Ab Initio Electron Energy Loss Spectroscopy

A dissertation submitted for the degree of *Doctor of Philosophy* at the University of Cambridge

Chris J. Pickard
Christ's College, Cambridge

September 1997

Preface

This dissertation describes work done between October 1994 and September 1997 at the Cavendish Laboratory, Cambridge, under the supervision of Dr. Mike Payne.

Except where stated otherwise, this dissertation is the result of my own work and contains nothing which is the outcome of work done in collaboration. This dissertation has not been submitted in whole or in part for any degree or diploma at this or any other university.

Chris J. Pickard
September 1997

Acknowledgements

I would first like to thank my supervisor, Mike Payne, who gave me the chance to work as I felt, total support and encouragement. I am also particularly indebted to the care he has taken in the reading of this thesis. Mick Brown, who made it possible for me to be fully a member of two groups in the Cavendish at once, has been an inspiration. His approach to physics, which I first experienced in the IB Solid State Physics course, has left its mark on me. Peter Rez, who originally suggested the use of CASTEP to calculate ELNES, has adopted me as one of his students — convincing me that I was doing something worthwhile. There have been many others without whom this work would have been impossible:

Phil Batson for interesting discussions, donation of experimental data and finding that crucial paper, Mark Gibbs, for going with the crazy idea of trying to use neural networks to solve the bandstructure problem, Brian Rafferty for the 3Ds of discussion, data and diagrams, Jun Yuan for his thorough knowledge of the theory, Eric Sandre for getting me to do things I might otherwise have not done and vice versa, Phillipe Moreau for discussions and encouragement, Matt Segall for getting me out of all sorts of trouble (on the computers), Rajiv Shah for help and advice of all kinds, Peter Haynes for assistance on the heavier theory side of things, Murray Jarvis for his invaluable help with pseudowavefunctions, Nanda Menon for angular integrals and entertainment, Pete Nellist for enthusiastic discussions, Rik Brydson for his interest and encouragement, Ming Hsien Lee for the pseudopotentials, Victor Milman for help with the optical code, everyone else in the Microstructural Physics and Theory of Condensed Matter Physics groups who allowed me to fill the days with coffee breaks, Charles Wildig, my friends from all over the world, and especially to my grandmother, who showed me how to really tackle the problem of eels.

This work was supported by an EPSRC studentship, with generous assistance towards travel from Christ's College, Cambridge.

This work is dedicated to my family in thanks for their unending support and love.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Electron Energy Loss Spectroscopy

Many modern Scanning Transmission Electron Microscopes (STEMs) are fitted with spectrometers which allow the energy lost by the electrons as they pass through the sample to be measured as an Electron Energy Loss Spectrum (EELS) with a resolution, at present, of about 0.3eV — see Figure 1.1 for a description of such a microscope[1]. Through the scanning nature of these microscopes, a very fine beam of electrons ($\sim$1Å) can be moved around the sample, allowing spectral properties to be measured at an atomic spatial resolution [6, 68].

Electron Energy Loss Spectroscopy probes the unoccupied electronic states of materials [27]. Energy is lost as transitions are made between the valence states and the unoccupied conduction band states, giving rise to what is known as Low Loss spectra. Or transitions may take place from tightly bound core states, producing the Energy Loss Near Edge Structure (ELNES) at the onset of the transitions and Extended Energy Loss Fine Structure (EXELFS) at higher energies. Figure 1.2 summarises these transitions. Knowledge of the spectral properties (i.e. the variation with energy and symmetry) of these states is important in several ways. Since the local potential directly effects the energy distribution of the electronic states the spectral properties are a *fingerprint* of the local structure (local since the focussed beam of electrons in a STEM forms a local probe). The unoccupied states also give indirect access to the occupied states, and hence to information about the bonding energetics [67]. Finally, since they dictate the electronic — e.g. optical and transport — properties of the system, direct observation of the unoccupied states themselves is of interest.

[1]With thanks to Brian Rafferty

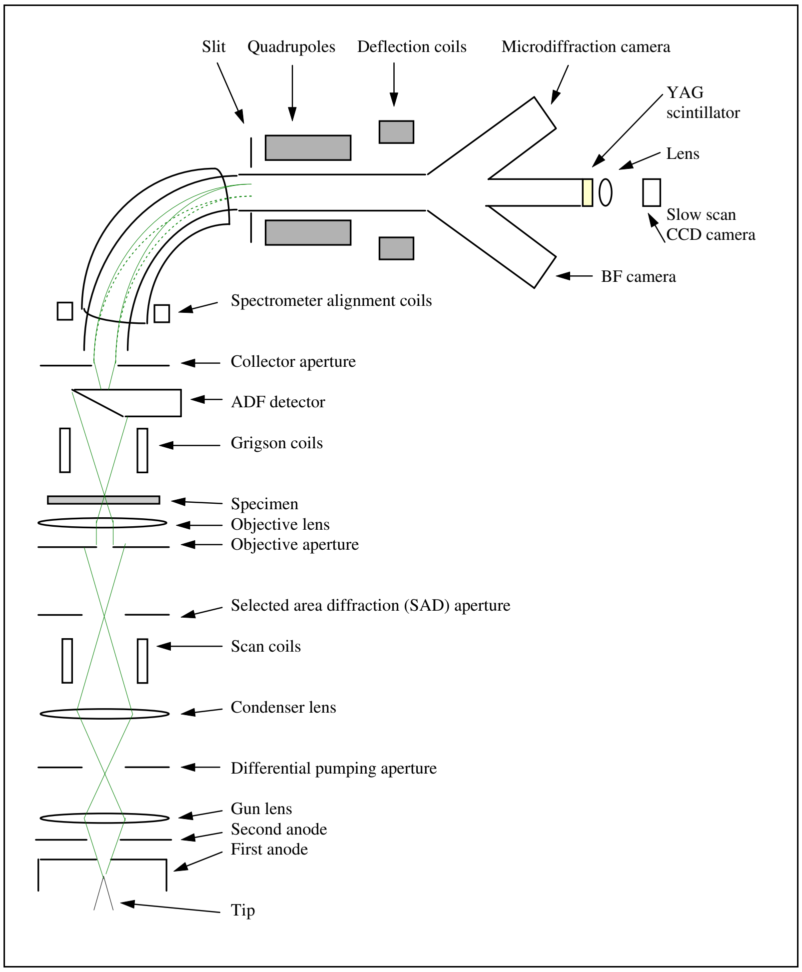


Figure 1.1: Schematic description of a Scanning Transmission Electron Microscope

With the development of computer power and theoretical tools the interpretation of ELNES can be taken beyond the fingerprint level through detailed calculation. If these calculations are well controlled they can demonstrate clearly the degree to which the energy loss spectra are predicted by a given level of theoretical approximation. If theory and experiment do not agree, the theory is then proved to be incomplete and must be developed further. Once a model of the spectra has been developed it may be used as an aid to intuition or as a quantitative tool in the interpretation of experimental results. In this thesis a scheme for the *ab initio* calculation of ELNES is presented, building on earlier work by Brohan [15], based on a Density Functional Theory (DFT) planewave pseudopotential total energy code — CASTEP [73].

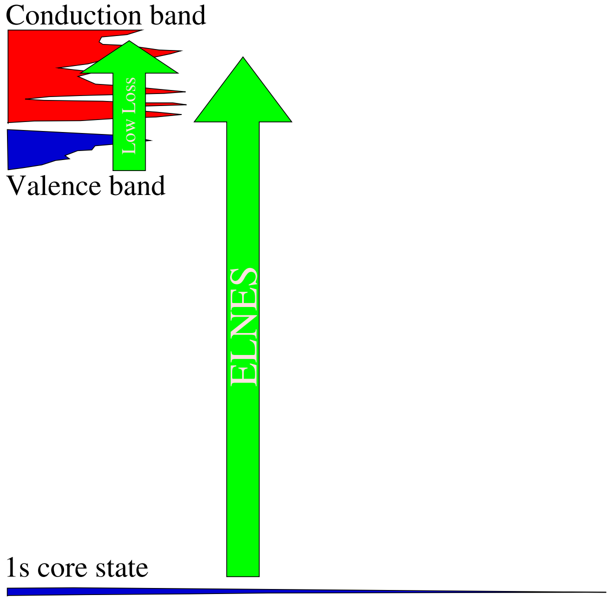


Figure 1.2: Pictorial representation of the energy loss process — in fact the results of an all electron calculation for Diamond. Blue denotes occupied states, and red unoccupied.

1.2 Energy Loss Near Edge Structure

The focus of this thesis is on the ELNES region of the energy loss spectrum. For a fast incident electron which loses an energy E and momentum $\mathbf{q}$ the differential cross section can be related to the imaginary part of the dielectric function, which within the single particle approximation is given for a periodic system by:

$$\varepsilon_2(\mathbf{q}, E) = \frac{4\pi e^2}{\Omega q^2} \sum_{n,\mathbf{k}} |\langle \psi^n_{\mathbf{k}} | e^{i\mathbf{q}.\mathbf{r}} | \mathrm{c} \rangle|^2 \delta(E^c_{\mathbf{k}} - E_{1\mathrm{s}} - E), \tag{1.1}$$

where $|c\rangle$ is a core state and Ω is the unit cell volume. In the scheme to be described in this thesis the final states and energies — $|\Psi^n_{\mathbf{k}}\rangle$ and $E^n_{\mathbf{k}}$ — derive from an *ab initio* electronic structure calculation. The sum is over $\mathbf{k}$ within the first Brillouin zone and unoccupied bands n, i.e. all final states. The two main requirements for the calculation of ELNES within this scheme are the evaluation of the density of states (DOS) term and the transition matrix elements. The theory behind the calculation of ELNES at this level of theory is described in detail in Chapter 2.

1.3 The planewave pseudopotential approach

In order to calculate the final electronic states and energies that appear in Equation 1.1 a DFT planewave total energy code (CASTEP) is used [73]. It is based on DFT within the Local Density Approximation. A planewave basis set and periodic boundary conditions are used (and hence summations over all states become sums over bands and an integration over the first Brillouin zone). Non-local pseudopotentials are used to overcome the computational problems associated with the rapidly varying wavefunctions within the core regions of the atoms. The technique is reviewed further in Chapter 2.

1.4 Brillouin zone integration

In order to calculate the ELNES to high accuracy the Brillouin zone integrations must themselves be performed accurately. In this thesis a highly efficient Brillouin Zone integration scheme is developed which requires a very low **k**-point sampling. It will be shown that it is important to use an extrapolative approach, since interpolative ones are hindered by band crossing at low sampling densities, introducing spurious singularities in the DOS. The information for the extrapolation is obtained using **k.p** perturbation theory to second order (thus the correct analytic behaviour at van Hove singularities is obtained) within a set of sub-cells into which the Brillouin zone is divided (efficiently chosen to make full use of symmetry at low sampling densities). Some corrections to the **k.p** expansion are required for the use of non-local pseudopotentials. The resulting piecewise quadratic representation of the bandstructure is converted directly to a DOS using the analytic quadratic approach of Methfessel [62]. This scheme is described in greater detail in Chapter 3.

1.5 ELNES matrix elements

In the pseudopotential approximation the core states are not explicitly calculated, and the pseudowavefunctions differ from the true wavefunctions within a radius r_c from the nucleus (r_c is typically of the order of 0.5Å). In order to evaluate the matrix elements the initial core states are taken from all electron calculations for isolated atoms — consistent with the frozen core approximation on which the pseudopotential approximation is based. In practice the dipole approximation is taken, but it is shown that it is straightforward to go further in this formalism. The matrix elements are then directly evaluated and corrected for the deviation of the pseudowavefunction from the true wavefunction within r_c using the Projector Augmented Wave approach of van de Walle and Blöchl [25]. In this way the matrix elements are evaluated quantitatively, resulting in an absolute prediction for the cross-section. This is in contrast to many previous bandstructure approaches to predicting the ELNES, which typically employ a straightforward local symmetry projection of the DOS [102].

1.6 Single particle core hole effects

The "single particle" core hole effects arise from the un-screening of the nuclear charge as an electron is excited from a localised core state. In a metal this potential is almost entirely re-screened and there is very little effect due to the core hole. In an insulator the screening of the hole by the valence electrons is only partially achieved. Physically, the excited atom can be thought of as an impurity in the material, with an associated distortion of the electronic structure (and no atomic relaxation on the time-scale of the excitation). In fact, impurity models — such as the Clogston-Wolff model — have been used to model core hole effects in systems where the screening is almost complete and hence the impurity potential differs only slightly from the unperturbed one. Weijs [100] studied the core hole effects in the X-ray absorption spectra of transition metal silicides in this way.

For many systems, the hybridisation of the unoccupied orbitals due to core hole potential cannot be treated perturbatively, and the effects must be calculated explicitly. Within a bandstructure approach, this is achieved by performing a supercell calculation in which there is a single excited potential in each cell. The spectra is then evaluated for that excited atom. The supercell must be made large enough that the neighbouring excited potentials do not interact with each other. Such calculations have previously been performed for Diamond [59] and Graphite [1] . In this thesis, while continuing to work within the pseudopotential approximation, the more common Z+1 approximation to the excited atom is improved upon by generating a special pseudopotential with partially occupied core states.

K-edges for Diamond, Graphite and cubic Boron Nitride are calculated and presented in Chapter 5. It will be shown that the inclusion of single particle core hole effects (through the supercell approximation) brings significant improvement to the calculated spectra in all these cases.

1.7 Previous approaches to the calculation of ELNES

Rez *et al* [81] present a review of the various theortetical and computational techniques currently available for the calculation of near edge structure. A summary of the hierarchy they established is shown in Table 1.1. Including the core hole effects, the planewave pseudopotential approach described in this work sits at the top of this hierarchy. While Rez *et al* claim that bandstructure methods will fail at higher energies due to "convergence problems", this work will demonstrate that the planewave pseudopotential approach shows extremely good agreement with experiment up to 60eV above the threshold (see Figure 5.11). Rez *et al* also expect that modifications due to core hole effects will be restricted to within $\sim$0.5eV of the threshold. However, the results presented in Chapter 5 show the core hole to have effects to relatively high energies, well into the conduction band.

1.8 Optical properties

While the focus of this work is primarily on high energy core loss spectra — the ELNES — it has proved natural to extend the techniques developed to the calculation of the optical properties of insulators. These optical properties are determined by transitions from the occupied valence to unoccupied conduction band states — the same as for the low loss region of an electron energy loss spectrum (see Figure 1.2). Appendix C is a self contained description of the calculation of the optical properties.

1.9 Summary

This thesis will show that ELNES can be calculated within the framework of a total energy pseudopotential code. The description of the wavefunctions in terms of planewaves allows the straightforward direct evaluation of the transition matrix elements, and corrections for the pseudopotentials can be included. Some of the computational expense of the technique is offset by efficient Brillouin zone integrations, and the importance of including the core hole is demonstrated. It is hoped that such a scheme, in concert with the structure prediction of these *ab initio* techniques will prove to be a powerful tool in the investigation of microstructure.

	Theory	Methods	Features
I	*Atomic calculation*	atomic multiplet [96], multiconfigurational Dirac Fock (MCDF)	basic edge shapes, L2,3 ratio, multiplet fine structure, crystal field splitting can be included
II	*Single scattering*	EXAFS and EXELFS methods [92]	reflections from neighbouring atoms, describes extended structure above ∼50eV
III	*Multiple scattering*	XANES methods [26]	multiple reflections, *not* self-consistent, poor description of threshold
IV	*Self-consistent band theory*	muffin tin — augmented plane wave (APW) [9], augmented spherical wave (ASW) [37, 100], Korringa-Kohn-Rostoker (KKR) [58] ...	self-consistent potentials, all electron, directly interpretable local angular momentum resolved DOS
		pseudopotential — pseudo-atomic orbital [102] and planewave [15]	pseudo-atomic orbitals give direct interpretation of DOS, but the inadequate basis leads to failure at high energies. Planewaves are discussed in the current work
V	*Inclusion of core effects*		See Chapter 5

Table 1.1: Hierarchy of the approaches to the calculation of near edge structure — following Rez *et al* [81]

Chapter 2

The theory of Electron Energy Loss Spectroscopy

In this chapter the theoretical background to the calculations performed in the remainder of this work is presented. As discussed in the introduction, the aim is to calculate how energy is lost by fast incident electrons as they are passed though thin samples in Scanning Transmission Electron Microscopes. In particular the transitions from core levels into the empty lower conduction band states will be examined — giving rise to the Energy Loss Near Edge structure (ELNES).

In Section 2.1 an expression for the inelastic cross section is presented and its relationship to the single particle electronic states is discussed. Section 2.2 describes the Density Functional Theory (DFT) based total energy pseudopotential approach to the evaluation of these single particle states. Finally, Section 2.3 examines the interpretation of the single particle states and their use in the calculation of the ELNES.

2.1 Cross section for inelastic electron scattering

In Electron Energy Loss Spectroscopy (EELS) high energy ($E \approx 100$keV) electrons are passed through thin samples. These electrons scatter in a variety of ways. The elastically scattered electrons are collected to form high magnification images of the sample. However, some of the electrons will lose a large amount of energy on their passage through the sample. The Coulomb potential of the fast electron causes excitations amongst the electrons in the sample. A spectrometer (see Figure 1.1) separates the transmitted electrons according to their energy loss, and the number of electrons counted for each energy loss is measured and used to produce a spectrum.

2.1.1 The inelastic scattering geometry

Figure 2.1 illustrates the scattering geometry for an electron of well defined incident energy E_o and momentum $\mathbf{k}_o$. As the electron passes through the sample it scatters through an angle θ, losing an energy ΔE and momentum $\mathbf{q}$. The momentum transfer can be broken down into components parallel and perpendicular to the incident direction, as shown in the diagram. Since the relevant length-scale for atoms in solids is 1Å, the corresponding relavent momentum transfers will be of order 1Å^{-1}. For 100keV electrons $\mathbf{k}_o \approx 200\text{Å}^{-1}$, which is much larger than $\mathbf{q}$. Since, in these circumstances, $q_\perp$ can be approximated by $k_o\theta$, collection angles should be of the order of $1\text{Å}^{-1}/200\text{Å}^{-1} \approx 5$ mrad which is indeed the case experimentally. The important point is that, by changing the scattering geometry, the momentum transfered to the sample during the energy loss process can be altered — hence the measurements become a function of the momentum transfer. However in the actual experimental arrangement, neither the incident nor collected scattered electrons are single planewaves. In the case of the STEM the incident electron is formed into a probe localised transversely to approximately atomic dimensions. Hence, the probe will contain transverse momentum components of the order of 1Å^{-1}. Also, a single planewave component is never collected — an integration over a range of angles, and hence momenta, is performed by the detector. The details of these integrations have been worked out [16] and are not considered further here, where the focus is on the underlying quantum mechanical excitations.

2.1.2 The partial differential cross section

In an EELS measurement it is the partial differential cross section $\frac{d^2\sigma}{d\Omega dE}$ that is measured — the fraction of the incident electrons which are scattered into a solid angle $d\Omega$ having lost an energy between ΔE and $\Delta E + dE$. This cross section can be expressed in terms of the macroscopic dielectric function $\varepsilon_M(\mathbf{q}, \omega)$ [76].

$$\frac{d^2\sigma}{d\Omega dE} = \frac{1}{(\pi e a_o)^2}\frac{1}{q^2} Im\left\{\frac{-1}{\varepsilon_M(\mathbf{q}, \omega)}\right\} \tag{2.1}$$

As mentioned above, an actual measurement will involve an integral over a range of solid angles. The quantity $Im\left\{-1/\varepsilon_M(\mathbf{q}, \omega)\right\}$ is known as the loss function, since it provides information about the energy loss process for an electron. This quantity is introduced

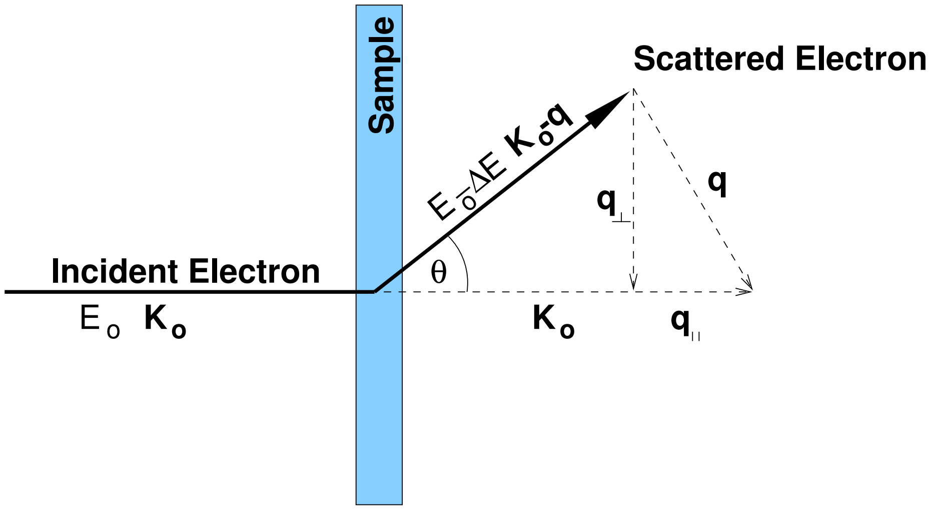


Figure 2.1: Scattering geometry of an electron energy loss event

again in Appendix C, where the focus is on the lower energy excitations of the electrons, while here we concentrate on the higher energy core loss excitations.

2.1.3 The dielectric function

The dielectric function $\varepsilon(\mathbf{q},\omega)$ describes the linear response of an electronic system to an applied external electric field. The real part $Re\{\varepsilon(\mathbf{q},\omega\} = \varepsilon_1(\mathbf{q},\omega)$ is related to the polarisability of the system, while the imaginary part $Im\{\varepsilon(\mathbf{q},\omega)\} = \varepsilon_2(\mathbf{q},\omega)$ describes real electronic transitions. Ehrenreich and Cohen [28] derived the following expression for the dielectric function in terms of the single particle states of the system $|\sigma\rangle$:

$$\varepsilon(\mathbf{q},\mathbf{q};\omega) = 1 - \frac{4\pi e^2}{q^2\Omega}\sum_{\sigma,\sigma'}\frac{f(\sigma)-f(\sigma')}{\hbar\omega - E_\sigma + E_{\sigma'} + i\delta}\left|\langle\sigma|e^{-i\mathbf{q}\cdot\mathbf{r}}|\sigma'\rangle\right|^2, \tag{2.2}$$

where $f(\sigma)$ is the Fermi-Dirac distribution function and Ω is the volume. One can use the relation $1/(x+i\delta) = P(1/x) - i\pi\delta(x)$ to obtain the imaginary part of the dielectric function:

$$\varepsilon_2(\mathbf{q},\mathbf{q};\omega) = \frac{4\pi e^2}{q^2\Omega}\sum_{\sigma,\sigma'}(f(\sigma)-f(\sigma'))\delta(\hbar\omega - E_\sigma + E_{\sigma'})\left|\langle\sigma|e^{-i\mathbf{q}\cdot\mathbf{r}}|\sigma'\rangle\right|^2 \tag{2.3}$$

The imaginary part of the dielectric function is therefore given by a joint density of states (JDOS) weighted by the appropriate matrix elements, $\langle\sigma|e^{-i\mathbf{q}.\mathbf{r}}|\sigma'\rangle$, which give rise to the selection rules for the transitions.

2.1.4 Core-level excitations

Figure 1.2 illustrates the distinction between low loss excitations between the valence and conduction band states (see Appendix C) and those from the tightly bound core states. In both cases the final states are the same — the energy of the excitations being determined by the initial states. A threshold occurs at an energy given by the difference between the core state concerned and the lowest unoccupied state accessible from that particular core. Since different core states are often well separated in energy, most thresholds are also well separated. It is usual to consider a single threshold at a time.

Since, $Im\{-1/\varepsilon\} = \varepsilon_2/(\varepsilon_1^2 + \varepsilon_2^2)$ and at high energies ε_1 is close to unity (there is no polarisation response at such high frequencies) while ε_2 is small, the loss function can be written as [31]:

$$Im\left\{\frac{-1}{\varepsilon(\mathbf{q},\omega)}\right\} = \varepsilon_2(\mathbf{q},\mathbf{q};\omega) = \frac{4\pi e^2}{q^2\Omega}\sum_{\sigma}\left|\langle\sigma|e^{-i\mathbf{q}.\mathbf{r}}|c\rangle\right|^2\delta(\hbar\omega - E_\sigma + E_c), \tag{2.4}$$

where $|c\rangle$ is a core state on the site of interest, and $|\sigma\rangle$ is the unoccupied final state, and E_c and E_σ are the respective energies of the states. In principle, as in Equation 2.3, the loss function is dominated by a joint density of states. But, the core states, being so tightly bound, are almost entirely non-dispersive (indeed, if they were dispersive they could hardly be termed core states). So, the loss function in fact measures the weighted unoccupied density of states — the Energy Loss Near Edge Structure (ELNES). The matrix element $\langle\sigma|e^{-i\mathbf{q}.\mathbf{r}}|c\rangle$, hereafter known as the ELNES matrix element, is discussed further in Chapter 4.

2.1.5 Beyond the single particle approximation

So far we have ignored any effects on the ELNES beyond the single particle approximation. There can be strong many-body effects due to the coupling of the core hole created on the excitation of an electron, and the final states (the so called multiplet structure [103]). Weaker, "single particle" core hole effects manifest themselves as a modification of the threshold. If these effects are weak they can be straightforwardly modelled in a perturbative approach. However, if the effects are stronger the core hole must be explicitly taken into account in the theoretical calculation. Chapter 5 shows how moderate core hole effects may be taken into account.

2.2 The planewave pseudopotential method

In this section the problem of evaluating quantum mechanically the ground-state electronic density and total energy of a system of interacing electrons for a given nuclear configuration is discussed. While the core of the methodology is presented here, more detailed reviews can be found in Jones and Gunnarsson [47] (on the Density Functional formalism) and in Payne *et al* [73] (on its applications to first principles molecular dynamics).

2.2.1 The many-body Schrödinger equation

The behaviour of a system of N electrons can be predicted simply by solving the Schrödinger Equation for the system:

$$\hat{H}\Psi = E\Psi \tag{2.5}$$

The problems arise in attempting to solve this equation. Ψ is the many-body wavefunction and is an anti-symmetric (to satisfy the Fermi statistics of electrons) function of the electron co-ordinates $\{\mathbf{r}_i : i = 1, N\}$, and the Hamiltonian, $\hat{H}$, is given by:

$$\hat{H} = -\frac{\hbar^2}{2m}\sum_i \nabla^2_{\mathbf{r}_i} + V_{ext}(\{\mathbf{r}_i\}) + V_{e-e}(\{\mathbf{r}_i\}) \tag{2.6}$$

The eigenvalue E is the total energy of the system, essentially determined by the external potential V_{ext} which describes the Coulomb interaction between the electrons and a given configuration of nuclei. The term V_{e-e} gives the electron-electron Coulomb interaction, and it is this term which introduces the coupling between the electronic co-ordinates, and precludes a straightforward separation of the many-body wavefunction which would make the solution of the problem very simple computationally.

2.2.2 Density Functional Theory

In Section 2.2.1 it was noted that the term V_{e-e} in the Schrödinger Equation introduces a coupling between the electronic co-ordinates of the many electrons in the system. This coupling is often referred to as "correlation" — as an electron moves the other electrons feel its Coulomb potential, experience a force and move in response. Hence the motion of the electrons is correlated. Although the physics is straightforward, the mathematics of the problem becomes intractable for all but the simplest systems. A step towards the solution of this problem was made by Hohenberg and Kohn [42]. They introduced the concept of the electronic density $n(\mathbf{r})$ as a basic variable, within the framework of density functional theory (DFT). They showed that the ground-state total energy E could be written as a functional of the ground-state electronic density,

$$E[n(\mathbf{r})] = F[n(\mathbf{r})] + \int V_{ext}(\mathbf{r})n(\mathbf{r})d^3\mathbf{r}. \tag{2.7}$$

where $F[n(\mathbf{r})]$ is a universal functional, so that it is $V_{ext}(\mathbf{r})$ which uniquely describes any particular physical system. Hohenberg and Kohn also showed that the density which minimised E in Equation 2.7 is the ground-state density. Unfortunately the functional is not known, and hence DFT is of little use in this form.

A practical scheme for DFT calculations became possible following the work of Kohn and Sham [51]. They chose to write the density in terms of a set of orthonormal functions, one for each of the N electrons in the system.

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2 \tag{2.8}$$

$F[n(\mathbf{r})]$ is separated into three terms:

$$F[n(\mathbf{r})] = T_S[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{XC}[n(\mathbf{r})], \tag{2.9}$$

where T_S is the kinetic energy term,

$$T_S = \sum_i \frac{\hbar^2}{2m} \int \phi_i^* \nabla^2 \phi_i d^3\mathbf{r}. \tag{2.10}$$

This is not equal to the true electronic kinetic energy for the system, but it is of similar magnitude and most importantly it can be computed exactly. It is known as the non-interacting kinetic energy. Previous approaches to approximating the functional $F[n(\mathbf{r})]$ (e.g. that of Thomas-Fermi [47]) attempted to approximate the kinetic energy purely in terms of the density. These approaches failed, since the kinetic energy makes up large part of the total energy, and the approximations used did not give an accurate enough value for the kinetic energy. The term E_H in Equation 2.9 describes the Coulomb energy of the electron density $n(\mathbf{r})$, which is the same as the electron-electron energy in the Hartree approximation:

$$E_H = \frac{1}{2}\int\int \frac{n(\mathbf{r}')n(\mathbf{r})}{|\mathbf{r}' - \mathbf{r}|} d^3\mathbf{r} d^3\mathbf{r}'. \tag{2.11}$$

Thus far the terms in the total energy have been defined to be exact. E_{XC} describes the rest of the contributions to the total energy, making up the difference between $T_S + E_H$ and the true functional F. It is known as the *exchange-correlation* energy, and is the only quantity that is approximated in the Kohn-Sham approach. The most common, and straightforward, approximation to E_{XC} is the local density approximation (LDA). Within the LDA, E_{XC} is written as,

$$E_{XC} = \int \epsilon_{xc}(n(\mathbf{r}))n(\mathbf{r}) d^3\mathbf{r}, \tag{2.12}$$

where $\epsilon_{xc}(n)$ is the exchange-correlation energy per unit volume of a homogeneous electron gas with a density of n. Monte Carlo total energy calculations have been performed for uniform electron gases at a variety of electron densities, and by subtracting T_S and E_H (which can both be straightforwardly evaluated), ϵ_{xc} can be extracted and parameterised [19, 74]. Given that this parameterisation is based on data for homogeneous charge densities, the LDA might be expected only to be strictly valid for systems in which the charge density is slowly varying, which is clearly not the case in a general covalently bound solid. However, experience has shown the LDA to be a very good approximation for a wide variety of systems in the solid state. This success can be attributed to the fact that the LDA adheres to the sum-rule for the exchange correlation hole [38]. It was the failure to ensure that this sum-rule was satisfied which initially caused many supposed improvements to the LDA to fail. Now generalised gradient approximations (GGA) have been developed which obey the sum rule and provide better descriptions of weak molecular bonds than the LDA [48, 75], however in this thesis calculations are performed exclusively within the LDA.

Having set up the formalism of Kohn and Sham, its practical implementation is now examined. Through Equation 2.8, the functional $E[n]$ has now been expressed in terms of a set of functions $\{\phi_i(\mathbf{r})\}$, thus minimising $E[n]$ with respect to this set, subject to the constraint that they remain orthonormal leads to the set of equations:

$$-\frac{\hbar^2}{2m}\nabla^2\phi_i + V_H\phi_i + V_{ext}\phi_i + V_{xc}\phi_i = \varepsilon_i\phi_i, \tag{2.13}$$

where,

$$V_H = \frac{\delta E_H}{\delta n(\mathbf{r})}, \tag{2.14}$$

and,

$$V_{xc} = \frac{\delta E_{XC}}{\delta n(\mathbf{r})}. \tag{2.15}$$

These equations are known as the Kohn-Sham equations and it can now be recognised what role the set $\{\phi_i(\mathbf{r})\}$ is to play. The Kohn-Sham equations clearly resemble non-interacting single particle Schrödinger equations — the $\{\phi_i\}$ being eigenstates and the Lagrange multiplier ε_i the corresponding eigenvalues. Thus the many-body problem described in Section 2.2.1 has been mapped to one of a system of non-interacting single particles. These single particle states will be interpreted in Section 2.3, and will turn out to be the states required in the evaluation of the inelastic cross section. The potentials V_H and V_{xc} depend on the charge density which, through Equation 2.8, depends of the Kohn-Sham eigenstates. Hence, the Kohn-Sham equations must be solved self-consistently — the potential and the resulting charge density must be consistent.

2.2.3 Periodic Boundary Conditions

For many years the electronic structure community focussed primarily on the properties of perfect crystalline solids. This has lead to the use of periodic boundary conditions in many electronic structure methods — as in the case of the total energy pseudopotential method. The use of periodic boundary conditions, through Bloch's Theorem (see below) allow the treatment of the very large number of electrons in a crystal. As solid state physicists have moved to the study of less perfectly ordered systems, periodic boundary conditions — although no longer strictly valid — have not been abandoned. They permit the use of the highly desirable planewaves basis (see Section 2.2.4) and give the choice as to whether the various terms in Equation 2.13 should be evaluated in real or reciprocal space [46]. However, to study such problems, aperiodicity must be approximated within the supercell approach.

Bloch's theorem

For a one-electron Hamiltonian — for example that in Equation 2.13 — if the potential has a lattice periodicity (i.e. $U(\mathbf{r}) = U(\mathbf{r}+\mathbf{R})$ for all $\mathbf{R}$ where $\mathbf{R}$ is a lattice vector), then

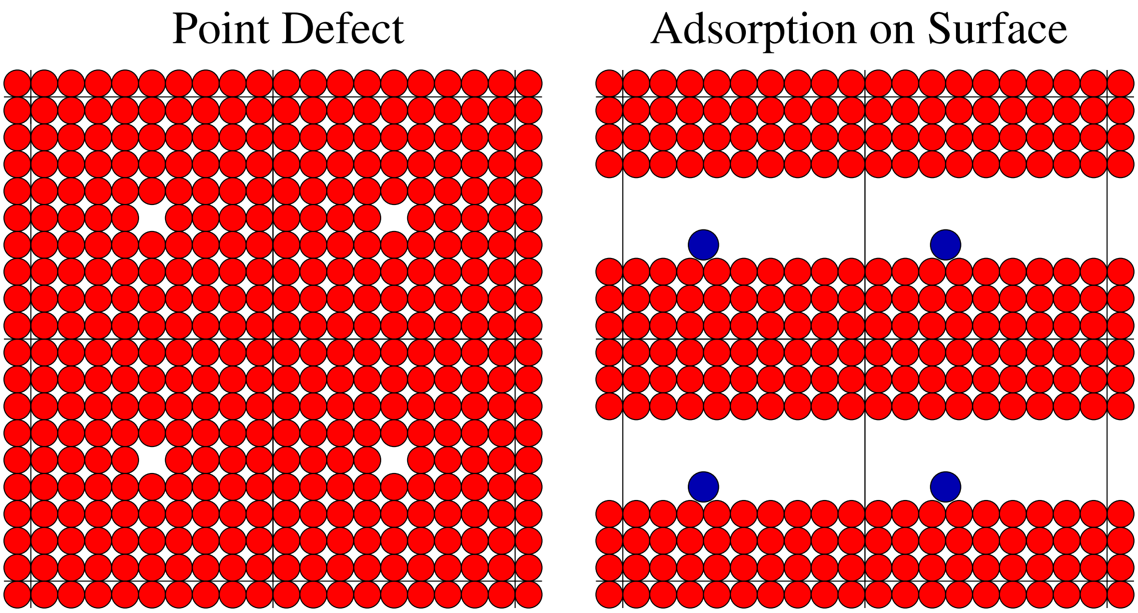


Figure 2.2: The supercell approximation —the modelling of aperiodicity within periodic boundary conditions.

the eigenstates of the Hamiltonian can be written as:

$$\Psi^n_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}.\mathbf{r}} u^n_{\mathbf{k}}(\mathbf{r}), \tag{2.16}$$

where $u^n_{\mathbf{k}}$ is a cell periodic functions such that $u^n_{\mathbf{k}}(\mathbf{r}) = u^n_{\mathbf{k}}(\mathbf{r}+\mathbf{R})$ for all lattice vectors $\mathbf{R}$. This implies that:

$$\Psi^n_{\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}.\mathbf{R}} \Psi^n_{\mathbf{k}}(\mathbf{r}) \tag{2.17}$$

On substituting $\Psi^n_{\mathbf{k}}(\mathbf{r})$ into Equation 2.13 a new set of eigenequations for $u^n_{\mathbf{k}}(\mathbf{r})$ is found, one for each value of the continuous variable $\mathbf{k}$. The problem of solving for an infinite number of electrons has become one of calculating for a finite number of bands at an infinite number of $\mathbf{k}$-points [2]. However, as will be discussed in Chapter 3, physical properties are expected to be smoothly varying functions of $\mathbf{k}$ and hence many integrals can be well approximated by a finite sampling of $\mathbf{k}$ [3, 20, 21, 65, 70].

The supercell approximation

The application of periodic boundary conditions forces periodicity on the system studied. This is significant, since many applications of electronic structure calculations are on systems which do not have full three dimensional translational symmetry — for example the study of defects, impurities or even the interaction of molecules and surfaces. In the supercell approximation, aperiodic systems are approximated by enclosing the region of interest in either bulk material (for a defect) or vacuum (for a molecule) and then periodically repeating this cell throughout space — see Figure 2.2. The supercell must be large enough for the fictitious interactions between neighbouring cells to be negligible. In Chapter 5, the treatment of core holes (which break the translational symmetry of even a perfect crystal) will be carried out using supercells.

2.2.4 The planewave basis set

In order to solve the eigenvalue problem of Equation 2.13 numerically the eigenstates $\{\phi_i\}$ must be represented by some basis set. While there are many possible choices, the one made here is to use planewaves as the basis. There are many advantages in the use of planewaves. They form a mathematically simple basis, giving a very direct representation of the electronic states. This has been particularly important in this work — being the main reason why it is straightforward to evaluate the ELNES matrix elements directly (see Chapter 4). Planewave calculations can be taken systematically to convergence as a function of the size of the basis (see below), and when forces are needed for molecular dynamics applications there is no need to consider the Pulay correction forces [78]. This is since the planewaves are not centred about atoms, and so they are not changed as the atoms move. In a similar way, it turns out that **k.p** perturbation theory is very much more straightforward in a planewave basis — see Chapter 3. However, planewave basis sets require many more basis states per atom than atom-centred orbitals — many hundreds per atom as opposed to of the order of ten. Without the use of pseudopotentials (described in Section 2.2.5) the use of planewaves would prove impractical in all but the very smallest systems (see the demanding all electron planewave calculations for Diamond in Chapter 4 and Appendix C).

The details of the planewave basis are now examined. Application of periodic boundary conditions ensures a discrete (but still infinite) basis set. The Kohn-Sham eigenstates are expressed as:

$$\Psi^n_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c^n_{\mathbf{k}}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G}).\mathbf{r}}, \tag{2.18}$$

where the sum is over all reciprocal lattice vectors $\mathbf{G}$. To truncate the basis set the sum is limited to a set of reciprocal lattice vectors contained within a sphere with a radius defined

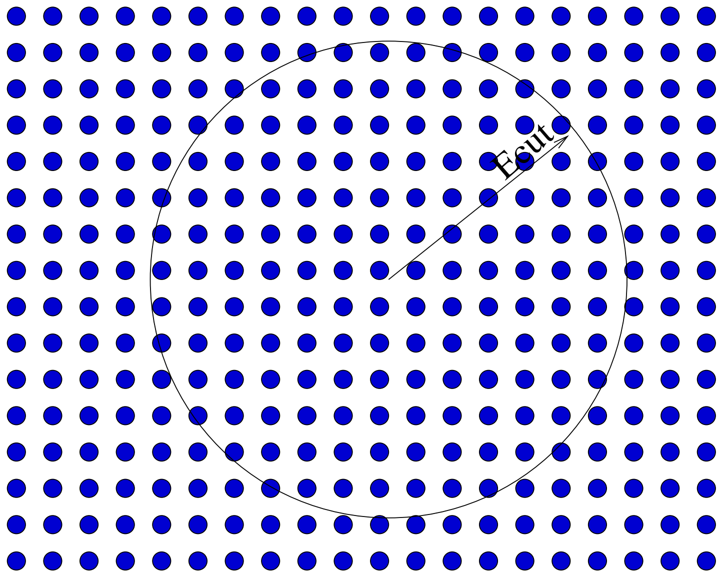


Figure 2.3: The planewave basis set is defined by the energy cutoff, E_{cut}

by the cutoff energy, E_{cut}:

$$\frac{\hbar^2|\mathbf{k}+\mathbf{G}|^2}{2m} \leq E_{cut}. \tag{2.19}$$

Hence, the basis set is defined by the maximum kinetic energy component it contains - see Figure 2.3. Physical quantities can be converged systematically by increasing E_{cut}.

While, in principle, all basis sets are equally accurate (so long as they are complete) there a many reasons why a particular set might be chosen. There can be a prejudice against the planewave basis — while the lack of a centre for a planewave removes the need to calculate Pulay corrections this lack of a centre can be looked at as a disadvantage. While a calculation performed using atom-centred orbitals as a basis (i.e. LCAO) produces wavefunctions with an immediate interpretation in terms of the mixing of those orbitals, the information contained within a wavefunction described in planewaves is less directly accessible. However, work performed by the author (in collaboration with Segall, Shah and Payne) [85, 86] shows that the straightforward planewave representation of wavefunctions allows the same chemical population analyses to be performed as in more traditional quantum chemistry approaches. And in this work, the evaluation of the correct ELNES matrix elements reveals the same site specific symmetry resolved information as do approaches using local orbitals, but in a more quantitative fashion — see Chapter 4.

2.2.5 The pseudopotential approximation

The electrons in an atom can be divided into two types — core electrons and valence electrons. The core electrons are tightly bound to the nucleus, while the valence electrons are more extended. A working definition for core electrons is that they are the ones which play no part in the interactions between atoms, while the valence electrons dictate most of the properties of the material. It is common to make the frozen core approximation. The core electrons are constrained not to differ from their free atomic nature when placed in the solid state environment. This reduces the number of electronic degrees of freedom in an all electron calculation. It is a very good approximation. A different, but physically related, approach is taken in the pseudopotential approximation [23, 41].

Since, in an all electron calculation, the valence electron wavefunctions must be orthogonal to the core wavefunctions they necessarily have strong oscillations in the region near the nucleus (see the all electron wavefunction in Figure 2.4). Given that a planewave basis set is to be used to describe the wavefunctions, these strong oscillations are undesirable — requiring many plane waves for an accurate description. Further, these oscillations are of very little consequence for the electronic structure in the solid, since they occur close to the nucleus, and interact little with the neighbouring atoms. In the pseudopotential approach only the valence electrons are explicitly considered, the effects of the core electrons being integrated within a new ionic potential. The valence wavefunctions need no longer be orthogonal to the core states, and so the orthogonality oscillations disappear, hence far fewer plane waves are required to describe the valence wavefunctions. This modified ionic potential, or pseudopotential, is constructed in the following way. An all electron DFT calculation is performed for an isolated atom. A core radius r_c is decided upon — chosen so that the core regions of neighbouring atoms will not overlap. The smaller the core radius the greater the transferability of the pseudopotential (i.e. the more chemical environments it will be valid in). The all electron valence wavefunctions are altered within r_c to remove the nodal structure. These new functions are the pseudowavefunctions. The Schrödinger equation is then inverted to find the potential that would produce these wavefunctions — this is the pseudopotential. See Figure 2.4 for a schematic representation of the potentials and wavefunctions (pseudo and true). It is usual to ensure that the charge within the core radius is the same for the pseudo and true wavefunctions — this is known as norm-conservation and simplifies many aspects of the implementation of pseudopotentials and also improves the energy range over which the pseudopotential is valid [39]. Since a pseudopotential must reproduce the correct phase shifts on scattering at the core, and

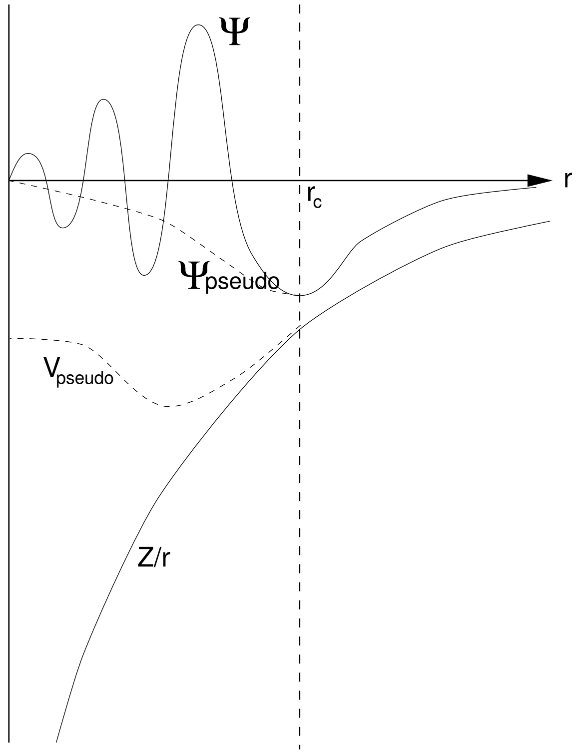


Figure 2.4: A schematic representation of the pseudopotential and pseudowavefunction. Not that the wavefunctions and potentials agree beyond the core radius, r_c and that the pseudowavefunctions are considerably smoother than the true wavefunctions within r_c — reducing the number of planewaves required.

these phase shifts will be different for different angular momentum states, in general a pseudopotential will be non-local, with different projectors for different angular momentum components. This non-locality introduces the requirement for the correction of the **k.p** and optical matrix elements (see Chapter 1.4 and Appendix C respectively).

In this work the non-local pseudopotentials are in the Kleinman-Bylander form [50]:

$$V = V_{loc} + \sum_{l,m} (V_l - V_{loc}) \hat{P}_{l,m}, \tag{2.20}$$

where the choice of V_{loc} is arbitrary. Using the approach of Lee [54], in some cases the potentials can be projector reduced. This consists of constructing two of the V_l to be very similar — both being eliminated by a single careful choice of V_{loc}. The method of Lee *et al* [55] has also been applied to optimise the pseudowavefunction with respect to planewave cutoff.

The neglect of the core electrons has specific consequences in this work. Since the ELNES results from transitions from those neglected core states, some corrections to the calculated transition matrix elements are required. This is discussed further in Chapter 4.

2.2.6 Minimisation of the energy functional

Equation 2.13 shows that the total energy functional can be minimised by solving a Schrödinger-like eigenvalue problem for a set of single particle wavefunctions $\{\phi_i : i = 1, N\}$. Clearly, standard matrix-diagonalisation techniques could be used, but it is highly inefficient in the case of a planewave approach. There are many more planewaves per atom than electronic states required. If M is the number of planewaves in a calculation, then typically $M/N > 100$. Hence, diagonalising a $M \times M$ matrix results in M eigenstates while only the lowest N are required. As an alternative, the total energy functional is minimised using the idea first proposed by Car and Parinello [17], the actual minimisation using a preconditioned conjugate gradient technique [73, 93]. These methods are implemented in the program CASTEP (CAmbridge Serial Total Energy Package), which has been used for all of the *ab initio* calculations in this work.

2.3 Bandstructure from the total energy method

As mentioned in Section 2.1, the cross section for the scattering of a fast electron by the electrons in a solid can be expressed in terms of single particle electronic states. The use of the Kohn-Sham eigenstates as these single particle states is justified and described in the following sections.

2.3.1 Kohn-Sham eigenstates and values

It is well known that the Kohn-Sham eigenstates resulting from the diagonalisation of the Hamiltonian in Equation 2.13 do not formally correspond to the single particle states required in the evaluation of the loss function [47]. It should be remembered that they were simply introduced as a tool for dealing with the kinetic energy term in the functional $F[n(\mathbf{r})]$; while is was convenient that the result was a mapping of the many-body problem onto one for single particles, those single particle states need not have any physical meaning. However, it is nearly universal to interpret these unoccupied Kohn-Sham eigenstates as the excited states required for our purposes, and with good reason. In many cases, where this interpretation is made, very good agreement is found between theory and experiment. This will be seen in this work both in Chapter 5 where calculated ELNES and Appendix C where calculated optical properties are compared to the relevant experimental data. It has been shown by Godby *et al* [36] that, for Silicon at least, the difference between the Kohn-Sham excitation energies and the correct quasiparticle energies can be summarised by a rigid shift of the conduction band upwards with respect to the valence band, the wavefunctions themselves being essentially unchanged. This rigid shift, or "scissor operator", is irrelevant for the calculation of ELNES at the current level of theory since the absolute energy of the threshold is not calculated. Reviews of the connection between DFT and excitation energies are presented by Godby [35] and Jones and Gunnarsson [47].

2.3.2 Performing a bandstructure calculation

Clearly, when a total energy calculation is performed there is no need to evaluate the unoccupied eigenstates — only the valence states contribute to the total energy. But, the calculation of the cross section for inelastic electron scattering requires the unoccupied states. So, in practice two calculations are performed. In the first the self consistent ground-state charge density is found, and then a subsequent calculation is performed using the fixed Hamiltonian resulting from that charge density. Many more eigenstates are found, at many more **k**-points, since accurate Brillouin zone integrations are required (see Chapter 3). In Appendix A the sensitivity of spectral calculations to the various parameters which must be chosen in the planewave pseudopotential approach is investigated.

2.3.3 The inelastic cross section in a crystal

Equation 2.4 gives an expression for the imaginary part of the dielectric function in terms of a sum over all final states. Since, in a sample, there will be of order 10^{20} relevant final states this is not the most useful form for the expression. As described in Section 2.2.3 the application of periodic boundary conditions allows such an expression to be converted to a sum over a finite number of bands and an infinite number of **k**-points within the first Brillouin zone. So,

$$\sum_{\sigma} |\langle\sigma|e^{-i\mathbf{q}.\mathbf{r}}|c\rangle|^2\delta(\hbar\omega - E_\sigma + E_c) = \sum_{n,\mathbf{k}} |\langle\Psi^n_{\mathbf{k}}|e^{-i\mathbf{q}.\mathbf{r}}|c\rangle|^2\delta(\hbar\omega - E_{n,\mathbf{k}} + E_c), \tag{2.21}$$

where $|\Psi^n_{\mathbf{k}}\rangle$ is the unoccupied single particle Kohn-Sham final state for band n and **k**-point **k**. Within the frozen core approximation $|c\rangle$ is a core state taken from an all electron atomic calculation — see Chapter 4 for more details on the matrix element evaluation. Since the core states are effectively non-dispersive, the above is simply the unoccupied density of states, weighted by the ELNES matrix element. Since the core state is taken to be on a single site, the resulting weighted density of states will be a symmetry projected (through the dipole or higher order operators) *local* density of states (LDOS). Chapter 3 describes the details of the Brillouin zone integration.

Symmetry is exploited in the Brillouin zone integrations. The irreducible wedge of a cubic system is 1/48 of the volume of the first Brillouin zone — if calculation is restricted to this volume then significant savings in computational effort can be made — see Section 3.8. It is important when the Brillouin zone integration in Equation 2.21 is performed to realise that the momentum transfer **q** breaks the symmetry, picking out a direction in the system. This must be taken into account by unfolding the matrix elements — calculating one for each symmetry related direction — and using this to weight the DOS contribution.

2.3.4 Lifetime and other broadening effects

The excited quasiparticles are not exact eigenstates, and the core hole has a finite lifetime. The result of both these effects is that some energy dependent Lorentzian broadening should be applied to the calculated spectra, for example as done by Weijs *et al* [100]. At the same time, the measurement process will result in some instrumental broadening, which can be summarised by a Gaussian of a certain width. In the calculated spectra presented in this work, the broadening is restricted to convolution with a Gaussian $e^{-E^2/2\gamma^2}$. Rather than attempting to exactly mimic the experimental spectra the width γ is chosen to remove distracting high resolution features from the spectra. Lifetime effects are currently ignored since they are only described empirically within present theoretical approaches.

Chapter 3

Brillouin Zone integrations

Most electronic properties of real systems in the solid state depend on sums over a computationally intractable number of electronic states. Section 2.2.3 discusses Bloch's theorem and its significance, noting the transformation of the electronic structure problem from one of calculating an infinite number of electronic states (for an infinite periodic system) to one of calculating for a finite number of bands at an infinite number of **k**-points. Most properties of a given system require some form of integration over these **k**-points. Straightforward sums over the states — e.g. for the total energy — are usually performed using a special point scheme (for example, that of Monkhorst and Pack [65]) for a modest number of **k**-points. However, energy resolved spectral properties (such as the ELNES calculated in this work) require more detailed integrations. This chapter describes the development of a new Brillouin zone integration scheme, designed specifically to perform well in spectral calculations, but which should also be useful in other situations, such as the evaluation of Green's Functions.

The requirements of an efficient Brillouin zone integration scheme are examined in Section 3.1, and some current methods are described in Section 3.2. The band crossing problem is examined in Section 3.3. The relative merits of interpolative and extrapolative approaches are discussed in Section 3.4. An outline of the proposed new piecewise quadratic extrapolative integration scheme is presented in Section 3.5. The generation of a local quadratic representation of a band from a single sample point using **k**.**p** perturbation theory is demonstrated in Section 3.6 and the theory is extended for the case of the non-local pseudopotentials used in this work. In Section 3.7 the problem of band crossing and "kissing" is examined, and a scheme to alleviate some of the effects is presented. Efficient division of the Brillouin zone into sub-cells is discussed in Section 3.8. A scheme for the conversion of the resulting piecewise quadratic representation analytically into a density of states (DOS) (following Methfessel *et al* [12, 62, 63]) is outlined in Section 3.9. Appendix B contains recipe for the quadratic analytic evaluation of the DOS that is hoped will be of use to future workers keen on using the method. Examples of the new technique are presented in Section 3.10, demonstrating the speed with which the fine structures in the DOS and other spectral properties can be obtained. Finally, in Section 3.11, the inclusion of dispersive matrix elements in the integration scheme is discussed.

3.1 What is required?

Examining experimental EELS spectra (see, for example, Figure 5.3) leads to the conclusion that any Brillouin zone integration should, at most, be required to produce features at a resolution around the 0.1eV level — a limit placed by the energy broadening effects introduced by core hole lifetimes and the finite lifetimes of the final quasiparticle states. It is not really physically meaningful to calculate EELS at a higher resolution. Further, current EELS spectrometers routinely achieve a resolution of only 0.3eV. Experimental investigations frequently concentrate on the shifts, distortions, and changes in weight in the spectral features. When one looks at calculated bandstructures (see, for example, Figure 3.1), it appears that the bands are essentially smoothly varying objects in **k**-space. It would therefore be hoped that only a very sparse sampling of the Brillouin zone would be required to obtain a good representation of the bands and hence the DOS. This proves to be untrue. Any sparse, high order sampling scheme has to confront the "band crossing problem" — see Section 3.3. This cripples any interpolative scheme at low sample densities. Past approaches have been to submit to this problem and simply calculate the DOS at the very high density of **k**-points required to alleviate its effects, and then smear the spectrum to the experimental resolution. This is clearly wasteful, but is offset by the fact that many of the electronic structure approaches used have relied on small basis sets (e.g. LMTO, LCAO, ...) and hence rapid matrix diagonalisation and small computational cost at each **k**-point. However, these approaches are necessarily restricted as to how far they can predict spectra into the conduction band, and they are not intrinsically as accurate as the more expensive plane-wave pseudopotential methods. It is clear that if resources can be saved in the Brillouin zone integrations then the savings can be "spent" on using plane-waves, larger supercells, more elaborate theory or even all electron planewave calculations (see Section 4.6.4 and Appendix C). The challenge is to develop a Brillouin zone integration

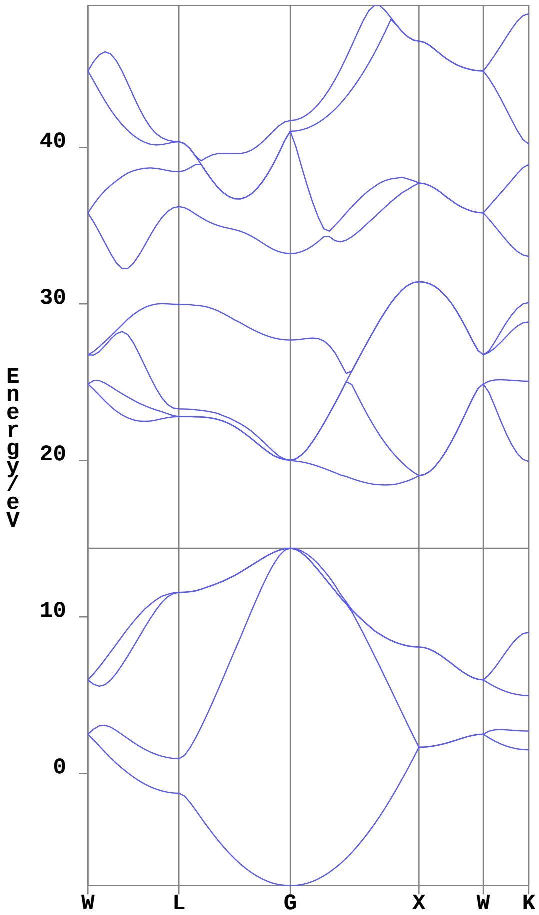


Figure 3.1: The bandstructure of Diamond, calculated in a primitive unit cell containing two atoms.

scheme which exploits the essential smoothness of the bands, while somehow avoiding the band crossing problem.

3.2 Current methods

Many Brillouin zone integration techniques have been developed over recent years. The most common and interesting ones are described below, with emphasis on their performance for predicting spectral properties (i.e. special point schemes will be ignored here).

3.2.1 Gaussian broadening

This is the simplest approach to Brillouin zone spectral integrations. **K**-points are evenly distributed throughout the Brillouin zone (either regularly or randomly as in a Monte Carlo method), and the DOS is formed according to:

$$D(E) \propto \sum_i e^{-(E-E_i)^2/2\sigma^2} \tag{3.1}$$

This method actually converges at van Hove singularities faster than the linear method described below. However, this only applies at the limit of full convergence, which is rarely required or achieved. It is, in practice, less accurate than the linear method for most purposes, but it is the method of choice for work that merely requires a rough idea of where the electronic states lie, at low resolution (e.g. looking for gap states in wide bandgap systems). It clearly does not suffer from the band crossing problem.

3.2.2 Linear interpolative tetrahedron method

The linear tetrahedron method is the staple Brillouin zone integration technique for the electronic structure community. Indeed, as recently as 1994 Blöchl *et al* [10] published an "improved" method, tidying up the tetrahedra generation procedure and introducing some higher order corrections. It is a relatively straightforward approach, and hence very attractive. The scheme works by dividing the irreducible wedge of the Brillouin zone into approximately equally volumed tetrahedra, then by calculating the band energies at each apex a linear representation of the band is constructed within the tetrahedron. Contributions to the DOS can then be evaluated analytically. However, this method suffers from the band crossing problem, and the poor representation of van Hove singularities due to the absence of second order band information.

3.2.3 Quadratic interpolative tetrahedron method

A series of papers by Methfessel *et al* [12, 62, 63] demonstrate the superior convergence properties of a method based on a piecewise quadratic representation of the Brillouin zone. This method converges more rapidly than both the Gaussian broadening and linear tetrahedron methods due to the more accurate treatment the van Hove singularities which occur when $|\nabla E(\mathbf{k})| = 0$. At these points $E(\mathbf{k})$ must be known to second order in $\mathbf{k}$ at least. However, this methodology does not appear to have been widely applied by the community. A probable reason for this it is particularly susceptible to the band crossing problem becuase it is a higher order interpolative scheme. And so, while convergence for a single band is impressive, the method becomes relatively expensive in real systems and the extra effort is not justified. It is, however, one of the building blocks for the method developed below.

3.2.4 Free energy minimisation and Gaussian processes

Several alternatives to the Brillouin zone integration scheme to be presented below were developed in the course of this work, in collaboration with Gibbs. The first was based on a mixture model approach, using free energy minimisation techniques, the second on free energy minimisation and Gaussian processes and the third on the theory of random matrices. They are described in detail in the thesis of Gibbs [34].

3.3 The band crossing problem

The so called "band crossing problem" is a well known impediment to the rapid convergence of interpolative methods. The simple one dimensional example shown in Figure 3.2 illustrates the changes that occur in the DOS given different band allocations for the interpolation. At a general **k**-point it is more likely that the bands do not actually cross — rather they can be described as "kissing", getting very close at a given point, but both the upper and lower bands remaining individually continuous. It is also unclear as to whether any "allocation" of bands will lead to the correct topology of the bandstructure. The bands cannot, in general, be uniquely identified because their topology is more complex than that of a set of interpenetrating sheets in four dimensions.

3.4 Extrapolation vs. interpolation

The vast majority of Brillouin zone integration schemes use an interpolative approach. However, an alternative is to extrapolate from a single point within the sub-cell of integration. This removes the need to allocate bands, circumventing the band crossing problem. This approach has been used by Müller *et al* [69] who developed a linear extrapolative method. They demonstrated the improvement over conventional schemes by calculating a free electron DOS using both a linear interpolative and extrapolative scheme. Using

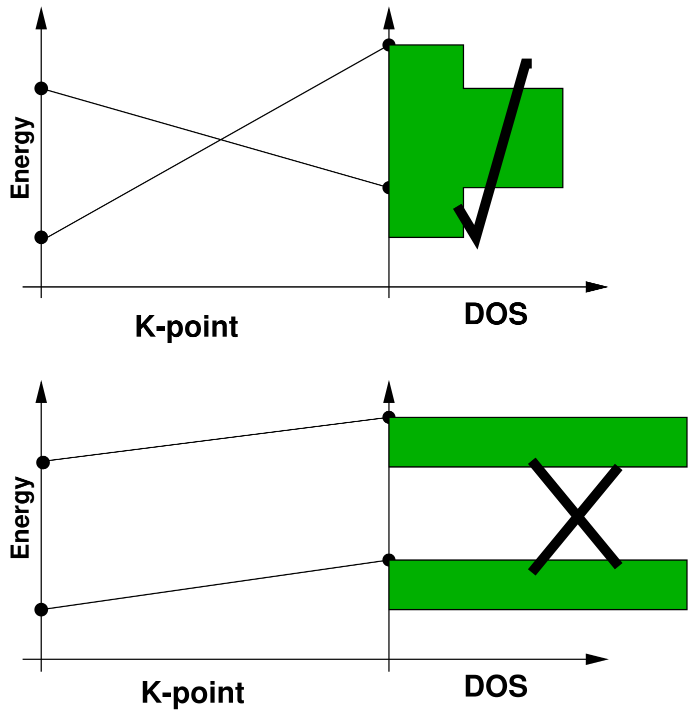


Figure 3.2: A simplified one dimensional example of the band crossing problem. An incorrect allocation of bands can cause dramatic changes in the calculated DOS. It is possible that the two bands "kiss" rather than cross, but it is clear that the top allocation will produce the more accurate DOS even in either case.

256 tetrahedra the extrapolative scheme has converged, while the interpolative one is a mess of spurious features (compare this to the calculation in Section 3.10.1 using a single integration cell). Any extrapolative method requires the functional form of a given band about a given **k**-point and for computational efficiency it is important that this can be found from information available at the single **k**-point — see Section 3.6.

3.5 An outline for a new integration scheme

To summarise so far, the prevalent scheme for spectral Brillouin zone integrations is currently the interpolative linear tetrahedron method. The low order interpolation fails to describe van Hove singularities efficiently, but is relatively simple. Since high **k**-point sampling densities are required for convergence the band crossing problem is not extreme. Other schemes have been presented in the literature, but all have apparently failed to make a great impact. The second order **k**.**p** approach of Wang *et al* [99] did not correctly implement Lowdin perturbation theory (as clarified by Loehr [57]). Also, it was thought of as an interpolative scheme, the DOS was evaluated by resampling and using a linear method, and as a result ended up using an immense number of sampling points (1357 in the irreducible wedge). Müller *et al* [69] used an extrapolative first order **k**.**p** scheme, clearly recognising the band crossing problem. It appears a sensible approach, and should have become more widely used. It is possible that the restriction of **k**.**p** theory to local potentials (with the emergence of non-local pseudopotential methods), and the problem of incomplete local basis sets, noted by Boykin [14], may have reduced its use. The developments discussed below, and the use of planewaves remove these barriers. Gradients can now be calculated very accurately within **k**.**p** perturbation theory (see Table 3.1) and so if a linear method is required this appears the most attractive approach. The quadratic interpolative method of Methfessel *et al* [12, 62, 63] was discussed in Section 3.2.3. In the light of the preceding arguments, the essential features of the new method are outlined as follows:

Extrapolative Extrapolating away from a single point alleviates the worst of the band crossing/"kissing" problems.

Piecewise quadratic To correctly treat the van Hove singularities, at least a second order expansion is required.

k.p perturbation theory Allows access to band information around a selected **k**-point, and hence gives the required quadratic expansion of the bands.

Band "kissing" Expansion about a "kissing" point reduces the radius of convergence. This is dealt with, to some extent, using Lowdin perturbation theory [49, 57].

Brillouin zone division The Brillouin zone is divided into arbitrary polyhedra of approximately equal volumes. The full symmetry of the system can be utilised.

Analytic quadratic The work of Methfessel *et al* [12, 62, 63] permits a direct conversion of the piecewise quadratic representation to a DOS contribution, avoiding the need for resampling.

Many of the individual elements of this proposed method are not new. However, to the author's knowledge they have never previously been brought together as a single technique. Some components (the treatment of "kissings", non-local pseudopotentials and the

Brillouin zone division) are novel developments. The technical details of the method are now described.

3.6 Obtaining the quadratic representation - k.p perturbation theory

As discussed in Section 3.4, if we are to employ an extrapolative scheme we need a technique for evaluating the quadratic representation of a band within a given cell from information available at a single **k**-point in that cell. **k.p** perturbation theory [49] offers just this.

3.6.1 An introduction to k.p perturbation theory

The central idea behind **k.p** perturbation theory lies at the heart of any attempt to understand the bandstructure of a periodic system. As a result it has been taught for many years to undergraduate students of solid state physics [2]. This idea is that knowledge of eigenstates and eigenvalues at any given point in the Brillouin zone implies knowledge of the bandstructure in the region surrounding that point (this concept resurfaces in modern day approaches to efficient electronic structure calculations [88]). Implicit in this is the concept that quantities (energies and matrix elements) vary smoothly throughout the Brillouin zone. This is crucial to any attempt to calculate properties that require integrations over the Brillouin zone.

k.p perturbation theory has a long history, stretching back to the early days of the theory of the electronic structure of the solid state [87]. It was initially used partly as a pedagogical instrument — for learning about the form of bands before calculated bandstructures were widely available. But by far the most important use of the method was as a tool in the parameterisation of bands using, as data, experimental observables such as effective masses and band gaps. The hope was to learn about as much of the bandstructure as possible from the restricted experimental data available, and and also to produce a compact representation of the bandstructure [18, 49]. However, recently **k.p** perturbation theory has not been so widely used. It has found a home in transport theory of heterostructures (it enables effective masses to be calculated) and Robertson *et al* [83] recently applied it to total energy pseudopotential calculations.

Possibly it is now time to return to the original spirit in which the theory was used. The output of the relatively expensive plane-wave pseudopotential calculations can be thought of as restricted experimental data — from which the maximum information about the band-structure is to be extracted, and a compact representation is to be produced. Kane [49] may have predicted this when he wrote:

> "... Perhaps a judicious combination of the k.p and pseudopotential methods will nevertheless provide the ultimate in accuracy. ..."

3.6.2 k.p approach to quadratic representation of a band

The Taylor expansion of a field in three dimensions around a point $\mathbf{k}_o$ (in this case — the band energy over a Brillouin zone) is:

$$E_n(\mathbf{k}_o + \mathbf{q}) = E_n(\mathbf{k}_o) + \mathbf{G}_n^T \mathbf{q} + \mathbf{q}^T \mathbf{B}_n \mathbf{q} + \cdots, \tag{3.2}$$

where $\mathbf{G}$ and $\mathbf{B}$ are given by:

$$G_{n,i} = \left.\frac{\partial E_n(\mathbf{k})}{\partial k_i}\right|_{\mathbf{k}_o} \quad B_{n,ij} = \left.\frac{1}{2}\frac{\partial^2 E_n(\mathbf{k})}{\partial k_i \partial k_j}\right|_{\mathbf{k}_o} \tag{3.3}$$

The perturbation expansion

Given that the Hamiltonian is cell periodic, the single particle wavefunctions $|\Psi^n_{\mathbf{k}}\rangle$, the eigenstates of $\hat{H}$, can be written as Bloch functions:

$$|\Psi^n_{\mathbf{k}}\rangle = e^{i\mathbf{k}.\mathbf{r}} |u^n_{\mathbf{k}}\rangle, \tag{3.4}$$

where the $|u^n_{\mathbf{k}}\rangle$ are cell periodic and are eigenstates of the Hamiltonian:

$$\hat{H}(\mathbf{k}) = e^{-i\mathbf{k}.\mathbf{r}} \hat{H} e^{i\mathbf{k}.\mathbf{r}}, \tag{3.5}$$

with eigenvalues $E_n(\mathbf{k})$. The perturbation potential, due to moving from $\mathbf{k}$ to $\mathbf{k} + \delta\mathbf{k}$, can be expressed as:

$$\begin{aligned} \Delta\hat{V}(\mathbf{k}) &= e^{-i\mathbf{k}.\mathbf{r}} \left(e^{-i\delta\mathbf{k}.\mathbf{r}} \hat{H} e^{i\delta\mathbf{k}.\mathbf{r}} - \hat{H} \right) e^{i\mathbf{k}.\mathbf{r}} \\ &= e^{-i\mathbf{k}.\mathbf{r}} \left((1 - i\delta\mathbf{k}.\mathbf{r} + \cdots) \hat{H} (1 + i\delta\mathbf{k}.\mathbf{r} + \cdots) - \hat{H} \right) e^{i\mathbf{k}.\mathbf{r}} \\ &= e^{-i\mathbf{k}.\mathbf{r}} i\delta\mathbf{k}.[\hat{H}, \hat{\mathbf{r}}] e^{i\mathbf{k}.\mathbf{r}} + O(\delta k^2) \end{aligned} \tag{3.6}$$

Hence, $\Delta\hat{V} = i\delta\mathbf{k}.[\hat{H},\hat{\mathbf{r}}]$. Writing $P^i_{mn} = \langle\Psi^m_{\mathbf{k}}|[\hat{H},\hat{r}_i]|\Psi^n_{\mathbf{k}}\rangle$, standard non-degenerate perturbation theory gives an expression for $|u^n_{\mathbf{k}+\delta\mathbf{k}}\rangle$:

$$|u^n_{\mathbf{k}+\delta\mathbf{k}}\rangle = |u^n_{\mathbf{k}}\rangle + i\sum_{m\neq n}\frac{\delta\mathbf{k}.\mathbf{P}_{mn}|u^m_{\mathbf{k}}\rangle}{E_n(\mathbf{k}) - E_m(\mathbf{k})} + \cdots \tag{3.7}$$

Now the gradient and curvatures of the band energies can be evaluated, noting that $E_n(\mathbf{k}) = \langle\Psi^n_{\mathbf{k}}|\hat{H}|\Psi^n_{\mathbf{k}}\rangle$. First, the gradient is given by:

$$\frac{\partial E_n(\mathbf{k})}{\partial k_i} = \langle\Psi^n_{\mathbf{k}}|\hat{H}|\frac{\partial\Psi^n_{\mathbf{k}}}{\partial k_i}\rangle + \langle\frac{\partial\Psi^n_{\mathbf{k}}}{\partial k_i}|\hat{H}|\Psi^n_{\mathbf{k}}\rangle, \tag{3.8}$$

where the differential, $|\frac{\partial\Psi^n_{\mathbf{k}}}{\partial k_i}\rangle$, is evaluated as:

$$|\frac{\partial\Psi^n_{\mathbf{k}}}{\partial k_i}\rangle = \frac{\partial(e^{i\mathbf{k}.\mathbf{r}}|u^n_{\mathbf{k}}\rangle)}{\partial k_i} = ir_i|\Psi^n_{\mathbf{k}}\rangle + e^{i\mathbf{k}.\mathbf{r}}|\frac{\partial u^n_{\mathbf{k}}}{\partial k_i}\rangle, \tag{3.9}$$

where $\{\mathbf{r}\}_i = r_i$, the components of the position operator. Using Equation 3.7,

$$e^{i\mathbf{k}.\mathbf{r}}|\frac{\partial u^n_{\mathbf{k}}}{\partial k_i}\rangle = e^{i\mathbf{k}.\mathbf{r}}\lim_{\delta k_i\to 0}\frac{|u^n_{\mathbf{k}+\delta\mathbf{k_i}}\rangle - |u^n_{\mathbf{k}}\rangle}{\delta k_i} = i\sum_{m\neq n}\frac{|\Psi^m_{\mathbf{k}}\rangle P^i_{mn}}{E_n(\mathbf{k}) - E_m(\mathbf{k})}, \tag{3.10}$$

and hence,

$$|\frac{\partial\Psi^n_{\mathbf{k}}}{\partial k_i}\rangle = ir_i|\Psi^n_{\mathbf{k}}\rangle + i\sum_{m\neq n}\frac{|\Psi^m_{\mathbf{k}}\rangle P^i_{mn}}{E_n(\mathbf{k}) - E_m(\mathbf{k})} \tag{3.11}$$

Making use of $\langle\Psi^n_{\mathbf{k}}|\Psi^m_{\mathbf{k}}\rangle = \delta_{nm}$,

$$\langle\Psi^n_{\mathbf{k}}|\hat{H}|\tfrac{\partial\Psi^n_{\mathbf{k}}}{\partial k_i}\rangle = i\langle\Psi^n_{\mathbf{k}}|\hat{H}\hat{r}_i|\Psi^n_{\mathbf{k}}\rangle \quad \langle\tfrac{\partial\Psi^n_{\mathbf{k}}}{\partial k_i}|\hat{H}|\Psi^n_{\mathbf{k}}\rangle = -i\langle\Psi^n_{\mathbf{k}}|\hat{r}_i\hat{H}|\Psi^n_{\mathbf{k}}\rangle, \tag{3.12}$$

and so,

$$\frac{\partial E_n(\mathbf{k})}{\partial k_i} = iP^i_{nn}. \tag{3.13}$$

The curvature of the bands are now evaluated. Differentiating the energy a second time, with respect to k_j:

$$\frac{\partial^2 E_n(\mathbf{k})}{\partial k_i\partial k_j} = i\langle\Psi^n_{\mathbf{k}}|[\hat{H},\hat{r}_i]|\frac{\partial\Psi^n_{\mathbf{k}}}{\partial k_j}\rangle + i\langle\frac{\partial\Psi^n_{\mathbf{k}}}{\partial k_j}|[\hat{H},\hat{r}_i]|\Psi^n_{\mathbf{k}}\rangle \tag{3.14}$$

In the same way as above, using Equation 3.11:

$$
\begin{aligned}
i\langle\Psi^n_{\mathbf{k}}|[\hat{H},\hat{r}_i]|\tfrac{\partial\Psi^n_{\mathbf{k}}}{\partial k_j}\rangle &= i\langle\Psi^n_{\mathbf{k}}|[\hat{H},\hat{r}_i]\left\{ir_i|\Psi^n_{\mathbf{k}}\rangle + i\sum_{m\neq n}\frac{|\Psi^m_{\mathbf{k}}\rangle P^j_{mn}}{E_n(\mathbf{k})-E_m(\mathbf{k})}\right\} \\
&= -\langle\Psi^n_{\mathbf{k}}|[\hat{H},\hat{r}_i]\hat{r}_j|\Psi^n_{\mathbf{k}}\rangle - \sum_{m\neq n}\frac{P^i_{nm}P^j_{mn}}{E_n(\mathbf{k})-E_m(\mathbf{k})} \\
i\langle\tfrac{\partial\Psi^n_{\mathbf{k}}}{\partial k_j}|[\hat{H},\hat{r}_i]|\Psi^n_{\mathbf{k}}\rangle &= \langle\Psi^n_{\mathbf{k}}|\hat{r}_j[\hat{H},\hat{r}_i]|\Psi^n_{\mathbf{k}}\rangle - \sum_{m\neq n}\frac{P^j_{nm}P^i_{mn}}{E_n(\mathbf{k})-E_m(\mathbf{k})}
\end{aligned}
\tag{3.15}
$$

And so we have the gradient and curvature:

$$
\begin{aligned}
\tfrac{\partial E_n(\mathbf{k})}{\partial k_i} &= iP^i_{nn} \\
\tfrac{\partial^2 E_n(\mathbf{k})}{\partial k_i\partial k_j} &= -\langle\Psi^n_{\mathbf{k}}|\left[[\hat{H},\hat{r}_i],\hat{r}_j\right]|\Psi^n_{\mathbf{k}}\rangle - \sum_{m\neq n}\frac{P^i_{nm}P^j_{mn}+P^j_{nm}P^i_{mn}}{E_n(\mathbf{k})-E_m(\mathbf{k})}
\end{aligned}
\tag{3.16}
$$

This is a general result, and is in agreement with other derivations [14, 49].

3.6.3 The commutators

An expression for $\hat{H}$ in the case of DFT is:

$$
\hat{H} = -\frac{\hbar^2\nabla^2}{2m} + \hat{V},
\tag{3.17}
$$

where $\hat{V}$ is some potential. Assuming for the moment that $\left[\hat{V},\hat{\mathbf{r}}\right]=0$, which is the case for local potentials, then we need to evaluate the commutators which occur in the expressions for the **k**.**p** expansion above:

$$
\begin{aligned}
\left[\hat{H},\hat{r}_i\right] &= \left[-\tfrac{\hbar^2\nabla^2}{2m},\hat{r}_i\right] = -i\tfrac{\hbar}{m}\hat{p}_i \\
\left[[\hat{H},\hat{r}_i],\hat{r}_j\right] &= \left[[-\tfrac{\hbar^2\nabla^2}{2m},\hat{r}_i],\hat{r}_j\right] = -\tfrac{\hbar^2}{m}\delta_{ij}
\end{aligned}
\tag{3.18}
$$

The operator $\hat{\mathbf{p}}$ is the momentum operator, from which the name "**k**.**p**" is derived. Using these expressions for the commutators, the **k**.**p** expansion agrees with the standard expressions [2]. There are however two situations in which the commutators are not so straightforward. One is in the case of incomplete of basis sets which has been examined by Boykin [14]. The problems in this significant case explain the problems of using **k**.**p** perturbation theory encountered in earlier work. But this does not appear to be a problem for the planewave basis employed in this work (see the accuracy of the tests in Section

3.6.5, in which a far from complete basis has been used and any effects due to incompleteness are ignored). It seems likely that this is due to the delocalised nature of the planewaves — recall the absence of the Pulay corrections to forces in a planewave basis[1]. However, this result remains to be proved rigorously. The other source of complication is more obvious, and concerns the use of non-local pseudopotentials. Since these potentials are used in this work it is important to take this into account. The effect of non-local pseudopotentials on the **k.p** expansion does not appear to have been treated, although the first order correction is closely connected to the correction required for optical matrix elements (see Chapter C) [79]. This is an important omission given the prevalence of such potentials in modern electronic structure calculations. The pseudopotential can clearly be divided into local and non-local components:

$$\hat{V} = \hat{V}_l + \hat{V}_{nl} \tag{3.19}$$

And so the full expressions for the commutators will be:

$$\begin{aligned} [\hat{H}, \hat{r}_i] &= -i\frac{\hbar}{m}\hat{p}_i + [\hat{V}_{nl}, \hat{r}_i] \\ \Big[[\hat{H}, \hat{r}_i], \hat{r}_j\Big] &= -\frac{\hbar^2}{m}\delta_{ij} + \Big[[\hat{V}_{nl}, \hat{r}_i], \hat{r}_j\Big] \end{aligned} \tag{3.20}$$

To demonstrate the fact that the non-local contribution is non-zero, consider the case of the following non-local pseudopotential:

$$\hat{V}_{nl} = \sum_l |l\rangle V_l \langle l|, \tag{3.21}$$

and hence,

$$\langle \Psi^n_{\mathbf{k}} | [\hat{V}_{nl}, \hat{r}_i] | \Psi^m_{\mathbf{k}} \rangle = \sum_l \left\{ \langle \Psi^n_{\mathbf{k}} | l \rangle V_l \langle l | \hat{r}_i | \Psi^m_{\mathbf{k}} \rangle - \langle \Psi^n_{\mathbf{k}} | \hat{r}_i | l \rangle V_l \langle l | \Psi^m_{\mathbf{k}} \rangle \right\}, \tag{3.22}$$

which is clearly non-zero in general.

3.6.4 Details of the matrix element evaluation

The matrix elements of the form $\langle \Psi^n_{\mathbf{k}} | \hat{p}_i | \Psi^m_{\mathbf{k}} \rangle$ are evaluated in reciprocal space (the plane waves making up the basis are eigenstates of the the momentum operator, and hence the

[1]They disappear. The force is given by $F_{x_i} = \frac{\partial}{\partial x_i} \langle \Psi | \hat{H} | \Psi \rangle$ which is simply $\langle \Psi | \frac{\partial \hat{H}}{\partial x_i} | \Psi \rangle$ in a planewave basis — the Hellmann-Feynman theorem [30] — there is no Pulay correction.

operator $\hat{p}_i$ is diagonal). The terms $\langle l|\hat{r}_i|\Psi_k^n\rangle$ and $\langle l|\hat{r}_i\hat{r}_j|\Psi_k^n\rangle$ required for the contributions from the non-local pseudopotentials may also be evaluated in reciprocal space in the following way. Consider the matrix element $\langle l|e^{i\mathbf{q}.\hat{\mathbf{r}}}|\Psi_{\mathbf{k}}^n\rangle$, for small $\mathbf{q}$:

$$\langle l|e^{i\mathbf{q}.\hat{\mathbf{r}}}|\Psi_{\mathbf{k}}^n\rangle = \langle l|\Psi_{\mathbf{k}}^n\rangle + i\langle l|\mathbf{q}.\hat{\mathbf{r}}|\Psi_{\mathbf{k}}^n\rangle - \frac{1}{2}\langle l|\,|\mathbf{q}.\hat{\mathbf{r}}|^2\,|\Psi_{\mathbf{k}}^n\rangle + O(q^3) \tag{3.23}$$

Now, let $\mathbf{q} = (q, 0, 0)$, then:

$$\langle l|e^{iq\hat{r}_1}|\Psi_{\mathbf{k}}^n\rangle - \langle l|e^{-iq\hat{r}_1}|\Psi_{\mathbf{k}}^n\rangle = 2iq\langle l|\hat{r}_1|\Psi_{\mathbf{k}}^n\rangle + O(q^3), \tag{3.24}$$

and so on. In a similar way, the matrix elements required for the double commutator may be found. Again, let $\mathbf{q} = (q, 0, 0)$, then:

$$\langle l|e^{iq\hat{r}_1}|\Psi_{\mathbf{k}}^n\rangle + \langle l|e^{-iq\hat{r}_1}|\Psi_{\mathbf{k}}^n\rangle - 2\langle l|\Psi_{\mathbf{k}}^n\rangle = -q^2\langle l|\hat{r}_1^2|\Psi_{\mathbf{k}}^n\rangle + O(q^4), \tag{3.25}$$

and so on. The cross terms (i.e. $\hat{r}_1\hat{r}_2$ etc.) may be generated as follows. Let $\mathbf{q} = (q, q, 0)$, then:

$$\langle l|e^{iq(\hat{r}_1+\hat{r}_2)}|\Psi_{\mathbf{k}}^n\rangle+\langle l|e^{-iq(\hat{r}_1+\hat{r}_2)}|\Psi_{\mathbf{k}}^n\rangle-2\langle l|\Psi_{\mathbf{k}}^n\rangle = -q^2(\langle l|\hat{r}_1^2|\Psi_{\mathbf{k}}^n\rangle+\langle l|\hat{r}_2^2|\Psi_{\mathbf{k}}^n\rangle+2\langle l|\hat{r}_1\hat{r}_2|\Psi_{\mathbf{k}}^n\rangle)+O(q^4), \tag{3.26}$$

and so, using Equation 3.25, $\langle l|\hat{r}_1\hat{r}_2|\Psi_{\mathbf{k}}^n\rangle$ etc. can be extracted. In effect this is a low order limit of the method used to parameterise the ELNES matrix elements when looking for non-dipole effects which is described in Section 4.2. Given the electronic structure method used must already have routines for finding $\langle l|\Psi_{\mathbf{k}}^n\rangle$, the advantage of this method is that evaluating $\langle l|e^{i\mathbf{q}.\hat{\mathbf{r}}}|\Psi_{\mathbf{k}}^n\rangle$ simply involves small shifts in the grid in reciprocal space on which $|l\rangle$ is calculated. This avoids tiresome evaluation of new harmonic projectors, and is readily extended to higher order. The size of q is chosen to be small so that the $O(q^3)$ and $O(q^4)$ terms are negligible, but large enough to avoid errors due to precision limitations.

3.6.5 Tests of the k.p method

To demonstrate the importance of the modifications suggested above, and to validate the technique, a test case is presented. A model system is considered, which is a simple cubic Carbon structure with lattice parameter of 2Å, a non-local pseudopotential with s,p and d projectors, and a low planewave cutoff of 400eV (600eV is recommended for converged total energies with this particular pseudopotential). This cutoff gives a small basis set comprising

Gradients	Band #1		Band #2	
	Mag.	Direction	Mag.	Direction
No Corrections	8.2603	(0.2708,0.5395,0.7972)	13.4168	(0.1236,0.2115,-0.9695)
$+1^{st}$ order correction	7.1653	(0.2685,0.5366,0.8000)	11.7139	(0.1274,0.2035,-0.9708)
Numerical diff.	7.1653	(0.2685,0.5366,0.8000)	11.7139	(0.1274,0.2035,-0.9708)

Table 3.1: Band gradient calculated using **k.p** and numerical differentiation. The magnitude and direction are presented separately. Taking the non-local correction into account the 1^{st} order **k** perturbation expansion is accurate. The units are eVÅ.

147 plane waves. In Tables 3.1 and 3.2 the calculated gradients and curvatures respectively are compared with those evaluated numerically. The non-local contributions are clearly essential for high accuracy calculations, as is (see Figure 3.3) the inclusion of all bands in the perturbation sum. This should, however, be kept in perspective. For the purposes of the Brillouin zone integration scheme presented here, few bands need be included and the 2^{nd} order correction can be safely neglected — see Section 3.10.5. Should higher accuracy ever be required it might be appropriate to consider "padding" the perturbation sum with plane waves (since these will be the approximate eigenstates at high energies). The smooth and steady change in curvature when more than 25 bands are included in the perturbation sum shown in Figure 3.3 hints that such an approximation is sensible.

3.7 Band "kissing" - and how to get around it

As mentioned in Section 3.3, one of the remaining problems of Brillouin zone integrations is what to do when bands cross, or get very close to each other — "kiss". For interpolative approaches, this manifests itself as a problem of band ordering. Although the band ordering problem is eliminated by adopting an extrapolative method, problems arise when sampling points land at or near "kissing" points (i.e. bands do not quite cross, there being small interactions between the approaching bands — see Figure 3.4). At these points the curvature can be extreme and the radius of convergence of the Taylor expansion is much reduced. Thus the piecewise quadratic method fails — and the extrapolation sends states to a wide range of incorrect energies (often leading to the closure of band gaps). Clearly, if there is to be any hope of a sparse sampling high order extrapolative approach to Brillouin zone integrations the problems associated with the "kissings" must be addressed.

There are many complex ways in which one could imagine these three dimensional sheets

Curvatures	Band #1		Band #2	
	Mag.	Direction	Mag.	Direction
No Corrections	5.7745	(-0.0004,-0.0387,0.9993)	-6.3724	(-0.0752,-0.5922,0.8023)
	6.6223	(-0.0822,-0.9959,-0.0386	4.7737	(-0.6819,0.6176,0.3919)
	6.8774	(0.9966,-0.0821,-0.0028)	5.3288	(0.7275,0.5177,0.4502)
$+1^{st}$ order	5.5963	(0.1963,0.4039,0.8935)	-4.7483	(-0.0483,-0.7201,0.6922)
	6.4416	(-0.3010,-0.8424,0.4469)	4.1147	(-0.8066,0.4369,0.3982)
	6.5306	(0.9332,-0.3567,-0.0438)	5.1480	(0.5891,0.5391,0.6019)
$+2^{nd}$ order	5.2357	(0.1888,0.3907,0.9010)	-4.3987	(-0.0432,-0.7397,0.6715)
	6.0304	(-0.2949,-0.8525,0.4315)	4.0840	(0.8498,-0.3807,-0.3646)
	6.1192	(0.9367,-0.3472,-0.0458)	5.1407	(-0.5253,-0.5549,-0.6451)
Numerical diff.	5.235(6)	(0.1888,0.3907,0.9010)	-4.399(0)	(-0.0432,-0.7397,0.6715)
	6.030(4)	(-0.2951,-0.8525,0.4315)	4.083(8)	(0.8498,-0.3806,-0.3646)
	6.119(2)	(0.9366,-0.3473,-0.0457)	5.141(1)	(-0.5253,-0.5549,-0.6451)

Table 3.2: Band curvature calculated using **k.p** and numerical differentiation. The magnitude and direction for each principal axis are presented separately. All 147 bands were included in the perturbation sum. The importance of both 1^{st} and 2^{nd} order corrections is clearly demonstrated. The resulting **k.p** curvatures are accurate (the simple numerical differentiation proved harder to converge). The units are in eVÅ^2, and the curvature for a free electron in the same units is 7.6201 eVÅ^2

in four dimensions — the bands — could cross or "kiss" each other. Given such a situation, it may appear to be an intractable problem to develop any scheme for reducing their effect on the Brillouin zone integration. One approach is, however, to simply to have a look at some representative bandstructures and see whether any generalisations can be made. As mentioned above, the "kissing" problem is revealed at points with extreme curvatures — the "kissing" bands get close to each other but rapidly turn away at the "kissing" point. The bandstructures are examined by looking for those bands at any particular **k**-point that exhibit extreme curvatures. Those bands which both curve strongly and lie close to each other in energy might be expected to be "kissing". Following this procedure it becomes clear that a vast majority of the "kissings" that occur are of the simplest kind — shown in Figure 3.4. These involve pairs of bands crossing as simple sheets. The upper one (in energy) curves strongly upwards, the lower downwards. An example is presented numerically in Table 3.3, in much the same form as the original data that lead to this discovery. The scheme presented below is designed to deal with these simplest "kissings". However, the situation is indeed more complex. In many cases nearly degenerate triplets of bands with extreme curvatures are found — which cannot be treated as a combination

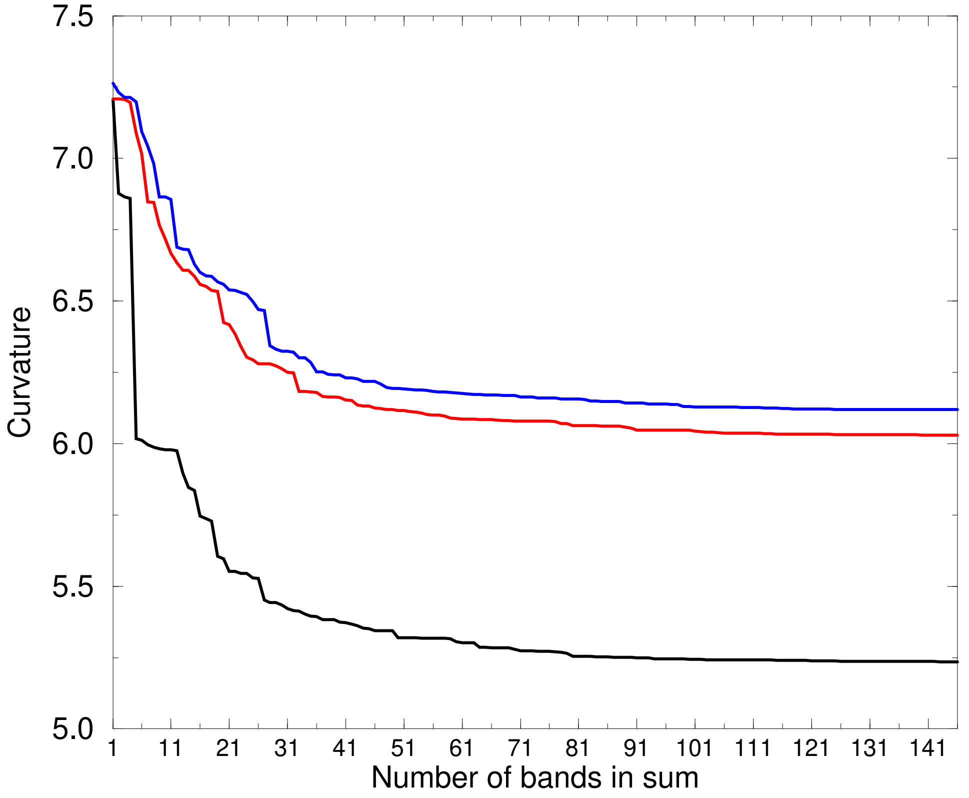


Figure 3.3: Convergence of the fully corrected **k.p** curvature with included bands for Band 1, in the directions of the three principal axes. Note that while convergence is initially rapid, the higher bands do make a significant contribution to the precision obtained.

of the simple "kissings". Dealing with these more complex cases is an area of current research.

3.7.1 The scheme

This method for Brillouin zone integration is designed for efficient calculation of experimentally observed spectral properties. Experimental and lifetime broadenings place a limit on the resolution required. It is clear that many of the band "kissings" observed will produce spectral features at a resolution too high to have any experimental consequence. As a

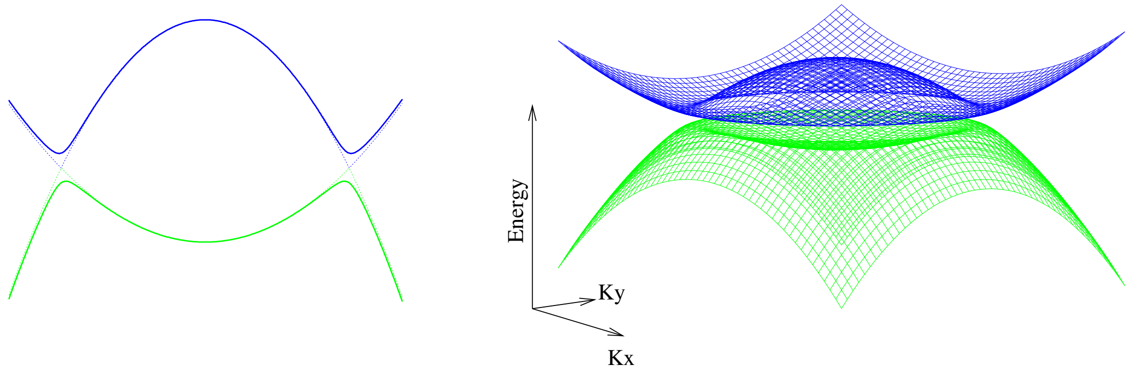


Figure 3.4: An example of one kind of frequently encountered band "kissing", for which there is some hope of accounting for in the extrapolative method. This example consists of two quadratically dispersive bands, intersecting each other and a repulsion occurring along the curve of intersection. To the left is a two dimensional projection, and on the right a three dimensional one (with two and one reciprocal space dimensions removed respectively). The full four dimensional situation is left to the reader's imagination.

result, a scheme is presented which deals with the more common "kissings" detailed above — the straightforward sheet "kissings". The approach is to make these true crossings — thus keeping the spectral weight in approximately the correct position, but ignoring the fine structure related to the band repulsion (i.e. a small change in the spectral weight due to the opening of a band gap).

The simple "kissings" to be treated are dominated by two interacting bands, the remaining bands playing a much weaker role. Since, as mentioned above, the second order Taylor expansion is insufficient in the region of the "kissing", higher order information about the behaviour of the bands is required. This is provided by a restricted form of Lowdin perturbation theory — see Kane [49] for a review. In fact, historically it has been common practice to use Lowdin perturbation theory to describe strongly interacting or degenerate bands in a **k**.**p** approach [89]. However, it has often been incorrectly applied, diagonalising non-Hermitian sub matrices [99]. This has been pointed out by Loehr [57]. It is likely that this went unnoticed given the large number of **k**-points used in these earlier **k**.**p** methods (e.g. 1357 in $\frac{1}{16}$ if the Brillouin zone in Wang and Callaway's [99] nickel bandstructure calculation). Given the low sampling density required here, there is little

m	k_x	k_y	k_z	B_1	B_2	B_3	E_i	i
3	-0.22	0.48	0.84	-38.7	-0.1	84.5	59.9	16
1	-0.76	-0.36	-0.53	-37.3	-9.7	-1.1	66.0	17
1	0.81	-0.49	0.28	-1495.5	-101.6	41.4	67.4	18
3	0.85	-0.42	0.29	-69.2	34.5	1507.9	67.5	19
3	0.43	0.88	-0.17	-37.2	8.7	136.3	69.0	20

Table 3.3: Curvature information indicating a simple band "kissing". The two highlighted rows (Band i=18,19) contain the proposed "kissing". Note, the energy eigenvalues are close, the curvatures are extreme and the directions of greatest curvature are approximately parallel.

room for such errors.

The fundamental problem in the present approach is to solve the band ordering problem, within the sub-cell in which the "kissing" occurs. This can only be done simply if we assume the bands effectively cross only in one dimension. This can be a good approximation if the sub-cells are small compared to the local curvatures of the bands. In such a case, a three dimensional crossing can be tracked along the direction of high curvature. If the approximation is good the directions will be the same for both bands and this is one of the conditions for detecting the treatable "kissings". The band ordering then becomes a one dimensional problem, and the the two bands may be unambiguously crossed (see Figure 3.5). New expansions for the two bands are constructed, using the higher order information from the Lowdin perturbation calculation. A more detailed recipe for carrying out the procedure is as follows:

1. Look at the stack of bands for each k-point, one at a time (it is still an extrapolative method).

2. Calculate the **k.p** expansion at each point (it is assumed that true degeneracies do not occur due to numerical inaccuracies).

3. Diagonalise the curvature tensor, and note the direction of greatest curvature.

4. The treatable "kissings" will consist of pairs of high curvature bands (see Table 3.3). For both bands the direction of high curvature will be close (modulo a sign), and one will curve positively and the other negatively due to the sign change in the energy denominator of the perturbation sum.

5. Having found a candidate "kissing", the bands are crossed by finding band energies either side of the suspect "kissings" using Lowdin perturbation theory (as corrected by Loehr). The following 2×2 Hamiltonian is diagonalised for the required $\mathbf{q}$ (where α_1 and α_2 refer to the "kissing" bands):

$$\begin{aligned} H_{\alpha_i\alpha_j} = \; & E_{\alpha_i}(\mathbf{k}_o)\delta_{\alpha_i\alpha_j} + i\mathbf{q}^T\mathbf{P}_{\alpha_i\alpha_j} - \tfrac{1}{2}\mathbf{q}^T\langle\Psi^n_{\mathbf{k}}| \left[[\hat{H},\hat{\mathbf{r}}],\hat{\mathbf{r}}\right] |\Psi^n_{\mathbf{k}}\rangle\mathbf{q} \\ & -\tfrac{1}{2}\sum_{m\neq\alpha_i,\alpha_j} \mathbf{q}.\mathbf{P}_{\alpha_i m}\mathbf{q}.\mathbf{P}_{m\alpha_j}\left[\frac{1}{E_{\alpha_i}(\mathbf{k}_o)-E_m(\mathbf{k}_o)} + \frac{1}{E_{\alpha_j}(\mathbf{k}_o)-E_m(\mathbf{k}_o)}\right] \end{aligned} \tag{3.27}$$

6. The new curvature in the high curvature direction is found (by evaluating the band energies assuming the new crossed allocation), and replaces the old curvature. The same is done with the gradient in that direction.

7. The gradients and curvatures are rotated back to the original co-ordinate frame.

8. Having looked through all bands and $\mathbf{k}$-points the procedure can be repeated, since there may be several "kissings" mixed together.

Clearly, some criteria as to what constitutes a strong curvature, and how close the high curvature directions must be are required. Setting these thresholds will determine the tolerance to "kissings", and thus appropriate values for these thresholds will depend on the resolution required in the final spectrum. At present, the treatment of this threshold remains rather crude, simply forming a weighted sum of number representing the closeness of the directions and strength of the curvatures. Setting the relative weights of these two contributions is somewhat an exercise in trial-and-error — more work is required to developed a robust approach to these thresholds.

3.8 Brillouin Zone Division

So far the expansion of the bandstructure of a given system about any point in the Brillouin zone has been discussed. This expansion is to second order only, and in anything other than the case of a free electron, it will be valid only about a finite region surrounding the expansion point. It is required, therefore, to divide the Brillouin zone into a set of "sub-cells" within which the expansion holds to the accuracy required. Thus, a piecewise

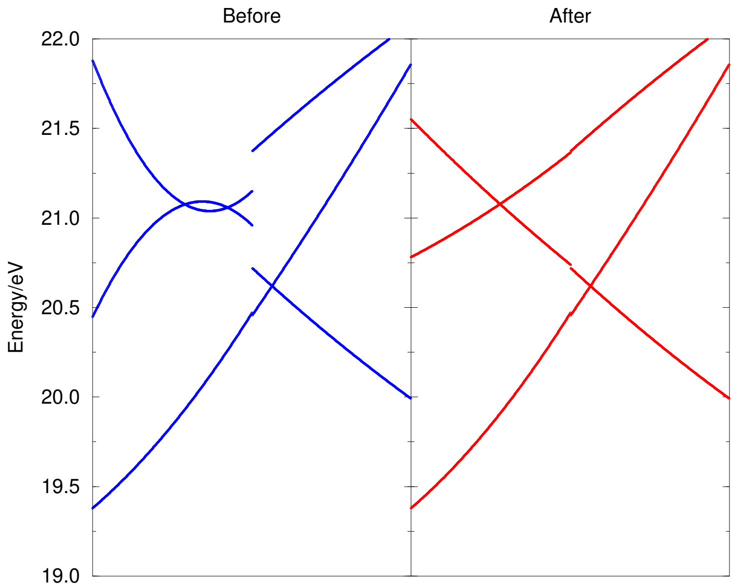


Figure 3.5: An example of the scheme at work. The crossing procedure described above clearly works well for this particular band "kissing" situation — the bands which were presumably previously "kissing" being made to truly cross. It is taken from the conduction band of Silicon, where such situations are common. As can be seen from the discontinuities in the bands, there are two **k**-points along the slice presented — not shown since the slices do not pass through them.

quadratic representation of the bandstructure is constructed. The total DOS for a given band might therefore be written:

$$D_{tot}(E) = \sum_i D_i(E), \tag{3.28}$$

where $D_i(E)$ is the DOS contribution of the i^{th} sub-cell. The challenge is to divide the Brillouin zone in the most efficient way, with an emphasis on good coverage for very low **k**-point sampling.

3.8.1 Commensurate sub-cells

One approach is to divide the Brillouin zone into commensurate sub-cells. This is achieved by forming a grid of points in terms of multiples of integer fractions of reciprocal lattice vectors, keeping those falling within the first Brillouin zone and constructing Wigner-Seitz proximity sub-cells about each point that is retained. For example, if $\mathbf{R}_{1,2,3}$ are reciprocal lattice vectors, then a point in the new grid is given by:

$$\mathbf{r} = \alpha\frac{\mathbf{R_1}}{l} + \beta\frac{\mathbf{R_2}}{m} + \gamma\frac{\mathbf{R_3}}{n}, \tag{3.29}$$

where α, β, and γ are variable integers, and l, m, and n fixed integers chosen to give an even distribution of **k**-points in reciprocal space (i.e. approximately spherical Winger-Seitz proximity sub-cells). If l, m, or n are even then the grid is shifted to balance the points about the origin in the associated direction. Figure 3.6 is an example of such a division of the first Brillouin zone into sub-cells. Notice that some parts of the sub-cells extend outside the Brillouin zone boundary, but may be brought back by the application of an appropriate translation by reciprocal lattice vectors.

If the system concerned has symmetry in addition to the inversion symmetry that will always be present, then the sub-cells may be symmetrised. However, it is clear that for low **k**-point density the symmetry will not be well used (at the extreme limit of a single **k**-point the sub-cell will be the entire first Brillouin zone, and symmetrisation will have no benefit. Also, such grids tend to produce points along high symmetry directions, making band "kissings" more common. Naively shifting the grid produces problems with symmetrisation. However, one may keep the cells fixed and shift the **k**-point of expansion within the sub-cell.

3.8.2 Full use of symmetry

Although many modern electronic structures calculations concentrate on systems with ever larger unit-cells, with ever reduced symmetry, it is clear that there are still many problems that could benefit from the use of symmetry such as core hole effects in bulk crystals as discussed in Chapter 5 , or any highly computationally intensive technique such as the GW method for calculation of quasiparticle spectra [35], or planewave all-electron calculations which are described in Section 4.6.4. To this end, a technique for efficiently making use of symmetry at low **k**-point densities is presented.

1. Find the first Brillouin zone for the system concerned using the Wigner-Seitz proximity cell definition for symmetry reasons.

2. Find the irreducible wedge, assuming the full symmetry of the lattice type. It is constructed from the Brillouin zone by the introduction of a set of planes which are dependent on the lattice type.

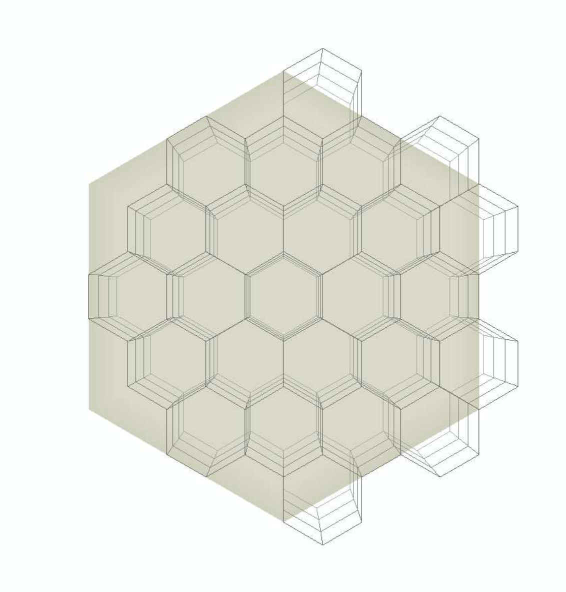


Figure 3.6: A hexagonal Wigner-Seitz Brillouin zone, divided into a commensurate set of mini-cells

3. Decide how many times to recursively halve the volume of the sub-cells — starting with the irreducible wedge. For example, n times will produce 2^n sub-cells and **k**-points. There is clearly complete freedom as to how to halve each sub-cell. It is important to try to make each sub-cell as close to spherical as possible so as to avoid "long" directions in which the second order Taylor expansion will be less well converged. So, the sub-cells are chopped perpendicular to the "long" direction as determined by consideration of the moment of inertia tensor of the cells which are treated as uniform solids, at a point where there is equal volume either side of the cut.

4. Having constructed the sub-cells, the **k**-points are placed at the centre of mass of each cell again treated as a uniform solid.

5. So far full lattice symmetry has been assumed. In the case of reduced symmetry that occur as atoms are placed on the lattice it is necessary to determine which symmetry operations generating the star remain after symmetrisation. The divided irreducible wedge, as formed above, is then repeated for each remaining part of the symmetrised star.

Examples of the application of the procedure outlined above are shown in Figure 3.7. It is clear that this approach makes full use of the system's symmetry. In the limit of a single **k**-point for a system with full symmetry, the one cell generated is the irreducible wedge. It should be noted that for any integration routine based on interpolation it is impossible to contemplate the use of a single **k**-point. Also, generally the **k**-points generated in this fashion do not fall in directions of high symmetry, reducing problems with band "kissing" and crossings. However, it should also be pointed out that this method is not particularly useful if high **k**-point sampling densities are required as the polyhedra produced get quite complex, and there is little gain due to symmetry at high density in comparison to the commensurate sub-cells described in Section 3.8.1. This approach is designed for situations of high symmetry, and for Brillouin zone integrations requiring low **k**-point sampling density.

3.9 Analytic quadratic density of states

Having obtained the piecewise quadratic representation of the band-structure, it still remains to perform Brillouin zone integrations for the DOS and other spectral properties. Clearly, one could re-sample the representation to find the band energies on a finer grid, and simply apply one of the standard methods for Brillouin zone integration described in Section 3.2. This is not appealing, since the correct analytic form for the bands around the critical points is lost. Furthermore, it would appear to be a computationally costly route. The ideal approach would be to take the quadratic representation, and turn it directly into a contribution to the total DOS. One may feel this is possible, since the DOS of a free electron (with its quadratic dispersion) has a well known, and simple, analytic form. The complication, which is significant, arises when the contribution to the DOS comes from a finite region in reciprocal space (i.e. the sub-cells described in Section 3.8). An expression for the density of states for a given band is:

$$D(E) = \int \delta(E(\mathbf{k}) - E) d^3\mathbf{k} = \int_S \frac{dS}{|\nabla_{\mathbf{k}} E(\mathbf{k})|}, \tag{3.30}$$

where S is a constant energy surface (CES) at energy E. For a spherical dispersion relationship $E(\mathbf{k}) = Bk^2$, $|\nabla E(\mathbf{k})| = 2Bk$ and hence:

$$D(E) = \int_S \frac{dS}{|\nabla_{\mathbf{k}} E(\mathbf{k})|} = \frac{4\pi k^2}{2Bk} = \frac{2\pi\sqrt{E}}{\sqrt{B^3}}, \tag{3.31}$$

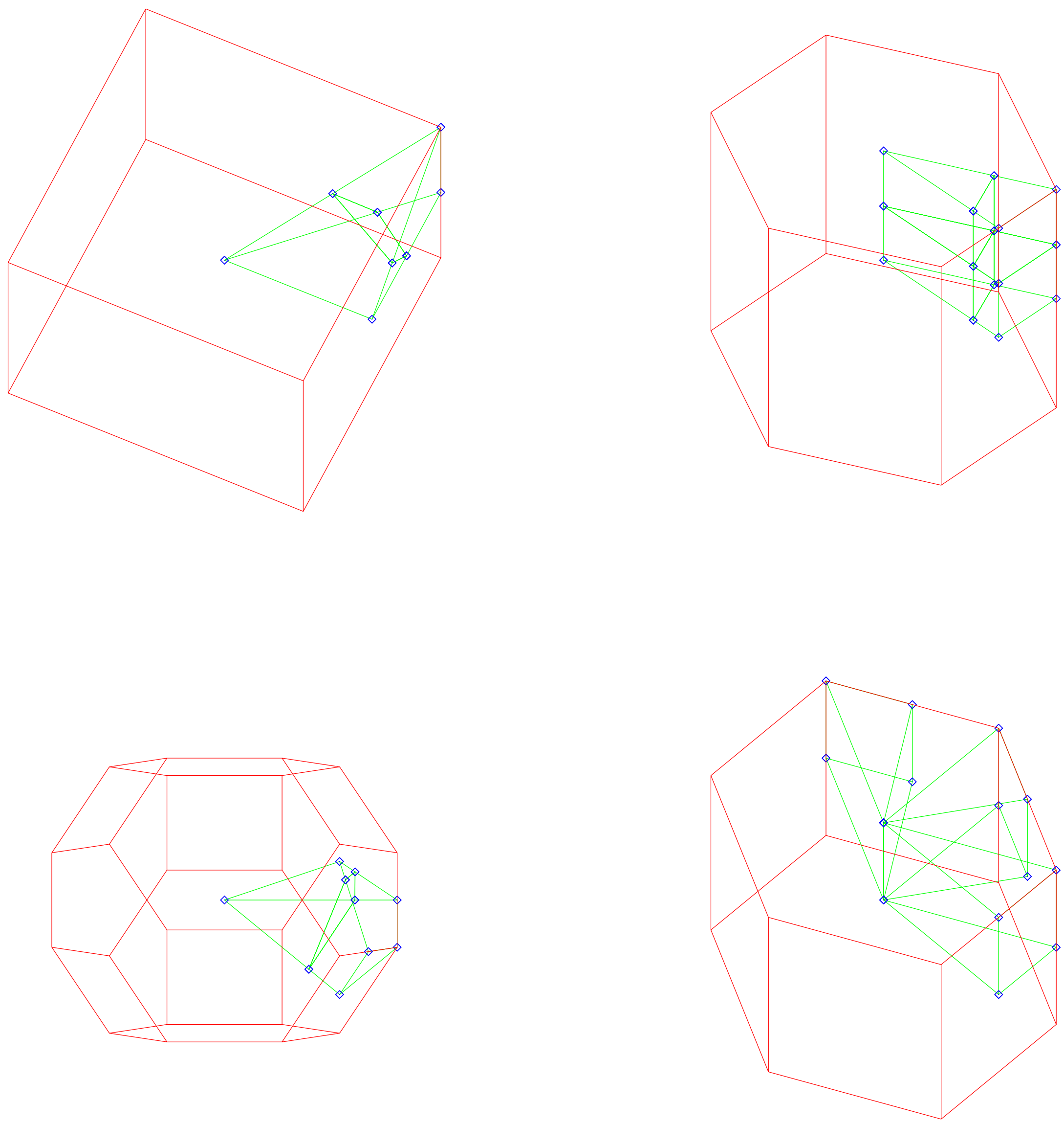


Figure 3.7: *Top Left*: A simple cubic Brillouin zone, with the full symmetry irreducible wedge divided into two sub-cells. *Bottom Left*: FCC lattice Brillouin zone with full cubic symmetry. *Top Right*: Hexagonal Brillouin zone, with full hexagonal symmetry and the irreducible wedge halved twice. *Bottom Right*: A hexagonal Brillouin zone with reduced symmetry, with the irreducible wedge undivided.

where spherical symmetry has been used to simplify the surface integral.

The problem is now extended to the situation in which the surface integration is restricted to a finite volume in reciprocal space (bounded by a convex polyhedron). For simplicity consider a cube, with the origin at its centre. Clearly, for small E the expression for the DOS will be unchanged. However, at some energy the CES must start to cut the polyhedron. At this energy and above the DOS will be reduced (since the included surface is reduced). This energy will later be known as a *singularity*. Singularities occur when the CES strikes the faces, edges and vertexes of the polyhedron. At some point the entire CES will be outside the polyhedron and the DOS will become zero. It may appear to be a difficult problem to track this change in the DOS and indeed this problem has been called *impossible* by Lehmann and Taut [56] but the work of Methfessel *et al* [12, 62, 63] presents a solution, and a scheme for its implementation. It involves consideration of differential geometry. As this is rather technical a detailed description of the implementation used in this work is presented in Appendix B.

3.10 The method at work

Having described the method in detail, there now follows some examples demonstrating the quality of Brillouin zone integrations that are possible with this scheme. It should be borne in mind that the technique is not designed for the ultimate in accuracy, but rather as a method for obtaining the "quality" of a spectra rapidly, to a given energy resolution. In certain situations the DOS converges rather slowly, for example in the S-band DOS (Figure 3.9). But in others it converges very rapidly, for example in the limit of free electrons (Figure 3.8). The convergence of a total DOS with the number of **k**-points is demonstrated and then the effect of introducing matrix elements on the convergence of an ELNES is investigated. Optical spectra are then presented. Finally, the effect of truncating the **k**.**p** perturbation sum is examined. It should be noted that in these examples the second order correction to the curvatures discussed is Section 3.6 is not applied.

3.10.1 Toy models

Free electron dispersion

The Brillouin zone integration method is applied after performing an electronic structure calculation with the ionic potentials set to zero. A simple cubic cell is used, and the

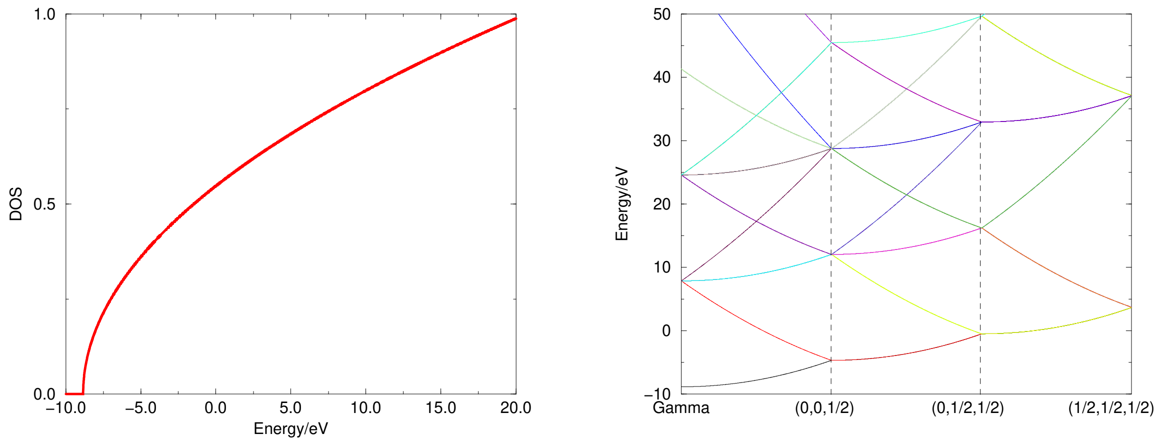


Figure 3.8: *Left:* Density of states for a free electron, calculated in a simple cubic cell using a single **k**-point, displaying the expected $(E-E_o)^{\frac{1}{2}}$ behaviour. *Right:* Corresponding free electron band-structure. The co-ordinates indicate points in the Brillouin zone — in units of reciprocal lattice vectors.

single integration sub-cell is the irreducible wedge. As expected (since the free electron dispersion is quadratic) the DOS is fully converged using a single **k**-point — see Figure 3.8. A selection of bands are shown — also extrapolated from a single sample point. This is not an entirely trivial result, since an interpolative approach would have had problems with band crossings. This result should be compared to the linear extrapolative free electron test of Müller *et al* [69] which required 256 tetrahedra for a similar level of convergence.

Simple cubic s-band test

The convergence is now examined for a simple cubic s-band, following Boon *et al* [12]. The dispersion relation is:

$$E(\mathbf{k}) = -\frac{1}{3}[\cos \pi k_x + cos\pi k_y + cos\pi k_z] \tag{3.32}$$

The gradient and curvature information is obtained directly from the dispersion relation. While Boon *et al* [12] looked at the extreme convergence properties and found the analytic quadratic method to work well, as it deals with the van Hove singularities, here the emphasis is on the behaviour at low sampling densities.

It is clear that, in comparison with the free electron case, convergence is slow (see

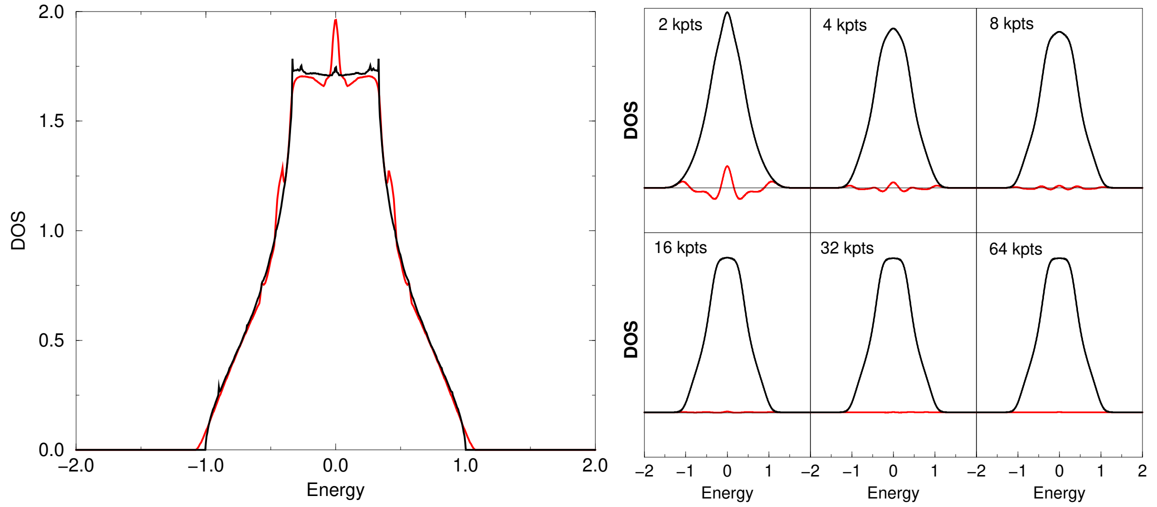


Figure 3.9: *Left:* S-band DOS with no smearing. The red curve is produced using 8 **k**-points in $\frac{1}{48}$ of the Brillouin zone, while the black is produced with 128 **k**-points. Notice the poor convergence on the plateau. *Right:* Convergence of s-band DOS with smearing. A Gaussian smear of 0.1 is applied, producing the black curves. The red curves are the differences of the unconverged DOS from one converged at 128 **k**-points

Figure 3.9). This is to be expected — the dispersion relation contains terms to all orders. Particular problems are encountered on the plateau of the DOS — the spikes being due to overlap in energy of the piecewise quadratic bands due to an inadequacy of the second order expansion. However, in keeping with the philosophy of the approach, the application of a Gaussian smearing to the DOS results in good convergence at low **k**-point sampling densities.

3.10.2 Total density of states of Diamond

Figure 3.10 displays a series of total DOS calculated for a two atom primitive cell of Diamond using an increasing number of **k**-points in the irreducible wedge of the Brillouin zone. The most important point is that the essential features of the spectra are found for very low **k**-point densities and somewhat surprisingly for as few as two points in the irreducible wedge. However, absolute convergence is slower due to inaccuracies in the band curvatures as a result of the neglect of the 2^{nd} order correction, and restricted number of bands in the perturbation sum. Both of these effects can be corrected. Given the low resolution of many experimental spectra, it clearly appears to be unnecessary to push the calculations to very high sampling densities. It should be pointed out that any interpolative method would require more than a single sample point just to perform the interpolation, regardless of the integration quality. Also, had the conventional cubic cell of Diamond (containing 8 atoms) been used then a single sample point would be expected to yield the same convergence as the four point calculation shown here.

To demonstrate the superiority of the extrapolative approach, the results of the current scheme are compared with results obtained from a LMTO bandstructure calculation and Blöchl's improved linear tetrahedral interpolative integration method [10] in Figure 3.11. Figures 3.10 and 3.11 are not directly comparable. Although the valence bands are very similar, the conduction band in the LMTO calculation does not extend far beyond the Fermi energy due to the inadequacy of the basis. Notice also that a second band gap opens fully in the LMTO calculation. However, the intention is to compare the quality of the Brillouin zone integration in the piecewise quadratic and linear approaches. While in the piecewise quadratic approach, a single **k**-point corresponds to a single integration sub-cell, in the linear tetrahedron method several tetrahedra can be generated for each **k**-point using symmetry, and the sharing of vertices. However, it is unfair to the linear tetrahedron method to base the comparison on number of integration sub-cells — it is better to compare the number of sample points, i.e. **k**-points. Clearly, the linear method fails completely for a single **k**-point. Convergence is slow, the linear method struggling to reproduce the form of the van Hove singularities. For two **k**-points in the piecewise quadratic method the onset of the valence band is essentially converged, while it takes about 120 in the linear method to represent the $E^{1/2}$ onset to the DOS.

3.10.3 Including matrix elements - ELNES

A sequence of Diamond K-edges are shown in Figure 3.12. Transitions into the occupied states are also included, although it is clear that these are not experimentally observable. The intention is to explore the effect of the inclusion of matrix elements in the Brillouin zone integrations, since in principle their dispersion should be taken into account as the integrations are performed — see Section 3.11. However, in practice their inclusion appears to have little effect on the convergence properties, and so one may conclude that the matrix element dispersion across sub-cells for core loss spectra is unimportant and the matrix elements vary relatively smoothly across the Brillouin zone.

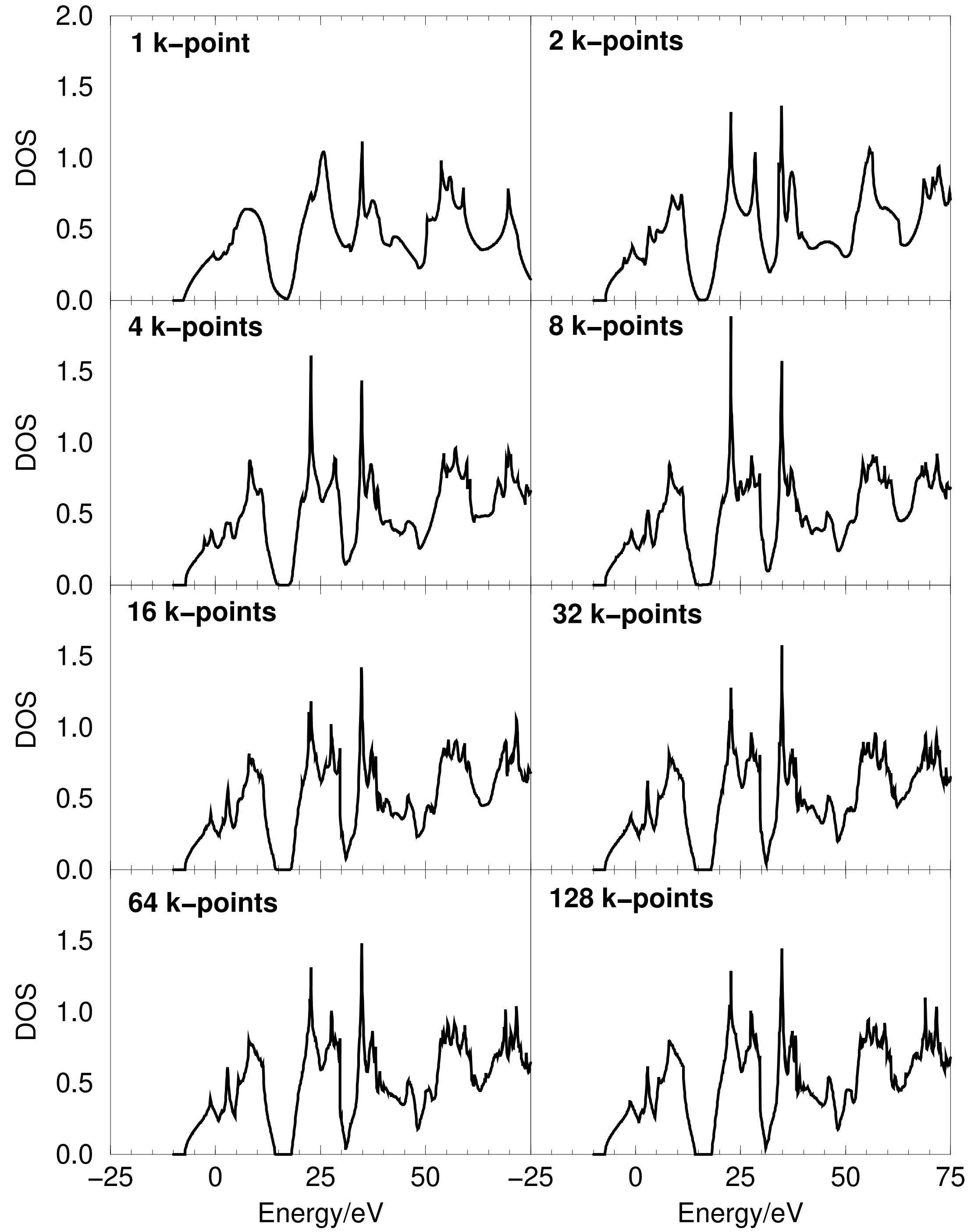


Figure 3.10: Convergence of Diamond DOS: Two atom diamond primitive cell, 24 bands, and an energy cutoff of 600eV. No smearing is applied. The **k**-points are contained within $\frac{1}{48}$ of the Brillouin zone. Notice that the dominant features appear very rapidly.

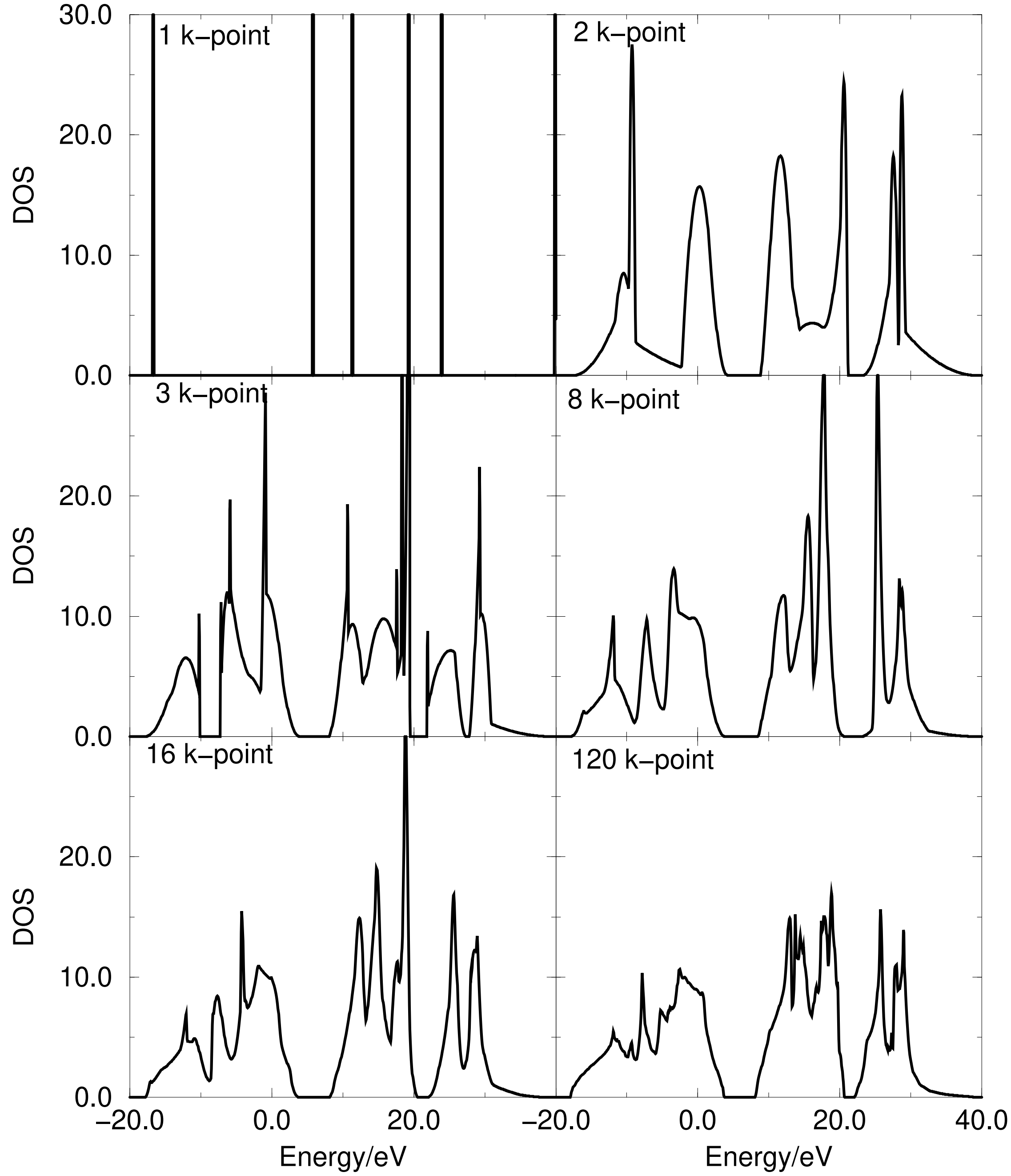


Figure 3.11: Convergence of Diamond DOS — using the linear tetrahedron method and LMTO: The use of symmetry means that, in general, several tetrahedra can be generated from a single **k**-point — 484 tetrahedra in the 120 **k**-point case.

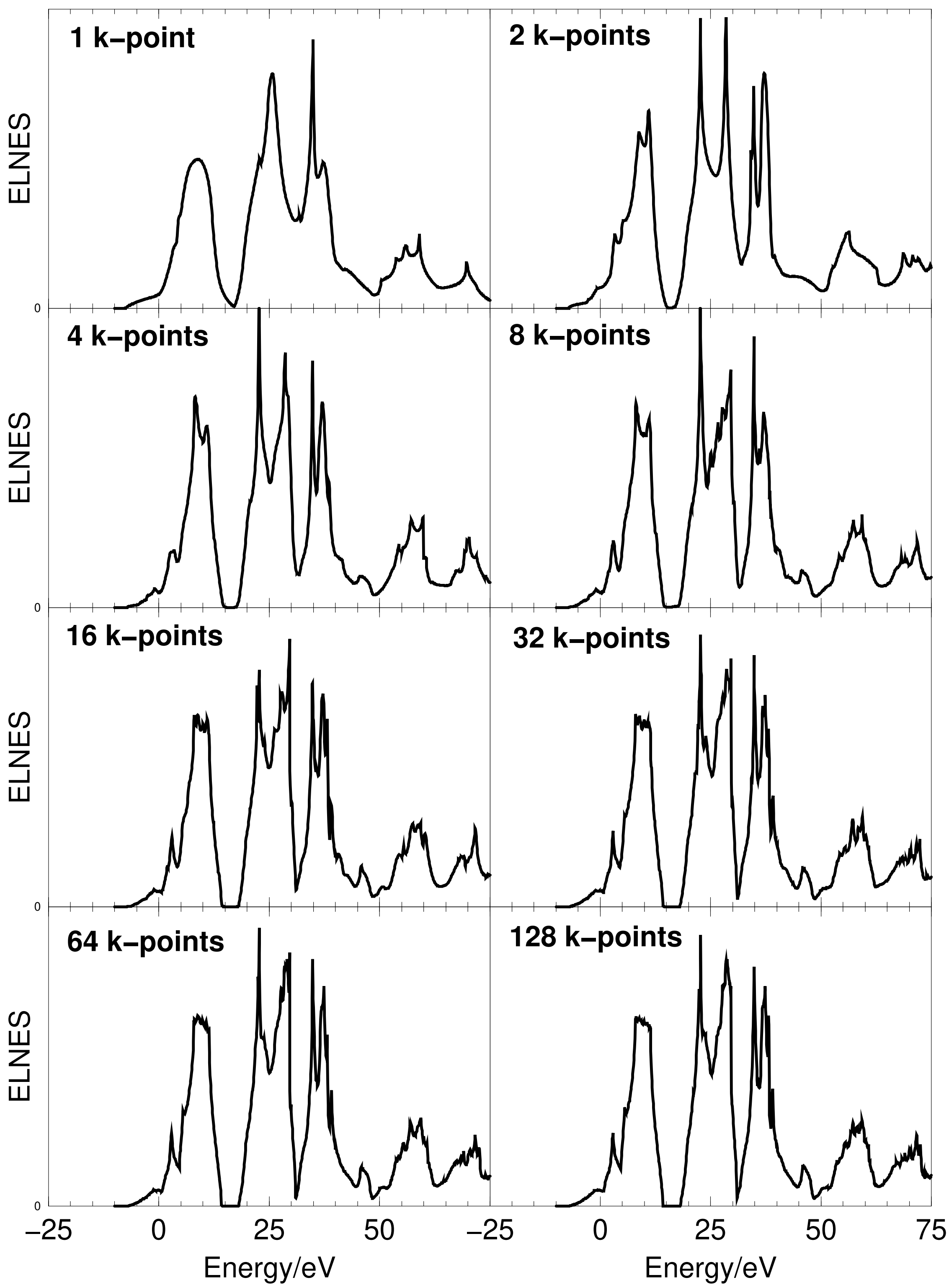


Figure 3.12: Convergence of Diamond K edge: Details as for the DOS. Occupied states are also shown. No smearing is applied. Notice the rate of convergence of the integration scheme is not significantly altered once the ELNES matrix elements are included.

3.10.4 Optical spectra

Figure 3.13 shows the convergence of the calculated optical dielectric function of Diamond as a function of the number of **k**-points in the irreducible wedge. The real and imaginary parts of the dielectric function are shown in red and black respectively. The method used for this calculation is described in detail in Appendix C, but basically the imaginary part of the dielectric constant is a matrix element weighted joint density of states between the valence and conduction bands. Convergence of the optical spectrum is rapid, being essentially complete by 16 **k**-points. However, in contrast to the tests described above, the low **k**-point evaluations show a reduced integrated weight in comparison to the converged results. This indicates the effect of matrix element dispersion. Section 3.11 describes how to overcome this problem.

3.10.5 Convergence with number of bands in k.p sum

As described in Section 3.6, the evaluation of the band curvature requires a perturbation sum over eigenstates. In any practical planewave scheme, this sum is truncated. A given electronic structure calculation will be performed for a fixed number of bands. There is a question as to the effect of this truncation on the accuracy of the resulting Brillouin zone integrations as discussed in see Section 3.6.5. Figure 3.14 shows the valence band DOS of Diamond, calculated using the **k.p** method for different numbers of bands. Truncation appears to have a fairly small effect on the calculated DOS, and the effects would largely disappear on smearing. This effect would increase for lower **k**-point sampling as the incorrect second order term results in a worse approximation to the bandstructure in larger sub-cells.

3.11 K-dependent matrix elements

So far the **k** dependence of the matrix elements that may have to be included in Brillouin zone integrations have been neglected. However, it was noted in Section 3.10.4 that this can have an effect on the convergence of the integral. While, in this work, such matrix element dispersions have not been taken into account (since our interest is primarily with core level spectroscopy, for which little dispersion was detected), Methfessel *et al* [63] provide a technique for dealing with such dispersion. An outline of their approach will be presented here for completeness. The Brillouin zone integrals with which we are concerned are of the

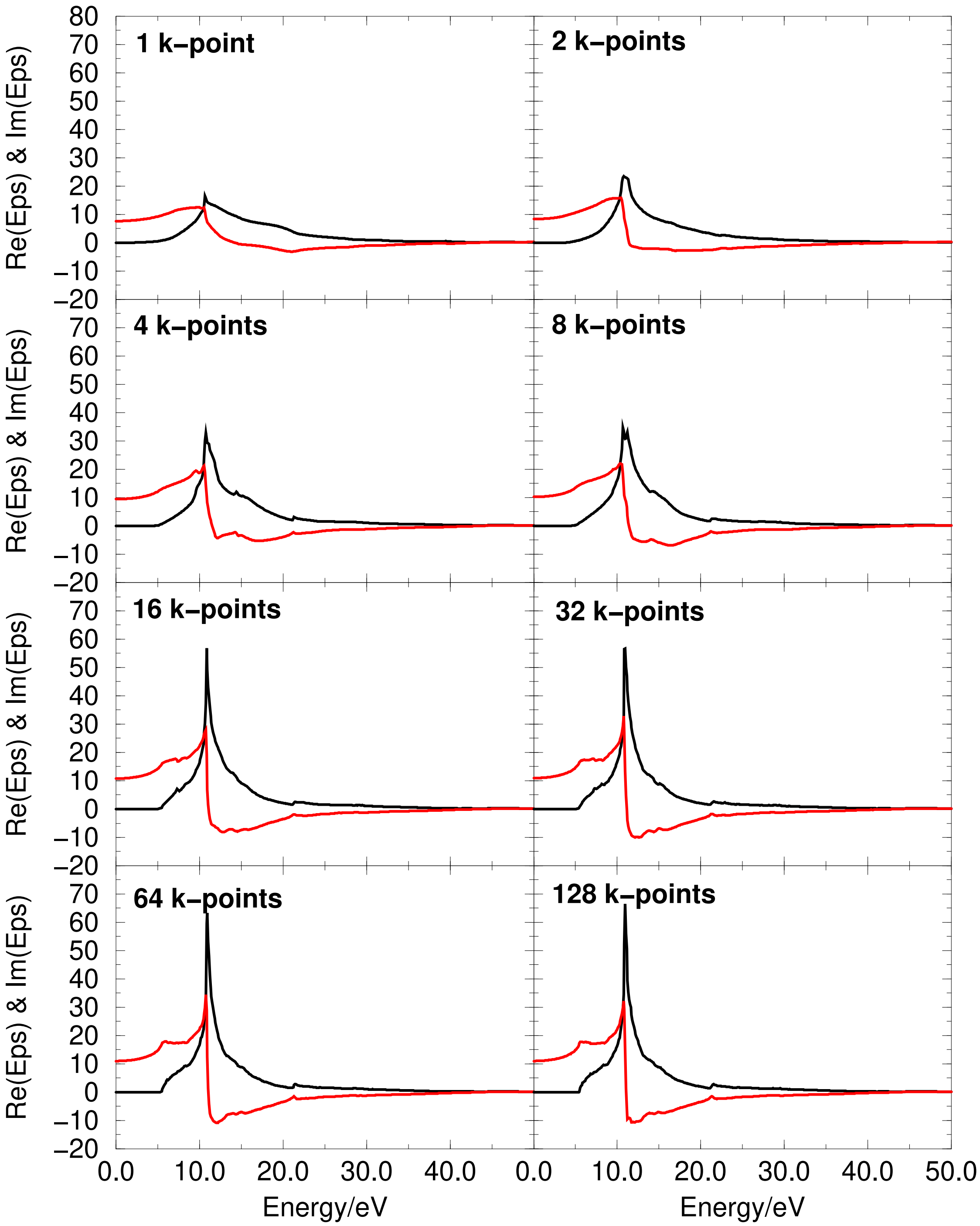


Figure 3.13: Convergence of the optical dielectric constant of Diamond: Calculation details as before. There is no smearing. The black curves show the imaginary part of the dielectric function $\epsilon(\omega)$ and the red curves show the real part. The convergence is rapid, and the calculation is nearly converged with 16 **k**-points. The 32, 64, and 128 **k**-point calculations are practically indistinguishable

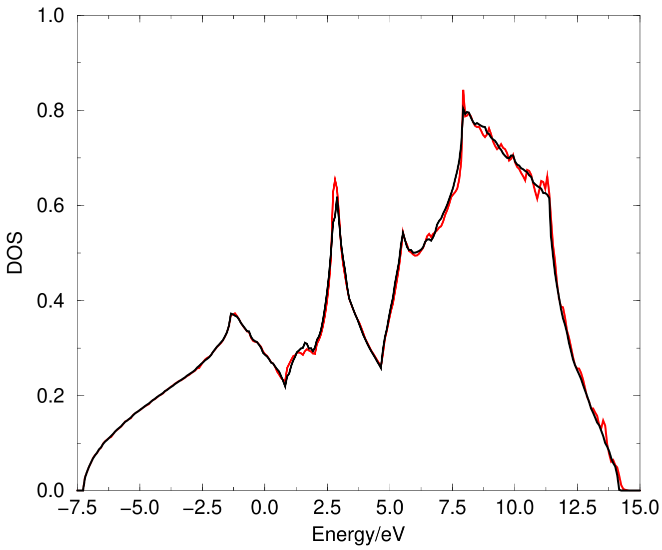


Figure 3.14: Convergence with number of bands in **k.p** sum: Valence band DOS of Diamond, using 600eV cutoff and 128 **k**-points in $\frac{1}{48}$ of the Brillouin zone of the primitive cell. There is no smearing. The red curve is a calculation using no conduction bands, while the black curve is with twenty.

form:

$$F(\varepsilon) = \int_{BZ} u(\mathbf{k})\delta(\varepsilon - E(\mathbf{k}))d^3\mathbf{k}, \tag{3.33}$$

which may be expressed as the surface integral over the constant energy surface S,

$$F(\varepsilon) = \int_S \frac{u(\mathbf{k})dS}{|\nabla E(\mathbf{k})|}. \tag{3.34}$$

The function $u(\mathbf{k})$ may be a transition probability, for example, $u(\mathbf{k}) = \left|\langle\Psi_i|\hat{O}|\Psi_f\rangle\right|^2$. Now, given that within any sub-cell $E(\mathbf{k})$ is expanded to second order only, $|\nabla E(\mathbf{k})|$ is known only to first order. Hence, it makes little sense to expand $u(\mathbf{k})$ (which is assumed to be smoothly varying, beyond first order. Hence, $u(\mathbf{k})$ can be written:

$$u(\mathbf{k}) = u_o + \mathbf{m}^T\mathbf{k} \tag{3.35}$$

And so:

$$\begin{aligned} F(\varepsilon) &= u_o \int_{BZ} \delta(\varepsilon - E(\mathbf{k})) d^3\mathbf{k} + \mathbf{m}^T \int_{BZ} \mathbf{k}\delta(\varepsilon - E(\mathbf{k})) d^3\mathbf{k} \\ &= u_o D(\varepsilon) + \mathbf{m}^T \mathbf{S}(\varepsilon) \end{aligned} \tag{3.36}$$

Retaining only the term $u_o D(\varepsilon)$ has been the approach taken in this work, simply taking the DOS contribution for the sub-cell and multiplying by the matrix element evaluated at the expansion point. The correction consists of the term $\mathbf{m}^T \mathbf{S}(\varepsilon)$. Many different Brillouin zone integrations can be performed for different functions $u(\mathbf{k})$ with a single evaluation of $D(\varepsilon)$ and $\mathbf{S}(\varepsilon)$. The paper by Methfessel *et al* [63] describes how to calculate the vector $\mathbf{S}(\varepsilon)$ in a similar, but rather more straightforward fashion to the evaluation of $D(\varepsilon)$ described in Appendix B.

Chapter 4

The ELNES matrix elements

As shown in Chapter 2 the matrix elements that determine the rate of inelastic scattering of electrons are given by:

$$M_{i\to f}(\mathbf{q}) = \langle \Psi_f | e^{i\mathbf{q}.\mathbf{r}} | \Psi_i \rangle, \tag{4.1}$$

where $|\Psi_i\rangle$ and $|\Psi_f\rangle$ are the initial and final electronic states respectively. In the case of ELNES, the initial state is a core state localised on a single site and so it is this matrix element that gives ELNES its site selectivity. The exponential can be expanded as a Taylor series:

$$M_{i\to f}(\mathbf{q}) = \langle \Psi_f | 1 + i\mathbf{q}.\mathbf{r} + \cdots | \Psi_i \rangle \tag{4.2}$$

Since the initial and final states are orthogonal, the well known *dipole approximation* for small q is obtained. This leads to symmetry selection rules — the *dipole selection rules* [98] — in which the orbital angular momentum can change only by ± 1 between the initial to final states. Thus, the ELNES is often thought of as a site selective symmetry projection of the density of states. Calculation of the DOS is examined in Chapter 3. This chapter is concerned with the evaluation of $M_{i\to f}(\mathbf{q})$ in the solid state, within the framework of the planewave pseudopotential method described in Section 2.2.

This work differs from many previous approaches to the calculation of ELNES in that, as discussed in Section 4.1, the matrix elements are directly evaluated, the symmetry projections arising as a natural consequence. Section 4.2 describes the multipole expansion of $M_{i\to f}(\mathbf{q})$, which may well be important for future calculations, as demand develops for a more quantitative theory develops that goes beyond the dipole approximation. However, the dipole approximation is adopted for the rest of this work. The second half of the chapter is concerned with the effect of the pseudopotential approximation on the evaluation of the matrix elements. In Section 4.3 the problems associated with this approximation are introduced and Section 4.4 discusses the nature of the core states. Sections 4.5 and 4.6 deal with the reconstruction of the all electron wavefunctions from the calculated pseudowavefunctions and the subsequent evaluation of the ELNES matrix elements. A scheme based on the Projector Augmented Wave method of Blöchl [11, 25] is employed. Finally, in Section 4.7 it is pointed out that having reconstructed the all electron wavefunctions it is straightforward to evaluate the all electron charge density, within the pseudopotential (frozen core) approximation.

4.1 Direct evaluation of the ELNES matrix elements

The direct evaluation of the ELNES matrix elements appears surprisingly rarely in the literature. Most previous and current approaches [15, 102] use the dipole approximation. The logic then runs — since the dipole approximation implies that the matrix elements perform a site selective symmetry projection, it simply remains to divide the density of states into s,p, and d local density of states (LDOS) — p-LDOS for a K-edge and s,d-LDOS for L2,3 etc. This is a natural, or easy, approach for many electronic structure methods, since the basis set employed is often written in terms of spherical harmonics located at the atomic sites of the system (e.g. LMTO, LCAO, etc.), so the decomposition is performed as part of the calculation. While this approach will reproduce much of the *quality* of a spectrum, it will be deficient in the quantitative prediction of the spectra. If the dipole approximation holds then the angular integrals in the projections will be perfectly correct, leading to the required symmetry selection rules. However, the radial integrals will be in error (the radial projector will have a significantly larger spatial extent than the actual core state from which the excitation is taking place). Not only will the total cross section be numerically incorrect, so will the energy dependence of the matrix elements. Further, this problem is compounded for the L2,3 edges. In this case, the excitation occurs from a 2p core level and the dipole allowed channels are the s and d projected final states. Since the final spectrum will be a mixture of the s and d-LDOS, it is essential to know the ratio in which they should be added. This is obtained if the matrix elements are evaluated but not otherwise. While the angular momentum decomposition is straightforward in most non-planewave codes, it is correspondingly more difficult to perform the direct evaluation since the wavefunctions are not represented simply. This could explain the reluctance to depart from this straightforward decomposition.

Once the final electronic states are known, and they are output from a planewave pseudopotential CASTEP calculation, and provided that the core states are assumed to be little different from their atomic form, it is straightforward to evaluate Equation 4.1. The final states are Bloch functions, which in the planewave approach are expanded as follows:

$$\langle \mathbf{r}|\Psi_f\rangle = \sum_{\mathbf{G}<\mathbf{G}_{cut}} C_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G}).\mathbf{r}}, \tag{4.3}$$

where $\mathbf{G}$ is a reciprocal lattice vector, and $\mathbf{k}$ is the $\mathbf{k}$-point at which the matrix element is to be evaluated. The initial states are localised on each atomic site (see Section 4.4 below) and are written as:

$$\langle \mathbf{r}|\Psi_i\rangle = \sigma(\mathbf{r}-\mathbf{R}), \tag{4.4}$$

where $\sigma(\mathbf{r}-\mathbf{R})$ is an atomic core state centred at site $\mathbf{R}$. So the matrix element can be evaluated as:

$$\begin{aligned} \langle \Psi_f|e^{i\mathbf{q}.\mathbf{r}}|\Psi_i\rangle &= \sum_{\mathbf{G}<\mathbf{G}_{cut}} C^*_{\mathbf{G}} \int e^{-i(\mathbf{k}+\mathbf{G}-\mathbf{q}).\mathbf{r}} \sigma(\mathbf{r}-\mathbf{R}) d^3\mathbf{r} \\ &= e^{-i(\mathbf{k}-\mathbf{q}).\mathbf{R}} \sum_{\mathbf{G}<\mathbf{G}_{cut}} C^*_{\mathbf{G}} e^{-i\mathbf{G}.\mathbf{R}} \tilde{\sigma}(\mathbf{q}-\mathbf{G}-\mathbf{k}), \end{aligned} \tag{4.5}$$

where $\tilde{\sigma}(\mathbf{q}) = \int e^{i\mathbf{q}.\mathbf{r}} \sigma(\mathbf{r}) d^3\mathbf{r}$ is tabulated for each core type and a polynomial interpolation is performed as required, in much the same way as for the non-local pseudopotential projections.

4.2 Multipole expansion

Given that any evaluation of an inelastic cross section will involve some integration over the momentum transfer (see Chapter 2) in principle $M_{i\to f}(\mathbf{q})$ must be known for many values of $\mathbf{q}$. It would be extremely costly to re-evaluate $M_{i\to f}(\mathbf{q})$ for each value of $\mathbf{q}$ in the integration. Thus, some form of parametrisation is required. As mentioned above, the most common approach is the dipole approximation. There is, however, no reason to restrict investigations to this lowest order approximation. There has been interest in the possibility of observing non-dipole effects in spectra, which are speculated to be more accessible to EELS than X-ray Absorption Spectroscopy (XAS) due to the much greater momentum carried by an electron of a given energy in comparison to a photon [31, 84].

A natural approach might be to simply parameterise $M_{i\to f}(\mathbf{q})$ by using a Taylor expansion to the required order. However, $M_{i\to f}(\mathbf{q})$ is not a completely arbitrary function of $\mathbf{q}$ and the Taylor expansion of $e^{i\mathbf{q}.\mathbf{r}}$ is known analytically. Inserting this into the expression for $M_{i\to f}(\mathbf{q})$ gives:

$$\langle f|e^{i\mathbf{q}.\mathbf{r}}|i\rangle = \langle f|i\rangle + i\langle f|\mathbf{q}.\mathbf{r}|i\rangle - \frac{1}{2}\langle f|(\mathbf{q}.\mathbf{r})^2|i\rangle - \frac{i}{6}\langle f|(\mathbf{q}.\mathbf{r})^3|i\rangle + \frac{1}{24}\langle f|(\mathbf{q}.\mathbf{r})^4|i\rangle + \cdots \quad (4.6)$$

Thus, the parameterisation is analogous to the familiar multipolar expansion of the matrix element, hence the interpolation is driven by the underlying physics.

4.2.1 Extraction of dipole matrix elements

In order to demonstrate the development of the technique for the parameterisation of the ELNES matrix elements, the dipole matrix elements are examined. If the assumption is made that $|\mathbf{q}|$ is small enough that the dipole approximation is a good one, the matrix element can be written as:

$$M(\mathbf{q}) = \langle f|e^{i\mathbf{q}.\mathbf{r}}|i\rangle = M_0 + iq_xM_x + iq_yM_y + iq_zM_z, \quad (4.7)$$

where $M_0 = \langle f|i\rangle$ and $M_x = \langle f|x|i\rangle$ etc. This can be written as a product between two vectors:

$$\begin{aligned} \mathbf{Q}(\mathbf{q}) &= (1, iq_x, iq_y, iq_z) \\ \mathbf{M} &= (M_0, M_x, M_y, M_z) \\ M(\mathbf{q}) &= \mathbf{Q}(\mathbf{q}).\mathbf{M} \end{aligned} \quad (4.8)$$

The extraction of the matrix elements is equivalent to finding the unknown vector $\mathbf{M}$. This can be cast as a linear algebra problem:

$$\begin{pmatrix} 1 & iq_{x_1} & iq_{y_1} & iq_{z_1} \\ 1 & iq_{x_2} & iq_{y_2} & iq_{z_2} \\ 1 & iq_{x_3} & iq_{y_3} & iq_{z_3} \\ 1 & iq_{x_4} & iq_{y_4} & iq_{z_4} \end{pmatrix} \begin{pmatrix} M_0 \\ M_x \\ M_y \\ M_z \end{pmatrix} = \begin{pmatrix} M(\mathbf{q}_1) \\ M(\mathbf{q}_2) \\ M(\mathbf{q}_3) \\ M(\mathbf{q}_4) \end{pmatrix} \quad (4.9)$$

The vector $\mathbf{M}$ is found by inverting the matrix on the left. Clearly the momentum vectors $\mathbf{q}_{i=1,4}$ must be carefully chosen so that the matrix is invertible and that the dipole approximation is good.

Once the vector $\mathbf{M}$ has been found, then the matrix element has been parametrised to first order. For any $\mathbf{q}$ such that the original expansion is valid, the matrix element can be recovered, to within some small error, by forming the vector $\mathbf{Q}(\mathbf{q})$ for the new $\mathbf{q}$ and then evaluating the scalar product with $\mathbf{M}$:

$$M(\mathbf{q}) = \mathbf{Q}(\mathbf{q}).\mathbf{M} \tag{4.10}$$

4.2.2 Higher order terms

The linear algebra approach is not vital in the simple example of the dipole matrix elements given above, but for higher order parameterisations it becomes essential. The expansion is now taken to the fourth order (and so the mono-, di-, quadro-, octo-, and 16-poles are obtained). The required vector $\mathbf{Q}(\mathbf{q})$ is:

$$\begin{aligned}\mathbf{Q}(\mathbf{q}) = \; & (1, \\ & iq_x, iq_y, iq_z, \\ & -\tfrac{1}{2}q_x^2, -q_xq_y, -\tfrac{1}{2}q_y^2, -q_xq_z, -q_yq_z, -\tfrac{1}{2}q_z^2, \\ & -\tfrac{i}{6}q_x^3, -\tfrac{i}{2}q_x^2q_y, -\tfrac{i}{2}q_xq_y^2, -\tfrac{i}{6}q_y^3, -\tfrac{i}{2}q_x^2q_z, -iq_xq_yq_z, -\tfrac{i}{2}q_y^2q_z, -\tfrac{i}{2}q_xq_z^2, -\tfrac{i}{2}q_yq_z^2, -\tfrac{i}{6}q_z^3, \\ & \tfrac{1}{24}q_x^4, \tfrac{1}{6}q_x^3q_y, \tfrac{1}{4}q_x^2q_y^2, \tfrac{1}{6}q_xq_y^3, \tfrac{1}{24}q_y^4, \tfrac{1}{6}q_x^3q_z, \tfrac{1}{2}q_x^2q_yq_z, \tfrac{1}{2}q_xq_y^2q_z, \tfrac{1}{6}q_y^3q_z, \tfrac{1}{4}q_x^2q_y^2, \\ & \tfrac{1}{2}q_xq_yq_z^2, \tfrac{1}{4}q_y^2q_z^2, \tfrac{1}{6}q_xq_z^3, \tfrac{1}{6}q_yq_z^3, \tfrac{1}{24}q_z^4)\end{aligned}$$

Finding a set of $\mathbf{q}$-points that give rise to an invertible matrix requires some trial and error at high order. The following is a suitable set of such $\mathbf{q}$-points:

$\mathbf{q}_1=\alpha(0,0,0)$	$\mathbf{q}_2=\alpha(1,0,0)$	$\mathbf{q}_3=\alpha(0,1,0)$	$\mathbf{q}_4=\alpha(0,0,1)$	$\mathbf{q}_5=\alpha(-1,0,0)$
$\mathbf{q}_6=\alpha(0,-1,0)$	$\mathbf{q}_7=\alpha(0,0,-1)$	$\mathbf{q}_8=\alpha(2,0,0)$	$\mathbf{q}_9=\alpha(0,2,0)$	$\mathbf{q}_{10}=\alpha(0,0,2)$
$\mathbf{q}_{11}=\alpha(-2,0,0)$	$\mathbf{q}_{12}=\alpha(0,-2,0)$	$\mathbf{q}_{13}=\alpha(0,0,-2)$	$\mathbf{q}_{14}=\alpha(1,1,0)$	$\mathbf{q}_{15}=\alpha(0,1,1)$
$\mathbf{q}_{16}=\alpha(1,0,1)$	$\mathbf{q}_{17}=\alpha(-1,1,0)$	$\mathbf{q}_{18}=\alpha(0,-1,1)$	$\mathbf{q}_{19}=\alpha(-1,0,1)$	$\mathbf{q}_{20}=\alpha(-1,-1,0)$
$\mathbf{q}_{21}=\alpha(0,-1,-1)$	$\mathbf{q}_{22}=\alpha(-1,0,-1)$	$\mathbf{q}_{23}=\alpha(2,2,0)$	$\mathbf{q}_{24}=\alpha(0,2,2)$	$\mathbf{q}_{25}=\alpha(2,0,2)$
$\mathbf{q}_{26}=\alpha(-2,2,0)$	$\mathbf{q}_{27}=\alpha(0,-2,2)$	$\mathbf{q}_{28}=\alpha(-2,0,2)$	$\mathbf{q}_{29}=\alpha(2,1,0)$	$\mathbf{q}_{30}=\alpha(0,2,1)$
$\mathbf{q}_{31}=\alpha(2,0,1)$	$\mathbf{q}_{32}=\alpha(1,1,1)$	$\mathbf{q}_{33}=\alpha(-1,1,1)$	$\mathbf{q}_{34}=\alpha(1,-1,1)$	$\mathbf{q}_{35}=\alpha(1,1,-1)$

The magnitude of α is chosen to be small enough that the expansion to fourth order is valid, but not so small as to cause numerical inaccuracies.

4.2.3 Tests of the multipole expansion

The results of a test of the multipole expansion are presented in Figure 4.1. The final states in the matrix element are derived from a pseudopotential calculation, and so are not

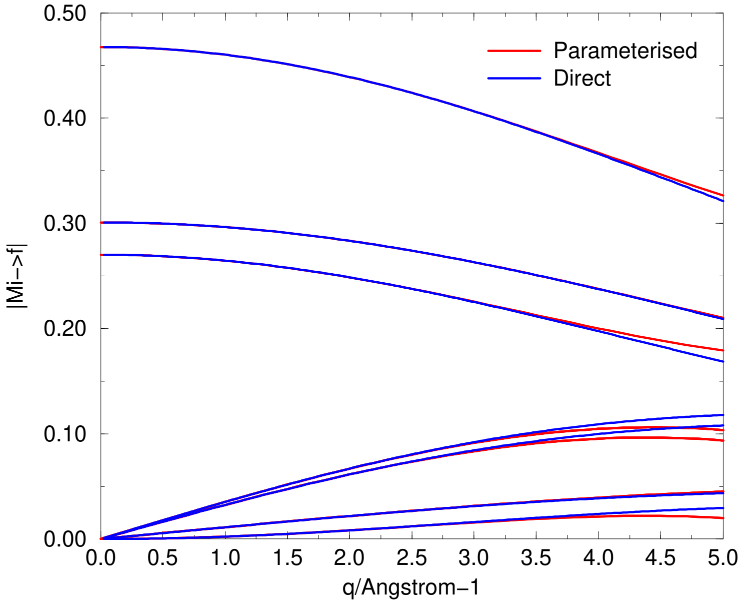


Figure 4.1: A test of the multipole expansion. The blue curves are directly evaluated, and the red ones extracted from the multipole expansion. The calculation is performed at the Γ-point of Diamond, and 14 bands are presented (degeneracy is present). The matrix element is $\sqrt{|x|^2 + |y|^2 + |z|^2}$ to take the arbitrary orientation of degenerate states into account. An α of 1Å^{-1} for the sample points was used. The expansion to fourth order appears to be valid up to $q = 4\text{Å}^{-1}$.

necessarily orthogonal to the core states (in this case the 1s level). This is evident from the non-zero monopole terms for some of the states. In Table 4.1 it is clear that (once the monopole terms are dropped) the dipole term indeed dominates for small q. However, for q not much greater than unity higher order transitions do become important. For 100keV incident electrons, and the Diamond K-edge at about 285eV, $q_{min} \approx 0.25\text{Å}^{-1}$ and $q_{max} \approx 1.6\text{Å}^{-1}$ for a 10mrad collection angle. However, the $\frac{1}{q^4}$ weighting of the cross-section ensures that non-dipole effects will be small for Diamond.

4.2.4 The dipole approximation

Given the conclusions of the previous section, under current experimental conditions — at least for Diamond — the dipole approximation is sufficient to provide quantitative prediction of the ELNES spectra. For the rest of this work the dipole approximation will be applied, despite the techniques being available to go to higher order. While it appears that the non-dipole effects must have some quantitative effect on the inelastic cross-section,

Band #	Monopole	Dipole/Å^{-1}	Quadrapole/Å^{-2}	Octopole/Å^{-3}	16-pole/Å^{-4}
1	0.1560	0.0000	0.0114	0.0002	0.0020
2,3,4	0.0000	0.0359	0.0011	0.0042	0.0006
5,6,7	0.0000	0.0330	0.0013	0.0037	0.0005
8	0.2700	0.0000	0.0148	0.0002	0.0018
9	0.1740	0.0000	0.0091	0.0002	0.0009
10,11	0.0000	0.0000	0.0046	0.0001	0.0015
12,13,14	0.0000	0.0113	0.0016	0.0010	0.0004

Table 4.1: Comparison of multipolar terms at the Γ-point of Diamond, and 14 bands are presented (degeneracy is present). The figures are supposed to represent the "magnitude" of each term — this is hard to define, the choice is taken to form, for example, for the dipole term $\sqrt{|M_x|^2 + |M_y|^2 + |M_z|^2}$. Some monopole terms are non-zero since the final states derive from a pseudopotential calculation.

the current state of theory and experiment does not justify the complications involved in their inclusion. For example, core hole effects on the threshold intensities have yet to be systematically included (although substantial progress towards this is presented in Chapter 5). As theory and experiment will move on, it is important to have shown that there is no barrier to the inclusion of higher order multipole effects in the future.

The investigation of non-dipole effects in the future will be highly desirable [31]. Non-dipole transitions will give access to states with new symmetries — allowing a more complete picture of the unoccupied electronic states. However, large momentum transfers are required to get appreciable non-dipole contributions, and these high momentum transfers lead to low counting rates and hence difficulties in the extraction of the signals. Success in this experimental endeavour will clearly set EELS apart from X-ray absorption spectroscopy.

4.3 The pseudopotential approximation

If an all electron calculation has been performed then all the information required for the evaluation of the ELNES matrix elements would be known — both the initial and final states would have been calculated (even if the core states had been constrained by a frozen core approximation). In this work, however, the pseudopotential approach has been employed. This introduces some very specific problems. Namely, the core levels are excluded from the electronic structure calculation and so the initial states are in principle unknown and the pseudowavefunctions within r_c of the ions bare no resemblance to the true all electron valence wavefunctions as illustrated in Figure 2.4.

4.4 The nature of the core states

A pseudopotential calculation explicitly neglects the core states. This can be a very good approximation for calculations concerned with the total energy, or electronic properties of a system which are dependent only on the valence and conduction bands such as the optical properties which will be discussed in Appendix C. However, some physical properties depend explicitly on the core states. Obvious examples are the core-level spectroscopies, including EELS, the focus of this work, where the core state is one of those involved in the absorption or emission process. Some thought must be given to the questions: can the properties be calculated within the pseudopotential approximation, and if so — how should the core levels be represented?

4.4.1 Why might we expect to get away with pseudopotentials?

This is really a question of how well the frozen core approximation holds. If the pseudopotential works at all, and the potentials are transferable, then it follows that those states which are treated as core states during the construction of the pseudopotential contribute little to the bonding. While core level spectroscopy depends fundamentally on the core levels, those core levels will differ very little from those in the free atomic environment. Some properties, for example core-level shifts, depend on the small changes in the core states in the solid state. Since such changes are small, it should be possible to treat them in a perturbative framework [72].

4.4.2 The choice of core states

The core states in a solid will not differ substantially from atomic core states placed at each site. But, since we are concerned with periodic systems we know that the wavefunctions must be Bloch functions (i.e. $\Psi_k(\mathbf{r}) = e^{i\mathbf{k}.\mathbf{r}}u_{\mathbf{k}}(\mathbf{r})$, where $u_{\mathbf{k}}(\mathbf{r})$ is cell periodic). Within the tight-binding approximation — expected to hold very well in the regime of weakly overlapping core states — the approximation to the core states is given by:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}_i} e^{i\mathbf{k}.\mathbf{R}_i}\sigma(\mathbf{r} - \mathbf{R}_i) \tag{4.11}$$

where a single atom per unit cell is assumed for simplicity. It is clear that within the tight-binding approximation the phase of the wavefunction is constant across each atomic site, the phase changes required according to the **k**-point occuring in the interstitial region. This is clearly shown in Figure 4.2, which displays the results of an all electron planewave calculation for the 1s core states of a fictitious Carbon system.

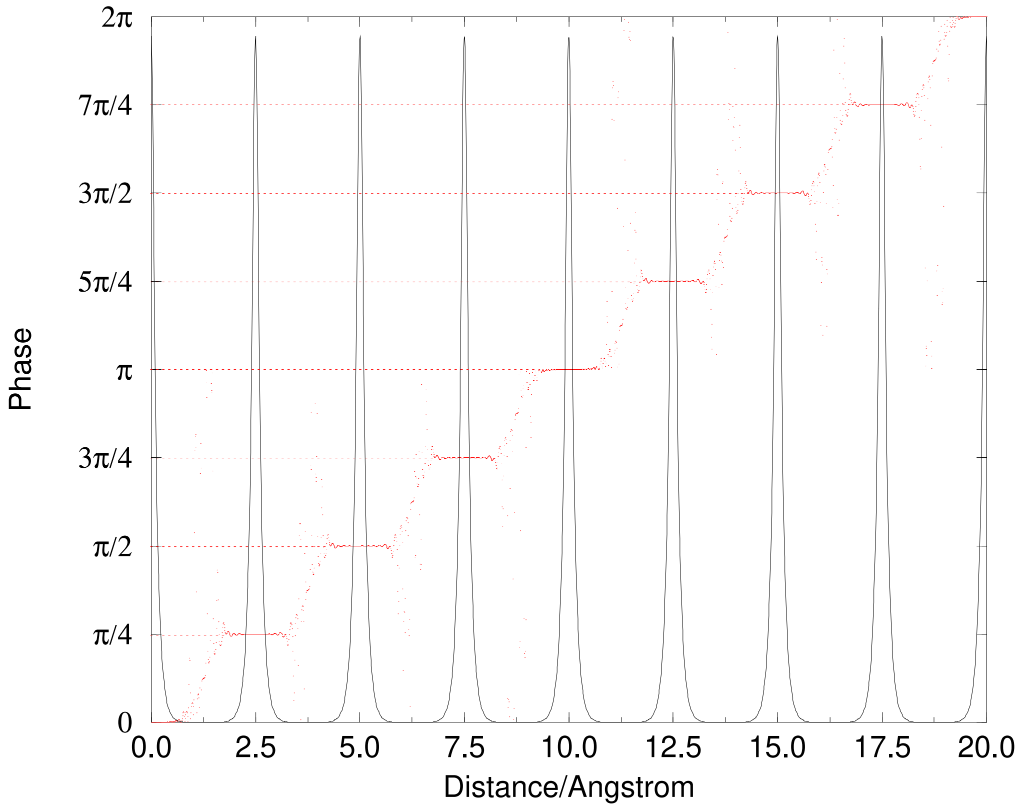


Figure 4.2: The spatial phase variation of a Carbon 1s core state. The system used was a single Carbon atom in a 2.5Å simple cubic cell, and the potential was a Z=6 Coulomb potential and a 9000eV cutoff was applied to the planewave basis set. The core wavefunctions were evaluated at a **k**-point $\mathbf{k} = (0, 0, 1/8)$ in units of reciprocal lattice vectors. The plot consists of the magnitude (solid black line) and phase (red dots) of the 1s band evaluated along the z-axis direction starting from the origin — which coincides with one of the atoms. The phase is ill defined in the regions where the wavefunction tends to zero. Note the increment of phase by $\pi/4$ from site to site in the interstitial regions — expected at $\mathbf{k} = (0, 0, 1/8)$, and the constant phase where the wavefunctions are non-zero.

4.4.3 Conservation of crystal momentum

In the tight-binding approximation the core states remain extended states — there is a coherent phase relationship between the states on different sites, even though the core state is highly localised on each site. This implies that any core level excitation must conserve crystal momentum (as do valence-conduction optical excitations). This can seen as follows. Writing $\langle \mathbf{r}|\Psi_f\rangle = e^{i\mathbf{k}_f\cdot\mathbf{r}}u_f(\mathbf{r})$ and $\langle \mathbf{r}|\Psi_i\rangle = \sum_j e^{i\mathbf{k}_i\cdot\mathbf{R}_j}\sigma(\mathbf{r}-\mathbf{R}_j)$:

$$M_{i\to f} = \underbrace{\sum_j e^{i(\mathbf{k}_i-\mathbf{k}_f+\mathbf{q})\cdot\mathbf{R}_j}}_{\text{crystal momentum conservation}} \underbrace{\int e^{i(\mathbf{q}-\mathbf{k}_f)\cdot\mathbf{x}}u_f^*(\mathbf{x})\sigma(\mathbf{x})d^3\mathbf{x}}_{\text{site and symmetry projection}}, \tag{4.12}$$

where $\mathbf{x} = \mathbf{r} - \mathbf{R}_j$ and the cell periodicity of $U_f(\mathbf{r})$ is utilised. The sum and the integral are independent, and may be evaluated separately. The site and symmetry projection is carried out by the integral. Crystal momentum is conserved by the sum, which is zero unless $\mathbf{k}_i - \mathbf{k}_f + \mathbf{q}$ is a reciprocal lattice vector.

However, this conservation of crystal momentum has no easily observable consequences. For a given $\mathbf{q}$ and $\mathbf{k}_f$, $\mathbf{k}_i$ can be found so that the crystal momentum conservation condition is satisfied. Since the core states are practically non-dispersive, and the integral has no dependence on $\mathbf{k}_i$ the resulting matrix element will be the same as that had the core states been treated as isolated states on each site (instead of the extended tight-binding states). This means that there is no need to be concerned as to whether the core states are extended or not in reality. This is an important point, and appears to have been assumed rather than stated in previous work [102]. The prevalent interpretation holds, that the core excitation occurs from a particular atom in the system, giving EELS its site selectivity, even though the formulation is in terms of periodically extended core states. It is also clear that it is important to have chosen the (correct) tight-binding form of the core states. In general, other choices would result in a dependence on $\mathbf{k}_i$ in the integral.

4.5 Reconstructing the wavefunctions

There are several schemes in the literature for the reconstruction of the full all electron wavefunctions from the corresponding pseudowavefunctions [8,25,33,64,94,97]. Each of the different schemes entail varying degrees of computational expense and effort of implementation. The aim should be to recover the majority of the correction to the wavefunctions and derived quantities with maximum simplicity. In some sense, there is a case for not doing anything. The most significant effect of the dipole matrix element is to introduce the dipole selection rules. Thus, the correct angular integrals will be performed — and so the symmetry projections are produced if the dipole matrix elements between atomic core states and pseudowavefunctions are evaluated. However, for truly quantitative results some form of correction must be applied.

Several of the approaches go too far for our requirements — accurately resolving within the core radius [33, 64, 94, 97]. Another approach is to orthogonalise the pseudowavefunctions to the core states of a free atom. This approach has been used in the past (with empirical pseudopotentials [33]) and recently by Bellaiche and Kunc in the calculation of Compton profiles for Lithium Hydride [8]. It was also adopted in the preliminary EELS calculations using CASTEP which were performed by Brohan [15], and in the initial stages of this work. However, this procedure proved to be unsatisfactory. Problems arise because the orthogonalisation procedure is not uniquely defined, and there are numerical instabilities related to the reorthogonalisation and normalisation up through the bands. Instead, the simple and elegant Projector Augmented Wave approach is employed.

4.6 The Projector Augmented Wave approach to the reconstruction

The Projector Augmented Wave (PAW) approach to the reconstruction of the all electron wavefunction is based on the electronic structure method developed by Blöchl [11]. The central observation that underlies the technique is that the pseudowavefunction is correct beyond the core radius of the pseudopotential, and that the core radius r_c is chosen so that the potential — to within a constant shift — is essentially atomic within it (if this were not the case, and the neighbouring atoms had an effect, then the potential would not be transferable). So, the wavefunction need only be corrected within r_c. Since the potential is approximately atomic near the nucleus, the wavefunction can be expanded in terms of the all electron partial waves (eigenstates of the atomic potential) within r_c. For an isolated atom only a single partial wave would be required for each state (the boundary conditions at r_c being trivially matched), but in the solid state the boundary conditions are altered by the influence of neighbouring atoms and a superposition of the partial waves is required to match them. At the same time as the addition of the all electron partial waves, the corresponding pseudo-partial waves must be subtracted. Since the full and pseudo- potential partial waves match beyond r_c (at least for the reference states used in the construction of the pseudopotential) the wavefunction should clearly be unchanged beyond r_c. The weights required for the superposition are determined by suitable projectors localised within r_c. This, and details of the implementation of the PAW reconstruction are discussed in Section 4.6.1.

This PAW reconstruction approach was originally suggested by Van de Walle and Blöchl for the calculation of hyperfine parameters of defects in semiconductors [25]. While the hyperfine parameters are critically dependent on the electron wavefunctions at (or close to) the nucleus, the authors recognised the importance of using a pseudopotential approach for much the same reasons that motivate this work: which is that all electron methods, such as linear-muffin-tin-orbital (LMTO) or full-potential linearised augmented-plane-wave (FLAPW) methods are greatly restricted by the type, symmetry and size of the system to be investigated.

4.6.1 The theory

In the following, a tilde indicates a pseudo-quantity — for example $|\tilde{\Psi}\rangle$, and an absence of a tilde denotes the all electron equivalent. The all electron wavefunction $|\Psi\rangle$ is expressed as a superposition of the corresponding pseudowavefunction $|\tilde{\Psi}\rangle$ and the partial waves of the isolated full and pseudo- potentials — $|\phi\rangle$ and $|\tilde{\phi}\rangle$ respectively:

$$|\Psi\rangle = |\tilde{\Psi}\rangle + \sum_i (|\phi_i\rangle - |\tilde{\phi}_i\rangle)\langle o_i|\tilde{\Psi}_i\rangle \tag{4.13}$$

The index i refers to both site and partial wave. In principle it runs over the entire partial wave spectrum, starting at the first pseudo-partial wave — the all electron core states are not included. A projector $|o_i\rangle$ is chosen so that the pseudowavefunction can be expressed as:

$$|\tilde{\Psi}\rangle = \sum_i \langle o_i|\tilde{\Psi}_i\rangle|\tilde{\phi}_i\rangle, \tag{4.14}$$

for $r < r_c$. A cutoff function $f(r)$ — to be chosen later — is employed to ensure that the weights are appropriate to the core region. The projector $|o_i\rangle$ is then written as:

$$|o_i\rangle = \sum_j f|\tilde{\phi}_j\rangle\alpha_{ji} \tag{4.15}$$

To ensure adherence to Equation 4.14 the following condition is applied:

$$\langle o_i|\tilde{\phi}_j\rangle = \sum_k \alpha_{ik}\langle\tilde{\phi}_k|f|\tilde{\phi}_j\rangle = \delta_{ij}, \tag{4.16}$$

which implies that:

$$\alpha_{ij} = (\mathbf{S}^{-1})_{ij}, \tag{4.17}$$

where $S_{ij} = \langle\tilde{\phi}_i|f|\tilde{\phi}_j\rangle$. The inclusion of the cutoff function $f(r)$ causes α_{ij} to deviate from δ_{ij} in general — i.e. the $|\tilde{\phi}_i\rangle$ are not orthonormal within the cutoff region.

Following Van de Walle and Blöchl [25], the formalism is simplified by making a definite choice for $f(r)$ — namely, it is unity for $r < r_c$ and zero otherwise. Restriction to a single partial wave per site and angular momentum leads to α_{ij} being diagonal. In practice, the partial waves are restricted to those reference states used in the pseudopotential generation. This ensures that $|\phi_i\rangle$ and $|\tilde{\phi}_i\rangle$ are identical beyond r_c. The numerical value of α_{ij} can be taken into account by redefining the partial waves. They are split in the following fashion:

$$|\phi\rangle = |\phi^{0\to r_c}\rangle + |\phi^{r_c\to\infty}\rangle, \tag{4.18}$$

where $|\phi^{0\to r_c}\rangle$ is exactly $|\phi\rangle$ within r_c but zero elsewhere. The states are then renormalised:

$$|\phi'\rangle = |\phi^{0\to r_c}\rangle/\sqrt{\langle\phi^{0\to r_c}|\phi^{0\to r_c}\rangle} \tag{4.19}$$

The final form of the simplified reconstruction scheme is:

$$|\Psi\rangle = |\tilde{\Psi}\rangle + \sum_i (|\phi'_i\rangle - |\tilde{\phi}'_i\rangle)\langle\tilde{\phi}'_i|\tilde{\Psi}\rangle \tag{4.20}$$

The pseudowavefunctions $|\tilde{\Psi}\rangle$ are obtained from a planewave pseudopotential CASTEP electronic structure calculation, and $\langle\tilde{\phi}'_i|\tilde{\Psi}\rangle$ are evaluated using the method described in Section 4.1 (with $\mathbf{q} = 0$). The corrected dipole matrix element can then be evaluated:

$$\langle\phi_c|r_\alpha|\Psi\rangle = \langle\phi_c|r_\alpha|\tilde{\Psi}\rangle + \sum_i (\langle\phi_c|r_\alpha|\phi'_i\rangle - \langle\phi_c|r_\alpha|\tilde{\phi}'_i\rangle)\langle\tilde{\phi}'_i|\tilde{\Psi}\rangle, \tag{4.21}$$

where $|\phi_c\rangle$ is the atomic core state from which the excitation takes place, and $\alpha = x, y, z$. $\langle\phi_c|r_\alpha|\tilde{\Psi}\rangle$ are the dipole matrix elements evaluated in Section 4.2.1. $(\langle\phi_c|r_\alpha|\phi'_i\rangle - \langle\phi_c|r_\alpha|\tilde{\phi}'_i\rangle)$ can be evaluated once (in real space) for each pseudopotential used. It should be emphasised that the pseudowavefunction matrix elements are corrected, as opposed to evaluating the correct matrix elements from reconstructed all electron wavefunctions which would presents a problem of representation because of the need to simultaneously describe the smooth bonding regions of the wavefunctions and the rapidly oscillating core regions, which is one of the fundamental electronic structure problems.

4.6.2 Limitations

There is a valid concern that the simplification of the formalism applied above (specifically the restriction to a single partial wave per angular momentum state) will begin to fail for states high above the threshold (as nodes begin to arise within r_c). In principle, the remedy would simply be to include more partial waves. However, contrary to the suggestion by Van de Walle and Blöchl [25], the full and pseudo- potential partial waves generally do not match beyond r_c throughout the entire atomic spectrum. This condition is only strictly enforced for the reference states used in the pseudopotential generation. This is of no consequence for their work — which was concerned only with the occupied electronic states. It does, however, complicate the extension of the PAW reconstruction scheme to higher energies. It also implies that the CASTEP pseudowavefunctions for the highly excited states are strictly incorrect beyond r_c. This is well known — but it is generally hoped that it is a small problem. In fact Figure 4.4 suggests that this is indeed the case, for Carbon at least, since the all electron and pseudo DOS agree up to about 40eV above the threshold. But strictly this is a comparison of eigenvalues rather than wavefunctions. A future remedy would be to construct pseudopotential which themselves have more than a single projector per angular momentum, with reference states at higher energies.

It is also obvious that the PAW approach is an approximation — it does not take core relaxation into account since it assumes an atomic-like potential within r_c. This approximation is consistent with the pseudopotential approximation and is expected to be valid.

4.6.3 Free atom tests of the PAW approach

A test of the PAW reconstruction scheme in the limit of an isolated atom is shown in Table 4.2. Isolated Carbon and Silicon atoms are modelled by placing a single atom in a large supercell 10Å on a side. In Carbon the 2p pseudowavefunction is very similar to its all electron counterpart (see Figure 4.3), and hence the correction of the 1s→2p matrix element is small. This should be contrasted to the case of the 1s→3p transition in Silicon.

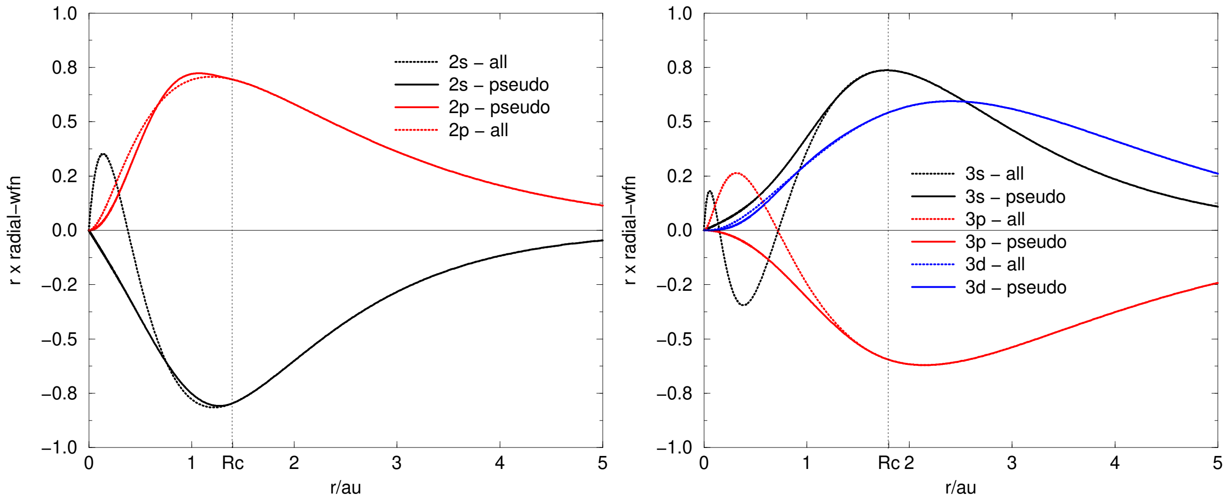


Figure 4.3: Comparison of all electron and pseudowavefunctions for Carbon and Silicon free atoms. *Left:* Carbon — $R_c = 1.4$au *Right:* Silicon — $R_c = 1.8$au, where R_c is the core radius. In both cases, notice that the all electron and pseudowavefunctions agree — as they should — beyond R_c. Also, the Carbon 2p wavefunctions are very similar for the full potentials and pseudopotentials. This explains why such a small correction is required for the Carbon K-edge (i.e. 1s→2p transition). The same can be seen to apply to the Silicon 3d wavefunctions.

	CASTEP Uncorrected	CASTEP Corrected	All electron
Carbon			
$1s \to 2p$	0.03997	0.04313	0.04316
Silicon			
$1s \to 3p$	0.00065	0.00472	0.00471
$2s \to 3p$	0.06207	0.03164	0.03168
$2p \to 3s$	0.09308	0.05213	0.05196
$2p \to 3d$	—	—	0.05262

Table 4.2: Reconstruction of atomic dipole matrix elements using the PAW approach. The CASTEP calculations are performed using the pseudopotential approximation and a supercell of 10Å to model the isolated atom. The Carbon pseudopotential had a core radius of 1.4au, and the Silicon, 1.8au. The reconstructed matrix elements compare well with those evaluated from an all electron real-space calculation. Note that the correction for Carbon is small. Tests show the CASTEP results to be converged with respect to cell size and cutoff, and so remaining errors are presumably caused by numerical inaccuracies. It was not possible to find Silicon 3d states using the supercell approximation. The units are Å^{-1}.

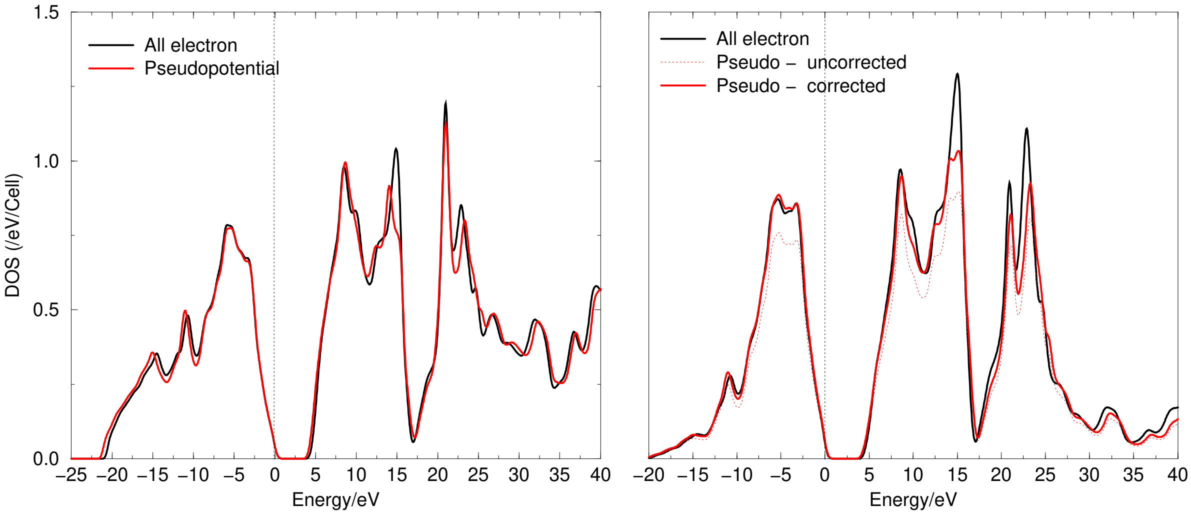


Figure 4.4: All electron total DOS (*left*) and K-edge (*right*) for Diamond, compared to a pseudopotential calculation. The dotted lines indicate the top of the valence band.

4.6.4 Application to the solid state

Testing the evaluation of the ELNES matrix elements in the solid state presents a problem. In the free atom case it is relatively straightforward to produce the all electron wavefunctions and hence matrix elements for reference. The pseudopotential approach was developed precisely because handling the core regions of the wavefunctions in the solid state is a challenging problem. While there are all electron electronic structure techniques available, it is advantageous to use the same code as the one used to do the pseudopotential calculations — changing the pseudopotential to the bare Coulomb potential. This allows a more direct comparison to be made, minimising the problems of different implementations and basis sets. Planewave all electron calculations are very costly. They are possible, but the system size and type are severely restricted, to the smallest cells and to first row elements. Fortunately, this is sufficient to validate and test the accuracy of our methodology. The matrix elements for dipole allowed transitions at the Γ-point of Diamond calculated from an all electron calculation (9000eV cutoff) and a pseudopotential calculation are compared in Table 4.3. There is evidence in the transitions from $1s \rightarrow$ Bands 12,13,14 that the PAW correction is insufficient at higher energies, as suggested in Section 4.6.2, indicating that more partial waves per angular momentum should be included.

In Figure 4.4 the calculation is performed for the full spectrum of states. The fact that DOS agrees well to at least 40eV above the Fermi level offers support to the use of pseudopotentials (constructed for use in the valence band) up into the conduction band in contrast to the expectation of Zeller [104] who suggests that the energy dependence of pseudopotentials was such that they would be of no use for the investigation of unoccupied electronic states. The small discrepancies below the Fermi level are possibly due to insufficient convergence of the all electron calculation at the 9000eV cutoff (the Coulomb potential causes slow convergence) or inaccuracies in the pseudopotentials. Note that the all electron calculations are computationally intensive, and the evaluation of the DOS in this case would be impossible without the efficient Brillouin zone integration scheme presented in Chapter 3. In this case, 16 **k**-points were used, and the resulting DOS smeared with a smearing Gaussian of width 0.25eV. Figure 4.4 also compares an all electron K-edge spectrum with that calculated using a pseudopotential, with and without the PAW correction. On application of the correction, the agreement is very good within the valence band, but gets progressively worse at higher energies, again providing evidence for the need for more partial waves. It is unfortunate that the correction required is not very dramatic, but the planewave all electron calculations are restricted to first row elements, which only have K-edges, and, as mentioned above, Figure 4.3 shows the pseudo- and all electron p-wavefunctions to be very similar — hence only a small correction is required in this case.

The energy dependence of the PAW correction to the matrix elements for Diamond and Silicon are shown in Figures 4.5 and 4.6 respectively — the ratio between the corrected and uncorrected K-edges being plotted alongside the K-edges themselves. The variation is smoothly varying, but as expected the magnitude of the correction is much larger for Silicon than Diamond. The magnitude of the fractional correction actually decreases with energy. This should be treated as suspicious. Table 4.3 shows that on comparison to the all electron matrix element, the correction actually should increase. This is further evidence of the requirement for the inclusion of more partial waves per angular momentum. The ratios would then no longer vary monotonically. The more complicated situation that

1s→Band #	Uncorrected	Corrected	All electron	Un/All
1s→2,3,4	0.0359	0.0388	0.0387	1.078
1s→5,6,7	0.0330	0.0355	0.0358	1.085
1s→12,13,14	0.0113	0.0120	0.0127	1.123

Table 4.3: Matrix element correction at the Γ-point of a Diamond primitive cell. The matrix element used is $\sqrt{|x|^2+|y|^2+|z|^2}$. The units are Å^{-1}. Bands 12-14 lie about 50eV above bottom of valence bands at Γ-point

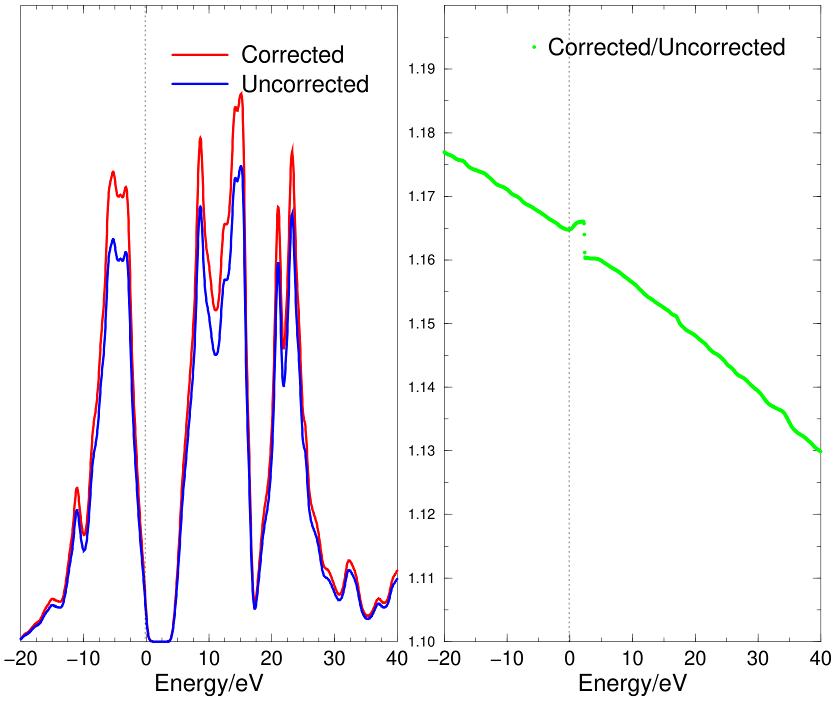


Figure 4.5: Diamond K-edge — before and after correction. The dotted line indicates the top of the valence band.

occurs for the Silicon L2,3-edge is shown in Figure 4.7. Through the dipole selection rules, the edge should be a mixture of s and d projected LDOS. Figure 4.3 shows that the 3d pseudowavefunction differs very little from the all electron one, while the 3s is quite different. Thus, a large correction in only expected where there is significant s-contribution. The correction is essentially constant below the Fermi level, indicating the projected DOS contains predominantly s-like states and very few d-like ones. Above the Fermi level the correction varies strongly with energy, indicating an energy dependent ratio of s to d-like states. The correction tends towards unity at higher energies, as the edge becomes d-dominated.

4.7 All electron charge density

Having reconstructed the all electron wavefunctions, it is trivial to extend the PAW method to the reconstruction of the all electron charge densities, within the pseudopotential approximation. This is of interest to those involved in the measurement of X-ray scattering

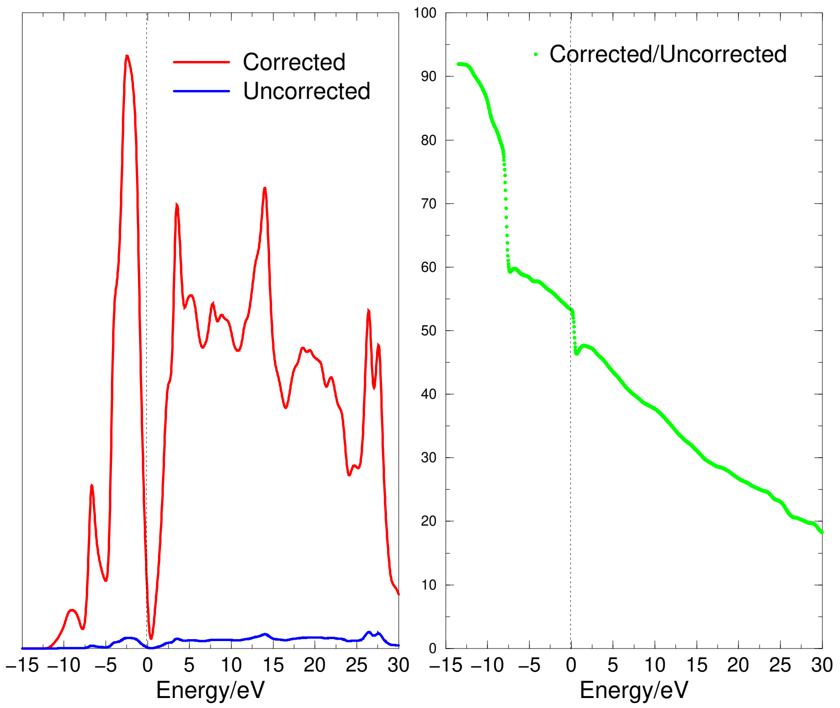


Figure 4.6: Silicon K-edge — before and after correction. The dotted line indicates the top of the valence band.

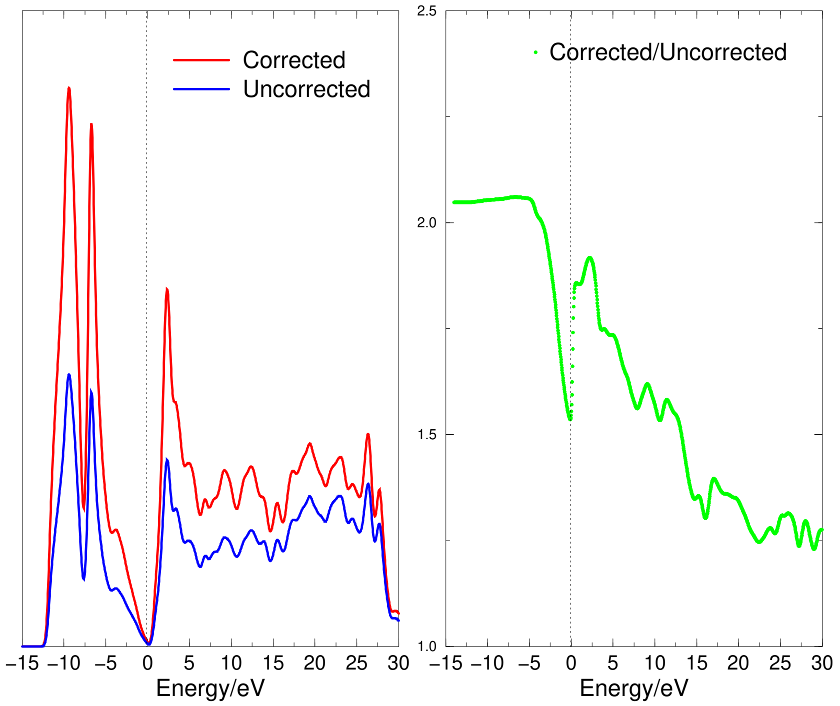


Figure 4.7: Silicon L2,3-edge — before and after correction. The dotted line indicates the top of the valence band.

factors etc. [82]. The total charge density is given by:

$$\rho_{\mathrm{solid}}(\mathbf{r}) = \tilde{\rho}_{\mathrm{solid}}(\mathbf{r}) + \rho_{\mathrm{core}}(\mathbf{r}) + \sum_{n\mathbf{k}} \left\{ \left| \sum_i \langle \mathbf{r}|\phi_i'\rangle\langle\tilde{\phi}_i'|\tilde{\Psi}_{n\mathbf{k}}\rangle \right|^2 - \left| \sum_i \langle \mathbf{r}|\tilde{\phi}_i'\rangle\langle\tilde{\phi}_i'|\tilde{\Psi}_{n\mathbf{k}}\rangle \right|^2 \right\} \quad (4.22)$$

The all electron charge density $\rho_{\mathrm{solid}}(\mathbf{r})$ is given by adding the atomic core charge densities at the appropriate sites $\rho_{\mathrm{core}}(\mathbf{r})$ to the pseudo-charge density $\tilde{\rho}_{\mathrm{solid}}(\mathbf{r})$. The correction is applied by replacing the pseudo-charge density within r_c by the all electron one. Since $|\phi_i'\rangle$ and $|\tilde{\phi}_i'\rangle$ are zero beyond r_c the charge density is unchanged outside the core regions (as required). Since only valence and core electrons contribute to the charge density, the formalism presented above is expected to give accurate results, at least consistent with the pseudopotential approximation. No higher partial waves are expected to be required in general.

Chapter 5

Core hole effects

In this chapter the planewave pseudopotential approach to the calculation of Energy Loss Near Edge Structure (ELNES) is applied to a selection of well characterised systems. Preliminary calculations for these systems found that the straightforward application of the "sudden" approximation (see below) leads to poor agreement with experiment close to the threshold. The correction of these deficiencies has required a more detailed look at the excitation process, and to the inclusion of so-called "single particle" core hole effects, namely the effects due to the interaction between the core hole and the excited electron which can be treated by single particle electronic structure methods.

The technical developments in Chapters 3 and 4, and the use of the planewave pseudopotential approach have allowed the systematic study of single particle core hole effects. Accurate Brillouin zone integrations and ELNES matrix elements, together with the robust convergence properties of the planewave pseudopotential approach (see Appendix A) permit deviations between the theoretical spectrum and experiment to be attributed to physical effects omitted in the theoretical description as opposed to artifacts of the calculation. The importance of this has been pointed out by Rez *et al* [81].

In Section 5.1 the physics underlying the single particle core hole effects is introduced, and then in Section 5.2 the details of performing the supercell calculations are described. The result of core hole calculations for Diamond, Graphite and cubic Boron Nitride are presented in Section 5.3. Greatly improved agreement with experiment is found in comparison with calculations performed within the sudden approximation.

5.1 Single particle core hole effects

This section discusses the physics behind the single particle core hole effects, and some background to the investigation.

5.1.1 What is not calculated?

It is important to distinguish the effects of the interaction between the core hole and the ejected electron that are treated in the approach to be presented from those which are ignored. All many-body effects, deriving from the coupling of the core hole and the excited electron which lead to multiplet structure [103] are ignored in this single particle approach. For example, this precludes the explicit calculation of the splitting in the L2,3 edges and the non-statistical ratios in intensity between transition metal white lines. The current investigation is simplified by placing the emphasis on K-edges, for which transitions originate from 1s core states. Initial attempts to combine bandstructure and multiplet calculations are described by Moreau *et al* [66].

5.1.2 The excitation process — a consideration of timescales

There is no hope of following the actual time dependent evolution of the core electron into its final excited state — which would, in principle, be the correct approach to the calculation of absorption spectra. As in the case of *ab initio* molecular dynamics where the electron and ionic co-ordinates are decoupled (in the Born-Oppenheimer approximation [13]), considerable simplifications can be made if the timescales for the various responses of the system can be separated.

As the core electron evolves from its initial to final state a number of things could be expected to happen to the system. Having removed an electron from the core, the electronic and ionic system will no longer be in their ground-state, and both will experience forces. Clearly, in any model of the core hole effects it is crucial to determine which parts of the system must be relaxed before attempting to calculate the ELNES. So, it is important to know on what timescale the excitation occurs, and how long the respective relaxations of the electronic and ionic systems take.

The time taken for the excitation of the core electron into its final state can be estimated from a knowledge of the approximate size of the final state and corresponding velocity of the excited electron. Assuming the final state to be of the order of 1nm in size, and that the velocity is close to the Fermi velocity at the threshold (i.e. about $10^6\mathrm{ms}^{-1}$) an excitation time is estimated to be of the order of $\tau_{excite} \sim 10^{-15}$s. The screening time for the electronic system is related to the plasmon energy, since it is the plasmons which give the response of the electronic system to a sudden "kick" — in this case the removal of a core electron. So, $\tau_{screen} \sim \hbar/E_p \sim 10^{-17}$s. The lattice responds much more slowly (being more massive), and the timescale is related to the Debye temperature — $\tau_{lattice} \sim 2\pi\hbar/k_b\Theta_D \sim 10^{-14}$s.

It can be seen that these timescales are well separated. In the time taken for the excitation of the core electron to take place the electronic system has time to respond, but the more massive lattice is essentially fixed in its groundstate configuration. This is good news from the perspective of using ELNES as a probe of microstructure. It would be unfortunate if the process of the measurement disorted the very microstructure that is of interest. In the remainder of this chapter, the extent to which the valence electrons can screen the core hole is investigated.

5.1.3 A model of the single particle core hole effects

In this work the screening of the core hole by valence electrons is assumed to be instantaneous, and the excitation from the core state is assumed to be fast enough that the lattice has no time to relax — the Frank-Condon principle [52]. The single particle core hole effects are modelled by removing a single electron from the core state from which the electron is excited and then allowing the valence electrons to relax to their groundstate in this new ionic potential. The core hole will then be screened in this approximation. The degree of the screening depends on the zero frequency dielectric response of the material. For a metal the screening will be almost complete, and the core hole effects on the unoccupied states will be minor. For semiconductors the screening will be good, getting progressively worse for insulators and so the core hole effects are expected to become more important. Techniques exist for dealing with the two extremes of weak and strong core hole effects and these are discussed in Section 5.1.5. For systems with moderate excitonic modification, neither perturbative nor atomic approaches are appropriate. In these situations the effects of the core hole on the unoccupied electronic structure must be explicitly calculated.

On inclusion of the core hole effects the absorption threshold is expected to be altered in the following way. As the core electron is ejected the nuclear charge will be partially unscreened. This has the effect of pulling states in the conduction band down towards the threshold energy, and causing a contraction of the wavefunctions at the threshold. Both these effects will increase the spectral weight at the threshold. In extreme cases, bound excitons will be formed, and may be pulled down below the continuum threshold to form discrete excitonic resonances.

5.1.4 Core holes in the presence of the swift electron

The situation is further complicated in the case of EELS due to the effect of the swift electron which causes the excitation of the core electron. On studying the differences between the Carbon K-edge of Diamond measured by EELS and X-ray Absorption Spectroscopy, Batson and Bruley [5, 7] concluded that the swift electron alters the potential seen by the excited electron. The core hole enhancement of the threshold is reduced in an EELS experiment. In fact, calculations of Silicon L2,3 edges by Weng *et al* [102] required no excitonic enhancement at the threshold to obtain agreement with EELS data, while it is clear that there should be some weak excitonic modification in the absence of the swift electron since Silicon is a semiconductor and the screening will be imperfect.

5.1.5 Other approaches to including core hole effects

For weak excitonic effects a "perturbative" approach such as the Clogston-Wolff method [37, 90, 95, 100] gives very good results. In this approach the effect of the core hole is parameterised by two numbers — the depth of the core hole (in energy) and the degree of hybridisation between the orbitals introduced by the core hole. These two parameters define a transform between the unperturbed symmetry projected DOS and the modified DOS in the presence of the core hole. Initially, it was hoped that this method could be extended to the case of moderate excitonic modification. If the two parameters could be found from a single first principles calculation and then turned out to be transferable, the core hole effects could have been dealt with straightforwardly. However, it turns out that in the case of moderate modification the hybridisation parameter is a strongly varying function of energy and so this approach failed.

It is possible to include core hole effects in the real space Green's function based multiple scattering approaches to the calculation of ELNES or XANES. The potential at the site of the excitation need simply be modified for the core hole. However, it is well known that the XANES approach fails at the threshold for many materials [61, 101, 105] — exactly where the core hole effects are expected to be strong.

For very strong core hole effects, the final states are highly localised and the problem essentially becomes atomic. In such situations, the thresholds can be modelled using atomic multiplet theory (to take the many-body effects of the interaction between the core hole and excited electron into account), including a crystal field to describe the perturbations introduced by the local atomic environment [96].

5.2 Performing a core hole calculation

This section describes the details of performing a core hole calculation, using the planewave pseudopotential approach developed in this thesis.

5.2.1 The supercell approximation

Just one core electron at particular site is excited at a given time. This breaks the translational symmetry of the crystal. Since the planewave approach employs periodic boundary conditions the supercell approximation must be applied to the calculation of core hole effects. The supercell is constructed by forming a large cell of the bulk material, and then choosing one of the atoms in each cell to be excited (see Section 5.2.2 below). The core hole is then screened by the valence electrons by performing a self-consistent calculation for the charge density. This is then followed by a bandstructure calculation for the unoccupied Kohn-Sham eigenstates (up to the energy to which the spectra is required). The spectra is calculated by forming a density of states projected by ELNES matrix elements. However, the calculation is performed for the excited site only. A new supercell calculation must be repeated for each inequivalent site in the supercell. The supercell should, in principle, be increased in size until the spectra converges — at this point the excited atoms in neighbouring cells will have ceased to interact. This interaction between neighbouring excited atoms is equivalent (in that it derives from the inability to deal computationally with an infinite number of atoms) to the problems cluster calculations have due to finite size effects.

It should be noted that there is no increase in computational cost for including core hole effects in the calculation of the ELNES for isolated impurities in a bulk system. The supercell will already have been converged with respect to interaction between neighbouring impurities, and so the core hole effects are introduced by simply replacing the potential of the impurity to an excited one.

Ahuja *et al* [1] recognised that the use of a supercell restricts the number of electrons available for screening (each cell has its own core hole to screen). A possible way to accelerate supercell convergence is to inject some extra electrons into the system. However, it appears unclear how to decide how many electrons would be needed from first principles, so this approach is not taken here.

5.2.2 Modelling the excited atom

The single particle core hole effects are modelled by performing an electronic structure calculation for the perturbed electronic states. These new electronic states are obtained by treating one of the atoms in a supercell as excited, lacking the core electron that has been promoted to the conduction band. This excited atom could be described in a number of ways as follows:

The Z+1 approximation

One could recognise that the removal of a single electron from a tightly bound core state can be approximated by increasing the nuclear charge by one. This is the so called "Z+1" approximation, and is widely used.

Constrained all electron calculations

Using an all electron electronic structure code it is straightforward to constrain the occupation of one of the core states on the excited atom to be reduced by one.

The pseudopotential approximation to the core hole

In the case of the pseudopotential approach used in this work the core states are not explicitly calculated. Their effect on the electronic structure is described by the pseudopotential. So, to model the excited atom an excited pseudopotential must be specially constructed, reducing the occupation of the core level of the reference configuration by one [59]. Two different excited pseudopotentials for Carbon were generated. The C*I potential is constructed using the atomic configuration $1s^1 2s^2 2p^2$, i.e. the ejected electron is assumed to be fully removed, while the C*II potential is constructed with the atomic configuration $1s^1 2s^2 2p^3$, the ejected electron promoted to the 2p level.

Comparison of the Z+1 and pseudopotential approximations

Within the frozen core approximation, the excited pseudopotential approach and constrained occupation in the all electron calculation are both expected to lead to accurate results. The Z+1 approximation is less accurate — it attempts to cancel out one of the core electrons by placing an additional charge at the nucleus, and so clearly the spatial variation of the resulting potential will be incorrect. For instance here is no distinction

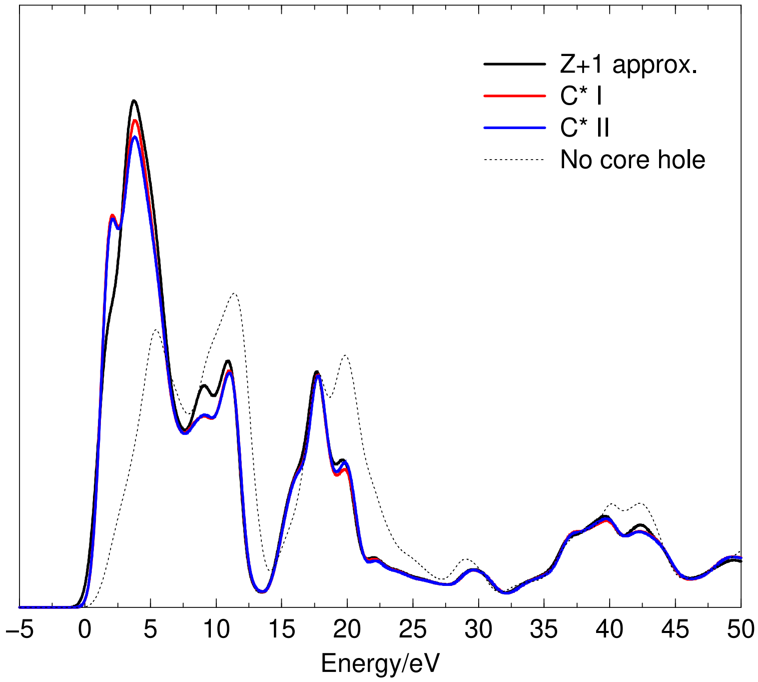


Figure 5.1: Comparison of the Z+1 approximation and two excited pseudopotentials for the K-edge of Diamond. A 16 atom supercell was used, and a 0.5eV Gaussian smear applied to the resulting spectra.

between 1s and 2p core holes in the Z+1 approximation. In Figure 5.1 a K-edge of Carbon in Diamond is shown using the Z+1 approximation (i.e. a Nitrogen pseudopotential), and the two different excited pseudopotentials, C*I and C*II. It is clear that the choice of the potential does not change the location of the peaks in energy, but it does effect the intensities of those peaks. The results for C*I and C*II are almost identical, and for the rest of this work potentials of type I are used. There is no advantage in using the less accurate Z+1 approximation since the excited pseudopotentials are straightforward to generate. However, it is clear that the Z+1 approximation does yield useful results, and so there is no need to dismiss the value of previous calculations based on it.

5.2.3 What to do with the excited electron?

During the excitation process the ejected core electron does not vanish from the system. In some average and time dependent way, it will contribute to the screening of the exposed nuclear charge. Given the current time independent approach to the absorption process,

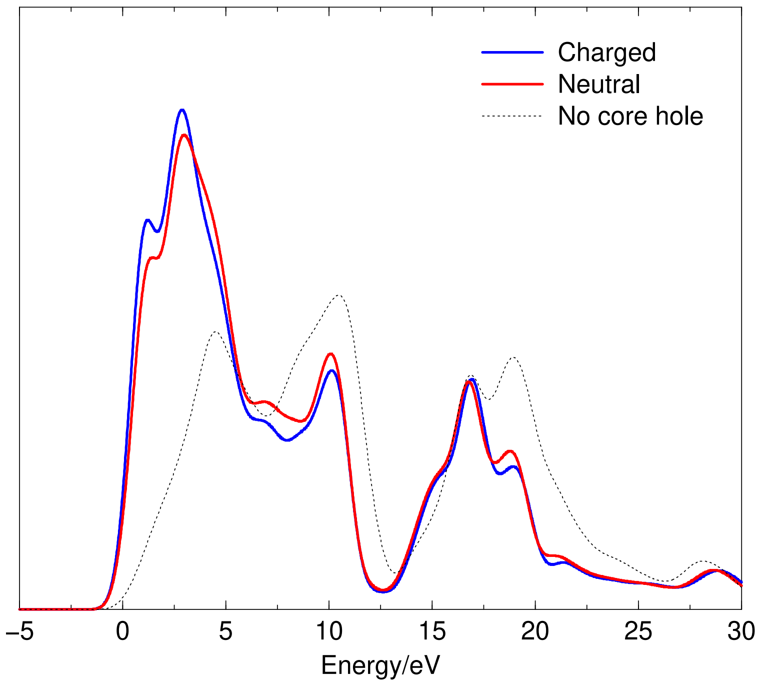


Figure 5.2: Charged vs. neutral supercell calculation for a 16 atom Diamond cell. A 0.5eV Gaussian smearing was applied to the resulting spectra. In the neutral cell, the excited electron is included in the self-consistent calculation of the charge density.

some approximation must be made. The two simplest approximations are either to entirely exclude the excited electron from the calculation, resulting in a charged supercell for an originally neutral system[1] or to include it as an extra valence electron, and which is thus able to participate fully in the screening of the core hole, resulting in a neutral supercell. These can be thought of as two extreme cases. In the first case there is no screening due to the excited electron. In the second case, the electron is allowed to relax to the ground-state in the self consistent calculation of the charge density, presumably close to the core hole, and so is expected to screen it highly efficiently. In most cases, the final states will extend over a relatively large region of space, and so less than full screening would be expected in practice.

In Figure 5.2 the results of charged and neutral supercell calculations for Diamond are compared. The results are broadly very similar and the energies of the peaks change very slightly. However, the intensities of some features are observed to change. As the extra

[1]In fact, in CASTEP charged supercells are made charge neutral by the addition of a compensating background charge.

screening electron is introduced in the neutral cell, the strength of the core hole is reduced and so the final states are less confined, leading to a reduction in overlap in the transition matrix element. Hence, the threshold intensity is expected to be reduced for the neutral cell, as observed in the calculations. At higher energies (20eV from the threshold and above) the two calculations agree closely.

The fact that the two alternative approaches give results that are so close to each other, and that both are significantly different from the calculation performed without a core hole suggests the core hole effects in the ELNES are not greatly sensitive to the detailed theoretical description of the screening due to the excited electron, at least for systems that experience moderate excitonic modification. The "correct" result is expected to lie somewhere between the two curves shown in Figure 5.2 and more likely closer to the result for the charged supercell, since the final states are relatively extended. For simplicity, in the rest of this work the excited electron is not included in the self consistent calculation of the charge density. This avoids the need to deal with partially filled bands. It would be expected that the calculated thresholds for systems with very strong excitonic modification would require a much more careful treatment of the excited electron. Also, as ever more quantitative calculations are required, further attention will have to be given to the role of the excited electron, even in systems such as Diamond.

5.3 Application to bulk systems

In this section the results of single particle core hole calculations for a selection of materials are compared to experimental electron energy loss spectra. The investigation is confined to materials in which the excitonic modification is expected to be significant, and so could not be treated by simple models for weak modification, such as the Clogston-Wolff model described in Section 5.1.5, but not extreme, in which case a more detailed understanding of the excitation process is required, or an atomic multiplet approach should be used. In Section 5.3.1 the K-edge of Carbon in Diamond is calculated and then in Section 5.3.2 Carbon is again studied, but this time in Graphitic form. This is a more complicated situation since the material displays a large degree of anisotropy in the electronic structure. Finally, in Section 5.3.3 the Boron and Nitrogen K-edges in cubic Boron Nitride are examined.

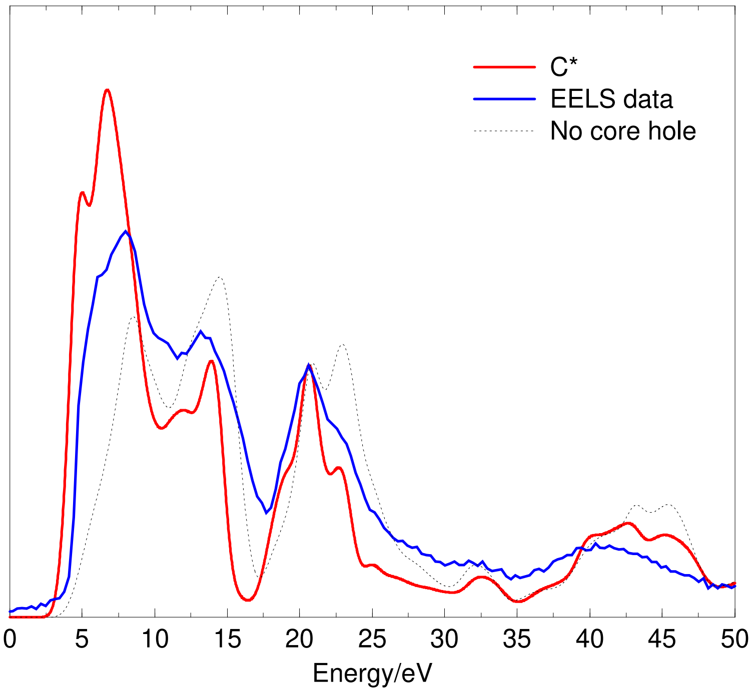


Figure 5.3: Carbon K-edge in Diamond: The red curve is the calculated spectra for a supercell of 16 atoms (2x2x2 primitive cell), one of which had an excited pseudopotential. A 0.5eV Gaussian smearing is applied to the calculated results. An arbitrary energy shift and scaling is applied to the experimental data to aid comparison.

5.3.1 The K-edge of Carbon in Diamond

Figure 5.3 shows the Carbon K-edge of Diamond, calculated in both the sudden approximation and the supercell approximation for the core hole. An experimental EELS spectrum is also shown for comparison[2].

The unoccupied electronic structure of Diamond may be thought of as comprising two thresholds. In addition to the true threshold, there is effectively another threshold 17eV higher in energy, where the density of states drops substantially. The sudden approximation reproduces the main features of the experimental spectrum up to at least 40eV. However, the relative weights of the peaks at 7 and 14eV on the first threshold and at 20 and 23eV on the second are incorrect. In both cases, the lower energy peaks experimentally have the greater weight experimentally, but the opposite is true in the calculation.

Including the core hole effect significantly improves the appearance of the two thresh-

[2]The Carbon K-edge in Diamond was taken by Brian Rafferty.

olds. In both cases the relative weights are reversed, so they are now in agreement with experiment. On the first threshold, clearly a new feature appears at the onset. It apparently correlates with a similar feature in the experimental spectra. However — Diamond is known to have an excitonic peak at the threshold. It is not visible in this spectra since it proves hard to separate from the zero loss peak in EELS measurements, but Batson [7] has successfully extracted it and compared it to X-ray absorption data. So, it is not clear whether this peak in the calculation is the exciton, not seen in this experimental data, or the feature that is seen. Similar calculations by Mauri and Car [59] did produce a bound exciton.

At slightly higher energies, between the first and second peaks of the first threshold, at about 10eV, another feature appears on inclusion of the core hole. This feature is also seen in the experimental spectrum. The shape of the second threshold, which is clearly made up of several peaks in the experimental EELS, is well modelled once the core hole is included.

Some attention should now be paid to the ways in which the calculated spectra deviate from the experiment. It does appear that the weight at the threshold is somewhat overestimated. However, as mentioned in Section 5.2.3 it is not yet clear the extent to which the excited electron should be allowed to screen. Also, Batson expects the swift electron, which causes the excitation in an EELS measurement, to contribute to the screening (see Section 5.1), again reducing the threshold intensity.

In summary, the inclusion of the single particle core hole effects substantially alters the K-edge of Carbon in Diamond, well into the conduction band. It provides a significant improvement over the sudden approximation, which itself reproduces many of the significant features in the spectrum up to at least 50eV into the conduction band.

5.3.2 The K-edge of Carbon in Graphite

Carbon is found in many stable forms. Above, the Carbon K-edge in Diamond was investigated. In this section the Carbon K-edge in Graphite will be examined.

An introduction to the electronic structure of Graphite

In Diamond the bonding is sp^3, leading to a three dimensional tetrahedrally co-ordinated structure. In Graphite the bonding is sp^2 and the structure is that of hexagonal sheets of Carbon atoms with the sheets weakly bound to each other. It is a highly anisotropic

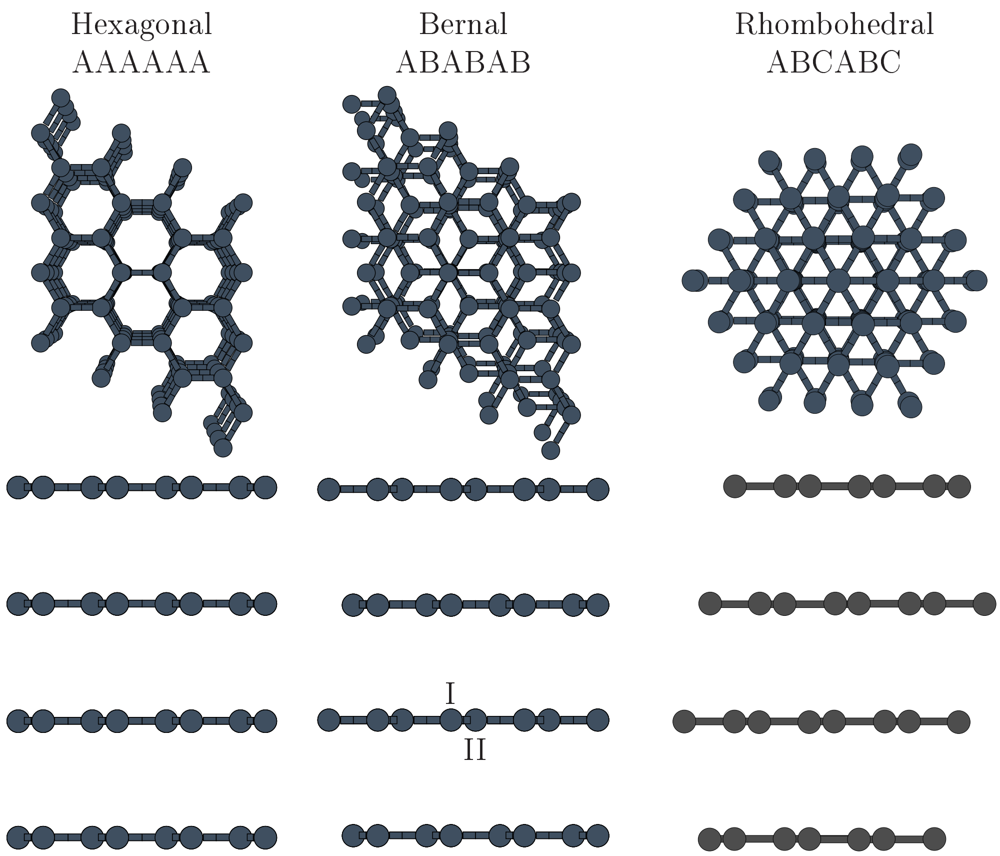


Figure 5.4: The three structures of Graphite. Bernal Graphite is the most commonly found, and has two inequivalent sites, labelled I and II, while in Hexagonal and Rhombohedral and Graphite all sites are equivalent.

structure and gives rise to a similarly anisotropic electronic structure which is reflected in the spectral properties.

Three possible structures for Graphite are presented in Figure 5.4. They differ in the stacking of the weakly bound sheets. Graphite is most often found in the Bernal form, and of the three types it is the only one in which there are inequivalent sites for the Carbon atoms. In the diagram they are labelled (arbitrarily) as Type I and II sites. The Type I sites have Carbon atoms above and below in the neighbouring sheets, while the Type II sites sit between two empty sites on the triangular lattices.

Figure 5.5 shows the variation in the electronic structure indicated by the total density of states for the three different forms of Graphite. The first two peaks in the DOS, at 2

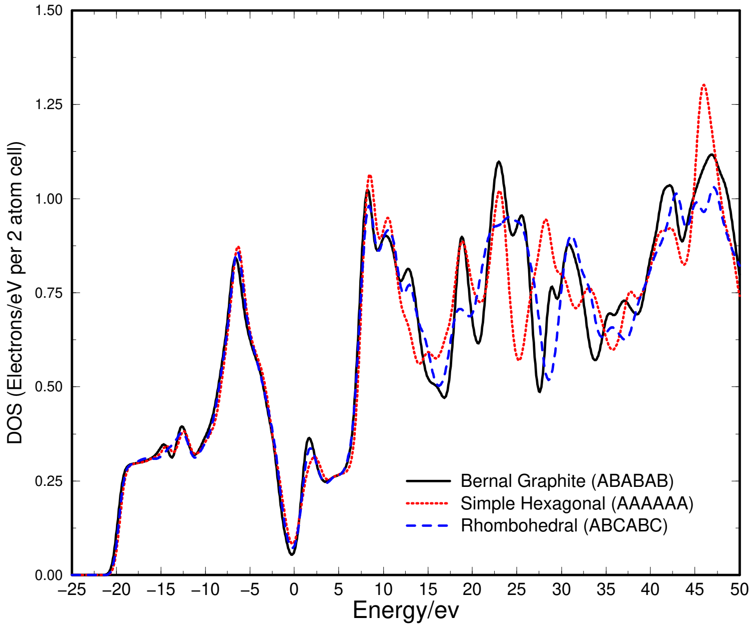


Figure 5.5: Total Density of States for Graphite for the three stacking geometries, smeared by a 0.5eV Gaussian. Note that, while very similar below E_f, there are differences at higher energies.

and 8eV, are known as the π^* and σ^* peaks respectively. The π^* peak derives from the anti-bonding branch of the band formed from the out of plane p_z orbitals, and the σ^* peak from the in-plane sp^2 hybrids. The DOS below the Fermi energy is very similar for all three forms. Above the Fermi energy there are noticeable differences in the DOS of the three forms. The π^* peak of the Simple Hexagonal Graphite is clearly broader than for the other two structures, and at even higher energies there are quite significant differences between the DOS of the three structures. It might be expected that an experimental investigation of the unoccupied DOS will easily distinguish between the various stacking geometries in Graphite (see below for a discussion in the context of the inclusion of the core hole).

Between the π^* and σ^* peaks a weak feature can be seen in the DOS. This feature is due the free electron like interlayer states, which are of great importance in the intercalation of alkali atoms into Graphite [22, 40, 43, 77]. This region of the DOS appears to have the $E^{1/2}$ rise associated with free electrons in three dimensions.

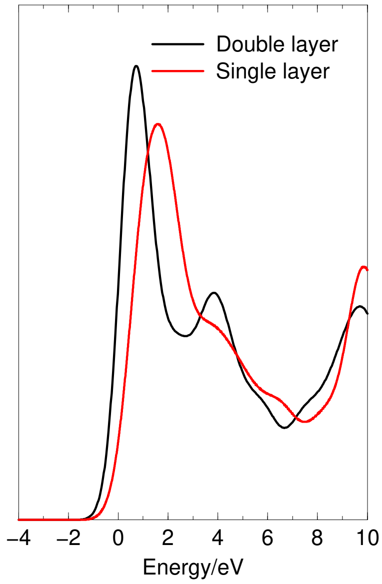


Figure 5.6: The π^* peak of Hexagonal Graphite with one and two layers in a supercell, 8 carbon atoms per layer, and a single excited atom. Clearly at least a double layer is required for convergence.

The inclusion of core hole effects

Supercell calculations for the K-edge of Carbon for each of the three structures were performed. Supercells with 8 atoms in a plane were used, giving a separation of 3 or 4 bond-lengths between excited atoms. Work by Ahuja *et al* and others [1, 60] suggests that the excitonic states are contained within 5 bond-lengths of the excited atom and so this supercell should be sufficient. Figure 5.6 shows the importance of including a least one hexagonal sheet between the sheets that contain excited atoms for the π^* peak.

In Figures 5.7 the black curves correspond to momentum transfer perpendicular to the graphitic planes, and the red curves to momentum transfer in-plane. The solid curves have been smeared by a 0.25eV Gaussian, while the dashed curves are unsmeared. The π^* and σ^* peaks are labelled. On the left, the K-edge calculated within the sudden approximation is presented. On the right are the results of the supercell core hole calculation, and experimental data is shown in blue[3]. The experimental spectra have been aligned with the σ^* peaks for comparison since the absolute energy of the threshold is not calculated.

In all cases the shape of the π^* peaks is improved on inclusion of the core hole becoming

[3]The experimental EELS for the Carbon K-edge of Graphite was taken by Batson and is for the “large” collector aperture setting described in Ref. [4].

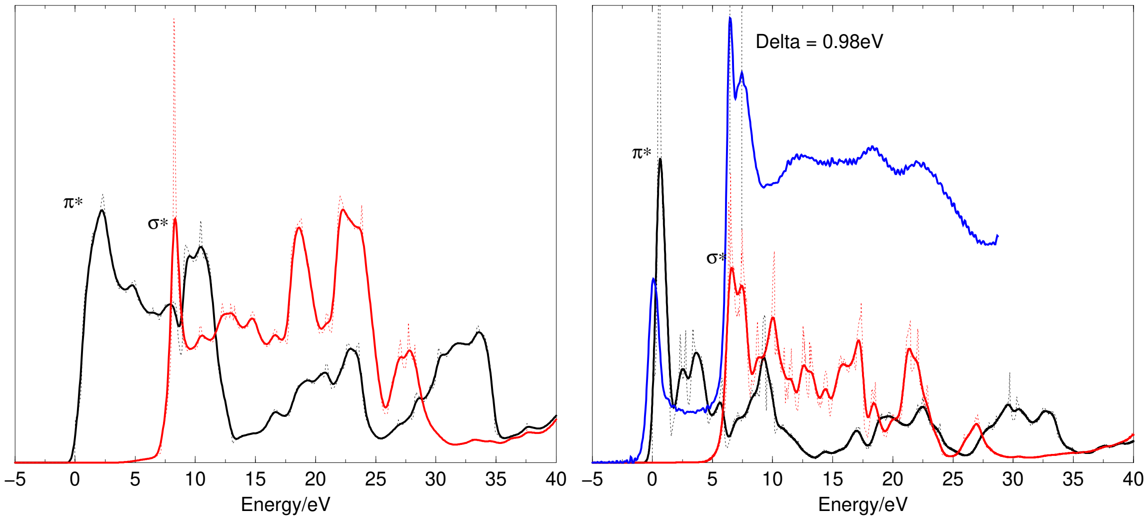


Figure 5.7: K-edge of Carbon for Simple Hexagonal Graphite. *Left:* The sudden approximation. *Right:* Core hole included in a 16 atom supercell — 8 atoms in a plane, and two layers.

sharper and moving down in energy towards the threshold. The σ^* peaks calculated within the sudden approximation are all single sharp peaks. In contrast the experimental spectra shows a splitting of the σ^* peak by about 1eV. This is reproduced in the core hole calculations. The origin of this double σ^* peak is controversial [1, 4], but these results would suggest that it is simply due to single particle core hole effects.

Above the σ^* peak the inclusion of core hole effects substantially improves the appearance of the calculated spectra. In the sudden approximation, there is a strong feature directly above the σ^* peak with the same symmetry as the π^* peak. This would lead to an extreme anisotropy in the spectra at this energy which is not observed experimentally [4, 53]. The core hole calculation reduces this feature, replacing it with a peak with σ^* symmetry as seen in the experimental spectra, albeit at too low an energy. At higher energy still, the features are consistently reproduced on inclusion of the core hole.

The localising effect of the core hole

Figure 5.8 compares the Carbon K-edges for Bernal and Rhombohedral Graphite calculated both within the sudden approximation and including core hole effects. In the sudden approximation there is a clear change in the intensity of the peak located at 19eV between the Bernal and Rhombohedral structures. However, the two spectra are practically indis-

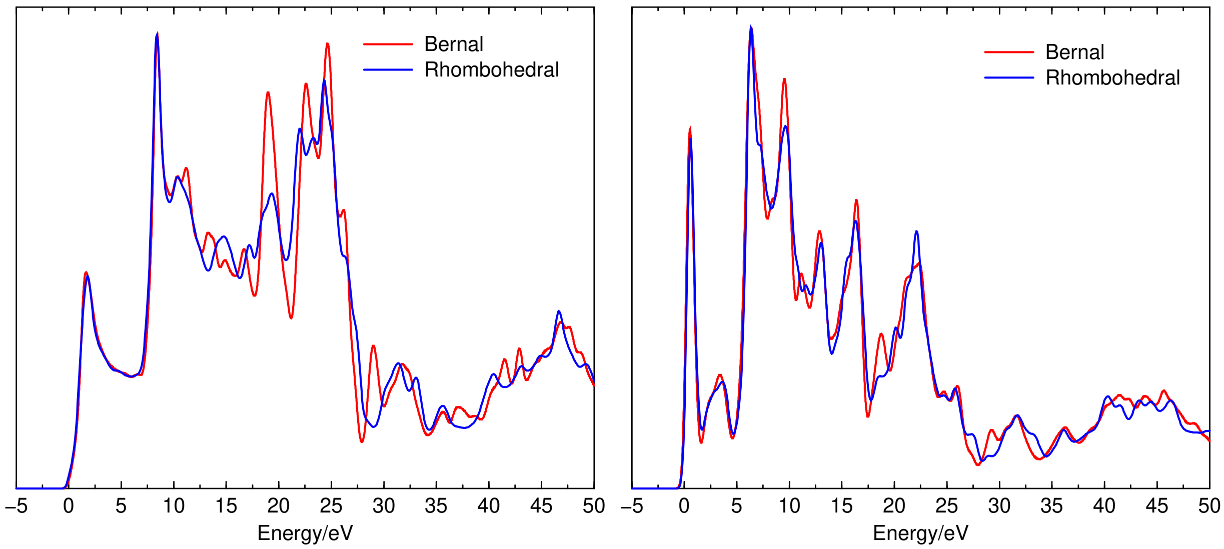


Figure 5.8: Bernal and Rhombohedral Graphite — with and without a core hole. The core hole calculation used 16 atom supercells. The spectra were averaged over all momentum transfers and smeared by a 0.25eV Gaussian. For Bernal Graphite a sum is performed over the Type I & II sites. *Left:* Within the sudden approximation — no core hole effects *Right:* Core hole effects included in the supercell approximation

tinguishable on including the core hole effects. This can be interpreted as the localising effect of the core hole potential — the final states do not explore such great distances under the influence of the attractive core hole and so the variable stacking of the Graphitic planes make less impact on the spectra. This indicates that studies of the unoccupied electronic states using core loss spectroscopy will display less obvious differences between the various Graphite structures than might be expected from straightforward calculations of the thresholds within the sudden approximation.

A puzzling peak

So far no comment has been made concerning the feature which appears in all the core hole calculations between the π^* and σ^* peaks at about 3eV from the threshold. It is clearly not seen in the experimental EELS spectrum presented here. In fact, such a feature, to the best of the author's knowledge, has not been observed in any EELS measurement of Graphite. However, as pointed out by Batson [4] it is often seen in X-ray absorption spectra. It has sometimes been attributed to impurities in the Graphite (by analogy to the C-H resonances found in hydrocarbons) [24], but it has also been attributed to interlayer

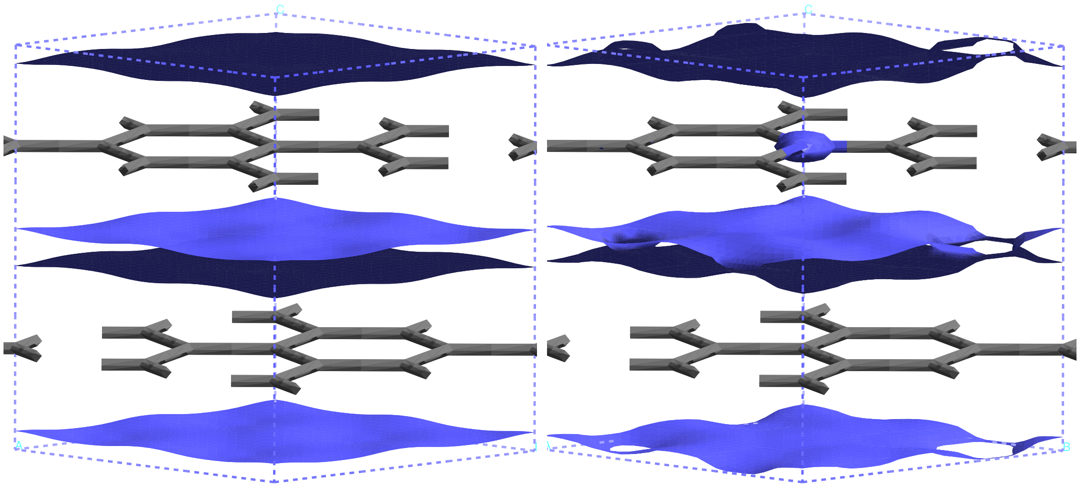


Figure 5.9: Coupling of the interlayer states to the excited atom in a 16 atom supercell of Bernal graphite. The unoccupied 39^{th} band at the Γ-point is shown as a density isosurface (at 0.03Å^{-3}) (the first 32 bands are occupied) *Left:* No core hole potential at 2.1eV above the lowest lying unoccupied state at the Γ-point. *Right:* Core hole potential on a single site at 3.7eV above the lowest lying unoccupied state at the Γ-point. Note, charge is depleted from the interlayer region, and collects at the excited site.

states [32]. It might appear impossible for interlayer states to contribute to the core loss spectrum. By their nature, the weight of the interlayer states is located between the layers and so there will be very little overlap with the core states which are localised in the planes. However, Figure 5.9 shows how one of these interlayer states (at an energy close to the feature concerned) is altered on the inclusion of a core hole. Clearly, the weight of this state is increased at the site of the excited atom, allowing for the possibility of transition into the interlayer state. Batson and Bruley [5, 7] point out possible differences between X-ray and EELS measurements due to the passage of the swift electron in EELS and the feature is not observed in EELS measurements. It will be interesting to investigate whether this feature is highly sensitive to details of the excitation process. The present results show it is not sensitive to the details of the Graphite structure.

Comparison of results with previous calculations

Supercell calculations of the K-edge of Graphite have also been performed by Ahuja *et al* [1]. While the calculations presented here agree that the inclusion of the core hole effects are important in the interpretation of the 1s absorption spectrum of Graphite, it is clear that a unified picture of those effects has not been reached. Ahuja uses full potential linear muffin-tin-orbitals (FPLMTO), supercells of up to 50 atoms per sheet and includes an extra electron "for screening". This extra electron is apparently to be supplied by the surrounding supercells, and does not model the screening effect of the excited electron. The effect that the inclusion of this extra electron has on the calculations in this work are shown in Figure 5.10. Their study was confined only to the study of a Graphene sheet — the claim being made that the main features of the electronic structure are determined by in-plane effects. This may be true below the Fermi level (as noted above) but it becomes progressively less true at higher energies. Also, only the π^* and σ^* peaks are discussed by Ahuja, and so it is a less general study of the absorption spectrum in comparison to the one presented here. Ahuja *et al* find the π^* peak to be split, which contradicts the result found here. This is most likely due to an insufficient Brillouin zone integration in the FPLMTO calculation. The splitting of the π^* peak is also seen in their ground-state DOS. The feature between the π^* and σ^* peaks is not present in the FPLMTO calculation nor, apparently, in the X-ray data presented in the same paper, but no explicit mention is made of that region of the spectrum. Finally, Ahuja *et al* do not find the splitting of the σ^* peak as presented here. This could be due to the restriction to Graphene, since the splitting is shown in our calculations (not shown) to be less extreme for Graphene than the other three dimensional Graphite structures. But as it could also be that the splitting would disappear for larger supercells, this is an area for future work. The interpretation of the σ^* peak remains controversial, but it is clear that tools are being developed that should help to resolve this question.

5.3.3 The K-edges of Boron and Nitrogen in cubic Boron Nitride

Boron Nitride (cBN) has the same zincblende structure as Diamond but with each of the Carbon atoms alternately replaced by Boron and Nitrogen. Boron has Z=5 and Nitrogen Z=7, hence cBN is isoelectronic to Diamond. The differences between cBN and Diamond can be thought of as arising from the introduction of an alternating potential which splits the many degeneracies present in bulk Diamond.

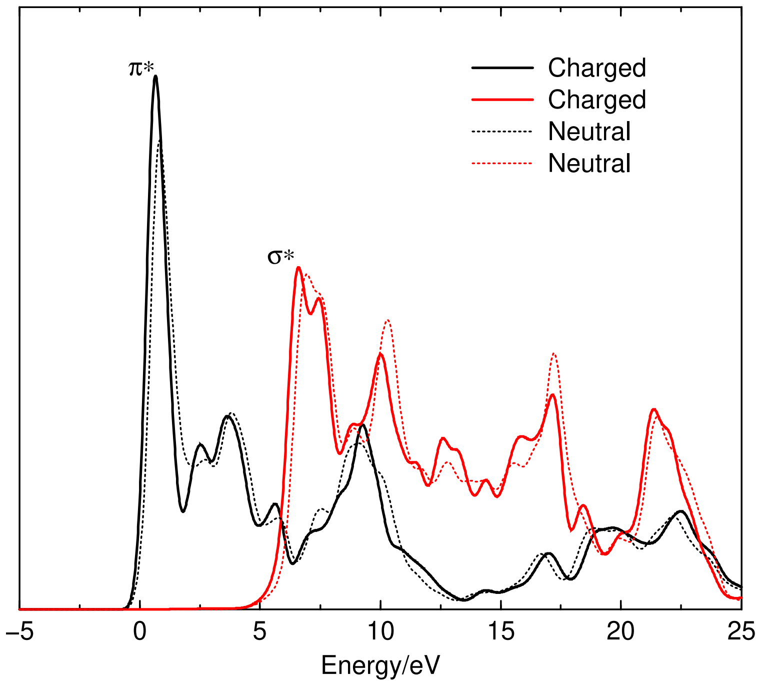


Figure 5.10: Charged and neutral supercell calculations for Hexagonal Graphite. Adding an extra electron to the 16 atom supercell makes little difference, apart from the expected small reduction in the excitonic effects.

Figure 5.11 shows calculated and experimental[4] Boron and Nitrogen K-edges in cubic Boron Nitride. As in the case of Diamond, the spectra calculated within the sudden approximation show many of the features present in the experimental spectra. At energies high above the threshold, above 20eV, the agreement is very good, even small features up to 60eV from the threshold are reproduced. However, nearer the threshold the experimental and calculated spectra deviate.

On performing supercell calculations to include the effects of the core holes (one for the Boron K-edge and one for the Nitrogen K-edge) remarkable agreement with experiment is observed. The weight and location of peaks up to about 30eV above the threshold is adjusted in both cases. At higher energies the calculations for the sudden approximation and core holes are extremely close — and thus the agreement with experiment at high energy is maintained. The weight at the threshold is enhanced and is also shifted to lower energy, as expected. There is more structure in the supercell calculations of the K-edges than observed in the experimental spectra. It is clear that there is some further structure

[4]The experimental spectra were provided by Gilles Hug.

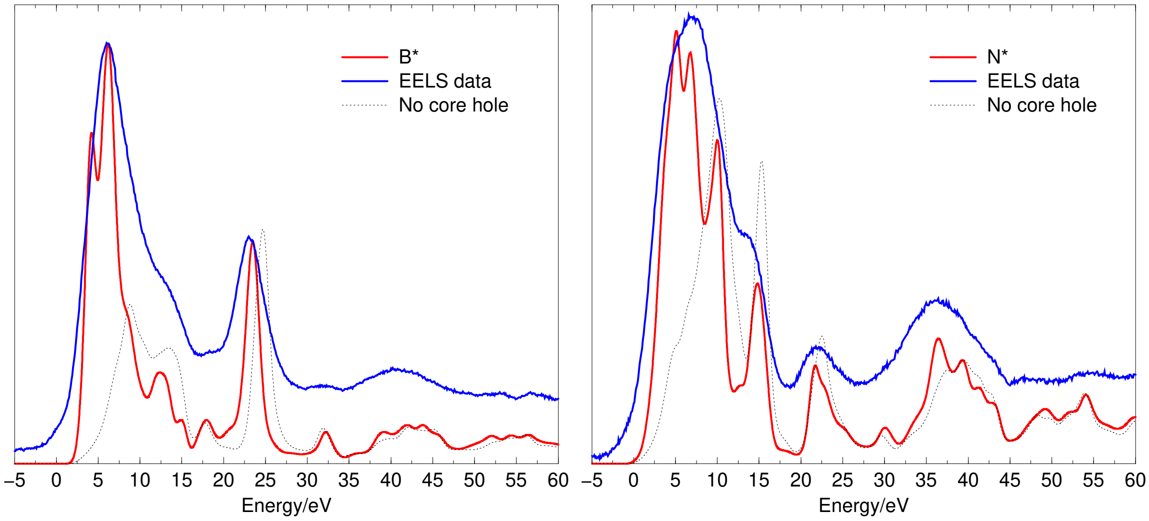


Figure 5.11: Boron and Nitrogen K-edges in cBN. *Left:* Boron K-edge *Right:* Nitrogen K-edge Two calculations were performed in a 16 atom supercell (2x2x2 primitive cell), one with a single excited Boron pseudopotential, and the other the an excited Nitrogen pseudopotential. A 0.5eV Gaussian smearing is applied to the calculated results, and the experimental data is shifted and scaled for comparison.

near the experimental threshold of the Nitrogen K-edge to be resolved.

The ELNES of cubic Boron Nitride is extremely well modelled within the current approach, for both the Boron and Nitrogen K-edges, while the sudden approximation is found to be clearly deficient near the threshold. However, at high energies both the sudden and supercell approximations give very good agreement with experiment.

5.4 Discussion

In this chapter, supercell calculations of core hole effects in ELNES in the planewave pseudopotential approach have been shown give very good agreement with experimentally measured spectra for materials containing Boron, Carbon or Nitrogen in both sp^2 and sp^3 hybridisations. Although there is every reason to suspect that the method will work equally well for K-edges in all systems with moderate excitonic effects, extensions to the systems already examined will provide a myriad of different structures for study in the immediate future. For example Fullerenes (BN as well as Carbon), the isoelectronic substitution of BN for C_2 in Carbon materials and vice versa, and Boron or Nitrogen doping of Diamond.

Work must continue, using larger supercells to ensure that the features observed are not artifacts of spurious interactions between neighbouring supercells. This appears to be especially important in the case of Graphite where there is some controversy as to the nature of many of the features in the spectrum. Once the supercells are converged, more detailed studies of the screening process can be undertaken leading eventually to an attempt to model stronger excitonic effects, for example in the Silicon L2,3 edge of Quartz, an understanding of which will give access to a whole range of silicate minerals. One advantage of studying materials with stronger core hole effects will be that the final states will be very localised, and hence the supercells required for convergence will be smaller. At the other end of the scale, work may be continued uisng simpler techniques in the way that Weijs *et al* [100] found *ab initio* parameters for the Clogston-Wolff method applied to transition metal silicides, a system with weak core hole effects. However, these effects are less localised and would need to be modelled in larger unit cells. Finally, as a long term aim, attempts should be made to combine the bandstructure methods described here with atomic multiplet calculations.

5.5 Conclusion

All the systems studied here have required calculations to be taken beyond the sudden approximation in order to obtain reasonable agreement with experiment in the threshold region of the ELNES. In all cases regions of the spectra affected by the core holes extend to at least 20eV above the threshold which is almost the entire region normally thought of as comprising the near edge structure. The modifications of the thresholds go beyond a simple enhancement of intensity and it has been shown that line shapes are also significantly altered.

Chapter 6

Conclusion

This thesis has described the development of a first principles approach to the calculation of the Energy Loss Near Edge Structure of Electron Energy Loss Spectra measured using Scanning Transmission Electron Microscopes. It is based on an *ab initio* planewave pseudopotential approach to the calculation of electronic structure. It has been shown to give very good agreement with experiment from the threshold up to energies traditionally covered by multiple scattering approaches. The calculations of the imaginary part of the dielectric function, which determines the energy loss at high energies, are quantitatively correct. The approach provides a framework for the investigation of single particle core hole effects, which have been shown to be significant for the systems investigated so far. The technique is not restricted according to the system size, symmetry and atomic constituents. It deals well with both open and close packed systems.

6.1 The planewave pseudopotential method

This approach to the calculation of ELNES was developed using the planewave pseudopotential code CASTEP to evaluate the single particle states. This is a widely used code, experimental electron microscopists are already using it to perform total energy relaxations structures of microstructure. In this context it is expected that the technique described in this thesis will become widely used in the community, and will be used in conjunction with other *ab initio* total energy studies.

The planewave basis set has specific benefits. It has been found that the method is robust with respect to convergence, and the straightforward representation of the wavefunctions has allowed the efficient direct evaluation of the various matrix elements required in this work.

6.2 Brillouin zone integrations

A highly efficient scheme for the evaluation of Brillouin zone integrations has been developed. Using this scheme the main spectral features are obtained with very low **k**-point sampling densities. As well as offsetting some of the computational expense associated with the planewave pseudopotential method, this approach to the Brillouin zone integrations has made it possible to perform previously intractable calculations such as spectroscopic all electron planewave calculations, which proved essential for the purposes of testing the all electron wavefunction reconstruction from pseudowavefunctions.

This new scheme is based on the extrapolation from a single **k**-point, as opposed to the more usual interpolative approaches. Extrapolation avoids difficulties with the allocation of bands, and the associated spurious features in the DOS. The information required for the extrapolation is obtained within second order **k.p** perturbation theory. The use of planewaves leads to a straightforward formulation compared to that required for local basis sets, and corrections can be easily applied to take the non-locality of the pseudopotentials into account. The fact the bands are expanded to second order ensures that the Van Hove singularities are treated exactly. The first Brillouin zone is efficiently divided into subcells, making full use of the symmetries of the system. The resulting piecewise quadratic representation of the bandstructure is analytically converted into the density of states. An attempt is also made to treat the so called “band kissing” problem.

For more accurate integrations it will be important to “pad” the eigenstate spectrum with planewaves (approximate eigenstates at high energies) to ensure the convergence of the perturbation sum. It also remains to take the dispersion of matrix elements into account for a general Brillouin zone integration. Finally, the case of more complex “band kissings” should be investigated, and a more elegant method for their detection is required.

6.3 ELNES matrix elements

The ELNES matrix elements are evaluated directly from the pseudowavefunctions and all electron atomic core wavefunctions. In contrast to a straightforward symmetry projection these matrix elements are quantitative. Errors due to the use of pseudopotentials are corrected using a scheme based on the Projector Augmented Wave approach. It was found that more projectors are required for the correction to apply at high energies, and so it will be necessary in the future to include these. At the same time, pseudopotentials could be generated with reference states in the conduction band. Currently, pseudopotentials are constructed to work well for ground-state properties and hence only for the valence electrons. The energy dependence of the pseudopotentials is currently good, as shown by the good agreement with experiment at high energy, but could be improved. A scheme for obtaining the non-dipole ELNES matrix elements has also been developed.

6.4 Core hole effects

Significant improvements in the spectra at the threshold energy were found after the inclusion of single particle core hole effects, using the supercell approximation and specially generated excited pseudopotentials. Importantly, the single particle core hole effects were found to extend well above the threshold in those cases studied, in contrast to the typical expectation that it is the threshold alone that is modified by the core hole.

A systematic study of the single particle core hole effects within the framework developed in this thesis can now be carried out. In fact, the results presented here show that it should be assumed that the core hole effects are important until proved otherwise. The excitation process should be studied in greater detail, investigating the connection between the bandstructure and atomic multiplet approaches. Also, the differences between EELS and XAS due to the swift electron in EELS should be examined in the light of the feature between the π^* and σ^* peaks of Graphite, seen in XAS but never in EELS.

6.5 Some immediate applications

There will many applications of this approach for the calculation of ELNES, but there are several which could be attempted immediately using the developments described in this thesis.

It was claimed above that these calculations are quantitative but this should be checked experimentally. It would be interesting to determine whether the absolute modification of the cross sections due to the core hole can be determined in this framework. This investigation should be carried out on a well characterised sample so that the experimental and theoretical situations closely match.

So far, the spatial resolution of the STEM has not been matched by theoretical calculations. Such calculations will be straightforward using the current approach. The ELNES for excitations from a number of sites across an interface could be calculated, and compared with experimental scans across such an interface. More advanced studies would involve the detailed modelling of the electron probe.

Finally, the possibility of observing non-dipole effects should be investigated theoretically. Since the non-dipole matrix elements have be shown to be accessible in the current approach, it should be straightforward to calculated the strength of the non-dipole contributions in various materials. A search could be performed for a material which would be most likely to display non-dipole contributions. An experimental measurement of non-dipole effects will clearly set EELS apart from XAS and make the probing of previously inaccessible unoccupied states possible.

6.6 Summary

Good progress has been made towards the theoretical modelling of the Energy Loss Near Edge Structure, and where the current theory has been shown to be deficient the technique presented in this thesis has been shown to give a framework for the investigation of the new physical effects which need to be included. In addition it is expected that this technique will prove valuable to analytical electron microscopists who wish to interpret the near edge structure to solve practical problems in physics and materials science.

Appendix A

Convergence of the spectral properties

While there is extensive experience in the use of CASTEP as a total energy code, there is considerably less concerning its use as a code for the calculation of spectral properties. It is well established, for example, which planewave energy cutoff is required to achieve a specified level of convergence in structural parameters, but it does not follow that this is the same as is required to converge spectral properties.

Since the degree of convergence required will always depend on the resolution of the experiment to which comparison is to be made, rather than make any firm statements, this appendix is intended to give an idea as to how spectral properties might be expected to change as a number of parameters are varied.

The test system in each of the following cases is a primitive unit cell of bulk Diamond, containing just two atoms. The spectral property examined is simply the total density of states. The convergence of the DOS with number of **k**-points has been studied elsewhere (in Section 3.10.2).

A.1 With planewave cutoff

The energy cutoff for the planewave basis set is an important parameter in planewave calculations. It determines the computational cost of the basis set, and hence the size of the calculation that must be performed. One of the significant features of a planewave code is that it has a basis set which can be defined by a single continuous parameter. Thus, convergence with basis set may be systematically examined. In contrast, the degree to which localised basis sets describe the wavefunctions vary discontinuously as the basis is changed.

In the example illustrated in Figure A.1 calculations for three cutoff energies are shown — 300, 500 and 700eV. The pseudopotential used is known to require a cutoff of between 600 and 700eV for the convergence of structural properties. It is clear that the valence band DOS is essentially converged using a cutoff of 300eV — from the perspective of an experimental measurement of that DOS. The structural properties can be significantly changed by a small distortion of the valence band, hence such calculations require a better basis set. However, all the major features in the spectra are present at 300eV, even well into the conduction band. Using a higher planewave cutoff simply shifts those features to their converged energies. This would not be true in the case of a localised basis set. If important angular momentum states are missed, then the corresponding spectral features would either be missing or severely distorted. In contrast, planewaves contain components of many angular momentum states and are not biased to any region in the unit cell. The general shape of a given state will be rapidly determined and the higher energy planewaves simply alter the fine detail. So, when performing a planewave calculation it is unlikely that a missing or incorrect feature can be blamed on an incomplete basis set, although error in the actual energies of the features might be caused by this. Thus it can be concluded that the planewave basis set is robust with respect to changes in its size for the calculation of spectral properties.

A.2 Number of k-points in self consistent calculation

The number of **k**-points included in the self consistent calculation of the ground-state charge density, which is required to define the fixed Hamiltonian for which the bandstructure calculation is to be performed, determines the accuracy to which that charge density is found.

In Figure A.1 it is clear that the DOS is very insensitive to the details in the charge density. Calculations are shown for 1x1x1, 2x2x2 and 3x3x3 Monkhorst Pack grids [65, 70] for the Brillouin zone integration of the charge density. There is only a significant difference between the 1x1x1 and 2x2x2 grids, and since the primitive Diamond cell described here is probably one of the smallest encountered it could be concluded that the 2x2x2 grid should be sufficient for most purposes. This is far too small a grid for a total energy calculation. The total energy is sensitive to the details of the Brillouin zone integration, particularly in

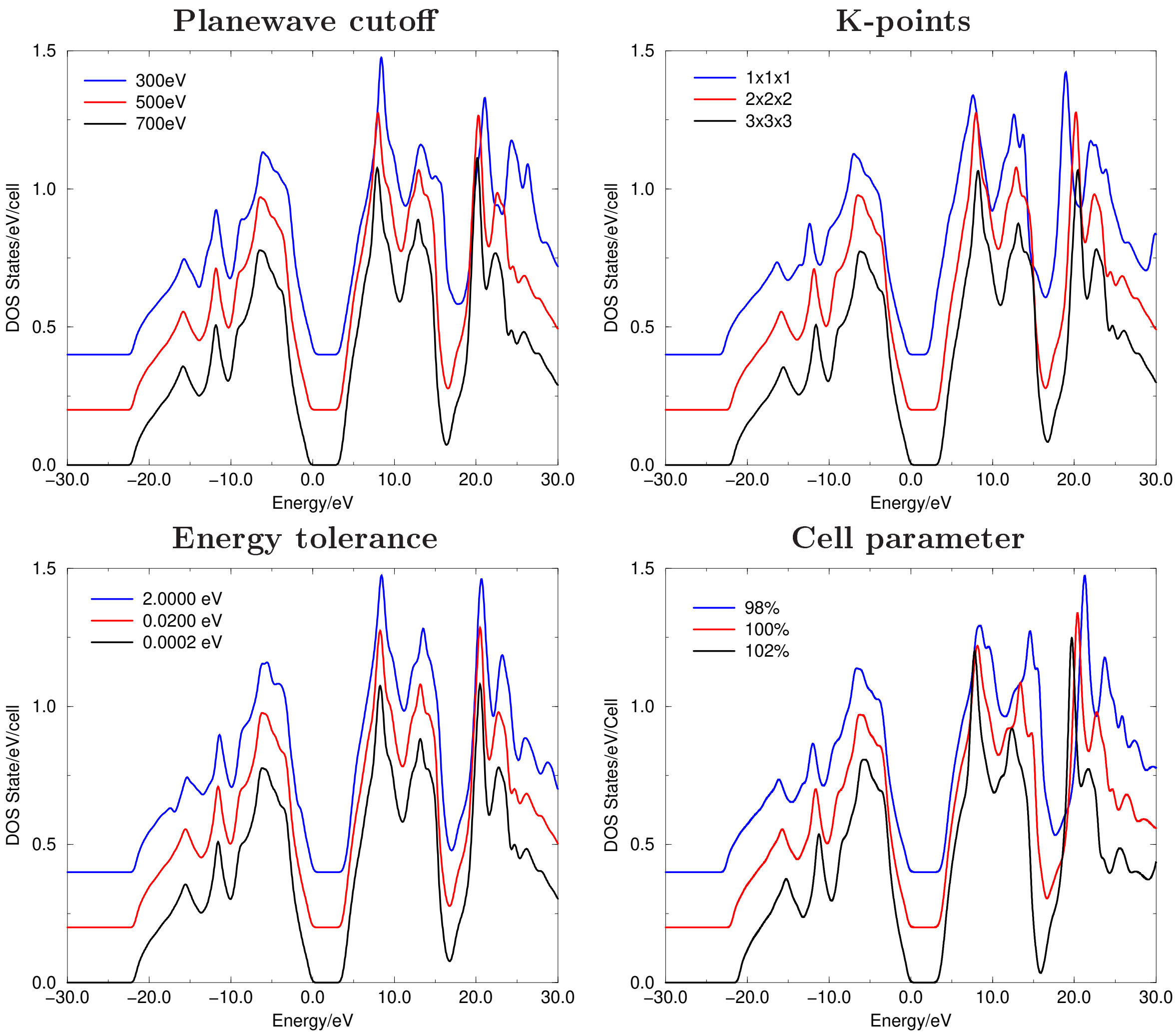


Figure A.1: A series of convergence tests — Total density of states for a primitive cell of Diamond with the Fermi energy in each case aligned and a Gaussian smearing applied of 0.25eV.

the integration of the kinetic energy over the Brillouin zone.

A.3 Cell parameter

In principle there should be no argument as to which cell parameter to use in the calculations since the value calculated by relaxing the system and the experimental value should be the same. However, it is well known that DFT implemented within the LDA or GGA gives good structural parameters, but that they are not exact. Commonly errors are of

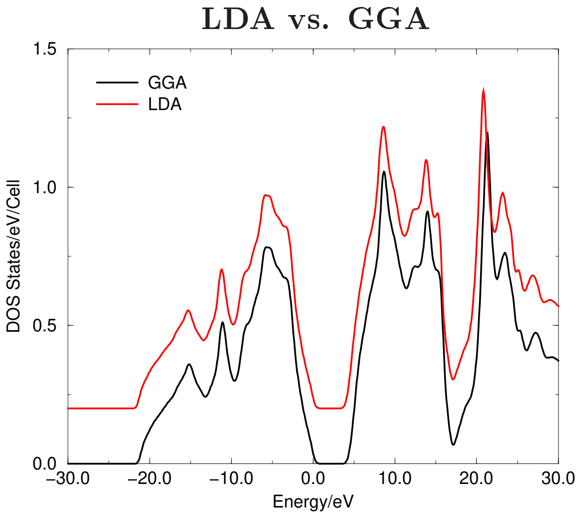


Figure A.2: Total density of states of Diamond, LDA vs. GGA

order 1%, but they can be significantly worse.

This presents a dilemma — which cell parameter should be used in the calculation of the spectral properties, experimental or calculated? To be totally consistent, it is clear that the relaxed value should be taken, but this work uses the experimental cell parameters. It is important to assess how much difference these two choices might make (clearly, if they made none it would indicate that the spectra are not sensitive to structural changes, and hence not useful for the investigation of microstructure!). In Figure A.1 the cell parameters are changed by $\pm 2\%$ from the experimental values. As expected, when the lattice is shrunk the valence bandwidth increases, by about 1eV, and correspondingly the bandwidth decreases as the lattice is expanded. More importantly, for the calculation of ELNES, features in the conduction band are also observed to move on the electron-volt scale.

A.4 LDA vs. GGA

The variation in the density of states with the level of approximation used for the exchange-correlation potential is shown in Figure A.2. Small differences in the conduction band can be seen, features shifting by the order of 1eV in energy.

A.5 Tolerances

CASTEP uses iterative minimisation to find the ground-state total energy and charge density, along with the individual wavefunctions and energies. An energy tolerance must be specified to halt the iterations. In Figure A.1 the effect of varying this tolerance is shown. It is clear that since the features are reproduced at a tolerance of 0.02eV it makes little sense to converge the eigenvalues to machine precision.

A.6 In summary

As a glance at the densities of states presented in this appendix reveals, the planewave approach is remarkably robust with respect to “sensible” changes in the parameter settings. Far larger changes are observed when the underlying physics is adjusted, for example when core hole effects are included as described in Chapter 5. This is a very important feature of the planewave approach. It makes it easier to determine whether the theory has to be improved, rather than simply changing the parameters of the calculation.

Appendix B

Implementation of the analytic quadratic integration method

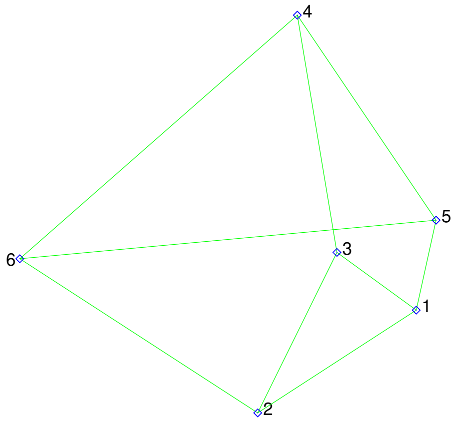
Although the idea behind the analytic quadratic DOS is simple enough, in analogy to the free electron DOS, the details of the mathematics behind the calculation are quite involved. However, the method may be implemented in a relatively straightforward fashion. What follows are the details of such an implementation. It is reproduced here mainly as an aid for future workers who may have use for the techniques but who do not wish to produce an implementation from scratch. There are also some novel developments such as the use of arbitrary polyhedra for the integration volume and correction of errors in the work of Boon *et al* [12].

B.1 Input

Within each sub-cell into which the Brillouin Zone is divided we have the dispersion relation,

$$E(\mathbf{q}) = E_o + \mathbf{G}^T\mathbf{q} + \mathbf{q}^T\mathbf{B}\mathbf{q}, \tag{B.1}$$

where $\mathbf{q}$ is the distance in reciprocal space from the k-point about which the expansion is performed. The gradient vector, $\mathbf{G}$, and the curvature tensor, $\mathbf{B}$ (which is symmetric), are obtained from the $\mathbf{k}.\mathbf{p}$ perturbation expansion described in Section 3.6, but may equally be obtained numerically by a more traditional interpolative scheme or numerical differentiation. Labels indicating the band and sub-cell number are suppressed for clarity. The details of the polyhedral sub-cell are required. The polyhedra are defined in terms of vertices, edges and faces. This represents a development from the original formulation in terms of tetrahedral sub-cells, in which the construction of faces and edges is unambiguous given the vertices. As an example, consider the way in which the following object, which is the irreducible wedge of the Diamond primitive cell, is defined:



Having generated a list of vertices, the nine edges are defined in terms of the joining of those vertices. The faces are described in a similar way. For example, the first edge is defined as the line joining vertex 1 to vertex 2. The first face is given by joining vertices 1,2, and 3. The entire construction is summarised:

Vertex #	k_x	k_y	k_z	Edge #	Edge Def.	Face #	Face Def.
1	1.15733	0.57866	0.00000	1	1→2	1	1→2→3
2	1.15733	0.00000	0.00000	2	2→3	2	4→5→1→3
3	1.15733	0.28933	0.28933	3	3→1	3	6→2→1→5
4	0.57866	0.57866	0.57866	4	4→5	4	6→5→4
5	0.86800	0.86800	0.00000	5	5→1	5	3→2→6→4
6	0.00000	0.00000	0.00000	6	3→4		
				7	6→2		
				8	5→6		
				9	4→6		

B.2 Preparation

The quadratic energy expansion is put in a form in which the origin coincides with the parabolic origin,

$$E(\mathbf{q}') = E_o' + \mathbf{q}'^T \mathbf{B}\mathbf{q}', \tag{B.2}$$

where $\mathbf{q}' = \mathbf{q} - \mathbf{k_o}$. The new parameters, E'_o and $\mathbf{k_o}$, are found by setting $\nabla_{\mathbf{q}} E(\mathbf{q}) = \mathbf{G} + 2\mathbf{B}\mathbf{q} = \mathbf{0}$:

$$\mathbf{k_o} = -\tfrac{1}{2}\mathbf{B}^{-1}\mathbf{G} \qquad E'_o = E_o - \tfrac{1}{4}\mathbf{G}\mathbf{B}^{-1}\mathbf{G} \tag{B.3}$$

The polyhedron's vertices are also shifted, although clearly the definitions of the face and edge are unaffected. In order to accelerate and simplify subsequent matrix manipulations, the problem (including the polyhedron) is rotated into the principal axis frame of $\mathbf{B}$. This will clearly not affect the results. E'_o is set to zero during the calculation of each cell's contribution to the DOS, to be replaced when the cell's contribution is added to the cumulative total DOS. At this point the following scalar product is introduced:

$$[\mathbf{a}, \mathbf{b}] = \mathbf{a}^T \mathbf{B} \mathbf{b} = \sum_i B_{ii} a_i b_i \tag{B.4}$$

It simplifies the evaluation of geometrical quantities, such as tangents, for the arbitrary metric defined by $\mathbf{B}$. When vectors are said to be orthogonal it is implied that they are orthogonal with respect to the above scalar product.

B.3 Sign and signature

The volume singularities which occur when $|\nabla_{\mathbf{k}} E(\mathbf{k}) = 0|$ i.e. at the origin in the new coordinates, can be classified according to how many negative elements $\mathbf{B}$ has. These classes were labelled M_0,M_1,M_2,M_3 by Van Hove [44]. Two of these classes may be obtained from the other two by inversion in energy. So if $det\mathbf{B}$ is negative we invert the energy grid and set $\mathbf{B} \rightarrow -\mathbf{B}$. This leaves two signatures, of type:

$$\underset{Spherical}{\begin{pmatrix} + & 0 & 0 \\ 0 & + & 0 \\ 0 & 0 & + \end{pmatrix}} \quad \underset{Hyperboloidal}{\begin{pmatrix} - & 0 & 0 \\ 0 & - & 0 \\ 0 & 0 & + \end{pmatrix}}$$

The two may be distinguished by comparing the quantities:

$$trace = B_{11} + B_{22} + B_{33} \tag{B.5}$$

$$cross = B_{11}B_{22} + B_{11}B_{33} + B_{22}B_{33} \tag{B.6}$$

If both *trace* and *cross* are positive then the signature is *spherical*, otherwise it is *hyperboloidal*.

B.4 Finding out about the singularities

The total number of possible singularities is:

$$NSTOT = 1 \times Volume + No.Vertices + No.Edges + No.Faces \tag{B.7}$$

The singularity energies and whether or not they contribute to the DOS have to be determined. The list of singularities is then sorted by energy.

B.5 Vertex singularities

All vertex singularities contribute, and the energies are given by:

$$E_\alpha = [\mathbf{a}, \mathbf{a}] \tag{B.8}$$

where $\mathbf{a}$ is the position of the vertex concerned. They occur when the CES passes through the vertices.

B.6 Edge singularities

The edge singularity for a given edge (defined by the vertices $\mathbf{a}$ and $\mathbf{b}$) occurs where the constant energy surface (CES) tangentially strikes the line defined by the edge concerned. This happens at a point on the line for which the vector joining the origin to it is orthogonal (with respect to the current metric) to the edge. This point can be written as:

$$\mathbf{w} = \mathbf{a} + \gamma\,(\mathbf{b} - \mathbf{a}) \tag{B.9}$$

But orthogonality requires $[\mathbf{w}, \mathbf{b} - \mathbf{a}] = 0$, and so we can find γ,

$$\gamma = \frac{-\,[\mathbf{a}, \mathbf{b} - \mathbf{a}]}{[\mathbf{b} - \mathbf{a}, \mathbf{b} - \mathbf{a}]} \tag{B.10}$$

The energy of the singularity is then given by $[\mathbf{w}, \mathbf{w}]$. If γ is positive and less than unity then the singularity contributes, i.e. $\mathbf{w}$ falls on the edge between $\mathbf{a}$ and $\mathbf{b}$. Each of the edges making up the polyhedron are examined in this way.

In the case of a hyperboloidal signature it is possible that $[\mathbf{b}-\mathbf{a}, \mathbf{b}-\mathbf{a}]$ is zero. This rarely occurs in real situations, but it can cause problems in highly symmetric test cases. The solution is to not include the singularity, and set $\mathbf{w} = \mathbf{a} - \mathbf{b}$.

B.7 Face singularities

A face singularity occurs where the CES tangentially strikes the plane of the face concerned. In close analogy with the case of the edge singularities, this happens at a point for which the vector joining it to the origin is orthogonal to the plane. The plane can be defined by any three points lying on it, so, since we are dealing with arbitrary polyhedra (not necessarily tetrahedra) we can pick any three vertices forming the plane (say $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$) to determine the point of singularity. Writing the singularity point as $\mathbf{z}$:

$$\mathbf{z} = \mathbf{a} + \lambda\,(\mathbf{b}-\mathbf{a}) + \mu\,(\mathbf{c}-\mathbf{a})\,, \tag{B.11}$$

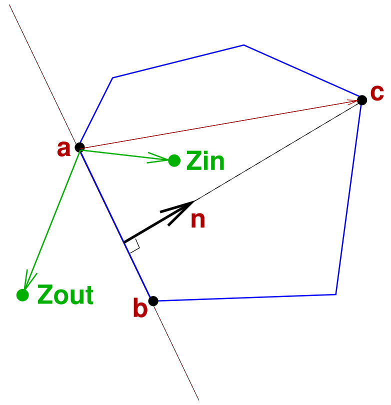
and requiring that $[\mathbf{z}, \mathbf{b}-\mathbf{a}]$ and $[\mathbf{z}, \mathbf{c}-\mathbf{a}]$ are both zero leads to two simultaneous equations for λ and μ. Using the following shorthand,

$$\begin{array}{lll} x = [\mathbf{b}-\mathbf{a}, \mathbf{b}-\mathbf{a}] & y = [\mathbf{c}-\mathbf{a}, \mathbf{b}-\mathbf{a}] & z = -\,[\mathbf{a}, \mathbf{b}-\mathbf{a}] \\ a = [\mathbf{b}-\mathbf{a}, \mathbf{c}-\mathbf{a}] & b = [\mathbf{c}-\mathbf{a}, \mathbf{c}-\mathbf{a}] & c = -\,[\mathbf{a}, \mathbf{c}-\mathbf{a}]\,, \end{array} \tag{B.12}$$

then,

$$\begin{array}{rcl} \lambda & = & (zb-cy)/(xb-ay) \\ \mu & = & (za-cx)/(ya-bx). \end{array} \tag{B.13}$$

The energy of the singularity is then $[\mathbf{z}, \mathbf{z}]$. Now, in the original formulation it was straightforward to determine whether the singularity was to be included. Each face was a triangle, and hence if $(1-\lambda-\mu), \lambda$, and μ were all positive $\mathbf{z}$ would fall within the face. Clearly in the current situation the same condition would ensure that $\mathbf{z}$ would fall within the triangle defined by the three points chosen on the face, which would exclude the singularity in many situations when it should be included. Thus a new scheme has been developed to solve this problem.



This proceeds as follows. Choose an edge, and label the two vertices $\mathbf{a}$ and $\mathbf{b}$. Pick a third distinct vertex on the face, $\mathbf{c}$. Then form a normal, $\mathbf{n}$, facing into the centre of the face:

$$\mathbf{n}' = (\mathbf{c} - \mathbf{a}) - \frac{(\mathbf{b}-\mathbf{a}).(\mathbf{c}-\mathbf{a})}{|\mathbf{b}-\mathbf{a}|^2}(\mathbf{b} - \mathbf{a}) \qquad \mathbf{n} = \frac{\mathbf{n}'}{|\mathbf{n}'|} \tag{B.14}$$

If $\mathbf{n}.(\mathbf{z} - \mathbf{a})$ is positive for all edges of the face then the singularity is to be included. Otherwise $\mathbf{z}$ falls outside the face, and does not contribute.

B.8 How to count the face singularities

In order to decide how to count the face singularities (i.e. the sign attributed to the discontinuity) we must decide whether the origin is on the "inside" or "outside" side of the plane defined by the face (inside or outside of the polyhedron). For the case of an arbitrary polyhedron we must choose three vertices of the face concerned, $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$. A normal to the face can be constructed,

$$\mathbf{n}' = (\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a}) \qquad \mathbf{n} = \frac{\mathbf{n}'}{|\mathbf{n}'|}. \tag{B.15}$$

A fourth vertex, $\mathbf{d}$, not in the current face, is chosen. If $\mathbf{n}.(\mathbf{d} - \mathbf{a})$ and $\mathbf{n}.(\mathbf{0} - \mathbf{a})$ are the same sign, then the origin is on the side of the "inside" of the polyhedron. We define a quantity, σ, for each face, and it is $+1$ for the origin on the "inside", and -1 for the origin on the "outside".

B.9 The volume singularity

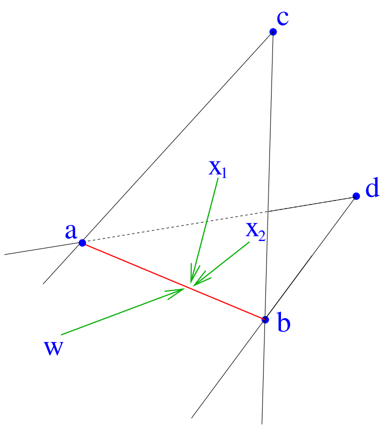
Clearly, if the origin is on the “inside” side of all the faces then the volume singularity is contained within the polyhedron, and it makes a contribution. By construction, the energy of the volume singularity is zero, since E'_o was subtracted earlier.

B.10 Evaluation of the tangents

In order to calculate the contribution to the DOS of the edge and vertex singularities some geometrical quantities must be evaluated — namely the “tangents” between the faces forming each edge (see Boon *et al* [12] for more detail). For each edge a vector, $\mathbf{w}$, has been found, joining the origin to the edge singularity. It is also clear that an edge is formed by the joining of two planes, or faces. Construct two vectors, $\mathbf{x_1}$ and $\mathbf{x_2}$, which lie in each of the two faces respectively and are orthogonal (in the sense defined above) to the edge. Write:

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{c}-\mathbf{a}+\lambda(\mathbf{b}-\mathbf{a}) \\ \mathbf{x}_2 &= \mathbf{d}-\mathbf{a}+\mu(\mathbf{b}-\mathbf{a}) \end{aligned} \tag{B.16}$$

and set $[\mathbf{x}_{1,2},\mathbf{b}-\mathbf{a}]=0$, solving for λ and μ.



Calculate the (hyperbolic) tangents of the angles between $\mathbf{w}$ and $\mathbf{x}_i$ for each of the two adjacent faces:

$$T_i = \pm\frac{\left\{\left|[\mathbf{x}_i,\mathbf{x}_i][\mathbf{w},\mathbf{w}]-[\mathbf{x}_i,\mathbf{w}]^2\right|\right\}^{\frac{1}{2}}}{[\mathbf{x}_i,\mathbf{w}]} \qquad i=1,2 \tag{B.17}$$

The sign is positive if $[\mathbf{b}-\mathbf{a},\mathbf{b}-\mathbf{a}]$ and σ (defined in Section B.8) for the face i are of the same sign, and negative otherwise.

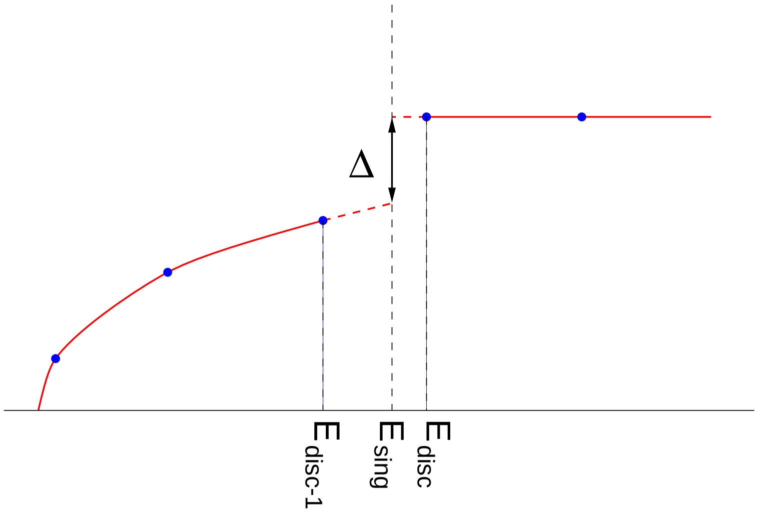
B.11 Contribution from face and volume singularities

In the preceding sections is was shown how to find the parameters which are necessary for the construction of the contribution from the current sub-cell to the DOS. This is now built up, starting with the contributions from the face and volume singularities.

The working array $D(E_i)$ is initially set to zero, the set of energies, E_i, are the set of discrete energies at which the DOS is required. The list of singularity energies should be sorted in increasing energy, noting which ones are to be included (as determined above). The edge and vertex singularities are put to one side for the moment. Working up through energy we introduce steps in $D(E_i)$ at the E_i immediately above each included singularity (labelled E_{disc}). If it is a volume singularity the step is of size $\pm 4\pi$, positive for a spherical dispersion, and negative for the hyperboloidal case. If it is an included face singularity then the step is $\pm 2\pi$. The sign for the spherical dispersion is positive if the σ (defined in Section B.8) for the face is negative, and negative otherwise. The opposite is true if it is a hyperboloidal dispersion. See the black curves in Figures B.1 and B.2. The discontinuous contribution to the DOS is formed by setting:

$$D(E_i) \to 2\sqrt{E_i} D(E_i) \tag{B.18}$$

The resulting discontinuous curves are show in red in Figures B.1 and B.2. It now remains to make the contribution to the DOS continuous. A naive approach is to simply work up through $D(E_i)$ setting $D(E_{disc})$ and above to be $D(E_i)+(D(E_{disc-1})-D(E_{disc}))$. Although this becomes correct for high grid densities, it introduces serious inaccuracies at low ones and so is not recommended. The correct approach is to calculate Δ, as indicated below.



$$\Delta = -2\sqrt{E_{sing}} \times Step, \tag{B.19}$$

where $Step$ is the step found above (i.e. $\pm 4\pi$ or $\pm 2\pi$). Then set $D(E_i)$, for $E_i = E_{disc}$ and above, to $D(E_i) + \Delta$.

B.12 Contribution from edge and vertex singularities

This is the most involved part of the implementation, requiring first the evaluation of the so called "primitives" — $f(\varepsilon)$, and $g(\varepsilon)$ — and then constructing the edge contributions using these primitives.

B.13 Evaluation of the primitives

For each edge we must determine two functions of energy, $f(\varepsilon)$, and $g(\varepsilon)$. As an input we have the following information:

1. ε, The energy at which to evaluate $f(\varepsilon)$, and $g(\varepsilon)$
2. E_{sing}, The energy of the edge singularity
3. $T_{1,2}$, The (hyperbolic) tangents between $\mathbf{w}$ and $\mathbf{x}_{1,2}$

Then apply the following:

1. If $T_i \neq 0$ then $s_i \to sgn(T_i)$ else $s_i = 0$
2. $y_i \to T_i\sqrt{\left|1 - \frac{E_{sing}}{\varepsilon}\right|}$
3. If $\varepsilon > 0$ then $y_i \to \arctan y_i - \frac{\pi}{2}s_i$ else $y_i \to \frac{1}{2}\ln\left|\frac{1+y_i}{1-y_i}\right|$
4. If $\varepsilon \neq E_{sing}$ then $x_i \to T_i^2 + sgn(\varepsilon - E_{sing})$ else $x_i \to T_i^2 + 1$
5. $x_i \to 2T_i\sqrt{\left|\frac{E_{sing}}{x_i}\right|}$
6. $xx_i \to \sqrt{\left|x_i\left(\frac{\varepsilon}{E_{sing}} - 1\right)\right|}$
7. If $sgn(E_{sing}x_i) > 0$ then $xx_i \to \arctan xx_i + \frac{\pi}{2}$ else $xx_i \to \frac{1}{2}\ln\left|\frac{1+xx_i}{1-xx_i}\right|$

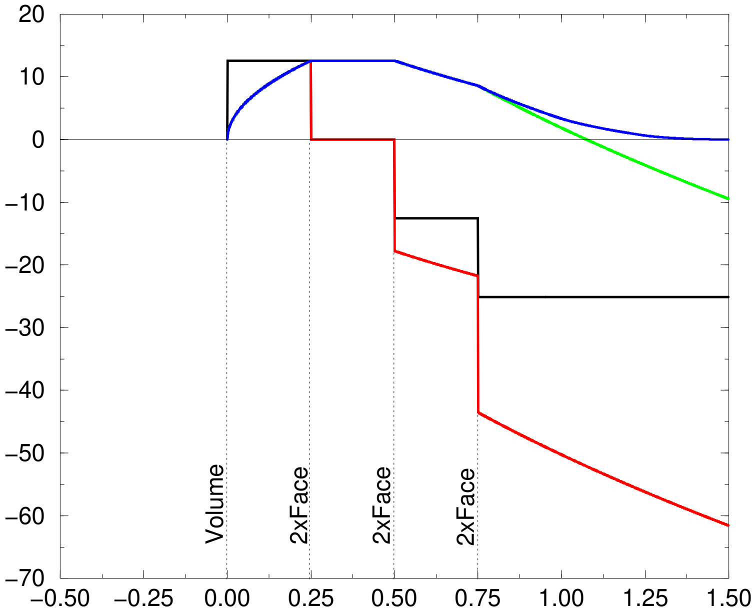


Figure B.1: The various contributions to the DOS for a spherical dispersion relation ($B_{11} = 1, B_{22} = 2, B_{33} = 3$ centred at (0,0,0) so is actually ellipsoidal), within a cube defined by planes at $\pm\frac{1}{2}$. The black curve indicates the step discontinuities due to volume and face singularities, the red on multiplication by $2\sqrt{E_i}$, the green on removing the discontinuities, and finally the blue on adding in the contribution due to edge singularities.

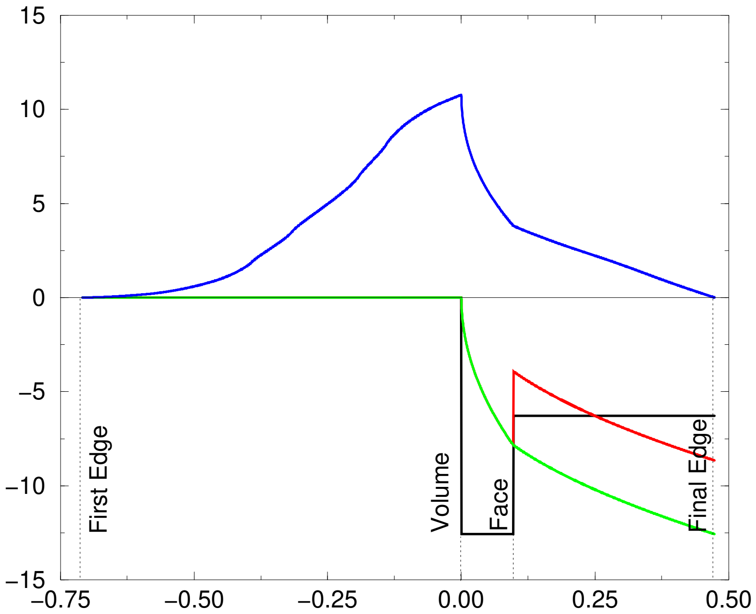


Figure B.2: As for the previous figure, but this time a hyperboloidal dispersion ($B_{11} = -1, B_{22} = -1, B_{33} = 1$ centred at (0.01,0.02,0.03))

Having found the quantities x_i, xx_i, and y_i for $i = 1, 2$ we can form the two primitives $f(\varepsilon)$ and $g(\varepsilon)$:

$$\begin{array}{lcc} f(\varepsilon) & \to & sgn(\varepsilon)\sum_{i=1,2} y_i \\ g(\varepsilon) & \to & 2\sqrt{|\varepsilon|}\sum_{i=1,2} y_i - \sum_{i=1,2} x_i xx_i \end{array} \tag{B.20}$$

It should be pointed out that the paper of Boon *et al* [12] contained an error. Using their definitions, the primitive $g(\varepsilon)$ should be written as,

$$g(\varepsilon) = 2\,|\varepsilon|\, f(\varepsilon) - 2\left|\frac{\varepsilon_o}{X}\right|^{\frac{1}{2}} TF\left[\left|X\left(\frac{\varepsilon}{\varepsilon_o} - 1\right)\right|\right]^{\frac{1}{2}}, \tag{B.21}$$

replacing,

$$g(\varepsilon) = 2\,|\varepsilon|^{\frac{1}{2}}\, f(\varepsilon) - 2\left|\frac{\varepsilon_o}{X}\right| TF\left[\left|X\left(\frac{\varepsilon}{\varepsilon_o} - 1\right)\right|\right]^{\frac{1}{2}}, \tag{B.22}$$

which appeared in the original work.

Having calculated the primitives, the edge contribution is constructed. The following is a recipe for that construction. There are two possible routes, depending on whether the edge singularity is included.

B.14 Edge singularity not included

Labelling the lower and higher vertex singularity energies as ε_1 and ε_2 respectively (with index labels n_1 and n_2 for the array D_{edge}), the following quantities are formed:

$$\begin{array}{ll} \alpha_1 = sgn(\varepsilon_1)2f(\varepsilon_1)\sqrt{|\varepsilon_1|} - g(\varepsilon_1) & \beta_1 = -2f(\varepsilon_1) \\ \alpha_2 = \alpha_1 + g(\varepsilon_2) - sgn(\varepsilon_2)2f(\varepsilon_2)\sqrt{|\varepsilon_2|} & \beta_2 = \beta_1 + 2f(\varepsilon_2) \end{array} \tag{B.23}$$

The array $D_{edge}(i)$ is set to zero for $i = 1, \ldots, n_1$. For $i = n_1 + 1, \ldots, n_2$:

$$\begin{array}{rl} \text{If } E_i \neq 0 \text{ then} & D_{edge}(i) = sgn(E_i)\beta_1\sqrt{|E_i|} + \alpha_1 + g(E_i) \\ \text{else} & D_{edge}(i) = \alpha_1 + g(E_i) \end{array}$$

For $i > n_2$:

$$\begin{array}{rl} \text{If } E_i \neq 0 \text{ then} & D_{edge}(i) = sgn(E_i)\beta_2\sqrt{|E_i|} + \alpha_2 \\ \text{else} & D_{edge}(i) = \alpha_2 \end{array}$$

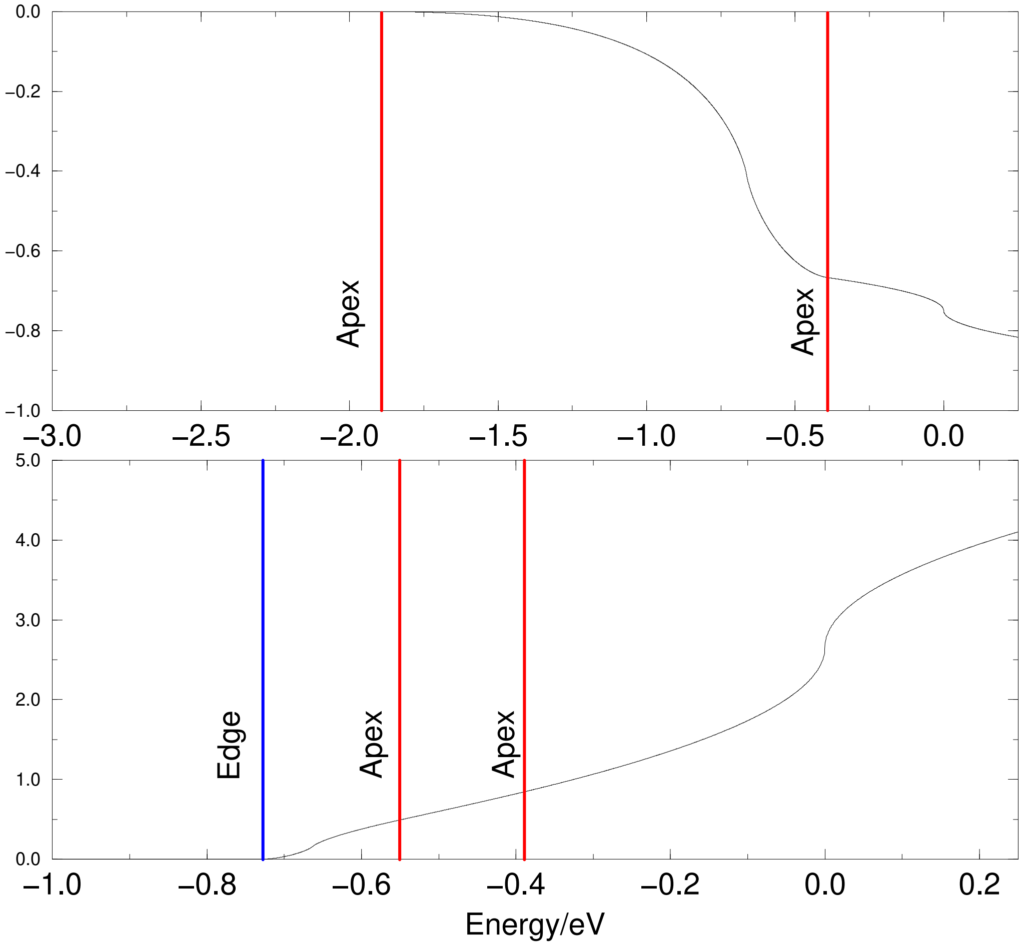


Figure B.3: Examples of the form of the edge contribution. The top curve is for a case when the edge singularity is not included, and the bottom included

B.15 Edge singularity included

The edge and two vertex singularities are ordered in increasing energy. The lower, middle, and higher singularity energies are labelled as ε_1, ε_2, and ε_3 respectively (with index labels n_1, n_2, and n_3 for the array D_{edge}). There are then three distinct situations:

1. The edge singularity is at ε_1, and so set $x = 2, y = 1$

2. The edge singularity is at ε_2, and if $|\varepsilon_2 - \varepsilon_1| < |\varepsilon_3 - \varepsilon_2|$ then set $x = 2, y = 1$ else $x = 1, y = 2$

3. The edge singularity is at ε_3, and so set $x = 1, y = 2$

The following quantities are constructed:

$$\begin{aligned}
\alpha_1 &= x\left(sgn(\varepsilon_1)2f(\varepsilon_1)\sqrt{|\varepsilon_1|} - g(\varepsilon_1)\right) & \beta_1 &= -2xf(\varepsilon_1)\\
\alpha_2 &= \alpha_1 - (x-y)\left(sgn(\varepsilon_2)2f(\varepsilon_2)\sqrt{|\varepsilon_2|} - g(\varepsilon_2)\right) & \beta_2 &= \beta_1 + 2(x-y)f(\varepsilon_2)\\
\alpha_3 &= \alpha_2 - y\left(sgn(\varepsilon_3)2f(\varepsilon_3)\sqrt{|\varepsilon_3|} - g(\varepsilon_3)\right) & \beta_3 &= \beta_2 + 2yf(\varepsilon_3)
\end{aligned} \tag{B.24}$$

The array $D_{edge}(i)$ is set to zero for $i = 1, \ldots, n_1$. For $i = n_1 + 1, \ldots, n_2$:

$$\begin{aligned}
\text{If } E_i \neq 0 \text{ then} \quad & D_{edge}(i) = sgn(E_i)\beta_1\sqrt{|E_i|} + \alpha_1 + xg(E_i)\\
\text{else} \quad & D_{edge}(i) = \alpha_1 + xg(E_i)
\end{aligned}$$

For $i = n_2, \ldots, n_3$:

$$\begin{aligned}
\text{If } E_i \neq 0 \text{ then} \quad & D_{edge}(i) = sgn(E_i)\beta_2\sqrt{|E_i|} + \alpha_2 + yg(E_i)\\
\text{else} \quad & D_{edge}(i) = \alpha_2 + yg(E_i)
\end{aligned}$$

For $i > n_3$:

$$\begin{aligned}
\text{If } E_i \neq 0 \text{ then} \quad & D_{edge}(i) = sgn(E_i)\beta_3\sqrt{|E_i|} + \alpha_3\\
\text{else} \quad & D_{edge}(i) = \alpha_3
\end{aligned}$$

B.16 And finally...

Having found all the contributions to the DOS, the edge, vertex, face and volume contributions are summed:

$$D(E_i) \rightarrow D(E_i) - \sum_{edge} D_{edge}(E_i) \tag{B.25}$$

If the original curvature tensor $\mathbf{B}$ had a negative determinant then the energy scale must be flipped back ($E_i \rightarrow -E_i$). Then the contribution is shifted by the original energy of the origin, E'_o. Finally the contribution is scaled by the factor $\frac{1}{4}(|det\mathbf{B}|)^{-\frac{1}{2}}$, and also by the appropriate matrix element if required.

Appendix C

Optical spectra

C.1 Introduction

The study of the interaction of light or fast electrons with solids give access to a rich source of information about the system. The interest might be in the optical properties themselves (for example, the colour or transmission properties of a material), some aspect of the electronic properties (for example, defect states), or as a probe of the structure itself.

The CASTEP total energy code can be used to study these properties. In order to calculate the total energy of a given system, Density Functional Theory requires the ground-state charge density to be evaluated self consistently. Having obtained this density, the Kohn-Sham equations can be solved for a fixed Hamiltonian,

$$[-\frac{\hbar^2}{2m}\nabla^2 + \hat{V}_{ion} + \hat{V}_H + \hat{V}_{XC}]\Psi_i = \varepsilon_i \Psi_i. \tag{C.1}$$

$\hat{V}_H$ and $\hat{V}_{XC}$ are the Hartree and exchange-correlation potentials (within LDA or GGA) respectively, and are fixed by the ground-state charge density [73]. Although the Kohn-Sham states and corresponding energies have no formal meaning, the equations are similar to those for the quasiparticles that actually take part in the electronic transitions. The approximation is made that the matrix elements between states and energy differences are close to the actual ones (see the discussion of the DFT band-gap error below in Section C.2.6). The response of the systems studied to a perturbation induced by a passing photon or electron is described in terms of these states.

In Section C.2 the theoretical background to the calculation of optical properties, and the limitations of the method are discussed. Section C.3 describes the procedure for producing the spectra. Validation tests are presented in Section C.4 for some real systems, illustrating some of the shortcomings of the present implementation of a DFT approach.

C.2 Theory

C.2.1 Optical properties

In general, the difference in the propagation of an electromagnetic wave through vacuum and some other material can be described by a *complex refractive index*.

$$N = n + ik \tag{C.2}$$

In vacuum N is real and equal to unity. For transparent materials it is purely real, the imaginary part being related to the *absorption coefficient* by,

$$\eta = \frac{2k\omega}{c} \tag{C.3}$$

η characterises the fraction of energy lost by the wave on passing through a unit thickness of the material concerned and can be derived by considering the rate of production of Joule heat in the sample.

The *reflection coefficient* can be obtained for the simple case of normal incidence onto a plane surface, by matching both the electric and magnetic fields at the surface,

$$R = \left|\frac{1-N}{1+N}\right|^2 = \frac{(n-1)^2 + k^2}{(n+1)^2 + k^2}. \tag{C.4}$$

However, when performing calculations of optical properties it is common to evaluate the *complex dielectric function*, and then express other properties in terms of it. The complex dielectric function $\varepsilon(\omega)$ is given by,

$$\varepsilon = \varepsilon_1 + i\varepsilon_2 = N^2, \tag{C.5}$$

and hence the relation between the real and imaginary parts of the refractive index and dielectric function is,

$$\varepsilon_1 = n^2 - k^2 \; \varepsilon_2 = 2nk. \tag{C.6}$$

A further frequent form for the expression of optical properties is the *optical conduc-*

tivity,

$$\sigma = \sigma_1 + i\sigma_2 = -i\frac{\omega}{4\pi}(\varepsilon - 1). \tag{C.7}$$

This is most useful for the discussion of the electronic properties of metals.

A further property we may calculate from the complex dielectric function it the *loss function* — see Chapter 2. It describes the energy lost by a point electron passing through a homogeneous dielectric material, and is given by,

$$Im\left\{\frac{-1}{\varepsilon(\omega)}\right\}. \tag{C.8}$$

In the low loss region of the electron energy loss spectrum the loss function is dominated by *plasmons* — collective excitations of the electronic system against the ionic background. These occur when the electronic system becomes unstable against small external perturbations. The real part ε_1 describes the screening of an external field within the material ($E_{int} = E_{ext}/\varepsilon_1$), and hence instabilities arise when $\varepsilon_1 = 0$. At high energies the loss function reduces to ε_2 and is dominated by real electronic transitions between core states and the unoccupied conduction band.

C.2.2 Connection to experiment

Experimentally, the most accessible optical parameters are the absorption $\eta(\omega)$, and the reflection $R(\omega)$ coefficients. In principle, given the knowledge of both these, the real and imaginary parts of N can be determined, through Equations C.3 and C.4. Equation C.5 allows expression in terms of the complex dielectric function. However, in practice, the experiments are more complicated than the case of normal incidence considered above. Polarisation effects must be accounted for, and the geometry can become quite involved (for example, transmission through multilayered films or incidence at a general angle). However, this is a problem of optics, and would needlessly complicate the microscopic origin of the optical properties. We consider only normal incidence, but polarisation effects are taken into account. For a more general discussion of the analysis of optical data see Ref [71]. Measurement of angularly resolved Low Loss EELS spectra allows the extraction of the loss function, and hence the real and imaginary parts of the dielectric function (using a Kramers-Kronig relation) [31, 80].

C.2.3 Connection to electronic structure

As mentioned in the introduction, the interaction of a photon with the electrons in the system under study is described in terms of time dependent perturbations of the ground-state electronic states. Transitions are caused between occupied and unoccupied states by the electric field of the photon (the magnetic field effect is weaker by a factor of $\frac{v}{c}$ where v is the electronic velocity) or the time varying Coulomb potential of the swift electron in EELS. When these excitations are collective they are known as plasmons which are most easily observed by passing a fast electron through the system rather than a photon, since transverse photons do not excite longitudinal plasmons. When the transitions are independent they are as single particle excitations. The spectra resulting from these excitations can be thought of as a joint density of states between the valence and conduction bands, weighted by the appropriate matrix elements which introduce the selection rules.

C.2.4 Evaluation of the dielectric function

We calculate the imaginary part of the dielectric function, which is given by,

$$\varepsilon_2(\mathbf{q} \to \mathbf{0}_{\hat{\mathbf{u}}}, \hbar\omega) = \frac{2e^2\pi}{\Omega\varepsilon_0} \sum_{\mathbf{k},v,c} |\langle\psi_{\mathbf{k}}^c|\hat{\mathbf{u}}.\mathbf{r}|\psi_{\mathbf{k}}^v\rangle|^2 \delta(E_{\mathbf{k}}^c - E_{\mathbf{k}}^v - E), \tag{C.9}$$

in the long wavelength limit so that the dipole appoximation can be applied for both optical and electron energy loss spectra where $\mathbf{u}$ is the vector defining the polarisation of the incident electric field or the momentum transfer of the swift electron in EELS.

This expression is similar to Fermi's Golden rule for time dependent perturbations, and $\varepsilon_2(\omega)$ can be thought of as detailing the real transitions between occupied and unoccupied electronic states. Since the dielectric function describes a causal response, the real and imaginary parts are linked by a Kramers-Kronig transform. This is used to obtain the real part, $\varepsilon_1(\omega)$. $\varepsilon_1(\omega) - 1$ is a measure of the polarizability of the electronic system.

C.2.5 Details of the computation

Brillouin Zone integrations

The current approach to integration over the Brillouin Zone involves taking a symmetrised Monkhurst-Pack grid, and smearing each energy level with a Gaussian spread function. As mentioned in Chapter 3, the piecewise quadratic approach to the Brillouin zone integrations could be applied here, but corrections due to the dispersion of the optical matrix elements would need to be implemented first.

Evaluation of Matrix Elements

The matrix elements that are required to describe the electronic transitions in Equation C.9 are $\langle\psi^c_{\mathbf{k}}|\mathbf{r}|\psi^v_{\mathbf{k}}\rangle$ [91], which are normally written as $\frac{1}{i\omega m}\langle\psi^c_{\mathbf{k}}|\mathbf{p}|\psi^v_{\mathbf{k}}\rangle$, allowing straightforward calculation in reciprocal space. However, this replacement depends on the use of local potentials, which are not common in modern planewave codes. The corrected form of the matrix elements for non-local pseudopotentials are,

$$\langle\psi^c_{\mathbf{k}}|\mathbf{r}|\psi^v_{\mathbf{k}}\rangle = \frac{1}{i\omega m}\langle\psi^c_{\mathbf{k}}|\mathbf{p}|\psi^v_{\mathbf{k}}\rangle + \frac{1}{\hbar\omega}\langle\psi^c_{\mathbf{k}}|[\hat{V}_{nl},\mathbf{r}]|\psi^v_{\mathbf{k}}\rangle, \tag{C.10}$$

where $\hat{V}_{nl}$ is the non-local pseudopotential. It should be noted that this is the same correction as the first order correction to the **k.p** matrix elements described in Chapter 3. Read and Needs [79] have emphasised its importance and noticed that many workers neglect it in the calculation of optical properties.

Polarisation

For materials that do not display full cubic symmetry, the optical properties will display some anisotropy. This can be included in the calculations by taking the polarisation of the electromagnetic field into account or the direction of the momentum transfer of the incident electron. When evaluating the dielectric function there are three options.

- *Polarised*

  This requires a vector to define the direction of the electric field vector for the light at normal incidence to the crystal.

- *Unpolarised*

  This takes the vector supplied as the direction of propagation of incident light at normal incidence to the crystal. The electric field vector is taken as an average over the plane perpendicular to this direction.

- *Polycrystal*

  No direction need be specified, the electric field vectors are taken as a fully isotropic average.

Scissor operator

As discussed in more detail below, the relative position of the conduction to valence bands are found to be in error when the Kohn-Sham eigenvalues are used. In an attempt to "fix" this a rigid shift of the conduction band is imposed. This is given the formal name of a *scissor operator*. Note, however, that the scissor operator is not included in the energy denominator when evaluating the matrix element using Equation C.10 [45]. The application of the scissor operator is required to produce a rigid upwards shift of the bandstructure and hence the correct onset for interband transitions. Through **k.p** perturbation theory the optical matrix elements are seen to be related to the band shape. If the scissor operator was included in the energy denominator the optical matrix elements would clearly change and hence so would the band shape which is not consistent with the simple rigid shift.

C.2.6 Limitations of the method

Local field effects

The level of approximation used here does not take any local field effects into account. These arise from the fact that the electric field experienced at a particular site in the system is screened by the polarisability of the system itself. So, the local field is different from the applied external field (i.e. the photon electric field). This can have a significant effect on the calculated spectra (see the example of bulk Silicon below). While the theoretical treatment of this effect is well known — it involves the evaluation of the full dielectric matrix [31] and it is prohibitively expensive to calculate for general systems at present. Hence, local field effects will be neglected.

Quasiparticles and the DFT bandgap

In order to calculate any spectral properties an identity between the Kohn-Sham eigenvalues and the quasiparticle energies has to be made. Although there is no formal connection between the two, the similarities between the Schrödinger-like equation for the quasiparticles and the Kohn-Sham equations provides the basis of this approach. For semiconductors, it has been shown computationally (comparing quasiparticle and LDA bandstructures) that most of the difference between Kohn-Sham eigenvalues and the true excitation energies can be accounted for by a rigid shift of the conduction band upwards with respect to the valence band [36]. This is attributed to a discontinuity in the exchange-correlation potential as the system goes from (N)-electron to (N+1)-electron during the excitation process [35]

Silicon	All electron	Pseudopotential
3s→3p	0.7034	0.7043
3p→3d	0.6764	0.6702

Table C.1: Optical matrix elements for an isolated atom, comparing all electron and pseudopotential calculations.

which is not included in the LDA within DFT since it is a ground-state theory. There can, in some systems, be considerable dispersion of this shift across the Brillouin Zone, in which case the scissor operator used here will be insufficient.

Excitonic effects

In connection with the absence of local field effects, excitonic effects are not treated in this formalism. This will be of particular importance for ionic crystals where such effects are large.

Others

Phonons and their optical effects in the infrared region have been neglected. There is also an intrinsic error in the matrix elements for optical transition due to the fact that pseudowavefunctions have been used (i.e. they deviate from the true wavefunctions in the core region), leading to small numerical errors. However, the selection rules will not be changed. See Table C.1 for a numerical test on an isolated atom (for which a straightforward comparison to all electron calculations may be made).

C.3 Practical details of the calculation

C.3.1 Producing optical spectra

The calculation of optical properties involves the following steps:

1. Perform a total energy calculation to evaluate the self-consistent charge density. This allows the construction of the fixed Hamiltonian since V_H and V_{xc} depend on the charge density.

2. For the fixed Hamiltonian, find the Kohn-Sham eigenvalues for many more **k**-points (for the Brillouin Zone integration) and many more unoccupied states than for the selfconsistent calculation. The matrix elements for transitions between the occupied and unoccupied states are evaluated at this stage.

3. From Equation C.9, $\varepsilon(\omega)$ is evaluated, for a given smearing to assist the Brillouin zone integration, given scissor operator to correct the bandgap error and polarisation.

4. Using the above dielectric function the optical properties required can be evaluated, following the methods given in Section C.2.1.

C.3.2 Effects of varying parameters on optical spectra

Following the discussion in Appendix A, a study of the convergence of the optical properties with various parameters is presented. Tests have been performed for a primitive cell of Silicon and the results are presented in Figure C.1. A few words about each of the tests follows.

Number of bands included

It is clear that $\varepsilon_2(\omega)$ converges rapidly with the number of conduction bands included. The contributions from successive bands can vary discontinuously due to selection rules.

Energy cutoff

While it is important to converge with planewave cutoff to obtain the correct energies for the spectral features, the qualitative form of those features converge rapidly, as discussed in Appendix A.

Number of k-points for SCF run

The optical properties converge rapidly with the number of **k**-points used in the SCF run, which implies a degree of insensitivity to the calculated charge density.

Number of k-points for Brillouin Zone integration

The importance of converging the Brillouin zone integration for the optical properties with **k**-point density is clear. Both energies and spectral features are strongly affected by the accuracy of this integration. The scheme described in Chapter 3 could be of assistance in the future.

C.3.3 The question of accuracy and precision

The effect of the various parameters on the degree of convergence achieved in the calculation of the optical properties has been described. It must, however, be kept in mind that full convergence (which will frequently lead to a highly strenuous calculation) will not always be appropriate. Experimental data is always only accurate to a given resolution, and the theory itself (see Section C.2.6) is limited in accuracy.

C.4 Some calculated spectra compared to experiment

C.4.1 Silicon

Figure C.2 shows the dielectric properties of bulk Silicon. Note that while there is a good general agreement with experiment [71], there is significant deviation at the threshold which can be attributed to the failure to include local field effects.

C.4.2 Diamond

Figure C.3 show the dielectric properties of Diamond. The agreement with experiment [71] is very good for this wide bandgap material.

C.4.3 Titanium Disulphide

Figure C.4 shows the results of calculations for the lamellar material TiS_2. While TiS_2 has anisotropic optical properties, angle resolved data is not available [29], so a polycrystalline calculation is shown. There is some controversy as to whether TiS_2 is a semiconductor or semimetal, but this DFT LDA calculation shows it to be a semimetal. The possibility of a bandgap error could be the cause for some deviation from experiment, as could the complications introduced by anisotropy. However, again there is good general agreement with experiment.

C.5 Comparison of all electron and pseudopotential calculations

The approach to the calculation of optical properties described in this appendix was initially intended to be used exclusively with pseudopotentials. However, as described in Chapter 4, if a sufficiently high planewave cutoff is used, in combination with the efficient Brillouin zone integration of Chapter 3, then all electron calculations are feasible for small systems using the CASTEP total energy code. Having performed such an all electron bandstructure calculation there is clearly no obstruction to evaluating the optical spectra. But since the all electron calculations include the core states, the resulting spectra will, at high energy, include the absorption thresholds due to transitions from the core states into the empty conduction bands. Thus, the X-ray (or EELS in the dipole approximation) absorption spectra is calculated. This offers an opportunity to test the approach given earlier in this thesis for the calculation of ELNES within the dipole approximation.

Figure C.5 shows such a calculation for a primitive cell of Diamond. The planewave cutoff for the all electron calculation was 9000eV, and 16 **k**-points were used. Note the K-edge threshold does not occur at the experimental value. This is most likely due to the incomplete convergence with planewave cutoff. The pseudopotential calculation — performed at a 600eV cutoff — is constructed from two parts. At low energy is a standard optical calculation — and shows good agreement with the all electron calculation. The K-edge at higher energies is calculated using the scheme described in this thesis. Since the energy of the core state is a unknown in the pseudopotential approximation, the threshold is aligned with that of the all electron calculation, but no other scaling is applied. Thus, it is again emphasised that the direct evaluation of the ELNES matrix elements within the pseudopotential approximation allows the absolute evaluation of the imaginary part of the dielectric function. The comparison serves to increase the confidence in the two different approaches to the calculation of spectral properties, since the same results are obtained via two different routes.

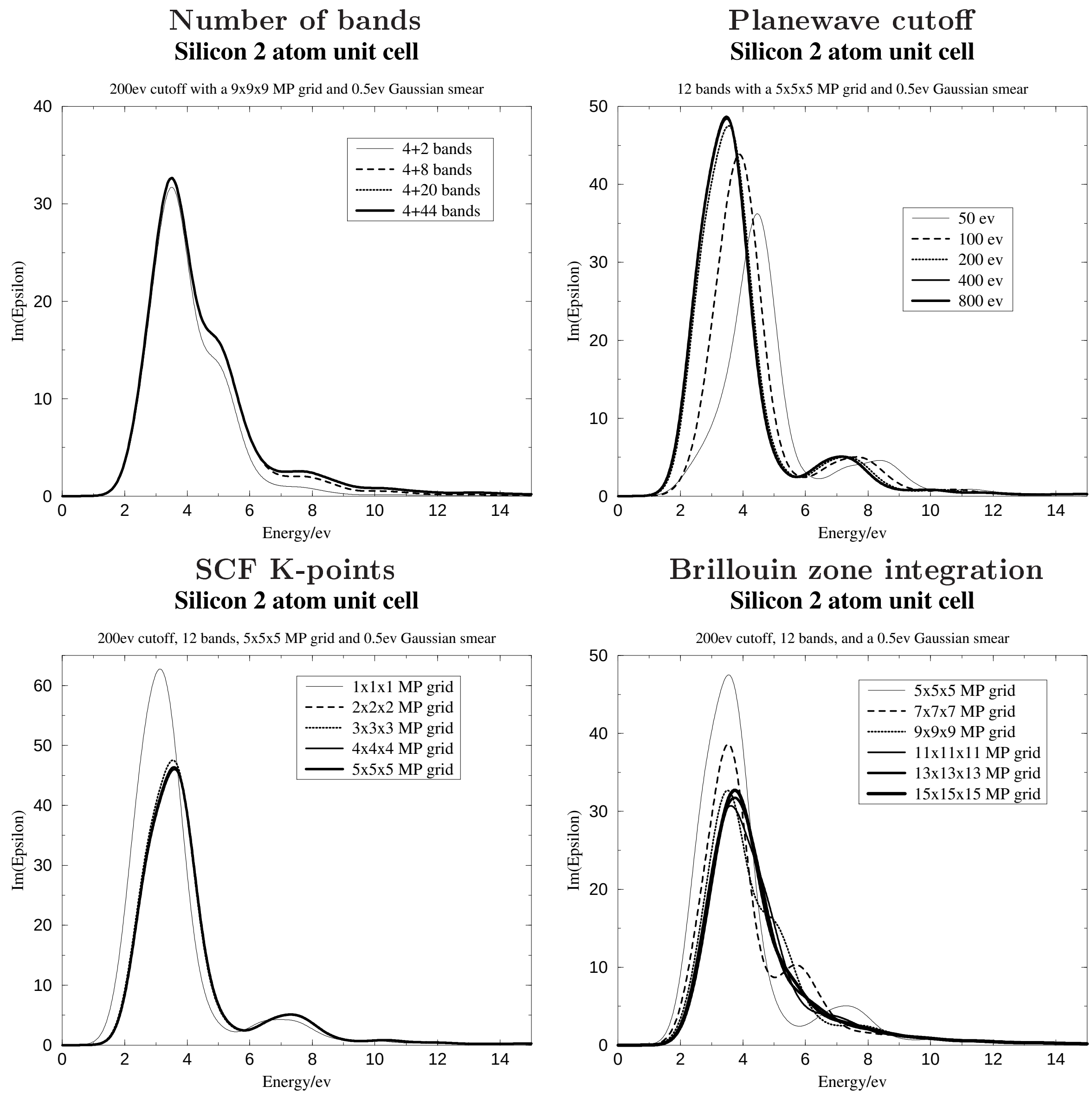


Figure C.1: A series of convergence tests - $\mathrm{Im}(\varepsilon(\omega))$.

Silicon

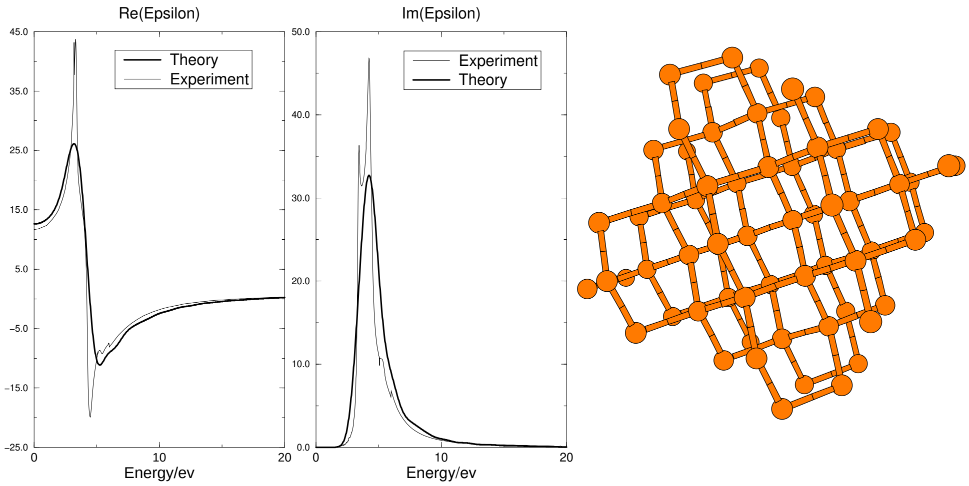


Figure C.2: The complex dielectric function was calculated with a 200ev cutoff, 24 bands, a 15x15x15 Monkhurst-Pack grid, a 0.5ev Gaussian smearing and a 0.5ev Scissor operator.

Diamond

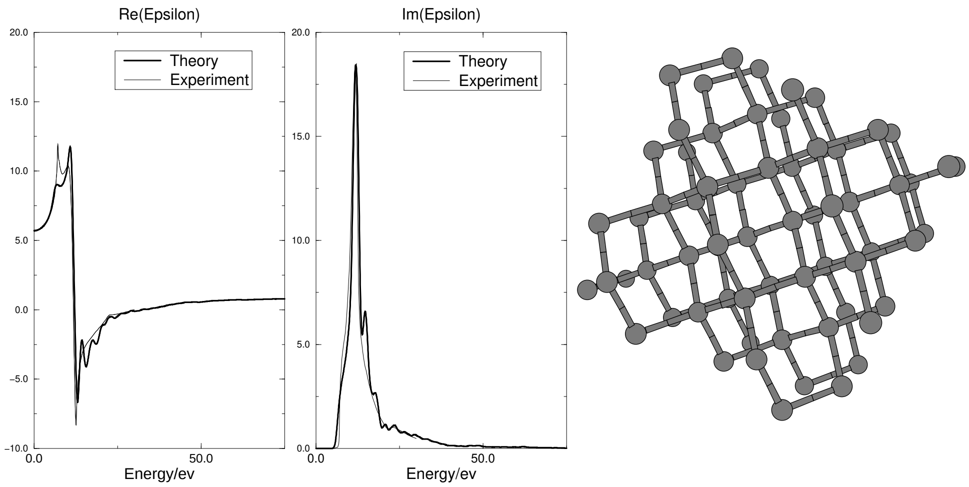


Figure C.3: The complex dielectric function was calculated with a 600ev cutoff, 24 bands, a 15x15x15 Monkhurst-Pack grid, a 0.5ev Gaussian smearing and a 0.7ev Scissor operator.

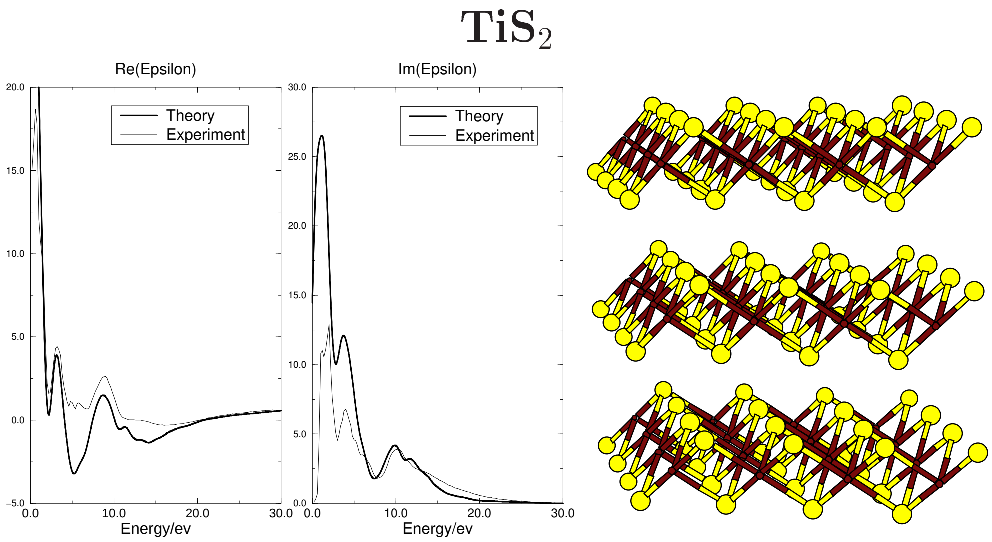


Figure C.4: The complex dielectric function was calculated with a 400ev cutoff, 32 bands, a 11x11x6 Monkhurst-Pack grid, a 0.5ev Gaussian smearing and a 0.0ev Scissor operator.

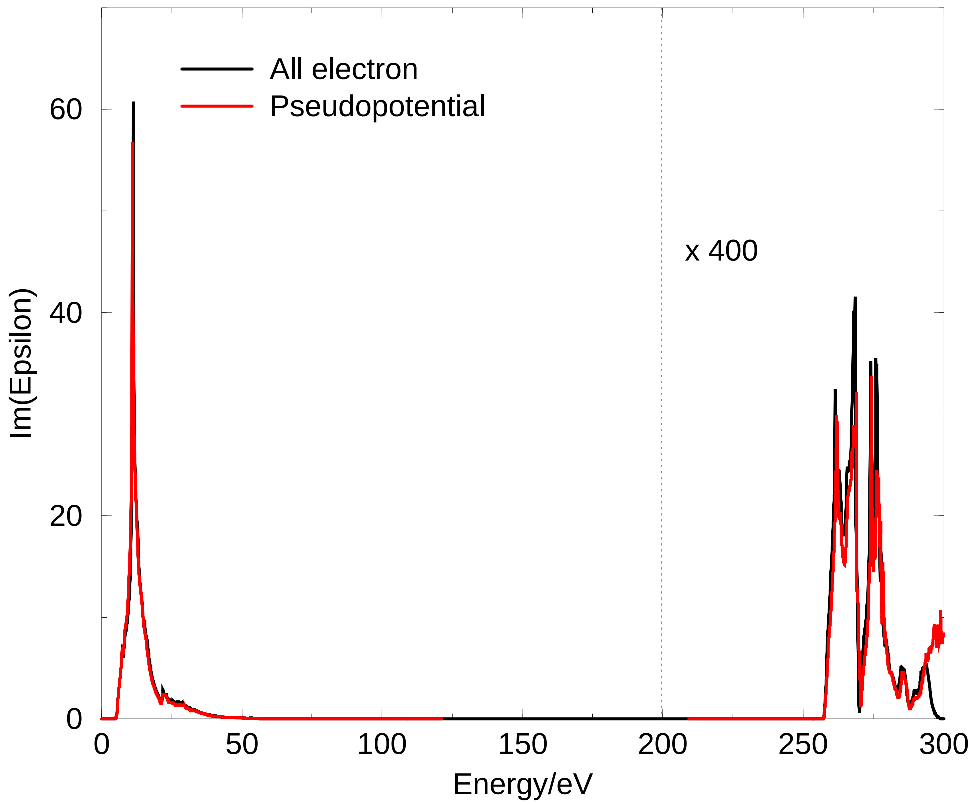


Figure C.5: Imaginary part of the dielectric function for Diamond — evaluated using both a bare coulomb potential and pseudopotential. The K-edge is scaled by 400 so as to be comparable to the low loss region.

Bibliography

[1] R. Ahuja, P. A. Brühwiler, J. M. Wills, B. Johansson, N. Martensson, and O. Eriksson. Theoretical and experimental study of the Graphite 1s X-ray absorption edges. *Phys. Rev. B*, 54(20):14396–14404, November 1996.

[2] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Saunders College, Philadelphia, 1976.

[3] A. Baldereschi. Mean value point in the Brillouin zone. *Phys. Rev. B*, 7:5212, 1973.

[4] P. E. Batson. Carbon 1s near-edge-absorption fine structure in graphite. *Phys. Rev. B*, 48(4):2608–2610, July 1993.

[5] P. E. Batson. Distortion of the core exciton by the swift electron and plasmon wake in spatially resolved electron-energy-loss scattering. *Phys. Rev. B*, 47(12):6898–6910, March 1993.

[6] P. E. Batson. Simultaneous stem imaging and electron energy-loss spectroscopy with atomic-column sensitivity. *Nature*, 366:727–728, 23/30 December 1993.

[7] P. E. Batson and J. Bruley. Dynamic screening of the core exciton by swift electrons in electron-energy-loss scattering. *Phys. Rev. Lett.*, 67(3):350–353, July 1991.

[8] L. Bellaiche and K. Kunc. Effect of pseudopotentials on Compton profiles and reconstruction of the true electron density. In *Abstract book*. DFT '95 Paris, August 1995. Abstract.

[9] P. Blaha, K. Schwarz, P. Sorantin, and S. B. Trickey. Full-potential, linearized augmented plane-wave programs for crystalline systems. *Comput. Phys. Commun.*, 59:399, 1990.

[10] P. E. Blöchl, O. Jepsen, and O.K. Andersen. Improved tetrahedron method for Brillouin-zone integrations. *Phys. Rev. B*, 49(23):16223–16233, June 1994.

[11] P.E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50(24):17953–17979, December 1994.

[12] M.H. Boon, M.S. Methfessel, and F.M. Mueller. Singular integrals over the Brillouin zone: the analytic quadratic method for the density of states. *J. Phys. C: Solid State Phys.*, 19:5337–5364, 1986.

[13] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Ann. Physik*, 84:457, 1927.

[14] T. B. Boykin. Incorporation of incompleteness in the **k.p** perturbation theory. *Phys. Rev. B*, 52(23):16317–16320, December 1995.

[15] P. Brohan. *Ab-initio studies of two-level systems in glasses and electron energy-loss spectra.* PhD thesis, Cambridge University, December 1993.

[16] N. D. Browning, J. Yuan, and L. M. Brown. Theoretical determination of angularly-integrated energy loss functions for anisotropic materials. *Phil. Mag. A*, 67(1):261–271, 1993.

[17] R. Car and M. Parrinello. Unified approach for molecular dynamics and density functional theory. *Phys. Rev. Lett.*, 55:2471, 1985.

[18] M. Cardona and F. H. Pollak. Energy-band structure of Germanium and Silicon : The **k.p** method. *Phys. Rev.*, 142(2):530–543, February 1966.

[19] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566, 1980.

[20] D. J. Chadi. Special points for Brillouin zone integrations. *Phys. Rev. B*, 16:1746, 1977.

[21] D. J. Chadi and M. L. Cohen. Special points in the Brillouin zone. *Phys. Rev. B*, 8:5747, 1973.

[22] J. C. Charlier, X. Gonze, and J. P. Michenaud. First-principles study of the electronic properties of graphite. *Phys. Rev. B*, 43(6):4579–4589, February 1991.

[23] M. L. Cohen and V. Heine. Pseudopotential theory of cohesion and structure. In *Solid State Physics, Vol. 24*, page 250. Academic Press, New York, 1970.

[24] G. Comelli, J. Stöhr, C. J. Robinson, and W. Jark. Structural studies of argon-sputtered amorphous carbon films by means of extended X-ray-absorption fine structure. *Phys. Rev. B*, 38(11):7511–7519, October 1988.

[25] C. G. Van de Walle and P. E. Blöchl. First-principles calculations of hyperfine parameters. *Phys. Rev. B*, 47(8):4244–4255, February 1993.

[26] P. J. Durham, J. B. Pendry, and C. H. Hodges. Calculation of X-ray absorption near-edge structure, XANES. *Comput. Phys. Commun.*, 25:193–205, 1982.

[27] R. F. Egerton. *Electron Energy Loss Sectroscopy in the Electron Microscope.* Plenum, 1986.

[28] H. Ehrenreich and M. H. Cohen. Self-consistent field approach to the many-electron problem. *Phys. Rev.*, 115(4):786–790, August 1959.

[29] L.A. Feldkamp, S.S. Shinozaki, C.A. Kukkonen, and S. P. Faile. Electron energy-loss spectroscopy of TiS_2. *Phys. Rev. B*, 19(4):2291–2294, February 1979.

[30] R. P. Feynman. Forces in molecules. *Phys. Rev.*, 56:340, 1939.

[31] J. Fink. Transmission electron energy-loss spectroscopy. In J. C. Fuggle and J. E. Inglesfield, editors, *Unoccupied Electronic State - Fundamentals for XANES, EELS, IPS and BIS*, volume 69 of *Topics in Applied Physics*, chapter 7, pages 204–241. Springer-Verlag, 1992.

[32] D. A. Fischer, R. M. Wentzcovitch, R. G. Carr, A. Continenza, and A. J. Freeman. Graphitic interlayer states: A carbon K near-edge X-ray-absorption fine-structure study. *Phys. Rev. B*, 44(3):1427–1429, July 1991.

[33] J. R. Gardner and N. A. W. Holzwarth. Pseudopotential inversion scheme. *Phys. Rev. B*, 33(10):7139–7143, May 1986.

[34] M. N. Gibbs. *Bayesian Gaussian processes for regression and classification.* PhD thesis, Cambridge University, September 1997.

[35] R. W. Godby. Exchange and correlation in solids. In J. C. Fuggle and J. E. Inglesfield, editors, *Unoccupied Electronic States — fundamentals for XANES, EELS, IPS and BIS*, volume 69 of *Topics in Applied Physics*, chapter 3, pages 51–87. Springer-Verlag, 1992.

[36] R. W. Godby, M. Schluter, and L. J. Sham. Accurate echange-correlation potential for Silicon and its discontinuity on an addition of an electron. *Phys. Rev. Lett.*, 56(22):2415–2418, June 1986.

[37] M. Grioni, J. F. van Acker, M. T. Czyžyk, and J. C. Fuggle. Unoccupied electronic structure and core-hole effects in the X-ray-absorption spectra of Cu_2O. *Phys. Rev. B*, 45(7):3309–3318, February 1992.

[38] O. Gunnarson and B. I. Lundqvist. Exchange and correlation in atoms, molecules and solids by the spin density functional formalism. *Phys. Rev. B*, 13:4274, 1976.

[39] D. R. Hamann, M. Schluter, and C. Chiang. Norm-conserving pseudopotentials. *Phys. Rev. Lett.*, 43:1494–1497, 1979.

[40] C. Hartwigsen, W. Witschel, and E. Spohr. Charge density and charge transfer in stage-1 alkali-graphite intercalation compounds. *Phys. Rev. B*, 55(8):4953–4959, February 1997.

[41] V. Heine. The pseudopotential concept. In *Solid State Physics, Vol. 24*, page 1. Academic Press, New York, 1970.

[42] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864, 1964.

[43] N. A. W. Holzwarth, S. G. Louie, and S. Rabii. X-ray form factors and the electronic structure of graphite. *Phys. Rev. B*, 26(10):5382–5390, November 1982.

[44] L. Van Hove. The occurrence of singularities in the eleastc frequency distribution of a crystal. *Phys. Rev.*, 89:1189, 1953.

[45] J. L. P. Hughes and J. E. Sipe. Calculation of second-order optical response in semiconductors. *Phys. Rev. B*, 53:10751–10763, 1996.

[46] J. Ihm, A. Zunger, and M. L. Cohen. Momentum space formalism for the total energy of solids. *J. Phys. C: Solid State Physics*, 12:4409, 1979.

[47] R. O. Jones and O. Gunnarsson. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61(3):689–746, July 1989.

[48] Y.-M. Juan and E. Kaxiras. Application of gradient corrections to density functional theory of atoms and solids. *Phys. Rev. B*, 48:14944–14952, 1993.

[49] E.O. Kane. Chapter 3: **k.p** method. *Semiconductors Semimetals*, 1:75–100, 1963.

[50] L. Kleinman and D.M. Bylander. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.*, 48(20):1425–1428, May 1982.

[51] W. Kohn and L. J. Sham. Self consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133, 1965.

[52] L. D. Landau and E. M. Lifshitz. *Quantum Mechanics — non-relativistic*, volume 3 of *Course of Theoretical Physics*. Pergamon Press, 1965.

[53] R. D. Leapman, P. L. Fejes, and J. Silcox. Orientation dependence of core edges from anisotropic materials determined by inelastic scattering of fast electrons. *Phys. Rev. B*, 28(5):2361–2373, September 1983.

[54] M.-H. Lee. *Improved optimised pseudopotentials and application to disorder in* $\gamma-$ Al_2O_3. PhD thesis, University of Cambridge, 1995.

[55] M.-H. Lee, J. S. Lin, M. C. Payne, V. Heine, V. Milman, and S. Crampin. Optimised pseudopotentials with kinetic energy filter tuning. *Phys. Rev. B*, 1997. (submitted).

[56] G. Lehmann and M. Taut. On the numerical claculation of the density of states and related properties. *Phys. Status Solidi*, 54:469, 1972.

[57] J. P. Loehr. Parameter consistency in multienergetic **k.p** models. *Phys. Rev. B*, 52(4):2374–2380, July 1995.

[58] J. M. Maclaren, S. Crampin, D. Vvedensky, and J. B. Pendry. Layer Korringa-Kohn-Rostoker technique for surface and interface electronic-properties. *Phys. Rev. B*, 40:12164–12175, 1989.

[59] F. Mauri and R. Car. First-principles study of excitonic self-trapping in diamond. *Phys. Rev. Lett.*, 75(17):3166–3169, October 1995.

[60] A. J. Maxwell, P. A. Brühwiler, D. Arvanitis, J. Hasselström, and N. Mårtensson. Length scales in core level excitation: C 1s X-ray absorption of C_{60}/Al. *Phys. Rev. Lett.*, 79(8):1567–1570, August 1997.

[61] D. G. McCulloch and R. Brydson. Carbon K-shell near-edge structure calculations for graphite using the multiple-scattering approach. *J. Phys.: Condens. Matter*, 8:3835–3841, 1996.

[62] M.S. Methfessel, M.H. Boon, and F.M. Mueller. Analytic-quaratic method of calculating the denstiy of states. *J. Phys. C: Solid State Phys.*, 16:l949–l954, 1983.

[63] M.S. Methfessel, M.H. Boon, and F.M. Mueller. Singular integrals over the Brillouin zone: inclusion of **k**-dependent matrix elements. *J. Phys. C: Solid State Phys.*, 20:1069–1077, 1987.

[64] B. Meyer, K. Hummler, C. Elsässer, and M. Fahnle. Reconstruction of the true wavefunctions from the pseudowavefunctions in a crystal and calculation of electric field gradients. *Journal of Physics - Condensed Matter*, 7(48):9201–9217, 1995.

[65] H. J. Monkhorst and J. D. Pack. Special poinits for Brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188–5192, June 1976.

[66] P. Moreau, C. Pickard, and T. Duguet. Anisotropy in EELS of TiS_2. In *Electron Microscopy and Analysis 1997*, Institute of Physics Conference Series, 1997.

[67] D. A. Muller, S. Subramanian, P. E. Batson, J. Silcox, and S. L. Sass. Structure, chemistry and bonding at grain-boundaries in Ni_3Al .1. The role of Boron in ductilizing grain-boundaries. *Acta Materialia*, 44(4):1647–1655, 1996.

[68] D. A. Muller, Y. Tzou, R. Raj, and J. Silcox. Mapping of sp^2 and sp^3 states of carbon at sub-nanometre spatial resolution. *Nature*, 366:725–727, 23/30 December 1993.

[69] J. E. Müller and J. W. Wilkins. Band-structure approach to the X-ray spectra of metals. *Phys. Rev. B*, 29(8):4331–4348, April 1984.

[70] J. D. Pack and H. J. Monkhorst. Special points Brillouin zone integrations - a reply. *Phys. Rev. B*, 16:1748, 1977.

[71] E. D. Palik, editor. *Handbook of Optical Constants of Solids.* Academic Press, Inc., 1985.

[72] A. Pasquarello, M. S. Hybertsen, and R. Car. Si 2p core-level shifts at the Si(001)-SiO_2 interface: A first principles study. *Phys. Rev. Lett.*, 74(6):1024–1027, February 1995.

[73] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos. Iterative minimisation techniques for *ab initio* total energy calculations. *Rev. Mod. Phys.*, 64(4):1045–1097, October 1992.

[74] J. Perdew and A. Zunger. Self interaction correction to density functional approximations for many electron systems. *Phys. Rev. B*, 23:5048, 1981.

[75] J. P. Perdew. In P. Zeische and H. Eschrig, editors, *Electronic Structure of Solids '91.* Akademie Verlag, Berlin, 1991.

[76] P. M. Platzman and P. A. Wolff. *Solid State Phys. Suppl.*, 13:1, 1973.

[77] M. Posternak, A. Baldereschi, A. J. Freeman, E. Wimmer, and M. Weinert. Prediction of electronic interlayer states in graphite and reinterpretation of alkali bands in graphite intercalation compounds. *Phys. Rev. Lett.*, 50(10):761–764, March 1983.

[78] P. Pulay. *Ab initio* calculation of force constants and equilibrium geometries in polyatomic molecules. I. theory. *Mol. Phys.*, 17:197, 1969.

[79] A. J. Read and R. J. Needs. Calculation of optical matrix elements with nonlocal pseudopotentials. *Phys. Rev. B*, 44(23):13071–13073, December 1991.

[80] P. Rez. Energy loss fine structure. In M. Disko, C. Ahn, and B. Fultz, editors, *Transmission Electron Energy Loss Spectroscopy in Materials Science*, EMPMD Monograph Series, chapter 5, page 110. The Minerals, Metals & Materials Society, 1992.

[81] P. Rez, J. Bruley, P. Brohan, M. C. Payne, and L. A. J. Garvie. Review of methods for calculating near edge structure. *Ultramicroscopy*, 59:159–167, 1995.

[82] P. Rez, J.C. Rendall, and M.C. Payne. Estimate of the accuracy of calculated charge densities. *Optik*, 98(3):135–140, 1995.

[83] I.J. Robertson and M.C. Payne. **k**-point sampling and the **k.p** method in pseudopotential total energy calculations. *J. Phys. : Condens. Matter*, 2:9837–9852, 1990.

[84] D. K. Saldin and Y. Ueda. Dipole approximation in electron-energy-loss spectroscopy: L-shell excitations. *Phys. Rev. B*, 46(9):5100–5109, September 1992.

[85] M. D. Segall, C. J. Pickard, R. Shah, and M. C. Payne. Population analysis in plane wave electronic structure calculations. *Mol. Phys.*, 89(2):571, 1996.

[86] M. D. Segall, R. Shah, C. J. Pickard, and M. C. Payne. Population analysis of plane wave electronic structure calculations of bulk materials. *Phys. Rev. B*, 54(23):16317, 1996.

[87] F. Seitz. *Modern Theory of Solids*, page 352. McGraw-Hill, New York, 1940.

[88] E. L. Shirley. Optimal basis sets for detailed Brillouin-zone integrations. *Phys. Rev. B*, 54(23):16464–16469, December 1996.

[89] W. Shockley. Energy band structures in semiconductors. *Phys. Rev.*, ?(?):173–174, 1950.

[90] W. Speier, J. F. van Acker, and R. Zeller. Electronic structure of Z+1 impurities in metals. *Phys. Rev. B*, 41(5):2753–2757, February 1990.

[91] A. F. Starace. Length and velocity formulas in approximate oscillator-strength calculations. *Phys. Rev. A*, 3(4):1242–1245, April 1971.

[92] B. K. Teo and D. C. Joy. *EXAFS Spectroscopy, techniques and applications*. Plenum, New York, 1981.

[93] M.P. Teter, M.C. Payne, and D. C. Allan. Solution of Schrödinger's equation for large systems. *Phys. Rev. B*, 40:12255, 1989.

[94] J. Vackář and A. Šimůnek. Modelling the valence electronic structure of the core region of an atom in a solid within a local-density approximation pseudopotential framework: reintroduction of the full nodal form. *J. Phy.: Condens. Matter*, 6:3025–3030, 1994.

[95] J. F. van Acker, W. Speier, J. C. Fuggle, and R. Zeller. Paramerization of the electronic structure of Z+1 impurities. *Phys. Rev. B*, 43(17):13916–13924, June 1991.

[96] G. van der Laan and B. T. Thole. Strong magnetic-X-ray dichroism in 2p absorption-spectra of 3d transition-metal ions. *Phys. Rev. B*, 43:13401, 1991.

[97] A. Šimůnek, J. Vackář, and G. Wiech. Local s, p and d charge distributions and X-ray emission bands of SiO_2: α-Quartz and Stishovite. *J. Phys.: Condens. Matter*, 5:867–874, 1993.

[98] D. D. Vvedensky. Theory of X-ray absorption fine structure. In J. C. Fuggle and J. E. Inglesfield, editors, *Unoccupied Electronic States- Fundamentals for XANES,EELS,IPS and BIS*, volume 69 of *Topics in Applied Physics*, chapter 4, pages 139–176. Springer-Verlag, 1992.

[99] C.S. Wang and J. Callaway. Band structure of nickel : Spin-orbit coupling, the Fermi surface, and the optical conductivity. *Phys. Rev. B*, 9(11):4897–4907, June 1974.

[100] P. J. W. Weijs, M. T. Czyzyk, J. F. van Acker, W. Speier, J. B. Goedkoop, H. van Leuken, H. J. M. Hendrix, R. A. de Groot, G. van der Laan, K. H. J. Buschow, G. Wiech, and J. C. Fuggle. Core-hole effects in the X-ray-absorption spectra of transition-metal silicides. *Phys. Rev. B*, 41(17):11899–11910, June 1990.

[101] X. Weng, P. Rez, and H. Ma. Carbon K-shell near-edge structure: Multiple scattering and band-theory calculations. *Phys. Rev. B*, 40(6):4175–4178, August 1989.

[102] X. Weng, P. Rez, and O. F. Sankey. Pseudo-atomic-orbital band theory applied to electron-energy-loss near-edge structures. *Phys. Rev. B*, 40(8):5694, September 1989.

[103] J. Zaanen. Theory of correlated holes (and electrons): from satellites to Luttinger liquids. In J. C. Fuggle and J. E. Inglesfield, editors, *Unoccupied electronic states — fundamentals for XANES, EELS, IPS and BIS*, volume 69 of *Topics in Applied Physics*, chapter 4, pages 89–137. Springer-Verlag, 1992.

[104] R. Zeller. Band-structure methods. In J. C. Fuggle and J. E. Inglesfield, editors, *Unoccupied Electronic States — fundamentals for XANES, EELS, IPS and BIS*, volume 69 of *Topics in Applied Physics*, chapter 2, pages 25–49. Springer-Verlag, 1992.

[105] Y. Zou and J. C. Tang. Multiple-scattering approaches to carbon K-shell near-edge X-ray absorption fine structure of graphite. *J. Phys.: Condens. Matter*, 6:2949–2956, 1994.