = -\frac{2i}{2\pi} \sum_{\mathbf{k}_1 n_1} \sum_{\mathbf{k}_2 n_2} \Psi_{\mathbf{k}_1 n_1}(\mathbf{r}_1) \Psi_{\mathbf{k}_2 n_2}^\dagger(\mathbf{r}_1) \Psi_{\mathbf{k}_1 n_1}^\dagger(\mathbf{r}_2) \Psi_{\mathbf{k}_2 n_2}(\mathbf{r}_2) \times \Theta_{\mathbf{k}_1 n_1, \mathbf{k}_2 n_2}(\omega). \quad (5.18)$$

The above expression can be used as starting point for the derivation of an additional symmetry property of $\chi^0(\mathbf{r}_1, \mathbf{r}_2; \omega)$, which is a direct consequence of time-reversal invariance. In systems for which $\hat{H}^\dagger = \hat{H}$, the wave function obtained by taking the complex conjugate of a Bloch state with crystalline momentum $\mathbf{k}$ is still an eigenstate of the Hamiltonian with the same eigenvalue, but reversed crystalline momentum $-\mathbf{k}$. More explicitly:

$$\begin{cases} \epsilon_{-\mathbf{k}} &= \epsilon_{\mathbf{k}}, \\ \Psi_{-\mathbf{k}}(\mathbf{r}) &= \Psi_{\mathbf{k}}^\dagger(\mathbf{r}). \end{cases} \quad (5.19)$$

The two symmetry relations in equation 5.19 can be exploited to rearrange the double summation in equation 5.18. After some algebra, one arrives at the important result that $\chi^0(\mathbf{r}_1, \mathbf{r}_2; \omega)$ is a symmetric function of the spatial coordinates, namely:

$$\chi^0(\mathbf{r}_1, \mathbf{r}_2; \omega) = \chi^0(\mathbf{r}_2, \mathbf{r}_1; \omega). \quad (5.20)$$

Once expressed in reciprocal space, equation 5.20 translates into the equality:

$$\chi^0_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}, \omega) = \chi^0_{-\mathbf{G}_2 - \mathbf{G}_1}(-\mathbf{q}, \omega). \quad (5.21)$$

which can be used to halve the number of independent $\mathbf{q}$ -points. Additional symmetries are present for frequencies along the imaginary axis. In this case, the complex imaginary shift $i\eta$ can be safely set to zero and consequently

$$\Theta_{\alpha\beta}(i\omega) = \Theta_{\alpha\beta}^\dagger(i\omega) = \Theta_{\beta\alpha}(-i\omega), \quad (5.22)$$

from which the following set of equalities readily follow

$$\left[\chi^0(\mathbf{r}_1, \mathbf{r}_2; i\omega) \right]^\dagger = \chi^0(\mathbf{r}_1, \mathbf{r}_2; -i\omega) = \chi^0(\mathbf{r}_2, \mathbf{r}_1; i\omega) = \chi^0(\mathbf{r}_1, \mathbf{r}_2; i\omega). \quad (5.23)$$

In $\mathbf{G}$ -space, equation 5.23 leads to the Hermiticity condition:

$$\chi^0_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, i\omega) = \left[\chi^0_{\mathbf{G}_2\mathbf{G}_1}(\mathbf{q}, i\omega) \right]^\dagger. \quad (5.24)$$

Note that both equations 5.22 and 5.24 still hold for real frequencies when ω is located within the frequency region where $\chi^0(\omega)$ has no poles.

After this preamble, all the symmetry properties of $\chi^0(\mathbf{r}_1, \mathbf{r}_2; \omega)$ in the frequency domain have been discussed; only the explicit expression of the plane wave representation, $\chi^0_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega)$, in terms of KS eigenvalues and eigenvectors still remains to be derived. Transforming equation 5.18 into Fourier space according to the conventions introduced in section A.1, one arrives at:

$$\chi^0_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega) = \frac{2}{V} \sum_{\substack{\mathbf{k} \\ b_1 b_2}} \left(f(\epsilon_{b_1\mathbf{k}-\mathbf{q}}) - f(\epsilon_{b_2\mathbf{k}}) \right) \frac{M_{\mathbf{G}_1}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) [M_{\mathbf{G}_2}^{b_1 b_2}(\mathbf{k}, \mathbf{q})]^\dagger}{\omega + \epsilon_{b_1\mathbf{k}-\mathbf{q}} - \epsilon_{b_2\mathbf{k}} - i\eta \text{sign}(\epsilon_{b_1\mathbf{k}-\mathbf{q}} - \epsilon_{b_2\mathbf{k}})}, \quad (5.25)$$

where the double sum over $\mathbf{k}$ -points in Eq. 5.18 has been reduced to a single integration over the full Brillouin zone by virtue of selection rules introduced by the conservation of the crystalline momentum. In equation 5.25, $M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q})$ is a shorthand notation to denote the so-called oscillator matrix elements, namely the matrix element of a plane wave between two wavefunctions of crystalline momentum $\mathbf{k}$ and $\mathbf{k} - \mathbf{q}$:

$$\begin{aligned} M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) &\equiv \langle \mathbf{k} - \mathbf{q}, b_1 | e^{-i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} | \mathbf{k}, b_2 \rangle = \sum_{\mathbf{G}'=1} u_{\mathbf{k}-\mathbf{q}b_1}^\dagger(\mathbf{G}') u_{\mathbf{k}b_2}(\mathbf{G} + \mathbf{G}') = \\ &= \int u_{\mathbf{k}-\mathbf{q}b_1}^\dagger(\mathbf{r}) e^{-i(\mathbf{G}-\mathbf{G}_0) \cdot \mathbf{r}} u_{\mathbf{k}b_2}(\mathbf{r}) d\mathbf{r} = \hat{\mathcal{F}}[u_{\mathbf{k}-\mathbf{q}b_1}^\dagger u_{\mathbf{k}b_2}](\mathbf{G} - \mathbf{G}_0). \end{aligned} \quad (5.26)$$

The point $\mathbf{k}$ is always supposed to be within the first Brillouin zone (BZ), while $\mathbf{k} - \mathbf{q} = \overline{\mathbf{k} - \mathbf{q}} + \mathbf{G}_0$ with $\overline{\mathbf{k} - \mathbf{q}}$ lying inside the BZ. $\mathbf{G}_0$ represents an appropriate, possibly null, reciprocal lattice vector depending on the external $\mathbf{q}$ -point. Finally the operator $\hat{\mathcal{F}}$ denotes a Fourier transform from real to reciprocal space. To avoid expensive convolutions in reciprocal space, the oscillator matrix elements in equation 5.26 are evaluated by performing a Fast Fourier Transform (FFT) of the real space product $u_{\overline{\mathbf{k}-\mathbf{q}b_1}}(\mathbf{r}) u_{\mathbf{k}b_2}(\mathbf{r})$. The wave function product is reconstructed starting from the symmetrical images of the two wave functions in the irreducible Brillouin zone (IBZ), by means of the symmetry relations reported in Eq. B.6 and Eq. B.7.

One point worth noting is that equation 5.25 requires the explicit calculation of the contribution due to both resonant and anti-resonant transitions between valence and conduction states. When time-reversal is assumed, the expression can be further simplified. Using the symmetry transformations of the wave functions reported in equation 5.19, it is possible to show that the oscillator matrix elements satisfy the relationship:

$$M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) = \langle \mathbf{k} - \mathbf{q}, b_1 | e^{-i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} | \mathbf{k}, b_2 \rangle = M_{\mathbf{G}}^{b_2 b_1}(-\mathbf{k} + \mathbf{q}, \mathbf{q}), \quad (5.27)$$

where the two band indices have been interchanged and an appropriate change of the crystalline momentum of the left and right wave functions has been done. Using Eq. 5.27, the sum over the two band indices in Eq. 5.25 can be rearranged so that the double sum over all states reduces to a simpler and faster summation only over the resonant transitions connecting occupied and empty states, plus an additional contribution that has to be considered only in the case of metallic occupations. The final expression, as implemented in the code, reads:

$$\begin{aligned} \chi^0_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega) = & \frac{2}{V} \sum_{\mathbf{k}}^{\text{BZ}} \sum_{b_1 > b_2} M_{\mathbf{G}_1}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \left[M_{\mathbf{G}_2}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \right]^\dagger \times \left(f(\epsilon_{b_1\mathbf{k}-\mathbf{q}}) - f(\epsilon_{b_2\mathbf{k}}) \right) \\ & \left\{ \frac{1}{\omega + \epsilon_{b_1\mathbf{k}-\mathbf{q}} - \epsilon_{b_2\mathbf{k}} - i\eta \text{sign}(\epsilon_{b_1\mathbf{k}-\mathbf{q}} - \epsilon_{b_2\mathbf{k}})} - \frac{1}{\omega - \epsilon_{b_1\mathbf{k}-\mathbf{q}} + \epsilon_{b_2\mathbf{k}} + i\eta \text{sign}(\epsilon_{b_1\mathbf{k}-\mathbf{q}} - \epsilon_{b_2\mathbf{k}})} \right\} + \\ & + \frac{2}{V} \sum_{\mathbf{k}}^{\text{BZ}} \sum_b \left(f(\epsilon_{b\mathbf{k}-\mathbf{q}}) - f(\epsilon_{b\mathbf{k}}) \right) \frac{M_{\mathbf{G}_1}^{bb}(\mathbf{k}, \mathbf{q}) \left[M_{\mathbf{G}_2}^{bb}(\mathbf{k}, \mathbf{q}) \right]^\dagger}{\omega + \epsilon_{b\mathbf{k}-\mathbf{q}} - \epsilon_{b\mathbf{k}} - i\eta \text{sign}(\epsilon_{b\mathbf{k}-\mathbf{q}} - \epsilon_{b\mathbf{k}})}. \quad (5.28) \end{aligned}$$

In the case of semiconductors or insulators, only the transitions between valence and conduction states contribute and, due to the presence of an energy gap in the excitation spectrum, equation 5.28 can be recast in the simpler form:

$$\begin{aligned} \chi^0_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega) = & \frac{2}{V} \sum_{\mathbf{k}}^{\text{BZ}} \sum_{c\nu} M_{\mathbf{G}_1}^{c\nu}(\mathbf{k}, \mathbf{q}) \left[M_{\mathbf{G}_2}^{c\nu}(\mathbf{k}, \mathbf{q}) \right]^\dagger \times \\ & \times \left\{ \frac{1}{\omega + \epsilon_{c\mathbf{k}-\mathbf{q}} - \epsilon_{v\mathbf{k}} - i\eta} - \frac{1}{\omega - \epsilon_{c\mathbf{k}-\mathbf{q}} + \epsilon_{v\mathbf{k}} + i\eta} \right\}, \quad (5.29) \end{aligned}$$

where ν and c run over the set of valence and conduction states, respectively. The new expression obtained after having rearranged the transitions leads to a speed up of a factor two with respect to the initial expression given in equation 5.25. Furthermore, it represents the starting point for the implementation of the parallel and scalable algorithm described in the next paragraph.

Few changes are required in order to generalize the formalism to systems with collinear magnetism. In this case, the irreducible polarizability (neglecting the change of the Fermi level in case of metals) can be expressed in terms of a 2×2 diagonal matrix:

$$\underline{\tilde{\chi}} = \begin{pmatrix} \tilde{\chi}^{\uparrow\uparrow} & 0 \\ 0 & \tilde{\chi}^{\downarrow\downarrow} \end{pmatrix} \quad (5.30)$$

where each component describes the linear response of one of the spin-components of the electron density, due to a variation in the total potential acting on that particular spin. Only transitions between occupied and empty electronic states with same spin have to be considered for the calculation of the two components of $\underline{\tilde{\chi}}$. Note, however, that the

screened interaction $W(\omega)$ felt by an electron does not depend on the spin as only the total polarizability, *i.e.* the trace of $\tilde{\chi}$, enters the definition of the dielectric matrix and therefore of $W(\omega)$.

5.3 Scaling of the RPA computation in Fourier space

This section briefly describes the different steps required to compute $\chi^0(\omega)$ in a reciprocal space representation, and gives a detailed analysis of the scaling of the computational effort. For simplicity, the discussion is limited to the case of a system with an energy gap in which the simpler expression given in equation 5.29 can be used. We shall use $N_{\mathbf{G}}$ to indicate the number of $\mathbf{G}$ vectors used to describe the two-point function $\chi^0(\omega)$, with N_{FFT} and N_{Ψ} the number of FFT points and the number of plane waves employed for the description of the wave functions, respectively. Let moreover N_{ν} and N_c indicate the number of valence and conduction electrons respectively, and $N_{\mathbf{k}}$ and $N_{\mathbf{q}}$ the number of $\mathbf{k}$ - and $\mathbf{q}$ -points sampling the full Brillouin zone.

The evaluation of the matrix elements $M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q})$ can be efficiently performed using the Fast-Fourier-Transform algorithm [86] with a scaling proportional to $N_{\text{FFT}} \ln(N_{\text{FFT}})$ whereas the direct evaluation of the convolution would scale as $N_{\mathbf{G}} N_{\Psi}$. To avoid aliasing errors [86], the FFT box should ideally enclose a sphere in reciprocal space whose radius is equal to the radius of the sphere for the wave functions plus the radius of the $\mathbf{G}$ -sphere where the oscillator matrix elements are calculated.¹ Equation 5.29 has to be evaluated for $N_{\mathbf{G}}^2$ pairs, each of them requiring an electron-hole summation for each frequency. The frequency dependence in the denominator cannot be factorized out of the sum over the transitions, therefore a separate accumulation for each ω is needed. Neglecting prefactors, the overall scaling of equation 5.29 is hence given by

$$N_{\mathbf{q}} \times \left[N_{\omega} N_{\mathbf{k}} (N_{\nu} N_c N_{\mathbf{G}}^2) + N_{\mathbf{k}} (N_{\nu} N_c) N_{\text{FFT}} \ln(N_{\text{FFT}}) \right], \quad (5.31)$$

with the first term dominant, especially if a large number of ω points is required. Equation 5.29 can be evaluated one $\mathbf{q}$ -point at a time, and the final results are immediately stored on an external file. Except for the storage of the wave functions and of the $\chi^0_{\mathbf{G}_1 \mathbf{G}_2}(\omega)$ sub-matrix, no additional large portion of memory has to be reserved during this step. The matrix elements $M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q})$ are calculated for each resonant transition, stored in a local buffer and their partial contribution to Eq. 5.29 is immediately accumulated. An important saving in computational effort comes from the truncation of the response matrix. In most of the cases, the off-diagonal elements of the response function can be assumed to be zero for $\mathbf{G}$ -vectors whose kinetic energy is much smaller than the plane wave cutoff used for the wave functions.

¹ Aliasing effects are important whenever the number of vectors $N_{\mathbf{G}}$ in the oscillators $M_{\mathbf{G}}$ is comparable to that used for the wave functions N_{Ψ} . This usually happens during the evaluation of the exchange part of the self-energy, whereas aliasing should be less severe for the case of $\chi^0(\omega)$, as the basis set needed to describe $\chi^0(\omega)$ is usually much smaller than the one required for the wave functions. According to our experience, aliasing effects are small once the plane wave basis set for the wave functions is sufficiently converged, and we usually opt for a minimal FFT grid after having performed initial convergence tests.

Two different steps are required to evaluate $\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega)$ in Fourier space representation: two matrix multiplications to build $\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega)$ (Eq. 5.5) and a matrix inversion to get $\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega)$. The matrices are dense, so the matrix multiplications cost $N_{\mathbf{G}}^3$ operations, with matrix inversion having a similar scaling. The computational effort of this part of the algorithm is thus given by:

$$N_{\mathbf{q}} \times (N_{\omega} N_{\mathbf{G}}^3). \quad (5.32)$$

Since the matrix elements of the irreducible polarizability converge slowly with respect to the number of empty states, the leading term defining the overall scaling for a single $\mathbf{q}$ -point is $N_{\omega} \times (N_{\mathbf{k}} N_{\nu} N_c N_{\mathbf{G}}^2)$, and we arrive at the conclusion that the time effort scales with the fourth power of the number of atoms.

If enough processing units are available, the factor $N_{\mathbf{q}}$ in Eq. 5.31 can be trivially absorbed by splitting the calculation of the $\mathbf{q}$ -points into several independent calculations. The results can then be gathered in a unique binary file by means of merging utilities. Unfortunately this poor man's parallelism over $\mathbf{q}$ -points is of no help in reducing the memory requirements. The total storage for all the occupied and unoccupied wave functions plus the whole complex matrix is indeed of the order of

$$(N_{\nu} + N_c)(N_{\Psi} + N_{\text{FFT}}) + 2N_{\mathbf{G}}^2, \quad (5.33)$$

if we assume that the wave functions in real space are also stored in memory to speed up the evaluation of the oscillator matrix elements. Many-body calculations hence have memory requirements much larger than those needed for standard ground-state applications, and an optimized implementation based on a parallel environment supporting the MPI protocol is needed. The algorithm greatly benefits from a coarse grained parallelism in which each processor stores the full set of valence states while conduction states are divided across the nodes. This particular distribution, which strongly differs from the standard approach used in the ground-state part of the ABINIT code, allows both CPU and memory saving since each processor has to calculate only the contribution to $\chi^0(\omega)$ arising from a smaller subset of transitions. The final results are then gathered on each processing unit by means of collective operations, and then saved on disk. Communication is needed only at the beginning and at the end of the calculation for a given $\mathbf{q}$. This algorithm presents good scalability both concerning CPU time and memory requirements, since an increasing number of conduction bands can be compensated for by just using more processors.

The remaining bottleneck is brought on by the storage of the occupied states and of the two-point function in reciprocal space, the only two parts of the present implementation that do not scale with the number of processing units. The first problem can be partly resolved, at the price of worsening performance, by opting for a less memory demanding approach in which only the wavefunctions in reciprocal space are stored, and an FFT from $\mathbf{G}$ to $\mathbf{r}$ is performed whenever a particular wave function is needed. The second limitation is more difficult to solve since it requires the distribution of the matrix $\chi^0_{\mathbf{G}_1\mathbf{G}_2}(\omega)$ on the different nodes using a ScaLapack grid [87, 88]. The matrix distribution should be done in conjunction with a parallelization of the FFT algorithm such

that also the bottleneck associated to the storage of the occupied states can be removed. Overcoming this limitation is of crucial importance for large-scale applications of the code and will be the subject of further developments.

The present implementation is thus best suited for crystalline compounds of medium size. Isolated or disordered systems can be simulated by means of the supercell technique [89], even though other approaches based on the real space imaginary time formulation [90, 91] or the mixed-space representation proposed in [92] will have a much better scaling under these particular conditions. Since, in the present work, we will mainly be concerned with crystalline structures of high symmetry, it is of fundamental importance to take advantage of any possible symmetry properties in order to reduce the computational effort as much as possible, in particular the scaling factors associated to the BZ sampling ($N_{\mathbf{q}}$ and $N_{\mathbf{k}}$). This topic is the subject of the next section.

5.4 Symmetries of the irreducible polarizability

This section describes how to take advantage of the constraints introduced by the time-reversal invariance and by the space group of the crystal in order to reduce the number of independent $\mathbf{k}$ - and $\mathbf{q}$ -points to be explicitly considered. First we discuss how to reduce the number of independent $\mathbf{q}$ -points needed for the description of $\chi^0(\omega)$. In section 5.4.2, the discussion is specialized to the case of magnetic space groups. In section 5.4.3, we then show how to reduce the number of $\mathbf{k}$ -points that have to be explicitly summed over in order to calculate $\chi^0(\omega)$.

5.4.1 Symmetries in $\mathbf{q}$ -space

So far, only the invariance of the two-point functions under an arbitrary translation of the Bravais lattice has been employed to expand $\chi^0(\mathbf{r}_1 \mathbf{r}_2; \omega)$ in a discrete Fourier series. The invariance of the crystal under the application of the symmetries of the space group introduces additional constraints on the Fourier representation of the two-point functions, constraints that can be exploited to reduce the number of non-equivalent $\mathbf{q}$ -points to be computed [82]. Since $\chi^0(\mathbf{r}_1 \mathbf{r}_2)$, or any other response function, has to be invariant under the application of a space group symmetry acting on $\mathbf{r}_1$ and $\mathbf{r}_2$, one obtains²

$$\begin{aligned} \chi^0(\mathbf{r}_1, \mathbf{r}_2) &= \frac{1}{V} \sum_{\mathbf{G}_1 \mathbf{G}_2}^{\mathbf{q}} e^{i(\mathbf{q} + \mathbf{G}_1) \cdot \mathbf{r}_1} \chi^0_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}) e^{-i(\mathbf{q} + \mathbf{G}_2) \cdot \mathbf{r}_2} = \hat{\mathcal{R}}_{\mathbf{t}} \left[\chi^0(\mathbf{r}_1, \mathbf{r}_2) \right] = \\ &= \chi^0(\mathcal{R}^{-1}(\mathbf{r}_1 - \mathbf{t}), \mathcal{R}^{-1}(\mathbf{r}_2 - \mathbf{t})) = \frac{1}{V} \sum_{\mathbf{G}_1 \mathbf{G}_2}^{\mathbf{q}} e^{i\mathcal{R}(\mathbf{q} + \mathbf{G}_1) \cdot (\mathbf{r}_1 - \mathbf{t})} \chi^0_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}) e^{-i\mathcal{R}(\mathbf{q} + \mathbf{G}_2) \cdot (\mathbf{r}_2 - \mathbf{t})}, \end{aligned} \quad (5.34)$$

² The conventions and notations used in this work to define the symmetries of the space group are discussed in section B. The frequency dependence is omitted, since all equations hold independently of ω .

from which, equating the Fourier component term by term, one arrives at:

$$\chi^0_{\mathbf{G}_1\mathbf{G}_2}(\mathcal{R}\mathbf{q} + \mathbf{G}_{\mathcal{R}}) = e^{i\mathbf{t}\cdot(\mathbf{G}_2-\mathbf{G}_1)} \chi^0_{\mathcal{R}^{-1}(\mathbf{G}_1+\mathbf{G}_{\mathcal{R}})} \mathcal{R}^{-1}(\mathbf{G}_2+\mathbf{G}_{\mathcal{R}})(\mathbf{q}). \quad (5.35)$$

In the above expressions $\mathcal{R}$ and $\mathbf{t}$ denote the rotational part and the fractional translation associated to the symmetry operation of the space group, respectively. A possibly null *umklapp* vector $\mathbf{G}_{\mathcal{R}}$ has been introduced such that $\mathcal{R}\mathbf{q} + \mathbf{G}_{\mathcal{R}}$ always lies within the first Brillouin zone. Equation 5.35 can be used to reconstruct the Fourier representation of any two-point function with the same symmetry of the crystal, starting from the irreducible wedge defined by the point group of the crystal. At high-symmetry $\mathbf{q}$ -points for which $\mathcal{R}\mathbf{q} = \mathbf{q}$, equation 5.35 provides a further relation among different elements of the same matrix. This property can be used to reduce the number of matrix elements which need to be calculated, or to optimize algorithms such as the Hilbert transform method described in section 5.6. In systems without an inversion center, time-reversal can be implemented according to equation 5.21.

5.4.2 Magnetic Symmetries

The symmetry properties described in the previous sections can be generalized to include crystals with a magnetic space group.³ In this case each atom in the magnetic unit cell, which is different from the chemical unit cell, is assigned a particular direction of the local magnetization, arbitrarily assumed to be along the z quantization axis. Each symmetry operation of the magnetic space group is characterized by an additional part defining whether the magnetization of the rotated atom is preserved in direction and amplitude or if it has to be reversed. The first set of operations will be said to have ferromagnetic character (FM), and a generic operation will be indicated with the symbol $\hat{\mathcal{R}}_{\mathbf{t}}$. The second ensemble will be said to have anti-ferromagnetic character (AFM), and the corresponding operations will be denoted with an underlined symbol, *i.e.* $\underline{\hat{\mathcal{R}}}_{\mathbf{t}}$. The action of a magnetic operation onto a wavefunction is given by

$$\hat{\mathcal{R}}_{\mathbf{t}} \Psi_{\mathbf{k}}^{\uparrow}(\mathbf{r}) \equiv \Psi_{\mathbf{k}}^{\uparrow}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = \Psi_{\mathcal{R}\mathbf{k}}^{\uparrow}(\mathbf{r}), \quad (5.36)$$

$$\underline{\hat{\mathcal{R}}}_{\mathbf{t}} \Psi_{\mathbf{k}}^{\uparrow}(\mathbf{r}) \equiv \Psi_{\mathbf{k}}^{\downarrow}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = \Psi_{\mathcal{R}\mathbf{k}}^{\downarrow}(\mathbf{r}). \quad (5.37)$$

The above symmetry relations, in particular equation 5.37, have several important consequences for the implementation, since only the $\mathbf{k}$ -points in the magnetic irreducible Brillouin zone have to be explicitly calculated. It is important to note that a ferromagnetic space group does not contain operations reversing the local magnetization, therefore it is not possible to relate wave functions of opposite spin by means of equation 5.37. In a ferromagnetic crystal, the two spin polarizations are completely independent, the ferromagnetic IBZ has the same number of irreducible points as the non magnetic one, hence no further reduction is possible. Only anti-ferromagnetic space groups allow a considerable saving both in computing time and memory, as the knowledge of a single spin

³ A brief introduction to magnetic groups is reported in section B.2 where the notations used in this work are also introduced. For simplicity, the discussion is limited to the case of collinear magnetism.

polarization, say the up component, is sufficient to completely describe the electronic properties. The application of an anti-ferromagnetic operation onto the Kohn-Sham density leads to the two following fundamental relations:

$$\hat{\mathcal{R}}_{\mathbf{t}} n^{\uparrow}(\mathbf{r}) = n^{\uparrow}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = n^{\uparrow}(\mathbf{r}), \quad (5.38)$$

$$\hat{\mathcal{R}}_{\mathbf{t}} n^{\downarrow}(\mathbf{r}) = n^{\downarrow}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = n^{\downarrow}(\mathbf{r}). \quad (5.39)$$

As expected, the total density $n^{\uparrow}(\mathbf{r}) + n^{\downarrow}(\mathbf{r})$ and consequently also the Hartree potential, are invariant under the application of any symmetry operation belonging to the magnetic space group. Similar equations also hold for the two spin components of the exchange-correlation potential. By exploiting equation 5.39, it is possible to derive the effect of an anti-ferromagnetic symmetry applied to the up-up channel of the irreducible polarizability. The final result reads:

$$\begin{aligned} \hat{\mathcal{R}}_{\mathbf{t}} \tilde{\chi}^{\uparrow\uparrow}(\mathbf{r}_1, \mathbf{r}_2) &= \tilde{\chi}^{\uparrow\uparrow}(\mathcal{R}^{-1}(\mathbf{r}_1 - \mathbf{t}), \mathcal{R}^{-1}(\mathbf{r}_2 - \mathbf{t})) = \\ &= \frac{\delta \left[\hat{\mathcal{R}}_{\mathbf{t}} n^{\uparrow} \right](\mathbf{r}_1)}{\delta \left[\hat{\mathcal{R}}_{\mathbf{t}} (v_{\text{tot}}^{\uparrow}) \right](\mathbf{r}_2)} = \frac{\delta n^{\downarrow}(\mathbf{r}_1)}{\delta v_{\text{tot}}^{\downarrow}(\mathbf{r}_2)} = \tilde{\chi}^{\downarrow\downarrow}(\mathbf{r}_1, \mathbf{r}_2), \end{aligned} \quad (5.40)$$

where we made the assumption that also the external perturbing potential transforms in the same way as the density under the application of an anti-ferromagnetic symmetry.

In reciprocal space, equation 5.40 leads to the important symmetry relation (*umklapp* processes are neglected for simplicity)

$$\tilde{\chi}_{\mathcal{R}\mathbf{G}_1\mathcal{R}\mathbf{G}_2}^{\downarrow\downarrow}(\mathcal{R}\mathbf{q}) = e^{i\mathcal{R}(\mathbf{G}_2 - \mathbf{G}_1) \cdot \mathbf{t}} \tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}^{\uparrow\uparrow}(\mathbf{q}), \quad (5.41)$$

where the horizontal bar is used to explicitly indicate that the fractional translation $\mathbf{t}$ and the rotational part $\mathcal{R}$ belong to a symmetry with anti-ferromagnetic character. Equation 5.41 establishes a connection between the matrix elements of the up-up and down-down component at different $\mathbf{q}$ -points related by the rotational part of an anti-ferromagnetic operation. The equation can be further simplified in case of the so-called magnetic space groups of type IV in which the set of anti-ferromagnetic symmetries can be obtained by simply composing the ferromagnetic subgroup with the generator constituted by a non-primitive fractional translation. Consequently, all symmetries belonging to a type IV magnetic group can be divided into pairs of ferromagnetic and anti-ferromagnetic operations having the same rotational part but different fractional translations. Starting from equation 5.41, it readily follows that

$$\tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}^{\downarrow\downarrow}(\mathbf{q}) = e^{i\mathcal{R}(\mathbf{G}_1 - \mathbf{G}_2) \cdot (\mathbf{t} - \underline{\mathbf{t}})} \tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}^{\uparrow\uparrow}(\mathbf{q}), \quad (5.42)$$

for each AFM-FM pair of symmetries with the same rotational part $\mathcal{R}$. This is quite a useful result because it implies that the down-down channel can be reconstructed by symmetry starting from the knowledge of the up-up component at the same $\mathbf{q}$ -point.

Equation 5.42 allows one to halve both the computational effort and the memory requirement since the calculation of the wavefunctions and oscillator matrix elements can be restricted to a single spin component. It is also worth noting that equation 5.42 has to hold for each AFM-FM pair of symmetries. This imposes a strict constraint on the matrix elements of χ^0 in reciprocal space as $\tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})$ has to be zero for each pair of $\mathbf{G}$ -vectors for which the non-symmorphic phase factor $e^{i\mathcal{R}(\mathbf{G}_1-\mathbf{G}_2)\cdot(\mathbf{t}-\mathbf{t})}$ is not constant.

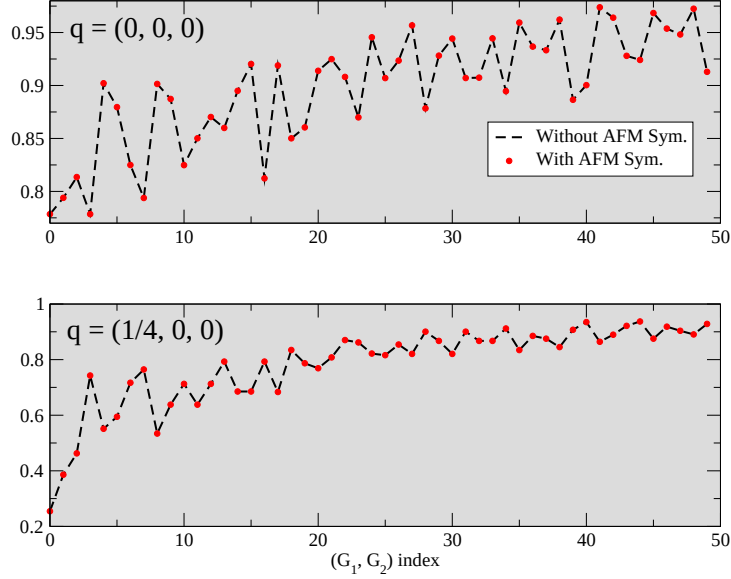


Figure 5.2: Comparison between selected matrix elements of $\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega = 0)$ of NiO in the type IV magnetic group, calculated with and without anti-ferromagnetic symmetries. The results obtained without taking advantage of anti-ferromagnetic symmetries (dashed line) have been calculated performing an independent calculation for each spin polarization. Red dots represent the results obtained employing the anti-ferromagnetic space group to reconstruct the down-down component of the polarizability by means of equation 5.42.

Figure 5.2 displays a comparison between selected matrix elements of the inverse dielectric matrix of NiO calculated with and without taking advantage of anti-ferromagnetic symmetries. On the scale of the plots, no significant difference can be noticed, the two curves lie almost on top of each other, confirming the validity of equation 5.42.

5.4.3 Symmetries in k-space

In the previous sections, it was shown how to reconstruct the irreducible polarizability at an arbitrary $\mathbf{q}$ -point in the full Brillouin zone, starting from the irreducible wedge defined by the full point group of the crystal. In this section, we shall discuss how to reduce by symmetry the number of $\mathbf{k}$ -points that have to be explicitly considered during

the evaluation of the integral over the BZ given in equation 5.28. It is anticipated that in this case, unlike the symmetrization in $\mathbf{q}$ -space of $\chi^0_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega)$, the computational gain depends on the external point $\mathbf{q}$. The set of independent $\mathbf{k}$ -points is indeed defined by the operations of the so-called little group of $\mathbf{q}$, namely the symmetry operations of the space group whose rotational part leaves $\mathbf{q}$ unchanged. In section B.3, it is shown that, with $\mathcal{R}$ being a proper or improper rotation belonging to the little group $\mathcal{L}_{\mathbf{q}}^+$ of $\mathbf{q}$, *i.e.* $\mathcal{R}\mathbf{q} = \mathbf{q}$, the matrix elements $M_{\mathbf{G}_1}^{b_1b_2}(\mathbf{k}, \mathbf{q})$ at different $\mathbf{k}$ -points satisfy the relation

$$M_{\mathbf{G}}^{b_1b_2}(\mathcal{R}\mathbf{k}, \mathbf{q}) = e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} M_{\mathcal{R}^{-1}\mathbf{G}}^{b_1b_2}(\mathbf{k}, \mathbf{q}), \quad \text{if } \mathcal{R}\mathbf{q} = \mathbf{q}. \quad (5.43)$$

If the system does not have an inversion center, time-reversal can be exploited

$$M_{\mathbf{G}}^{b_1b_2}(-\mathcal{R}\mathbf{k}, \mathbf{q}) = e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} [M_{-\mathcal{R}^{-1}\mathbf{G}}^{b_1b_2}(\mathbf{k}, \mathbf{q})]^\dagger, \quad \text{if } -\mathcal{R}\mathbf{q} = \mathbf{q}. \quad (5.44)$$

Hereafter, $\mathcal{L}_{\mathbf{q}}^-$ will denote the set of space group operations whose rotational part requires a composition with the time-reversal operator in reciprocal space (a simple inversion) to preserve the external $\mathbf{q}$. Furthermore, let $IBZ_{\mathbf{q}}$ indicate the irreducible wedge associated to the set of operations in the union $\mathcal{L}_{\mathbf{q}}^+ \cup \mathcal{L}_{\mathbf{q}}^-$. Using Eq. 5.43 and Eq. 5.44, it is possible to restrict the sum over the full Brillouin zone present in equation 5.29 to a faster summation over the $\mathbf{k}$ -points in $IBZ_{\mathbf{q}}$:

$$\begin{aligned} \chi^0_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega) = & \frac{2}{V} \sum_{\nu} \sum_{\mathbf{k}}^{IBZ_{\mathbf{q}}} \left(\frac{1}{\omega + \epsilon_{\nu\mathbf{k}-\mathbf{q}} - \epsilon_{\mathbf{c}\mathbf{k}} + i\eta} - \frac{1}{\omega - \epsilon_{\nu\mathbf{k}-\mathbf{q}} + \epsilon_{\mathbf{c}\mathbf{k}} - i\eta} \right) \times \\ & \left\{ \sum_{\hat{\mathcal{R}}_{\mathbf{t}}}^{\mathcal{L}_{\mathbf{q}}^+} \tilde{w}^+(\hat{\mathcal{R}}_{\mathbf{t}}, \mathbf{k}) e^{-i(\mathbf{G}_1-\mathbf{G}_2)\cdot\mathbf{t}} M_{\mathcal{R}^{-1}\mathbf{G}_1}^{\nu c}(\mathbf{k}, \mathbf{q}) \left[M_{\mathcal{R}^{-1}\mathbf{G}_2}^{\nu c}(\mathbf{k}, \mathbf{q}) \right]^\dagger + \right. \\ & \left. + \sum_{\hat{\mathcal{R}}_{\mathbf{t}}}^{\mathcal{L}_{\mathbf{q}}^-} \tilde{w}^-(\hat{\mathcal{R}}_{\mathbf{t}}, \mathbf{k}) e^{-i(\mathbf{G}_1-\mathbf{G}_2)\cdot\mathbf{t}} \left[M_{-\mathcal{R}^{-1}\mathbf{G}_1}^{\nu c}(\mathbf{k}, \mathbf{q}) \right]^\dagger M_{-\mathcal{R}^{-1}\mathbf{G}_2}^{\nu c}(\mathbf{k}, \mathbf{q}) \right\}, \quad (5.45) \end{aligned}$$

where the weights $\tilde{w}^+(\hat{\mathcal{R}}_{\mathbf{t}}, \mathbf{k})$ and $\tilde{w}^-(\hat{\mathcal{R}}_{\mathbf{t}}, \mathbf{k})$ are normalized to 1 and defined such that each point in the star of $\mathbf{k}$ is appropriately reconstructed. The use of symmetries considerably reduces the computational demand in case of $\mathbf{q}$ -points located on high symmetry planes or along high-symmetry directions. For instance, 256 $\mathbf{k}$ -points in the full zincblende Brillouin zone are reduced to 40 for $\mathbf{q} = X$. At the gamma point, the entire set of symmetries can be utilized and only 16 $\mathbf{k}$ -points have to be considered. For $\mathbf{q}$ -points on the surface of the Brillouin zone, an *umklapp* vector, $\mathbf{G}_{\mathcal{R}}$, might be required to bring the rotated point back to the first Brillouin zone, *i.e.* $\mathcal{R}\mathbf{q} = \mathbf{q} + \mathbf{G}_{\mathcal{R}}$. Since *umklapp* processes are important only for $\mathbf{q}$ -points on the Brillouin zone boundary, this particular eventuality has not been taken into account during the symmetrization in order to keep the notation as simple as possible. The generalization of equation 5.45 to account for a possible *umklapp* process is, however, straightforward to implement, and the interested reader is referred to section B.3 where the required equations are reported.

One subtlety we wish to point out is that the symmetrized equation 5.45 has been obtained assuming that both the wave functions entering the expression for the oscillator matrix elements are separately nondegenerate. In the case of degenerate states the action of the symmetry operation in the degenerate subspace should in general be included when reconstructing wave functions at $\mathbf{k}$ -points related by symmetry (equation B.8 in appendix B).

However, in the expression for the irreducible polarizability given in 5.45, valence-, conduction-, and therefore any degenerate set of states are summed over. Since the linear transformation $D_{\alpha\beta}(\mathcal{R})$ in equation B.8 can be assumed to be unitary, the states enter in complex conjugate products and all the states in the degenerate complex are summed, the matrices can be factorized and they drop out of the final results by virtue of the orthogonality theorem of group theory [93]. For this reason, there is no need to take into account how each symmetry property acts within the degenerate subspaces, and the simplified symmetrized expression given in equation B.14 can be safely used. We will return to this point in section 5.7.1.1 during the discussion of the symmetrization of the matrix elements of the self-energy operator. It is anticipated that, due to the mathematical structure of $\hat{\Sigma}$, the same “cancellation” property does not hold anymore. Hence special care has to be taken in order to assure that the quasiparticle energies present the correct energy degeneracy as required by the space group of the crystal.

5.5 Plasmon-Pole Models

One of the major computational efforts in self-energy calculations is represented by the calculation of the frequency dependence of the screened interaction, which is needed for the evaluation of the convolution given in equation 5.1. Due to the ragged behavior of $G(\omega)$ and $W(\omega)$ along the real axis, several real frequencies are in principle required to converge the results. On the other hand, since the fine details of $W(\omega)$ are integrated over, it is reasonable to expect that approximated models, able to capture the main physical features of the screened interaction, should give sufficiently accurate results with a considerably reduced computational effort. This is the basic idea behind the so-called Plasmon-Pole Model (PPM) in which the frequency dependence of $W(\omega)$ is modeled in terms of analytic expressions depending on coefficients with a direct physical interpretation. The coefficients of the model are derived from first principles, *i.e.* without any adjustable external parameters, either by enforcing exact relations or by making sure that *ab initio* calculated quantities are correctly reproduced (for example the static limit $W(\omega = 0)$).

Four different plasmon-pole techniques are available in the ABINIT code, each of them has its own advantages and drawbacks. The aim of the next two sections is to present the mathematical structure of the different models as well as the basic equations employed to calculate the plasmon-pole coefficients. We shall start by discussing two famous and widespread approximations that have been successfully used in the first pioneering implementations of the *GW* formalism: the plasmon-pole model of Godby-Needs [94] and the model proposed by Hybertsen and Louie [56]. Then we present two

more recent approaches due to Von der Linden Horsch [95] and Engel Farid [96] which are computationally more efficient and expected to be more accurate. A final summary of the pros and cons of the plasmon-pole method concludes the discussion.

5.5.1 Godby-Needs and Hibertsen-Louie Plasmon-Pole Models

Experiments and first-principles analysis reveal that the imaginary part of $W(\omega)$ is generally characterized by a sharp peak in correspondence to a plasmon excitation at the plasmon frequency. For cases where there is not a well-defined peak, direct calculations show that the imaginary part tends to be small and fluctuates in sign as a function of ω . A typical example is the homogeneous electron gas where $\text{Im}W(\mathbf{q}, \omega)$, in the limit $\mathbf{q} \rightarrow 0$, is characterized by a sharp plasmon excitation. For finite $\mathbf{q}$, however, the spectrum of the homogeneous gas contains also contributions due to electron-hole excitations located at lower energies which eventually merge with the plasmon excitations as $\mathbf{q}$ gets larger. As a first approximation, it is therefore reasonable to neglect any possible electron-hole contribution, assuming that all the weight in $\text{Im}W(\mathbf{q}, \omega)$ resides in the plasmon excitation. In its simplest form, the PPM approximation is given by [61]:

$$\text{Im} \epsilon^{-1}(\mathbf{q}, \omega) = A(\mathbf{q}) \left[\delta(\omega - \omega(\mathbf{q})) - \delta(\omega + \omega(\mathbf{q})) \right], \quad (5.46)$$

where $A(\mathbf{q})$ and $\omega(\mathbf{q})$ define the amplitude and the position of the peak, respectively. In crystalline systems, the model has to be generalized to take into account the periodicity of the system and local field effects. In the generalized plasmon-pole approximation proposed in [56], each matrix component of the inverse dielectric matrix acquires an additional dependence on the $(\mathbf{G}_1, \mathbf{G}_2)$ pair of vectors of the reciprocal lattice:

$$\text{Im} \epsilon_{\mathbf{G}_1 \mathbf{G}_2}^{-1}(\mathbf{q}, \omega) = A_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}) \left[\delta(\omega - \omega_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q})) - \delta(\omega + \omega_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q})) \right]. \quad (5.47)$$

The corresponding real part is then obtained by means of the Kramers-Kronig relation [97] and the final result reads⁴

$$\text{Re} \epsilon_{\mathbf{G}_1 \mathbf{G}_2}^{-1}(\mathbf{q}, \omega) = \delta_{\mathbf{G}_1 \mathbf{G}_2} + \frac{\Omega_{\mathbf{G}_1 \mathbf{G}_2}^2(\mathbf{q})}{\omega^2 - \tilde{\omega}_{\mathbf{G}_1 \mathbf{G}_2}^2(\mathbf{q})}. \quad (5.48)$$

At this point, we only need to define a set of physical constraints that can be employed to find the parameters without ambiguity. Unfortunately, there is no unique way to define the plasmon parameters. Hence, the final equations for the plasmon coefficients will depend on the particular physical properties or on the exact relations one aims to reproduce.

In the original implementation of the *GW* part of ABINIT, the plasmon-pole parameters were derived such that the inverse dielectric matrix is correctly reproduced at two

⁴ Here we work with the causal version of $\epsilon^{-1}(\omega)$ such that the Kramers-Kronig relation, relating cause and effect, can be properly used. The time-ordered expression is then obtained from the retarded version by just appropriately shifting the poles in the complex plane.

different frequencies: the static limit and an additional imaginary point located at $i\omega_P$, where $\omega_P = \sqrt{4\pi\rho}$ is the plasma frequency of the system (according to the Sommerfeld model [98]) and ρ is the number of electrons per volume. After some algebra, the following set of equations defining the plasmon-pole coefficients can be derived:

$$\begin{cases} A_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) = \epsilon_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega = 0) - \delta_{\mathbf{G}_1\mathbf{G}_2} = -\frac{\Omega_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q})}{\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q})} \\ \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}^2 = \omega_P^2 \left[\frac{A_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})}{\epsilon_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega = 0) - \epsilon_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, i\omega_P)} - 1 \right] \\ \Omega_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q}) = -A_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q}). \end{cases} \quad (5.49)$$

This particular approach for finding the plasmon-pole parameters dates back to the work of Godby and Needs [94]. In the following discussion, the set of expressions in equation 5.49 will be referred to as the GN plasmon-pole model. The method has proved to be accurate and reliable for several different applications, and is still the default model used for QP calculations in ABINIT. On the other hand, many of the plasmon-pole based calculations reported in the literature rely on the original model proposed by Hybertsen and Louie (HL) in their famous *GW* implementation [56]. For the sake of comparison with previously published results, we have thus decided to implement this approach as well. In this case, the plasmon-pole parameters are calculated so as to reproduce the static limit exactly and to fulfill a generalized *f*-sum rule relating the imaginary part of the many-body inverse dielectric matrix to the plasma frequency and the charge density [56, 99, 100]:

$$\int_0^{+\infty} \text{Im} \epsilon_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega) d\omega = \frac{\pi\omega_P^2}{2} \frac{(\mathbf{q} + \mathbf{G}_1) \cdot (\mathbf{q} + \mathbf{G}_2)}{|\mathbf{q} + \mathbf{G}_1|^2} \frac{n(\mathbf{G}_1 - \mathbf{G}_2)}{n(\mathbf{G} = 0)}. \quad (5.50)$$

The model gives a good description of both the low-energy behavior via the reproduction of the static limit and the high energy behavior via the reproduction of the first frequency moment. Unlike the GN method, only a single frequency – the static limit – has to be explicitly calculated. The final expressions for the parameters read:

$$\begin{cases} A_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) = -\frac{\pi}{2} \frac{\Omega_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q})}{\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q})} \\ \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q}) = \frac{\Omega_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q})}{\delta_{\mathbf{G}_1\mathbf{G}_2} - \epsilon_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega = 0)} \\ \Omega_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q}) = \omega_P^2 \frac{(\mathbf{q} + \mathbf{G}_1) \cdot (\mathbf{q} + \mathbf{G}_2)}{|\mathbf{q} + \mathbf{G}_1|^2} \frac{n(\mathbf{G}_1 - \mathbf{G}_2)}{n(\mathbf{G} = 0)}. \end{cases} \quad (5.51)$$

The plasmon-pole coefficients have to be evaluated only for $\mathbf{q}$ -points in the irreducible Brillouin zone. The effective bare plasma frequency $\Omega_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})$ transforms in the same way as the inverse dielectric matrix under the application of a symmetry operation:

$$\Omega_{\mathbf{G}_1\mathbf{G}_2}(\mathcal{R}\mathbf{q}) = e^{-i(\mathbf{G}_1-\mathbf{G}_2)\cdot\mathbf{t}} \Omega_{\mathcal{R}^{-1}\mathbf{G}_1\mathcal{R}^{-1}\mathbf{G}_2}(\mathbf{q}), \quad (5.52)$$

while the plasmon frequencies, as expected, are invariant under symmetry. In case of the Godby-Needs model, the symmetry relation 5.52 derives directly from the analogous symmetry properties of the inverse dielectric matrix via Eq. 5.49. In the HL PPM, equation 5.52 can be straightforwardly obtained, by starting from the definition of $\Omega_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})$ and using the following symmetry:

$$n(\mathbf{G}_1 - \mathbf{G}_2) = e^{-i(\mathbf{G}_1-\mathbf{G}_2)\cdot\mathbf{t}} n(\mathcal{R}^{-1}(\mathbf{G}_1 - \mathbf{G}_2)), \quad (5.53)$$

relating the Fourier components of the electronic density. The effect of time-reversal symmetry on $\Omega_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})$ can be derived starting from equation 5.23.

Models based on the two expressions 5.47 and 5.48 have a number of undesirable features, despite their success. The total number of poles is large since it equals the square of the number of plane waves used to represent $\epsilon_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega)$, so expensive tables of $N_{\mathbf{G}}^2$ elements have to be stored in memory during the entire calculation. Moreover, for some elements with $\mathbf{G}_1 \neq \mathbf{G}_2$, the plasmon poles $\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})$ can become very small or even imaginary which is somewhat unphysical. As discussed in [56], complex frequencies usually occur when the imaginary part does not have a well defined peak and oscillates in sign; therefore the contribution given by these term is expected to be small. Hybertsen and Louie suggested neglecting these unphysical results during the evaluation of the self-energy and this is also the approach followed in our implementation. Nevertheless, even in simple systems such as Si, the number of complex plasmon frequencies has been found to represent a significant part of the total number of poles, sometimes greater than half the total number. According to our experience, even the generalization of the HL PPM proposed in [101] for systems without an inversion center does not help reduce the occurrence of such unphysical poles. The GN method, on the other hand, seems to not be affected by this serious deficiency, even though the better stability is partly paid for by the fact that two independent calculations of $\chi^0(\omega)$ have to be performed, and the stability of the final quasiparticle corrections should be checked against the second imaginary frequency employed for the fit.

5.5.2 Von der Linden-Horsch and Engel-Farid Plasmon-Pole Models

A different model, proposed by von der Linden and Horsch, circumvents the problem of imaginary plasmon frequencies by using the dielectric-bandstructure approach [95]. A plasmon-pole model with $N_{\mathbf{G}}$ parameters is derived from the spectral decomposition of the symmetrized inverse dielectric matrix [95, 102]:

$$\tilde{\epsilon}^{-1}(\mathbf{q}, \omega) = \sum_p |\Phi_p(\mathbf{q}, \omega)\rangle \lambda_p^{-1}(\mathbf{q}, \omega) \langle \Phi_p(\mathbf{q}, \omega)|, \quad (5.54)$$

where p labels the real eigenvalues $\lambda_p(\mathbf{q}, \omega)$. The plasmon-pole model is obtained assuming that the frequency dependence is solely contained in its eigenvalues, and is of the form:

$$\lambda_p^{-1}(\mathbf{q}, \omega) = 1 + \frac{\zeta_p(\mathbf{q}) \omega_p(\mathbf{q})}{2} \left[\frac{1}{\omega - \omega_p(\mathbf{q}) + i\delta} - \frac{1}{\omega + \omega_p(\mathbf{q}) - i\delta} \right]. \quad (5.55)$$

The eigenvectors in equation 5.54, and the plasmon-pole strengths, $\zeta_p(\mathbf{q})$, are assumed to be frequency-independent, and are determined by comparison with the static limit $\tilde{\epsilon}^{-1}(\mathbf{q}, \omega = 0)$ calculated from first principles:

$$\begin{cases} |\Phi_p(\mathbf{q}, \omega)\rangle = |\Phi_p(\mathbf{q}, \omega = 0)\rangle \equiv |\Phi_p^0(\mathbf{q})\rangle \\ \zeta_p(\mathbf{q}) = 1 - \lambda_p^{-1}(\mathbf{q}, \omega = 0). \end{cases} \quad (5.56)$$

Finally, Johnson's f -sum rule [99] permits one to find the plasmon-pole frequencies:

$$\omega_p(\mathbf{q})^2 = \frac{\omega_P^2}{\zeta_p(\mathbf{q})} \sum_{\mathbf{G}_1 \mathbf{G}_2} \frac{[\Phi_p^0(\mathbf{q}, \mathbf{G}_1)]^\dagger}{|\mathbf{q} + \mathbf{G}_1|} (\mathbf{q} + \mathbf{G}_1) \cdot (\mathbf{q} + \mathbf{G}_2) \frac{n(\mathbf{G}_1 - \mathbf{G}_2)}{n(\mathbf{G} = 0)} \frac{\Phi_p^0(\mathbf{q}, \mathbf{G}_2)}{|\mathbf{q} + \mathbf{G}_2|}. \quad (5.57)$$

Also in this case, symmetry properties can be used to reconstruct the plasmon parameters in the full Brillouin zone starting from their values in the irreducible wedge. From:

$$\tilde{\epsilon}(\mathbf{q}, \omega = 0) |\Phi_p^0(\mathbf{q})\rangle = \lambda_p(\mathbf{q}, \omega = 0) |\Phi_p^0(\mathbf{q})\rangle, \quad (5.58)$$

and using the symmetry properties of the dielectric matrix, one obtains:

$$\begin{cases} \Phi_p^0(\mathcal{R}\mathbf{q}, \mathbf{G}) = \Phi_p^0(\mathbf{q}, \mathcal{R}^{-1}\mathbf{G}) \\ \lambda_p(\mathcal{R}\mathbf{q}) = \lambda_p(\mathbf{q}) \\ \omega_p(\mathcal{R}\mathbf{q}) = \omega_p(\mathbf{q}) \\ \zeta_p(\mathcal{R}\mathbf{q}) = \zeta_p(\mathbf{q}). \end{cases} \quad (5.59)$$

In ABINIT, time-reversal can be exploited in systems without an inversion center. Using the symmetry properties of the static limit of the symmetrized inverse dielectric matrix,

the following set of equations can be derived:

$$\begin{cases} \Phi_p^0(-\mathcal{R}\mathbf{q}, \mathbf{G}) = \left[\Phi_p^0(\mathbf{q}, \mathcal{R}^{-1}\mathbf{G}) \right]^\dagger \\ \lambda_p(-\mathcal{R}\mathbf{q}) = \lambda_p(\mathbf{q}) \\ \omega_p(-\mathcal{R}\mathbf{q}) = \omega_p(\mathbf{q}) \\ \zeta_p(-\mathcal{R}\mathbf{q}) = \zeta_p(\mathbf{q}), \end{cases} \quad (5.60)$$

where the last three relations derive from the reality of the eigenvalues.

One of the disadvantages of the von der Linden and Horsch approach is that it satisfies the f -sum rule only for the diagonal elements. Engel and Farid have derived a plasmon-pole model with $N_{\mathbf{G}}$ parameters, whose eigenvalues and eigenvectors are frequency dependent, by explicitly constructing an approximation to the reducible polarizability $\chi(\omega)$ [96]. This approximation is exact both in the static- and high-frequency limit, and can be expressed as:

$$\chi_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega) = \left[\omega^2 L^{-1}(\mathbf{q}) + \chi^{-1}(\mathbf{q}, \omega = 0) \right]_{\mathbf{G}_1\mathbf{G}_2}^{-1}, \quad (5.61)$$

where

$$L_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) \equiv (\mathbf{q} + \mathbf{G}_1) \cdot (\mathbf{q} + \mathbf{G}_2) n(\mathbf{G}_1 - \mathbf{G}_2). \quad (5.62)$$

The parameters can be obtained by diagonalizing the matrix in equation 5.61, which is equivalent to the solution of the following generalized eigenvalue problem:

$$\chi(\mathbf{q}, \omega = 0) |x_p(\mathbf{q})\rangle = -\frac{1}{\omega_p^2(\mathbf{q})} L(\mathbf{q}) |x_p(\mathbf{q})\rangle. \quad (5.63)$$

The eigenvectors $x_p(\mathbf{q})$ are normalized such that:

$$\langle x_p(\mathbf{q}) | L(\mathbf{q}) | x_{p'}(\mathbf{q}) \rangle = \delta_{pp'}. \quad (5.64)$$

Defining the scaled eigenvectors $|y_p(\mathbf{q})\rangle$ as:

$$|y_p(\mathbf{q})\rangle = L(\mathbf{q}) |x_p(\mathbf{q})\rangle, \quad (5.65)$$

one finds the following expression [96] for the spectral representation of $\chi(\omega)$:

$$\chi_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega) = \sum_p \frac{y_p(\mathbf{q}, \mathbf{G}_1) y_p(\mathbf{q}, \mathbf{G}_2)^\dagger}{\omega^2 - (\omega_p(\mathbf{q}) - i\eta)^2}, \quad (5.66)$$

The symmetrized inverse dielectric matrix is readily obtained by starting from equation 4.32. The final result is:

$$\begin{aligned}
\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega) &= \delta_{\mathbf{G}_1\mathbf{G}_2} + 4\pi \frac{\chi_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega)}{|\mathbf{q} + \mathbf{G}_1||\mathbf{q} + \mathbf{G}_2|} = \\
&= \delta_{\mathbf{G}_1\mathbf{G}_2} + 4\pi \sum_p \frac{y_p(\mathbf{q}, \mathbf{G}_1)}{|\mathbf{q} + \mathbf{G}_1|} \frac{y_p(\mathbf{q}, \mathbf{G}_2)^\dagger}{|\mathbf{q} + \mathbf{G}_2|} \frac{1}{\omega^2 - (\omega_p(\mathbf{q}) - i\eta)^2}. \quad (5.67)
\end{aligned}$$

The computational advantage of the vHL and EF models over the HL and the GN plasmon-pole models in the context of *GW* calculations is considerable, mainly because the plasmon poles depend on a single plane wave index. Moreover, the numerator in $\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega)$ is given as a direct product of two vectors rather than a full matrix. The problem of small or imaginary plasmon energies affecting the HL PPM does not arise [95, 96], and no additional adjustment is required to deal with systems without an inversion center.

5.5.3 Pros and Cons of the Plasmon-Pole Technique

The most important advantage in using a plasmon-pole model is due (as shown in section 5.7.2.1) to the elimination of frequency convolution defining the self-energy, which can be carried out analytically once the plasmon parameters are known. This leads to considerable computational savings, especially during the computation of the irreducible polarizability, since only one or two frequencies are needed. Therefore, PPM techniques are the ideal tool for initial convergence studies. Besides, the PPM approximation usually proves to be accurate to within 0.1-0.2 eV for states close to the Fermi level [31], when compared to results obtained with a costly numerical integration of the self-energy. On the other hand, the accuracy worsens for states far from the gap, especially for low-lying states where a broadened resonance in the screened interaction can be encountered.

A serious drawback of the PPM method is its limited applicability to systems other than those with *s* and *p* electrons. For more complex systems where the plasmon excitations merge with the particle-hole excitations, it is not clear anymore that simple expressions such as Eq. 5.48 and Eq. 5.47 or expressions in which the frequency dependence is only contained in the eigenvalues are still appropriate. Another, distinct, limitation is that the imaginary part of the self-energy is zero except at the plasmon-poles. Consequently, neither the life-times of the quasiparticles nor the spectral function are accessible. To extract these properties, it is necessary to avoid the use of the PPM methods and calculate the frequency dependence of $W(\omega)$ explicitly, at the price of a drastic increase in the computational effort. In the next section, we therefore present an efficient approach based on the spectral representation of the irreducible polarizability which allows one to remove the bottleneck associated to the evaluation of $\chi^0(\omega)$, the most CPU demanding part in the case of G_0W_0 calculations.

5.6 Spectral Method for the Polarizability

We have seen in section 5.2 that the computational effort for the evaluation of the irreducible polarizability (Eq. 5.28) scales linearly with respect to the number of frequencies N_ω , albeit with a large prefactor which increases with the fourth power of the number of atoms. The main reason for the linear scaling is that the dependence on ω in the denominator of equation 5.28 cannot be factorized out of the sum over the transitions, hence a distinct and expensive summation has to be performed for each frequency, separately. This computational bottleneck can be removed, under certain circumstances, by employing an efficient algorithm proposed in [103], and subsequently revisited in [104], in which only spectral representation of $\chi^0(\omega)$ has to be evaluated in terms of electronic transitions between valence and conduction states, while the real part is efficiently retrieved by means of a less expensive frequency integration.

The spectral representation of the polarizability, $\tilde{\chi}^S(\omega)$, can be expressed in terms of a sum of delta peaks, weighted by the oscillator matrix elements:

$$\begin{aligned} \tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}^S(\mathbf{q}, \omega') = \\ = \frac{2}{V} \sum_{\substack{\mathbf{k} \\ b_1 b_2}} \left[f(\epsilon_{b_1\mathbf{k}-\mathbf{q}}) - f(\epsilon_{b_2\mathbf{k}}) \right] \times M_{\mathbf{G}_1}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \left[M_{\mathbf{G}_2}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \right]^\dagger \text{sign}(\omega') \delta(\omega' - \epsilon_{b_1\mathbf{k}-\mathbf{q}} + \epsilon_{b_2\mathbf{k}}), \end{aligned} \quad (5.68)$$

while the polarizability matrices can be obtained through the Hilbert transform:

$$\chi_{\mathbf{G}_1\mathbf{G}_2}^0(\mathbf{q}, \omega) = \int_0^{+\infty} \tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}^S(\mathbf{q}, \omega') \times \left(\frac{1}{\omega - \omega' + i\delta} - \frac{1}{\omega + \omega' - i\delta} \right) d\omega'. \quad (5.69)$$

Evaluating $\tilde{\chi}^S(\omega')$ at the frequency ω' is rather efficient since only those states satisfying $\omega' = \epsilon_{b_1\mathbf{k}-\mathbf{q}} - \epsilon_{b_2\mathbf{k}}$ for the given ω' have to be considered. $\tilde{\chi}^S(\omega')$ is calculated on a user-specified discrete grid $\{\omega'_i, i = 1, N_{\omega'}\}$, and the δ -function is approximated by means of a triangular function centered at the energy transition $\Delta_{n'n}^{\mathbf{kq}} = \epsilon_{n'\mathbf{k}-\mathbf{q}} - \epsilon_{n\mathbf{k}}$ using the following expression [104]:

$$\delta(\omega_i - \Delta_{n'n}^{\mathbf{kq}}) = \begin{cases} 0 & \text{if } \Delta_{n'n}^{\mathbf{kq}} < \omega_{i-1} \\ \frac{\Delta_{n'n}^{\mathbf{kq}} - \omega_{i-1}}{\omega_i - \omega_{i-1}} & \text{for } \omega_{i-1} < \Delta_{n'n}^{\mathbf{kq}} \leq \omega_i \\ \frac{\omega_{i+1} - \Delta_{n'n}^{\mathbf{kq}}}{\omega_{i+1} - \omega_i} & \text{for } \omega_i \leq \Delta_{n'n}^{\mathbf{kq}} < \omega_{i+1} \\ 0 & \text{if } \Delta_{n'n}^{\mathbf{kq}} > \omega_{i+1} \end{cases} \quad (5.70)$$

This particular definition allows us to preserve the norm of the delta distribution independently on the type of mesh used for ω' ; both linear meshes or meshes intensified around the plasma frequency might be used.⁵ Since, according to equation 5.70, each electron-hole pair only contributes to two single ω' frequencies, the CPU time required to evaluate $\tilde{\chi}^S(\omega')$ is just twice that required by a single static evaluation based on the standard Adler-Wiser expression reported in equation 5.29. This is independent of the (large) number of ω' points that have to be used in order to correctly describe the behavior of the delta function. The price to be paid, however, is that a large table for the imaginary part (Eq. 5.68) has to be stored in memory and a Hilbert transform has to be performed for each $(\mathbf{G}_1, \mathbf{G}_2)$ pair. The storage of $\tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}^S(\omega')$ can be avoided if a partial Hilbert transform is done for each transition, but this choice would lead to a dramatic worsening of the performance since the number of transitions in real systems (and thus the number of Hilbert transforms to be done for each $(\mathbf{G}_1, \mathbf{G}_2)$ pair) is usually of the order of thousands or higher. In the present implementation, the computational gain is thus governed by the effort required by the frequency integral given in equation 5.69. For systems in which a large plane wave basis set is needed to describe the spatial variations of $\chi^0(\mathbf{r}_1, \mathbf{r}_2; \omega)$, the CPU time needed for the Hilbert transform will exceed the one for the standard Adler-Wiser formulation. The position of the crossover point between the two algorithms is system- and hardware-dependent and, moreover, heavily dependent on the particular implementation. An optimal distribution of both computing load and memory can be achieved by distributing the conduction states across the nodes. By virtue of the particular distribution of the data, each processor has to store and treat only a subset of the entire range of transition energies. The algorithm therefore presents good scalability since the number of ω' frequencies that have to be stored and considered in the Hilbert transform decreases with the number of processors that are used.

The symmetrized expression for the imaginary part in which the BZ summation is performed only in the irreducible wedge defined by the external $\mathbf{q}$ reads:

$$\begin{aligned} \tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}^S(\mathbf{q}, \omega') = & \frac{2}{V} \sum_{b_1 b_2} \sum_{\mathbf{k}}^{IBZ_{\mathbf{q}}} \text{sign}(\omega') \delta(\omega' - \epsilon_{b_1\mathbf{k}-\mathbf{q}} + \epsilon_{b_2\mathbf{k}}) \left(f(\epsilon_{b_1\mathbf{k}-\mathbf{q}}) - f(\epsilon_{b_2\mathbf{k}}) \right) \times \\ & \left\{ \sum_{\hat{\mathcal{R}}_{\mathbf{t}}}^{\mathcal{L}_{\mathbf{q}}^+} \tilde{w}^+(\hat{\mathcal{R}}_{\mathbf{t}}, \mathbf{k}) e^{-i(\mathbf{G}_1 - \mathbf{G}_2) \cdot \mathbf{t}} M_{\mathcal{R}^{-1}\mathbf{G}_1}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \left[M_{\mathcal{R}^{-1}\mathbf{G}_2}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \right]^\dagger + \right. \\ & \left. + \sum_{\hat{\mathcal{R}}_{\mathbf{t}}}^{\mathcal{L}_{\mathbf{q}}^-} \tilde{w}^-(\hat{\mathcal{R}}_{\mathbf{t}}, \mathbf{k}) e^{-i(\mathbf{G}_1 - \mathbf{G}_2) \cdot \mathbf{t}} \left[M_{-\mathcal{R}^{-1}\mathbf{G}_1}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \right]^\dagger M_{-\mathcal{R}^{-1}\mathbf{G}_2}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \right\}. \quad (5.71) \end{aligned}$$

⁵ We have also implemented a gaussian approximant for the δ function, following [103], but we did not find significant improvements in the systems studied so far. The rate of converge of the two algorithms with respect to $N_{\omega'}$ is similar, but the method based on the triangular approximant (5.70) is somewhat easier to use and considerably faster, as only two frequency points are needed. Unlike the gaussian approximant of [103], no convergence study on the value of the gaussian broadening has to be performed, and the accuracy of the results only depends on the frequency step used for the ω' mesh.

At high symmetry $\mathbf{q}$ -points, the symmetry relation given in equation 5.35 can be exploited such that the Hilbert transform has to be performed only for the set of independent $(\mathbf{G}_1, \mathbf{G}_2)$ pairs. An additional, distinct, computing saving is obtained if the analytical continuation of $\chi^0(\omega)$ along the imaginary axis is required. In this case, the Hilbert transform can be performed only for, say, the upper triangle of the matrix, since the lower triangle can be reconstructed by enforcing the matrix to be Hermitian.

Figure 5.3 reports an application of the method to the calculation of the inverse dielectric matrix in α -quartz SiO_2 . Once converged with respect to the number of ω' points, the results obtained using the Hilbert transform method are very close to those from the standard Adler-Wiser expressions. Numerical tests have shown that using the two algorithms for evaluating $\tilde{\epsilon}^{-1}(\omega)$ leads to changes in the QP corrections of the order of a few tenths of meV. This is expected since the calculation of the self-energy requires an integration over frequency, and the final results are supposed to be insensitive to the fine details of the integrand. Moreover, when sophisticated algorithms such as the contour deformation method (Sect. 5.7.2.2), or the analytic continuation (Sect. 5.7.2.3), are employed to calculate QP energies, the frequency convolution defining the self-energy is partly or entirely evaluated along the imaginary axis where, the Hilbert transform algorithm turns out to converge fast with respect to $N_{\omega'}$, giving results very close to the Adler-Wiser values (not shown).

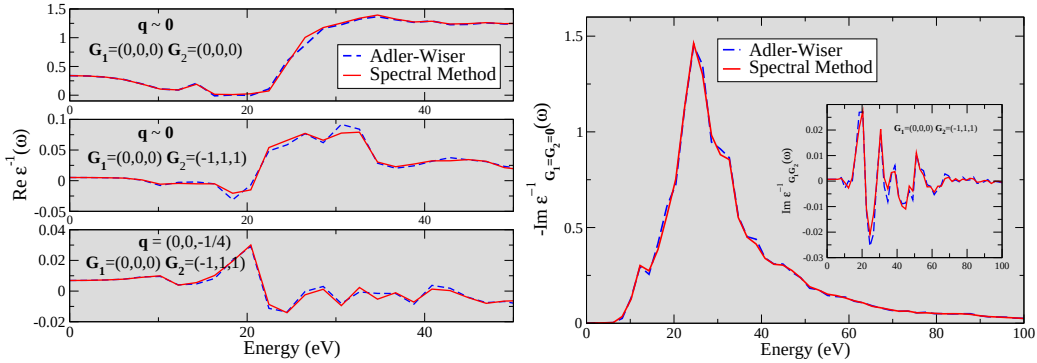


Figure 5.3: Left panel: Comparison between selected matrix elements of $\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega)$ of α -quartz SiO_2 calculated using the standard Adler-Wiser expression given in equation 5.29, and the method based on the Hilbert transform (Eq. 5.68 and Eq. 5.69). Right panel: Comparison between the electron energy loss function of α -quartz SiO_2 calculated using the two methods. The inset shows the imaginary part of an off-diagonal element of $\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q} \rightarrow 0, \omega)$.

The Hilbert transform method also represents a reliable and fast algorithm for analyzing optical properties including local field effects, as shown in [105]. In this case, however, special care has to be taken in order to properly account for the commutator between the position operator $\mathbf{r}$ and the nonlocal part of the Hamiltonian, as described

in the two appendices C and E.

5.7 Evaluation of the GW self-energy

In the previous sections, several methods used to calculate the screened interaction (or to model it from first principles) have been discussed. Once $W(\omega)$ is available, the self-energy operator can be constructed and the quasiparticle energies can be obtained by solving the Dyson-like equation given in equation 4.53. For convenience, the basic equations are summarized below.

$$\Sigma(\mathbf{r}_1, \mathbf{r}_2; \omega) = \frac{i}{2\pi} \int e^{i\omega'\delta^+} G(\mathbf{r}_1, \mathbf{r}_2; \omega + \omega') W(\mathbf{r}_1, \mathbf{r}_2; \omega') d\omega', \quad (5.72)$$

where

$$W_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}, \omega) = \tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega) \tilde{v}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) = \tilde{v}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) + \underbrace{(\tilde{\epsilon}_{\mathbf{G}_1\mathbf{G}_2}^{-1}(\mathbf{q}, \omega) - \delta_{\mathbf{G}_1, \mathbf{G}_2})}_{\tilde{\alpha}_{\mathbf{G}_1\mathbf{G}_2}^c(\mathbf{q}, \omega)} \tilde{v}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}). \quad (5.73)$$

In equation 5.73, the screened interaction has been split into a static term given by the symmetrized bare Coulomb interaction:

$$\tilde{v}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) \equiv \frac{4\pi}{|\mathbf{q} + \mathbf{G}_1||\mathbf{q} + \mathbf{G}_2|}, \quad (5.74)$$

and a second dynamic contribution arising from the frequency-dependent part of the inverse dielectric matrix, $\tilde{\alpha}_{\mathbf{G}_1\mathbf{G}_2}^c(\mathbf{q}, \omega)$. Substituting definition 5.73 into equation 5.72 yields

$$\Sigma(\mathbf{r}_1, \mathbf{r}_2; \omega) \equiv \Sigma_x(\mathbf{r}_1, \mathbf{r}_2) + \Sigma_c(\mathbf{r}_1, \mathbf{r}_2; \omega), \quad (5.75)$$

where the bare exchange term, $\hat{\Sigma}_x$, arises from the static part of the screened interaction and the dynamic operator $\hat{\Sigma}_c(\omega)$ accounts for correlation effects beyond $\hat{\Sigma}_x$. The formalism can be easily generalized to collinear magnetism. In this case, since the projection of the spin is a constant of the motion, $\hat{\Sigma}$ cannot connect states with different spin. Therefore the self-energy operator is given by a 2×2 diagonal matrix:

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}^{\uparrow\uparrow} = iG^{\uparrow\uparrow}W & 0 \\ 0 & \hat{\Sigma}^{\downarrow\downarrow} = iG^{\downarrow\downarrow}W \end{pmatrix} \quad (5.76)$$

where the screened interaction, which is not spin dependent, is calculated starting from the trace of the irreducible polarizability given in equation 5.30. As expected, equation 5.76 reduces to the standard formulation in case of spin-compensated systems.

At this point, all the ingredients needed for the calculation of the quasiparticle energies and amplitudes are available, and only the Dyson-like equation has to be solved.

Finding all the QP energies at a particular $\mathbf{k}$ -point with the standard perturbative approach would require the knowledge of the plane wave representation $\Sigma_{\mathbf{G}_1\mathbf{G}_2}^{\mathbf{k}}(\omega)$ for a set of frequencies around the KS band energies, such that a linear expansion in ω can be performed. The evaluation of the nonlocal operator in the plane wave basis set, however, would be prohibitively expensive, as it would scale with the fourth power of the number of atoms [56]. If, as is usually the case, only a few quasiparticle energies are required, it is much more efficient to evaluate the matrix elements of $\hat{\Sigma}$ only between selected Bloch states, since this part scales only with the third power of the number of atoms [106, 107, 56].

In the forthcoming sections we shall present the equations needed to evaluate the matrix elements of the self-energy in the KS representation. First we discuss how to evaluate the exchange part, then we proceed to present the different techniques that can be used to handle the more complicated frequency-dependent contribution.

5.7.1 Exchange Part of the Self-Energy

For the exchange part, the only contribution to the convolution in the frequency domain comes from the dynamic part of the Green's function. Substituting the Lehmann representation of the propagator in equation 5.72 yields

$$\begin{aligned}\Sigma_x(\mathbf{r}_1, \mathbf{r}_2) &\equiv \frac{i}{2\pi} \int G(\mathbf{r}_1, \mathbf{r}_2; \omega + \omega') v(\mathbf{r}_1, \mathbf{r}_2) e^{i\omega'\delta} d\omega' = \\ &= \frac{i}{2\pi} \sum_s \Psi_s(\mathbf{r}_1) \Psi_s^\dagger(\mathbf{r}_2) v(\mathbf{r}_1, \mathbf{r}_2) \times \underbrace{\int \frac{e^{i\omega'\delta^+} d\omega'}{\omega + \omega' - \epsilon_s + i\eta \text{sign}(\epsilon_s - \mu)}}_{2\pi i \theta(\mu - \epsilon_s)},\end{aligned}\quad (5.77)$$

where the integral on the right hand side has been evaluated using the residue theorem, by closing the integration path in the upper half of the complex plane, where the integrand goes to zero exponentially, and the limit $\delta \rightarrow 0^+$ has been taken at the end of the derivation. The integrand presents a singularity located at $z = -\omega + \epsilon_s - i\eta \text{sign}(\epsilon_s - \mu)$, but only the poles corresponding to occupied states (positive imaginary part) contribute to the integral. The resulting operator is static, Hermitian, and turns out to have the same mathematical structure as the Fock operator in Hartree-Fock theory, albeit constructed with quasiparticle amplitudes and energies. In crystalline systems, characterized by an energy gap in the neutral excitation energies, $\Sigma_x(\mathbf{r}_1, \mathbf{r}_2)$ assumes the form

$$\Sigma_x(\mathbf{r}_1, \mathbf{r}_2) = - \sum_{\mathbf{k}} \sum_{\nu}^{\text{BZ occ}} \Psi_{n\mathbf{k}}(\mathbf{r}_1) \Psi_{n\mathbf{k}}^\dagger(\mathbf{r}_2) v(\mathbf{r}_1, \mathbf{r}_2). \quad (5.78)$$

The matrix elements of $\hat{\Sigma}_x$ between two wavefunctions are given by:

$$\langle b_1 \tilde{\mathbf{k}} | \hat{\Sigma}_x | b_2 \tilde{\mathbf{k}} \rangle = - \frac{4\pi}{V} \sum_{\nu}^{\text{occ}} \sum_{\mathbf{q}}^{\text{BZ}} \sum_{\mathbf{G}} \frac{[M_{\mathbf{G}}^{\nu b_1}(\tilde{\mathbf{k}}, \mathbf{q})]^\dagger M_{\mathbf{G}}^{\nu b_2}(\tilde{\mathbf{k}}, \mathbf{q})}{|\mathbf{q} + \mathbf{G}|^2}. \quad (5.79)$$

The diagonal elements are always negative and have the simpler form:

$$\langle b_1 \tilde{\mathbf{k}} | \hat{\Sigma}_x | b_1 \tilde{\mathbf{k}} \rangle = -\frac{4\pi}{V} \sum_{\nu}^{\text{occ}} \sum_{\mathbf{q}}^{\text{BZ}} \sum_{\mathbf{G}} \frac{|M_{\mathbf{G}}^{b_1 b_1}(\tilde{\mathbf{k}}, \mathbf{q})|^2}{|\mathbf{q} + \mathbf{G}|^2}. \quad (5.80)$$

The computational effort needed for the evaluation of Eq. 5.79 or Eq. 5.80 is proportional to $N_{\nu} \times N_{\mathbf{G}} \times N_{\mathbf{q}}$, but one should always keep in mind that, due to the long range of the bare Coulomb interaction, the convergence with respect to the number of plane waves used in the oscillator matrix elements $M_{\mathbf{G}}$ is usually slow, much slower than the convergence of the correlation part, where the presence of $\tilde{\alpha}_{\mathbf{G}_1 \mathbf{G}_2}^c(\mathbf{q}, \omega)$ introduces a cutoff in $\mathbf{G}$ -space.

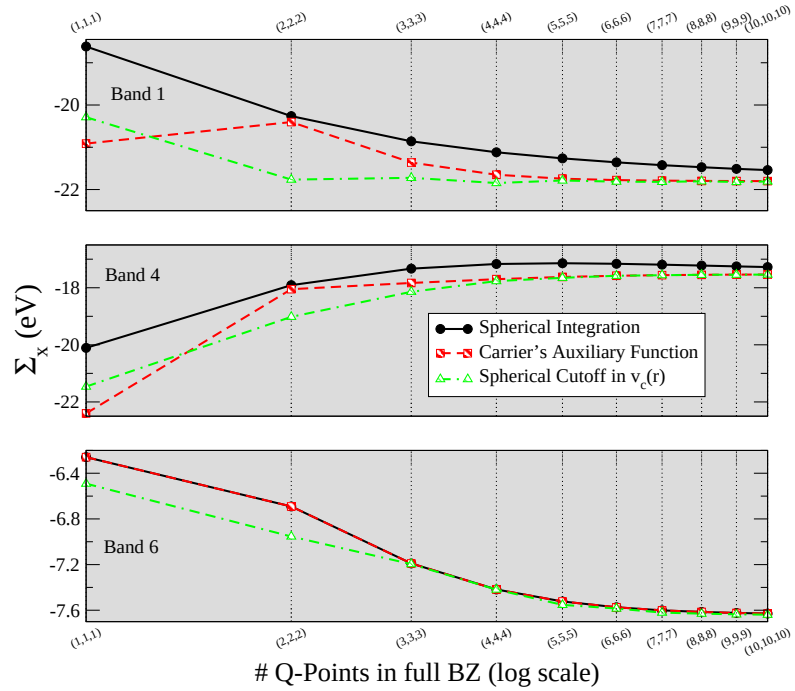


Figure 5.4: Effect of the different treatments of the Coulomb singularity on the convergence of the diagonal matrix elements of $\hat{\Sigma}_x$ with respect to the $\mathbf{q}$ -point sampling the full BZ of SiC. The spherical integration used in [56], the auxiliary function proposed by Carrier [108], and a spherical cutoff in the bare Coulomb interaction [109] are compared. The dashed and solid lines are there to guide the eyes. The $\mathbf{k}$ -point meshes are indicated by parentheses.

A point worth noting is the presence in equation 5.80 of the Coulomb singularity for $|\mathbf{q} + \mathbf{G}| \rightarrow 0$. From a mathematical point of view, the integral is well-defined since the singularity is integrable in three-dimensional space once the thermodynamical limit, *i.e.* $N_{\mathbf{q}} \rightarrow \infty$, is reached. On the other hand, only a finite number of $\mathbf{q}$ -points can be used for practical applications, and a careful numerical treatment is needed to avoid an

exceedingly slow convergence with respect to the BZ sampling. In the original work of Hybertsen and Louie [56], the contribution to the integral given by the small box Ω_Γ around Γ is approximated as:

$$\int_{\Omega_\Gamma} \frac{f(\mathbf{q})}{|\mathbf{q}|^2} d\mathbf{q} \approx f(\mathbf{q}=0) \int_{\Omega_\Gamma} \frac{d\mathbf{q}}{|\mathbf{q}|^2}, \quad (5.81)$$

where $f(\mathbf{q})$ is assumed to be smooth around $\mathbf{q} = 0$. Then Ω_Γ is taken to be spherical and the singularity is integrated out analytically. This approach might, however, hinder the convergence with respect to $N_\mathbf{q}$, especially in anisotropic systems where the small volume around the Γ point is highly nonspherical. Another possible approach consist in introducing an auxiliary function with the same asymptotic behavior for $\mathbf{q} \rightarrow 0$ whose integral can be analytically performed. The auxiliary function should have the same symmetries of the crystal such that the special $\mathbf{k}$ -points technique can be exploited [110]. Expressions for selected Bravais lattices have been reported in [111, 112] but they are now superseded by the auxiliary function proposed in [108]:

$$f(\mathbf{q}) = 2\pi^2 \left\{ 4 \sum_{j=1}^3 [\mathbf{b}_j \sin(\mathbf{a}_j \cdot \mathbf{q}/2)] \cdot [\mathbf{b}_j \sin(\mathbf{a}_j \cdot \mathbf{q}/2)] + \right. \\ \left. + 2 \sum_{j=1}^3 [\mathbf{b}_j \sin(\mathbf{a}_j \cdot \mathbf{q})] \cdot [\mathbf{b}_{j+1} \sin(\mathbf{a}_{j+1} \cdot \mathbf{q})] \right\}^{-1}, \quad (5.82)$$

which is applicable to arbitrary lattices. In equation 5.82, $\mathbf{b}_j$ are the reciprocal lattice vectors with $\mathbf{b}_4 = \mathbf{b}_1$ for a compact notation, and the $\mathbf{a}_j$ are the corresponding lattice vectors spanning the unit cell. Another interesting technique to handle the Coulomb singularity has recently been proposed by Spencer [109] in the framework of total energy calculations within the exact-exchange formalism [113]. In this case, the true Coulomb interaction is truncated within a sphere of radius R whose volume is equal to that of the Born-von Kármán crystal and the Coulomb term in equation 5.80 is replaced by

$$v_0(\mathbf{q}, \mathbf{G}) = \frac{4\pi}{|\mathbf{q} + \mathbf{G}|^2} [1 - \cos(|\mathbf{q} + \mathbf{G}|R)]. \quad (5.83)$$

Expression 5.83 reduces to the true bare Coulomb interaction in the limit $N_\mathbf{q} \rightarrow \infty$, and has a well-defined finite limit for $|\mathbf{q} + \mathbf{G}| \rightarrow 0$. To analyze the accuracy of the three different methods, we have compared the convergence of the diagonal elements of $\hat{\Sigma}_x$ in SiC with respect to the number of sampled $\mathbf{q}$ -points. The results are shown in Figure 5.4. Although SiC is a relatively simple and isotropic system, the spherical integration of Hybertsen and Louie [56] is marred by slow convergence and is outperformed by the other two methods. The spherical cutoff in the Coulomb interaction improves upon the auxiliary method, especially for low-lying states. The improved performance of the spherical cutoff for occupied states is consistent with the better convergence in the exact-exchange total energy observed in [109].

5.7.1.1 Symmetrization of $\hat{\Sigma}_x$

For the calculation of the matrix elements of $\langle b_1 \tilde{\mathbf{k}} | \hat{\Sigma}_x | b_2 \tilde{\mathbf{k}} \rangle$ the point group of the external wave-vector $\tilde{\mathbf{k}}$ defines the irreducible wedge. Let us suppose that a particular rotation preserves $\tilde{\mathbf{k}}$ (*i.e.* $\mathcal{R}\tilde{\mathbf{k}} = \tilde{\mathbf{k}} + \mathbf{G}_{\mathcal{R}}$) and let $\mathcal{L}_{\mathbf{q}}^+$ be the subset of such rotations (*i.e.* the small group of $\tilde{\mathbf{k}}$). If time-reversal symmetry can be exploited, the set of rotations $\mathcal{R}$ such that $-\mathcal{R}\tilde{\mathbf{k}} = \tilde{\mathbf{k}} + \mathbf{G}_{\mathcal{R}}$ can also be included. This additional ensemble will henceforth be indicated with the symbol $\mathcal{L}_{\mathbf{q}}^-$. Using the symmetry relationships of the oscillator matrix elements from equation B.3, the sum defining the matrix elements of $\hat{\Sigma}_x$ can be rewritten as:

$$\begin{aligned} \langle b_1 \tilde{\mathbf{k}} | \hat{\Sigma}_x | b_2 \tilde{\mathbf{k}} \rangle &= -\frac{4\pi}{V} \sum_{\nu}^{\text{occ}} \sum_{\mathbf{q}}^{IBZ_{\mathbf{k}}} \sum_{\mathbf{G}} \left\{ \sum_{\hat{\mathcal{R}}_{\mathbf{t}} \in \mathcal{L}_{\mathbf{q}}^+} \tilde{w}^+(\hat{\mathcal{R}}_{\mathbf{t}}, \mathbf{q}) \frac{[M_{\mathcal{R}^{-1}\mathbf{G}}^{\nu b_1}(\tilde{\mathbf{k}}, \mathbf{q})]^{\dagger} M_{\mathcal{R}^{-1}\mathbf{G}}^{\nu b_2}(\tilde{\mathbf{k}}, \mathbf{q})}{|\mathbf{q} + \mathcal{R}^{-1}\mathbf{G}|^2} + \right. \\ &\quad \left. + \sum_{\hat{\mathcal{R}}_{\mathbf{t}} \in \mathcal{L}_{\mathbf{q}}^-} \tilde{w}^-(\hat{\mathcal{R}}_{\mathbf{t}}, \mathbf{q}) \frac{M_{-\mathcal{R}^{-1}\mathbf{G}}^{\nu b_1}(\tilde{\mathbf{k}}, \mathbf{q}) [M_{-\mathcal{R}^{-1}\mathbf{G}}^{\nu b_2}(\tilde{\mathbf{k}}, \mathbf{q})]^{\dagger}}{|\mathbf{q} - \mathcal{R}^{-1}\mathbf{G}|^2} \right\} = \\ &= -\frac{4\pi}{V} \sum_{\mathbf{q}}^{IBZ_{\mathbf{k}}} \sum_{\nu}^{\text{occ}} \left(w^+(\mathbf{q}) + w^-(\mathbf{q}) \right) \text{Re} \Lambda(\tilde{\mathbf{k}}, \mathbf{q})^{\nu b_1 b_2} + i \left(w^+(\mathbf{q}) - w^-(\mathbf{q}) \right) \text{Im} \Lambda(\tilde{\mathbf{k}}, \mathbf{q})^{\nu b_1 b_2} \end{aligned} \quad (5.84)$$

where $w^+(\mathbf{q})$, $w^-(\mathbf{q})$ are the number of distinct points in the full Brillouin zone which are related to $\mathbf{q}$ through the symmetry operations in $\mathcal{L}_{\mathbf{q}}^+$ and $\mathcal{L}_{\mathbf{q}}^-$, respectively. Furthermore, the following definition has been introduced:

$$\Lambda(\tilde{\mathbf{k}}, \mathbf{q})^{\nu b_1 b_2} \equiv \sum_{\mathbf{G}} \frac{[M_{\mathbf{G}}^{\nu b_1}(\tilde{\mathbf{k}}, \mathbf{q})]^{\dagger} M_{\mathbf{G}}^{\nu b_2}(\tilde{\mathbf{k}}, \mathbf{q})}{|\mathbf{q} + \mathbf{G}|^2}. \quad (5.85)$$

One thing worth noting is that, contrary to the symmetrized expression obtained for the irreducible polarizability in equation 5.45, here the action of the symmetry operations in the degenerate subspace has to be considered when reconstructing wave functions at symmetrical $\mathbf{k}$ -points. This is due to the presence of a single sum over bands, so the unitary transformation cannot be factorized and removed by invoking the orthogonality theorem of group theory [93]. An analogous problem occurs also for the correlation part $\hat{\Sigma}_c(\omega)$. A proper symmetrization of the expressions therefore requires the knowledge and the correct inclusion of the unitary transformation given in equation B.8, for each symmetry operation and each set of degenerate states. Such a calculation would be expensive as it would require the knowledge of the wave functions in the full Brillouin zone, and the correct inclusion of the unitary transformation would slow down the calculation of the QP corrections. A symmetric invariant for the diagonal matrix elements of $\hat{\Sigma}$ can be cheaply constructed, however, by simply averaging the QP results within the degenerate

subspace. This is the solution proposed in the seminal work of Hybertsen and Louie [56] and it is also the approach implemented in the *GW* code of ABINIT.

5.7.2 The Correlation Part of the Self-Energy

The expression for the correlation part, $\Sigma_c(\omega)$, in the frequency-space domain is given by:

$$\begin{aligned} \Sigma_c(\mathbf{r}_1, \mathbf{r}_2; \omega) = & \frac{i}{2\pi V} \sum_{\mathbf{q}} \sum_{\mathbf{G}_1 \mathbf{G}_2} \sum_s \Psi_s(\mathbf{r}_1) e^{i(\mathbf{q} + \mathbf{G}_1) \cdot \mathbf{r}_1} \Psi_s^\dagger(\mathbf{r}_2) e^{-i(\mathbf{q} + \mathbf{G}_2) \cdot \mathbf{r}_2} \times \\ & \times \underbrace{\tilde{v}_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}) \int_{-\infty}^{+\infty} \frac{\tilde{\alpha}_{\mathbf{G}_1, \mathbf{G}_2}^c(\mathbf{q}, \omega') e^{i\omega' \delta} d\omega'}{\left(\omega + \omega' - \epsilon_s + i\eta \text{sign}(\epsilon_s - \mu)\right)}}_{\mathcal{J}_{\mathbf{G}_1 \mathbf{G}_2}^s(\mathbf{q}, \omega)}, \end{aligned} \quad (5.86)$$

where all dynamical effects are now contained in the integral $\mathcal{J}_{\mathbf{G}_1 \mathbf{G}_2}^s(\mathbf{q}, \omega, s)$, and s is a shorthand notation for crystalline momentum and band index. Unlike the exchange contribution, the matrix elements of $\hat{\Sigma}_c(\omega)$ between two Bloch states are now expressed in terms of a summation over the entire set of states, both empty and occupied

$$\langle b_1 \tilde{\mathbf{k}} | \hat{\Sigma}_c(\omega) | b_2 \tilde{\mathbf{k}} \rangle = \frac{i}{2\pi V} \sum_{\mathbf{q}} \sum_{\mathbf{G}_1 \mathbf{G}_2} \sum_{n=1}^{+\infty} \left[M_{\mathbf{G}_1}^{nb_1}(\tilde{\mathbf{k}}, \mathbf{q}) \right]^\dagger M_{\mathbf{G}_2}^{nb_2}(\tilde{\mathbf{k}}, \mathbf{q}) \tilde{v}_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}) \mathcal{J}_{\mathbf{G}_1 \mathbf{G}_2}^{n\tilde{\mathbf{k}}-\mathbf{q}}(\mathbf{q}, \omega). \quad (5.87)$$

For the forthcoming discussion, in particular for the symmetrization in $\mathbf{q}$ -space of equation 5.87, it is useful to analyze how $\mathcal{J}_{\mathbf{G}_1 \mathbf{G}_2}^{n\tilde{\mathbf{k}}-\mathbf{q}}(\mathbf{q}, \omega)$ transforms in $\mathbf{q}$ -space under the application of a symmetry operation of the crystal. Due to the presence of the crystalline momentum $\tilde{\mathbf{k}} - \mathbf{q}$, only those symmetries belonging to the point group of the external $\tilde{\mathbf{k}}$ can be employed. Once this constraint is fulfilled, the symmetry properties of $\mathcal{J}_{\mathbf{G}_1 \mathbf{G}_2}^{n\tilde{\mathbf{k}}-\mathbf{q}}(\mathbf{q}, \omega)$ can be easily derived from the knowledge of the symmetry transformation of the inverse dielectric matrix. The frequency integral at a generic point $\mathcal{R}\mathbf{q}$ (for simplicity supposed to be in the full Brillouin zone), can be related to the value calculated at the irreducible point $\mathbf{q}$ by means of:

$$\mathcal{J}_{\mathbf{G}_1 \mathbf{G}_2}^{n\mathcal{R}(\mathbf{k}-\mathbf{q})}(\mathcal{R}\mathbf{q}, \omega) = e^{i(\mathbf{G}_2 - \mathbf{G}_1) \cdot \mathbf{t}} \mathcal{J}_{\mathcal{R}^{-1}\mathbf{G}_1 \mathcal{R}^{-1}\mathbf{G}_2}^{n\mathbf{k}-\mathbf{q}}(\mathbf{q}, \omega), \quad \text{if } \mathcal{R}\mathbf{k} = \mathbf{k} + \mathbf{G}_{\mathcal{R}}. \quad (5.88)$$

In systems without an inversion center, it is also possible to take advantage of time-reversal symmetry at the price of having to perform a transposition of the matrix in $\mathbf{G}$ -space:

$$\mathcal{J}_{\mathbf{G}_1 \mathbf{G}_2}^{n-\mathcal{R}(\mathbf{k}-\mathbf{q})}(-\mathcal{R}\mathbf{q}, \omega) = e^{i(\mathbf{G}_2 - \mathbf{G}_1) \cdot \mathbf{t}} \mathcal{J}_{-\mathcal{R}^{-1}\mathbf{G}_2 \mathcal{R}^{-1}\mathbf{G}_1}^{n\mathbf{k}-\mathbf{q}}(\mathbf{q}, \omega), \quad \text{if } -\mathcal{R}\mathbf{k} = \mathbf{k} + \mathbf{G}_{\mathcal{R}}. \quad (5.89)$$

The above relation complicates the symmetrization in $\mathbf{q}$ -space of $\hat{\Sigma}_c(\omega)$ somewhat because the inverse dielectric matrix is not Hermitian, along the real axis, and neither is its integral $\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega)$. Nevertheless, it is still possible to exploit symmetries, in particular the invariance under time reversal, by making a decomposition of $\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega)$ into a Hermitian and an anti-Hermitian part:

$$\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega) = \mathcal{A}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega) + i\mathcal{B}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega), \quad (5.90)$$

where

$$\begin{cases} \mathcal{A}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega) \equiv \frac{\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega) + [\mathcal{J}_{\mathbf{G}_2\mathbf{G}_1}^s(\mathbf{q}, \omega)]^\dagger}{2} \\ \mathcal{B}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{k}, \mathbf{q}, \omega) \equiv \frac{\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega) - [\mathcal{J}_{\mathbf{G}_2\mathbf{G}_1}^s(\mathbf{q}, \omega)]^\dagger}{2i}. \end{cases} \quad (5.91)$$

The decomposition given in equation 5.90, in conjunction with Eq. 5.88 and Eq. 5.89, allows us to reconstruct by symmetry all the matrix elements involved in the calculation of $\hat{\Sigma}_c(\omega)$. The derivation of the symmetrized expression is summarized below.

$$\begin{aligned} \langle b_1 \tilde{\mathbf{k}} | \Sigma_c(\omega) | b_2 \tilde{\mathbf{k}} \rangle &= \frac{2i}{V} \sum_{m=1}^{+\infty} \sum_{\mathbf{q}} \sum_{\mathbf{G}_1\mathbf{G}_2}^{IBZ_{\mathbf{k}}} \\ &\left\{ \sum_{\hat{\mathbf{R}} \in \mathcal{L}_{\mathbf{q}}^+} \tilde{w}^+(\hat{\mathbf{R}}_{\mathbf{t}}, \mathbf{k}) \frac{[M_{\mathcal{R}^{-1}\mathbf{G}_1}^{mb_1}(\tilde{\mathbf{k}}, \mathbf{q})]^\dagger}{|\mathbf{q} + \mathcal{R}^{-1}\mathbf{G}_1|} \frac{M_{\mathcal{R}^{-1}\mathbf{G}_2}^{mb_2}(\tilde{\mathbf{k}}, \mathbf{q})}{|\mathbf{q} + \mathcal{R}^{-1}\mathbf{G}_2|} \times \mathcal{J}_{\mathcal{R}^{-1}\mathbf{G}_1\mathcal{R}^{-1}\mathbf{G}_2}^{m\tilde{\mathbf{k}}-\mathbf{q}}(\mathbf{q}, \omega) + \right. \\ &+ \left. \sum_{\hat{\mathbf{R}} \in \mathcal{L}_{\mathbf{q}}^-} \tilde{w}^-(\hat{\mathbf{R}}_{\mathbf{t}}, \mathbf{k}) \frac{M_{-\mathcal{R}^{-1}\mathbf{G}_1}^{mb_1}(\tilde{\mathbf{k}}, \mathbf{q})}{|\mathbf{q} - \mathcal{R}^{-1}\mathbf{G}_1|} \frac{[M_{-\mathcal{R}^{-1}\mathbf{G}_2}^{mb_2}(\tilde{\mathbf{k}}, \mathbf{q})]^\dagger}{|\mathbf{q} - \mathcal{R}^{-1}\mathbf{G}_2|} \times \mathcal{J}_{-\mathcal{R}^{-1}\mathbf{G}_2-\mathcal{R}^{-1}\mathbf{G}_1}^{m\tilde{\mathbf{k}}-\mathbf{q}}(\mathbf{q}, \omega) \right\} = \\ &= \frac{2i}{V} \sum_{m=1}^{+\infty} \sum_{\mathbf{q}}^{IBZ_{\mathbf{k}}} \left\{ \left(w^+(\mathbf{q}) \Xi(\tilde{\mathbf{k}}, \mathbf{q})^{mb_1b_2} + w^-(\mathbf{q}) [\Xi(\tilde{\mathbf{k}}, \mathbf{q})^{mb_1b_2}]^\dagger \right) + \right. \\ &\quad \left. + i \left(w^+(\mathbf{q}) \Pi(\tilde{\mathbf{k}}, \mathbf{q})^{mb_1b_2} + w^-(\mathbf{q}) [\Pi(\tilde{\mathbf{k}}, \mathbf{q})^{mb_1b_2}]^\dagger \right) \right\} \quad (5.92) \end{aligned}$$

where the following quantities have been introduced

$$\Xi(\tilde{\mathbf{k}}, \mathbf{q})^{mb_1b_2} \equiv \sum_{\mathbf{G}_1\mathbf{G}_2} \frac{[M_{\mathbf{G}_1}^{mb_1}(\tilde{\mathbf{k}}, \mathbf{q})]^\dagger}{|\mathbf{q} + \mathbf{G}_1|} \frac{M_{\mathbf{G}_2}^{mb_2}(\tilde{\mathbf{k}}, \mathbf{q})}{|\mathbf{q} + \mathbf{G}_2|} \times \mathcal{A}_{\mathbf{G}_1\mathbf{G}_2}^{m\tilde{\mathbf{k}}-\mathbf{q}}(\mathbf{q}, \omega), \quad (5.93)$$

and

$$\Pi(\tilde{\mathbf{k}}, \mathbf{q})^{mb_1 b_2} \equiv \sum_{\mathbf{G}_1 \mathbf{G}_2} \frac{[M_{\mathbf{G}_1}^{mb_1}(\tilde{\mathbf{k}}, \mathbf{q})]^\dagger}{|\mathbf{q} + \mathbf{G}_1|} \frac{M_{\mathbf{G}_2}^{mb_2}(\tilde{\mathbf{k}}, \mathbf{q})}{|\mathbf{q} + \mathbf{G}_2|} \times \mathcal{B}_{\mathbf{G}_1 \mathbf{G}_2}^{m\tilde{\mathbf{k}}-\mathbf{q}}(\mathbf{q}, \omega). \quad (5.94)$$

The expression in equation 5.92 can be further simplified if only the diagonal matrix elements of $\hat{\Sigma}_c(\omega)$ are needed. In this particular case, both terms Ξ and Π are real-valued, and one arrives at the important conclusion that only the Hermitian part of $\mathcal{J}(\omega)$ contributes to the real part of $\hat{\Sigma}(\omega)$ and, therefore, to the QP energy corrections. Similarly, only the anti-Hermitian term $i\mathcal{B}$ contributes to the imaginary part of the correlated self-energy.

It is worth noting that also $\hat{\Sigma}_c(\omega)$ presents an integrable Coulomb singularity which—in the present implementation—is treated by employing the spherical integration as described in [56]. Since the singularity for $|\mathbf{q} + \mathbf{G}| \rightarrow 0$ in $\hat{\Sigma}_c(\omega)$ is expected to be milder than that of $\hat{\Sigma}_x$, a spherical approximation should be accurate enough, at least in case of isotropic crystalline compounds. On the other hand, for strongly anisotropic systems such as surfaces, the presently available approach is not completely satisfying, and might be a source of numerical error due to the non-analytical behavior of the screened interaction for $|\mathbf{q}| \rightarrow 0$ [80]. The offset- Γ method proposed by Kotani and co-workers in [9] would improve the accuracy of the integration in simple systems but is also faces some difficulties in treating anisotropic systems [11]. A more refined treatment in which the non-analytical behavior of the screened interaction is fully taken into account has been proposed recently in [114], and is currently under development.

Equation 5.92 leads to significant computational savings, depending on the external $\tilde{\mathbf{k}}$ -point, but we still have to find a reliable and efficient approach for evaluating $\mathcal{J}_{\mathbf{G}_1 \mathbf{G}_2}^s(\mathbf{q}, \omega)$. The calculation of the frequency integral presents several difficulties since both the Green's function and the screened interaction have branch cuts in the complex plane infinitesimally below and above the real frequency axis. Consequently, the behavior of the integrand along the real axis is rather ragged, and a straightforward integration is expected to give accurate results only if an exceedingly dense frequency mesh is used. Three different techniques are presently available in the ABINIT code for the evaluation of the frequency dependent part $\mathcal{J}_{\mathbf{G}_1 \mathbf{G}_2}^s(\mathbf{q}, \omega)$:

- 1 Integration with the plasmon-pole technique
- 2 Integration with Contour Deformation (CD)
- 3 Analytic Continuation (AC) from imaginary to real axis via Padé extrapolation

The following sections contain a brief description of the implementation of the three different techniques as well as a discussion of their pros and cons. We start with the discussion of the plasmon-pole approach, presenting the equations that have to be implemented for the four different plasmon models described in section 5.5. We then proceed

to present the equations needed for the contour deformation and the analytic continuation techniques. A comparison of the accuracy and reliability of the CD and the AC method concludes the discussion.

5.7.2.1 Plasmon-Pole

The plasmon-pole technique described in section 5.5 allows one to perform the integration analytically. The explicit expression for $\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega)$ depends on the particular model employed. For the HL and GN plasmon-pole method, inserting the expression for the real part of $\tilde{\epsilon}^{-1}(\mathbf{q}, \omega)$ (Eq. 5.48) in equation 5.86 yields

$$\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega) = \Omega_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q}) \int \frac{e^{i\omega'\delta} d\omega'}{\left(\omega + \omega' - \epsilon_s + i\eta \text{sign}(\epsilon_s - \mu)\right) \left(\omega'^2 - (\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) - i\eta)^2\right)}, \quad (5.95)$$

where the complex shift $i\eta$ has been introduced in order to obtain the time-ordered version of $\tilde{\epsilon}^{-1}(\omega)$. The integration can be carried out, via the residue techniques, by closing the contour in the upper half of the complex plane. The integrand has poles located in the complex plane at:

$$\omega' = \begin{cases} p_1^\pm = \pm(\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) - i\eta) \\ p_2 = \epsilon_s - \omega - i\eta \text{sign}(\epsilon_s - \mu), \end{cases} \quad (5.96)$$

therefore only the contributions given by the pole at p_1^- and the residue at p_2 for $\epsilon_s < \mu$ contribute to the final result. The residues of the integrand are readily obtained

$$\begin{cases} \text{Res}_{p_1^-} = -\left[2\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})(\omega' - \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) - \epsilon_s)\right]^{-1} \\ \text{Res}_{p_2} = \left[(-\omega - \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) + \epsilon_s)(-\omega + \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) + \epsilon_s)\right]^{-1}, \end{cases} \quad (5.97)$$

and the final expression for the frequency dependent part reads

$$\begin{aligned} \mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega) &= 2\pi i \Omega_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q}) \\ &\left[-\frac{1}{2\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})(-\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) + \omega - \epsilon_s)} + \frac{\theta(\mu - \epsilon_s)}{(-\omega - \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) + \epsilon_s)(-\omega + \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) + \epsilon_s)} \right] = \\ &= -\frac{i\pi \Omega_{\mathbf{G}_1\mathbf{G}_2}^2(\mathbf{q})}{\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})} \left[\frac{\theta(\mu - \epsilon_s)}{\omega + \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) - \epsilon_s - i\delta} + \frac{\theta(\epsilon_s - \mu)}{\omega - \tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) - \epsilon_s + i\delta} \right] \end{aligned} \quad (5.98)$$

As expected, the poles of $\Sigma_c(\omega)$ are located infinitesimally above the real axis for $\omega < \mu$ and infinitesimally below for $\omega > \mu$. The poles arising from empty states are at

$\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) + \epsilon_s$, and those coming from occupied states at $-\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) + \epsilon_s$. There are no poles in an interval of width $2\tilde{\omega}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q})$ around the Fermi level μ .

For the von der Linden-Horsch model, the frequency integral is obtained by summing over the plasmon-pole bands:

$$\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega) = \sum_p \Phi_p^0(\mathbf{q}, \mathbf{G}_1) \left[\Phi_p^0(\mathbf{q}, \mathbf{G}_2) \right]^\dagger \int \frac{\lambda_p^{-1}(\mathbf{q}, \omega) e^{i\omega'\delta} d\omega'}{\left(\omega + \omega' - \epsilon_s + i\eta \text{sign}(\epsilon_s - \mu) \right)}. \quad (5.99)$$

As anticipated, unlike the expression for the GN and HL method reported in equation 5.98, here the integral $\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega)$ is given by the direct product of two vectors rather than a full matrix. The expression for the frequency dependent part in the Engel-Farid plasmon-pole model has a very similar mathematical structure:

$$\begin{aligned} \mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q}, \omega) = 4\pi \sum_p \frac{y_p(\mathbf{q}, \mathbf{G}_1)}{|\mathbf{q} + \mathbf{G}_1|} \frac{y_p^\dagger(\mathbf{q}, \mathbf{G}_2)}{|\mathbf{q} + \mathbf{G}_2|} \times \\ \times \int \frac{e^{i\omega'\delta} d\omega'}{\left(\omega + \omega' - \epsilon_s + i\eta \text{sign}(\epsilon_s - \mu) \right) \left(\omega'^2 - (\tilde{\omega}_p(\mathbf{q}) - i\eta)^2 \right)} \end{aligned} \quad (5.100)$$

Thanks to the similarity between the mathematical expression of the VLH and EF plasmon-model, the expression for the matrix elements of $\hat{\Sigma}_c(\omega)$ can be recast in a single compact formula. The final result reads:

$$\langle b_1 \tilde{\mathbf{k}} | \hat{\Sigma}_c(\omega) | b_2 \tilde{\mathbf{k}} \rangle = \frac{1}{V} \sum_{\mathbf{q}} \sum_{n=1}^{+\infty} \sum_p \alpha_p(\mathbf{q}) \frac{\beta_p^{nb_1}(\tilde{\mathbf{k}}, \mathbf{q}) \left[\beta_p^{nb_2}(\tilde{\mathbf{k}}, \mathbf{q}) \right]^\dagger}{\omega - \epsilon_{n\mathbf{k}-\mathbf{q}} + \omega_p(\mathbf{q}) \text{sign}(\mu - \epsilon_{n\mathbf{k}-\mathbf{q}})}, \quad (5.101)$$

where $\alpha_p(\mathbf{q}) = 2\pi z_p(\mathbf{q})$, $\tilde{\omega}_p(\mathbf{q})$ for the von der Linden-Horsch method and $\alpha_p(\mathbf{q}) = 1$ for the Engel-Farid PPM. In equation 5.101, the matrix element $\beta_p^{mn}(\mathbf{k}, \mathbf{q})$ is defined as

$$\beta_p^{mn}(\mathbf{k}, \mathbf{q}) \equiv \sum_{\mathbf{G}} [M_{\mathbf{G}}^{mn}(\mathbf{k}, \mathbf{q})]^\dagger \Theta_p(\mathbf{q}, \mathbf{G}), \quad (5.102)$$

where the scaled plasmon-pole eigenvectors $\Theta_p(\mathbf{q}, \mathbf{G})$ are given by

$$\Theta_p(\mathbf{q}, \mathbf{G}) \equiv \frac{\Phi_p^0(\mathbf{q}, \mathbf{G})}{|\mathbf{q} + \mathbf{G}|}, \quad (5.103)$$

for the Von der Linden-Horsch model, and by

$$\Theta_p(\mathbf{q}, \mathbf{G}) \equiv \frac{4\pi}{\sqrt{2\omega_p(\mathbf{q})}} \frac{y_p(\mathbf{q}, \mathbf{G})}{|\mathbf{q} + \mathbf{G}|^2} \quad (5.104)$$

for the Engel-Farid approach.

5.7.2.2 Contour Deformation

The contour deformation method was proposed in order to avoid having to deal with quantities close to the real axis as much as possible [115, 9, 64]. The integral over the real frequency axis in equation 5.86 can indeed be transformed into an integral over the contour depicted in blue in figure 5.5, plus contributions coming from the poles of the integrand lying inside the contour:

$$\Sigma_c(\omega) = \frac{i}{2\pi} \left\{ 2\pi i \sum_{z_p}^{\mathcal{C}} \lim_{z \rightarrow z_p} G(z) W^c(z) (z - z_p) - \int_{-\infty}^{+\infty} G(\omega + i\omega') W^c(i\omega') d(i\omega') \right\}. \quad (5.105)$$

In the above equation, the first sum is restricted to the poles lying inside the path $\mathcal{C}$. $W^c(z)$ represents the frequency dependent part of the screened interaction, whose expression in reciprocal space is given by:

$$W^c(\mathbf{q}, \omega)_{\mathbf{G}_1 \mathbf{G}_2} \equiv \tilde{\alpha}_{\mathbf{G}_1 \mathbf{G}_2}^c(\mathbf{q}, \omega) \tilde{v}_{\mathbf{G}_1 \mathbf{G}_2}(\mathbf{q}). \quad (5.106)$$

The integration along the imaginary axis is expected to converge quickly with respect to the number of sampled frequencies. Only the residues of the integrand have to be evaluated at the complex poles contributed by the Green's function whose frequency dependence is known.

In what follows, we report the basic equations used to calculate equation 5.105. To keep the notation as simple as possible, only the contribution given by a single state s is considered. The final result is obtained by summing over all the states entering the definition of the Green's functions, taking into account the spatial dependence arising from the QP amplitudes. Taking advantage of the RPA symmetry $W^c(-i\omega') = W^c(i\omega')$, the contribution of a single state s to the integral along the imaginary axis is calculated according to:

$$\begin{aligned} i \int_{-\infty}^{+\infty} \frac{W_{\mathbf{G}_1 \mathbf{G}_2}^c(\mathbf{q}, i\omega')}{\omega + i\omega' - \epsilon_s} d\omega' &= i \int_0^{+\infty} W_{\mathbf{G}_1 \mathbf{G}_2}^c(\mathbf{q}, i\omega') \left[\frac{1}{\omega + i\omega' - \epsilon_s} + \frac{1}{\omega - i\omega' - \epsilon_s} \right] d\omega' = \\ &= 2i \int_0^{+\infty} \frac{(\omega - \epsilon_s) W_{\mathbf{G}_1 \mathbf{G}_2}^c(\mathbf{q}; i\omega')}{(\omega - \epsilon_s)^2 + \omega'^2} d\omega'. \end{aligned} \quad (5.107)$$

The total contribution from all poles is instead given by [41]:

$$\sum_{z_p}^{\mathcal{C}} [\dots] = \sum_s \left[-\theta(\mu - \epsilon_s) \theta(\epsilon_s - \omega) + \theta(\omega - \epsilon_s) \theta(\epsilon_s - \mu) \right] \tilde{W}^c(|\epsilon_s - \omega| - i\eta). \quad (5.108)$$

Notice that the frequency integral in equation 5.107 is Hermitian in reciprocal space, as it involves the integration of the screened interaction along the imaginary axis. As a

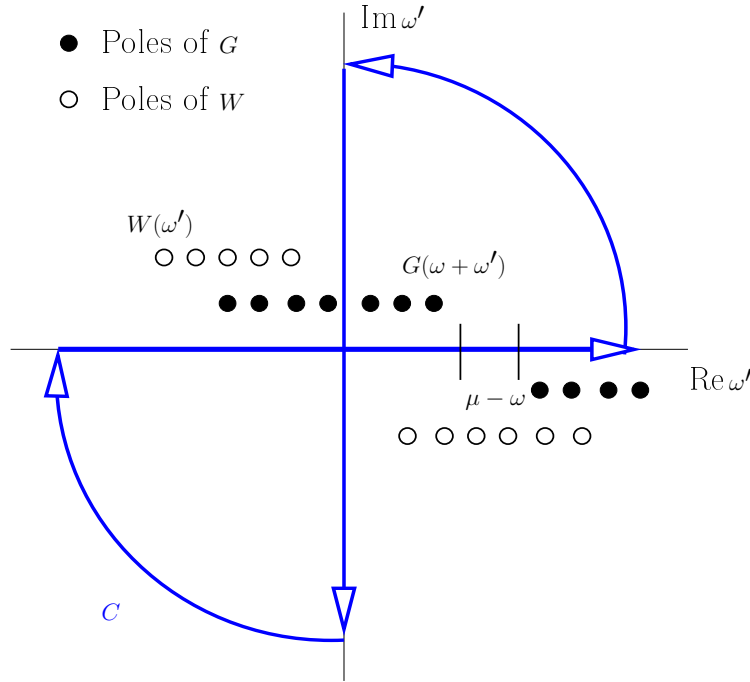


Figure 5.5: Schematic representation of the contour of integration in the complex ω' plane used to evaluate $\mathcal{J}_{\mathbf{G}_1\mathbf{G}_2}^s(\mathbf{q},\omega)$. The poles of the integrand are shown as circles. Only the poles due to Green's function that lie inside the path contribute to the final result.

consequence, equation 5.107 gives a contribution only to the real part of the self-energy, and therefore to the real part of the QP energies. The imaginary part of the self-energy comes from the contribution given by the poles along the real axis, more precisely, from the anti-Hermitian part of equation 5.108. The computational effort required by the CD method is obviously larger than that required by the PPM technique since a high resolution of ω' is required in order to allow an accurate estimate of the contribution given by the poles. Nevertheless, the contour deformation technique is more accurate than the PPM approximation, and is preferable, especially if quasiparticle corrections for low-lying states are required or if the system under study contains *d*- or *f*-electrons, where the use of the PPM approach is questionable. Another, distinct, advantage of the CD method over PPM techniques is that the CD method gives access to the imaginary part of the self-energy and it is therefore possible to obtain the matrix elements of spectral function between Bloch states according to [2]:

$$\langle \Psi_{m\mathbf{k}} | \hat{A}(\omega) | \Psi_{m\mathbf{k}} \rangle = \frac{1}{\pi} \frac{|\text{Im} \langle \Psi_{m\mathbf{k}} | \Sigma(\omega) | \Psi_{m\mathbf{k}} \rangle|}{\left[\omega - \epsilon_{m\mathbf{k}} - \text{Re} \langle \Psi_{m\mathbf{k}} | \Sigma(\omega) | \Psi_{m\mathbf{k}} \rangle \right]^2 + \left[\text{Im} \langle \Psi_{m\mathbf{k}} | \Sigma(\omega) | \Psi_{m\mathbf{k}} \rangle \right]^2}. \quad (5.109)$$

5.7.2.3 Analytic Continuation

Another possible approach to the calculation of $\Sigma_c(\omega)$ is to evaluate the matrix elements of the correlation part for a set of purely imaginary frequencies. The self-energy at imaginary frequency is then analytically continued to the real axis by fitting the obtained values to a Padé approximant. Special care has to be taken so that the branch cuts of $\Sigma(\omega)$ are not crossed while performing the analytic continuation. This method was originally proposed by Godby and co-workers in [116, 117, 118], and is routinely employed to calculate quasiparticle energies using the formulation of the *GW* equations in real space and imaginary time [119, 120, 90]. The analytic continuation of the self-energy along the imaginary axis is obtained by performing an integration along the contour depicted in figure 5.6, so that the rather ragged behavior of $G(\omega)$ and $W(\omega)$ is completely avoided, and there is no need to take precautions for handling the poles.

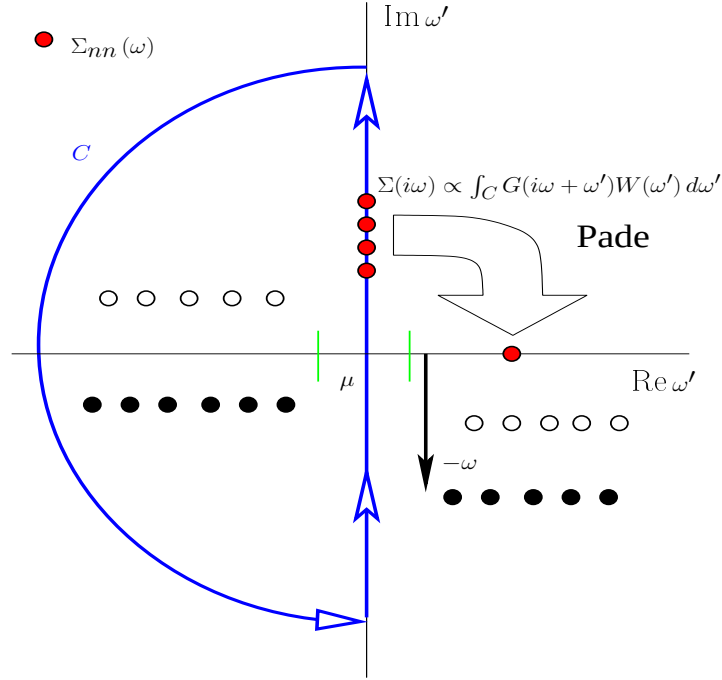


Figure 5.6: Schematic representation of the contour of integration in the complex ω' plane for the frequency integral defining the analytic continuation of $\Sigma(\omega)$ to imaginary frequencies. The contour is chosen such that it encloses the poles to the left of the imaginary axis of both $W(\omega)$ and $G(\omega)$. The energy zero is chosen to be in the middle of the band gap. The poles of the integrand are shown as circles.

Using that, in the RPA, $W(-i\omega) = W(i\omega)$, the self-energy at imaginary frequency can be obtained as follows:

$$\begin{aligned}
\Sigma_c(\mathbf{r}_1, \mathbf{r}_2, i\omega) &= \frac{i^2}{2\pi} \int_0^{+\infty} \left[G(\mathbf{r}_1, \mathbf{r}_2, i\omega + i\omega') + G(\mathbf{r}_1, \mathbf{r}_2, i\omega - i\omega') \right] \tilde{W}^c(\mathbf{r}_1, \mathbf{r}_2, i\omega') d\omega' = \\
&= -\frac{2}{V} \sum_s \sum_{\mathbf{q}} \frac{e^{i(\mathbf{q}+\mathbf{G}_1)\cdot\mathbf{r}_1} \Psi_s(\mathbf{r}_1)}{|\mathbf{q} + \mathbf{G}_1|} \frac{e^{-i(\mathbf{q}+\mathbf{G}_2)\cdot\mathbf{r}_2} \Psi_s^\dagger(\mathbf{r}_2)}{|\mathbf{q} + \mathbf{G}_2|} \times \\
&\quad \times \underbrace{\int_0^{+\infty} \tilde{W}_{\mathbf{G}_1\mathbf{G}_2}^c(\mathbf{q}, i\omega') \left(\frac{1}{i(\omega + \omega') - \epsilon_s} + \frac{1}{i(\omega - \omega') - \epsilon_s} \right) d\omega'}_{\mathcal{A}_{\mathbf{G}_1\mathbf{G}_2}^s(\omega)}. \quad (5.110)
\end{aligned}$$

The integrand entering the definition of $\mathcal{A}_{\mathbf{G}_1\mathbf{G}_2}^s(\omega)$ is rather smooth, and the integration can be performed accurately by using a standard Gauss-Legendre quadrature scheme [86]. First, the indefinite integral is transformed to an integration over a finite domain, by making the standard variable substitution:

$$\begin{cases} z = \frac{1}{x+1} \\ \int_0^{+\infty} f(x) dx = \int_0^1 f\left(\frac{1}{z} - 1\right) \frac{dz}{z^2}, \end{cases} \quad (5.111)$$

then the integration is approximated by a weighted sum over a Gauss-Legendre mesh with weights w_α :

$$\begin{aligned}
\mathcal{A}_{\mathbf{G}_1\mathbf{G}_2}^{n\mathbf{k}-\mathbf{q}}(\omega) &= \int_0^{+\infty} \tilde{W}_{\mathbf{G}_1\mathbf{G}_2}^c(\mathbf{q}, i\omega') \times \left[\frac{1}{i(\omega + \omega') - \epsilon_{n\mathbf{k}-\mathbf{q}}} + \frac{1}{i(\omega - \omega') - \epsilon_{n\mathbf{k}-\mathbf{q}}} \right] d\omega' = \\
&= 2 \sum_{\alpha} \frac{w_\alpha}{z_\alpha^2} \tilde{W}_{\mathbf{G}_1\mathbf{G}_2}^c(\mathbf{q}, i(\frac{1}{z_\alpha} - 1)) \times \frac{i\omega - \epsilon_{n\mathbf{k}-\mathbf{q}}}{(i\omega - \epsilon_{n\mathbf{k}-\mathbf{q}})^2 + (\frac{1}{z_\alpha} - 1)^2}. \quad (5.112)
\end{aligned}$$

The advantages of the analytic continuation with respect to the contour deformation method is that the integration is performed only along the imaginary axis, where the integrand is well behaved and therefore a few points should be enough. Another, distinct, advantage is that the method theoretically gives access to the frequency dependence of the self-energy. This implies that the quasiparticle equation given in 4.53 can be solved by employing standard root finding algorithms (*e.g.* the Newton-Raphson algorithm [86]) without the need of performing a linear expansion around the KS energy. Besides, the analytic continuation represents the ideal approach to the evaluation of the spectral function defined in equation 5.109, because the frequency dependence of $\Sigma(\omega)$ can be obtained with a considerable reduction in computational effort once the Padé coefficients are known. The main drawback of the AC method is related to the validity of the Padé extrapolation. The self-energy has two different branch cuts in the complex plane, and the Padé rational expression, – a function with only a finite number of poles – is able to provide an accurate description of the ω -dependence of $\Sigma(\omega)$ only in a limited region around the band gap, whose extent cannot, unfortunately, be predicted *a priori*.

5.8 Contour Deformation vs Analytic Continuation

To validate the reliability of the AC approach, we have performed test calculations for two prototype compounds: Si and SiC. The two tables 5.2 and 5.4 display a comparison of different QP quantities obtained in the two approaches: the value of the correlation part evaluated at the KS energy, $\Sigma_c(\epsilon_{\text{KS}})$, the renormalization factor, Z , and the QP corrections to the LDA eigenvalues *i.e.* $\Delta^{\text{QP}} = \epsilon^{\text{QP}} - \epsilon^{\text{KS}}$. To avoid possible spurious effects due to slow convergence, all the calculations have been performed with over-converged parameters reported in the caption of each table.

Comparing the results, for the description of the states close to the gap, we notice that CD and AC are in good agreement with each other as the difference between the direct gaps is always smaller than 0.1 eV. Far from the chemical potential the agreement is somewhat less satisfying, albeit still reasonable. In the region far from the gap, the AC approach is not as reliable as the CD integration, since the renormalization factor Z obtained after having performed the analytical continuation departs from results obtained in other independent implementations. In this particular energy region, the Padé extrapolation, turns out to be quite sensitive to the parameters (the maximum imaginary frequency and the number of sampled points) used for evaluating $\Sigma_c(i\omega)$ along the imaginary axis.

To investigate the problem further, it is useful to make a comparison between the real part of $\Sigma_c(\omega)$ obtained with the CD and the AC approach for selected states at high symmetry $\mathbf{k}$ -points. The results for Si and SiC are reported in figure 5.7. For energy states close to the gap, the AC results are very close to the values obtained with the CD method within a frequency window of around 10 eV around the reference KS energy. This is enough for an accurate evaluation of the QP energies in the AC approach since only the knowledge of $\Sigma_c(\epsilon^{\text{KS}})$ and of its first derivative is needed. On the other hand, the results obtained with the AC approach are not sufficiently accurate for the calculation of the spectral function, which requires a good description of the frequency dependence of $\Sigma_c(\omega)$ over a wide range of frequencies. For low-lying states, $\Sigma_c(\omega)$ calculated with the AC approach shows unphysical oscillations (left panel in Fig. 5.7), and the value extrapolated at the KS energy departs significantly from the corresponding CD results. These spurious oscillations can be traced back to a non perfect convergence with respect to the maximum frequency in $\Sigma_c(i\omega)$ sampled along the imaginary axis. As a rule of thumb, indeed, a reliable Padé extrapolation at a real frequency $\tilde{\omega}$ requires the evaluation of $\Sigma_c(i\omega)$ at least up to $2i\tilde{\omega}$. This explains the discrepancies observed in Si and SiC for states far from the gap. We therefore conclude that the AC method is quite efficient and reliable for calculating band gaps where few imaginary points are enough to converge the results, whereas additional computational improvements are needed to render the fit more stable and efficient in the region far from the gap. A possible solution might be to perform a nonlinear fit using a multipole model function with a possible additional decaying exponential term as fitting expression. This approach is commonly employed in the space and time formulation of the GW equations [90, 91] and is presently under development.

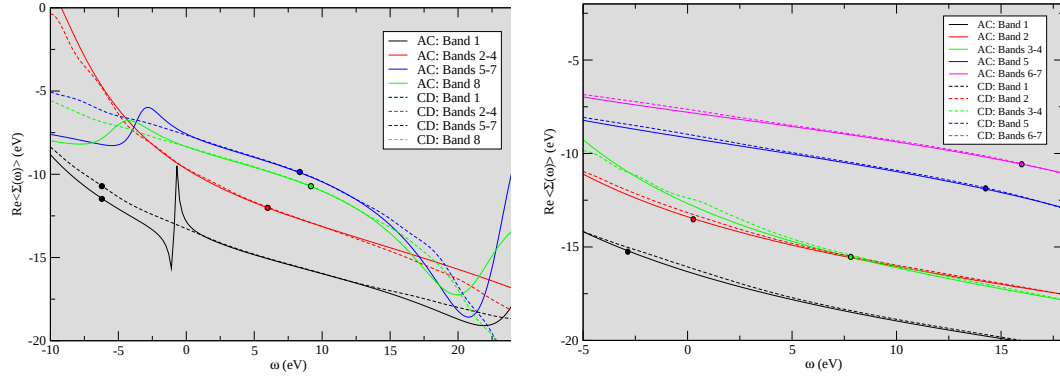


Figure 5.7: Left panel: Real part of $\Sigma_c(\omega)$ in Si as function of ω at $\mathbf{k} = \Gamma$. The dots indicate the position of the KS energies used in the standard perturbative method for the linear expansion of the self-energy and the evaluation of the Z factor. Note the oscillations for the first band at Γ introduced by the Padé extrapolation. Right panel: Same as the left panel, but relative to the self-energy of SiC at $\mathbf{k} = L$. No oscillations are observed in the AC curves but the frequency dependence of $\Sigma_c(\omega)$ for the first band departs significantly from that obtained using the CD approach.

$\mathbf{k}$	CD $\Sigma_c(\epsilon_{KS})$	AC $\Sigma_c(\epsilon_{KS})$	CD Z	AC Z	CD Δ^{QP}	AC Δ^{QP}
Γ_{1v}	6.528	6.917	0.642	0.579	-0.180	0.103
Γ'_{25v}	0.493	0.439	0.764	0.763	-0.536	-0.574
Γ_{15c}	-4.060	-4.105	0.760	0.765	0.122	0.088
Γ'_{2c}	-4.672	-4.710	0.750	0.758	0.073	0.045
O_g	3.185	3.189				
X_{1v}	4.480	4.396	0.686	0.665	-0.330	-0.367
X_{4v}	1.711	1.644	0.733	0.745	-0.529	-0.583
X_{1c}	-3.613	-3.656	0.782	0.780	0.150	0.116
X_{4c}	-6.943	-7.002	0.713	0.682	-0.127	-0.163
O_g	4.182	4.202				
L'_{2v}	5.380	5.495	0.680	0.628	-0.300	-0.190
L_{1v}	3.958	3.817	0.689	0.685	-0.337	-0.425
L'_{3v}	1.001	0.948	0.754	0.757	-0.551	-0.591
L_{1c}	-3.863	-3.910	0.771	0.772	0.102	0.066
L_{3c}	-4.381	-4.417	0.759	0.767	0.177	0.152
O_g	3.342	3.345				

Table 5.2: Comparison of the QP results obtained in Si using the CD and the AC method (all in eV). The calculation have been performed with a $6 \times 6 \times 6$ $\mathbf{k}$ -mesh, 250 bands for the sum in G and W , a cut-off of 4 Hartree for W , 8 Hartree for the wavefunctions, 6 Hartree for Σ_x . We used 50 Gauss-Legendre points for the AC integration and 12 frequency points for $\Sigma(i\omega)$ with a spacing of 1 eV. For the CD method, the frequency integral has been evaluated with 50 points along the imaginary axis and 75 points along the real axis on an homogeneous frequency mesh up to 75 eV. O_g denotes the direct gap.

k	<i>CD</i> $\Sigma_c(\epsilon_{\text{KS}})$	<i>AC</i> $\Sigma_c(\epsilon_{\text{KS}})$	<i>CD</i> <i>Z</i>	<i>AC</i> <i>Z</i>	<i>CD</i> Δ^{QP}	<i>AC</i> Δ^{QP}
Γ_{1v}	7.515	7.667	0.703	0.657	-0.915	-0.700
Γ_{15v}	1.539	1.443	0.788	0.796	-0.718	-0.798
Γ_{1c}	-5.341	-5.387	0.776	0.789	0.218	0.184
Γ_{15c}	-5.107	-5.122	0.795	0.811	0.327	0.320
O_g	0.936	0.982				
X_{1v}	6.128	5.881	0.732	0.706	-0.916	-1.089
X_{3v}	4.255	4.042	0.730	0.743	-0.851	-1.013
X_{5v}	2.447	2.363	0.783	0.777	-0.897	-0.966
X_{1c}	-3.840	-3.884	0.799	0.822	0.280	0.244
O_g	5.698	5.732				
L_{1v}	6.521	6.354	0.732	0.694	-0.964	-0.997
L_{1v}	4.500	4.295	0.730	0.738	-0.844	-0.987
L_{3v}	1.842	1.760	0.783	0.784	-0.808	-0.869
L_{1c}	-4.848	-4.903	0.799	0.805	0.283	0.240
L_{3c}	-5.076	-5.096	0.798	0.813	0.434	0.425
O_g	1.091	1.110				

Table 5.4: Comparison of the QP results obtained in SiC using the CD and the AC method (all in eV). The calculation has been performed with a $6 \times 6 \times 6$ **k**-mesh, 250 bands for the sum in G and W , a cut-off 4 Hartree for W , 8 Hartree for the wavefunctions and 6 Hartree for Σ_x . We used 50 Gauss-Legendre points for the AC integration and 12 frequency points for $\Sigma(i\omega)$ with a spacing of 1 eV. For the CD method, the frequency integral has been evaluated with 50 points along the imaginary axis and 75 points along the real axis on an homogeneous frequency mesh up to 75 eV. O_g denotes the direct gap.

5.9 QPSCGW using Kohn-Sham states as basis set

As discussed in section 4.10, the *QPSCGW* approach of Faleev and co-workers [16, 59, 11] requires the diagonalization of the static and Hermitian approximation to the self-energy, $\tilde{\Sigma}$, given in equation 4.62. Diagonalizing $\tilde{\Sigma}$ in the reciprocal space representation would not be practical as the calculation of the self-energy scales with the fourth power of the number of atoms [56]. The computational effort can be reduced significantly if a finite number of Kohn-Sham eigenstates is employed to describe the quasiparticle amplitudes. In most cases, the KS wavefunctions are not so different from the QP amplitudes, hence, KS states are expected to constitute an accurate basis set for the solution of the problem. Following this idea, the quasiparticle amplitudes, $|\Phi_{m\mathbf{k}}\rangle$, are expanded in terms of KS states, $|\Psi_{n\mathbf{k}}\rangle$, according to:

$$|\Phi_{m\mathbf{k}}\rangle = \sum_n \mathcal{U}_{mn}^{\mathbf{k}} |\Psi_{n\mathbf{k}}\rangle, \quad (5.113)$$

where $\mathcal{U}^{\mathbf{k}}$ is the unitary matrix obtained by diagonalizing the approximated self-energy $\tilde{\Sigma}$. Using a Hermitian and static approximation instead of the true self-energy solves several mathematical issues associated to the calculation of $\hat{\Sigma}(\omega)$ which, being non-Hermitian and energy-dependent, should have left and right energy-dependent eigenvectors. Nevertheless, *QPSCGW* is still challenging and much more demanding than the standard G_0W_0 approach because both diagonal and off-diagonal matrix elements of $\hat{\Sigma}(\omega)$ for all bands included in the basis set have to be explicitly calculated. Usually, all occupied bands are included in the basis unless the system contains deep core states that are supposed to maintain atomic-like character. An additional set of empty states has to be included in order to improve the accuracy of the basis set expansion. Note that the set of empty states should be chosen such that all the degeneracies at each $\mathbf{k}$ -point are included, because this assures that the Kohn-Sham basis set will have the correct symmetry. Since the cost of the calculation of a single matrix element of $\hat{\Sigma}$ scales with the third power of the number of atoms, and since the number of basis set functions increases linearly with the size of the system, the overall scaling of *QPSCGW* is of the order of N_{at}^5 and thus even more costly than the calculation of $\chi^0(\omega)$.

Group theory can, as usual, help reduce the computational effort because $\Sigma(\mathbf{r}_1, \mathbf{r}_2)$ has the symmetry of the crystal, and strict selection rules can be derived for its matrix elements. The invariance of $\Sigma(\mathbf{r}_1, \mathbf{r}_2)$ under any translation of the Bravais lattice leads to the first selection rule already implicitly employed in the preceding equations: matrix elements between two wavefunctions of different $\mathbf{k}$ are null. Group theory permits one to derive another important property: the self-energy cannot connect states belonging to different irreducible representations [93]. An equivalent statement would be that only Bloch states with the same symmetry can enter the linear expansion given in equation 5.113.

In the KS representation, the $\hat{\Sigma}$ operator is thus given by a sparse matrix and, as a consequence, only matrix elements between states with same symmetry have to be explicitly calculated. A significant time saving can hence be achieved by just classifying

the initial KS wave functions according to the irreducible representation they belong to. As a side effect, all densities and potentials that have to be self-consistently calculated during the solution of the *QPSCGW* equations will preserve the correct symmetry during the self-consistent iteration, thus reducing the occurrence of spurious degeneracy lift introduced by round-off errors or numerical instabilities.

The irreducible representation associated to a set of degenerate states at a high symmetry $\mathbf{k}$ -point can be obtained as follows: The application of a symmetry operation on a Bloch wave function with momentum $\mathbf{k}$ belonging to the set of degenerate bands $\mathcal{C}_{n\mathbf{k}}$, produces a new eigenstate with the same energy but different crystalline momentum $\mathcal{R}\mathbf{k}$ (see also Eq. B.8). This new eigenstate is given by an appropriate linear combination of the eigenstates of the Hamiltonian with momentum $\mathcal{R}\mathbf{k}$ according to:

$$\hat{\mathcal{R}}_{\mathbf{t}} \Psi_{n\mathbf{k}}(\mathbf{r}) \equiv \Psi_{n\mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = \sum_{\alpha \in \mathcal{C}_{n\mathbf{k}}} D_{\alpha n}(\mathcal{R}) \Psi_{\alpha\mathcal{R}\mathbf{k}}(\mathbf{r}). \quad (5.114)$$

It readily follows that, if $\hat{\mathcal{R}}_{\mathbf{t}}$ belongs to the little group of $\mathbf{k}$ *i.e.* $\mathcal{R}\mathbf{k} = \mathbf{k} + \mathbf{G}_{\mathcal{R}}$, then:

$$D_{\alpha\beta}(\mathcal{R}_{\mathbf{t}}) = e^{-i(\mathbf{k} + \mathbf{G}_{\mathcal{R}}) \cdot \mathbf{t}} \int e^{i\mathbf{G}_{\mathcal{R}} \cdot \mathbf{r}} u_{\alpha\mathbf{k}}^{\dagger}(\mathbf{r}) u_{\beta\mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) d\mathbf{r}. \quad (5.115)$$

It is possible to demonstrate [93] that the set of matrices $D_{\alpha\beta}(\mathcal{R}_{\mathbf{t}})$ in equation 5.115 forms an irreducible representation of the little group of $\mathbf{k}$ and they provides all the information needed to classify the electronic states of the Hamiltonian at the given $\mathbf{k}$ -point.⁶

⁶Additional information is, however, needed if $\mathbf{k}$ is at the zone border, and the little group of $\mathbf{k}$ contains non-symmorphic operations.

Chapter 6

GW Calculations in Silicon Dioxide

6.1 On the Reliability of the Different Plasmon-Pole Models

This section presents a comparative study of the accuracy of the four different plasmon-pole models described in section 5.5. Understanding the accuracy and the reliability of the different plasmon-pole models is important, since it allows us to select the most accurate (and possibly most efficient) approximation to use for production calculations in complicated systems. To the best of our knowledge, most of the comparative studies reported so far have focused on the description of the fundamental gap while it would be interesting to know the accuracy of the different PPM techniques over a wider range of energies. This is especially important if, for instance, the PPM approximation is used during the first iterations of the *QPSCGW* method in order to speed up the approach to the self-consistent solution. In this case, the costly numerical integration would be used only for the last few steps when most of the QP energies and QP amplitudes are already close to convergence.

The reliability of the different models is analyzed by comparing the relative position of the QP energies versus the corresponding results obtained with the contour-deformation technique described in section 5.7.2.2. The four different prototype compounds Si, C in the diamond structure, SiC are considered to be representative of typical *s* and *p* compounds. For each system, we report the mean relative error (MRE) and the mean absolute relative error (MARE). All calculations have been performed employing standard norm-conserving pseudopotentials. The input parameters, reported in the caption of each table, ensure the numerical convergence of the QP energy differences to within 0.05 eV.

Our findings for the investigated systems can be summarized as follows. For the description of the gaps and the relative position of the QP energies in the region close to the chemical potential, the plasmon-pole model of Godby-Needs turns out to be in good agreement with the results obtained with the contour deformation technique. The agreement worsens for the low-lying states, especially at the bottom of the conduction bands where the largest discrepancies are observed, but this is a well known deficiency of any plasmon-pole based calculation. Despite the difference at the bottom of the conduction

bands, the overall agreement between GN and CD results is good as evidenced by the values of the MARE and MRE, which are systematically smaller than the corresponding results obtained with the other three plasmon-pole models. Next comes the Engel-Farid approach which, although not as accurate as the GN method in the gap region, presents the important advantage of being closer to the CD values especially for the bottom of valence bands. The model of Hybertsen-Louie systematically gives QP band gaps larger than the CD results. Since all crystalline compounds considered have small gaps, this overestimate is within the numerical uncertainty of GW due to external factors such as the use of the pseudopotential approximation or the core-valence decoupling. However, this overestimation is proportional to the initial KS gap size and it is expected to be larger in the case of wide gap insulators. A similar trend was observed also in other compounds, such as SiO_2 , which is discussed in the next section.

k	LDA	GN PPM	HL PPM	vLH PPM	EF PPM	CD
$\Gamma_{1\nu}$	-12.08	-11.44	-11.96	-11.95	-11.84	-11.77
$\Gamma_{25\nu}$	0.0	0.0	0.0	0.0	0.0	0.0
Γ_{15c}	2.53	3.16	3.21	3.20	3.21	3.11
Γ_{2c}	3.25	3.84	3.90	3.90	3.87	3.86
X_{1v}	-7.88	-7.74	-7.97	-7.97	-7.88	-7.70
X_{4v}	-2.91	-2.95	-2.99	-3.00	-2.97	-2.92
X_{1c}	0.59	1.22	1.26	1.26	1.27	1.25
L_{2v}	-9.70	-9.43	-9.72	-9.72	-9.61	-9.49
L_{1v}	-7.08	-6.97	-7.17	-7.17	-7.08	-6.91
L_{3v}	-1.22	-1.25	-1.27	-1.27	-1.25	-1.24
L_{1c}	1.47	2.07	2.10	2.11	2.10	2.09
L_{3c}	3.32	4.00	4.07	4.07	4.07	4.01
MARE	-	1.12%	2.10%	2.14%	1.47%	-
MRE	-	-0.24%	2.11%	2.14%	1.47%	-

Table 6.2: Calculated QP band structure at points of high symmetry for Si (all in eV). The GW calculations have been performed using the four different plasmon-pole models described in the text and the numerical integration of the self-energy using the CD method. The BZ is sampled with a $6 \times 6 \times 6$ $\mathbf{k}$ -mesh, 160 bands for G and W , 8 Hartree for W , 12 Hartree for the wavefunctions, 12 Hartree for Σ_x . For the CD method we used 50 frequencies both along the real and the imaginary axis.

k	LDA	GN PPM	HL PPM	vLH PPM	EF PPM	CD
Γ_{1v}	-21.36	-22.08	-22.38	-22.74	-22.52	-22.16
Γ'_{25v}	0.0	0.0	0.0	0.0	0.0	0.0
Γ_{15c}	5.59	7.29	7.38	7.36	7.39	7.31
Γ'_c	13.50	14.72	14.954	14.97	14.95	14.66
L'_{2v}	-15.50	-16.42	-16.83	-16.80	-16.69	-16.22
L_{1v}	-13.36	-14.06	-14.17	-14.22	-14.07	-13.73
L'_{3v}	-2.79	-2.96	-2.99	-2.99	-2.97	-2.95
L_{1c}	8.45	10.27	10.40	10.37	10.39	10.29
L_{3c}	9.03	10.25	10.39	10.39	10.40	10.28
L'_{2c}	15.52	17.34	17.62	17.59	17.59	17.28
MARE	-	0.65%	1.82%	1.95%	1.63%	-
MRE	-	0.40%	1.82%	1.95%	1.63%	-

Table 6.4: Calculated QP bands structure at points of high symmetry for diamond (all in eV). The GW calculations have been performed using the four different plasmon-pole models described in the text and the numerical integration of the self-energy using the CD method. The BZ is sampled with $6 \times 6 \times 6$ $\mathbf{k}$ -mesh, 160 bands for G and W , 8 Hartree for W , 20 Hartree for the wavefunctions, 20 Hartree for Σ_x . For the CD method we used 50 frequencies both along the real and the imaginary axis.

k	LDA	GN PPM	HL PPM	vLH PPM	EF PPM	CD
Γ_{1v}	-15.39	-15.35	-15.82	-16.08	-15.76	-15.58
Γ_{15v}	0.0	0.0	0.0	0.0	0.0	0.0
Γ_{1c}	6.32	7.18	7.26	7.30	7.32	7.26
Γ_{15c}	7.21	8.17	8.37	8.23	8.37	8.25
X_{1v}	-10.22	-10.58	-10.83	-10.92	-10.83	-10.47
X_{3v}	-7.84	-8.18	-8.26	-8.43	-8.19	-7.97
X_{5v}	-3.21	-3.44	-3.45	-3.51	-3.42	-3.39
X_{1c}	<u>1.31</u>	<u>2.21</u>	<u>2.35</u>	<u>2.21</u>	<u>2.38</u>	<u>2.31</u>
X_{3c}	4.18	5.21	5.36	5.21	5.39	5.30
X_{5c}	13.84	14.91	15.37	15.28	15.24	14.92
L_{2v}	-11.72	-11.98	-12.29	-12.43	-12.28	-11.97
L_{1v}	-8.60	-8.91	-9.00	-9.20	-8.93	-8.72
L_{3v}	-1.06	-1.16	-1.16	-1.18	-1.16	-1.15
L_{1c}	5.38	6.28	6.40	6.34	6.44	6.38
L_{3c}	7.15	8.22	8.43	8.29	8.43	8.30
L_{1c}	9.94	11.14	11.43	11.35	11.38	11.12
MARE	-	1.37%	1.94%	2.71%	1.87%	-
MRE	-	-0.24%	1.94%	1.78%	1.87%	-

Table 6.6: Calculated QP bands-structure at points of high symmetry for β -SiC (all in eV). The GW calculations have been performed using the four different plasmon-pole models described in the text and the numerical integration of the self-energy using the CD method. The BZ is sampled with a $6 \times 6 \times 6$ **k**-mesh, 160 bands for G and W , 8 Hartree for W , 20 Hartree for the wavefunctions, 20 Hartree for Σ_x . For the CD method we used 50 frequencies both along the real and the imaginary axis.

6.2 GW Results for α -SiO₂

Silicon dioxide is a wide gap semiconductor of crucial importance for technological applications. It is the central ingredient in metal-oxide semiconductor structures in modern transistors, and is widely used as a raw material for the manufacturing of optical fibers, glasses, ceramics, precision frequency oscillators, etc. Technological applications generally involve the amorphous phase of silicon dioxide, a-SiO₂, and several efforts have been made in the past decades to understand its optical properties, especially in the presence of defects or vacancies. First-principles studies on a-SiO₂ [121] have showed that excitonic effects play an important role. Therefore, a complete description of the optical properties of this compound would require the solution of the Bethe-Salpeter equation [7, 8] which, however, goes beyond the scope of the present analysis.

In the present study, we aim to understand how the electronic properties of silicon dioxide are described at the *GW* level as well as the effect of *GW* self-consistency on the quasiparticle band structure. The second goal is testing the accuracy and reliability of the several approximations involved in the solution of the *GW* equations: the approximation used for $v_{xc}(\mathbf{r})$, the parameters employed to generate the pseudopotentials, and the different plasmon-pole approximations.

Performing a *GW* calculation for the amorphous phase would be challenging, and, moreover, it would complicate the analysis of the different approximations involved in the calculation. For these reasons, we will focus on the simpler crystalline hexagonal phase, α -SiO₂, in the following discussion. Understanding the physics of α -SiO₂ represents an important preparatory step for the first-principles study of the more challenging amorphous phase, which is presently underway in our group.

Ordinary low-temperature alpha quartz is the most stable silica polymorph at ambient conditions up to 3 GPa, and persists as a metastable state at higher pressure. Quartz has a trigonal structure (space group P3₂21). Its hexagonal unit cell contains three SiO₂ molecules whose atoms are arranged in the following reduced coordinates [122]:

$$\text{Si: } (-u, -u, \frac{1}{3}), \quad (u, 0, 0), \quad (0, u, \frac{2}{3}).$$

$$\begin{aligned} \text{O: } \quad & (x, y, z), \quad (y - x, -x, z + \frac{1}{3}), \quad (-y, x - y, z + \frac{2}{3}), \\ & (x - y, -y, -z), \quad (y, x, -z + \frac{2}{3}), \quad (-x, y - x, -z + \frac{1}{3}). \end{aligned}$$

A schematic representation of the position of the atoms in the unit cell is shown in figure 6.1. The crystalline structure of α -SiO₂ has a structure formed by corner shared silicate tetrahedral units, the tetrahedron of oxygen atoms about a silicon atom is almost regular with the Si-O distance being 1.61 Å. Apart from two silicon neighbors, each oxygen has six neighboring adjacent oxygens at distances between 2.60 Å and 2.67 Å.

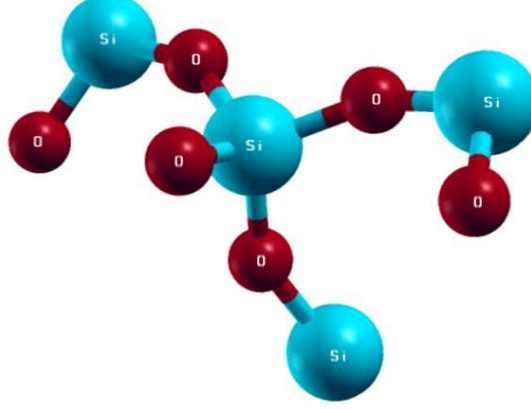


Figure 6.1: Schematic representation of the arrangement of the three SiO_2 molecules in the hexagonal unit cell of $\alpha\text{-SiO}_2$.

As already discussed in section 4.11, reliable many-body calculations require pseudopotentials able to accurately reproduce the scattering properties of the all-electron atoms over a wide range of energies, including empty states. For this reason, particular care has been used to generate the pseudopotentials used in the present work. Figure 6.2 and figure 6.3 summarize the parameters used for the pseudopotential generation, and provide a quantitative analysis of the transferability of the pseudo atoms. A comparison between the all-electron wave functions and the pseudized versions is shown in the left panel of each figure. The right panel of each figure shows a comparison between the logarithmic derivatives of the all-electron and the pseudized atom over a wide range of energies around the reference energy. In order to reduce the plane wave cutoff needed for the description of the nonlocal part of the Hamiltonian as much as possible, the optimized pseudopotential method discussed in [123] is used. Henceforth, our generated pseudopotentials will be referred to as OP pseudopotentials.

From the right panel of figure 6.2, we can see that the scattering properties of the oxygen pseudo-atom are excellent for all the angular channels, up to 2 Ry above the vacuum level. This high degree of transferability is paid for in terms of computing effort because the generated pseudopotential requires around 45 Hartree for the converge structural and electronic properties (not shown). Unfortunately, generating a highly transferable pseudopotential for Si has been proved much more complicated. The quality of the logarithmic derivative of silicon (right panel of Fig. 6.3) was found to depend strongly on the reference atomic configuration as well as on the choice of the angular channel used for the local part of the fully separable form of Kleinman and Bylander [124, 125]. After several attempts, the best transferability, in particular for the d -channel,

was obtained using $[\text{Ne}] 3s^2 3p^1 3d^1$ as the reference atomic configuration (as suggested in [121]). The transferability of the p -channel of Si is not as good as that of the other angular momenta, but it seems it is quite difficult to generate a Si pseudopotential having good transferability over a large energy range without having more than one projector per angular channel. The publicly available pseudopotential generators for ABINIT are not able to generate multi-projector pseudopotentials.

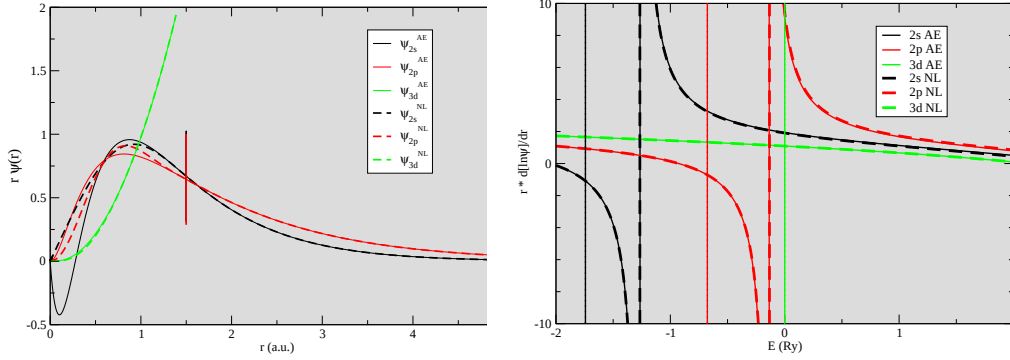


Figure 6.2: (Original in color) Left panel: All-electron wave functions (solid line) and corresponding pseudized images (dashed line) used to generate the norm-conserving pseudopotential for the oxygen atom used in this work. Vertical bars mark the position of the core radius for the pseudization of the all-electron orbitals. Right panel: The logarithmic derivatives for the three different angular channels calculated at $r = 2.0$ a.u. The d channel is used as the local part to generate the Kleinman-Bylander form. Vertical bars indicate the position of the reference eigenvalue per each l .

The trustworthiness of our OP pseudopotentials was tested by comparing the optimized structural parameters with those calculated using standard Troullier-Martins pseudopotentials available on the ABINIT web site (hereafter referred to as TM pseudopotentials). A comparison of the results is reported in table 6.10, where the values obtained with LDA and GGA functionals are also compared. First of all, we notice that the effect of the approximation used for $v_{xc}(\mathbf{r})$ is pronounced. As already pointed out by Hamann in [126], GGA structural parameters are less accurately predicted with respect to the LDA. Our LDA results lead to a consistently more accurate description of all atomic coordinates, in agreement with previous findings [127, 126], and are in excellent agreement¹ with the structural parameters employed in [121]. Therefore, in the following discussion, only the results obtained using the LDA will be presented and discussed.

Figure 6.4 reports the band structure and the electronic density of states (DOS), calculated at the theoretical LDA lattice parameters. Excitonic effects should be taken

¹ Private communication of S. Ismail-Beigi.

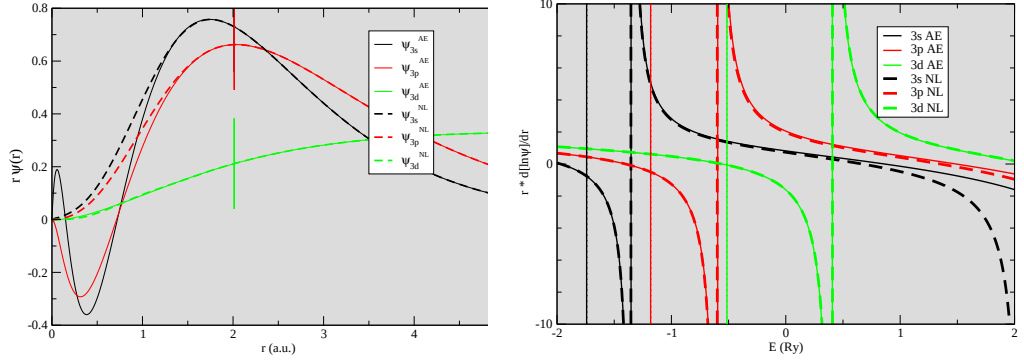


Figure 6.3: (Original in color) Left panel: All-electron wave functions (solid lines) and corresponding pseudized images (dashed lines) used to generate the norm-conserving pseudopotential for the silicon atom used in this work. Vertical bars mark the position of the core radius for the pseudization of the all-electron orbitals. Right panel: The logarithmic derivatives for the three different angular channel calculated at $r = 3.0$ a.u. The d channel is used as the local part to generate of the Kleinman-Bylander form. Vertical bars indicate the position of the reference eigenvalue per each l .

PP / v_{xc}	a	c	u	x	y	z
TM / LDA	9.102	10.067	0.462	0.410	0.281	0.109
OP / LDA	9.161	10.087	0.467	0.413	0.271	0.116
TM / GGA	9.469	10.388	0.477	0.415	0.253	0.130
OP / GGA	9.486	10.404	0.478	0.415	0.251	0.131
Exp	9.284	10.213	0.465	0.415	0.272	0.120

Table 6.10: Theoretical structural parameters of α -SiO₂ for different approximated $v_{xc}(\mathbf{r})$, and different setups used to generate the pseudopotentials. a and c are the parameters of the hexagonal unit cell in atomic units. The parameters u , x , y , and z define the position of the atoms in the unit cell as given in the text. The values have been obtained from symmetry preserving structural relaxations with space group P3₂21. Experimental lattice parameters and atomic positions are from [122].

into account when comparing band structure results with experiment, but nevertheless the 6.4 eV optical gap at Γ is clearly underestimated because the first peak in the ex-

perimental absorption spectrum is located at around 10.3 eV [128]. The analysis of the orbital-projected DOS (not shown) reveals that the valence bands correspond to p -like oxygen states, while the bottom conduction bands have both silicon and oxygen s and p character.

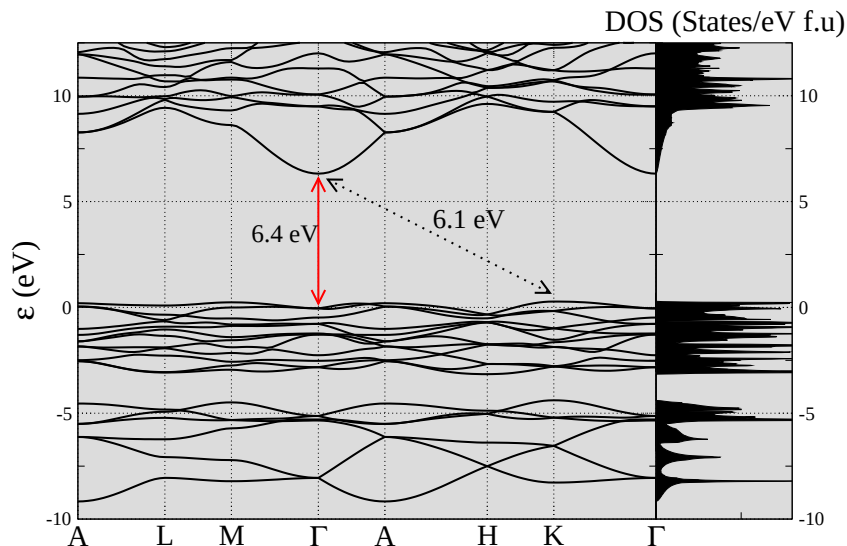


Figure 6.4: Band structure of α -SiO₂ and electronic density of states calculated at the theoretical LDA parameters. The values for the optical gap at Γ and the fundamental gap between Γ and K are displayed.

In what follows, we summarize the main results obtained by performing a GW calculation starting from the LDA results. The QP energies are obtained by using a $6 \times 6 \times 6$ Gamma-centered $\mathbf{k}$ -mesh to sample the full BZ. The screened interaction is expanded in a plane wave basis set with a maximum kinetic energy of 10 Ha. The GW results have been found to converge extremely slowly with respect to the number of empty states. To converge the QP gaps within 0.1 eV accuracy, we had to include 1200 bands in the screening $W(\omega)$ and 800 states in the calculation of $\Sigma(\omega)$. Since the slow convergence of the GW results with respect to the number of empty states is expected to hold also in the amorphous phase, the use of parallel algorithms such as the three levels of parallelization described in [129] or refined techniques such as the extrapolar method recently proposed in [130] will likely be needed to obtain converged QP results in α -SiO₂ with a reasonable computing effort.

The main results of our G_0W_0 calculations for α -SiO₂ are reported in figure 6.5, where the G_0W_0 band structure obtained with the HL and GN plasmon-pole model are displayed and compared to the LDA band dispersions. First of all, we note that G_0W_0 leads to an important opening of the gap: the nearly direct gap at Γ is opened up significantly from the LDA value of 6.4 eV to about 9.5 eV if the GN model is used. It is also quite surprising that the discrepancy of the optical gaps due to the use of the

two plasmon-pole techniques (GN vs HL) is larger than the value commonly cited in the literature. This is usually reported to be around 0.1-0.2 eV for states close to the chemical potential [31]. The optical gap at Γ obtained with the HL PPM (≈ 9.9 eV) is in almost perfect agreement with the results obtained in [121] where the same type of PPM has been employed, but it is in significant disagreement with the values coming from the GN method (≈ 9.5 eV). To clarify this discrepancy, we have performed G_0W_0 calculations using the more refined plasmon-pole models of Von der Linden-Horsch and Engel-Farid described in section 5.5.2. The final results have been checked against the more accurate values obtained by numerically integrating the self-energy using both the contour deformation method (Sect. 5.7.2.2) and the analytic continuation (Sect. 5.7.2.3). Figure 6.6 compares the values of the optical gaps at Γ and A calculated using the different approaches. The AC and CD optical gaps are in excellent agreement with each other and extremely close to the GN PPM results. The optical gaps obtained with the other two plasmon-pole models (vLH and EF) represent an improvement over the GN PPM although they do not have the same accuracy of the GN PPM and systematically overestimate the AC and CD results. The fact that two independent integration schemes for $\hat{\Sigma}(\omega)$ gave results almost identical to the Godby-Needs PPM, leads us to conclude that the plasmon-pole of Hybersten-Louie introduces spurious contributions in the evaluation of the self-energy matrix elements. These spurious effects lead to a fortuitous enlargement of the optical gaps. A possible cause of error in the HL model might be ascribed to the well known problem associated with the occurrence of imaginary plasmon-frequencies. Indeed, we found that around 70% of the HL plasmon frequencies are complex in α -SiO₂, and this despite the fact that the HL PPM implemented in ABINIT has been generalized following the improved version proposed in [101] for systems without spatial inversion. It is important to stress that no imaginary plasmon-frequencies were found during the GN calculation.

To have a more general overview of the accuracy of the GN and HL methods, it is useful to compare the corresponding values of the QP corrections as function of the Kohn-Sham energies. The results are plotted in figure 6.7. It is interesting to note that the Godby-Needs PPM describes both the gap and the absolute position of the quasiparticle energies as well. The agreement between CD and the GN PPM worsen for low-lying states (not shown), but this is a well-know deficiency of any plasmon-pole based calculation. On the contrary, the HL QP corrections are systematically underestimated, with the underestimation being more pronounced for valence states. This produces the widening of the final value for the G_0W_0 gaps observed in our results. We therefore conclude that the Godby-Needs plasmon-pole method represents a reliable and effective approach to calculate QP corrections in silicon dioxide, and that it can be used to study more challenging systems based on α -SiO₂, for instance to calculate the position of defects within the gap. Similar conclusions were also reached in a recent work [131], where the GN PPM has been successfully used for the calculation of band-offsets in materials.

At this point, it is worth investigating how Faleev's self-consistency affects the QP band structure of α -SiO₂. Figure 6.8 compares the G_0W_0 QP band structure with the one obtained by updating both the eigenvalues and eigenfunctions entering the Green's

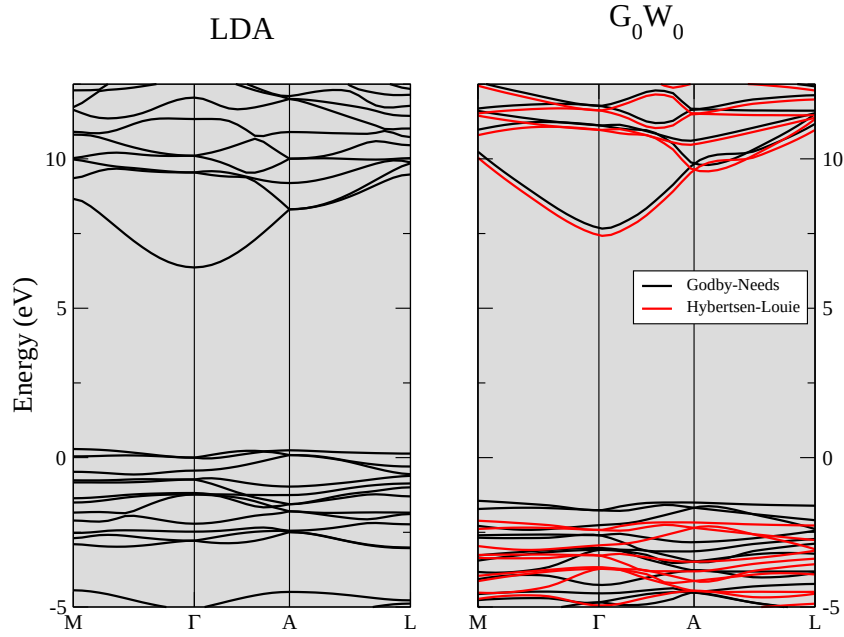


Figure 6.5: (Original in color) Left panel: LDA band structure of α -SiO₂. Right panel: G_0W_0 energy bands of α -SiO₂ calculated using the Godby-Needs (black solid lines) and the Hybertsen-Louie (red solid lines) plasmon-pole model. All calculations have been performed at the theoretical LDA parameters. In both graphics, the zero of energy coincides with the top of the LDA band structure.

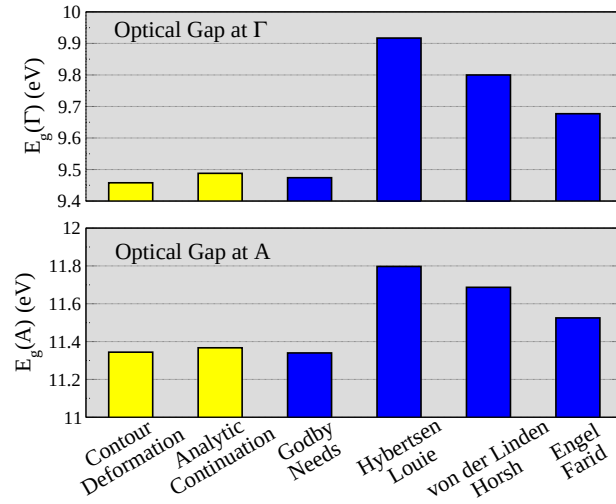


Figure 6.6: Optical gap at Γ and A in α -SiO₂ calculated using the AC, CD and the four different plasmon-pole methods described in section 5.5.

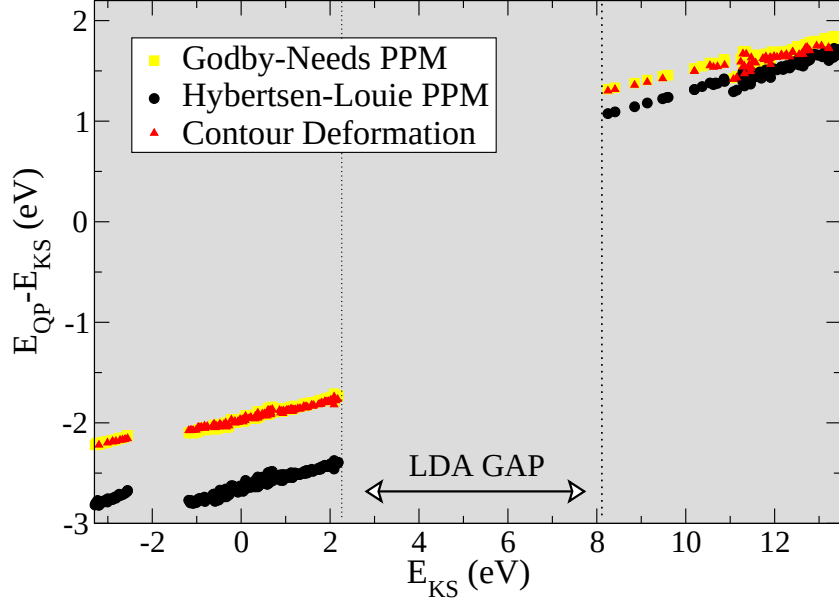


Figure 6.7: G_0W_0 quasiparticle corrections as a function of the Kohn-Sham energy calculated using the CD method (red triangles), the GN (yellow squares) and the HL plasmon-pole model (black circles).

function (GW_0). GW_0 leads to an overall lowering of the valence states while the conduction bands are only slightly affected. This produces a further opening of the gap, which now turns out to be close to the 10.1 eV gap obtained with the standard G_0W_0 method based on the HL plasmon-pole. The dispersion of the bands is slightly affected by the partial self-consistency on G and the calculation of the overlap between QP and GGA wave functions has revealed that, at least for the high symmetry $\mathbf{k}$ -points considered, the GGA orbitals are very close to the converged QP amplitudes. Therefore, most of the opening of the gap can be attributed to the updating of the energies entering the definition of the Green's function. To support this hypothesis, we have also performed the so-called e- GW_0 calculation in which the wave functions have been kept fixed at the GGA level, and results similar to the much more demanding GW_0 approach to self-consistency have been obtained (not shown).

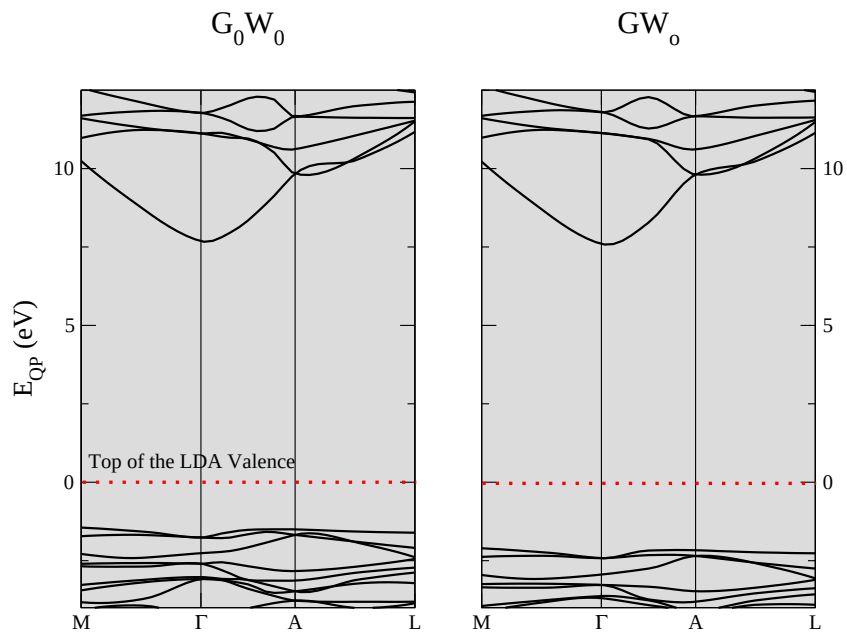


Figure 6.8: Comparison between the QP band structure of α -SiO₂ obtained using the standard one-shot approach, and using Faleev's *QPSCGW* method with only partial self-consistency on the Green's function (both on energies and wavefunctions (GW_0)).

Chapter 7

The GW method within the PAW formalism

7.1 Introduction

In this chapter, we briefly review the projector augmented wave (PAW) formalism introduced and implemented for the first time by Blöchl in 1994 [15]. The material is organized as follows. Section 7.2 gives a brief overview of the main concepts underlying the PAW method, stressing similarities and differences between PAW and other first principles approaches based either on the pseudopotential theory or AE methods. The equations needed to calculate matrix elements of arbitrary local or semi-local operators are discussed in section 7.3. Section 7.4 presents the expression for the different electronic charges that have to be considered due to the particular partitioning of the unit cell used in the PAW approach. The mathematical expression of the PAW Hamiltonian operator is introduced and discussed in section 7.5. The changes needed to implement the G_0W_0 method within PAW are discussed in section 7.6, 7.7, and 7.8. The technical details needed to implement the $QPSCGW$ of Faleev and co-workers are finally gathered in section 7.10.

In the present discussion, we follow the notations introduced in reference [132], whose approach slightly departs from the original formulation given by Blöchl in [15] and other subsequent implementations [133, 134]. The methodological approach followed in the ABINIT code closely follows the one described in reference [132], and is extensively detailed in reference [135].

7.2 The PAW formalism

The projector augmented wave method is designed to efficiently and accurately calculate the electronic properties of materials within the framework of density functional theory [15]. The method takes advantage of several ideas and techniques developed in the past decades both in the pseudopotential and in the all-electron community. From

the pseudopotential approach [63, 70], it inherits the idea of substituting the true Kohn-Sham wave function $\Psi(\mathbf{r})$ with a pseudized image $\tilde{\Psi}(\mathbf{r})$ that can be efficiently expanded with a finite basis set. Similarly to many all-electron (AE) approaches, PAW employs atomic orbitals to describe the all-electron wave function $\Psi(\mathbf{r})$ in the regions close to the atomic cores, thus retaining information about the correct nodal structure of the electronic orbitals. This clearly represents an important advantage with respect to the pseudopotential formalism because *s*-, *p*-, *d*- and *f*-electrons can be treated on the same footing without any significant loss in accuracy. Besides, the use of a localized basis set permits the evaluation of electronic matrix elements using the true wave functions. This enables an accurate description of physical quantities sensitive to variations in the nuclear region such as, for instance, magnetic and optical properties.

The mapping between the true wave function $|\Psi\rangle$ with its complete and complex nodal structure around the nuclei and the fictitious smooth pseudo wave functions $|\tilde{\Psi}\rangle$ is defined by the linear transformation:

$$|\Psi\rangle = \hat{\mathcal{T}} |\tilde{\Psi}\rangle. \quad (7.1)$$

Except for the requirement of linearity, no further constraint on $\hat{\mathcal{T}}$ is required and the mapping can be chosen such that $|\tilde{\Psi}\rangle$ is rapidly convergent once expanded in a finite basis set, in our case plane waves and regular FFT meshes in real space.

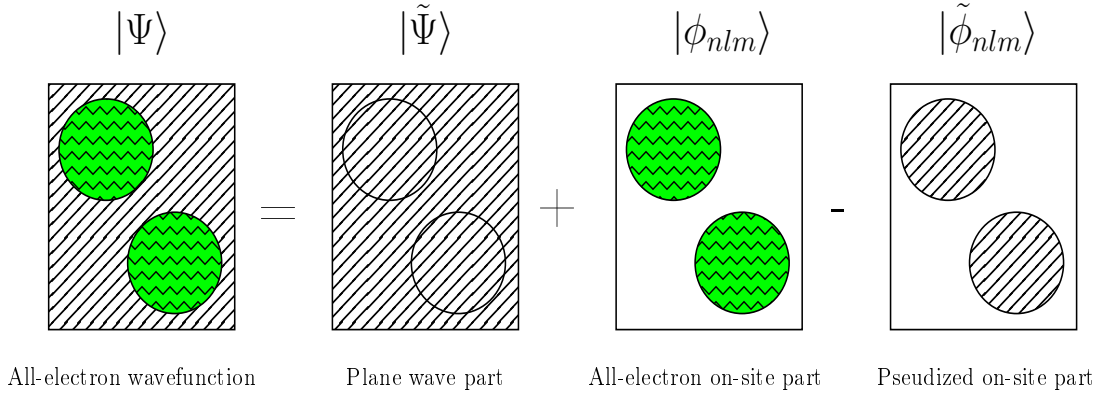


Figure 7.1: Graphical representation of the partitioning of the unit cell into a “smooth” region described with a homogeneous FFT mesh and non-overlapping spheres. Within each sphere, the set of atomic-like orbitals $\{\phi_{nlm}\}$ and $\{\tilde{\phi}_{nlm}\}$ is used as a basis set to describe the onsite true wave function $|\Psi\rangle$ and the associated pseudized image $|\tilde{\Psi}\rangle$, respectively. Inside each augmentation region, the on-site representation of $|\tilde{\Psi}\rangle$ exactly cancels the contribution given on the FFT mesh provided that the basis set $\{\tilde{\phi}_{nlm}\}$ is complete.

Since the true and the pseudized version of the wave function are expected to differ mainly in the atomic region where the Coulomb potential is strong, it is natural to assume that $\hat{\mathcal{T}}$ will differ from the identity operator only in a relatively small portion of space around each ion. The operator $\hat{\mathcal{T}}$ has to modify the smooth auxiliary function $|\tilde{\Psi}\rangle$, in

this region alone, to project out the true wave function $|\Psi\rangle$. Following the augmented wave method [136], the simulation space is thus partitioned into non-overlapping atom-centered spheres and an interstitial region. A schematic representation of the division of the unit cell employed in the PAW method is shown in figure 7.1. The partitioning makes it possible to rewrite the linear transformation in equation 7.1 as the identity operator plus a sum of localized terms, $\hat{\mathcal{T}}_a$, only acting within the atomic spheres Ω_a centered on atomic sites a , that is:

$$\hat{\mathcal{T}} = \hat{1} + \sum_a \hat{\mathcal{T}}_a. \quad (7.2)$$

The linear transformation within each augmentation region Ω_a is defined by specifying a set of functions, $\{\phi_i^a\}$, which form a complete basis set within Ω_a . This set of functions serves as a basis set for the expansion of the true electronic wave function in each augmentation region with as yet unknown coefficients c_i^a :

$$|\Psi\rangle = \sum_i c_i^a |\phi_i^a\rangle \quad \text{in} \quad \Omega_a. \quad (7.3)$$

A possible and natural choice for the basis set $\{\phi_i^a\}$ is the solutions of the radial Schrödinger equation for the isolated atom. In this case the index i is a contracted notation for the atomic position $\mathbf{R}_a$, the angular momentum quantum numbers (l, m) , and an additional index n used to label solutions with different energy. For each AE partial wave $|\phi_i^a\rangle$, an auxiliary pseudo partial wave $|\tilde{\phi}_i^a\rangle$ is defined according to:

$$|\phi_i^a\rangle = (\hat{1} + \hat{\mathcal{T}}_a) |\tilde{\phi}_i^a\rangle, \quad (7.4)$$

from which one obtains:

$$\hat{\mathcal{T}}_a |\tilde{\phi}_i^a\rangle = |\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle. \quad (7.5)$$

In analogy to the norm-conserving [63], or to the ultrasoft scheme [70], the AE partial waves $|\phi_i^a\rangle$ and their auxiliary pseudo counter parts, $|\tilde{\phi}_i^a\rangle$, are therefore identical beyond the radius r_c^a of the augmentation region, that is:

$$\tilde{\phi}_i^a(\mathbf{r}) = \phi_i^a(\mathbf{r}) \quad \text{for} \quad r \geq r_c^a. \quad (7.6)$$

By virtue of the exact cancellation in the interstitial region, it is possible to include unnormalized solutions of the Schrödinger equation in the localized basis set. These correspond to unbound states of the reference atom and are used to improve the flexibility of the basis set expansion. As we will see in the following discussion, the exact cancellation also has important consequences regarding the evaluation of the matrix elements of local or semilocal operators.

In each augmentation sphere, the pseudized wave function, $|\tilde{\Psi}\rangle$, is expanded in terms of smooth partial waves according to:

$$|\tilde{\Psi}\rangle = \sum_i \langle \tilde{p}_i^a | \tilde{\Psi} \rangle |\tilde{\phi}_i^a\rangle \quad \text{in} \quad \Omega_a, \quad (7.7)$$

where the coefficients of the linear combination are obtained by projecting $|\tilde{\Psi}\rangle$ onto some, for the moment unspecified, atom-centered projector functions $|\tilde{p}_i^a\rangle$. Since equation 7.7 has to hold for any arbitrary function, it follows that the PAW projectors must satisfy the relation:

$$\sum_i |\tilde{\phi}_i^a\rangle \langle \tilde{p}_i^a| = \hat{1} \quad \text{in} \quad \Omega_a. \quad (7.8)$$

It is possible to demonstrate [15] that a sufficient condition for the fulfillment of condition 7.8 is that the projectors obey the following orthogonality property:

$$\langle \tilde{p}_i^a | \tilde{\phi}_j^a \rangle = \delta_{ij}. \quad (7.9)$$

Despite this constraint, the choice of the projectors $|\tilde{p}_i^a\rangle$ is quite arbitrary. For instance, neither the projectors nor the partial waves need to be orthogonal among themselves. Consequently, it is possible to exploit all the techniques developed in the past years within the pseudopotential formalism [63, 70, 123, 71, 137], in order to generate a set of smooth auxiliary functions, $\{\tilde{\phi}_i^a\}$, and smooth projectors, $\{\tilde{p}_i^a\}$, which are efficient with a finite plane wave basis set [138]. Combining the equations 7.2, 7.5 and 7.7, one arrives at the final expression for the linear transformation:

$$\hat{\mathcal{T}} = \hat{1} + \sum_a \sum_i \left(|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle \right) \langle \tilde{p}_i^a|. \quad (7.10)$$

The true wave function can hence be expressed as

$$|\Psi\rangle = |\tilde{\Psi}\rangle + \sum_i \sum_a \left(|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle \right) \langle \tilde{p}_i^a | \tilde{\Psi} \rangle = |\tilde{\Psi}\rangle + \sum_a |\Psi_a^1\rangle - |\tilde{\Psi}_a^1\rangle \quad (7.11)$$

where

$$|\Psi_a^1\rangle \equiv \sum_i \langle \tilde{p}_i^a | \tilde{\Psi} \rangle |\phi_i^a\rangle, \quad (7.12)$$

and

$$|\tilde{\Psi}_a^1\rangle \equiv \sum_i \langle \tilde{p}_i^a | \tilde{\Psi} \rangle |\tilde{\phi}_i^a\rangle, \quad (7.13)$$

represent the onsite representation of the true and pseudized wave function inside the augmentation sphere Ω_a . To understand the expression for the true wave function given in equation 7.11, let us concentrate on the different regions used to partition the unit cell as sketched in figure 7.1. In the interstitial region, $\Psi(\mathbf{r}) = \tilde{\Psi}(\mathbf{r})$ by construction since, according to equation 7.6, the partial waves are identical outside the augmentation spheres. Inside the augmentation sphere Ω_a the true wave function is – by virtue of the exact cancellation between the auxiliary wave function and its partial wave expansion – instead given by $\Psi(\mathbf{r}) = \Psi_a^1(\mathbf{r})$, that is, built up from partial waves having the correct nodal structure. It is worth noticing that the pseudo wave function $|\tilde{\Psi}\rangle$ has a physical

meaning only in the interstitial region. Inside the augmentation spheres, $|\tilde{\Psi}\rangle$ is only a computational tool and a bad approximation to the true wave function. Even the norm of the all-electron wave function is not correctly reproduced, as we will see in the next section 7.3.

At this point, only the explicit expressions for the augmentation waves, $|\phi_i^a\rangle$, $|\tilde{\phi}_i^a\rangle$ and the PAW projectors $|\tilde{p}_i^a\rangle$ have to be defined. Due to the spherical symmetry of the reference atomic configuration, the PAW partial waves and the projector functions can be factorized into an angular and radial part according to:

$$\phi_i^a(\mathbf{r}) = \frac{\phi_{n_i l_i}^a(r)}{r} y_{m_i}^{l_i}(\hat{\mathbf{r}}), \quad (7.14)$$

$$\tilde{\phi}_i^a(\mathbf{r}) = \frac{\tilde{\phi}_{n_i l_i}^a(r)}{r} y_{m_i}^{l_i}(\hat{\mathbf{r}}), \quad (7.15)$$

$$\tilde{p}_i^a(\mathbf{r}) = \frac{\tilde{p}_{n_i l_i}^a(r)}{r} y_{m_i}^{l_i}(\hat{\mathbf{r}}). \quad (7.16)$$

In the above expressions, $y_m^l(\hat{\mathbf{r}})$ denotes a real spherical harmonic whose basic properties are summarized in appendix D. The “smooth” pseudo-image $|\tilde{\Psi}\rangle$ is expanded in a plane wave basis set:

$$|\tilde{\Psi}_{n\mathbf{k}}\rangle = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} \tilde{u}_{n\mathbf{k}}(\mathbf{G}) |\mathbf{k} + \mathbf{G}\rangle, \quad (7.17)$$

where Ω is the volume of the unit cell and $|\mathbf{k} + \mathbf{G}\rangle$ is a plane wave. Thanks to the smoothness of the pseudized wave function, the projection of $\tilde{\Psi}(\mathbf{r})$ onto $\tilde{p}_i(\mathbf{r})$ can be accurately evaluated in reciprocal space. Such a task is achieved by expanding the plane wave basis set in terms of real spherical harmonics and spherical Bessel functions using the Rayleigh expansion reported in equation D.18. The final result reads:

$$\langle \tilde{p}_i^a | \tilde{\Psi} \rangle = \frac{4\pi i^{l_i}}{\sqrt{\Omega}} \sum_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{R}_a} y_{m_i}^{l_i}(\widehat{\mathbf{k}+\mathbf{G}}) \tilde{u}(\mathbf{G}) \tilde{p}_{n_i l_i}^a(|\mathbf{k} + \mathbf{G}|), \quad (7.18)$$

where the Fourier components of the PAW radial projectors are given by:

$$\tilde{p}_i^a(|\mathbf{k} + \mathbf{G}|) = 4\pi (-i)^{l_i} e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{R}_a} y_{m_i}^{l_i}(\widehat{\mathbf{k}+\mathbf{G}}) \int_0^{r_c^a} \tilde{p}_{n_i l_i}^a(r) j_l(|\mathbf{k} + \mathbf{G}|r) dr. \quad (7.19)$$

7.2.1 Remarks on the notation and terminology

After this preamble, all the notations and concepts needed for discussing the PAW methodology and the *GW* implementation have been defined. Throughout the rest of the chapter, unless explicitly noted, the atom index will usually be omitted and implicitly absorbed in the index i . It is also worth clarifying a delicate point that might be a cause of misunderstanding in the following discussion.

The PAW method described so far constitutes a virtually exact implementation of density functional theory in an all-electron and full-potential framework. On the other hand, most of the PAW implementations reported so far in the literature rely on the *frozen core* approximation in which inner core electrons are considered to be frozen in the atomic configuration. Only valence states are allowed to relax during the self-consistent solution of the KS equations. The *frozen core* approximation significantly reduces the computational effort, has been proved accurate for total energy calculations [60, 133, 139], and will be adopted also in the present work. Nevertheless, we will continue using the term “all-electron function” or interchangeably “true wave function” to indicate the wave function obtained from the auxiliary pseudized version $|\tilde{\Psi}\rangle$ by means of the linear transformation given in equation 7.1. One should always keep in mind, however, that $|\Psi\rangle$ equals the AE solution only if core relaxation effects are fully taken into account. We refer the reader to the recent work reported in [140] for a discussion of the technical details needed for the implementation of the PAW formalism with relaxed cores.

7.3 Expectation values

The matrix elements of a local or semi-local operator $\hat{A}$ between two AE wave functions can be efficiently and accurately evaluated by employing the linear transformation $\hat{\mathcal{T}}$, as given in equation 7.10. Using the decomposition of the AE wave function reported in equation 7.11, one arrives at the following result:

$$\begin{aligned} \langle \Psi | \hat{A} | \Psi \rangle &= \langle \tilde{\Psi} | \hat{A} | \tilde{\Psi} \rangle + \sum_a \langle \Psi_a^1 | \hat{A} | \Psi_a^1 \rangle - \langle \tilde{\Psi}_a^1 | \hat{A} | \tilde{\Psi}_a^1 \rangle + \\ &+ \underbrace{\sum_a \langle \Psi_a^1 - \tilde{\Psi}_a^1 | \hat{A} | \tilde{\Psi} - \tilde{\Psi}_a^1 \rangle + \langle \tilde{\Psi} - \tilde{\Psi}_a^1 | \hat{A} | \Psi_a^1 - \tilde{\Psi}_a^1 \rangle + \sum_{a \neq a'} \langle \Psi_a^1 - \tilde{\Psi}_a^1 | \hat{A} | \Psi_{a'}^1 - \tilde{\Psi}_{a'}^1 \rangle}_{=0 \text{ if the partial wave basis set is complete and } \hat{A} \text{ is local or semilocal}} \end{aligned} \quad (7.20)$$

Only the first line of equation 7.20 has to be explicitly evaluated. The two summations on the second line can be safely neglected for sufficiently local operators as long as the partial wave expansion is complete. This important property can be seen by the following argument: $\Psi_a^1(\mathbf{r}) - \tilde{\Psi}_a^1(\mathbf{r})$ vanishes by construction in the interstitial region while $\tilde{\Psi}(\mathbf{r}) - \tilde{\Psi}_a^1(\mathbf{r})$ vanishes inside the PAW sphere once the partial wave expansion is converged. In no region of space are both functions, $\Psi_a^1(\mathbf{r}) - \tilde{\Psi}_a^1(\mathbf{r})$ and $\tilde{\Psi}(\mathbf{r}) - \tilde{\Psi}_a^1(\mathbf{r})$, simultaneously nonzero. The functions from different atomic sites are never nonzero in the same region provided that the augmentation spheres do not overlap. Note, however, that for truly nonlocal operators the terms on the second line in equation 7.20 would have to be explicitly evaluated.

Under all the above mentioned assumptions, equation 7.20 can be simplified as:

$$\langle \Psi | \hat{A} | \Psi \rangle = \langle \tilde{\Psi} | \hat{A} | \tilde{\Psi} \rangle + \sum_{ij} \langle \tilde{\Psi} | \tilde{p}_i \rangle \left[\langle \phi_i | \hat{A} | \phi_j \rangle - \langle \tilde{\phi}_i | \hat{A} | \tilde{\phi}_j \rangle \right] \langle \tilde{p}_j | \tilde{\Psi} \rangle, \quad (7.21)$$

where the onsite representation of the AE and PS wave functions (Eq. 7.12 and Eq. 7.13, respectively) have been used. The first term in equation 7.21 has the same mathematical structure as the expression present in the pseudopotential formalism. As it involves only the “smooth” part of wave function, it can be evaluated either in real or reciprocal space, depending on the nature of $\hat{A}$, by changing representation through FFT techniques. The second term involves the onsite matrix elements of the $\hat{A}$ operator between AE and PS partial waves. It can be evaluated either by employing a radial and angular mesh in real space or by expanding the operator $\hat{A}$ in terms of angular momenta. An important consequence of equation 7.21 is that the set of pseudo wave functions $\{\tilde{\Psi}_\alpha\}$ satisfy the orthogonality condition:

$$\langle \tilde{\Psi}_\alpha | \hat{O} | \tilde{\Psi}_\beta \rangle = \langle \Psi_\alpha | \Psi_\beta \rangle = \delta_{\alpha\beta}, \quad (7.22)$$

where the overlap operator $\hat{O}$ – the pseudized version of the identity – is given by:

$$\hat{O} \equiv \hat{1} + \sum_{ij} |\tilde{p}_i\rangle \left[\langle \phi_i | \phi_j \rangle - \langle \tilde{\phi}_i | \tilde{\phi}_j \rangle \right] \langle \tilde{p}_j|. \quad (7.23)$$

In analogy to the ultrasoft formalism [70], the pseudo wave functions, $|\tilde{\Psi}_\alpha\rangle$, are not normalized, hence additional localized charges have to be introduced to correctly describe the multipoles of the AE density. The description of the total electronic charge in the PAW formalism is the subject of the next section.

7.4 Charge Densities in the PAW formalism

The electronic charge at point $\mathbf{r}$ is defined as the expectation value of the operator $\sum_i \delta(\mathbf{r} - \mathbf{r}_i)$. By means of equation 7.20, one sees that the AE charge density is given by the sum of three different contributions¹

$$n(\mathbf{r}) = \sum_{n\mathbf{k}} f_{n\mathbf{k}} |\Psi_{n\mathbf{k}}(\mathbf{r})|^2 = \tilde{n}(\mathbf{r}) + n^1(\mathbf{r}) - \tilde{n}^1(\mathbf{r}). \quad (7.24)$$

with $f_{n\mathbf{k}}$ the occupation number of each state. $\tilde{n}(\mathbf{r})$ is the so-called soft pseudo charge-density. It is directly obtained from the pseudized wave functions on a homogeneous real-space grid using:

$$\tilde{n}(\mathbf{r}) = \sum_{n\mathbf{k}} f_{n\mathbf{k}} |\tilde{\Psi}_{n\mathbf{k}}(\mathbf{r})|^2, \quad (7.25)$$

which is completely analogous to the equation used in the norm-conserving approach. $n^1(\mathbf{r})$ and $\tilde{n}^1(\mathbf{r})$ are the onsite charge densities needed to restore the correct total charge.

¹ Hereafter, the following convention is used. Quantities with a tilde are directly derived from the pseudized wave functions. Expressions with the superscript “1” are one-center quantities evaluated on a radial grid. Terms arising from the on-site representation of the smooth wave functions are denoted with the superscript “1” and an additional tilde.

They extend up to the PAW cutoff radius around each atom, and are evaluated on a radial support grid. The final expression for the onsite contributions to the charge reads:

$$n^1(\mathbf{r}) = \sum_{ij} \rho_{ij} \langle \phi_i | \mathbf{r} \rangle \langle \mathbf{r} | \phi_j \rangle, \quad (7.26)$$

$$\tilde{n}^1(\mathbf{r}) = \sum_{ij} \rho_{ij} \langle \tilde{\phi}_i | \mathbf{r} \rangle \langle \mathbf{r} | \tilde{\phi}_j \rangle, \quad (7.27)$$

with ρ_{ij} the occupancies of each augmentation channel (i, j) . The ρ_{ij} are calculated from the pseudo wave functions by applying the projector functions:

$$\rho_{ij} = \sum_{n\mathbf{k}} f_{n\mathbf{k}} \langle \tilde{\Psi}_{n\mathbf{k}} | \tilde{p}_i \rangle \langle \tilde{p}_j | \tilde{\Psi}_{n\mathbf{k}} \rangle. \quad (7.28)$$

For a complete set of projectors, the charge density $\tilde{n}^1(\mathbf{r})$ is exactly the same as $\tilde{n}(\mathbf{r})$ within each PAW sphere, so the all-electron valence charge density is simply given by $n^1(\mathbf{r})$. On the other hand, since the set of partial waves has to be truncated for practical applications, the cancellation within the augmentation regions is always inexact, and special numerical techniques are needed for an accurate treatment of the long-range electrostatic interaction. Following an approach commonly used in the LAPW method [141], a soft compensation charge, $\hat{n}(\mathbf{r})$, strictly localized in the augmentation regions, is added to the soft charge densities $\tilde{n}(\mathbf{r})$ and $\tilde{n}^1(\mathbf{r})$ [132]:

$$n(\mathbf{r}) = \underbrace{(\tilde{n}(\mathbf{r}) + \hat{n}(\mathbf{r}))}_{\text{FFT Mesh}} + \underbrace{(n^1(\mathbf{r}) - \tilde{n}^1(\mathbf{r}) - \hat{n}(\mathbf{r}))}_{\text{Onsite Description}}. \quad (7.29)$$

The compensation charge $\hat{n}(\mathbf{r})$ is chosen such that the moments of the electronic charge within each augmentation region are zero, namely that:

$$\int_{\Omega_{\mathbf{R}}} (n^1(\mathbf{r}) - \tilde{n}^1(\mathbf{r}) - \hat{n}(\mathbf{r})) |\mathbf{r} - \mathbf{R}|^l \mathcal{Y}_m^l(\widehat{\mathbf{r} - \mathbf{R}}) d\mathbf{r} = 0, \quad (7.30)$$

where $\Omega_{\mathbf{R}}$ is the volume of the augmentation sphere. Fulfilling the above condition is of crucial importance for the reliability and the efficiency of the PAW method. Indeed, only in this case is the Coulomb potential generated by the on-site charge density $n^1(\mathbf{r}) - \tilde{n}^1(\mathbf{r}) - \hat{n}(\mathbf{r})$ in one sphere null in the other augmentation regions, and only relatively simple one-center integrals have to be considered. Equation 7.30 is automatically satisfied by defining the compensation charge $\hat{n}(\mathbf{r})$ according to:

$$\hat{n}(\mathbf{r}) = \sum_{ij} \sum_{LM} \rho_{ij} \hat{Q}_{ij}^{LM}(\mathbf{r}), \quad (7.31)$$

where

$$\hat{Q}_{ij}^{LM}(\mathbf{r}) = q_{ij}^{LM} g_L(r) \mathcal{Y}_m^L(\hat{\mathbf{r}}), \quad (7.32)$$

and

$$q_{ij}^{LM} = \sqrt{4\pi} \mathcal{G}_{l_i m_i, l_j m_j}^{LM} \int_0^{r_c^a} (\phi_{n_i l_i} \phi_{n_j l_j} - \tilde{\phi}_{n_i l_i} \tilde{\phi}_{n_j l_j}) r^L dr. \quad (7.33)$$

$g_L(r)$ is the so-called shape function, which is localized in the sphere and has to satisfy the condition:

$$\int_0^{r_c^a} r^{2+L} g_L(r) dr = 1. \quad (7.34)$$

The symbol $\mathcal{G}_{l_i m_i, l_j m_j}^{lm}$ is used to denote the integration over the solid angle of three real spherical harmonics – the so-called Gaunt coefficient [142] – whose basic properties are discussed in section D.2. Note that, in equation 7.29, the first term in parenthesis is treated on the FFT mesh. For this reason, the shape functions, $g_L(r)$, have to be chosen such that $\hat{n}(\mathbf{r})$ is smooth both in the real-space- and the Fourier-domain. This assures that an accurate description can be achieved by sampling the PAW spheres with an affordably dense FFT mesh.

As already mentioned, the PAW method implemented in the ABINIT code relies on the frozen core approximation and therefore the charge density in equation 7.24 has to be interpreted as the contribution given by the valence electrons. For the description of the core charge, following [132], we introduce four different quantities $n_c(\mathbf{r})$, $\tilde{n}_c(\mathbf{r})$, $n_{Zc}(\mathbf{r})$ and $\tilde{n}_{Zc}(\mathbf{r})$. $n_c(\mathbf{r})$ denotes the charge density of the frozen core in the reference atomic configuration. The partial electronic core density $\tilde{n}_c(\mathbf{r})$ is equivalent to the frozen core AE charge $n_c(\mathbf{r})$ outside a certain radius r_{pc} which lies inside the augmentation region. $\tilde{n}_c(\mathbf{r})$ is used to calculate nonlinear core corrections as described in [143]. $n_{Zc}(\mathbf{r})$ represents the charge density of the nuclei plus the frozen AE charge $n_c(\mathbf{r})$, that is:

$$n_{Zc}(\mathbf{r}) = n_Z(\mathbf{r}) + n_c(\mathbf{r}). \quad (7.35)$$

Finally, the pseudized core density $\tilde{n}_{Zc}(\mathbf{r})$ is a charge distribution that is equivalent to $n_{Zc}(\mathbf{r})$ outside the core region and have the same moments as $n_{Zc}(\mathbf{r})$ inside the core region. The PAW all-electron total charge density $n_T(\mathbf{r})$ is thus partitioned in three terms:

$$n_T = n + n_{Zc} = \underbrace{(\tilde{n} + \hat{n} + \tilde{n}_{Zc})}_{\tilde{n}_T} + \underbrace{(n^1 + n_{Zc})}_{n_T^1} - \underbrace{(\tilde{n}^1 + \hat{n} + \tilde{n}_{Zc})}_{\tilde{n}_T^1}. \quad (7.36)$$

To treat the exchange-correlation energy and potential, it is advantageous to use the following decomposition for the electron-only charge:

$$n_v + n_c = (\tilde{n} + \hat{n} + \tilde{n}_c) + (n^1 + n_c) - (\tilde{n}^1 + \hat{n} + \tilde{n}_c), \quad (7.37)$$

where $\tilde{n}_c(\mathbf{r})$ is the soft pseudized core charge.

7.5 PAW Hamiltonian

Starting from the decomposition of the density given in equation 7.36, the total energy per unit cell can be expressed as a functional of the pseudo wave functions and pseudo densities [132]. The final expression consists of the three different contributions:

$$E = \tilde{E} + E^1 - \tilde{E}^1, \quad (7.38)$$

which are computed as follows:²

$$\begin{aligned} \tilde{E} = & -\frac{1}{2} \sum_{n\mathbf{k}} f_{n\mathbf{k}} \langle \Psi_{n\mathbf{k}} | \Delta | \Psi_{n\mathbf{k}} \rangle + E_{\text{xc}}[\tilde{n} + \hat{n} + \tilde{n}_c] + E_H[\tilde{n} + \hat{n}] + \\ & + \int v_H[\tilde{n}_{Zc}](\mathbf{r})(\tilde{n} + \hat{n}) d\mathbf{r} + U_{I-I}(\{\mathbf{R}\}) \end{aligned} \quad (7.39)$$

$$E^1 = -\frac{1}{2} \sum_{ij} \rho_{ij} \langle \phi_i | \Delta | \phi_j \rangle + E_{\text{xc}}[n^1 + n_c] + E_H[n^1] + \int_{\Omega_{\mathbf{R}}} v_H[n_{Zc}](\mathbf{r}) n^1(\mathbf{r}) d\mathbf{r} \quad (7.40)$$

$$\begin{aligned} \tilde{E}^1 = & -\frac{1}{2} \sum_{ij} \rho_{ij} \langle \tilde{\phi}_i | \Delta | \tilde{\phi}_j \rangle + E_{\text{xc}}[\tilde{n}^1 + \hat{n} + \tilde{n}_c] + E_H[\tilde{n}^1 + \hat{n}] + \\ & + \int_{\Omega_{\mathbf{R}}} v_H[\tilde{n}_{Zc}](\mathbf{r})(\tilde{n}^1 + \hat{n})(\mathbf{r}) d\mathbf{r} \end{aligned} \quad (7.41)$$

In the above expressions, the contribution due to the ion-ion interaction and the Hartree energy of the core electron densities are not included since they are constant in our particular case. The PAW Hamiltonian operator is obtained by differentiating the energy functional with respect to the pseudo density operator $\tilde{n} = \sum_{n\mathbf{k}} f_{n\mathbf{k}} |\tilde{\Psi}_{n\mathbf{k}}\rangle \langle \tilde{\Psi}_{n\mathbf{k}}|$. The final result reads [132]:

$$\hat{H} = -\frac{1}{2} \Delta + \tilde{v}_{\text{eff}}(\mathbf{r}) + \sum_{ij} |\tilde{p}_i\rangle \underbrace{\left[\hat{D}_{ij} + D_{ij}^1 - \tilde{D}_{ij}^1 \right]}_{D_{ij}} \langle \tilde{p}_j|, \quad (7.42)$$

where the coefficients D_{ij} , also called PAW pseudopotential strengths, define the importance of each ij channel to the nonlocal part of the PAW Hamiltonian. Each pseudopo-

² Henceforth, the following conventions are used: Integrals evaluated inside the augmentation sphere on a radial mesh are denoted with $\int_{\Omega_{\mathbf{R}}}$, the symbol $\int_{\Omega}$ is used to indicate an integration performed on the FFT mesh.

tential strength D_{ij} is the sum of three different contributions:

$$\begin{cases} D_{ij}^1 = \langle \phi_i | -\frac{1}{2}\Delta + v_{\text{eff}}^1 | \phi_j \rangle \\ \tilde{D}_{ij}^1 = \langle \tilde{\phi}_i | -\frac{1}{2}\Delta + \tilde{v}_{\text{eff}}^1 | \tilde{\phi}_j \rangle + \sum_{LM} \int_{\Omega_R} \tilde{v}_{\text{eff}}^1(\mathbf{r}) \hat{Q}_{ij}^{LM}(\mathbf{r}) d\mathbf{r} \\ \hat{D}_{ij} = \sum_{LM} \int_{\Omega} \tilde{v}_{\text{eff}}(\mathbf{r}) \hat{Q}_{ij}^{LM}(\mathbf{r}) d\mathbf{r}, \end{cases} \quad (7.43)$$

where the different effective potentials are defined by:

$$\begin{cases} \tilde{v}_{\text{eff}}(\mathbf{r}) = v_H[\tilde{n} + \hat{n} + \tilde{n}_{Zc}] + v_{\text{xc}}[\tilde{n} + \hat{n} + \tilde{n}_c] \\ v_{\text{eff}}^1(\mathbf{r}) = v_H[n^1 + n_{Zc}] + v_{\text{xc}}[n^1 + n_c] \\ \tilde{v}_{\text{eff}}^1(\mathbf{r}) = v_H[\tilde{n}^1 + \hat{n} + \tilde{n}_{Zc}] + v_{\text{xc}}[\tilde{n}^1 + \hat{n} + \tilde{n}_c]. \end{cases} \quad (7.44)$$

The potential $\tilde{v}_{\text{eff}}(\mathbf{r})$ arises from soft densities, and can be accurately described employing a relatively coarse FFT mesh. The two potentials $v_{\text{eff}}^1(\mathbf{r})$ and $\tilde{v}_{\text{eff}}^1(\mathbf{r})$ are generated by the onsite AE and PS charges, respectively. At this point, it is important to note that the Hartree and the XC potentials generated by the onsite densities are treated in different ways. The Coulomb potential is always expanded in terms of angular momenta as described in paragraph 7.6.2. The evaluation of the onsite exchange-correlation potential is more involved: The radial dependence of $v_{\text{xc}}(\mathbf{r})$ is always described using a radial support mesh while the angular variations can be described either in terms of angular momenta and real spherical harmonics or, alternatively, by means of a discretized angular mesh. This rather technical point is discussed in more detail in the next section 7.6.

It is worth noting that, contrarily to the norm-conserving formalism, the PAW strengths D_{ij} have an implicit functional dependence on the density through the various effective potentials introduced in equation 7.43. Therefore the D_{ij} 's are not simple constants, and have to be updated during the self-consistent solution of the KS equation. This is similar in spirit to the ultra-soft formalism, and indeed, both the ultra-soft and the norm-conserving Hamiltonian can be obtained from the PAW Hamiltonian operator (Eq. 7.42) under well-defined approximations [15, 132]. Within the PAW method, as in the ultrasoft theory, the Kohn-Sham orbitals and energies are calculated by solving the generalized eigenvalue problem [15]:

$$\hat{H}\tilde{\Psi}(\mathbf{r}) = \epsilon \hat{O}\tilde{\Psi}(\mathbf{r}), \quad (7.45)$$

where the overlap operator $\hat{O}$ is defined by equation 7.23.

7.5.1 Treatment of the Different Contributions to D_{ij}

In order to facilitate the forthcoming discussion on the GW implementation, it is useful to split the total pseudopotential strengths D_{ij} into three different contributions:

$$D_{ij} = D_{ij}^0 + (D_{ij})^H + (D_{ij})^{\text{xc}}. \quad (7.46)$$

D_{ij}^0 is obtained during the generation of the PAW pseudopotential, and is constant during the self-consistent cycle. We do not comment further on this term since it plays no role neither in case of one-shot G_0W_0 calculation nor in case of $QPSCGW$. The interested reader is referred to [135], where the explicit expression of D_{ij}^0 in terms of n_{Zc} , $\tilde{n}_{Zc}$ and PAW partial waves is provided.

The exchange-correlation contribution to the total D_{ij} can be recast in the form:

$$(D_{ij})^{\text{xc}} = (D_{ij}^1)^{\text{xc}} - (\tilde{D}_{ij}^1)^{\text{xc}} + \hat{D}_{ij}^{\text{xc}}, \quad (7.47)$$

where the first two terms are defined by:

$$(D_{ij}^1)^{\text{xc}} = \langle \phi_i | v_{\text{xc}} [n^1 + n_c] | \phi_j \rangle, \quad (7.48)$$

and

$$(\tilde{D}_{ij}^1)^{\text{xc}} = \langle \tilde{\phi}_i | v_{\text{xc}} [\tilde{n}^1 + \hat{n} + \tilde{n}_c] | \tilde{\phi}_j \rangle + \sum_{LM} q_{ij}^{LM} \int_{\Omega_R} g_L(r) \mathfrak{y}_M^L(\Omega) v_{\text{xc}} [\tilde{n} + \hat{n} + \tilde{n}_c](\mathbf{r}) d\mathbf{r}. \quad (7.49)$$

Both equation 7.48 and equation 7.49 require an integration inside the PAW spheres which can be performed either by employing a Gauss-Legendre integration scheme or through an expansion in terms of momenta, as described in section 7.6.1. The last XC-like contribution in equation 7.47 – the so-called “hat” term – is given by:

$$\hat{D}_{ij}^{\text{xc}} = \sum_{LM} q_{ij}^{LM} \int_{\Omega} g_L(r) \mathfrak{y}_M^L(\Omega) v_{\text{xc}} [\tilde{n}^1 + \hat{n} + \tilde{n}_c](\mathbf{r}) d\mathbf{r}. \quad (7.50)$$

In order to guarantee an accurate sampling of the spherical regions, equation 7.50 is evaluated on an augmented FFT mesh much denser than that employed for the smooth wave functions.

The decomposition of the Hartree-like contribution, $(D_{ij})^H$, is more complicated as several different terms arising from both core and valence densities are involved. Some of these terms are directly imported from the atomic reference configuration, whereas others have an explicit functional dependence on the valence densities, and have to be updated during the self-consistent iterations. Also the Hartree contribution to the total D_{ij} can – like D_{ij}^{xc} – be decomposed into:

$$(D_{ij})^H = (D_{ij}^1)^H - (\tilde{D}_{ij}^1)^H + (\hat{D}_{ij})^H. \quad (7.51)$$

The first two terms are defined as:

$$(D_{ij}^1)^H = \langle \phi_i | v_H [n^1 + n_{Zc}] | \phi_j \rangle, \quad (7.52)$$

and

$$(\tilde{D}_{ij}^1)^H = \langle \tilde{\phi}_i | v_H [\tilde{n}^1 + \tilde{n}_{Zc} + \hat{n}] | \tilde{\phi}_j \rangle + \sum_{LM} \int_{\Omega_{\mathbf{R}}} \hat{Q}_{ij}^{LM}(\mathbf{r}) v_H [\tilde{n}^1 + \hat{n}](\mathbf{r}) d\mathbf{r}. \quad (7.53)$$

Since only onsite densities and potentials are involved, the two contributions are evaluated by means of an expansion in terms of angular momenta as described in section 7.6.2. The “hat” term is, in contrast, given by:

$$\hat{D}_{ij}^H = \sum_{LM} \int \hat{Q}_{ij}^{LM}(\mathbf{r}) v_H [\tilde{n} + \hat{n} + \tilde{n}_{Zc}](\mathbf{r}) d\mathbf{r} = \sum_{LM} q_{ij}^L \int g^L(r) r^2 \mathcal{Y}_M^L(\hat{\mathbf{r}}) v_H [\tilde{n} + \hat{n} + \tilde{n}_{Zc}](\mathbf{r}) d\mathbf{r}, \quad (7.54)$$

and is always evaluated on an augmented FFT mesh.

7.6 Treatment of v_{xc} and v_H in PAW

This section describes the different numerical techniques that can be used to evaluate the onsite contribution of the exchange-correlation potential and the Hartree term. The first paragraph (7.6.1) discusses the treatment of the exchange-correlation potential. The expressions for the onsite matrix elements of $v_{xc}(\mathbf{r})$ needed in case of G_0W_0 calculations are also presented. In paragraph 7.6.2, we then proceed to discuss how to calculate the onsite Hartree potential which has to be self-consistently updated when the *QPSCGW* approach of Faleev et al. [16] is used.

7.6.1 Onsite Exchange-Correlation Potentials

Two different methods are available in the ABINIT code to describe the onsite densities and the corresponding exchange-correlation potentials generated inside the augmentation spheres. In the first approach, all quantities are discretized using a radial support grid and an angular mesh. All integrations are then performed using a Gauss-Legendre scheme [86] over the solid angle. For instance, the matrix element of the onsite $v_{xc}(\mathbf{r})$ potential between two PAW partial waves is calculated according to:

$$\langle \phi_i | v_{xc} | \phi_j \rangle = \sum_{\hat{\omega}}^{N_{\hat{\omega}}} w(\hat{\omega}) \mathcal{Y}_{m_i}^{l_i}(\hat{\omega}) \mathcal{Y}_{m_j}^{l_j}(\hat{\omega}) \int v_{xc}(\hat{\omega}r) \phi_{n_i l_i}(r) \phi_{n_j l_j}(r) dr \quad (7.55)$$

where $\hat{\omega}$ is one of the $N_{\hat{\omega}}$ points (θ, ϕ) sampled on the unitary sphere, and $w(\hat{\omega})$ is the associated weight satisfying the normalization condition:

$$\sum_{\hat{\omega}}^{N_{\hat{\omega}}} w(\hat{\omega}) = 1. \quad (7.56)$$

The main disadvantage of the Gauss-Legendre approach is that evaluating the gradient of the density is not straightforward and, moreover, it is prone to numerical errors. For such a reason, the description of $v_{xc}(\mathbf{r})$ in terms of a discretized angular mesh is available only for LDA functionals.

More accurate results can be obtained by expanding the onsite densities and potentials in terms of angular momenta³

$$n(\mathbf{r}) = \sum_{LM} n^{LM}(r) \mathcal{Y}_M^L(\hat{\mathbf{r}}), \quad v_{xc}(\mathbf{r}) = \sum_{LM} v_{xc}^{LM}(r) \mathcal{Y}_M^L(\hat{\mathbf{r}}) \quad (7.57)$$

where the (l, m) moments are defined by:

$$n^{LM}(r) \equiv \int_{\Omega_{\mathbf{R}}} n(\mathbf{r}) \mathcal{Y}_M^L(\Omega) d\Omega \quad v_{xc}^{LM}(r) \equiv \int_{\Omega_{\mathbf{R}}} v_{xc}(\mathbf{r}) \mathcal{Y}_M^L(\Omega) d\Omega. \quad (7.58)$$

The onsite matrix elements between two partial waves are readily obtained using the expansion of $v_{xc}(\mathbf{r})$ given in equation 7.57:

$$\langle \phi_i | v_{xc} | \phi_j \rangle = \sum_{LM} \mathcal{G}_{l_i m_i l_j m_j}^{LM} \int v_{xc}^{LM}(r) \phi_{n_i l_i}(r) \phi_{n_j l_j}(r) dr. \quad (7.59)$$

The above expression provides all the information needed for evaluating the matrix elements of the exchange-correlation potential in the PAW formalism. Since LDA (GGA) is a local (semilocal) functional of the density, the matrix elements of the exchange-correlation potential between two all-electron Bloch states can be rewritten as:

$$\begin{aligned} \langle \Psi | v_{xc} [n_v + n_c] | \Psi \rangle &= \langle \tilde{\Psi} | v_{xc} [\tilde{n} + \hat{n} + \tilde{n}_c] | \tilde{\Psi} \rangle + \\ &+ \langle \Psi^1 | v_{xc} [n^1 + n_c] | \Psi^1 \rangle - \langle \tilde{\Psi}^1 | v_{xc} [\tilde{n}^1 + \hat{n} + \tilde{n}_c] | \tilde{\Psi}^1 \rangle. \end{aligned} \quad (7.60)$$

The first term in equation 7.60 is evaluated on the FFT mesh, and does not require any additional coding as compared to the case of norm-conserving pseudopotentials. The evaluation of the second line of equation 7.60 involves the one-center contributions of the XC potential, and can be obtained from the knowledge of the matrix elements defined in equation 7.59. By expressing the onsite representation of the AE wave functions in terms of the partial waves, one arrives at the final result:

³ All the results reported in this work have been obtained using this approach.

$$\begin{aligned} \langle \Psi | v_{xc}[n + n_c] | \Psi \rangle &= \langle \tilde{\Psi} | v_{xc}[\tilde{n} + \hat{n} + \tilde{n}_c] | \tilde{\Psi} \rangle + \\ &+ \sum_{ij} \langle \tilde{\Psi} | \tilde{p}_i \rangle \left[(D_{ij}^1)^{xc} - (\tilde{D}_{ij}^1)^{xc} + \hat{D}_{ij}^{xc} \right] \langle \tilde{p}_j | \tilde{\Psi} \rangle, \end{aligned} \quad (7.61)$$

where the different XC contributions to the total PAW strengths have already been introduced and discussed in section 7.5. Equation 7.61 defines the matrix element of the total exchange-correlation, namely the XC potential generated both by core and valence electrons. As already discussed in section 4.11, the appropriate exchange-correlation contribution due to the valence has to be subtracted out from the self-energy when computing QP energies in a perturbative approach based on the frozen core approximation (Eq. 4.55). The simplest approach to the core-valence decoupling consists in assuming that the core-valence exchange-correlation and core-polarization contributions of a valence state is given by [56]:

$$\langle \Psi | v_{xc}^{\text{core-val}} | \Psi \rangle \approx \langle \Psi | v_{xc}[n_v + n_c] | \Psi \rangle - \langle \Psi | v_{xc}[n_v] | \Psi \rangle. \quad (7.62)$$

In this case, the valence-electron exchange-correlation contribution is removed from the valence self-energy using the expression (henceforth referred to as KS decoupling):

$$\langle \Psi | \hat{\Sigma} - v_{xc}[n_v] | \Psi \rangle, \quad (7.63)$$

where $v_{xc}[n_c + n_v] = v_{xc}[n_c] + v_{xc}[n_v]$ has been implicitly assumed. The valence-only contribution to the exchange-correlation potential needed in equation 7.63 is calculated using an expression similar to equation 7.61. In this particular case, only the contributions arising from the PS and AE valence electrons have to be considered.

The KS core-valence decoupling expressed by equation 7.63 is also the standard approach followed in the pseudopotential formalism, but it may be a source of error due to the nonlinearity of the XC potential. More importantly, equation 7.63 is hard to justify from a physical point of view.

In chapter 4, we saw that the contribution to the self-energy due to core electrons reduces to a localized Fock-like operator if core polarization effects are neglected. Therefore, a more accurate and more physical approach to the core-valence decoupling can be obtained by directly computing the core-valence exchange interaction at the Fock level (hereafter referred to as HF decoupling [14]). The exchange operator of the core states is defined by:

$$\Sigma_x^{\text{core}}(\mathbf{r}_1, \mathbf{r}_2) = - \sum_a \sum_{nlm} \frac{\Phi_{nlm}^a(\mathbf{r}_1) [\Phi_{nlm}^a(\mathbf{r}_2)]^\dagger}{|\mathbf{r}_1 - \mathbf{r}_2|}, \quad (7.64)$$

where $\Phi_{nlm}^a(\mathbf{r})$ are core orbitals centered on atomic site a . For sufficiently large PAW spheres, $\Phi_{nlm}^a(\mathbf{r})$ can be assumed to be strictly localized inside the augmentation region, and thus the evaluation of the matrix elements of $\hat{\Sigma}_x^{\text{core}}$ only requires the onsite representation of the true valence functions. It is important to stress that the same approach

would not be reliable in the pseudopotential formalism since $\hat{\Sigma}_x^{\text{core}}$ probes the nodal shape of the valence wave functions in the regions close to the ions.

k	$v_{\text{xc}}[n_v]$ Present	$v_{\text{xc}}[n_v]$ [14]	$v_{\text{xc}}[n_v + n_c] - \Sigma_x^c$ [14]
$\Gamma_{1\nu}$	-10.60	-10.57	-10.61
$\Gamma_{25\nu}$	-11.46	-11.45	-11.60
Γ_{15c}	-10.20	-10.19	-10.34
Γ_{2c}	-11.23	-11.16	-11.26
X_{1v}	-10.99	-10.95	-11.02
X_{4v}	-10.75	-10.74	-10.85
X_{1c}	-9.19	-9.15	-9.25
X_{3c}	-10.83	-10.84	-11.09
$L_{2'v}$	-10.98	-10.95	-11.01
L_{1v}	-10.37	-10.33	-10.41
$L_{3'v}$	-11.21	-11.20	-11.33
L_{1c}	-10.32	-10.28	-10.39
L_{3c}	-9.80	-9.78	-9.92

Table 7.2: Matrix elements of the exchange-correlation potential of the valence states $\langle \Psi | v_{\text{xc}}[n_v] | \Psi \rangle$ compared with previously published results (all in eV). The matrix elements obtained using the Hartree-Fock decoupling, *i.e.* $\langle \Psi | v_{\text{xc}}[n_v + n_c] | \Psi \rangle - \langle \Psi | \Sigma_x^{\text{core}} | \Psi \rangle$ are also reported for the sake of comparison.

A comparison of the matrix elements calculated with the two approaches is reported in table 7.2 for the case of crystalline silicon. First of all, we wish to note that our values for the valence-only matrix elements of the XC potential are in good agreement with the results reported in [14], and so validating our implementation. It is also quite surprising that using the Hartree-Fock decoupling in Si leads to QP results whose average deviation from the traditional KS method is less than 0.1 eV, thus well within the typical numerical uncertainty of *GW* calculations, commonly quoted to be $\approx 0.2\text{-}0.3$ eV. The small deviation provides an *a posteriori* justification to the KS core-valence decoupling (Eq. 7.62) for the particular case of crystalline silicon.

On the other hand, the agreement between the KS and the HF decoupling is less satisfying in more complicated systems, particularly in case of atoms with shallow cores. Recent calculations reported in the literature [104, 144] have indeed revealed that using the HF decoupling gives results that are closer to AE calculations while keeping most of the electrons frozen in the core. Having the possibility of performing HF decoupling within ABINIT will considerably improve the accuracy and the efficiency of the *GW* implementation and is presently under development. The only technical issue is related to the definition of the core orbitals used to construct Σ_x^{core} (Eq. 7.64). Strictly speaking, the core wave functions should be orthogonal to all the remaining states, both empty and

occupied. Therefore the core wave functions should not be directly imported from the atomic environment, but should be obtained from a relaxed core calculation in which the core electrons are relaxed with an inner self-consistent loop for spherical onsite Hamiltonians [140].

7.6.2 Onsite Hartree Potentials

The Hartree terms generated by the charges localized within the augmentation spheres are best evaluated using an expansion in terms of angular momenta. Given a charge distribution of angular momentum l ,

$$n(\mathbf{r}) = f(r) y_M^L(\hat{\mathbf{r}}) \quad (7.65)$$

the Poisson equation:

$$\Delta v(\mathbf{r}) = -4\pi n(\mathbf{r}), \quad (7.66)$$

can be solved by expanding the Coulomb kernel in terms of real spherical harmonics via the expression reported in equation D.19. The final expression reads:

$$\begin{aligned} v_H[n](\mathbf{r}) &= \sum_{lm} \frac{4\pi}{(2l+1)} y_m^l(\hat{\mathbf{r}}) \int_0^\infty r'^2 f(r') \frac{r_{<}^l}{r_{>}^{l+1}} dr' \int y_m^l(\hat{\mathbf{r}}') y_M^L(\hat{\mathbf{r}}') d\hat{\mathbf{r}}' = \\ &= \frac{4\pi}{(2L+1)} \frac{y_M^L(\hat{\mathbf{r}})}{r^{L+1}} \int_0^\infty r'^2 f(r') \frac{r_{<}^L}{r_{>}^{L+1}} dr', \quad (7.67) \end{aligned}$$

with $r_{<} \equiv \min(r, r')$ and $r_{>} \equiv \max(r, r')$. Thanks to the linearity of the Hartree potential as a functional of the density, the solution for a generic charge density expanded in terms on angular momenta (Eq. 7.57) can be obtained as a linear superposition of the solutions expressed by equation 7.67. In what follows, we report the explicit expression for the terms involving the valence charge, *i.e.* the only quantities that have to be updated in order to solve the *QPSCGW* equations. The matrix elements of the Hartree potentials generated by the onsite AE and PS valence charge densities are given by:

$$\begin{aligned} \langle \phi_i | v_H[n^1] | \phi_j \rangle - \langle \tilde{\phi}_i | v_H[\tilde{n}^1] | \tilde{\phi}_j \rangle &= \sum_{st} \sum_{LM} \frac{4\pi}{(2L+1)} \rho_{st} \mathcal{G}_{l_i m_i l_j m_j}^{LM} \mathcal{G}_{l_s m_s l_t m_t}^{LM} \\ &\left[\iint \phi_{n_i l_i}(r) \phi_{n_j l_j}(r) \frac{r_{<}^L}{r_{>}^{L+1}} \phi_{n_s l_s}(r') \phi_{n_t l_t}(r') dr dr' - \right. \\ &\quad \left. - \iint \tilde{\phi}_{n_i l_i}(r) \tilde{\phi}_{n_j l_j}(r) \frac{r_{<}^L}{r_{>}^{L+1}} \tilde{\phi}_{n_s l_s}(r') \tilde{\phi}_{n_t l_t}(r') dr dr' \right], \quad (7.68) \end{aligned}$$

where all the terms are constants except for the occupancies ρ_{ij} which have to be recalculated at each iteration, starting from the new wave functions using equation 7.28.

The one-center contributions due to the electrostatic potential generated by the compensation charge is:

$$\langle \tilde{\phi}_i | v_H[\hat{n}] | \tilde{\phi}_j \rangle = \sum_{st} \rho_{st} \sum_{LM} q_{st}^{LM} \mathcal{G}_{l_i m_i l_j m_j}^{LM} \hat{V}_{n_i l_i n_j l_j}^L, \quad (7.69)$$

with

$$\hat{V}_{n_i l_i n_j l_j}^L \equiv \int \tilde{\phi}_{n_i l_i}(r) \hat{v}_L(r) \tilde{\phi}_{n_j l_j}(r) dr, \quad (7.70)$$

and

$$\hat{v}_L(\mathbf{r}) \equiv \frac{4\pi}{2L+1} \left\{ \frac{1}{r^{L+1}} \int_0^r s^{L+2} g_L(s) ds + r^L \int_r^{+\infty} s^{1-L} g_L(s) ds \right\}. \quad (7.71)$$

7.7 Oscillator matrix elements within PAW

In chapter 5, we saw that the oscillator matrix elements introduced in equation 5.27 constitute the basic ingredients required to evaluate both the irreducible polarizability, $\chi^0(\omega)$, and the matrix elements of the self-energy $\hat{\Sigma}(\omega)$. Within the PAW formalism, the oscillator strengths can be obtained by means of the following equation:

$$\begin{aligned} \langle \Psi_{b_1 \mathbf{k}-\mathbf{q}} | e^{-i(\mathbf{q}+\mathbf{G}) \cdot \mathbf{r}} | \Psi_{b_2 \mathbf{k}} \rangle &= \langle \tilde{\Psi}_{b_1 \mathbf{k}-\mathbf{q}} | e^{-i(\mathbf{q}+\mathbf{G}) \cdot \mathbf{r}} | \tilde{\Psi}_{b_2 \mathbf{k}} \rangle + \sum_{ij} \langle \tilde{\Psi}_{b_1 \mathbf{k}-\mathbf{q}} | \tilde{p}_i \rangle \langle \tilde{p}_j | \tilde{\Psi}_{b_2 \mathbf{k}} \rangle \\ &= e^{-i(\mathbf{q}+\mathbf{G}) \cdot \mathbf{R}_i} \underbrace{\left[\langle \phi_i | e^{-i(\mathbf{q}+\mathbf{G}) \cdot (\mathbf{r}-\mathbf{R}_i)} | \phi_j \rangle - \langle \tilde{\phi}_i | e^{-i(\mathbf{q}+\mathbf{G}) \cdot (\mathbf{r}-\mathbf{R}_i)} | \tilde{\phi}_j \rangle \right]}_{4\pi \sum_{lm} (-i)^l \mathcal{Y}_m^l(\hat{\mathbf{q}+\mathbf{G}}) \mathcal{G}_{l_i m_i l_j m_j}^{lm} \int_0^{r_c} j_l(|\mathbf{q}+\mathbf{G}|r) (\phi_{n_i l_i} \phi_{n_j l_j} - \tilde{\phi}_{n_i l_i} \tilde{\phi}_{n_j l_j}) dr} \end{aligned} \quad (7.72)$$

where the Rayleigh expansion (Eq. D.18) is used to express the plane wave in terms of Bessel functions $j_l(x)$ and real spherical harmonics $\mathcal{Y}_m^l(\hat{\mathbf{G}})$. The first term on the right hand side of equation 7.72 is analogous to the expression obtained in the norm-conserving case. Only the onsite contribution involving a sum over all atoms in the unit cell and the different (n, l, m) channels per each type have to be added. A significant memory saving is achieved by noting that the integrals involving the spherical Bessel function in equation 7.72 depend on $\mathbf{q}$ and $\mathbf{G}$ only through the modulus $|\mathbf{q} + \mathbf{G}|$. One can thus precalculate and tabulate the set of integrals:

$$F_{n_i l_i n_j l_j}^l(x) \equiv \int_0^{r_c} j_l(|x|r) (\phi_{n_i l_i} \phi_{n_j l_j} - \tilde{\phi}_{n_i l_i} \tilde{\phi}_{n_j l_j}) dr, \quad (7.73)$$

on a sufficiently large mesh in reciprocal space before entering the main loop over the electronic transitions. Then, the value at a generic $\mathbf{q} + \mathbf{G}$ vector is obtained by using an interpolating spline technique [86]. The derivative of $F_{n_i l_i n_j l_j}^l(x)$ for $x \rightarrow 0$, needed for the

spline interpolation, is readily obtainable from the asymptotic behavior of the spherical Bessel functions (equation 10.1.4 of [145]):

$$j_l(x) \underset{x \rightarrow 0}{\approx} \frac{x^l}{(2l+1)!!}, \quad (7.74)$$

$$\frac{\partial}{\partial q} \int f(r) j_l(2\pi q r) dr \underset{q \rightarrow 0}{\approx} \frac{2\pi l (2\pi q)^{l-1}}{(2l+1)!!} \int r^l f(r) dr. \quad (7.75)$$

7.8 On the Symmetrization of the PAW projections

Another point worth mentioning is the symmetrization of the PAW projections $\langle \tilde{p}_i^a | \tilde{\Psi}_{n\mathbf{k}} \rangle$ for arbitrary $\mathbf{k}$ -points in the Brillouin zone. As discussed in chapter 5, the BZ integration defining $\chi^0(\omega)$ as well as the matrix elements of $\hat{\Sigma}(\omega)$ can be reduced by symmetry only to an appropriate irreducible zone, which is, in general, larger than the irreducible wedge defined by the full point group of the crystal. As a consequence, the PAW matrix elements $\langle \tilde{p}_i^a | \tilde{\Psi}_{n\mathbf{k}} \rangle$ obtained from a band structure calculation performed in the IBZ have to be properly expanded into the full BZ before performing a *GW* calculation.

A possible solution to the problem would consist in symmetrizing the wave function $|\tilde{\Psi}_{n\mathbf{k}}\rangle$ in the full BZ using the set of equations given in equation B.6, and then calculating the overlap with the PAW projectors $\tilde{p}_i(\mathbf{r})$ via equation 7.18. This approach, on the other hand, would be CPU demanding due to the large number of bands generally considered in many-body calculations. Fortunately, it is possible to obtain a closed expression relating the matrix elements in the full Brillouin zone to their symmetrical images in the irreducible wedge. The derivation of the symmetrizing expression is briefly summarized in what follows.

The reciprocal space expression for the overlap between a wave function with crystalline momentum $\mathbf{k}$ located in the IBZ and a PAW projector is:

$$\langle \tilde{p}_i^a | \tilde{\Psi}_{n\mathbf{k}} \rangle = \frac{4\pi i^{l_i}}{\sqrt{\Omega}} \sum_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G}) \cdot \mathbf{R}_a} y_{m_i}^{l_i}(\widehat{\mathbf{k}+\mathbf{G}}) \tilde{u}_{n\mathbf{k}}(\mathbf{G}) \tilde{p}_{n_i l_i}^a(|\mathbf{k}+\mathbf{G}|). \quad (7.76)$$

The matrix element at the point $\mathcal{R}\mathbf{k}$, located in the full BZ, is obtained by considering how the periodic part of the wave function $\tilde{u}_{n\mathbf{k}}(\mathbf{G})$ and the real spherical harmonic $y_m^l(\widehat{\mathbf{k}+\mathbf{G}})$ transform upon the application of the rotation $\mathcal{R}$ (Eq. B.6 and Eq. D.16, respectively). The final result is:

$$\begin{aligned}
\langle \tilde{p}_i^a | \tilde{\Psi}_{n\mathbf{R}\mathbf{k}} \rangle &= \\
\frac{4\pi i^{l_i}}{\sqrt{\Omega}} \sum_{\mathbf{G}} e^{i(\mathbf{k}+\mathcal{R}^{-1}\mathbf{G})\cdot\mathcal{R}^{-1}\mathbf{R}_a} y_{m_i}^{l_i} \left(\mathcal{R}(\widehat{\mathbf{k}+\mathcal{R}^{-1}\mathbf{G}}) \right) \tilde{u}_{n\mathbf{R}\mathbf{k}}(\mathbf{G}) \times \tilde{p}_{n_i l_i}^a(|\mathbf{k}+\mathcal{R}^{-1}\mathbf{G}|) &= \\
= \frac{4\pi i^{l_i}}{\sqrt{\Omega}} \sum_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathcal{R}^{-1}(\mathbf{R}_a-\mathbf{t})} y_{m_i}^{l_i} \left(\mathcal{R}(\widehat{\mathbf{k}+\mathbf{G}}) \right) \tilde{u}_{n\mathbf{k}}(\mathbf{G}) \tilde{p}_{n_i l_i}^a(|\mathbf{k}+\mathbf{G}|) &= \\
= e^{i\mathbf{k}\cdot\mathbf{L}} \sum_{\alpha} D_{\alpha m_i}^{l_i}(\mathcal{R}^{-1}) \langle \tilde{p}_{n_i l_i \alpha}^{a'} | \tilde{\Psi}_{n\mathbf{k}} \rangle & \quad (7.77)
\end{aligned}$$

So the matrix elements in the BZ are given by an appropriate linear combination of the projections at the symmetrical $\mathbf{k}$ -point in the IBZ. In equation 7.77, a' is the symmetric image of atom a , supposed to be located within the unit cell, *i.e.* $\mathcal{R}^{-1}(\mathbf{R}_a - \mathbf{t}) = \mathbf{R}^{a'} + \mathbf{L}$, where $\mathbf{L}$ is, a possibly null, direct lattice vector needed to bring the rotated atom back into the first unit cell. Finally, $D_{m'm}(\mathcal{R})$ is the matrix describing the rotation of the real spherical harmonic whose explicit expression can be found in section D.3. In systems without inversion symmetry, it is possible to take advantage of time-reversal symmetry to reduce the number of points further:

$$\langle \tilde{p}_i^a | \tilde{\Psi}_{n-\mathbf{k}} \rangle = \langle \tilde{p}_i^a | \tilde{\Psi}_{n\mathbf{k}} \rangle^\dagger. \quad (7.78)$$

Equations 7.77 and 7.78 allows significant CPU saving especially when a large number of empty states have to be symmetrized and the unit cell contains several atoms. The matrix elements in the full BZ can be obtained with inexpensive matrix-vector multiplications on small tables without the need of performing a full symmetrization of the wave functions, either in real or reciprocal space. It is important to stress that the symmetry relations hold only for non-degenerate states. As already discussed in section 5.4.3, the simplified symmetrizing expression can be safely used during the calculation of the irreducible polarizability, while an average of the QP corrections for degenerate states is needed to preserve the correct band degeneracies in the self-energy (Sect. 5.7.1.1).

7.9 Testing the GW + PAW Implementation

To validate our implementation, we have performed test *GW* calculations for crystalline silicon. Our results are compared with recently published calculations obtained within similar GW codes employing the PAW method: the original implementation of Arnaud and Alouani based on Blöchl's code [14], and the Vasp package [104]. The matrix elements of the valence-only XC potential in Si have already been compared in table 7.2, where good agreement with Arnaud's work was found. Here we focus on the final results for the QP energies of silicon, whose values are grouped in table 7.4. The parameters used for our calculation are reported in the caption of the table. They are very similar to the ones used in [104]. The chosen setup ensures the convergence of the QP energies to within 0.05 eV.

$\mathbf{k}$	E_{QP} Present	E_{QP} [14]	E_{QP} [104]
Γ_{1v}	-11.73	-11.96	-11.75
Γ_{25v}	0.0	0.0	0.00
Γ_{15v}	3.10	3.15	3.17
Γ_{2c}	4.00	4.19	4.04
X_{1v}	-7.66	-7.92	-7.68
X_{4v}	-2.88	-3.01	-2.88
X_{1c}	1.12	1.15	1.19
X_{3c}	10.44	10.75	10.54
L_{2v}	-9.46	-9.61	-9.46
L_{1v}	-6.87	-7.17	-6.87
L_{3v}	-1.21	-1.26	-1.22
L_{1c}	2.06	2.16	2.10
L_{3c}	3.90	3.95	3.93

Table 7.4: Quasiparticle energies of Si at high symmetry $\mathbf{k}$ -points (all in eV). The calculation has been performed using 160 bands both for Σ and $\tilde{\chi}$, and a $6 \times 6 \times 6$ $\mathbf{k}$ -point grid centered on Γ . The self-energy is evaluated using the CD method described in section 5.7.2.2.

As can be seen from table 7.2, the agreement with the other implementations is quite satisfying. Our results compare very well with those obtained with the Vasp code (Ref. [104]) while a somewhat larger discrepancy is found with respect to the first implementation of Arnaud, which relies on the plasmon-pole approximation. The good agreement with Vasp results is somewhat expected because the parameters used for the GW calculation as well as the integration scheme for $\hat{\Sigma}(\omega)$ are similar. As concerns the discrepancy with the results reported in [14], we believe that the main source of it can be attributed to the use of the PPM approximation in the cited paper. This hypothesis is confirmed by the fact that subsequent plasmon-pole-free calculations reported by the same authors in [64] gave results that are closer to ours.

A more stringent test for the reliability of the ABINIT implementation is represented by the calculation of the frequency dependence of the self-energy, whose values are quite sensitive to the computational details. Figure 7.2 reports the real and imaginary part of $\hat{\Sigma}_c(\omega)$ for the first eight bands of silicon at $\mathbf{k} = \Gamma$. The two panels can be directly compared to Fig. 1 of [64]. In this case the agreement is also good, the only noticeable difference being that our curves for $\Sigma_c(\omega)$ are, generally speaking, less smooth. This is essentially because, in [64], the self-energy along the real axis is obtained by means of a Padé extrapolation, whereas we opted for the more accurate, but also more computationally demanding CD integration. The spectral function $A(\omega)$ (Eq. 5.109) for the first eight bands of silicon at $\mathbf{k} = \Gamma$ is displayed in figure 7.3. The sharp peaks and the incoherent

part compares well with [64], while only the absolute value of $A(\omega)$ is different due to a missing π prefactor in the expression for the spectral function used by the authors (Eq 20 of [64]).

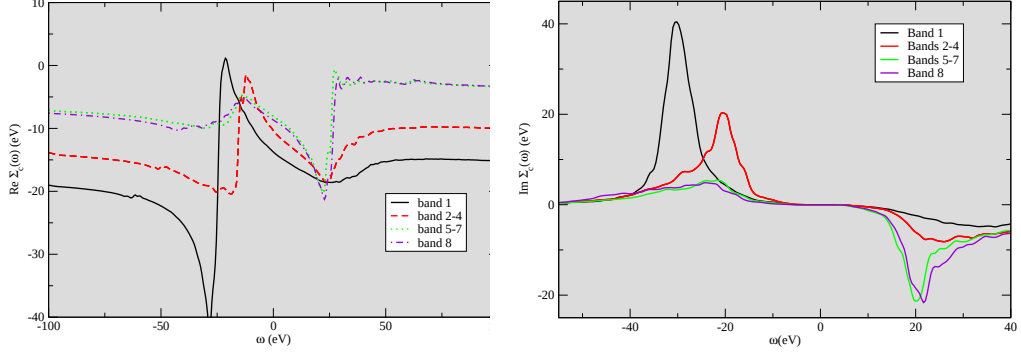


Figure 7.2: Left panel: Real part of $\Sigma_c(\omega)$ in silicon shown for the first eight bands at $\mathbf{k} = \Gamma$. Right panel: Imaginary part of $\Sigma_c(\omega)$. In both pictures, the zero of the energy is at the center of the band gap.

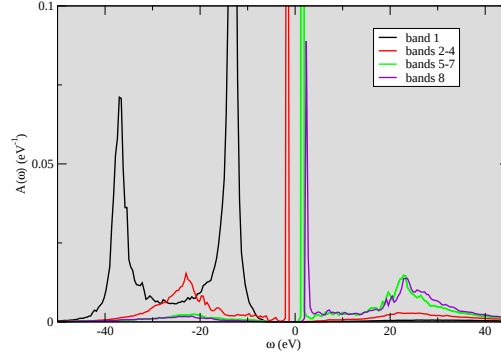


Figure 7.3: Spectral function $\langle \Psi_{n\mathbf{k}} | \hat{A}(\omega) | \Psi_{n\mathbf{k}} \rangle$ for the first eight bands of silicon at $\mathbf{k} = \Gamma$. The zero of the energy is at the center of the band gap. The sharp peaks are QP poles, their weight corresponds to the renormalization factor Z defined in equation 4.58.

We have also performed calculations for SiC and C in the diamond structure in which a partial self-consistency has been obtained by substituting the energies entering the denominator of the Green's function with QP energies, (hereafter referred to as e- G_0W_0). A comparison of selected QP transitions is given in the two tables 7.6 and 7.8.

Another feature of the implementation worth testing concerns the calculation of optical properties. According to Adler [42] and Wiser [43], the macroscopic dielectric function, $\epsilon_M^{\text{LF}}(\omega)$, can be directly related to the inverse of the microscopic dielectric matrix

k	LDA	G_0W_0	$e - G_0W_0$	Expt.
$\Gamma'_{25v} \rightarrow \Gamma_{1c}$	5.59 (5.53)	7.30 (7.39)	7.43 (7.41)	7.3
$\Gamma'_{25v} \rightarrow X_{1c}$	4.74 (4.61)	6.12 (6.19)	6.25 (6.27)	
$\Gamma'_{25v} \rightarrow L_{1c}$	8.46 (8.38)	10.23 (10.36)	10.38 (10.38)	
E_g	4.10 (4.01)	5.59 (5.60)	5.72 -	5.48

Table 7.6: Quasiparticle optical band gaps for diamond at Γ , L , and X (all in eV). The BZ is sampled with $6 \times 6 \times 6$ $\mathbf{k}$ -mesh, 160 bands for G and W , 8 Hartree for W , 25 Hartree for the wavefunctions, 20 Hartree for Σ_x . Data in parenthesis are from [14]. Experimental results are from [146]

k	LDA	G_0W_0	$e - G_0W_0$	Expt.
$\Gamma'_{25v} \rightarrow \Gamma_{15c}$	6.26 (6.25)	7.42 (7.32)	7.51 (7.45)	
$\Gamma'_{25v} \rightarrow X_{1c}$	1.30 <u>(1.28)</u>	2.17 <u>(1.80)</u>	2.25 <u>(1.89)</u>	<u>2.39</u>
$\Gamma'_{25v} \rightarrow L_{1c}$	5.37 (5.34)	6.50 (6.45)	6.63 (6.56)	4.2

Table 7.8: Quasiparticle optical band gaps for SiC at Γ , L , and X (all in eV). The BZ is sampled with $6 \times 6 \times 6$ $\mathbf{k}$ -mesh, 160 bands for G and W , 8 Hartree for W , 25 Hartree for the wavefunctions, 20 Hartree for Σ_x . The values for the fundamental gap are underlined. Data in parenthesis are from [64]. Experimental results are from [146]

by means of:

$$\epsilon_M^{\text{LF}}(\omega) = \lim_{\mathbf{q} \rightarrow 0} \frac{1}{\epsilon_{00}^{-1}(\mathbf{q}, \omega)} \quad (7.79)$$

The optical absorption spectrum – the quantity one can compare with experiments – is given by the imaginary part $\epsilon_M^{\text{LF}}(\omega)$. Note that equation 7.79 differs from

$$\epsilon_M^{\text{NLF}}(\omega) = \lim_{\mathbf{q} \rightarrow 0} \epsilon_{00}(\mathbf{q}, \omega) \quad (7.80)$$

due to the so called local-field effects introduced by the presence of the crystalline environment

Both equation 7.79 and equation 7.80 require the evaluation of the limit $\mathbf{q} + \mathbf{G} \rightarrow 0$ in the oscillator matrix elements defined in equation 7.72. The discussion of the technical details needed to evaluate the optical limit within the PAW formalism are relegated to Appendix E. Here, we only wish to discuss the outcome of our test calculations.

Figure 7.4 shows the optical absorption spectrum of Si and GaAs with and without local field effects. Figure 7.4 can be directly compared to Fig 1 of Ref. [147]. Both the calculation for Si and GaAs were performed by employing the Hilbert transform method described in section 5.6. The imaginary part of each matrix element was evaluated in energy intervals of 0.1 eV up to 200 eV, and 8000 $\mathbf{k}$ points were used to sample the full Brillouin zone. Our results compare well with the spectra calculated in [147]. The fine details are not perfectly reproduced, but this is mainly because, in reference [147], the tetrahedron method is employed to perform the integration in $\mathbf{k}$ -space. In contrast, our results have been obtained by smearing the contribution of each transition with a Lorentzian broadening (the complex shift present in the denominator of equation 5.29). A value of $\delta = 0.01$ eV was used in our calculation to reproduce the amplitude of the peaks reported in the referenced paper.

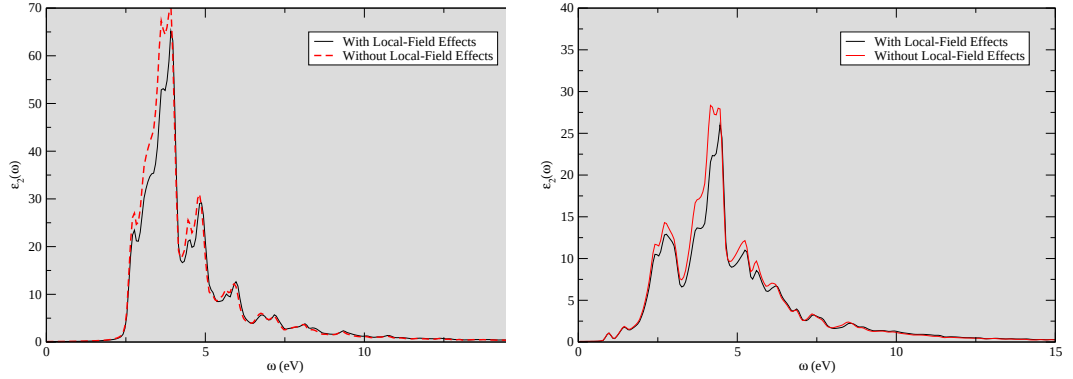


Figure 7.4: Left panel: Calculated LDA imaginary part of the dielectric function of Si. To be compared with Fig 1 of [147] Right panel: Calculated LDA imaginary part of the dielectric function of GaAs.

7.10 QPSCGW with PAW

In this section, the changes needed to implement the constrained self-consistent *GW* method proposed by Faleev and co-workers [16] are briefly described. The results obtained with our implementations are then compared to recently published calculations. For the sake of completeness, the main ideas underlying the quasiparticle self-consistent *GW* method are briefly summarized before exposing the technical details of the PAW implementation.

In the *QPSCGW* method, one has to diagonalize the operator:

$$\hat{H}_{\text{QPSCGW}} = \hat{h}_0[n](\mathbf{r}_1) + \tilde{\Sigma}(\mathbf{r}_1, \mathbf{r}_2), \quad (7.81)$$

where the bare Hamiltonian is generally defined as:

$$\hat{h}_0[n](\mathbf{r}) = -\frac{1}{2}\Delta^2 + v_H[n](\mathbf{r}) + v_{\text{ext}}(\mathbf{r}). \quad (7.82)$$

An Hermitian, energy-independent approximation to the self-energy is constructed according to:

$$\tilde{\Sigma}_{ij}^{\mathbf{k}} \equiv \frac{1}{2} \text{Herm} \left[\Sigma_{ij}^{\mathbf{k}}(\epsilon_{i\mathbf{k}}) + \Sigma_{ij}^{\mathbf{k}}(\epsilon_{j\mathbf{k}}) \right], \quad (7.83)$$

where Herm means that only the Hermitian part has to be considered. Then the approximated self-energy operator is diagonalized providing a new set of quasiparticle amplitudes and energies. From this new set a new density, and the corresponding Hartree potential is generated. A new approximated self-energy is constructed, and the procedure is iterated to self-consistency.

As already described in section 4.10, the ABINIT code employs KS orbitals, $|\Psi_{n\mathbf{k}}\rangle$, to expand the QP amplitudes, $|\Phi_{n\mathbf{k}}\rangle$, such that the expensive calculation of the plane wave representation of $\hat{\Sigma}(\omega)$ is avoided. An analogous approach is followed also for the PAW implementation. Following this idea, the pseudized version of the QP amplitudes are related to the KS eigenstates by means of the following relation:

$$|\tilde{\Phi}_{m\mathbf{k}}\rangle = \sum_n \mathcal{U}_{nm}^{\mathbf{k}} |\tilde{\Psi}_{n\mathbf{k}}\rangle, \quad (7.84)$$

where $\mathcal{U}^{\mathbf{k}}$ is a unitary transformation. Similarly, the overlap of the QP pseudo amplitudes with the PAW projector functions have to be consistently redefined in terms of the KS quantities:

$$\langle \tilde{p}_i | \tilde{\Phi}_{m\mathbf{k}} \rangle = \sum_n \mathcal{U}_{nm}^{\mathbf{k}} \langle \tilde{p}_i | \tilde{\Psi}_{n\mathbf{k}} \rangle. \quad (7.85)$$

In the PAW method, a complication arises because the pseudized version of the mean-field Hamiltonian has a functional dependence on the quasiparticle density not only through the Hartree potential, but also through the pseudopotential strengths D_{ij} . As a consequence, the D_{ij} terms – more precisely the Hartree-like contributions – also have to be

updated during the *QSPSCGW* iterations to reflect the changes in the QP valence density. In what follows, we provide a brief description of the way the different contributions have to be handled. One has for the total PAW pseudo strengths:

$$D_{ij} \approx D_{ij}^0 + (D_{ij})^H + (D_{ij}^{\text{xc}})_{\text{val}} + (D_{ij}^{\text{xc}})_{\text{core}}, \quad (7.86)$$

where we have assumed that the exchange-correlation potential can be separated into core and valence contributions.⁴ The update of the Hartree-like part, $(D_{ij})^H$, is straightforward thanks to the linearity of the Hartree potential as a functional of the different densities. A new value for the $(D_{ij})^H$ term is obtained starting from the expression reported in section 7.6.2, by just replacing the KS occupancies with the corresponding QP values calculated from the new quasiparticle amplitudes using equation 7.28.

The treatment of the XC-like contribution to D_{ij} is more involved, since the valence contribution to the XC part has to be subtracted from the total D_{ij} in order to obtain the bare Hamiltonian. Under the assumption of linearity of $v_{\text{xc}}[n](\mathbf{r})$, one has:

$$\begin{aligned} (D_{ij}^{\text{xc}})_{\text{val}} = & \langle \phi_i | v_{\text{xc}}[n^1] | \phi_j \rangle - \langle \tilde{\phi}_i | v_{\text{xc}}[\tilde{n}^1 + \hat{n}] | \tilde{\phi}_j \rangle - \sum_{LM} \int_{\Omega_{\mathbf{R}}} v_{\text{xc}}[\tilde{n}^1 + \hat{n}] \hat{Q}_{ij}^{LM}(\mathbf{r}) d\mathbf{r} + \\ & + \sum_{LM} \int_{\Omega} v_{\text{xc}}[\tilde{n} + \hat{n}] \hat{Q}_{ij}^{LM}(\mathbf{r}) d\mathbf{r} \end{aligned} \quad (7.87)$$

and

$$\begin{aligned} (D_{ij}^{\text{xc}})_{\text{core}} = & \langle \phi_i | v_{\text{xc}}[n_c] | \phi_j \rangle - \langle \tilde{\phi}_i | v_{\text{xc}}[\tilde{n}_c] | \tilde{\phi}_j \rangle - \sum_{LM} \int_{\Omega_{\mathbf{R}}} v_{\text{xc}}[\tilde{n}_c] \hat{Q}_{ij}^{LM}(\mathbf{r}) d\mathbf{r} + \\ & + \sum_{LM} \int_{\Omega} v_{\text{xc}}[\tilde{n}_c] \hat{Q}_{ij}^{LM}(\mathbf{r}) d\mathbf{r}. \end{aligned} \quad (7.88)$$

Finally, the bare PAW Hamiltonian can be rewritten as:

$$\begin{aligned} \hat{h}_0 = & -\frac{1}{2}\Delta + v_H[\tilde{n}^{\text{QP}} + \hat{n}^{\text{QP}} + \tilde{n}_{Zc}] + v_{\text{xc}}[\tilde{n}_c] + \\ & + \sum_{ij} |\tilde{p}_i\rangle \left[D_{ij}^0 + (D_{ij})^H + (D_{ij}^{\text{xc}})_{\text{core}} \right] \langle \tilde{p}_i|. \end{aligned} \quad (7.89)$$

Contrary to the case of norm-conserving pseudopotentials in which $\hat{h}_0$ is obtained from the KS eigenvalues by subtracting the XC potential, in PAW the different contributions

⁴ Equation 7.86 is not exact due to the nonlinearity of the XC potential, but represents a reasonable approximation for atoms in which core and valence are spatially well separated. The approximation would be less reliable for the case of atoms with shallow cores which, however, are usually treated as valence when performing *GW* calculations.

to $\hat{h}_0$ have to be explicitly evaluated at each *QPSCGW* iteration. The implementation of Faleev's method within PAW is more involved but, despite the complication of the formalism, the computing effort needed to assemble the Hamiltonian in equation 7.89 represents only a small fraction of the total CPU time required to calculate the matrix elements of $\hat{\Sigma}(\omega)$.

In table 7.10, the fundamental gaps of selected prototype systems calculated with the present implementation are compared to results available in the literature. The values obtained with the present implementation are in reasonable – although not perfect agreement – with the gaps reported in [144]. A possible source of discrepancies might be attributed to the use of the plasmon-pole approximation versus the numerical integration of the self-energy employed in the Vasp code. Another source of discrepancy is due to the different PAW setup being employed. The calculation in [104] uses a high-quality PAW basis set with three projectors per angular channel, while our set of PAW partial waves is more efficient, but less transferable, since only two projectors per each l are used.

	GGA	G_0W_0	e- GW_0	GW_0
Si [144]	0.62	1.12	1.20	1.28
Present	0.59	1.29	1.39	1.39
Exp [97]	1.17	-	-	-
SiC [144]	1.35	2.27	2.43	2.64
Present	1.35	2.13	2.25	2.47
Exp [97]	2.40	-	-	-
C [144]	4.12	5.50	5.68	5.99
Present	4.14	5.60	5.75	5.82
Exp [97]	5.48	-	-	-

Table 7.10: Comparison between the fundamental gap of Si, SiC and C calculated with the present PAW implementation and the Vasp code [144] (all in eV). e- GW_0 indicates a *QPSCGW* calculation in which only the energies are updated in the Green's function while the wave functions are treated at the KS level. The results labeled with GW_0 are obtained by performing a *QPSCGW* calculation in which the screened interaction is treated at the RPA level.

Chapter 8

Wannier Interpolation

8.1 Introduction

In this chapter, we present an efficient and reliable interpolation scheme for quasiparticle energies based on the use of the MLWFs introduced by Marzari and Vanderbilt in [17]. The MLWFs are functions localized in real-space, spanning the same Hilbert subspace as that described by the set of Bloch eigenstates used for their construction. The interpolation technique based on Wannier functions does not reduce the overall scaling of self-energy calculations, but facilitates the analysis of band structure dispersions and the study of integrated quantities such as, for instance, the QP density of states. The application of MLWFs is by no means limited to many-body calculations. Among the several possible applications outside the realm of *GW*, one particularly worth mentioning is the interpolation of electron-phonon matrix elements presented in [148, 149].

8.2 The Wannier Representation

A set of Wannier Functions (WFs), $|n\mathbf{R}\rangle$, labeled by the Bravais lattice vectors $\mathbf{R}$ and the band index n , are constructed from Bloch states by performing the unitary transformation:

$$w_n(\mathbf{r} - \mathbf{R}) = \langle \mathbf{r} | n\mathbf{R} \rangle = \frac{1}{\Omega_{\text{BZ}}} \int_{\text{BZ}} e^{-i\mathbf{k}\cdot\mathbf{R}} \psi_{n\mathbf{k}}(\mathbf{r}) d\mathbf{k}, \quad (8.1)$$

where Ω_{BZ} is the volume of the Brillouin zone. WFs represent a convenient basis set since they span the same Hilbert subspace as the initial Bloch functions and, in addition, are both orthogonal and localized in real-space. For this reason, WFs are frequently used for formal discussions and demonstrations in which these properties are employed without even the need for an explicit construction. WFs also play a central role in the theory of electronic polarizability [150, 151, 152] and represent a promising approach for the development of linear scaling algorithms for electronic structure calculations [153]. However, their widespread use in the *ab initio* community has been hindered by their nonuniqueness. Bloch wavefunctions are invariant under a different choice of phase. For

a single isolated band, for example, it is always possible to modify the phase according to:

$$u_{n\mathbf{k}} \rightarrow e^{i\phi_n(\mathbf{k})} u_{n\mathbf{k}}, \quad (8.2)$$

whereas in the case of an isolated group of bands, the allowed transformations are of the more general form:

$$u_{n\mathbf{k}} = \sum_{m=1}^N U_{mn}^{\mathbf{k}} u_{m\mathbf{k}}, \quad (8.3)$$

where the unitary matrix $U_{mn}^{\mathbf{k}}$ mixes bands with the same wave-vector $\mathbf{k}$. Unfortunately, the Wannier transform defined in equation 8.1 is not gauge invariant. Consequently both the shape and the extent of the Wannier functions will in general depend on the particular gauge, and this affects the degree of localization and the quality of the Wannier basis set.

Marzari and Vanderbilt [17] developed a method for constructing MLWFs from an isolated group of bands by finding the particular gauge which minimizes the sum of the quadratic spreads defined as:

$$\Omega = \sum_{n=1}^N \langle r^2 \rangle_n - \langle \mathbf{r} \rangle_n^2, \quad (8.4)$$

where the sum is over the chosen group of bands and

$$\langle \mathbf{r} \rangle_n = \int \mathbf{r} |w_n(\mathbf{r})|^2 d\mathbf{r}. \quad (8.5)$$

The technique for obtaining MLWFs has been subsequently generalized by Sousa and co-workers [154] to the case of entangled bands, *i.e.* a set of bands that cannot be separated in energy from the remaining states. In this particular case, an energy window – the outer window – is introduced, and a set of optimally connected Bloch states is selected by performing a unitary transformation among the states falling within the energy window:

$$|u_{n\mathbf{k}}^{(\text{opt})}\rangle = \sum_{m \in N_{\text{win}}^{\mathbf{k}}} \mathcal{U}_{mn}^{\mathbf{k}} |u_{m\mathbf{k}}\rangle. \quad (8.6)$$

In equation 8.6, $N_{\text{win}}^{\mathbf{k}}$ indicates the number of Bloch states within the outer window at a given $\mathbf{k}$ -point while $\mathcal{U}^{\mathbf{k}}$ is a rectangular $N \times N_{\text{win}}^{\mathbf{k}}$ matrix. As discussed in [154], the matrices $\mathcal{U}^{\mathbf{k}}$ are obtained by minimizing the gauge invariant part of the spread 8.4 within the outer energy window. Once the matrices $\mathcal{U}^{\mathbf{k}}$ are found, the standard procedure for isolated bands can be used to obtain MLWFs spanning this optimal subspace. It should be noted that the energy bands of this optimal subspace may not correspond to any of the original energy bands due to the mixing among states. In order to exactly preserve the properties of a system in a given energy range, a second energy window is introduced, such that all states lying within the inner energy window are included unchanged in the optimal subspace.

8.3 Wannier Interpolation of Matrix Elements

In what follows, we describe the methodological aspects for the interpolation of a generic periodic operator using a set of MLWFs [155]. Our goal is to use MLWFs to interpolate the matrix elements of a generic operator $\hat{O}(\mathbf{q})$ onto a dense mesh in the BZ. We assume that the unitary $U_{mn}(\mathbf{q})$ is already known, and that the matrix elements to be interpolated are the results of a first-principles calculation.

The Wannier interpolation is a two-step procedure. In the first step, one calculates the matrix elements of the operator between two Wannier functions, in the second step, the matrix elements in real-space are Fourier-transformed to $\mathbf{k}$ -space on an arbitrarily dense mesh. The mathematical details of the interpolation are described in what follows.

At each $\mathbf{q}$, the $N_{\mathbf{q}} \times M$ matrix $U(\mathbf{q})$ relates the original set of $N_{\mathbf{q}}$ Bloch eigenvectors within the outer window to the new set of M orthonormal Bloch-like states

$$|u_{n\mathbf{q}}^{(W)}\rangle = \sum_{m=1}^{N_{\mathbf{q}}} U_{mn}(\mathbf{q}) |u_{m\mathbf{q}}\rangle, \quad (8.7)$$

that vary smoothly with $\mathbf{q}$. These states are labeled with a superscript (W) to indicate that the MLWFs are obtained from them by the direct Fourier transform:

$$|n\mathbf{R}\rangle = \frac{1}{N_{\mathbf{q}}} \sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot\mathbf{R}} |u_{n\mathbf{q}}^{(W)}\rangle, \quad (8.8)$$

where the sum runs over the grid of *ab initio* $N_{\mathbf{q}}$ points, and $|n\mathbf{R}\rangle$ is the Wannier function “belonging” to the unit cell indicated by the lattice vector $\mathbf{R}$. The matrix elements of the Hamiltonian in the maximally localized Wannier gauge, $H_{mn}^{(W)}(\mathbf{q})$, can be obtained from the knowledge of the unitary transformation U according to:

$$H_{mn}^{(W)}(\mathbf{q}) = \langle u_{n\mathbf{q}}^{(W)} | \hat{H}(\mathbf{q}) | u_{m\mathbf{q}}^{(W)} \rangle = \left[U^*(\mathbf{q}) H(\mathbf{q}) U(\mathbf{q}) \right]_{nm}, \quad (8.9)$$

where $H_{nm}(\mathbf{q}) = \delta_{nm} \epsilon_n(\mathbf{q})$ is a diagonal matrix containing the KS eigenvalues $\epsilon_n(\mathbf{q})$. Here, the following notation has been introduced:

$$\hat{H}(\mathbf{q}) = e^{-i\mathbf{q}\cdot\mathbf{r}} \hat{H} e^{i\mathbf{q}\cdot\mathbf{r}}. \quad (8.10)$$

The matrix $H_{mn}^{(W)}(\mathbf{q})$ is diagonalized by finding an $M \times M$ unitary matrix, $I(\mathbf{q})$, such that

$$I^\dagger(\mathbf{q}) H^{(W)}(\mathbf{q}) I(\mathbf{q}) = H^{(H)}(\mathbf{q}) \quad (8.11)$$

where

$$\left[H^{(H)}(\mathbf{q}) \right]_{nm} = \delta_{nm} \epsilon_n^{(H)}(\mathbf{q}), \quad (8.12)$$

with $\epsilon_n^{(H)}(\mathbf{q})$ identical to the original *ab initio* eigenvalues. The matrix elements of the operator $\hat{O}$ in the Wannier gauge are defined by

$$O_{mn}^{(W)}(\mathbf{q}) = \langle u_{n\mathbf{q}}^{(W)} | \hat{O}(\mathbf{q}) | u_{m\mathbf{q}}^{(W)} \rangle = [U^\dagger O(\mathbf{q}) U(\mathbf{q})]_{nm}. \quad (8.13)$$

The discrete Fourier transform of equation 8.13 gives the matrix elements of $\hat{O}$ between Wannier functions, *i.e.* :

$$O_{nm}^{(W)}(\mathbf{R}) = \langle n\mathbf{0} | \hat{O} | m\mathbf{R} \rangle, \quad (8.14)$$

with $\mathbf{0}$ being the origin of the real-space lattice, and:

$$O_{nm}^{(W)}(\mathbf{R}) = \frac{1}{N_{\mathbf{q}}} \sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot\mathbf{R}} O_{nm}^{(W)}(\mathbf{q}). \quad (8.15)$$

The inverse Fourier transform:

$$O_{nm}^{(W)}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} O_{nm}^{(W)}(\mathbf{R}), \quad (8.16)$$

yields the interpolation of the matrix elements of the operator in the Wannier gauge at an arbitrary point $\mathbf{k}$. Since $|O_{nm}^{(W)}(\mathbf{R})|$ is expected to decay exponentially if the MLWFs are exponentially localized, equation 8.16 should require few lattice points to converge, and relatively coarse *ab initio* $\mathbf{q}$ -meshes should be sufficient to obtain accurate interpolations.¹ The final step consists of transforming the matrix elements in equation 8.16 from the Wannier to the Hamiltonian gauge. The transformation is performed by finding the unitary matrix Λ which diagonalizes the Hamiltonian in the Wannier gauge at the $\mathbf{k}$ -point of interest, that is:

$$\Lambda^\dagger(\mathbf{k}) H^{(W)}(\mathbf{k}) \Lambda(\mathbf{k}) = \delta_{nm} \epsilon_n(\mathbf{q}). \quad (8.17)$$

8.4 Wannier interpolation of *GW* results

The ABINIT code already has an interface to the Wannier90 package [157] that can be used to obtain MLWFs starting from a standard band structure calculation. An interpolation of *GW* results based on the use of Wannier functions has been recently proposed and implemented in ABINIT by Hamann [158].

¹ In the original paper of Marzari and Vanderbilt [17], the authors pointed out that all MLWFs obtained in their study interestingly turned out to be real and exponentially localized. However, they were not able to provide a rigorous proof of the exponential localization of MLWFs for the general three-dimensional case. This long-standing question has only recently been solved, and the exponential localization of Wannier functions in two or three dimensions has been proved for all insulators with time-reversal symmetry [156]. An explicit condition for the reality of the Wannier functions has also been identified.

In this work, we have extended Hamann’s implementation to the case of PAW and, more importantly, we have developed a library which implements the Wannier interpolation described in the previous section. Our implementation does not rely on the Wannier90 package anymore as it only requires an external binary file containing the unitary transformation defining the Wannier gauge. Our work is not original since it is based on preexisting implementations, but presents the big advantage of being modular and reusable for the interpolation of arbitrary $\mathbf{k}$ -dependent matrix elements. This opens the way towards disparate applications ranging from the simple interpolation of QP energies up to more involved applications such as the interpolation of oscillator matrix elements or the study of optical properties in the long-wavelength limit. The latter quantities are known to converge slowly with respect to the BZ sampling [81]. The implementation of some of these approaches is presently underway.

In this paragraph, we report test calculations performed in silicon within the PAW method. The results are gathered in figure 8.1 which presents the interpolated LDA and G_0W_0 band dispersion as well as the corresponding density of states. The initial *ab initio* calculations have been performed on an $8 \times 8 \times 8$ $\mathbf{k}$ -mesh and the values have been subsequently interpolated on the $L - \Gamma - X - K - \Gamma$ path. An interpolating $24 \times 24 \times 24$ $\mathbf{k}$ -mesh is used for the calculation of the density of states by employing the tetrahedron method [39, 159]. The Wannier functions used for the interpolation are shown in figure 8.1, they correspond to the first eight energy bands of silicon and have an average spread of the centers (Eq. 8.4) of around 14 Bohr. The density of the *ab initio* $8 \times 8 \times 8$ $\mathbf{k}$ -mesh assures that all the generated Wannier functions are well localized within the Born-von Kármán supercell, hence spurious effects due to the superposition of periodically repeated orbitals is expected to be negligible.

A point worth noting is that the localization of the Wannier functions strongly depends on the set of bands included in the MLWF calculation. For instance, a set of four Wannier orbitals corresponding to the valence states of Si (figure 8.3) leads to a much better localization with an average spread of 6.8 Bohr. Numerical tests have revealed that interpolating the energies of the first four bands using the eight functions displayed in figure 8.2 or the first four Wannier functions spanning the valence manifold in figure 8.3 produces almost indistinguishable results. On the other hand, we cannot exclude *a priori* that more complicated systems will need much denser *ab initio* meshes in order to obtain accurate Wannier interpolation. According to our experience, finding a set of well localized Wannier functions is not an easy task, especially if a large number of states is considered. Whether this difficulty is related to the algorithmic procedures used to find the optimal Wannier gauge or if it is an intrinsic problem due to the delocalization of the conduction states is still an open issue deserving further investigation.

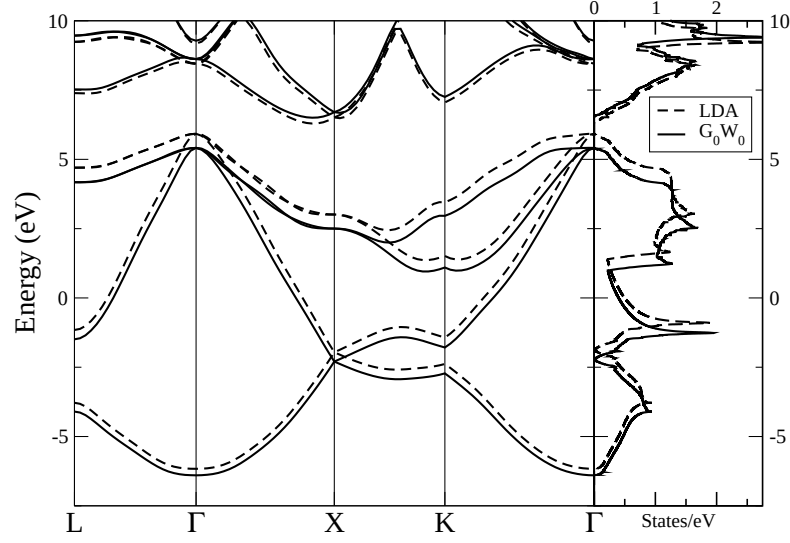


Figure 8.1: LDA and G_0W_0 band energies and density of states of Si. The results have been interpolated using the first eight Wannier functions shown in figure 8.2 using an $8 \times 8 \times 8$ *ab initio* $\mathbf{k}$ -mesh.

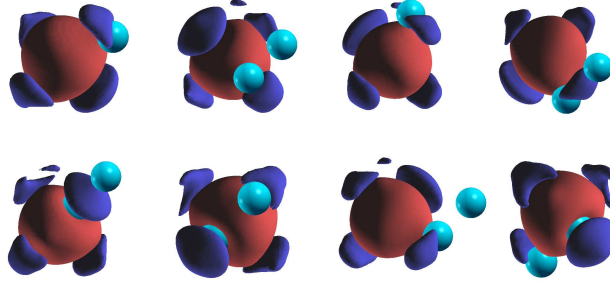


Figure 8.2: Isosurfaces for the first eight Wannier functions in Si. The MLFs have been obtained with the disentanglement procedure. An outer energy window of 17 eV and a set of “frozen” states spanning an energy range up 10 eV is used. The average spread of the Wannier functions is 14 Bohr.

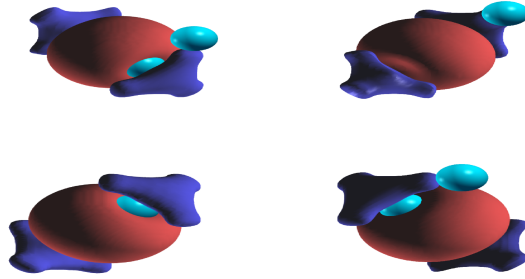


Figure 8.3: Isosurfaces associated to the first four states spanning the valence manifold of Si. The average spread of the Wannier centers is 6.8 Bohr.

Chapter 9

Conclusions and Perspectives

In the preceding pages, we have presented the result of three and half years of work spent mainly struggling with equations, compilers and debuggers rather than running calculations to calculate the properties of materials at the nanoscale. As it is always difficult to write down a unique conclusion after such long period of study, we think it is worth summarizing here what has been the historical development of this work, from the original project to the final achievements. We started our PhD project with the idea of understanding the optical properties of amorphous silicon dioxide in the presence of defects and vacancies, but we immediately had to face several difficulties. The *GW* code was not parallelized, and the memory requirements were exceedingly large for the computing facilities available at the time. Even the study of crystalline systems of moderate size was too computationally demanding, as the code was not able to take advantage of symmetry properties. Hence, we had to choose between two possibilities: either using the inadequate code at hand for painful calculations or taking the long road via a thorough rewriting of the code to make it more modular, faster and scalable. As the reader might have gathered from the overall discussion presented in the preceding chapters, we opted for the latter choice. Most of our work in the past two years has been devoted to removing bottlenecks, and this thesis summarizes some of the improvements that have been added, although many other aspects, projects and collaborations have been omitted.

We would have desired to have many more physical results before the end of our thesis but, of course, one should stop somewhere. We are aware that the originality of the presented results is somehow limited, but it is also true that scientific results and publications are generally the outcome of a patient and sometimes painful preparatory phase (usually assigned to PhD students) in which previous results are reproduced, sometimes revisited or even criticized. Only after this preliminary step, which is necessary to fill the gap with respect to previous works, one can extend what has already been done and, finally, tackle more complicated and interesting physical phenomena. According to this point of view, this PhD thesis constitutes the seed for more important applications, and we hope this work will be used as a basis for future further developments, by us or by our colleagues.

At present, many of the bottlenecks of the previous implementation have been removed, and the *GW* code can be applied to more challenging and interesting systems. Symmetry properties, including anti-ferromagnetic symmetries, are extensively employed in each phase of the calculation, thus reducing the computing effort and the memory consumption. Four different plasmon-pole models are available to facilitate the comparison with results reported in the literature. Two different numerical integration schemes can be used to analyze the frequency dependence of the self-energy. The Hilbert transform method for the irreducible polarizability can be efficiently used to study the frequency dependence of $\hat{\Sigma}$ as well as optical properties, allowing for an important computational savings. This renders the ABINIT code a powerful tool for the study of optical properties of materials, and represents a first step towards the implementation of TDDFT-based approaches for the analysis of absorption spectra in isolated and, with the due precautions, extended systems.

The new implementation takes advantage of a scalable coarse grained MPI parallelism which, besides, permits a significant reduction of the memory requirements. This opens the way towards the analysis of challenging large-scale problems. In particular, the analysis of the optical properties of amorphous silicon dioxide we were interested in at the beginning of the project is now feasible, and is presently underway in our group. This challenging *ab initio* study can benefit from the analysis we have performed for the *GW* energies in the crystalline hexagonal phase of α -SiO₂. Our calculations have revealed that the optical gaps as well as the absolute positions of the QP energies of α -SiO₂, are quite sensitive to the particular plasmon-pole approximation employed. We have been able to reproduce previously published results using the plasmon-pole model of Hybertsen and Louie, but we have also found that the HL results differ from those obtained by numerically evaluating the self-energy with two different and independent algorithms. A careful comparison of the four different plasmon-pole models described in this work has revealed that the original model of Godby and Needs provides accurate results both for absolute and relative QP energies of α -SiO₂. On the one hand, this finding calls for a better understanding of the approximations involved in the different approaches and might be the subject of a forthcoming paper. On the other hand, the proved reliability of the GN model constitutes an important starting point for our analysis of the *GW* energies of amorphous SiO₂ in presence of defects since the GN model is expected to give accurate QP levels close to the gap, eliminating the need for an exceedingly costly numerical integration of $\hat{\Sigma}(\omega)$. Besides the GN PPM represents a reliable tool for the calculation of band-offsets in related materials.

We have also developed other algorithmic improvements in order to facilitate the calculation of quasiparticle properties. A general technique for the interpolation matrix elements in $\mathbf{k}$ space which uses maximally localized Wannier functions has been described and implemented. The interpolating algorithms have been compiled into a flexible library that can be used for various applications even outside the realm of *GW* calculations, for instance it might be used to interpolate electron-phonon matrix elements and calculate superconducting properties which are known to converge slowly with respect to the sampling of the Fermi surface. Among several other developments not described in this

manuscript, we wish to mention the implementation of a cut-off for the Coulomb interaction, whose main utility is to speed up the convergence of the QP energies in nanosystems (clusters and wires), or surfaces simulated through the supercell technique. The technical details of the implementation have not been reported in this work in order to spare the reader the umpteenth equation, but the cut-off technique is presently used by our colleagues in a conjoint project to calculate the quasiparticle band structures of Si and Ge nanowires.

Unfortunately, having fast and efficient algorithms is sometimes not enough to describe the correct physics. Many studies we are presently performing, either in our group or within external collaborations, deal with d electrons or atoms with shallow cores where standard calculations based on the norm-conserving approach would be cumbersome and plagued by several difficult to control approximations. This is the reason why we decided to extend the implementation to take advantage of the improvements introduced by the PAW formalism. Compared to the norm-conserving case, the PAW method allows us to get results closer to all-electron calculations, solving many issues related to the pseudization of the wave functions in the region close to atomic cores, while leading to significant CPU savings as the plane wave energy cut-off used are generally smaller. Both G_0W_0 and the $QPSCGW$ scheme of Faleev have been implemented and the correctness of the implementation has been validated by comparing our quasiparticle results and optical spectra for prototype compounds against recent studies reported in the scientific literature. All these improvements are presently used by us or by colleagues in our group to perform GW calculations in a variety of systems. It is important to stress that all the algorithms and the numerical techniques described in this thesis are available both in the case of norm-conserving pseudopotentials and the PAW method. This allows us to achieve a coherent implementation which enables a detailed comparison of the different approximations involved in GW calculations not yet performed until now.

Last but not least, we achieved the difficult task of setting a bridge between our developments and the work done by the other ABINIT developers. A promising link with other codes belonging to the ETSF network has also been established. Hopefully, the effort spent in improving the modularity and the integration of the GW part with the main code will pay off in terms of an improved interoperability as well as ease of development. In the mid-term, this will facilitate the extension of the code so that we can benefit from other techniques and approaches developed in the ABINIT community. Among the several possible applications, we wish to emphasize a planned interplay between GW and dynamical mean field theory, and a possible extension of the PAW implementation to include excitonic effects, either via the Bethe-Salpeter approach or TDDFT. The inclusion of non-collinear magnetism with a self-consistent treatment of the spin-orbit interaction at the many-body level is presently under development, and a preliminary implementation will be available soon.

In the long-term, it would be interesting to take advantage of the linear scaling of modern approaches to DFT based on the use of Daubechies wavelets. This kind of basis set represents a very promising and appealing method for the study of optical properties in large-scale systems, and a code based on the use of this basis set is already interfaced

with ABINIT.

To best of our knowledge, the *GW* code developed in this work represents the third implementation of the many-body formalism within the projector augmented wave method, the first to be freely and publicly released under GPL license. The ABINIT package is freely accessible and routinely used by about one thousand users spread all around the world, and it is encouraging to see that the number of colleagues employing its many-body part is progressively increasing day by day, as revealed by the increasing number of mails received on the forum. This motivates us to keep on working to improve the code making it a little bit faster and more scalable.

Appendix A

Fourier Transform

This appendix reports the conventions used in this work for the Fourier representation in frequency- and momentum-space. The volume of the unit cell is denoted with Ω , while V is the total volume of the crystal simulated employing Born-von K rman periodic boundary condition [98].

A.1 Conventions

$$f(\omega) = \int f(t) e^{i\omega t} dt, \quad f(t) = \frac{1}{2\pi} \int f(\omega) e^{-i\omega t} d\omega \quad (\text{A.1})$$

$$f(\mathbf{k}) = \int f(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d\mathbf{r}, \quad f(\mathbf{r}) = \frac{1}{(2\pi)^3} \int f(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} d\mathbf{k} \quad (\text{A.2})$$

A.2 Properties

$$\mathcal{F}[f(t)g(t)](\omega) = \frac{1}{2\pi} \int f(\omega - \omega') g(\omega') d\omega' = (f * g)(\omega) = (g * f)(\omega) \quad (\text{A.3})$$

$$\mathcal{F}[f(t)g(-t)](\omega) = \frac{1}{2\pi} \int f(\omega + \omega') g(\omega') d\omega' \quad (\text{A.4})$$

A.3 Fourier series for periodic lattice quantities

$$u(\mathbf{r}) = \sum_{\mathbf{G}} u(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}, \quad u(\mathbf{G}) = \frac{1}{\Omega} \int_{\Omega} u(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} d\mathbf{r} \quad (\text{A.5})$$

$$f(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{V} \sum_{\mathbf{G}_1, \mathbf{G}_2}^{\mathbf{q}} e^{i(\mathbf{q}+\mathbf{G}_1)\cdot\mathbf{r}_1} f_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) e^{-i(\mathbf{q}+\mathbf{G}_2)\cdot\mathbf{r}_2} \quad (\text{A.6})$$

$$f_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}) = \frac{1}{V} \iint_V e^{-i(\mathbf{q}+\mathbf{G}_1)\cdot\mathbf{r}_1} f(\mathbf{r}_1, \mathbf{r}_2) e^{i(\mathbf{q}+\mathbf{G}_2)\cdot\mathbf{r}_2} d\mathbf{r}_1 d\mathbf{r}_2 \quad (\text{A.7})$$

Appendix B

Symmetry Properties

B.1 Effect of space group symmetries on electron energies and wavefunctions

A crystal structure is fully specified by the three primitive vectors $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ of the translation group and by the basis vectors $\mu_1, \mu_2, \dots, \mu_j$ which determine the position of the atoms in the primitive unit cell. All the symmetry operations which leave the crystal unchanged constitute a space group. Besides the translation symmetry operations, a space group contains proper or improper rotations followed by an appropriate fractional displacement. There are 230 in total [93]. A generic element of the space group will be denoted in the following with $\hat{\mathcal{R}}_{\mathbf{t}}$ where $\mathcal{R}$ is a real orthonormal matrix associated to a proper or improper rotation while $\mathbf{t}$ is the corresponding fractional translation. If all the fractional translations are zero the space group is said to be *symmorphic*. The application of a symmetry operation $\hat{\mathcal{R}}_{\mathbf{t}}$ to a vector $\mathbf{A}_i$ defining the position of an atom in the unit cell gives:

$$\hat{\mathcal{R}}_{\mathbf{t}} \mathbf{A}_i \equiv \mathcal{R}^{-1}(\mathbf{A}_i - \mathbf{t}) = \mathbf{A}_j + \mathbf{R}, \quad (\text{B.1})$$

where $\mathbf{A}_j$ indicates an atom in the same unit cell of the same type as $\mathbf{A}_i$ which may be coincident with $\mathbf{A}_i$, and $\mathbf{R}$ is a suitable lattice translation (possibly zero). The application of the symmetry operation $\hat{\mathcal{R}}_{\mathbf{t}}$ on a generic function $F(\mathbf{r})$ of the three spatial coordinates is conventionally defined by:

$$\hat{\mathcal{R}}_{\mathbf{t}} F(\mathbf{r}) \equiv F(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})). \quad (\text{B.2})$$

Since $\hat{\mathcal{R}}_{\mathbf{t}}$ commutes with the Hamiltonian $\hat{H}$ of the crystal, it readily follows that, given $\Psi_{n\mathbf{k}}(\mathbf{r})$ eigenstate of $\hat{H}$, also $\hat{\mathcal{R}}_{\mathbf{t}} \Psi_{n\mathbf{k}}(\mathbf{r})$ is eigenstate of the Schrödinger problem with the same eigenvalue:

$$\begin{cases} [\hat{\mathcal{R}}_{\mathbf{t}}, \hat{H}] = 0 \\ \hat{H} \Psi_{n\mathbf{k}}(\mathbf{r}) = \epsilon_{n\mathbf{k}} \Psi_{n\mathbf{k}}(\mathbf{r}) \end{cases} \implies \hat{H} \hat{\mathcal{R}}_{\mathbf{t}} \Psi_{n\mathbf{k}} = \hat{\mathcal{R}}_{\mathbf{t}} \hat{H} \Psi_{n\mathbf{k}} = \epsilon_{n\mathbf{k}} \hat{\mathcal{R}}_{\mathbf{t}} \Psi_{n\mathbf{k}}. \quad (\text{B.3})$$

Although $\hat{\mathcal{R}}_{\mathbf{t}} \Psi_{n\mathbf{k}}(\mathbf{r})$ has the same eigenenergy as $\Psi_{n\mathbf{k}}(\mathbf{r})$, its crystalline momentum is different. The operation $\hat{\mathcal{R}}_{\mathbf{t}}$ of the space group transforms a Bloch function with vector $\mathbf{k}$ into a new Bloch wave function of crystalline momentum $\mathcal{R}\mathbf{k}$. This important property can be seen as follows:

$$\begin{aligned} \left[\hat{\mathcal{R}}_{\mathbf{t}} \Psi_{n\mathbf{k}} \right] (\mathbf{r} + \mathbf{R}) &= \Psi_{n\mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r} + \mathbf{R} - \mathbf{t})) = e^{i\mathbf{k} \cdot \mathcal{R}^{-1}(\mathbf{r} + \mathbf{R} - \mathbf{t})} u_{n\mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = \\ &= e^{i\mathbf{t} \cdot \mathcal{R}^{-1}\mathbf{k} \cdot \mathbf{r}} \Psi_{n\mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = e^{i\mathcal{R}\mathbf{k} \cdot \mathbf{r}} \hat{\mathcal{R}}_{\mathbf{t}} \Psi_{n\mathbf{k}}(\mathbf{r}), \end{aligned} \quad (\text{B.4})$$

where $\mathbf{R}$ is an arbitrary vector of the direct Bravais lattice and the invariance of the periodic part of the Bloch wave function has been exploited.¹ For a nondegenerate state, one obtains:

$$\Psi_{\mathcal{R}\mathbf{k}}(\mathbf{r}) = \Psi_{\mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})). \quad (\text{B.5})$$

The above equation can be used to reduce the number of $\mathbf{k}$ -points and matrix elements that have to be evaluated and stored since the information in the full Brillouin zone can be reconstructed by symmetry from an appropriate irreducible wedge. The set of equations below summarizes the most useful relationships commonly employed:

$$\begin{cases} \epsilon_{\mathcal{R}\mathbf{k}} &= \epsilon_{\mathbf{k}} \\ u_{\mathcal{R}\mathbf{k}}(\mathbf{r}) &= e^{-i\mathcal{R}\mathbf{k} \cdot \mathbf{t}} u_{\mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) \\ u_{\mathcal{R}\mathbf{k}}(\mathbf{G}) &= e^{-i(\mathcal{R}\mathbf{k} + \mathbf{G}) \cdot \mathbf{t}} u_{\mathbf{k}}(\mathcal{R}^{-1}\mathbf{G}). \end{cases} \quad (\text{B.6})$$

The time invariance of the Hamiltonian, $\hat{H}^\dagger = \hat{H}$, might introduce additional constraints:

$$\begin{cases} \epsilon_{n\mathbf{k}} &= \epsilon_{-n\mathbf{k}}, \\ u_{n\mathbf{k}}(\mathbf{r}) &= u_{n-\mathbf{k}}^\dagger(\mathbf{r}), \\ u_{n\mathbf{k}}(\mathbf{G}) &= u_{n-\mathbf{k}}^\dagger(-\mathbf{G}). \end{cases} \quad (\text{B.7})$$

It is important to stress that the set of equations in B.6 hold only in case of nondegenerate states. In the presence of degeneracy, the application of a symmetry operation on a Bloch wave function with momentum $\mathbf{k}$ belonging to the set of degenerate states $\mathcal{C}_{n\mathbf{k}}$, produces a new eigenstate with same energy and crystalline momentum $\mathcal{R}\mathbf{k}$. The new eigenstate is given by an appropriate linear combination of the degenerate states with wave vector $\mathcal{R}\mathbf{k}$. More explicitly:

$$\hat{\mathcal{R}}_{\mathbf{t}} \Psi_{n\mathbf{k}}(\mathbf{r}) = \Psi_{n\mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = \sum_{\alpha \in \mathcal{C}_{n\mathbf{k}}} D_{\alpha n}(\mathcal{R}) \Psi_{\alpha\mathcal{R}\mathbf{k}}(\mathbf{r}), \quad (\text{B.8})$$

¹ The last equality in B.4 follows from the orthogonality of the $\mathcal{R}$ matrix when referred to in a Cartesian frame of reference. The orthogonality of the rotation matrix does not hold anymore if, as usually done in the ABINIT code, the symmetry operations are expressed in reduced coordinates. In this case, the correct matrix to use for operations in reciprocal space is given by the transpose of $\mathcal{R}^{-1}$.

where $D_{\alpha\beta}(\mathcal{R})$ is the unitary transformation associated with $\mathcal{R}$ [93]. Equation B.8 reduces to B.5 if the state $\Psi_{n\mathbf{k}}(\mathbf{r})$ is nondegenerate since, in this particular case, $D_{\alpha\beta}(\mathcal{R}) = \delta_{\alpha\beta}$.

B.2 Magnetic Space Groups

An extension of the concept of symmetry is needed in order to explain the magnetic properties of crystals [160]. In addition to the spatial arrangement of atoms, the orientation of the atomic magnetic moment also becomes important in such systems. It may turn out that the usual spatial operation, while leaving the crystal unchanged in regard to its geometrical structure, will reverse the orientation of spins. Let the symbol $\mathcal{J}$ indicate the operation of reversing all spins, and let $\mathcal{E}$ denote the identity operator. A combined operation consisting of an ordinary symmetry operation followed by $\mathcal{J}$ is a new type of symmetry operation called a complementary operation. The rules for operator composition in the standard nonmagnetic group can be easily extended to include $\mathcal{J}$:

$$\text{If } \hat{\mathcal{A}}_1 \cdot \hat{\mathcal{A}}_2 = \hat{\mathcal{A}}_3 \text{ then } \begin{cases} (\mathcal{J}\hat{\mathcal{A}}_1) \cdot (\mathcal{J}\hat{\mathcal{A}}_2) = \hat{\mathcal{A}}_3 \\ (\mathcal{J}\hat{\mathcal{A}}_1) \cdot \hat{\mathcal{A}}_2 = \mathcal{J}\hat{\mathcal{A}}_3 \\ \hat{\mathcal{A}}_1 \cdot (\mathcal{J}\hat{\mathcal{A}}_2) = \mathcal{J}\hat{\mathcal{A}}_3. \end{cases} \quad (\text{B.9})$$

Magnetic groups are obtained by replacing some of the symmetry elements of the initial non-magnetic space group by their complementary operations so that the resulting ensemble forms a group with respect to the new algebra. Magnetic space groups are sometimes referred to as Shubnikov groups, and can be classified according to three different categories [161, 162]:

1. Groups which include $\mathcal{J}$ explicitly, so-called grey groups or Shubnikov type IV (230 in number). Each group can be obtained by taking the direct product of each of the conventional space groups with the group $(\mathcal{E}, \mathcal{J})$
2. Groups which do not include $\mathcal{J}$ either explicitly or in conjunction with a conventional symmetry operation, also named colorless groups. There are 230 and they are indistinguishable from the conventional space groups described in section B.1
3. Groups which do not include $\mathcal{J}$ explicitly, but which contain complementary symmetry operations. Also called mixed groups or Shubnikov type III. There are 1191 in total.

Henceforth, in order to keep the notation as simple as possible, a symmetry operation containing $\mathcal{J}$ will be denoted with an underlined symbol $\hat{\underline{\mathcal{R}}}_{\mathbf{t}}$, and will be said to have anti-ferromagnetic character. On the contrary, symmetry operations which preserve the orientation of the spin projection will be denoted using the standard notation $\hat{\mathcal{R}}_{\mathbf{t}}$, and

will be said to have ferromagnetic character. The action of a magnetic symmetry on a nondegenerate Bloch state has to be generalized according to:

$$\hat{\mathcal{R}}_{\mathbf{t}} \Psi_{\mathbf{k}}^{\sigma}(\mathbf{r}) \equiv \Psi_{\mathbf{k}}^{\sigma}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = \Psi_{\mathcal{R}\mathbf{k}}^{\sigma}(\mathbf{r}), \quad (\text{B.10})$$

$$\hat{\mathcal{R}}_{\mathbf{t}} \Psi_{\mathbf{k}}^{\sigma}(\mathbf{r}) \equiv \Psi_{\mathbf{k}}^{-\sigma}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = \Psi_{\mathcal{R}\mathbf{k}}^{-\sigma}(\mathbf{r}), \quad (\text{B.11})$$

where $\sigma = \pm \frac{1}{2}$. Using the above symmetry relations, one can verify that the two spin components of the electron density transform according to:

$$\hat{\mathcal{R}}_{\mathbf{t}} n^{\sigma}(\mathbf{r}) = n^{\sigma}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = n^{\sigma}(\mathbf{r}), \quad (\text{B.12})$$

$$\hat{\mathcal{R}}_{\mathbf{t}} n^{\sigma}(\mathbf{r}) = n^{\sigma}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) = n^{-\sigma}(\mathbf{r}), \quad (\text{B.13})$$

from which it follows that the total charge and, consequently, the Hartree potential are invariant under the application of any operation of the magnetic group. Similar symmetry relationships hold for the two components of the exchange-correlation potential.

B.3 Symmetries of the oscillator matrix elements

This section is a compendium of the symmetry properties of the oscillator matrix elements. These properties are used to reduce the integration over the full Brillouin zone to an appropriate irreducible wedge.

B.3.1 Symmetrization in k-space

Let us suppose that a particular symmetry operation $\hat{\mathcal{R}}_{\mathbf{t}}$ preserves $\mathbf{q}$ (*i.e.* $\mathcal{R}\mathbf{q} = \mathbf{q}$). Using the set of symmetry relationships given in B.6, one obtains:

$$\begin{aligned} M_{\mathbf{G}}^{b_1 b_2}(\mathcal{R}\mathbf{k}, \mathbf{q}) &= \int \Psi_{b_1 \mathcal{R}\mathbf{k} - \mathbf{q}}^{\dagger}(\mathbf{r}) e^{-i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} \Psi_{b_2 \mathcal{R}\mathbf{k}}(\mathbf{r}) d\mathbf{r} = \\ &= \int \Psi_{b_1 \mathcal{R}(\mathbf{k} - \mathbf{q})}^{\dagger}(\mathbf{r}) e^{-i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} \Psi_{b_2 \mathcal{R}\mathbf{k}}(\mathbf{r}) d\mathbf{r} = \\ &= \int \Psi_{b_1 \mathbf{k} - \mathbf{q}}^{\dagger}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) e^{-i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} \Psi_{b_2 \mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r} - \mathbf{t})) d\mathbf{r} = \\ &= e^{-i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{t}} \int \Psi_{b_1 \mathbf{k} - \mathbf{q}}^{\dagger}(\mathbf{r}) e^{-i\mathcal{R}^{-1}(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} \Psi_{b_2 \mathbf{k}}(\mathbf{r}) d\mathbf{r} = e^{-i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{t}} M_{\mathcal{R}^{-1}\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}). \quad (\text{B.14}) \end{aligned}$$

The generalization to the case in which time-reversal, or a non-null *umklapp* vector is needed in order to bring the $\mathbf{q}$ -point back onto itself, is obtained in a very similar way. For completeness, the final results are summarized below:

$$M_{\mathbf{G}}^{b_1 b_2}(\mathcal{R}\mathbf{k}, \mathbf{q}) = e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} M_{\mathcal{R}^{-1}\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \quad \text{if } \mathcal{R}\mathbf{q} = \mathbf{q} \quad (\text{B.15})$$

$$M_{\mathbf{G}}^{b_1 b_2}(\mathcal{R}\mathbf{k}, \mathbf{q}) = e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} M_{\mathcal{R}^{-1}(\mathbf{G}-\mathbf{G}_{\mathcal{R}})}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \quad \text{if } \mathcal{R}\mathbf{q} = \mathbf{q} + \mathbf{G}_{\mathcal{R}} \quad (\text{B.16})$$

$$M_{\mathbf{G}}^{b_1 b_2}(-\mathcal{R}\mathbf{k}, \mathbf{q}) = e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} \left[M_{-\mathcal{R}^{-1}\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \right]^{\dagger} \quad \text{if } -\mathcal{R}\mathbf{q} = \mathbf{q} \quad (\text{B.17})$$

$$M_{\mathbf{G}}^{b_1 b_2}(-\mathcal{R}\mathbf{k}, \mathbf{q}) = e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} \left[M_{-\mathcal{R}^{-1}(\mathbf{G}-\mathbf{G}_{\mathcal{R}})}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \right]^{\dagger} \quad \text{if } -\mathcal{R}\mathbf{q} = \mathbf{q} + \mathbf{G}_{\mathcal{R}}. \quad (\text{B.18})$$

B.3.2 Symmetrization in $\mathbf{q}$ -space

For the calculation of the self-energy, one needs to know how the oscillator matrix elements transform in $\mathbf{q}$ -space. In this case, only the symmetry operations which preserve $\mathbf{k}$ can be exploited. Let suppose that a particular operation belongs to the little group of $\mathbf{k}$, *i.e.* $\mathcal{R}\mathbf{k} = \mathbf{k}$. Using equation B.6, one arrives at:

$$\begin{aligned} M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathcal{R}\mathbf{q}) &= \int \Psi_{b_1\mathbf{k}-\mathcal{R}\mathbf{q}}^{\dagger}(\mathbf{r}) e^{-i(\mathcal{R}\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} \Psi_{b_2\mathbf{k}}(\mathbf{r}) d\mathbf{r} = \\ &= \int \Psi_{b_1\mathcal{R}(\mathbf{k}-\mathbf{q})}^{\dagger}(\mathbf{r}) e^{-i(\mathcal{R}\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} \Psi_{b_2\mathcal{R}\mathbf{k}}(\mathbf{r}) d\mathbf{r} = \\ &= \int \Psi_{b_1\mathbf{k}-\mathbf{q}}^{\dagger}(\mathcal{R}^{-1}(\mathbf{r}-\mathbf{t})) e^{-i(\mathcal{R}\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} \Psi_{b_2\mathbf{k}}(\mathcal{R}^{-1}(\mathbf{r}-\mathbf{t})) d\mathbf{r} = \\ &= e^{-i(\mathcal{R}\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} \int \Psi_{b_1\mathbf{k}-\mathbf{q}}^{\dagger}(\mathbf{r}) e^{-i(\mathbf{q}+\mathcal{R}^{-1}\mathbf{G})\cdot\mathbf{r}} \Psi_{b_2\mathbf{k}}(\mathbf{r}) d\mathbf{r} = e^{-i(\mathcal{R}\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} M_{\mathcal{R}^{-1}\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \end{aligned} \quad (\text{B.19})$$

The generalization to the case in which time-reversal, or a non-null *umklapp* vector is needed in order to bring the $\mathbf{k}$ -point back onto itself, is obtained in a very similar way. The final results are summarized below:

$$M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathcal{R}\mathbf{q}) = e^{-i(\mathcal{R}\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} M_{\mathcal{R}^{-1}\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \quad \text{if } \mathcal{R}\mathbf{k} = \mathbf{k} \quad (\text{B.20})$$

$$M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathcal{R}\mathbf{q}) = e^{-i(\mathcal{R}\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} M_{\mathcal{R}^{-1}\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \quad \text{if } \mathcal{R}\mathbf{k} = \mathbf{k} + \mathbf{G}_{\mathcal{R}} \quad (\text{B.21})$$

$$M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, -\mathcal{R}\mathbf{q}) = e^{-i(-\mathcal{R}\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} \left[M_{-\mathcal{R}^{-1}\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \right]^{\dagger} \quad \text{if } -\mathcal{R}\mathbf{k} = \mathbf{k} \quad (\text{B.22})$$

$$M_{\mathbf{G}}^{b_1 b_2}(\mathbf{k}, -\mathcal{R}\mathbf{q}) = e^{-i(-\mathcal{R}\mathbf{q}+\mathbf{G})\cdot\mathbf{t}} \left[M_{-\mathcal{R}^{-1}\mathbf{G}}^{b_1 b_2}(\mathbf{k}, \mathbf{q}) \right]^{\dagger} \quad \text{if } -\mathcal{R}\mathbf{k} = \mathbf{k} + \mathbf{G}_{\mathcal{R}}. \quad (\text{B.23})$$

Note that, contrary to the symmetrization in $\mathbf{k}$ space described in the previous section, here the presence of a non null *umklapp* vector does not introduce any additional complication.

Appendix C

Optical properties within the pseudopotential formalism

C.1 Norm-conserving Pseudopotentials in the Kleyman Bylander form

This appendix reports basic definitions and equations needed to evaluate the matrix elements of the commutator between the position operator, $\mathbf{r}$, and the nonlocal part of the Hamiltonian in the particular case in which norm-conserving pseudopotentials are used. Providing a consistent introduction to the theory of Norm-Conserving (NC) pseudopotentials is beyond the purposes of this section, a more complete discussion of the topic can be found in specialized articles available in the literature [63, 163, 124, 66, 71, 164, 125]. For our purposes, we can summarize by saying that a pseudopotential is constructed in order to replace the atomic all-electron potential such that core states are eliminated and valence electrons are described by nodeless pseudo wavefunctions whose representation in the Fourier space decays rapidly. Modern norm-conserving pseudopotentials are constructed following the procedure described in [63, 66, 71] such that the scattering properties of the all-electron atom are reproduced around the reference energy configuration, up to first order in energy. In modern *ab initio* codes, the fully separable form proposed by Kleyman and Bylander (KB) is usually employed as it drastically reduces the number of operations required for the application of the nonlocal part of the Hamiltonian as well as the memory required to store the operator [124]. In the KB form, the interaction between a valence electron and an ion of type α is described by means of the operator:¹

$$\hat{v}_{\text{PS}}^{\alpha} = \hat{v}_{\text{loc}}^{\alpha}(\mathbf{r}) + \sum_{lm} |\chi_{lm}^{\alpha}\rangle E_l^{\alpha} \langle \chi_{lm}^{\alpha}|, \quad (\text{C.1})$$

¹For simplicity, the discussion is limited to pseudopotentials with a single projector per angular channel.

where $\hat{v}_{\text{loc}}^\alpha(\mathbf{r})$ is a purely local potential with a Coulomb tail γ/r . The so-called Kleinman-Bylander energies, E_l^α , measure the strength of the nonlocal component with respect to the local part. The projectors $\chi_{lm}^\alpha(\mathbf{r})$ are short-ranged functions, expressed in terms of a complex spherical harmonic $Y_m^l(\theta, \phi)$ multiplied by an l - and atom-dependent radial function

$$\chi_{lm}^\alpha(\mathbf{r}) = \frac{1}{r} \chi_l^\alpha(r) Y_m^l(\theta, \phi). \quad (\text{C.2})$$

The total nonlocal part of the Hamiltonian is obtained by summing the different atom-centered contributions. The final expression reads:

$$V_{\text{nl}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{\substack{\mathbf{R} \\ \alpha \mathbf{t}_\alpha}} \sum_{lm} \langle \mathbf{r}_1 - \mathbf{R} - \mathbf{t}_\alpha | \chi_{lm}^\alpha \rangle E_l^\alpha \langle \chi_{lm}^\alpha | \mathbf{r}_2 - \mathbf{R} - \mathbf{t}_\alpha \rangle, \quad (\text{C.3})$$

where $\mathbf{R}$ runs over the sites of the Bravais lattice, and $\mathbf{t}_\alpha$ over the positions of the atoms with the same type α located inside the unit cell. Due to the nonlocality of the operator, its Fourier transform depends on two separate indices, $\mathbf{k} + \mathbf{G}_1$ and $\mathbf{k} + \mathbf{G}_2$, instead of the simple difference $\mathbf{G}_1 - \mathbf{G}_2$. The shorthand notation $\mathbf{K} = \mathbf{k} + \mathbf{G}$ is used in the following to simplify the derivation. Using the Rayleigh expansion of planewaves in terms of spherical harmonics $Y_m^l(\theta, \phi)$ and spherical Bessel functions $j_l(Kr)$:

$$e^{i\mathbf{K} \cdot \mathbf{r}} = 4\pi \sum_{lm} i^l j_l(Kr) Y_m^l(\hat{\mathbf{K}})^\dagger Y_m^l(\hat{\mathbf{r}}), \quad (\text{C.4})$$

the Fourier representation of the projector defined by equation C.2 can be expressed as:

$$\langle \mathbf{K} | \chi_{lm}^\alpha \rangle = \frac{4\pi}{\Omega^{1/2}} (-1)^l Y_m^l(\hat{\mathbf{K}}) \int_0^\infty r j_l(Kr) \chi_l^\alpha(r) r \, d = \frac{4\pi}{\Omega^{1/2}} (-1)^l Y_m^l(\hat{\mathbf{K}}) F_l^\alpha(K). \quad (\text{C.5})$$

Finally, the expression for the total nonlocal operator in reciprocal space is:

$$V_{\text{nl}}(\mathbf{K}, \mathbf{K}') = \frac{(4\pi)^2}{\Omega} \sum_{\alpha \mathbf{t}_\alpha} \sum_{lm} e^{-i(\mathbf{K}-\mathbf{K}') \cdot \mathbf{t}_\alpha} Y_m^l(\hat{\mathbf{K}}) Y_m^l(\hat{\mathbf{K}}')^\dagger E_l^\alpha F_l^\alpha(K) F_l^\alpha(K'). \quad (\text{C.6})$$

C.2 Treatment of the long-wavelength limit

The treatment of the Coulomb singularity in equation 5.5 deserves some clarification, especially if dielectric properties are calculated within the pseudopotential formalism due to the presence of a fully nonlocal term in the Kohn-Sham Hamiltonian. As shown in [80], in the long-wavelength limit, *i.e.* for $\mathbf{q} \rightarrow 0$, the so-called head ($\mathbf{G}_1 = \mathbf{G}_2 = 0$) and wings ($\mathbf{G}_1 = 0$ or $\mathbf{G}_2 = 0$) of the inverse dielectric matrix are non-analytical functions of $\mathbf{q}$. In a non-metallic system, both the head and wings of the irreducible polarizability tend to 0 for $\mathbf{q} \rightarrow 0$ thus leading to an indeterminate form $\frac{0}{0}$ that has to be properly

analyzed. A linear expansion up to the first order in q of the oscillator matrix elements results in:

$$\langle \Psi_{\mathbf{k}-\mathbf{q}b_1} | e^{-i\mathbf{q}\cdot\mathbf{r}} | \Psi_{\mathbf{k}b_2} \rangle \underset{\mathbf{q} \rightarrow 0}{\approx} -i\mathbf{q} \cdot \langle \Psi_{\mathbf{k}b_1} | \mathbf{r} | \Psi_{\mathbf{k}b_2} \rangle + \mathcal{O}(q^2), \quad (\text{C.7})$$

where the difference between the two wave functions $|\Psi_{\mathbf{k}-\mathbf{q}}\rangle$ and $|\Psi_{\mathbf{k}}\rangle$ has been neglected because it only introduces terms that are $\mathcal{O}(q^2)$. Unfortunately, the above expression cannot be directly used because the matrix elements of the position operator $\mathbf{r}$ are ill-defined when wavefunctions obey Born-von Kármán periodic boundary conditions. Following the transversal gauge approach proposed in [81, 82], the integral involving the $\mathbf{r}$ operator can be rewritten as

$$\langle \Psi_{\mathbf{k}b_1} | \mathbf{r} | \Psi_{\mathbf{k}b_2} \rangle = \frac{\langle \Psi_{\mathbf{k}+\mathbf{q}b_1} | [\hat{H}, \mathbf{r}] | \Psi_{\mathbf{k}b_2} \rangle}{\epsilon_{\mathbf{k}b_1} - \epsilon_{\mathbf{k}b_2}}. \quad (\text{C.8})$$

The matrix elements of the commutator $[\hat{H}, \mathbf{r}]$ are well-defined even within Born-von Kármán periodic boundary conditions, and the long-wavelength limits can be evaluated analytically. In a full-potential approach, the only nonlocal operator in the Hamiltonian is the kinetic term (semilocal) whose commutator can be straightforwardly derived:

$$[\Delta, \mathbf{r}] = 2\nabla. \quad (\text{C.9})$$

The same approach presents some difficulties when pseudopotentials are used because the KS Hamiltonian:

$$\hat{H} = -\frac{1}{2}\Delta + V_H(\mathbf{r}) + V_{\text{loc}}(\mathbf{r}) + V_{\text{nl}}(\mathbf{r}, \mathbf{r}') + V_{\text{ext}}(\mathbf{r}), \quad (\text{C.10})$$

now contains the fully nonlocal term $V_{\text{nl}}(\mathbf{r}, \mathbf{r}')$ whose contribution should be properly taken into account. This leads to the more complicated expression:

$$\langle b_1, \mathbf{k} - \mathbf{q} | e^{-i(\mathbf{q}\cdot\mathbf{r})} | b_2, \mathbf{k} \rangle \underset{\mathbf{q} \rightarrow 0}{\approx} \frac{\langle b_1, \mathbf{k} | -i\mathbf{q} \cdot \nabla + i\mathbf{q} \cdot [V_{\text{nl}}, \mathbf{r}] | b_2, \mathbf{k} \rangle}{\epsilon_{b_2\mathbf{k}} - \epsilon_{b_1\mathbf{k}}}. \quad (\text{C.11})$$

The term involving the momentum operator can be efficiently evaluated in reciprocal space in $N_{\mathbf{G}}$ operations:

$$-i\mathbf{q} \cdot \langle b_1, \mathbf{k} | \nabla | b_2, \mathbf{k} \rangle = \mathbf{q} \cdot \sum_{\mathbf{G}} u_{b_1, \mathbf{k}}^\dagger(\mathbf{G}) [\mathbf{k} + \mathbf{G}] u_{b_2, \mathbf{k}}(\mathbf{G}). \quad (\text{C.12})$$

The matrix elements at the point $\mathcal{R}\mathbf{k}$ in the full Brillouin zone can be reconstructed from the symmetrical image $\mathbf{k}$ in the IBZ by means of the symmetry relationships give in B.6. The final result reads:

$$\begin{aligned}
-i\mathbf{q} \cdot \frac{\langle \mathcal{R}\mathbf{k} b_1 | \nabla | \mathcal{R}\mathbf{k} b_2 \rangle}{\epsilon_{\mathcal{R}\mathbf{k} b_2} - \epsilon_{\mathcal{R}\mathbf{k} b_1}} &= \frac{\mathbf{q}}{\epsilon_{\mathbf{k} b_2} - \epsilon_{\mathbf{k} b_1}} \cdot \sum_{\mathbf{G}} u_{\mathbf{k} b_1}^\dagger(\mathbf{G}) \mathcal{S} \mathbf{G} u_{\mathbf{k} b_2}(\mathbf{G}) = \\
&= \frac{\mathcal{R}^{-1} \mathbf{q}}{\epsilon_{\mathbf{k} b_2} - \epsilon_{\mathbf{k} b_1}} \cdot \sum_{\mathbf{G}} u_{\mathbf{k} b_1}^\dagger(\mathbf{G}) \mathbf{G} u_{\mathbf{k} b_2}(\mathbf{G}), \quad (\text{C.13})
\end{aligned}$$

where possible phase factors associated to non-symmorphic operations cancel out each other. The treatment of the nonlocal part of the Hamiltonian is more involved. Employing the identity:

$$\begin{aligned}
(\nabla_{\mathbf{K}} + \nabla_{\mathbf{K}'}) \left[\frac{1}{V} \int e^{-i\mathbf{K} \cdot \mathbf{r}} V_{\text{nl}}(\mathbf{r}, \mathbf{r}') e^{i\mathbf{K}' \cdot \mathbf{r}'} d\mathbf{r} d\mathbf{r}' \right] &= \\
&= \frac{i}{V} \int e^{-i\mathbf{K} \cdot \mathbf{r}} \left[\mathbf{r}' V_{\text{nl}}(\mathbf{r}, \mathbf{r}') - \mathbf{r} V_{\text{nl}}(\mathbf{r}, \mathbf{r}') \right] d\mathbf{r} d\mathbf{r}', \quad (\text{C.14})
\end{aligned}$$

the matrix element of the commutator $[V_{\text{nl}}, \mathbf{r}]$ between to plane wave basis functions can be rewritten as:

$$i\langle \mathbf{K} | [V_{\text{nl}}, \mathbf{r}] | \mathbf{K}' \rangle = (\nabla_{\mathbf{K}} + \nabla_{\mathbf{K}'}) V_{\text{nl}}(\mathbf{K}, \mathbf{K}'), \quad (\text{C.15})$$

where the notation $\mathbf{K} \equiv \mathbf{k} + \mathbf{G}$ has been introduced. Following the conventions for the pseudopotential operator introduced in section C.1, the matrix elements in reciprocal space of the nonlocal part of the Hamiltonian can finally be expressed as

$$V_{\text{nl}}(\mathbf{K}, \mathbf{K}') = \frac{(4\pi)^2}{\Omega} \sum_{\alpha \mathbf{t}_\alpha} \sum_{lm} e^{-i(\mathbf{K}-\mathbf{K}') \cdot \mathbf{t}_\alpha} Y_m^l(\hat{\mathbf{K}}) Y_m^{l\dagger}(\hat{\mathbf{K}}') E_{l\alpha}^{\text{KB}} F_{l\alpha}(K) F_{l\alpha}(K') \quad (\text{C.16})$$

where $\mathbf{t}_\alpha$ runs over the positions of the atoms with the same type α located inside the unit cell, and l and m are the angular and magnetic angular momentum associated to the pseudopotential projectors. The form factor $F_{l\alpha}(K)$ can be easily evaluated starting from the knowledge of the radial part of the pseudopotential projector.

In the approach proposed in previous works [82, 165], the sum over the magnetic angular momentum m in equation C.16 is eliminated and expressed in terms of Legendre polynomials, $\mathcal{P}_l(\mathbf{K} \cdot \mathbf{K}')$ by employing the addition theorem for complex spherical harmonics given in equation D.9. The price to pay is that the resulting matrix elements cannot be factorized anymore in terms of two functions which separately depend only on $\hat{\mathbf{K}}$ and $\hat{\mathbf{K}}'$. The same holds also for the derivatives given in equation C.14. As a consequence, both the computational effort required to apply the commutator in reciprocal space as well as the memory requirements scale quadratically with respect to the number of plane waves used to describe the wavefunctions. The correct inclusion of the commutator in the optical limit is therefore prohibitively expensive, especially in the case

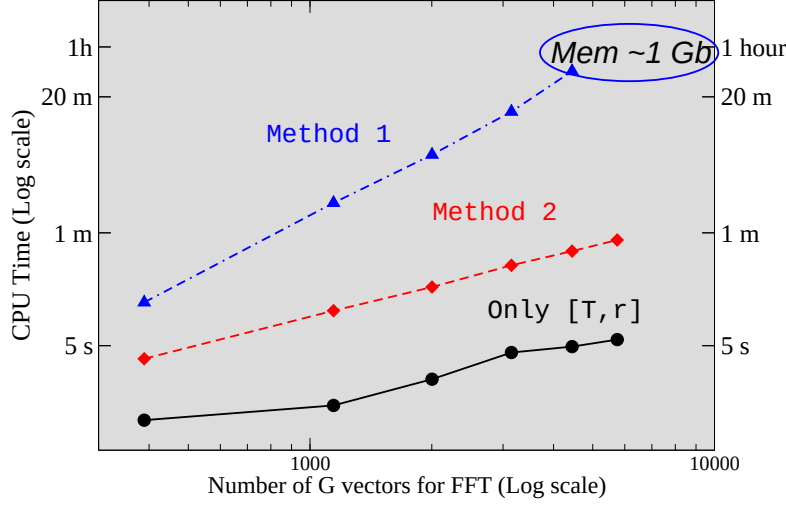


Figure C.1: Computing time required by the inclusion of the commutator $[\hat{H}, \hat{T} + V_{\text{nl}}]$ in $\alpha\text{-SiO}_2$ as function of the number of plane waves used to describe the periodic part of the wave functions. Three different approaches are compared. In the first one (black solid line) only the contribution given by the kinetic operator is included. Red dashed and blue dashed-dotted lines represent the CPU time required when the non-local term is properly included using equation (B8) in [82] or the present expression given in C.22, respectively. The method of [82] has the worst scaling and, more importantly, it is extremely memory demanding: for $N_\Psi \approx 5000$ the memory needed to store the matrix elements of the commutator is already of the order of 1 Gigabyte.

of hard pseudopotentials or systems with large unit cells, where the contribution given by the $\mathbf{q} \rightarrow 0$ limit is expected to be important.

A faster and less memory demanding algorithm can be implemented if the use of Legendre polynomials is completely avoided, and the dependence on $Y_m^l(\theta, \phi)$ in equation C.16 is maintained. A similar approach was originally implemented in the two codes DP [166] and EXC [167] and subsequently employed also in the GW space and time code gwst [114]. By differentiating equation C.14, one arrives at:

$$\begin{aligned}
(\nabla_{\mathbf{K}} + \nabla_{\mathbf{K}'}) V_{nl}(\mathbf{K}, \mathbf{K}') &= \frac{(4\pi)^2}{\Omega} \sum_{\alpha \mathbf{t}_\alpha} \sum_{lm} e^{-i(\mathbf{K}-\mathbf{K}') \cdot \mathbf{t}_\alpha} E_{l\alpha}^{\text{KB}} \times \\
&\times \left\{ \nabla_{\mathbf{K}} \left(F_{l\alpha}(K) Y_m^l(\hat{\mathbf{K}}) \right) F_{l\alpha}(K') Y_m^{l\dagger}(\hat{\mathbf{K}}') + F_{l\alpha}(K) Y_m^l(\hat{\mathbf{K}}) \nabla_{\mathbf{K}'} \left(F_{l\alpha}(K') Y_m^{l\dagger}(\hat{\mathbf{K}}') \right) \right\} = \\
&= \frac{(4\pi)^2}{\Omega} \sum_{\alpha \mathbf{t}_\alpha} \sum_{lm} e^{-i(\mathbf{K}-\mathbf{K}') \cdot \mathbf{t}_\alpha} E_{l\alpha}^{\text{KB}} \left\{ \left(F'_{l\alpha}(K) \hat{\mathbf{K}} Y_m^l(\hat{\mathbf{K}}) + F_{l\alpha}(K) \nabla_{\mathbf{K}} Y_m^l(\hat{\mathbf{K}}) \right) \times \right. \\
&\times F_{l\alpha}(K') Y_m^{l\dagger}(\hat{\mathbf{K}}') + F_{l\alpha}(K) Y_m^l(\hat{\mathbf{K}}') \left(F'_{l\alpha}(K') \hat{\mathbf{K}}' Y_m^{l\dagger}(\hat{\mathbf{K}}') + F_{l\alpha}(K') \nabla_{\mathbf{K}'} Y_m^{l\dagger}(\hat{\mathbf{K}}') \right) \Big\} = \\
&= \sum_{\alpha \mathbf{t}_\alpha} \sum_{lm} \vec{f}_{nld}(\mathbf{K}, \alpha, \mathbf{t}_\alpha, l, -m)^\dagger f_{nl}(\mathbf{K}', \alpha, l, -m) + f_{nl}(\mathbf{K}, \alpha, \mathbf{t}_\alpha, l, -m)^\dagger \vec{f}_{nld}(\mathbf{K}', \alpha, l, -m),
\end{aligned} \tag{C.17}$$

where the following shorthand notations have been introduced:

$$f_{nl}(\mathbf{K}, \alpha, \mathbf{t}_\alpha, l, m) \equiv \frac{4\pi}{\Omega^{\frac{1}{2}}} e^{i\mathbf{G} \cdot \mathbf{t}_\alpha} Y_m^l(\hat{\mathbf{K}}) F_{l\alpha}(K) |E_{l\alpha}^{\text{KB}}|^{\frac{1}{2}} \text{sign}(E_{l\alpha}^{\text{KB}}) \tag{C.18}$$

$$\vec{f}_{nld}(\mathbf{K}, \alpha, \mathbf{t}_\alpha, l, m) \equiv \frac{4\pi}{\Omega^{\frac{1}{2}}} e^{i\mathbf{G} \cdot \mathbf{t}_\alpha} \nabla_{\mathbf{K}} \left(Y_m^l(\hat{\mathbf{K}}) F_{l\alpha}(K) \right) |E_{l\alpha}^{\text{KB}}|^{\frac{1}{2}}. \tag{C.19}$$

Using the relations:

$$\begin{cases} \nabla_{\mathbf{K}} f(K) = f'(K) \hat{\mathbf{K}} \\ \nabla_{\mathbf{K}} f(\theta, \phi) = \left(\frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial K_x} + \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial K_x}, \dots, \frac{\partial f}{\partial \theta} \frac{\partial \theta}{\partial K_z} + \frac{\partial f}{\partial \phi} \frac{\partial \phi}{\partial K_z} \right) \end{cases} \tag{C.20}$$

one obtains:

$$\nabla_{\mathbf{K}} \left(F_{l\alpha}(K) Y_m^l(\hat{\mathbf{K}}) \right) = F'_{l\alpha}(K) \hat{\mathbf{K}} Y_m^l(\hat{\mathbf{K}}) + F_{l\alpha}(K) \nabla_{\mathbf{K}} Y_m^l(\hat{\mathbf{K}}). \tag{C.21}$$

The final result for the matrix elements of the commutator thus reads:

$$\begin{aligned}
i \mathbf{q} \cdot \langle b_1, \mathbf{k} | [V_{nl}, \mathbf{r}] | b_2, \mathbf{k} \rangle &= \langle b_1, \mathbf{k} | (\nabla_{\mathbf{K}} + \nabla_{\mathbf{K}'}) V_{nl}(\mathbf{K}, \mathbf{K}') | b_2, \mathbf{k} \rangle = \\
&= \frac{\mathbf{q}}{\Omega} \cdot \sum_{\alpha, l, m} \left\{ \sum_{\mathbf{G}} u_{b_1, \mathbf{k}}^\dagger(\mathbf{G}) \vec{f}_{nld}^\dagger(\mathbf{k} + \mathbf{G}, \alpha, \mathbf{t}_\alpha, l, m) \sum_{\mathbf{G}} u_{b_2, \mathbf{k}}(\mathbf{G}) f_{nl}(\mathbf{k} + \mathbf{G}, \alpha, \mathbf{t}_\alpha, l, m) + \right. \\
&\quad \left. \sum_{\mathbf{G}} u_{b_1, \mathbf{k}}^\dagger(\mathbf{G}) f_{nl}^\dagger(\mathbf{k} + \mathbf{G}, \alpha, \mathbf{t}_\alpha, l, m) \sum_{\mathbf{G}} u_{b_2, \mathbf{k}}(\mathbf{G}) \vec{f}_{nld}(\mathbf{k} + \mathbf{G}, \alpha, \mathbf{t}_\alpha, l, m) \right\}. \tag{C.22}
\end{aligned}$$

The advantage of the present formulation is that only the relatively small vectors $f_{nl}(\mathbf{K}, \alpha, \mathbf{t}_\alpha, l, m)$ and $\vec{f}_{nld}(\mathbf{K}, \alpha, \mathbf{t}_\alpha, l, m)$ have to be stored in memory. Furthermore, the evaluation

of Eq. C.22 requires only fast vector-vector multiplications instead of slow matrix-vector operations. Figure C.1 displays the results of a benchmark calculation in which the algorithm described in this section and the approach proposed in [82] have been used to calculate the optical properties of α -SiO₂.

Appendix D

Complex and Real Spherical Harmonics

D.1 Complex Spherical Harmonics

Complex spherical harmonics, Y_m^l , are defined as the eigenfunctions of the orbital angular momentum operators, $\hat{L}^2$ and $\hat{L}_z$. They can be labeled by two quantum numbers, l and m , related to the eigenvalues of $\hat{L}^2$ and $\hat{L}_z$:

$$\hat{L}^2 Y_m^l(\theta, \phi) = l(l+1) Y_m^l(\theta, \phi), \quad (\text{D.1})$$

and

$$\hat{L}_z Y_m^l(\theta, \phi) = m Y_m^l(\theta, \phi). \quad (\text{D.2})$$

Their explicit expression is:

$$Y_m^l(\theta, \phi) = (-1)^m N_{lm} P_m^l(\cos \theta) e^{im\phi}, \quad (\text{D.3})$$

where l takes non-negative integer values and the possible values for m are integers from $-l$ to l . The two symbols θ and ϕ are used to denote angular spherical coordinates. N_{lm} is the normalization factor:

$$N_{lm} = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} \quad (\text{D.4})$$

with P_m^l being the associated Legendre polynomials. Complex spherical harmonics are orthogonal functions on the sphere, *i.e.* :

$$\langle Y_m^l | Y_{m'}^{l'} \rangle_\Omega = \delta_{ll'} \delta_{mm'}, \quad (\text{D.5})$$

where $\langle . | . \rangle_\Omega$ denotes the integration over the unit sphere. Furthermore, they form a complete basis set since any square-integrable function $f(\theta, \phi)$ can be expanded in series according to:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l a_{lm} Y_m^l(\theta, \phi). \quad (\text{D.6})$$

Complex spherical harmonics satisfy two useful properties commonly employed in practical applications. The parity properties:

$$Y_m^{l\dagger}(\hat{\mathbf{r}}) = (-1)^m Y_{-m}^l(\hat{\mathbf{r}}), \quad (\text{D.7})$$

$$Y_m^l(-\hat{\mathbf{r}}) = (-1)^l Y_m^l(\hat{\mathbf{r}}), \quad (\text{D.8})$$

and the addition theorem:

$$\sum_m Y_m^l(\hat{\mathbf{r}}) Y_m^{l\dagger}(\hat{\mathbf{r}}') = \frac{2l+1}{4\pi} \mathcal{P}_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}'), \quad (\text{D.9})$$

where $\mathcal{P}_l(q)$ are Legendre polynomials. Equation D.9 is generally employed in the application of the nonlocal part of the Hamiltonian in the case of norm-conserving pseudopotentials.

D.2 Real spherical harmonics

Real Spherical Harmonic (RSH) are obtained by combining complex conjugate functions associated to opposite values of m . RSH are the most adequate basis functions for calculations in which atomic symmetry is important since they can be directly related to the irreducible representations of the subgroups of D_3 [168]. Moreover, being real, they have half the memory requirement of complex spherical harmonics. This is clearly an advantage if high angular momenta are needed or several RHS values have to be stored in memory. A possible definition for RHS is [168]:

$$\mathcal{Y}_m^l \equiv \begin{cases} \frac{(-1)^m}{\sqrt{2}} (Y_m^l + Y_m^{l\dagger}) & m > 0 \\ Y_0^l & m = 0 \\ \frac{(-1)^m}{i\sqrt{2}} (Y_{|m|}^l - Y_{|m|}^{l\dagger}) & m < 0. \end{cases} \quad (\text{D.10})$$

Equation D.10 can be rewritten in matrix notation as

$$\underline{\mathcal{Y}}^l = U^l \underline{Y}^l, \quad (\text{D.11})$$

where $\underline{y}^l$ and $\underline{Y}^l$ are column vectors containing real and spherical harmonics ordered by increasing m values. U^l is a unitary matrix whose explicit expression can be found in [168].

The basic properties of RSH can be easily derived from the properties of complex spherical harmonics by means of equation D.10.

Reality:

$$y_m^l(\hat{\mathbf{r}}) = y_m^l(\hat{\mathbf{r}}), \quad (\text{D.12})$$

parity property:

$$y_m^l(-\hat{\mathbf{r}}) = (-1)^l y_m^l(\hat{\mathbf{r}}), \quad (\text{D.13})$$

and orthonormality:

$$\langle y_m^l | y_{m'}^{l'} \rangle_\Omega = \delta_{ll'} \delta_{mm'}. \quad (\text{D.14})$$

Certain on-site matrix elements include the integration of three real spherical harmonics, which can be evaluated exactly:¹

$$\mathcal{G}_{l_i m_i l_j m_j}^{LM} \equiv \int y_{m_i}^{l_i} y_M^L y_{m_j}^{l_j} d\Omega. \quad (\text{D.15})$$

The above integral is known in the literature under the name of Gaunt coefficient [142]. Due to selection rules, the integral in equation D.15 is nonzero only if $m_i + m_j = M$ and $|l_i + l_j| \geq (L, L + 2, L + 4, \dots) \geq |l_i - l_j|$.

D.3 Symmetry transformation of RSH

The use of symmetry properties is of fundamental importance to simplify the determination of electronic structure as it greatly reduces the number of non-equivalent integrals that have to be computed, as well as the size of any tables stored in memory. Frequently one has to evaluate the value of a RSH at a rotated point after the application of a symmetry operation belonging to the space group of the crystal. This usually occurs when $\mathbf{k}$ -dependent matrix elements have to be symmetrized in the full Brillouin zone starting from the knowledge of their symmetrical image in the irreducible wedge. The effect of a proper or improper rotation on a RSH can be deduced from:

$$\hat{\mathcal{R}} y_m^l(\hat{\mathbf{r}}) = y_m^l(\mathcal{R}^{-1}\hat{\mathbf{r}}) = \sum_{\alpha} D_{\alpha m}^l(\mathcal{R}) y_m^l(\hat{\mathbf{r}}). \quad (\text{D.16})$$

¹ The expression reported for $\mathcal{G}$ follows the internal conventions used inside the ABINIT code and slightly differs from the one reported in standard text-books.

That is, spherical harmonics of given l are transformed into a linear combination of RHS of same l where the coefficients are given by:

$$D_{\alpha m}^l(\mathcal{R}) \equiv \langle y_\alpha^l | \hat{\mathcal{R}} | y_m^l \rangle. \quad (\text{D.17})$$

D.4 Useful expansions in terms of RSH

By means of the unitary transformation given in D.10, it is possible to rewrite the Rayleigh expansion of a plane wave in terms of RSH as

$$e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} = 4\pi \sum_{lm} i^l j_l(|\mathbf{k}+\mathbf{G}|r) y_m^l(\widehat{\mathbf{k}+\mathbf{G}}) y_m^l(\hat{\mathbf{r}}) \quad (\text{D.18})$$

with $j_l(x)$ the spherical Bessel function of order l . In a similar way, the electrostatic potentials can be expanded in a real spherical harmonics basis set. The final expression is very similar to the one obtained in the case of complex spherical harmonics:

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_{l=0}^{\infty} \frac{4\pi}{2l+1} \sum_{m=-l}^{m=l} \frac{r_{<}^l}{r_{>}^{l+1}} y_m^l(\hat{\mathbf{r}}) y_m^l(\hat{\mathbf{r}}'), \quad (\text{D.19})$$

with $r_{<} \equiv \min(r, r')$ and $r_{>} \equiv \max(r, r')$.

Appendix E

Optical properties within PAW

This appendix contains the equations required to evaluate optical properties in the long-wavelength limit within the PAW approach when the transversal gauge is used. Section E.1 treats the case of a standard Kohn-Sham Hamiltonian with a local potential. The changes needed for the explicitly nonlocal LDA+U Hamiltonian are discussed in section E.2.

E.1 Matrix elements of the gradient operator between PAW partial waves

As we have seen in Appendix C, the linear expansion of the oscillator matrix elements for $\mathbf{q} \rightarrow 0$ requires the evaluation of the commutator of the Hamiltonian with the position operator:

$$\begin{aligned} M_{\mathbf{G}=0}^{vc}(\mathbf{k}, \mathbf{q}) &= \langle \Psi_{v\mathbf{k}-\mathbf{q}} | e^{-i\mathbf{q} \cdot \mathbf{r}} | \Psi_{c\mathbf{k}} \rangle = \frac{\langle \Psi_{v\mathbf{k}-\mathbf{q}} | [\hat{H}, e^{-i\mathbf{q} \cdot \mathbf{r}}] | \Psi_{c\mathbf{k}} \rangle}{\epsilon_{v\mathbf{k}-\mathbf{q}} - \epsilon_{c\mathbf{k}}} = \\ &= -i\mathbf{q} \cdot \frac{\langle \Psi_{v\mathbf{k}} | [\hat{H}, \mathbf{r}] | \Psi_{c\mathbf{k}} \rangle}{\epsilon_{v\mathbf{k}} - \epsilon_{c\mathbf{k}}} + \mathcal{O}(q^2). \quad (\text{E.1}) \end{aligned}$$

Since PAW, unlike the pseudopotential theory, is a full-potential formalism in which only the pseudized version of the Hamiltonian contains a nonlocal term, the commutator in equation E.1 reduces to the contribution due to the semi-local kinetic term, *i.e.* $-\nabla$. The matrix element of the semi-local ∇ operator can be expressed in terms of pseudized wavefunctions and onsite terms according to:

$$\langle \Psi_{v\mathbf{k}} | \nabla | \Psi_{c\mathbf{k}} \rangle = \langle \tilde{\Psi}_{v\mathbf{k}} | \nabla | \tilde{\Psi}_{c\mathbf{k}} \rangle + \sum_{ij} \langle \tilde{\Psi}_{v\mathbf{k}} | \tilde{p}_i \rangle \langle \tilde{p}_j | \tilde{\Psi}_{c\mathbf{k}} \rangle \left[\langle \phi_i | \nabla | \phi_j \rangle - \langle \tilde{\phi}_i | \nabla | \tilde{\phi}_j \rangle \right]. \quad (\text{E.2})$$

Since the partial waves are given in terms of radial functions multiplied by real spherical harmonics, the one-center matrix elements in equation E.2 are best evaluated by

expressing the gradient operator in spherical coordinates:

$$\nabla = \hat{\mathbf{r}} \frac{\partial}{\partial r} + \frac{\hat{\phi}}{r} \frac{\partial}{\partial \phi} + \frac{\hat{\theta}}{r \sin \theta} \frac{\partial}{\partial \theta} \quad (\text{E.3})$$

with

$$\begin{cases} \hat{\mathbf{r}} = (\cos \theta \sin \phi, \sin \theta \sin \phi, \cos \phi), \\ \hat{\theta} = (-\sin \theta, \cos \theta, 0), \\ \hat{\phi} = (\cos \theta \cos \phi, \sin \theta \cos \phi, -\sin \phi). \end{cases} \quad (\text{E.4})$$

By using Eq. E.3 and Eq. E.4, the following expressions for the Cartesian components of the one-center matrix are derived:

$$\begin{aligned} \langle \phi_i | \nabla | \phi_j \rangle_x = & \int r \phi_{n_i l_i} \frac{\partial(\phi_{n_j l_j} / r)}{\partial r} dr \int \cos \theta \sin \phi y_{m_i}^{l_i} y_{m_j}^{l_j} d\Omega - \\ & - \int \frac{\phi_{n_i l_i} \phi_{n_j l_j}}{r} dr \int \frac{\sin \theta}{\sin \phi} y_{m_i}^{l_i} \frac{\partial y_{m_j}^{l_j}}{\partial \theta} d\Omega + \\ & + \int \frac{\phi_{n_i l_i} \phi_{n_j l_j}}{r} dr \int \cos \theta \cos \phi y_{m_i}^{l_i} \frac{\partial y_{m_j}^{l_j}}{\partial \phi} d\Omega, \quad (\text{E.5}) \end{aligned}$$

$$\begin{aligned} \langle \phi_i | \nabla | \phi_j \rangle_y = & \int r \phi_{n_i l_i} \frac{\partial(\phi_{n_j l_j} / r)}{\partial r} dr \int \sin \theta \sin \phi y_{m_i}^{l_i} y_{m_j}^{l_j} d\Omega + \\ & + \int \frac{\phi_{n_i l_i} \phi_{n_j l_j}}{r} dr \int \frac{\cos \theta}{\sin \phi} y_{m_i}^{l_i} \frac{\partial y_{m_j}^{l_j}}{\partial \theta} d\Omega + \\ & + \int \frac{\phi_{n_i l_i} \phi_{n_j l_j}}{r} dr \int \sin \theta \cos \phi y_{m_i}^{l_i} \frac{\partial y_{m_j}^{l_j}}{\partial \phi} d\Omega, \quad (\text{E.6}) \end{aligned}$$

$$\begin{aligned} \langle \phi_i | \nabla | \phi_j \rangle_z = & \int r \phi_{n_i l_i} \frac{\partial(\phi_{n_j l_j} / r)}{\partial r} dr \int \cos \phi y_{m_i}^{l_i} y_{m_j}^{l_j} d\Omega - \\ & - \int \frac{\phi_{n_i l_i} \phi_{n_j l_j}}{r} dr \int \sin \phi y_{m_i}^{l_i} \frac{\partial y_{m_j}^{l_j}}{\partial \phi} d\Omega. \quad (\text{E.7}) \end{aligned}$$

All the above integrals are tabulated at the beginning of the calculation, and then used to evaluate the matrix elements of the gradient operator (Eq. E.2) once the projections $\langle \tilde{p}_i | \Psi_{n\mathbf{k}} \rangle$ are known.

E.2 Optical limit within LDA+U

In the LDA+U formalism, the single-electron Hamiltonian is explicitly non-local, even in the all-electron representation, due to the presence of the additional U term. Starting from the definition of the LDA+U Hamiltonian the commutator can be written as:

$$[\hat{H}^{\text{LDA}} + \hat{H}^U, e^{-i\mathbf{q}\cdot\mathbf{r}}] = i\mathbf{q} \cdot \nabla - i\mathbf{q} \cdot \sum_{m_1 m_2} V_{m_1 m_2}^\sigma [\mathcal{P}_{m_1 m_2}, \mathbf{r}] + \mathcal{O}(q^2). \quad (\text{E.8})$$

The first term in the above equation has already been described in the previous section. Only the second contribution requires clarification. By using the property $(\mathcal{P}_{m_1 m_2})^\dagger = \mathcal{P}_{m_2 m_1}$, the contribution of the U term of the Hamiltonian to the matrix elements can be written as:

$$\begin{aligned} \lim_{q \rightarrow 0} M_U^{vc}(\mathbf{k}, \mathbf{q}) &= \frac{-i\mathbf{q}}{\epsilon_{v\mathbf{k}-\mathbf{q}} - \epsilon_{c\mathbf{k}}} \sum_{ij} \langle \tilde{\Psi}_{v\mathbf{k}}^\sigma | \tilde{p}_i \rangle \langle \tilde{p}_j | \tilde{\Psi}_{c\mathbf{k}}^\sigma \rangle \\ &\quad \sum_{m_1 m_2} V_{m_1, m_2}^\sigma \underbrace{\left(\langle \mathcal{P}_{m_2 m_1} \phi_i | \mathbf{r} | \phi_j \rangle - \langle \phi_i | \mathbf{r} | \mathcal{P}_{m_1 m_2} \phi_j \rangle \right)}_{\delta_{l_i l_U} \delta_{m_1 m_i} \langle \phi_{n_i l_i m_2} | \mathbf{r} | \phi_j \rangle - \delta_{l_j l_U} \delta_{m_2 m_j} \langle \phi_i | \mathbf{r} | \phi_{n_j l_j m_1} \rangle}, \quad (\text{E.9}) \end{aligned}$$

where the difference between the two wave functions $|\tilde{\Psi}_{v\mathbf{k}-\mathbf{q}}\rangle$ and $|\tilde{\Psi}_{v\mathbf{k}}\rangle$ has been neglected because it only introduces terms that are $\mathcal{O}(q^2)$. The matrix elements of the position operator between two PAW partial waves are easily obtained by expressing the position operation $\mathbf{r}$ in terms of real spherical harmonics:

$$\mathbf{r} = 2\sqrt{\frac{\pi}{3}} r \mathcal{Y}_m^1, \quad m = 1, 2, 3 \quad (\text{E.10})$$

from which one arrives at:

$$\langle \phi_i | \mathbf{r} | \phi_j \rangle = \mathbf{R}_0 \langle \phi_i | \phi_j \rangle + 2\sqrt{\frac{\pi}{3}} \langle \phi_i | r \mathcal{Y}_m^1 | \phi_j \rangle, \quad (\text{E.11})$$

where $\mathbf{R}_0$ denotes the position of the PAW augmentation sphere within the unit cell.

Appendix F

List of Publications

1. Electrons and phonons in the ternary alloy $CaAl_{2-x}Si_x$ as a function of composition. M. Giantomassi, L. Boeri, and G.B. Bachelet Physical Review B, **72**, 224512, (2005)
2. Unresolved problems in superconductivity of CaC_6 . I.I. Mazin, L. Boeri, O.V. Dolgov, A.A. Golubov, G.B. Bachelet, M. Giantomassi, and O.K. Andersen. Physica C, **460**, 116-120 (2007)
3. Electron-phonon interaction in graphite intercalation compounds. L. Boeri, G.B. Bachelet, M. Giantomassi, and O.K. Andersen Physical Review B, **76**, 064510, (2007)
4. Pressure effects on the superconducting transition in $nH-CaAlSi$. L. Boeri, J. S. Kim, M. Giantomassi, F. S. Razavi, S. Kuroiwa, J. Akimitsu, and R. K. Kremer Physical Review B, **77**, 144502, (2008)
5. ABINIT: First-principles approach to material and nanosystem properties. X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Cote, T. Deutsch, L. Genovese, Ph. Ghosez, M. Giantomassi, S. Goedecker, D. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M. Oliveira, T. Rangel, Y. Pouillon, G.-M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M. Verstraete, G. Zerah, and J. Zwanziger. Submitted to Computer Physics Communication.
6. Implementation of optical matrix elements, including a real-space cutoff, in ABINIT C. Motta, M. Giantomassi, M. Cazzaniga, K. Gaál-Nagy, and X. Gonze In preparation. To be submitted to Computer Physics Communication.
7. Reliability and accuracy of the plasmon-pole approximation for GW calculations. R. Shaltaf, M. Giantomassi, G.-M. Rignanese and X. Gonze. In preparation. To be submitted to Physical Review B.

8. Full quasi-particle band structures of Si and Ge nanowires. H. Peelaers, B. Partoens, F. Peeters, M. Giantomassi, T. Rangel, E. Goosens, G.-M. Rignanese, X. Gonze In preparation. To be submitted to Physical Review B.
9. Implementation of the GW approximation within the projector augmented-wave method in the ABINIT code. Application to the study of many-body effects in NiO. M. Giantomassi et al. In preparation. To be submitted to Computational Material Science.

Bibliography

- [1] E. Runge and E.K.U. Gross. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.*, 52:997–1000, 1984.
- [2] G. Onida, L. Reining, and A. Rubio. Electronic excitations: density-functional versus many-body green’s-function approaches. *Rev. Mod. Phys.*, 74:601–659, 2002.
- [3] A.A. Abrikosov, L.P. Gorkov, and E. Dzyaloshinskii. *Methods of quantum field theory in statistical physics*. Dover, New York, 1975.
- [4] A.L. Fetter and J.D. Walecka. *Quantum Theory of Many-Particle Systems*. McGraw-Hill, New York, 1971.
- [5] L. Hedin. New method for calculating the one-particle Green’s function with application to the electron-gas problem. *Phys. Rev.*, 139:A796–A823, 1965.
- [6] L. Hedin. On correlation effects in electron spectroscopies and the GW approximation. *J. Phys. Condens. Matter*, 11:R489–R528, 1999.
- [7] E.E. Salpeter and H.A. Bethe. A relativistic equation for bound-state problems. *Phys. Rev.*, 84:1232–1242, 1951.
- [8] S. Albrecht, L. Reining, R. Del Sole, and G. Onida. Ab initio calculation of excitonic effects in the optical spectra of semiconductors. *Phys. Rev. Lett.*, 80:4510–4513, 1998.
- [9] T. Kotani and M. van Schilfgaarde. All-electron GW approximation with the mixed basis expansion based on the full-potential LMTO method. *Solid State Communications*, 121:461 – 465, 2002.
- [10] W. Ku and A.G. Eguiluz. Band-gap problem in semiconductors revisited: Effects of core states and many-body self-consistency. *Phys. Rev. Lett.*, 89:126401, 2002.
- [11] T. Kotani, M. van Schilfgaarde, and S.V. Faleev. Quasiparticle self-consistent GW method: A basis for the independent-particle approximation. *Phys. Rev. B*, 76:165106, 2007.

- [12] R. Gómez-Abal, X. Li, M. Scheffler, and C. Ambrosch-Draxl. Influence of the core-valence interaction and of the pseudopotential approximation on the electron self-energy in semiconductors. *Phys. Rev. Lett.*, 101:106404, 2008.
- [13] M.L. Tiago, S. Ismail-Beigi, and S.G. Louie. Effect of semicore orbitals on the electronic band gaps of Si, Ge, and GaAs within the GW approximation. *Phys. Rev. B*, 69:125212, 2004.
- [14] B. Arnaud and M. Alouani. All-electron projector-augmented-wave GW approximation: Application to the electronic properties of semiconductors. *Phys. Rev. B*, 62:4464–4476, 2000.
- [15] P.E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, 1994.
- [16] S.V. Faleev, M. van Schilfgaarde, and T. Kotani. All-electron self-consistent GW approximation: Application to Si, MnO, and NiO. *Phys. Rev. Lett.*, 93:126406, 2004.
- [17] M. Nicola and D. Vanderbilt. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B*, 56:12847–12865, 1997.
- [18] M. Born and Oppenheimer J.M. *Ann. Physik*, 84:457, 1927.
- [19] L.H. Thomas. *Proc. Cambridge. Philos. Soc.*, 23:542, 1927.
- [20] E. Fermi. *Atti Accad. Naz. Lincei, Cl. Sci. Mat. Nat. Rend.*, 6:602, 1927.
- [21] D.R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part I. theory and methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24:89–110, 1928.
- [22] V. Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik A Hadrons and Nuclei*, 61:126–148, 1930.
- [23] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 1964.
- [24] M. Levy. *Proc. Natl. Acad. Sci. USA*, 76:6062, 1979.
- [25] M. Levy. Electron densities in search of hamiltonians. *Phys. Rev. A*, 26:1200–1208, 1982.
- [26] W. Kohn and L.J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965.
- [27] J.P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996.

- [28] W. Kohn. Nobel lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.*, 71:1253–1266, 1999.
- [29] X. Gonze. First-principles responses of solids to atomic displacements and homogeneous electric fields: Implementation of a conjugate-gradient algorithm. *Phys. Rev. B*, 55:10337–10354, 1997.
- [30] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73:515–562, 2001.
- [31] W.G. Aulbur, L. Jonsson, and J.K. Wilkins. *Quasiparticle calculations in solids*. Academic press inc., 525 B Street, suite 1900, San Diego, CA 92101-4495 USA, 2000.
- [32] M. Levy, J.P. Perdew, and V. Sahni. Exact differential equation for the density and ionization energy of a many-particle system. *Phys. Rev. A*, 30:2745–2748, 1984.
- [33] J.P. Perdew and M. Levy. Physical content of the exact Kohn-Sham orbital energies: Band gaps and derivative discontinuities. *Phys. Rev. Lett.*, 51:1884–1887, 1983.
- [34] L. J. Sham and M. Schlüter. Density-functional theory of the energy gap. *Phys. Rev. Lett.*, 51:1888–1891, 1983.
- [35] L. J. Sham and M. Schlüter. Density-functional theory of the band gap. *Phys. Rev. B*, 32:3883–3889, 1985.
- [36] G.D Mahan. *Many Particle Physics*. Plenum, New York, 1990.
- [37] R.D. Mattuck. *A guide to Feynman diagrams in the many-body problem*. Dover, New York, 1992.
- [38] V.M. Galitskii and A.B. Migdal. *Sov. Phys. JETP*, 7:96, 1958.
- [39] G. Lehmann and M. Taut. On the numerical calculation of the density of states and related properties. *Phys. Stat. Sol. b*, 54:469, 1972.
- [40] G. Onida, L. Reining, R.W. Godby, R. Del Sole, and W. Andreoni. Ab initio calculations of the quasiparticle and absorption spectra of clusters: The sodium tetramer. *Phys. Rev. Lett.*, 75:818–821, 1995.
- [41] F. Bruneval. *Exchange and correlation in the electronic structure of solids, from silicon to cuprous oxide: GW approximation and beyond*. PhD thesis, Ecole Polytechnique, 2005.
- [42] S.L. Adler. Quantum theory of the dielectric constant in real solids. *Phys. Rev.*, 126:413–420, 1962.

- [43] N. Wiser. Dielectric constant with local field effects included. *Phys. Rev.*, 129:62–69, 1963.
- [44] P.C. Martin and J. Schwinger. Theory of many-particle systems. i. *Phys. Rev.*, 115:1342–1373, 1959.
- [45] G. Baym and L.P. Kadanoff. Conservation laws and correlation functions. *Phys. Rev.*, 124:287–299, 1961.
- [46] G. Baym. Self-consistent approximations in many-body systems. *Phys. Rev.*, 127:1391–1401, 1962.
- [47] B. Holm and U. von Barth. Fully self-consistent GW self-energy of the electron gas. *Phys. Rev. B*, 57:2108–2117, 1998.
- [48] B. Holm and F. Aryasetiawan. Total energy from the Galitskii-Migdal formula using realistic spectral functions. *Phys. Rev. B*, 62:4858–4865, 2000.
- [49] P. García-González and R.W. Godby. Self-consistent calculation of total energies of the electron gas using many-body perturbation theory. *Phys. Rev. B*, 63:075112, 2001.
- [50] E.L. Shirley. Self-consistent GW and higher-order calculations of electron states in metals. *Phys. Rev. B*, 54:7758–7764, 1996.
- [51] W.-D. Schöne and A.G. Eguiluz. Self-consistent calculations of quasiparticle states in metals and semiconductors. *Phys. Rev. Lett.*, 81:1662–1665, 1998.
- [52] P. Sánchez-Friera and R.W. Godby. Efficient total energy calculations from self-energy models. *Phys. Rev. Lett.*, 85:5611–5614, 2000.
- [53] P. García-González and R.W. Godby. Many-body GW calculations of ground-state properties: Quasi-2d electron systems and van der Waals forces. *Phys. Rev. Lett.*, 88:056406, 2002.
- [54] F. Aryasetiawan, T. Miyake, and K. Terakura. Total energy method from many-body formulation. *Phys. Rev. Lett.*, 88:166401, 2002.
- [55] M.S. Hybertsen and S.G. Louie. First-principles theory of quasiparticles: Calculation of band gaps in semiconductors and insulators. *Phys. Rev. Lett.*, 55:1418–1421, 1985.
- [56] M.S. Hybertsen and S.G. Louie. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Phys. Rev. B*, 34:5390–5413, 1986.
- [57] J.-L. Li, G.-M. Rignanese, and S.G. Louie. Quasiparticle energy bands of NiO in the GW approximation. *Phys. Rev. B*, 71:193102, 2005.

- [58] F. Bruneval, N. Vast, and L. Reining. Effect of self-consistency on quasiparticles in solids. *Phys. Rev. B*, 74:045102, 2006.
- [59] M. van Schilfgaarde, T. Kotani, and S. Faleev. Quasiparticle self-consistent GW theory. *Phys. Rev. Lett.*, 96:226402, 2006.
- [60] U. von Barth and C.D. Gelatt. Validity of the frozen-core approximation and pseudopotential theory for cohesive energy calculations. *Phys. Rev. B*, 21:2222–2228, 1980.
- [61] L Hedin and S. Lundqvist. *Solid State Physics vol. 23*. Academic, New York, 1969.
- [62] J.C. Phillips. Generalized Koopmans’ theorem. *Phys. Rev.*, 123:420–424, 1961.
- [63] D.R. Hamann, M. Schlüter, and C. Chiang. Norm-conserving pseudopotentials. *Phys. Rev. Lett.*, 43:1494–1497, 1979.
- [64] S. Lebegue, B. Arnaud, M. Alouani, and P. E. Blochl. Implementation of an all-electron GW approximation based on the projector augmented wave method without plasmon pole approximation: Application to Si, SiC, AlAs, InAs, NaH, and KH. *Phys. Rev. B*, 67:155208, 2003.
- [65] K. Delaney, P. García-González, A. Rubio, P. Rinke, and R.W. Godby. Comment on band-gap problem in semiconductors revisited: Effects of core states and many-body self-consistency. *Phys. Rev. Lett.*, 93:249701, 2004.
- [66] D.R. Hamann. Generalized norm-conserving pseudopotentials. *Phys. Rev. B*, 40:2980–2987, 1989.
- [67] P.E. Blöchl. Generalized separable potentials for electronic-structure calculations. *Phys. Rev. B*, 41:5414–5416, 1990.
- [68] A. Marini, G. Onida, and R. Del Sole. Quasiparticle electronic structure of copper in the GW approximation. *Phys. Rev. Lett.*, 88:016403, 2001.
- [69] E. Luppi, H.-C. Weissker, S. Bottaro, F. Sottile, V. Veniard, L. Reining, and G. Onida. Accuracy of the pseudopotential approximation in ab initio theoretical spectroscopies. *Phys. Rev. B*, 78:245124, 2008.
- [70] D. Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41:7892–7895, 1990.
- [71] N. Troullier and J.L. Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43:1993–2006, 1991.
- [72] F. Mauri, B.G. Pfrommer, and S.G. Louie. Ab initio theory of NMR chemical shifts in solids and liquids. *Phys. Rev. Lett.*, 77:5300–5303, 1996.

- [73] C.J. Pickard and F. Mauri. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys. Rev. B*, 63:245101, 2001.
- [74] M. Taillefumier, D. Cabaret, A.-M. Flank, and F. Mauri. X-ray absorption near-edge structure calculations with the pseudopotentials: Application to the k edge in diamond and α -quartz. *Phys. Rev. B*, 66:195107, 2002.
- [75] E.L. Shirley and R.M. Martin. GW quasiparticle calculations in atoms. *Phys. Rev. B*, 47:15404–15412, 1993.
- [76] E.L. Shirley and R.M. Martin. Many-body core-valence partitioning. *Phys. Rev. B*, 47:15413–15427, 1993.
- [77] E.L. Shirley, X. Zhu, and S.G. Louie. Core polarization in solids: Formulation and application to semiconductors. *Phys. Rev. B*, 56:6648–6661, 1997.
- [78] X. Gonze, J. M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G. M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, M. Torrent, A. Roy, M. Mikami, Ph. Ghosez, J. Y. Raty, and D. C. Allan. First-principles computation of material properties: the abinit software project. *Computational Materials Science*, 25:478 – 492, 2002.
- [79] X. Gonze, G.-M. Rignanese, M. Verstraete, J.-M. Beuken, Y. Pouillon, R. Caracas, F. Jollet, M. Torrent, G. Zerah, M. Mikami, Ph. Ghosez, M. Veithen, J.-Y. Raty, V. Olevano, F. Bruneval, L. Reining, R. Godby, G. Onida, D.R. Hamann, and D.C. Allan. A brief introduction to the abinit software package. *Zeit. Kristallogr.*, 220:558–562, 2005.
- [80] R.M. Pick, M.H. Cohen, and R.M. Martin. Microscopic theory of force constants in the adiabatic approximation. *Phys. Rev. B*, 1:910–920, 1970.
- [81] S. Baroni and R. Resta. Ab initio calculation of the macroscopic dielectric constant in silicon. *Phys. Rev. B*, 33:7017, 1986.
- [82] M.S. Hybertsen and S.G. Louie. Ab initio static dielectric matrices from the density-functional approach. *Phys. Rev. B*, 35:5585, 1986.
- [83] A. Baldereschi and E. Tosatti. Mean-value point and dielectric properties of semiconductors and insulators. *Phys. Rev. B*, 17:4710–4717, 1978.
- [84] G.E. Engel, B. Farid, C.M.M. Nex, and N.H. March. Calculation of the GW self-energy in semiconducting crystals. *Phys. Rev. B*, 44:13356–13373, 1991.
- [85] E.K.U Gross, J. Dobson, and M. Petersilka. *Density-Functional Theory of time-dependent phenomena in Density-Functional Theory II*. Springer, Berlin, 1996.
- [86] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in Fortran 77 The art of scientific Computing. Second Edition*. Cambridge University Press, 1986.

- [87] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and Sorensen D. *LAPACK Users' Guide, Third Edition*. SIAM, 3600 University City Science Center Philadelphia, PA 19104-2688, 1999.
- [88] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley. *ScaLAPACK Users' Guide*. SIAM, 3600 University City Science Center, Philadelphia, PA 19104-2688., 1997.
- [89] M.C Payne, M.P Teter, D.C Allan, T.A Arias, and J.D Joannopolos. Iterative minimization techniques for ab initio total energy calculations: molecular dynamics and conjugate-gradients. *Rev. Mod. Phys.*, 64:1045–1097, 1992.
- [90] M.M. Rieger, L. Steinbeck, I.D White, H.N. Rojas, and R.W Godby. The GW space-time method for the self-energy of large systems. *Computer Physics Communications*, 117:211 – 228, 1999.
- [91] L. Steinbeck, A. Rubio, L. Reining, M. Torrent, I.D. White, and R.W. Godby. Enhancements to the GW space-time method. *Computer Physics Communications*, 125:105–118, 2000.
- [92] X. Blase, A. Rubio, S.G. Louie, and M.L. Cohen. Mixed-space formalism for the dielectric response in periodic systems. *Phys. Rev. B*, 52:R2225–R2228, 1995.
- [93] F. Bassani and G. Pastori Parravicini. *Electronic and optical transitions in solids*. Pergamon Press, 1975.
- [94] R.W. Godby and R.J. Needs. Metal-insulator transition in Kohn-Sham theory and quasiparticle theory. *Phys. Rev. Lett.*, 62:1169–1172, 1989.
- [95] W. von der Linden and P. Horsch. Precise quasiparticle energies and Hartree-Fock bands of semiconductors and insulators. *Phys. Rev. B*, 37:8351–8362, 1988.
- [96] G.E. Engel and B. Farid. Generalized plasmon-pole model and plasmon band structures of crystals. *Phys. Rev. B*, 47:15931–15934, 1993.
- [97] C. Kittel. *Introduction to Solid State Physics, 6th ed.* Wiley, New York, 1986.
- [98] N.W. Ashcroft and N.D. Mermin. *Solid State Physics*. Cornell University, 1976.
- [99] D.L. Johnson. Local field effects and the dielectric response matrix of insulators: A model. *Phys. Rev. B*, 9:4475–4484, 1974.
- [100] M. Taut. Frequency moments of the dielectric function for an inhomogeneous electron gas. *J. Phys. C: Solid State Phys*, 18:2677–2690, 1985.

- [101] S.B. Zhang, D. Tománek, M.L. Cohen, S.G. Louie, and M.S. Hybertsen. Evaluation of quasiparticle energies for semiconductors without inversion symmetry. *Phys. Rev. B*, 40:3162–3168, 1989.
- [102] R. Hott. GW-approximation energies and Hartree-Fock bands of semiconductors. *Phys. Rev. B*, 44:1057–1065, 1991.
- [103] T. Miyake and F. Aryasetiawan. Efficient algorithm for calculating noninteracting frequency-dependent linear response functions. *Phys. Rev. B*, 61:7172–7175, 2000.
- [104] M. Shishkin and G. Kresse. Implementation and performance of the frequency-dependent GW method within the PAW framework. *Phys. Rev. B*, 74:035101, 2006.
- [105] L. Caramella, G. Onida, F. Finocchi, L. Reining, and F. Sottile. Optical properties of real surfaces: Local-field effects at oxidized Si(100)(2 x 2) computed with an efficient numerical scheme. *Phys. Rev. B*, 75:205405, 2007.
- [106] G. Strinati, H. J. Mattausch, and W. Hanke. Dynamical correlation effects on the quasiparticle bloch states of a covalent crystal. *Phys. Rev. Lett.*, 45:290–294, 1980.
- [107] G. Strinati, H.J. Mattausch, and W. Hanke. Dynamical aspects of correlation corrections in a covalent crystal. *Phys. Rev. B*, 25:2867–2888, 1982.
- [108] P. Carrier, S. Rohra, and A. Görling. General treatment of the singularities in Hartree-Fock and exact-exchange Kohn-Sham methods for solids. *Phys. Rev. B*, 75:205126, 2007.
- [109] J. Spencer and A. Alavi. Efficient calculation of the exact exchange energy in periodic systems using a truncated coulomb potential. *Phys. Rev. B*, 77:193110, 2008.
- [110] H.J. Monkhorst and J.D. Pack. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13:5188–5192, 1976.
- [111] F. Gygi and A. Baldereschi. Self-consistent Hartree-Fock and screened-exchange calculations in solids: Application to silicon. *Phys. Rev. B*, 34:4405–4408, 1986.
- [112] B. Wenzien, G. Cappellini, and F. Bechstedt. Efficient quasiparticle band-structure calculations for cubic and noncubic crystals. *Phys. Rev. B*, 51:14701–14704, 1995.
- [113] A. Görling and M. Levy. Exact Kohn-Sham scheme based on perturbation theory. *Phys. Rev. A*, 50:196–204, 1994.
- [114] C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, R.W. Godby, and M. Scheffler. Dielectric anisotropy in the GW space-time method. *Computer Physics Communications*, 176:1 – 13, 2007.

- [115] F. Aryasetiawan. *Advances in Condensed Matter Science vol.1*, volume 1. Gordon & Breach, London, 2000.
- [116] R.W. Godby, M. Schlüter, and L.J. Sham. Accurate exchange-correlation potential for silicon and its discontinuity on addition of an electron. *Phys. Rev. Lett.*, 56:2415–2418, 1986.
- [117] R.W. Godby, M. Schlüter, and L.J. Sham. Quasiparticle energies in GaAs and AlAs. *Phys. Rev. B*, 35:4170–4171, 1987.
- [118] R.W. Godby, M. Schlüter, and L.J. Sham. Trends in self-energy operators and their corresponding exchange-correlation potentials. *Phys. Rev. B*, 36:6497–6500, 1987.
- [119] R.W. Godby, M. Schlüter, and L.J. Sham. Self-energy operators and exchange-correlation potentials in semiconductors. *Phys. Rev. B*, 37:10159–10175, 1988.
- [120] H.N. Rojas, R.W. Godby, and R.J. Needs. Space-time method for ab initio calculations of self-energies and dielectric response functions of solids. *Phys. Rev. Lett.*, 74:1827–1830, 1995.
- [121] S. Ismail-Beigi and S.G. Louie. Self-trapped excitons in silicon dioxide: Mechanism and properties. *Phys. Rev. Lett.*, 95:156401, 2005.
- [122] R.W.G Wyckoff. *Crystal Structures*. Interscience, New York, 1963.
- [123] A.M. Rappe, K.M. Rabe, E. Kaxiras, and J.D. Joannopoulos. Optimized pseudopotentials. *Phys. Rev. B*, 41:1227–1230, 1990.
- [124] L. Kleinman and D.M. Bylander. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.*, 48:1425–1428, 1982.
- [125] M. Fuchs and M. Scheffler. Ab initio pseudopotentials for electronic structure calculations of poly-atomic systems using density-functional theory. *Computer Physics Communications*, 119:67–98, 1999.
- [126] D.R. Hamann. Generalized gradient theory for silica phase transitions. *Phys. Rev. Lett.*, 76:660–663, 1996.
- [127] J.R. Chelikowsky, H.E. King, N. Troullier, J. Lus Martins, and J. Glinnemann. Structural properties of alpha-quartz near the amorphous transition. *Phys. Rev. Lett.*, 65:3309–3312, 1990.
- [128] H.R. Philipp. Optical transitions in crystalline and fused quartz. *Solid State Communications*, 4:73 – 75, 1966.
- [129] F. Bottin, S. Leroux, A. Knyazev, and G. Zérah. Large-scale ab initio calculations based on three levels of parallelization. *Computational Materials Science*, 42:329–336, 2007.

- [130] F. Bruneval and X. Gonze. Accurate GW self-energies in a plane-wave basis using only a few empty states: Towards large systems. *Phys. Rev. B*, 78:085125, 2008.
- [131] R. Shaltaf, G.-M. Rignanese, X. Gonze, F. Giustino, and Pasquarello. A. Band offsets at the *si/sio₂* interface from many-body perturbation theory. *Phys. Rev. Lett.*, 100:186401, 2008.
- [132] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758–1775, 1999.
- [133] N.A.W. Holzwarth, G.E. Matthews, R.B. Dunning, A.R. Tackett, and Y. Zeng. Comparison of the projector augmented-wave, pseudopotential, and linearized augmented-plane-wave formalisms for density-functional calculations of solids. *Phys. Rev. B*, 55:2005–2017, 1997.
- [134] N.A.W. Tackett, A. R. and Holzwarth and G.E. Matthews. A projector augmented wave (PAW) code for electronic structure calculations, Part II: pwpaw for periodic solids in a plane wave basis. *Computer Physics Communications*, 135:348 – 376, 2001.
- [135] M. Torrent, F. Jollet, G. Bottin, G. Zerah, and X. Gonze. Implementation of the projector augmented-wave method in the ABINIT code: Application to the study of iron under pressure. *Computational Materials Science*, 42:337–351, 2008.
- [136] O.K. Andersen. Linear methods in band theory. *Phys. Rev. B*, 12:3060–3083, 1975.
- [137] R.D. King-Smith, M.C. Payne, and J.S. Lin. Real-space implementation of non-local pseudopotentials for first-principles total-energy calculations. *Phys. Rev. B*, 44:13063–13066, 1991.
- [138] N.A.W. Holzwarth, A.R. Tackett, and G.E. Matthews. A projector augmented wave (PAW) code for electronic structure calculations, Part I: atompaw for generating atom-centered functions. *Computer Physics Communications*, 135:329 – 347, 2001.
- [139] A. Kiejna, G. Kresse, J. Rogal, A. De Sarkar, K. Reuter, and M. Scheffler. Comparison of the full-potential and frozen-core approximation approaches to density-functional calculations of surfaces. *Phys. Rev. B*, 73:035404, 2006.
- [140] M. Marsman and G. Kresse. Relaxed core projector-augmented-wave method. *The Journal of Chemical Physics*, 125:104101, 2006.
- [141] E. Wimmer, H. Krakauer, M. Weinert, and A. J. Freeman. Full-potential self-consistent linearized-augmented-plane-wave method for calculating the electronic structure of molecules and surfaces: *o₂* molecule. *Phys. Rev. B*, 24:864–875, 1981.
- [142] M. Rotenber, Bivins. R., Metropolis. N, and J.K. Wooten. *The 3-j and 6-j symbols*. Technology Press, Massachusetts Institute of Technology, Cambridge, MA, 1959.

- [143] S.G. Louie, S. Froyen, and M.L. Cohen. Nonlinear ionic pseudopotentials in spin-density-functional calculations. *Phys. Rev. B*, 26:1738–1742, 1982.
- [144] M. Shishkin and G. Kresse. Self-consistent GW calculations for semiconductors and insulators. *Phys. Rev. B*, 75:235102, 2007.
- [145] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. 1970.
- [146] K.H. Hellwege and O. Madelung. *Numerical Data and Functional Relationships in Science and Technology*. Landolt-Börnstein, New Series, Group III, Vols 17 and 22, 1982.
- [147] B. Arnaud and M. Alouani. Local-field and excitonic effects in the calculated optical properties of semiconductors from first-principles. *Phys. Rev. B*, 63:085208, 2001.
- [148] F. Giustino, Marvin L.C., and S.G. Louie. Electron-phonon interaction using wannier functions. *Phys. Rev. B*, 76:165108, 2007.
- [149] F. Giustino, M.L. Cohen, and S.G. Louie. Small phonon contribution to the photoemission kink in the copper oxide superconductors. *Nature*, 452:975–978, 2008.
- [150] R.D. King-Smith and D. Vanderbilt. Theory of polarization of crystalline solids. *Phys. Rev. B*, 47:1651–1654, 1993.
- [151] D. Vanderbilt and R.D. King-Smith. Electric polarization as a bulk quantity and its relation to surface charge. *Phys. Rev. B*, 48:4442–4455, 1993.
- [152] R. Resta. Macroscopic polarization in crystalline dielectrics: the geometric phase approach. *Rev. Mod. Phys.*, 66:899–915, 1994.
- [153] S. Goedecker. Linear scaling electronic structure methods. *Rev. Mod. Phys.*, 71:1085–1123, 1999.
- [154] I. Souza, N. Marzari, and D. Vanderbilt. Maximally localized Wannier functions for entangled energy bands. *Phys. Rev. B*, 65:035109, 2001.
- [155] J.R. Yates, X. Wang, D. Vanderbilt, and I. Souza. Spectral and Fermi surface properties from Wannier interpolation. *Phys. Rev. B*, 75:195121, 2007.
- [156] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari. Exponential localization of Wannier functions in insulators. *Phys. Rev. Lett.*, 98:046402, 2007.
- [157] A.A Mostofi, J.R Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari. Wannier90: A tool for obtaining maximally-localised Wannier functions. *Computer Physics Communications*, 178:685 – 699, 2008.

- [158] D.R. Hamann and D. Vanderbilt. Maximally localized wannier functions for GW quasiparticles. *Phys. Rev. B*, 79:045109, 2009.
- [159] P.E. Blöchl, O. Jepsen, and O.K. Andersen. Improved tetrahedron method for Brillouin-zone integration. *Phys. Rev. B*, 49:16223, 1994.
- [160] L.D Landau and E.M Lifshitz. *Electrodynamics of continuous media*. Pergamon Press, Oxford, 1960.
- [161] S. Bhagavantam. *Crystal symmetry and physical properties*. Academic Press. London and New York, 1966.
- [162] A.N. Shubnikov. Symmetry and antisymmetry of finite figures. *Izd. Acad. Sci. USSR*, 1951.
- [163] G.B. Bachelet, D.R. Hamann, and M. Schlüter. Pseudopotentials that work: From H to Pu. *Phys. Rev. B*, 26:4199–4228, 1982.
- [164] X. Gonze, R. Stumpf, and M. Scheffler. Analysis of separable potentials. *Phys. Rev. B*, 44:8503–8513, 1991.
- [165] V. Olevano, M. Palummo, G. Onida, and R. Del Sole. Exchange and correlation effects beyond the LDA on the dielectric function of silicon. *Phys. Rev. B*, 60:14224–14233, 1999.
- [166] V. Olevano, L. Reining, and F. Sottile. Dp code, <http://www.dp-code.org/>. Technical report, CNRS.
- [167] L. Reining, V. Olevano, F. Sottile, S. Albrecht, and G. Onida. Exc code, <http://theory.polytechnique.fr/codes/exc/>. Technical report, CNRS, France.
- [168] M.A. Blanco, M. Florez, and M. Bermejo. Evaluation of the rotation matrices in the basis of real spherical harmonics. *Journal of molecular structure*, 419:19–27, 1997.