Aalborg University
Department of Physics and Nanotechnology

Microscopic Theory of Linear & Nonlinear Optical Response: Zinc-blende semiconductors

Mads Lund Trolle
Master Thesis

Department of Physics and Nanotechnology
Skjernvej 4A
9220 Aalborg Øst
Phone: +45 9940 9215
http://physics.aau.dk

Title: Microscopic Theory of Linear & Nonlinear Optical Response:
Zinc-blende semiconductors
Author: Mads Lund Trolle
Supervisor: Thomas Garm Pedersen
Number of Copies: 4
Number of Pages: 116
Number of Appendices: 2

Abstract

The empirical pseudopotential method, both including and excluding spin-orbit interaction, is applied to the calculation of the optical properties of the zinc-blende semiconductors GaAs, InSb, Ge and Si. The linear-analytic tetrahedron method is applied for the necessary Brillouin zone integration, and spectra of the linear susceptibility for frequencies in the optical range is calculated for the aforementioned materials. Optical spectra including spatial dispersion, i.e. including the photon wave vector, is calculated and presented for Ge and Si. The bulk dipole second harmonic susceptibility tensor is calculated for GaAs and InSb, followed by investigation of the effects of spatial dispersion on the second harmonic response tensor of Si and Ge. Depletion-electric-field-induced second harmonic generation (DEFISHG) is considered, and a model of the surface depletion layer is presented using parameters for GaAs. The DEFISHG tensor is estimated for GaAs, InSb, Ge and Si. To investigate effects from semiconductor surfaces and semiconductor thin-films, the empirical pseudopotential model is adapted for two dimensional thin-films, neglecting reconstruction. Results for Ge and Si is presented as surface band structures and density of states calculations. Using this model, linear spectra of Ge and Si thin-films is calculated for both the in-plane and out-of-plane response. A method for calculating surface second harmonic generation is presented, and results for Ge is shown. The macroscopic, nonlinear optical properties of a dielectric, semi-infinite medium with a surface and an infinite slab of finite thickness is reviewed, and expressions relating the second harmonic susceptibility tensor to the fundamental beam of a second harmonic generation experimental setup is derived. Methods and procedures used for experimental second harmonic generation measurements on a GaAs wafer during this project is presented, and the results are related to the theoretical results calculated here, in addition to results from other authors who include excitonic effects.

Institut for fysik og nanoteknologi
Skjernvej 4A
9220 Aalborg Øst
Telefon: +45 9940 9215
http://physics.aau.dk

Title: Mikroskopisk Modellering af den Optiske Respons:
Zinc-blende halvlederer
Forfatter: Mads Lund Trolle
Vejleder: Thomas Garm Pedersen
Antal kopier: 4
Sidetal: 116
Antal appendikser: 2

Resumé

Den empiriske pseudopotentialmetode anvendes som udgangspunkt til simulation af zinc-blende halvlederne GaAs, InSb, Si og Ge, både inkluderende of ekskluderende spin-bane kobling. De lineære optiske egenskaber undersøges ved at beregne den frekvensafhængige elektriske susceptibilitet. De nødvendige Brillouinzoneintegraler udføres vha. den lineær-analytiske tetrahedron metode, og de optiske spektrer for GaAs, InSb, Ge og Si præsenteres. Effekten af rumlig dispersion, dvs. inklusion of fotonens bølgevektor, undersøges for Ge og Si, og resultater præsenteres som frekvens og fotonbølgevektorafhængige responsfunktioner. Bulk andenharmonisk generation modelleres i dipoltilnærmelsen for GaAs og InSb, og for Ge og Si inkluderende rumlig dispersion. Elektrisk feltinduceret andenharmonisk generation som følge af overfladeudtømningslaget i Ge, Si, InSb og GaAs undersøges. For at undersøge halvlederes optiske overfladeeffekter, samt for at undersøge halvledertyndfilms optiske respons, modificeres den empiriske pseudopotentialmetode så tyndfilm kan beskrives. Vha. denne model beregnes den lineære optiske susceptibilitet for polarisation både i og ud af tyndfilmsplanet. En metode til beregning af andenharmonisk generation fra overflader udledes, og resultater for Ge præsenteres. Den makroskopiske, ulineære, optiske model for grænseflade mellem et semiuendeligt dielektrikum og vakuum beregnes, og tilfældet for et ulineært medium af endelig tykkelse betragtes. Til sidst præsenteres de eksperimentelle metoder samt resultater. Disse perspektiveres til de beregnede spektrer samt spektrer fra anden kilde, der medtager eksitoner.

Preface

The text at hand documents my master thesis in nanophysics and materials in which I have explored the realm of computational semiconductor response theory.

The work have been motivated, on one hand, by the technological applicability of semiconductors and the methods used for their theoretical and experimental characterization, but also from a desire to answer a question which have been bugging me since my first introductory course on quantum mechanics and solid state physics: What kind of information, with real world technological implications, can actually be derived from the abstract concept of a band structure?

In this text, some of the linear and nonlinear optical properties of the semiconductors Si, Ge, GaAs and InSb have been investigated theoretically by application of quantum mechanical response theory. Heavy numerical work was done in Fortran 2003, using the Intel Fortran compiler and the Intel Math Kernel Library on a Linux platform, while light numerical work was done in Matlab. Several thousand lines of code was written during the project, and I would be pleased to forward parts of it to any who might be interested*.

I would like to express my gratitude towards Prof. Kjeld Pedersen, for invaluable aid concerning the experimental work of this project, and especially to Prof. Thomas Garm Pedersen for many fruitful discussions, for always having an open door and taking the time to explain the intricacies of nanophysics.

*email: mlt07@student.aau.dk

Contents

1 Introduction

It would be difficult to imagine a world absent of semiconductors; every electronic apparatus, be it communications device or toaster, every optical system, be it photo detector or television screen, is today dependent on the remarkable material attributes of these compounds. In terms of its electronic properties, the defining feature of a semiconductor is its band gap, separating occupied and empty states, small enough for the exponentially decaying tail of a Fermi-Dirac distribution function populating electronic states to extend into the conduction bands upon heating or application of external fields, thus providing means of e.g. tuning the conductivity by external stimuli. This concept is readily applied in the field effect transistor; perhaps the most influential semiconductor application, whose development paved the way for the silicon revolution of the digital era. In such devices an external gate potential is used to regulate the resistance of a semiconductor junction, allowing the ”off” and ”on” switching necessary for logical circuits. In junctions between P and N doped semiconductors, a purely statistical ”diffusion current” driven by the electron/hole concentration gradient will tend to create a static electric field perpendicular to the interface, creating a diode structure. These types of PN diodes are useful in terms of their purely electrical characteristics , however, they also display remarkably versatile abilities as emitters of both coherent and incoherent electromagnetic radiation and as optical detectors and power generation devices. Basically, the light emitting optical properties of these devices are due to the large concentration of minority carriers in the junction region, hence facilitating effective recombination centers where electrons and holes recombine upon excitation of a photon. The photovoltaic applications of the PN junction essentially works by the opposite rationale; photons are absorbed upon generation of electron/hole pairs in the depletion region. The electric field of the junction will then tend to separate the charge carriers, producing a current source which can be applied for power generation or detector purposes.

With the advent of fabrication techniques allowing industrial production of nano-engineered systems, such as the lithographic methods applied by e.g. Intel for integrated circuit production, a question naturally arises: What can we actually accomplish by engineering on the nanoscale? Theoretical models based on the foundations of modern physics developed in the previous century and its subsequent application on crystalline solids can be applied as a direct answer to this question. Purely analytical models, while being instructive on the fundamental properties of nanoscale systems, can never hope to fully explain more complicated systems of technological interest. However, with the advent of modern computers numerical simulations of complicated systems is now possible, yielding results agreeing well with experiments. While successful modeling of systems on the nanoscale is scientifically rewarding in its own right, it is also clear that such work is of obvious technological interest.

There are many possible avenues to explore in this context, and this work shall mainly be concerned with modeling the optical properties of the arch typical group IV semiconductors silicon (Si), germanium (Ge), and the group III-V semiconductors gallium arsenide (GaAs) and indium antimonide (InSb). These are chosen because the first three represent examples of semiconductors extensively used throughout the industry in addition to representing direct gap (GaAs) and indirect gap (Si and Ge) semiconductors, while InSb is an example of

a compound displaying a very large spin-orbit splitting, which will be discussed later.

State of the art

The objective of this section is to provide an overview on what has been done in the area of modeling the optical response of semiconductors within the one-electron, band-theoretical framework. In 1961 Adler[1] published an extensive text on the quantum theory of the dielectric constant of solids. This article was one of the first to collect the full band theory of the linear optical response in real solids, expressing the dielectric tensor within the density matrix formalism, building on the work of Cohen & Ehrenreich (1959)[2] and Nozieres & Pines (1958)[3]. It was recognized that both the transverse, longitudinal and off-diagonal terms of the response tensor yielded finite results, with the response becoming isotropic for cubic systems in the "optical limit" of vanishing photon wavevector. Umklapp-type scattering processes was recognized, leading to so called "local-field effects". In the area of linear optics, the work by Adler is still the formalism of choice when calculating the optical response of materials from one-electron band structures. The foundations of these microscopic models is the one-electron band structure. Before the development of computers, simplified models had to be considered, such as the classical Lorentz oscillator model. The first good band structures appeared in the late 60's using the empirical pseudopotential approach[4]. Following this, linear dielectric constants of various materials could be calculated in the optical limit , with examples being GaAs[5], ZnTe and ZnSe[6] and Si[7], where the latter includes the so called "local field-effects". Non-linear optics follows naturally from taking the perturbation theory used in deriving the linear response to higher orders. With the observation of second harmonic generation in 1961 by Franken *et al.*[8], following the development of high-powered laser systems, a surge of interest within the field of nonlinear optics ensued. In Frankens experiment, the optical field at 348nm generated by the second harmonic polarization response of quartz to a fundamental pump beam at 695nm from a ruby laser was measured. However, characterization of the frequency dependent microscopic nonlinear response by e.g. reflection spectroscopy is made more complicated than for the linear case due to increased complexity of the optical system. Extensive work in this area was pioneered by e.g. N. Bloembergen *et al.*[9;10]. Here it was shown from Maxwell's equations that the harmonically generated fields should be understood in terms of inhomogeneous coupled wave equations, where the nonlinear polarization depending explicitly on elements of the nonlinear susceptibility tensor entered as a source term. Solving the boundary value problem of the resulting fields yielded the radiated harmonic fields in terms of the incoming fundamental fields and the nonlinear response tensor.

The microscopic theory of the nonlinear susceptibility tensor was calculated numerically with the introduction of computers in the 1970's, by e.g. Bethune *et al.*[11], however without sufficient accuracy to achieve convergent results. Later, researchers such as Sipe and Moss calculated the second harmonic response tensors of a wide range of semiconductors[12;13]. These calculations do not agree as well with experiment as do *ab initio* modeling of the linear optical properties. In addition, experimental investigation of the fundamentally appealing case of harmonically generated frequencies near the absorption edge (i.e. the edge where the harmonic response tensor becomes purely real) remains scarce, even though this is the textbook example for understanding the linear response of semiconductors. This could be due to the demanding optical systems required, i.e. fundamental wavelengths in the μm range is needed whereas detectors must work at multiples of this frequency.

This work

The text at hand is concerned with the theory behind ,and numerical simulation of, the linear an non-linear optical processes of zinc-blende semiconductors. This report have been divided into the following chapters; in chapter 2, the dipole-allowed linear optical properties of the semiconductors mentioned will be reviewed based on band structures calculated by the empirical pseudo potential method (EPM) and Brillouin zone integrations performed by the linear-analytic tetrahedron method (LATM), first neglecting the effects of spin-orbit interaction. These effects will be introduced in the EPM potential, providing a non-local correction to the pseudopotential and the effects of this on the linear optical spectra is considered. Moving beyond the dipole approximation, the non-local, so called quadrupole effects are taken into account by considering the response to photons of finite wave vectors, and the significance of these effects are considered with respect to nano-optics. In chapter two, second harmonic generation will be considered and the second harmonic response tensors of the semiconductors under investigation will be calculated using the LATM generalized for nonlinear response calculations. Again the effects of including spin-orbit interaction will be considered relative to neglecting them. Moving on, the nonlocal, quadrupole effects in nonlinear optics is considered. In chapter 4, the effects of confining the dimensions of a semiconductor will be reviewed by introducing a method of handling finite systems with the EPM. Linear optical spectra for semiconductor thin-films is presented, and second harmonic generation from surfaces is calculated. In chapter 5, an experimental method, and the the macroscopic optics necessary to understand the experimental results, for determining the absolute second harmonic response is presented and experimental results for GaAs is presented. The derivation of the optical response tensors applied throughout this report using quantum mechanical perturbation theory is found in Appendix A due to the tedious nature of the calculations. If the reader is interested, he is encouraged to reading the appendix following this introduction, however references are made to the appropriate sections wherever necessary. In Appendix B the linear-analytic tetrahedron method (LATM) applied to three-dimensional systems in linear response theory is found, due to the subject being covered in an earlier project.

2 Linear Optical Properties

In the present chapter, the linear bulk properties of the semiconductors silicon, germanium, indium-antimonide and gallium-arsenide will be reviewed. The optical spectra will be simulated numerically from band structures calculated using the EPM and the well known formula for the optical susceptibility in the single particle, random phase approximation[1] derived in App A.3

$$\chi''_\omega = \frac{e^2}{4\varepsilon_0 m^2 \pi^2 \omega^2} \sum_{c,v} \int |p^z_{cv}|^2 \delta(E_{cv} - \hbar\omega)\ d^3k, \quad T \to 0\text{K} \tag{2.1}$$

where e is the (positive) fundamental charge quantum, m is the free electron mass, ω is the field frequency, $\hbar$ Planck's reduced constant, ε_0 the vacuum permittivity, T temperature and a factor $\times 2$ have been included for spin-summation. c and v annotate unoccupied and occupied bands, respectively. p^z_{cv} represent the matrix element $p^z_{cv} = \langle c\vec{k}|\hat{p}_z|v\vec{k}\rangle$, whereas E_{cv} represent the energy difference $E_{cv} = E_c(\vec{k}) - E_v(\vec{k})$. The integration domain is the Brillouin zone, and the integral will be solved using the LATM, which is derived in App. B. Both the EPM and the LATM have been reviewed in previous work[14] based on the articles[4;15]. Having determined the imaginary part of the response, the real part is found by Kramers-Kronig transformation[16]

$$\chi'_\omega = \frac{2}{\pi}\, PV \int_0^\infty \frac{\omega' \chi''_{\omega'}}{\omega'^2 - \omega^2}\, d\omega'$$

where PV indicate that the Cauchy principal value of the integral is taken. The energy eigenvalues needed to evaluate E_{cv} is found by the EPM, where a Schrödinger-like equation is solved using an empirically determined pseudopotential instead of the true potential

$$\left[\frac{\hat{p}^2}{2m} + V_{ps}(\vec{r})\right] |\alpha\vec{k}\rangle = E_\alpha(\vec{k})|\alpha\vec{k}\rangle, \tag{2.2}$$

using a plane wave basis

$$|\alpha\vec{k}\rangle = \sum_{\vec{G}} f^\alpha_{\vec{k}-\vec{G}} e^{i(\vec{k}-\vec{G})\cdot\vec{r}}. \tag{2.3}$$

It is noted that the pseudo Schrödinger equation above can be derived from the true Schrödinger equation by application of a plane wave basis orthogonalized to the localized core wavefunctions[17] (a so called "orthogonalized plane wave" (OPW) basis); since V_{ps} is determined empirically without regarding its OPW origins, we shall not consider this complication further aside from noting that the pseudo-wavefunctions are identical to the true wave functions outside the ionic core region, but much smoother than the true wave functions near the nuclei*. Insertion of (2.3) into (2.2) yields upon rearrangement

$$\frac{\hbar^2[\vec{k}-\vec{G}]^2}{2m} f^\alpha_{\vec{k}-\vec{G}} + \sum_{\vec{G}'} V_{\vec{G}-\vec{G}'} f^\alpha_{\vec{k}-\vec{G}'} = E_\alpha(\vec{k}) f^\alpha_{\vec{k}-\vec{G}}, \tag{2.4}$$

*A fact which becomes important when considering physical phenomena arising from this localized region of space, such as spin-orbit coupling

	a,[Å]	$v^S_{\sqrt{3}}$	$v^S_{\sqrt{8}}$	$v^S_{\sqrt{11}}$	$v^A_{\sqrt{3}}$	$v^A_{\sqrt{4}}$	$v^S_{\sqrt{11}}$
Si[18]	5.43	-3.0371	0.7508	0.9847	0	0	0
Ge[4]	5.66	-3.128	0.136	0.816	0	0	0
GaAs[19]	5.65	-3.196	0.160	0.680	1.020	0.816	0.680
InSb[4]	6.48	-2.720	0	0.544	0.816	0.680	0.136

Table 2.1: Parameters used for the bulk band structures in this report. All form factors are in units of eV, and the superscripts on the compounds are citations.

where $V_{\vec{G}}$ is the Fourier transform of the pseudopotential $V_{ps}(\vec{r}) = \sum_{\vec{G}} V_G e^{i\vec{G}\cdot\vec{r}}$. In zinc-blende crystals, an expansion of V_{ps} in atomic (local) contributions allow the Fourier components to be written on the form

$$V_{\vec{G}} = v^S_G \cos\vec{G}\cdot\vec{\tau} + iv^A_G \sin\vec{G}\cdot\vec{\tau}, \quad \vec{\tau} \equiv \frac{a}{8}(1,1,1),$$

where $v^S_G \equiv (v^{(1)}_G + v^{(2)}_G)/2$ is the symmetric form factor whereas $v^A_G \equiv (v^{(1)}_G - v^{(2)}_G)/2$ is the anti-symmetric form factor. The superscripts (1) and (2) denote the two basis atoms of the zinc-blende crystal structure and a is the lattice constant. Spherical symmetry of the atomic potentials causes only the length of the reciprocal vectors to be important, and the form factors are empirically determined parameters given by e.g. Cohen & Bergstresser (1966)[4]. The parameters used in this project is given in Tab. 2.1, and the band structures used is shown as the black lines of Fig. 2.3 (the red lines indicate band structures where spin-orbit interaction is included, to be discussed shortly). Solving (2.4) returns the (pseudo) wave functions explicitly on a plane wave form, allowing evaluation of p^z_{cv}

$$p^z_{cv} = -\hbar \sum_{\vec{G}} f^{c*}_{\vec{k}-\vec{G}} f^v_{\vec{k}-\vec{G}} G_z.$$

where $f^\alpha_{\vec{k}-\vec{G}}$ is returned as eigenvectors to the Hamilton matrix. Furthermore, the integration domain can be reduced to the irreducible Brillouin zone by performing the substitution

$$|p^z_{cv}|^2 \to \frac{48}{3}\left(|p^x_{cv}|^2 + |p^y_{cv}|^2 + |p^z_{cv}|^2\right) \equiv |p_{cv}|^2$$

valid for cubic solids. This procedure was checked numerically by comparison with results obtained upon integrating the entire Brillouin zone without substituting the matrix elements, and was found to give the same result. In the LATM, the integration domain is divided into contributions from tetrahedral subdomains. For each tetrahedron, the energy eigenvalues are interpolated based their values at the tetrahedron vertices, allowing analytic solution of the energy-linearized integral over the tetrahedron volume[†]. Results for a few semiconductors together with experimental results from Aspnes *et al.* (1983)[20] are shown in Fig. 2.1; the general trend of the curves are seen to agree well with experiment, however several discrepancies are apparent. The calculated peaks are generally much higher and narrower than for experimental values, which can be attributed broadening effects such as electron-phonon interaction. This is expected, since (2.1) is derived in the limit of vanishing broadening; a phenomenological broadening factor can be reintroduced by convolution of the spectrum with a Lorentzian line shape. The peak positions are not entirely accurate relative to the experimental spectra, which is ascribed to the simplicity of the local empirical

[†] see App. B

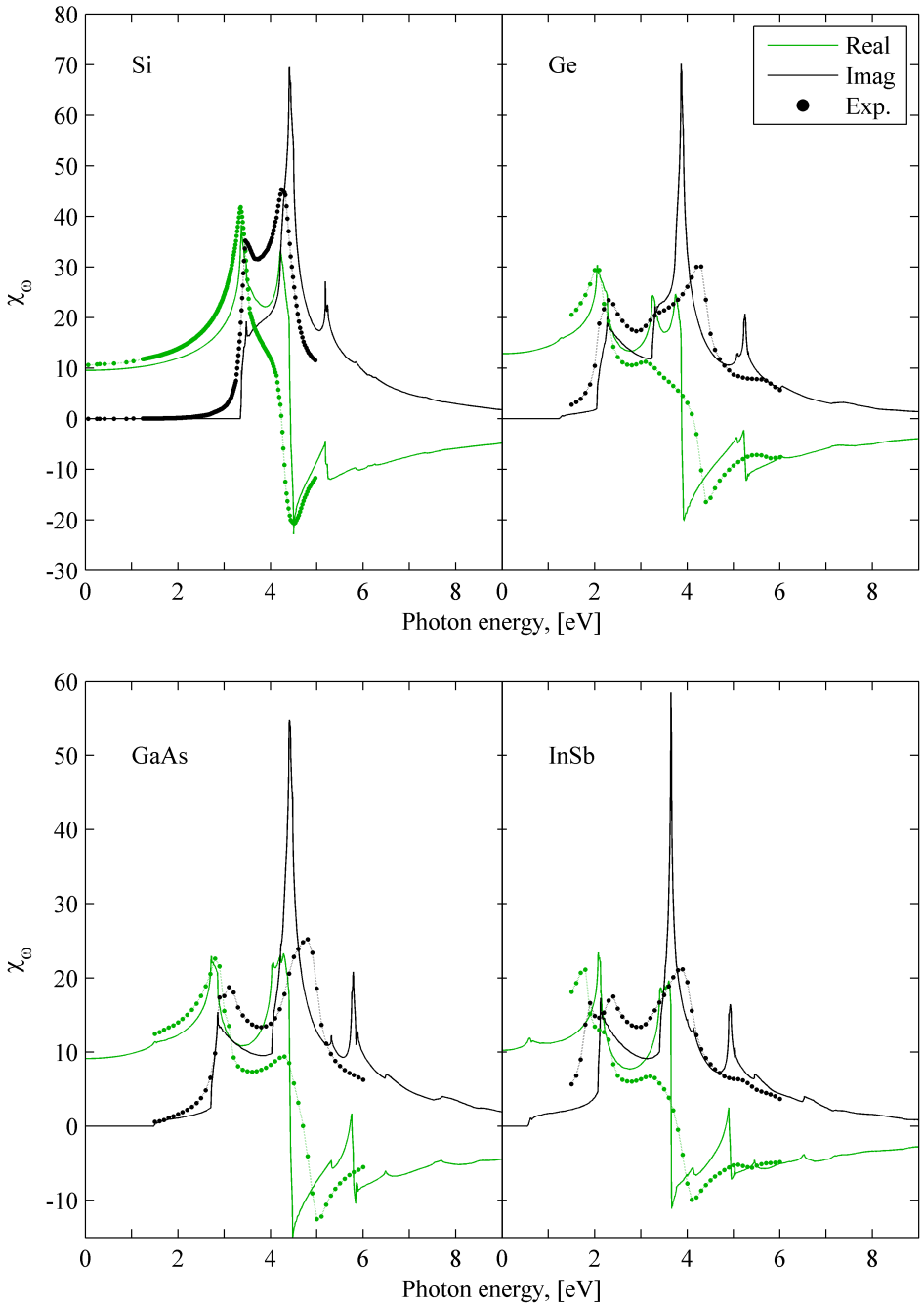


Figure 2.1: Linear spectra calculated using (2.1) from EPM band structures, neglecting spin-orbit coupling, using a basis of 137 plane waves and form factors from Tab. 2.1. 98.304 tetrahedral subdivision of the irreducible zone and energy intervals of 1meV was applied. Experimental spectra are from Aspnes *et al.* (1983)[20], except for Silicon, which is from the handbook by Palik[21].

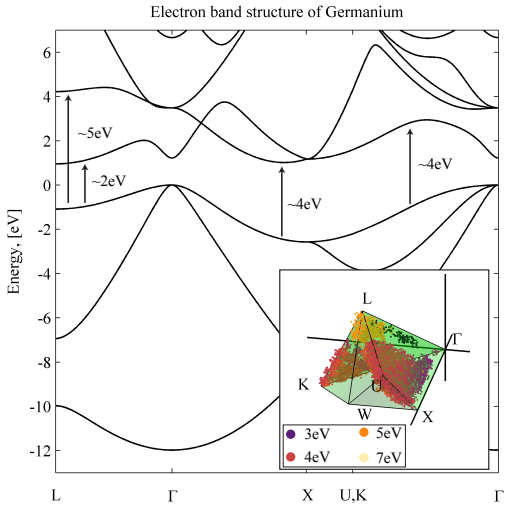


Figure 2.2: Band structure of Ge calculated by the local empirical pseudopotential method using a 137 plane wave basis. Inset: A sample of the k-points from the irreducible Brillouin zone having the largest contribution to the spectrum of Ge in Fig. 2.1. The color coding is used to indicate E_{cv}.

pseudopotentials, which fails to take several key phenomena into account, e.g. non-local and excitonic effects in addition to spin-orbit interaction. The latter is responsible for the splitting of the first peak in the experimental spectra of the heavier compounds, i.e. Si (#14 in the periodic table) displays almost no splitting, GaAs (#31/33) and Ge (#32)[‡] does while InSb (#49/51)displays a spin orbit splitting of almost 1eV. It should also be apparent that all compounds have several features in common; an absorption edge and three main peaks, where the second peak has the highest oscillator strength and where the third peak is only weakly represented in the experimental spectra. These features are intrinsic to the zinc-blende crystal structure, and originates from similar transitions. In the inset of Fig. 2.2 a set of sample points taken from within the irreducible Brillouin zone of Ge is shown. These points represent the top 0.5% contributors to the spectrum, sorted according to $\overline{|p_{cv}|^2} \times g$, where g is the contribution to the joint density of states of the tetrahedron containing the point and $\overline{|p_{cv}|^2}$ is the average value of $|p_{cv}|^2$ at the four tetrahedron vertices – hence these points represent the largest contributors to the peaks. The points have been color-coded according to E_{cv}, as shown in the legend. Comparing with the spectrum of Ge shown in Fig. 2.1, it is clear that the absorption edge is very weak. In a direct gap semiconductor, the transitions near the zone center responsible for the onset of absorption is termed the E0 transitions. The peak near 2eV arises from transitions predominantly along the ΓL-line termed E1 transitions. The global peak near 4eV is due to transitions at the ΓXL- and ΓKL-faces termed the E2 peak, where the E2 transitions are found along the ΓL-line. The 5eV peak is a consequence of transtions near the L-point. A few transitions near 7eV is also included, however any peak in this region is quenched by the ω^{-2} proportionality of χ''_ω. The E1 peak intensity is due to the parallel dispersion of the highest valence bands and the lowest conduction band along parts of the ΓL- and ΓK-lines. This gives a constant E_{cv} in $\vec{k}$, and thus a van Hove singularity in the joint density of states. Similar considerations can be applied for the two other main peaks. The shoulder on the E2 peak is due to the band gap narrowing slightly between the Γ and X points. Similar analyses leads to the same conclusions for the

[‡]The splitting for Ge is more apparent from the plot in the original article[20], however the data-points reported in the article does not allow the exact same plot here.

origins of the E0, E1, E2 and E3 peaks of the other semiconductors.

2.1 Spin-orbit interaction

In addition to their orbital angular momentum, with clear analogies in classical mechanics, spin is an intrinsic attribute of quantum particles with no such simple correspondences. In the absence of spin-dependent terms in the Hamiltonian, spin is included by counting all states as doubly degenerate. In the relativistic picture, however, the orbital angular momentum of an electron interacts with its spin angular momentum to give a correction term $\hat{H}_{so}$ to the non-relativistic Hamiltonian[17]

$$\hat{H}_{so} = \frac{\hbar}{4m^2c^2}\hat{\vec{\sigma}} \cdot \nabla V(\vec{r}) \times \hat{\vec{p}}. \tag{2.5}$$

where $\hat{\vec{\sigma}}$ is the dimensionless spin-operator $\hat{\vec{\sigma}} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z)$, $V(\vec{r})$ is the true electronic potential and c is the speed of light in vacuum, hinting at the relativistic origins of the spin-orbit term. To implement spin-orbit interaction in the EPM, matrix elements of $\hat{H}_{so}$ are needed. However, in the EPM the deep-well parts of the true potential near the atomic-nuclei were truncated to remove the core states from the problem, allowing a convenient description of the electronic states in a plane-wave basis. Since ∇V in (2.5) diverges near the nuclei, the strongest correction is expected to come from this region – hence, by using the EPM, we have effectively removed the means by which to include spin-orbit interaction from our formalism. To reintroduce the dependency of the core states, it is necessary to revisit the original OPW states $\varphi_{\mathrm{OPW}}(\vec{k})$ used in the derivation of the pseudopotential. A plane wave orthogonalized to the electronic core states $|c\rangle$, with spin-polarization s, can be written

$$\begin{aligned} |\varphi_{\mathrm{OPW}}(\vec{k}), s\rangle = |\varphi_{\mathrm{OPW}}(\vec{k})\rangle|s\rangle &= |\vec{k}\rangle|s\rangle - \sum_{c,s'} |c\rangle|s'\rangle\langle c|\vec{k}\rangle\langle s'|s\rangle \\ &= |\vec{k}\rangle|s\rangle - \sum_{c} |c\rangle|s\rangle\langle c|\vec{k}\rangle. \end{aligned} \tag{2.6}$$

$|s\rangle$ and $|\vec{k}\rangle$ denoting spin states and plane wave states of wave vector $\vec{k}$, respectively. Hence, the spin-orbit matrix element becomes

$$H^{so}_{\vec{k}s,\vec{k}'s'} = \langle \varphi_{\mathrm{OPW}}(\vec{k}), s|\hat{H}_{so}|\varphi_{\mathrm{OPW}}(\vec{k}'), s'\rangle,$$

and upon insertion of (2.6)

$$\begin{aligned} H^{so}_{\vec{k}-\vec{G}s,\vec{k}-\vec{G}'s'} = &\langle s|\langle \vec{k}-\vec{G}|\hat{H}_{so}|\vec{k}-\vec{G}'\rangle|s'\rangle + \sum_c \langle s|\langle \vec{k}-\vec{G}|\hat{H}_{so}|c\rangle|s'\rangle\langle c|\vec{k}-\vec{G}'\rangle \\ &+ \sum_c \langle \vec{k}-\vec{G}|c\rangle\langle s|\langle c|\hat{H}_{so}|\vec{k}-\vec{G}'\rangle + \sum_{c,c'} \langle s|\langle \vec{k}-\vec{G}|c\rangle\langle c|\hat{H}_{so}|c'\rangle\langle c'|\vec{k}-\vec{G}'\rangle|s'\rangle. \end{aligned}$$

Following Balanski[17], the dominant correction is expected to originate from the core-region, why the double sum is taken as the main contribution. Hence, keeping only these terms and separating spatial and spin parts of $\hat{H}_{so}$

$$\begin{aligned} \hat{H}_{so} &= \hat{\vec{\Lambda}}(\vec{r}) \cdot \hat{\vec{\sigma}}, \\ \hat{\vec{\Lambda}}(\vec{r}) &\equiv \frac{\hbar}{4m^2c^2}\nabla V(\vec{r}) \times \hat{\vec{p}}, \end{aligned}$$

we get

$$H^{so}_{\vec{k}-\vec{G}s,\vec{k}-\vec{G}'s'} \approx \sum_{c,c'} \langle s|\hat{\vec{\sigma}}|s'\rangle \cdot \langle \vec{k}-\vec{G}|c\rangle \langle c|\hat{\vec{\Lambda}}(\vec{r})|c'\rangle \langle c'|\vec{k}-\vec{G}'\rangle.$$

The potential gradients are only significant near the nuclei, why they can be expanded in an atomic basis

$$\nabla V(\vec{r}) = \sum_{i,j} \nabla v_j(\vec{r}-\vec{R}_i-\vec{\tau}_j),$$

$\vec{R}_i$ and $\vec{\tau}_j$ being lattice and basis vectors, respectively. Similarly for the core states

$$|c\rangle = \sum_{i,j} |ij\alpha\rangle,$$

where $|ij\alpha\rangle$ denote a state of type α located on the j'th atom of the i'th unit cell, hence

$$\begin{aligned} H^{so}_{\vec{k}-\vec{G}s,\vec{k}-\vec{G}'s'} = \frac{\hbar}{4m^2c^2} \langle s|\hat{\vec{\sigma}}|s'\rangle \cdot \sum_{ij\alpha} \sum_{i'j'\beta} \sum_{i''j''} \langle \vec{k}-\vec{G}|ij\alpha\rangle \langle ij\alpha|\nabla v_j(\vec{r}-\vec{R}_{i''}-\vec{\tau}_{j''}) \\ \times \hat{\vec{p}}|i'j'\beta\rangle \langle i'j'\beta|\vec{k}-\vec{G}'\rangle|s'\rangle. \end{aligned}$$

The localization of the atomic core states mean that their amplitude decrease rapidly with distance to the corresponding atomic positions. The same is true for the atomic potential gradients, why the main contributions come from terms where $ij = i'j' = i''j''$. In addition, $\hat{\vec{\Lambda}}_j$ shall denote $\hat{\vec{\Lambda}}$ with $\nabla V(\vec{r})$ substituted by $\nabla v_j(\vec{r})$

$$\begin{aligned} H^{so}_{\vec{k}-\vec{G}s,\vec{k}-\vec{G}'s'} &\approx \frac{\hbar}{4m^2c^2} \langle s|\hat{\vec{\sigma}}|s'\rangle \cdot \sum_{ij} \sum_{\alpha\beta} \langle \vec{k}-\vec{G}|ij\alpha\rangle \langle j\alpha|\nabla v_j(\vec{r}-\vec{R}_i-\vec{\tau}_j) \times \hat{\vec{p}}|j\beta\rangle \langle ij\beta|\vec{k}-\vec{G}'\rangle \\ &= \langle s|\hat{\vec{\sigma}}|s'\rangle \cdot \sum_{j} \sum_{\alpha\beta} \langle j\alpha|\hat{\vec{\Lambda}}_j|j\beta\rangle \sum_{i} \langle \vec{k}-\vec{G}|ij\alpha\rangle \langle ij\beta|\vec{k}-\vec{G}'\rangle. \end{aligned}$$

Considering the overlap integrals

$$\begin{aligned} S_j &= \sum_i \langle \vec{k}-\vec{G}|ij\alpha\rangle \langle ij\beta|\vec{k}-\vec{G}'\rangle \\ &= \frac{1}{\Omega^2} \sum_i \int e^{-i(\vec{k}-\vec{G})\cdot\vec{r}} \varphi_\alpha(\vec{r}-\vec{R}_i-\vec{\tau}_j) d^3r \int e^{i(\vec{k}-\vec{G}')\cdot\vec{r}} \varphi^*_\beta(\vec{r}-\vec{R}_i-\vec{\tau}_j) d^3r \\ &= \frac{1}{\Omega^2} \sum_i e^{i(\vec{k}-\vec{G})\cdot(\vec{R}_i+\vec{\tau}_j)} \int e^{-i(\vec{k}-\vec{G})\cdot(\vec{r}-\vec{R}_i-\vec{\tau}_j)} \varphi_\alpha(\vec{r}-\vec{R}_i-\vec{\tau}_j) d^3r \\ &\quad \times e^{-i(\vec{k}-\vec{G}')\cdot(\vec{R}_i+\vec{\tau}_j)} \int e^{i(\vec{k}-\vec{G}')\cdot(\vec{r}-\vec{R}_i-\vec{\tau}_j)} \varphi^*_\beta(\vec{r}-\vec{R}_i-\vec{\tau}_j) d^3r. \end{aligned}$$

The integrals are recognized as the Fourier transforms of the core wave functions

$$\tilde{\varphi}_\alpha(\vec{k}) = \int \varphi_\alpha(\vec{r}) e^{-i\vec{k}\cdot\vec{r}}\, d^3r,$$

allowing S_j to be rewritten

$$\begin{aligned} S_j &= \frac{1}{\Omega^2} \sum_i e^{i(\vec{G}-\vec{G}')\cdot(\vec{R}_i+\vec{\tau}_j)} \tilde{\varphi}_\alpha(\vec{k}-\vec{G}) \tilde{\varphi}^*_\beta(\vec{k}-\vec{G}') \\ &= \frac{N}{\Omega^2} e^{i(\vec{G}-\vec{G}')\cdot\vec{\tau}_j} \tilde{\varphi}_\alpha(\vec{k}-\vec{G}) \tilde{\varphi}^*_\beta(\vec{k}-\vec{G}'). \end{aligned}$$

Hence,

$$H^{so}_{\vec{k}-\vec{G}s,\vec{k}-\vec{G}'s'} = \sum_j \sum_{\alpha\beta} S_j \langle s|\hat{\vec{\sigma}}|s'\rangle \cdot \langle j\alpha|\hat{\vec{\Lambda}}_j|j\beta\rangle.$$

This derivation have so far been based on Balanski *et al.*[17], however in that source the continued derivation is skipped; this gap will be mended in following. First, we note that, upon writing out the sum over j for a zinc-blende type crystal (denoting the basis atoms A and B), $H^{so}_{\vec{k}-\vec{G}s,\vec{k}-\vec{G}'s'}$ can be written

$$\begin{aligned} H^{so}_{\vec{k}-\vec{G}s,\vec{k}-\vec{G}'s'} &= \langle s|\hat{\vec{\sigma}}|s'\rangle \cdot \sum_{\alpha,\beta} \left\{ S_A \langle A\alpha|\hat{\vec{\Lambda}}_A|A\beta\rangle + S_B \langle B\alpha|\hat{\vec{\Lambda}}_B|B\beta\rangle \right\} \\ &= \langle s|\hat{\vec{\sigma}}|s'\rangle \cdot \sum_{\alpha,\beta} \left\{ \vec{f}_A e^{i(\vec{G}-\vec{G}')\cdot\vec{\tau}} + \vec{f}_B e^{-i(\vec{G}-\vec{G}')\cdot\vec{\tau}} \right\}, \end{aligned}$$

where $\vec{\tau} = (1,1,1)\,a/8$, choosing the origin mid-way between the basis atoms so $\vec{\tau}_A = -\vec{\tau}$ and $\vec{\tau}_B = +\vec{\tau}$. The function $\vec{f}_A$ is defined as

$$\vec{f}_A \equiv \tilde{\varphi}^A_\alpha(\vec{k}-\vec{G})\tilde{\varphi}^{A*}_\beta(\vec{k}-\vec{G}')\langle A\alpha|\hat{\vec{\Lambda}}_A|A\beta\rangle,$$

and similarly for $\vec{f}_B$. If Euler's formula is used to rewrite the complex exponentials, we get

$$H^{so}_{\vec{k}-\vec{G}s,\vec{k}-\vec{G}'s'} = 2\langle s|\hat{\vec{\sigma}}|s'\rangle \cdot \sum_{\alpha,\beta} \left\{ \vec{f}^S S^S(\vec{G}-\vec{G}') + \vec{f}^A S^A(\vec{G}-\vec{G}') \right\},$$

where $S^S(\vec{k})$ and $S^A(\vec{k})$ are the symmetric and anti-symmetric structure factors

$$S^S(\vec{k}) = \cos(\vec{k}\cdot\vec{\tau}), \quad S^A(\vec{k}) = i\sin(\vec{k}\cdot\vec{\tau}),$$

whereas $\vec{f}^S$ and $\vec{f}^A$ are the corresponding spin-orbit factors

$$\vec{f}^S = \frac{1}{2}\left(\vec{f}_A + \vec{f}_B\right), \quad \vec{f}^A = \frac{1}{2}\left(\vec{f}_A - \vec{f}_B\right).$$

We now consider the matrix element $\langle A\alpha|\Lambda_A|A\beta\rangle$

$$\langle A\alpha|\Lambda_A|A\beta\rangle = \frac{\hbar}{4m^2c^2}\langle A\alpha|\nabla v_A(\vec{r}+\vec{\tau})\times\hat{\vec{p}}|A\beta\rangle. \tag{2.7}$$

$v(\vec{r})$ is approximately a centro-symmetric potential $v(\vec{r}) = v(r)$, hence, the core states can be written as a product between a radial wave function $R_{n,\ell}$ and a spherical harmonic $Y_{\ell,m}$

$$|\alpha\rangle = R_{n,\ell}(r)Y_{\ell,m}(\theta,\phi), \quad |\beta\rangle = R_{n',\ell'}(r)Y_{\ell',m'}(\theta,\phi).$$

Furthermore $\hat{\vec{\Lambda}}$ can be rewritten in terms of spherical coordinates

$$\hat{\vec{\Lambda}} = \frac{\hbar}{4m^2c^2}\nabla v(r)\times\hat{\vec{p}} = \frac{\hbar}{4m^2c^2r}\frac{\partial v(r)}{\partial r} r\hat{r}\times\hat{\vec{p}} = \frac{\hbar}{4m^2c^2r}\frac{\partial v(r)}{\partial r}\hat{\vec{l}} = \xi(r)\hat{\vec{l}},$$

$\hat{\vec{l}}$ being the orbital angular momentum operator. Hence, the matrix element becomes

$$\begin{aligned} \langle \alpha A|\Lambda|\beta A\rangle &= \int R_{n,\ell}Y_{\ell,m}\,\xi(r)\hat{\vec{l}}\,R_{n',\ell'}Y_{\ell',m'}d^3r \\ &= \int_0^\infty R_{n,\ell}\,\xi(r)\,R_{n',\ell'}r^2\,dr \int_0^{2\pi}\int_0^\pi Y_{\ell,m}\,\hat{\vec{l}}\,Y_{\ell',m'}\sin(\theta)\,d\theta d\phi. \end{aligned} \tag{2.8}$$

Only the outermost p-states ($\ell = 1$) are considered, as in Cohen & Chelikowsky (1976)[22]. Considering first the angular part, which for brevity will be written $\langle\alpha|\hat{\vec{\ell}}|\beta\rangle$, where the dimensionless orbital angular momentum operator $\hbar\hat{\vec{\ell}} = \hat{\vec{l}}$ have been introduced. The angular behavior of a p-state can be expressed in a basis of three spherical harmonics

$$Y_{1,-1} = \sqrt{\frac{3}{8\pi}}\sin(\theta)e^{-i\phi}, \quad Y_{1,0} = \sqrt{\frac{3}{4\pi}}\cos(\theta), \quad Y_{1,1} = -\sqrt{\frac{3}{8\pi}}\sin(\theta)e^{i\phi},$$

which can be assembled into linearly independent functions displaying the dumbbell like behavior of a p-state aligned along the cartesian axes

$$\begin{aligned} |p_x\rangle = |x\rangle &= \frac{1}{\sqrt{2}}(Y_{1,-1} - Y_{1,1}) = \sqrt{\frac{3}{4\pi}}\sin(\theta)\cos(\phi), \\ |y\rangle &= \frac{i}{\sqrt{2}}(Y_{1,-1} + Y_{1,1}) = \sqrt{\frac{3}{4\pi}}\sin(\theta)\cos(\phi), \\ |z\rangle &= Y_{1,0}, \end{aligned}$$

where a look at the functions clearly reveal their alignment along the x,y,z-axes. The angular momentum operators from basic quantum mechanics[23] are

$$\left.\begin{aligned} \ell_+ &= \ell_x + i\ell_y \\ \ell_- &= \ell_x - i\ell_y \end{aligned}\right\} \Rightarrow \begin{cases} \ell_x = (\ell_+ + \ell_-)/2 \\ \ell_y = (\ell_+ - \ell_-)/2i \end{cases},$$

whereas ℓ_z is the operator of the eigenvalue ℓ. ℓ_+ and ℓ_- are the angular momentum creation and annihilation operators, respectively. Considering the example

$$\hat{\vec{\ell}}|z\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} (Y_{1,1} + Y_{1,-1}) \\ (Y_{1,1} - Y_{1,-1})/i \\ 0 \end{bmatrix} = \begin{bmatrix} -i|y\rangle \\ i|x\rangle \\ 0 \end{bmatrix},$$

the following matrix elements are easily evaluated

$$\langle x|\hat{\vec{\ell}}|z\rangle = i\hat{y}, \quad \langle y|\hat{\vec{\ell}}|z\rangle = -i\hat{x}, \quad \langle z|\hat{\vec{\ell}}|z\rangle = 0,$$

using the orthornormality of the spherical harmonics. Written more compactly this becomes $\langle\alpha|\hat{\vec{\ell}}|z\rangle = -i\hat{\alpha}\times\hat{z}$. Or generally,

$$\langle\alpha|\hat{\vec{\ell}}|\beta\rangle = -i\hat{\alpha}\times\hat{\beta}, \quad \alpha,\beta \in \{x,y,z\}. \tag{2.9}$$

Thus, we know how to perform the angular part of the Λ-matrix element. For the moment we shall forget about the radial integral, and call it λ – hence

$$\langle\alpha|\Lambda|\beta\rangle = -i\lambda\ \hat{\alpha}\times\hat{\beta}.$$

Considering now the overlap between a plane wave and a core state $\tilde{\varphi}_\alpha(\vec{q})$, we note that upon shifting the spherical harmonics to cartesian coordinates $|\alpha\rangle$ can be written

$$|\alpha\rangle = \sqrt{\frac{3}{4\pi}}R_{n,\ell}(r)\frac{\alpha}{r} = f(r)\alpha, \quad \alpha \in \{x,y,z\}.$$

With this, the overlap becomes

$$\begin{aligned} \tilde{\varphi}_\alpha(\vec{q}) &= \int e^{-i\vec{q}\cdot\vec{r}}\alpha f(r)\ d^3r \\ &= i\frac{\partial}{\partial q_\alpha}\int e^{-i\vec{q}\cdot\vec{r}}f(r)\ d^3r \\ &= i\frac{\partial}{\partial q_\alpha}f(q) \end{aligned}$$

$f(q)$ being the Fourier transform of $f(r)$. Using the chain rule to substitute $\partial/\partial q_\alpha \to q_\alpha q^{-1} \times \partial/\partial q$ allows

$$\tilde{\varphi}_\alpha(\vec{q}) = i\frac{q_\alpha}{q}\frac{\partial}{\partial q}f(q) \equiv q_\alpha F(q).$$

Now, collecting the results, we get

$$\begin{aligned}\vec{f}_A &= \tilde{\varphi}^A_\alpha(\vec{k}-\vec{G})\langle A\alpha|\hat{\Lambda}_A|A\beta\rangle\tilde{\varphi}^A_\beta(\vec{k}-\vec{G}') \\ &= -i\lambda_A|F_A(\vec{k}-\vec{G})|^2(\vec{k}-\vec{G})_\alpha\ (\hat{\alpha}\times\hat{\beta})\ (\vec{k}-\vec{G}')_\beta \\ &= -i\lambda_A|F_A(\vec{k}-\vec{G})|^2\ (\vec{k}-\vec{G})\times(\vec{k}-\vec{G}').\end{aligned}$$

Hence,

$$H^{so}_{\vec{k}-\vec{G}s,\vec{k}-\vec{G}'s'} = -i\left\{S^S(\vec{G}-\vec{G}')\lambda^S + S^A(\vec{G}-\vec{G}')\lambda^A\right\}\ \vec{K}\times\vec{K}'\cdot\langle s|\hat{\bar{\sigma}}|s'\rangle,$$

with the definitions

$$\begin{aligned}\vec{K} &\equiv \vec{k}-\vec{G}, \\ \lambda^S &\equiv \lambda_A|F_A(\vec{k}-\vec{G})|^2 + \lambda_B|F_B(\vec{k}-\vec{G})|^2, \\ \lambda^A &\equiv \lambda_A|F_A(\vec{k}-\vec{G})|^2 - \lambda_B|F_B(\vec{k}-\vec{G})|^2.\end{aligned}$$

With this, the entire Hamilton matrix becomes

$$H_{ms,ns'} = K_{\vec{k}-\vec{G}_m}\delta_{m,n}\delta_{s,s'} + V_{\vec{G}_m-\vec{G}_n}\delta_{s,s'} + H^{so}_{\vec{k}-\vec{G}_m s,\vec{k}-\vec{G}_n s'},$$

$K_{\vec{k}-\vec{G}_m}$ denoting kinetic energy matrix elements and $V_{\vec{G}_m-\vec{G}_n}$ local pseudopotential matrix elements as in the previous section. Finally, it is noted that $\langle s|\hat{\bar{\sigma}}|s'\rangle$ becomes[23]

$$\begin{aligned}\langle\uparrow|\hat{\bar{\sigma}}|\uparrow\rangle &= \hat{z}, \\ \langle\uparrow|\hat{\bar{\sigma}}|\downarrow\rangle &= \hat{x}-i\hat{y}, \\ \langle\downarrow|\hat{\bar{\sigma}}|\uparrow\rangle &= \hat{x}+i\hat{y}, \\ \langle\downarrow|\hat{\bar{\sigma}}|\downarrow\rangle &= -\hat{z}.\end{aligned} \tag{2.10}$$

Thus the Hamilton matrix can be written

$$\mathbf{H} = \begin{bmatrix}\mathbf{V}+\mathbf{K}+\vec{A}\cdot z & \mathbf{V}+\vec{A}\cdot(\hat{x}-i\hat{y}) \\ \mathbf{V}+\vec{A}\cdot(\hat{x}+i\hat{y}) & \mathbf{V}+\mathbf{K}-\vec{A}\cdot z\end{bmatrix}.$$

To investigate the basic behavior of this system with, the case of no-spin (i.e. $\lambda^A = \lambda^B = 0$) is investigated; in this case, $\mathbf{H}$ becomes a block-diagonal matrix

$$\mathbf{H} = \begin{bmatrix}\mathbf{H}_0 & 0 \\ 0 & \mathbf{H}_0\end{bmatrix}.$$

with eigenvalues doubly degenerate those of $\mathbf{H}_0$ It is noted that the most important feature of spin-orbit interaction in linear optical spectra are the splitting of the E1 peak into peaks termed E1 and E1+Δ, where Δ is the splitting energy of the top valence bands along the ΓL-line. Spin-orbit interaction has a tendency to split the doubly degenerate bands, although only those allowed by the specific form of $(\vec{G}_m\times\vec{G}_n)\cdot\vec{z}$. In practice, the spin-orbit parameters λ^A and λ^B can be calculated, although they are always significantly adjusted by

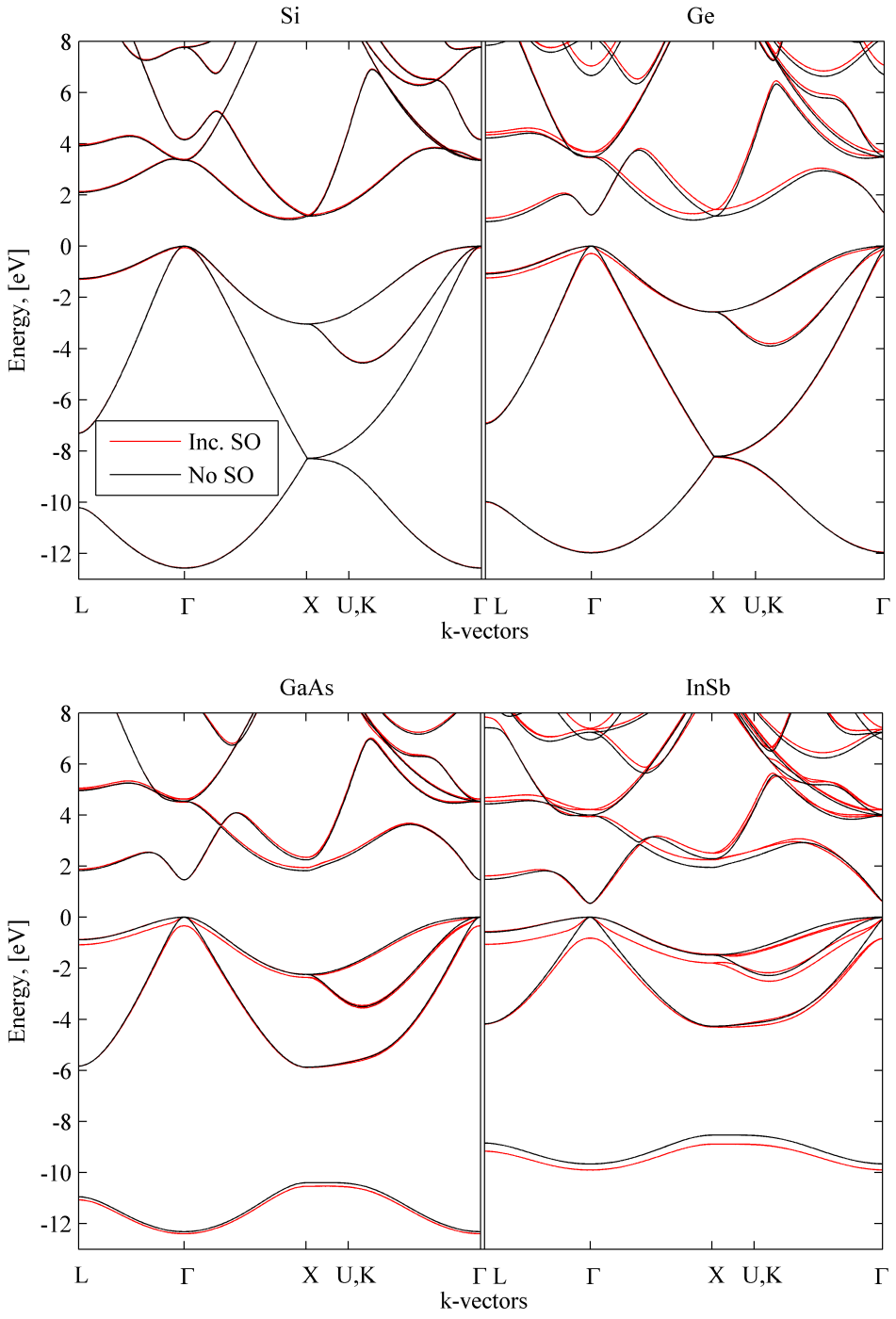


Figure 2.3: Band structures used in this project, with the red and black lines denoting the result including and neglecting spin-orbit interaction, respectively.

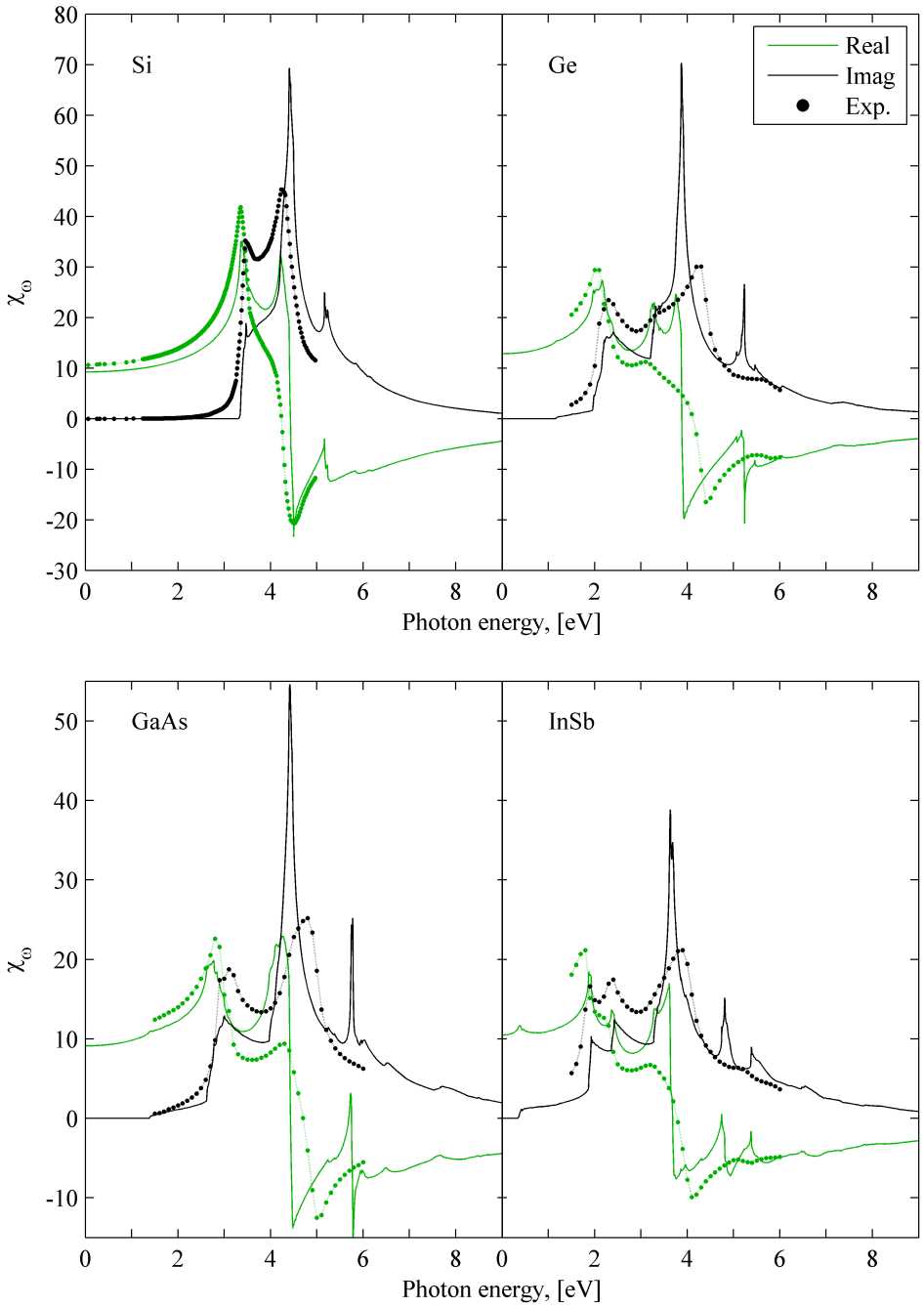


Figure 2.4: Linear spectra calculated from (2.1) including spin-orbit interaction. Experimental spectra are from Aspnes *et al.* (1983)[20], except for Silicon, which is from the handbook by Palik[21]. Form factors are Tab. 2.1 and spin-orbit parameters fitted to give the experimental spin-orbit splittings is given in Tab 2.2. All spectra are calculated using 98.304 tetrahedra in the irreducible zone, energy intervals of 1meV and a basis of 137 plane waves.

	Δ_{SO}, [eV]	$\lambda_s \times 10^{-3}$	λ_s/λ_a	F_{SO}
Si	0.055	2.592	NaN	1.00969
Ge	0.29	9.059	NaN	1.0470
GaAs	0.34	7.268	-20.556	1.0263
InSb	0.82	11.876	-7.363	1.0713

Table 2.2: Parameters used to include spin-orbit coupling

an empirically determined parameter[22]. Here we shall neglect the $\vec{k}$-dependency of these parameters, and determine them empirically. For centro-symmetric compounds Si and Ge, λ^A is identically zero giving only a single parameter λ^S which is fitted using a secant routine[24] to give the experimentally observed spin-orbit splitting factors[16] (i.e. the difference between the highest and second highest valence bands) at Γ denoted by Δ_{SO} in Tab 2.2. The usual form factors will have to be adjusted simultaneously in order to maintain the band gap; this is most easily done by simply multiplying all form factors by a common factor F_{SO}, which was seen to maintain the general form of the band structure, but change the direct band gap. The factors used are also given in Tab. 2.2.

For InSb and GaAs, the ratio λ^S/λ^A is taken to be the one obtained by evaluating the the functions at Γ. Gorsczyca *et al.* (1991)[25] present expressions from which λ^S/λ^A values can be calculated yielding for GaAs: -20.555, InSb: -7.3636. Optical spectra are shown in Fig. 2.4 calculated from (2.1) where c and v now run over twice as many states as before, since spin have been included explicitly in the potential, doubling the sice of the Hamilton matrix. Thus, expression should be divided by 2 before proceeding.

2.2 Spatial dispersion

In this section, we shall move beyond the dipole approximation and consider the optical wave vector $\vec{q}$ finite. To do this, the expression used for the optical susceptibility (2.1) must be generalized. This is done in the appendix, repeating A.32 here

$$\tilde{\chi}''_{\omega,\vec{q}} \equiv \frac{\hbar^2 e^2}{16\omega^2 \varepsilon_0 m^2 \pi^2} \sum_{c,v} \int \mathbf{M}_{c,v}(\vec{k},\vec{q}) \delta(E_{c\vec{k}+\vec{q},v\vec{k}} - \hbar\omega) d^3k, \tag{2.11}$$

where $\mathbf{M}_{c,v}(\vec{k},\vec{q})$ is a tensor defined by

$$\mathbf{M}_{c,v}(\vec{q},\vec{k}) = \langle v\vec{k}| \left[e^{i\vec{q}\cdot\vec{r}}\nabla + \nabla e^{i\vec{q}\cdot\vec{r}} \right] |c\vec{k}+\vec{q}\rangle \otimes \langle c\vec{k}+\vec{q}| \left[e^{-i\vec{q}\cdot\vec{r}}\nabla + \nabla e^{-i\vec{q}\cdot\vec{r}} \right] |v\vec{k}\rangle.$$

$\tilde{\chi}''_{\omega,\vec{q}}$ should be understood as the response in reciprocal space, i.e. as

$$\vec{\mathcal{P}}_{\omega,\vec{q}} = \varepsilon_0 \tilde{\chi}_{\omega,\vec{q}} \vec{\mathcal{E}}_{\omega,\vec{q}}, \tag{2.12}$$

where the polarization density $\vec{\mathcal{P}}_{\omega,\vec{q}}$ and the electric field $\vec{\mathcal{E}}_{\omega,\vec{q}}$ are plane waves of wave vector $\vec{q}$ and frequency ω. To evaluate (2.11), it is realized that energy band differences and matrix elements between states displaced by $\vec{q}$ in k-space is needed. To evaluate the integral over these by the LATM, the Schrödinger equation is simply solved for each tetrahedron vertex in addition to a matching point displaced by $\vec{q}$ relative to the vertex. As is apparent from the results shown in Fig. 2.5, including the optical wave vector may lead to photon facilitated absorption (i.e. non-vanishing imaginary part of the susceptibility) below the direct band

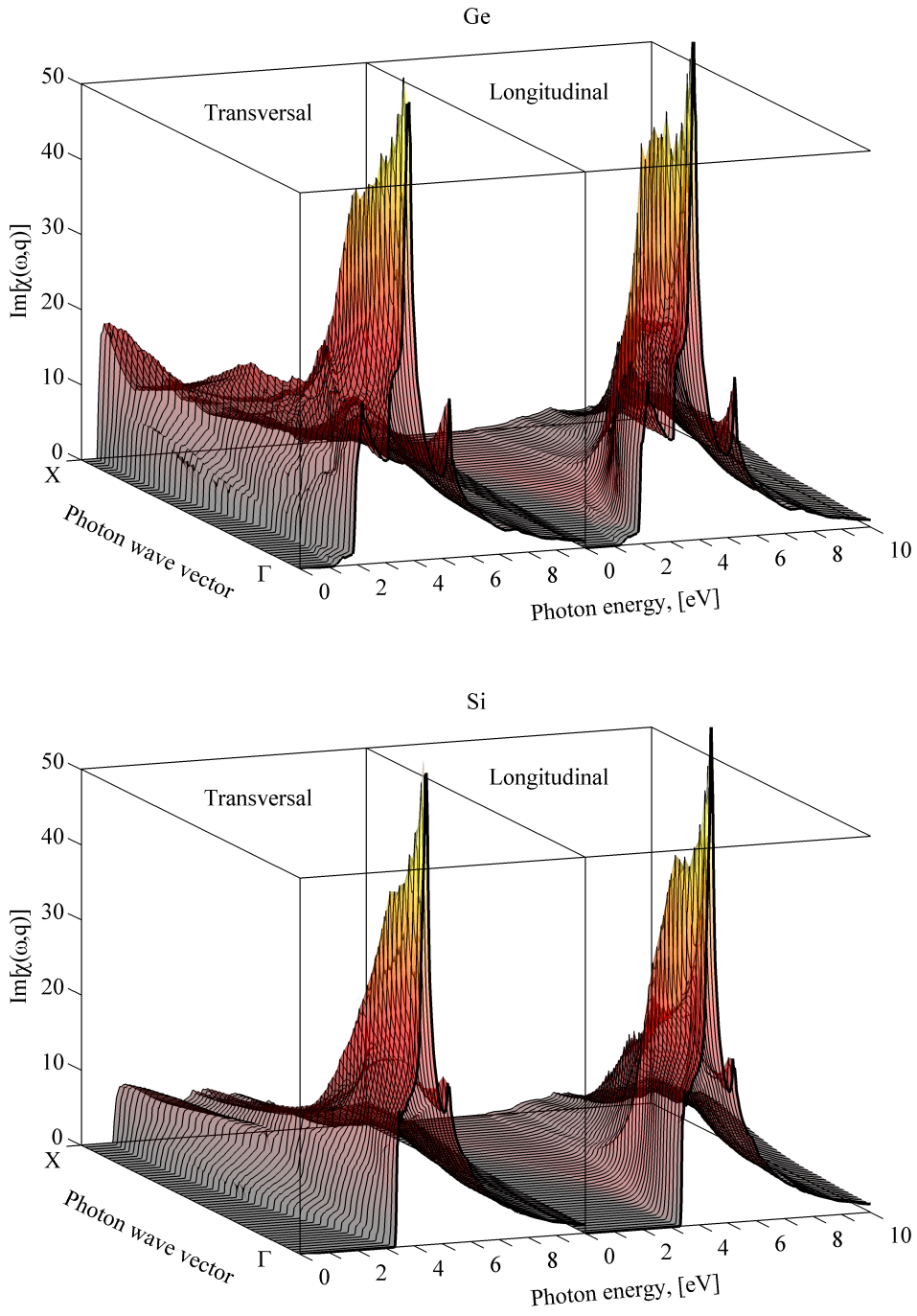


Figure 2.5: Spectra calculated using (2.11) from EPM band structures, neglecting spin-orbit coupling, using a basis of 137 plane waves and form factors from Tab. 2.1.

gap in indirect band gap materials, such as Si and Ge. Indirect transitions are, of cause, always present due to phonon-assisted transitions, giving rise to the weak absorption below the direct band gap. Photons disperse according to the well known relation $\omega/c = q$, making the q-vectors roughly 10^{-3} the size of the Brillouin zone, and thereby the q-axis of Fig. 2.5. Surface plasmons bound to metal/dielectric nanostructures provide a route of breaking with the photonic dispersion relation, allowing electric fields of extreme spatial variation bound to metal-dielectric interfaces. In the theory at hand, this concept translates into Fourier transforming the spatial variation of the plasmonic near-fields, yielding a q-spectrum of the electric field (which would be a δ-function for a plane wave). As an example, the following case is considered; a surface plasmon propagating in the x-direction, with its electric field decaying from a metallic surface at $z = 0$ into a silicon medium towards positive z. Thus the field can be written

$$\vec{\mathcal{E}}_\omega = \vec{\mathcal{E}}_0(\omega) e^{-sz+i\beta x}, \quad z > 0, \tag{2.13}$$

where it is noted that s is complex $s = s' + is''$, all primed s's real. The spatial variation of the electric field for $z > 0$ is Fourier transformed, yielding

$$\vec{\mathcal{E}}_{\omega,\vec{q}} = \frac{\vec{\mathcal{E}}_0(\omega)}{2\pi} \frac{1}{s + iq_z} \delta(q_x - \beta)\delta(q_y).$$

The decay length of a surface plasmon $\delta_z = 1/s'$ can be shown to be on the nanometer scale for thin metallic films embedded in a high-index dielectric, e.g. Si. With $s \sim 10^9 m^{-1}$, $\mathcal{E}_{\omega,\vec{q}}$ remains finite for extremely large wave vectors $k_z \sim 10^{10} m^{-1}$, significant compared to the scale of the Brillouin zone typically measured in units of $2\pi/a \sim 10^{10} m^{-1}$. A more thorough analysis of a surface plasmon mode bound to Ag film of thickness h embedded in Si can be made by application of the model by Burke *et al.*[26]; here the electric field amplitude of a bound mode is considered to decay exponentially away from the two Ag/Si interfaces at $z = 0$ and $z = h$. Solving the boundary value problems for the electric field and using one of the boundary conditions for the $\mathcal{H}$-field allow an expression for the electric field in s_1,s_3, the decay coefficients into Si, and s_2, the decay coefficient into Ag. Rewriting the exponential expression using hyperbolic identities allows a surface plasmon field, propagating in the x-direction by the propagating constant β, to be written on the form

$$\vec{\mathcal{E}}_\omega(\vec{r}) = \mathcal{E}_0 \left[\mathcal{E}_x \hat{x} + \mathcal{E}_z \hat{z}\right] e^{i\beta x},$$

with

$$\mathcal{E}_x = \frac{i}{\omega\varepsilon\varepsilon_0} \begin{cases} s_1 e^{s_1 z}, & z < 0 \\ s_2 \left(\sinh s_2 z + \dfrac{s_1 \varepsilon_m}{s_2 \varepsilon} \cosh s_2 z \right), & h > z > 0 \\ -s_3 \left(s_2 \cosh s_2 h + \dfrac{s_1 \varepsilon_m}{s_2 \varepsilon} \sinh s_2 h \right) s^{-s_3 z}, & z > 0 \end{cases}, \tag{2.14}$$

$$\mathcal{E}_z = -\frac{\beta}{\omega\varepsilon\varepsilon_0} \begin{cases} e^{s_1 z}, & z < 0 \\ \cosh s_2 h + \dfrac{s_1 \varepsilon_m}{s_2 \varepsilon} \sinh s_2 z, & h > z > 0 \\ \left(s_2 \cosh s_2 h + \dfrac{s_1 \varepsilon_m}{s_2 \varepsilon} \sinh s_2 h \right) s^{-s_3 z}, & z > 0 \end{cases}, \tag{2.15}$$

The decay coefficients couple for thin films, giving rise to symmetric and antisymmetric modes, and are found by solving the dispersion relation

$$\left[s_2^2 \varepsilon^2 + s_1 s_3 \varepsilon_m^2\right] \tanh s_2 h + \left[s_2 s_3 + s_1 s_2\right] \varepsilon\varepsilon_m = 0$$

which arises from invoking the last boundary condition for the $\mathcal{H}$-field. For the fields to satisfy the wave equation, the decay coefficients and the propagation constant must satisfy

$$\begin{aligned} s_1^2 &= \beta^2 - \varepsilon\frac{\omega^2}{c^2} \\ s_2^2 &= \beta^2 - \varepsilon_m\frac{\omega^2}{c^2} \\ s_3^2 &= \beta^2 - \varepsilon\frac{\omega^2}{c^2}, \end{aligned} \tag{2.16}$$

where the positive solutions are chosen for bound modes. A Drude model is applied for Ag $\varepsilon_m = \varepsilon_\infty - \omega_p^2/(\omega(\omega + i\Gamma))$, and $\varepsilon = 12$ appropriate for Si at longer wavelengths, where the dielectric constant varies slowly. Thus, solving the dispersion relation and introducing the skin depth of the field into Si $\delta_z^{(Si)} = s^{-1}$, the skin depth into Ag $\delta_z^{(Ag)} = s_2^{-1}$ and the propagation length $\delta_x = (2\beta)^{-1}$ allow Fig. 2.6, where the dashed line represent the antisymmetric solutions (refering to the parity of the $\mathcal{H}$-field with respect to $z = h/2$) to the dispersion relation, whereas the full lines represent symmetric solutions. As was mentioned before, some modes display decay coefficients on the nanometer scale. Since this is an eigenmode analysis, the fields cannot be related to an incoming field to investigate how effects of spatial dispersion might be probed. Instead, the absorption in Si is considered relative to the absorption in Ag film due to ohmic losses. The time-averaged absorbed power due to time-harmonic field can be expressed as

$$P_{abs}(\omega) = \frac{\omega}{2}\mathrm{Im}\left[\int \mathcal{P}_\omega(\vec{r}) \cdot \vec{\mathcal{E}}_\omega(\vec{r})^*\, d^3r\right], \tag{2.17}$$

and rewritten to Fourier space by Fourier transforming the spatial dependence of the electric field and polarization density

$$P_{abs}(\omega) = \pi\omega\mathrm{Im}\left[\int \mathcal{P}_{\omega,\vec{q}} \cdot \vec{\mathcal{E}}^*_{\omega,\vec{q}}\, d^3q\right].$$

Applying (2.12)

$$P_{abs}(\omega) = \pi\varepsilon_0\omega\mathrm{Im}\left[\int \tilde{\chi}_{\omega,\vec{q}}\vec{\mathcal{E}}_{\omega,\vec{q}} \cdot \vec{\mathcal{E}}^*_{\omega,\vec{q}}\, d^3q\right]$$

and writing out the components of the tensor product, neglecting off-diagonal elements (which all turn out to be zero) for z directed along the cubic [100] axis of silicon

$$P_{abs}(\omega) = \pi\varepsilon_0\omega\mathrm{Im}\left[\int |\mathcal{E}^x_{\omega,\vec{q}}|^2\tilde{\chi}^{xx}_{\omega,\vec{q}} + |\mathcal{E}^z_{\omega,\vec{q}}|^2\tilde{\chi}^{zz}_{\omega,\vec{q}}\, d^3q\right] \tag{2.18}$$

Now, considering that the electric fields decay exponentially in Si on the from given by (2.14) and (2.15) and that the Fourier transform of an exponential field is given by (2.13),the remaining task is purely algebraic, why we skip directly to the result

$$P^{(Si)}_{abs}(\omega) = \mathcal{P}_0\Omega^{1/3}\int_0^\infty \frac{|e_x|^2\chi^{xx\prime\prime}_{\omega,\vec{q}} + |e_z|^2\chi^{zz\prime\prime}_{\omega,\vec{q}}}{|s + iq_z|^2}dq_z \tag{2.19}$$

with the definitions

$$\mathcal{P}_0 \equiv \frac{\omega|\mathcal{E}_0|^2\varepsilon_0}{2\pi}, \quad e_x \equiv \frac{is_1}{\omega\varepsilon\varepsilon_0}, \quad e_z \equiv \frac{\beta}{\omega\varepsilon_0\varepsilon},$$

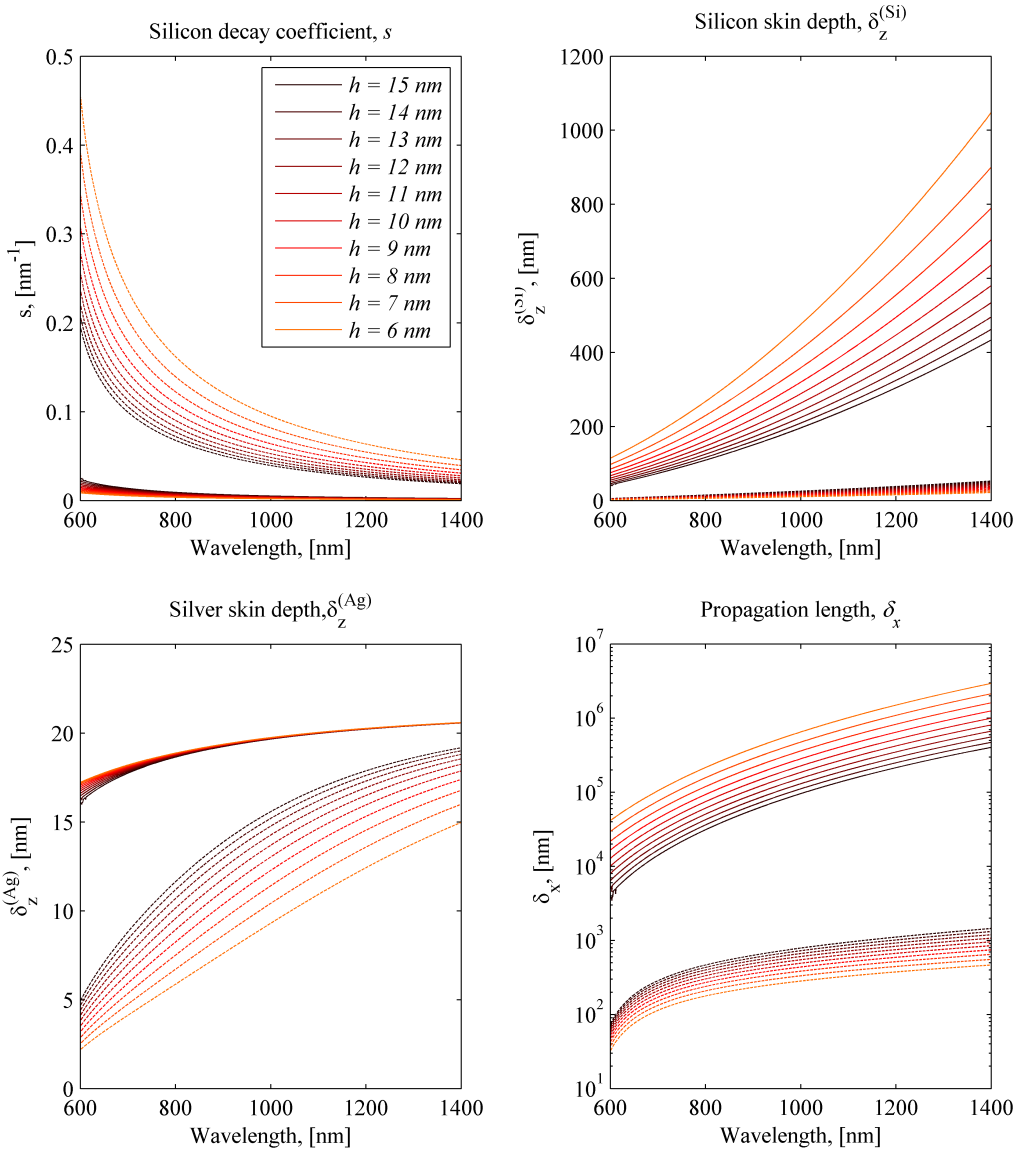


Figure 2.6: Characteristics of a Ag thin-film embedded in Si using $\hbar\Gamma = 0.0212$eV, $\hbar\omega_p = 9.1476$eV and $\varepsilon_\infty = 3.5616$eV found by fitting to the experimental results by Christy & Johnson[27] in the free-electron regime below 2.5eV.

where $\Omega^{1/3}$ represents the width of the sample following form ´the fact that the plane wave is not restricted in the y-direction. The integration over x is taken on the small, yet still macroscopic, interval dx. The absorption due to longitudinal and transversal fields are seperated:

$$P_L^{(Si)} = \mathcal{P}_0\Omega^{1/3}\int_0^\infty \frac{|e_z|^2\chi^{zz\prime\prime}_{\omega,\vec{q}}}{|s+iq_z|^2}dq_z, \quad P_T^{(Si)} = \mathcal{P}_0\Omega^{1/3}\int_0^\infty \frac{|e_x|^2\chi^{xx\prime\prime}_{\omega,\vec{q}}}{|s+iq_z|^2}dq_z.$$

Noting that the longitudinal response plotted in Fig. 2.5 becomes several orders of magnitude smaller than the transversal response for the energy range below the direct band gap under investigation here, it seems safe to ignore the longitudinal response provided $|e_z| \leq |e_x|$. Thus, investigating the ratio $|e_z|/|e_x|$:

$$\frac{|e_z|}{|e_x|} = \frac{|\beta|}{|s_1|} = \frac{|\sqrt{s_1^2+\varepsilon k_0^2}|}{|s_1|}.$$

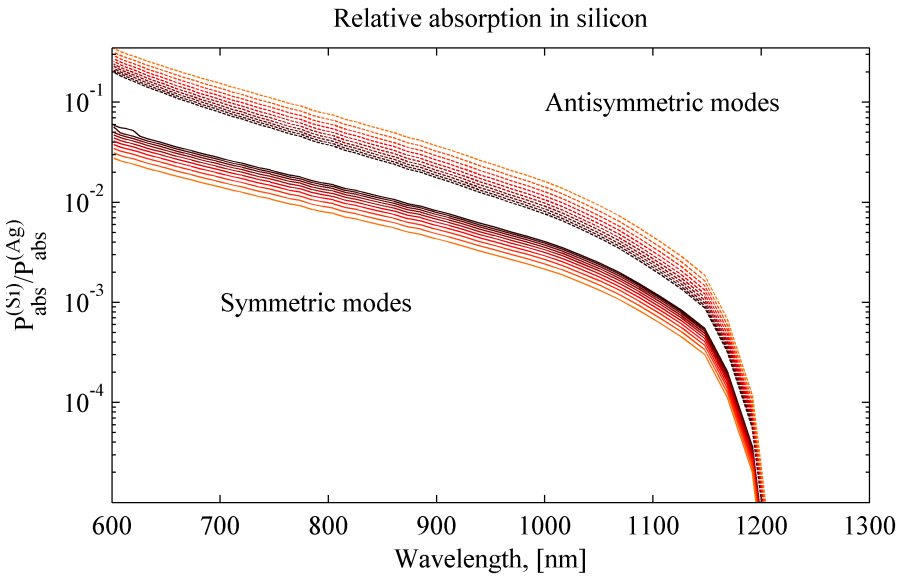


Figure 2.7: Absorption in Si relative to absorption in Ag. Colors denote film thickness as in Fig. 2.6(a).

For photon energies below the direct band gap $k_0 \sim 10^7 \mathrm{m}^{-1}$, whereas s_1 span the region from $\sim 10^9 \mathrm{m}^{-1}$ for the antisymmetric modes to $\sim 10^7 \mathrm{m}^{-1}$ for the symmetric modes. Hence, in all cases $|e_x|$ is on the same order of magnitude or larger than $|e_z|$, why the transversal absorption must dominate, i.e.

$$P_{abs}^{(Si)}(\omega) \approx P_T^{(Si)}(\omega).$$

This is the opposite tendency to what is observed in the analogous case for plasmons bound to small metal nanoparticles embedded in Si[28]. The integral in $P_L^{(Si)}$ is evaluated numerically by a simplex method from the spectra of Fig. (2.5). Calculating the absorption in the Ag film, neglecting effects of including spatial dispersion, can similarly be accomplished by the field profiles given by (2.14) and (2.15) and the absorption integral (2.17) in real space. However, this becomes a rather tedious exercise why the result will simply be stated[§]

$$P_{abs}^{(Ag)} = \frac{|\mathcal{E}_0|^2 \Omega^{1/3} dx}{8\varepsilon_0 \omega} \varepsilon_m'' \left[\left| \frac{s_2}{\varepsilon_m} \right|^2 I_x + \left| \frac{\beta}{\varepsilon_m} \right|^2 I_z \right] \tag{2.20}$$

with

$$\begin{aligned} I_x \equiv & \Big[\frac{1}{s_2'} \sinh(2s_2'h)(1+|a|^2) + \frac{1}{s_2''} \sin(2s_2''h)(|a|^2-1) \\ & + 2\Big(\frac{a'}{s_2'} (\cosh(2s_2'h)-1) - \frac{a''}{s_2''} (\cos(2s_2''h)-1) \Big) \Big], \\ I_z \equiv & \Big[\frac{1}{s_2'} \sinh(2s_2'h)\big(1+|a|^2\big) + \frac{1}{s_2''} \sin(2s_2''h)\big(1-|a|^2\big) \\ & + 2\Big(\frac{a'}{s_2'} (\cosh(2s_2'h)-1) + \frac{a''}{s_2''} (\cos(2s_2''h)-1) \Big) \Big]. \end{aligned} \tag{2.21}$$

The absorbed power in Si relative to the Ag film is shown in Fig. 2.7, where it is seen that decay coefficients on the nanometer scale of the antisymmetric modes does indeed lead

[§]For consistency, the full calculation is included in Appendix C

to significant absorption ($\sim$ 20% relative absorbed power in Si) below the direct bandgap relative to ohmic propagation losses in the Ag film. Recent work on the analogous case of metallic nanoparticles embedded in a Si medium showed that little incresed absorption below the band gap was found[28]. The difference between that case and here is, however, that while a surface plasmon bound to a metal nanoparticle radiate, a surface plasmon bound to a metallic waveguide propagates at low radiation losses. Thus, symmetric modes can, in ideal cases, propagate up to several hundred microns (hundreds of nanometers for the antisymmetric modes).

3 Non-Linear Optical Properties

In linear optics one investigates a response to an applied field of frequency ω oscillating at the same frequency. In non-linear optics, however, an applied monochromatic field can generate a polarization response oscillating at multiples of its frequency, driving a field of this multiple harmonic mode. Here we shall be concerned with second harmonic generation (SHG), which, simply put, is the investigation of the field generated by a current oscillating at a frequency 2ω due to an applied field of frequency ω. In the frequency domain, the polarization oscillating at 2ω ($\vec{\mathcal{P}}_{2\omega}$) is coupled to the field oscillating at ω ($\vec{\mathcal{E}}_{\omega}$) through the relationship

$$\vec{\mathcal{P}}_{2\omega} = \sum_{a,b,c} \varepsilon_0 \chi^{(2)}_{abc}(\omega) \mathcal{E}^b_{\omega} \mathcal{E}^c_{\omega}\, \hat{a},$$

where a,b and c annotate the Cartesian basis x,y,z. The imaginary part of the second harmonic response function can in semiconductors be expressed as[29] (see App. A.4.1)

$$\chi^{(2)\prime\prime}_{abc}(\omega) = \frac{e^3}{8\pi^2 m^3 \hbar^2 \varepsilon_0 \omega^3} \sum_{c,v,l} \int \Big[\frac{P_{vcl}}{\omega - \omega_{lv}} \delta\big(2\omega - \omega_{cv}\big) + \Big(\frac{P_{vlc}}{\omega + \omega_{cl}} + \frac{P_{clv}}{\omega + \omega_{lv}} \Big) \delta\big(\omega - \omega_{cv}\big) \Big] d^3k, \quad l \neq v,c \tag{3.1}$$

where $P_{cvl} = \mathrm{Im}\left[p^a_{cv} \big(p^b_{vl} p^c_{lc} + p^c_{vl} p^b_{lc} \big) \right]$. As in the linear case, the real part of the response is found by Kramers-Kronig transformation. To evaluate (3.1), the energy eigenvalues are determined by the EPM as in last section whereas the Brillouin zone integral warrants a more elaborate discussion due to the seemingly singular fractions of (3.1).

3.1 Brillouin zone integrals in non-linear optics

To solve (3.1), we need to perform integrals of the kind

$$I(\omega) = \int \frac{P\, \delta\big(\alpha\omega - \omega_{cv}\big)}{\omega_{cv} - \tilde{\omega}}\, d^3k, \quad \alpha \in \{1,2\}$$

where P and $\tilde{\omega}$ represent the appropriate matrix elements and band energy differences, respectively. Proceeding as in the LATM, we divide the integral into contributions from a set of N tetrahedra, spanning the irreducible zone to give

$$I(\omega) = \sum_{n=1}^{N} \int_{V_n} \frac{P\, \delta\big(\alpha\omega - \omega_{cv}\big)}{\omega_{cv} - \tilde{\omega}}\, d^3k \equiv \sum_{n=1}^{N} I_n(\omega) \tag{3.2}$$

where n indicates a specific tetrahedron and V_n annotates the volume of the n'th tetrahedron. In the standard LATM[15], the energy eigenvalues and matrix elements are linearized based on their values at the tetrahedron vertices. If the same approach is applied here, the argument of the δ-function is linearized

$$\omega_{cv}(\vec{k}) = \nabla_k \omega_{cv}(\vec{k}) \cdot (\vec{k} - \vec{k}_1) + \omega_{cv}(\vec{k}_1), \tag{3.3}$$

as is the factor

$$h(\omega,\vec{k}) = \nabla_k h(\omega,\vec{k})\cdot(\vec{k}-\vec{k}_1) + h(\omega,\vec{k}_1),$$

where

$$h(\omega,\vec{k}) \equiv \frac{P}{\omega_{cv}-\tilde{\omega}}.$$

However, since $h(\omega,\vec{k})$ may be singular within a given tetrahedron, the tetrahedral subdivision will have to be sufficiently fine-meshed to allow the linear interpolation to capture this behavior. In addition, even though h is often singular, the δ-function effectively removes many of these singularities from the problem – a behavior which is also demanding on the mesh-quality. To avoid this trouble retaining the advantages of the linear interpolation scheme, the band energy difference $\tilde{\omega}(\vec{k})$in $h(\omega,\vec{k})$ will have to be interpolated instead of h itself. Using the latter method, the integration over the tetrahedral domain can either be performed analytically[30;31], or by a numerical Monte Carlo integration[12;13]. The latter of these amounts to a linearized sampling technique, with the artifacts expected for such integration procedures, whereas the former is a fully analytic method giving much smoother results. All three methods was implemented and evaluated during this project. Common for the methods presented next is the desire to solve the integral

$$I_n(\omega) = \int_{V_n} \frac{P\,\delta(\alpha\omega-\omega_{cv})}{\omega_{cv}/\alpha-\tilde{\omega}}. \tag{3.4}$$

with P and $\hbar\tilde{\omega}$ again denoting appropriate matrix elements and energy band differences, respectively.

3.1.1 The linearized Monte-Carlo integration method

Random sampling integration applied by e.g. Moss *et al.*[12], or Monte-Carlo integration, is based on the rule that the integration of some function $f(\vec{k})$ over a specific domain V_n can be approximated by

$$I_n(\omega) = \int_{V_n} f(\vec{k})\,d^3k = V_n\langle f\rangle,$$

where $\langle f\rangle$ is the statistical mean value of a large amount of random samples of $f(\vec{k})$, $\vec{k}$ inside V_n[24]. In the present case, $f(\vec{k})$ is the integrand of (3.4). All k-dependent quantities, i.e. P, ω_{cv} and $\tilde{\omega}$, are linearized as in (3.3) based on their values at the tetrahedron vertices. A set of random points in V_n is then generated followed by evaluation of $f(\vec{k})$ using the linearized functions at each of the sample points. Finding $\langle f\rangle$ is then simple, although the singular nature of the δ-function requires the substitution of this with a broadened function, e.g. a Gaussian or a triangular approximation. Similarly, the fraction becomes singular at points where $\omega_{cv}=\tilde{\omega}$; to allow numerical evaluation, a phenomenological dampening factor is introduced, corresponding to Lorentzian broadening. Hence, appropriate substitutions are[29]

$$\delta(x) \to \begin{cases} \dfrac{2}{d}\left(1-\left|\dfrac{x}{2d}\right|\right), & \left|\dfrac{x}{2d}\right| \le 1 \\ 0 & \left|\dfrac{x}{2d}\right| > 1 \end{cases}$$

where d is half the width of the energy bins used in the summation. As mentioned, a Gaussian could also have been used, however no significant change of the convergence rate was found. A square approximation to the δ-function, i.e.

$$\delta(x) \to \left\{ \begin{array}{ll} \dfrac{1}{2a}, & |x| \le a \\ 0, & |x| > a \end{array} \right\}$$

was found to give a very slow convergence, with many instabilities. A Lorentzian broadening of the fraction amounts to the substitution

$$\frac{1}{\omega_{cv}-\tilde{\omega}} \rightarrow \frac{\omega_{cv}-\tilde{\omega}}{(\omega_{cv}-\tilde{\omega})^2+\gamma^2},$$

where γ should be chosen small enough not to change the result with subsequent broadening of the spectrum. These substitutions must also be used when using the "standard" LATM.

This method have the advantage of being easy to implement, especially for complicated matrix elements and energy fractions, as for third harmonic generation[32], and much faster than a standard sampling technique where the eigenvalue problem is solved at each k-point. The method is, however, both slower and less precise than LATM-based routines.

3.1.2 The linear-analytic tetrahedron method in second harmonic calculations

The following generalization of the LATM to nonlinear calculations is based on the work of Moss *et al.* (1987)[30]. The integral (3.4) is rewritten according to

$$\int \ldots d^3k = \int\int_{S_n} \frac{\ldots}{|\nabla_k \omega_{cv}|}\, dS\, d\omega_{cv}$$

where S_n is a surface of constant energy $\hbar\omega_{cv}$ contained within the n'th tetrahedron. If $\omega_{cv}-\tilde{\omega}$ were slowly varying and non-vanishing, the integration over the surface S_n could be approximated by simple multiplication by its area A_n. A_n would then be parametrized in ω_{cv}, followed by the trivial (due the the δ-function) ω_{cv}-integration; this is done in App. B for the linear response function.
Here, however, the possibility of $\omega_{cv} \approx \tilde{\omega}$ makes the fraction a rapidly varying function. If we similarly to S_n define a plane of constant $\tilde{\omega}$ inside the tetrahedron, the intersection of this plane with S_n will form a line L_n where both ω_{cv} and $\tilde{\omega}$ are constant; with this in mind, the integral is once again rewritten

$$\int_{S_n} \ldots dS = \int\int_{L_n} \frac{\ldots}{|\nabla_{||}\tilde{\omega}|}\, dl\, d\tilde{\omega}$$

where $\nabla_{||}\tilde{\omega}$ is the projection of $\nabla_k\tilde{\omega}$ onto S_n. Hence, using the definitions $\vec{b} \equiv \nabla_k\omega_{cv}$ and $\vec{d} \equiv \nabla_{||}\tilde{\omega}$, I_n becomes

$$\begin{aligned} I_n &= \int\int_{\tilde{\omega}_{s'}}^{\tilde{\omega}_{u'}} \int_{L_n} \frac{\delta(\alpha\omega-\omega_{cv})}{\omega_{cv}/\alpha-\tilde{\omega}} \frac{1}{|\vec{b}||\vec{d}|}\, dl\, d\tilde{\omega}\, d\omega_{cv} \\ &= \int\int_{\tilde{\omega}_{s'}}^{\tilde{\omega}_{u'}} \frac{\delta(\alpha\omega-\omega_{cv})}{\omega_{cv}/\alpha-\tilde{\omega}} \frac{L_n}{|\vec{b}||\vec{d}|}\, d\tilde{\omega}\, d\omega_{cv}, \end{aligned} \tag{3.5}$$

where the constancy of ω_{cv} and $\tilde{\omega}$ on L_n was used. The $\tilde{\omega}$ integration limits is from the smallest value of $\tilde{\omega}$ on S_n ($\tilde{\omega}_{s'}$) to the highest ($\tilde{\omega}_{u'}$). Thus, the task at hand is to find the length L_n in terms of ω_{cv} and $\tilde{\omega}$ and to perform the $\tilde{\omega}$ integration. This is facilitated by linearizing ω_{cv} according to

$$\omega_{cv} = \vec{b}\cdot(\vec{k}-\vec{k}_1) + \omega_{cv}(\vec{k}_1), \tag{3.6}$$

which allows S_n to become a simple triangle or quadrilateral spanned by points on the tetrahedron edges of identical energy, as shown in Fig. 3.1(a). The tetrahedron vertices at

$\{\vec{k}_1,\vec{k}_2,\vec{k}_3,\vec{k}_4\}$ have energies $\hbar\omega_i = \hbar\omega_{cv}(\vec{k}_i)$ and are sorted according to $\omega_1 \leq \omega_2 \leq \omega_3 \leq \omega_4$. The shape of S_n varies in three distinct regions; in Region I, $\omega_1 \leq \omega_{cv} \leq \omega_2$, and S_n is a triangular plane spanned by points lying on the $\vec{k}_1\vec{k}_2$-, $\vec{k}_1\vec{k}_3$- and $\vec{k}_1\vec{k}_4$-edges. In Region II, $\omega_2 \leq \omega_{cv} \leq \omega_3$, S_n resembles a quadrilateral consisting of a triangle as defined for Region I with one of its vertices extending outside the tetrahedron (to $\tilde{k}_s$ in Fig. 3.1(a)) cut by the tetrahedron face. In Region III, $\omega_3 \leq \omega_{cv} \leq \omega_4$, S_n is a triangle spanned by points on the $\vec{k}_4\vec{k}_1$-, $\vec{k}_4\vec{k}_2$- and $\vec{k}_4\vec{k}_3$-lines. The points spanning S_n, $\{\vec{k}'_s,\vec{k}'_t,\vec{k}'_u\}$ are also sorted according to $\omega_{s'} \leq \omega_{t'} \leq \omega_{u'}$. In the following, Region I will be considered and the result for this region will be generalized to Region II and Region III. Hence, the set $\{s,t,u\}$ have the same elements as the set $\{2,3,4\}$ with the exact correspondence fixed by the ordering of both $\{\omega_{s'},\omega_{t'},\omega_{u'}\}$ and $\{\omega_2,\omega_3,\omega_4\}$. A simple linear interpolation of $\tilde{\omega}$ in ω_{cv} along the tetrahedron edges using $\tilde{\omega}_1$ as reference yields

$$\begin{aligned}
\tilde{\omega}_{s'} &= \tilde{\omega}_1 + (\omega_{cv}-\omega_1)\frac{\tilde{\omega}_{s1}}{\omega_{s1}},\\
\tilde{\omega}_{t'} &= \tilde{\omega}_1 + (\omega_{cv}-\omega_1)\frac{\tilde{\omega}_{t1}}{\omega_{t1}},\\
\tilde{\omega}_{u'} &= \tilde{\omega}_1 + (\omega_{cv}-\omega_1)\frac{\tilde{\omega}_{u1}}{\omega_{u1}},
\end{aligned} \tag{3.7}$$

where $\omega_{u1} = \omega_u - \omega_1$.
If $\tilde{\omega}_{s'} \leq \tilde{\omega} \leq \tilde{\omega}_{t'}$, L_n can be found by interpolation

$$L_n = \left|(\vec{k}'_u - \vec{k}'_s)\left(\frac{\tilde{\omega}-\tilde{\omega}_{s'}}{\tilde{\omega}_{u'}-\tilde{\omega}_{s'}}\right) + (\vec{k}'_t - \vec{k}'_s)\left(\frac{\tilde{\omega}-\tilde{\omega}_{s'}}{\tilde{\omega}_{t'}-\tilde{\omega}_{s'}}\right)\right|,$$

Extracting common factors of the two terms and using (3.7) allows L_n to be written more conveniently

$$L_n = \left((\tilde{\omega}-\tilde{\omega}_1) - (\omega_{cv}-\omega_1)\frac{\tilde{\omega}_{s1}}{\omega_{s1}}\right)\Gamma_{us,ts}, \tag{3.8}$$

where

$$\begin{aligned}
\Gamma_{us,ts} &\equiv \left|\frac{\vec{k}_u\omega_{s1}-\vec{k}_s\omega_{u1}}{\Delta_{su}} - \frac{\vec{k}_t\omega_{s1}-\vec{k}_s\omega_{t1}}{\Delta_{st}}\right|,\\
\Delta_{su} &\equiv \omega_{s1}\tilde{\omega}_{u1} - \tilde{\omega}_{s1}\omega_{u1},\\
\Delta_{st} &\equiv \omega_{s1}\tilde{\omega}_{t1} - \tilde{\omega}_{s1}\omega_{t1}.
\end{aligned}$$

For $\tilde{\omega}_{t'} \leq \tilde{\omega} \leq \tilde{\omega}_{u'}$

$$L_n = \left[(\tilde{\omega}-\tilde{\omega}_1) - (\omega_{cv}-\omega_1)\frac{\tilde{\omega}_{u1}}{\omega_{u1}}\right]\Gamma_{su,tu} \tag{3.9}$$

which follows from considering $\vec{k}'_u$ as reference instead of $\vec{k}'_s$; the difference of these two cases is illustrated in Fig. 3.1(b). $\Gamma_{us,ts}$ and $\Gamma_{su,tu}$ can also be expressed

$$\Gamma_{us,ts} = 6V_n\left|\frac{\omega_{s1}}{\Delta_{su}\Delta_{st}}\right||\vec{b}||\vec{d}|, \quad \Gamma_{su,tu} = 6V_n\left|\frac{\omega_{u1}}{\Delta_{us}\Delta_{ut}}\right||\vec{b}||\vec{d}|.$$

Using (3.8) and (3.9) in (3.5) allows

$$\begin{aligned}
I_n = \int\Bigg\{&\int_{\tilde{\omega}_{s'}}^{\tilde{\omega}_{t'}}\left(\frac{\tilde{\omega}-\tilde{\omega}_1}{\omega_{cv}/\alpha-\tilde{\omega}} - \frac{\omega_{cv}-\omega_1}{\omega_{cv}/\alpha-\tilde{\omega}}\frac{\tilde{\omega}_{s1}}{\omega_{s1}}\right)d\tilde{\omega}\,\Gamma_{us,ts}\\
&+\int_{\tilde{\omega}_{t'}}^{\tilde{\omega}_{u'}}\left(\frac{\tilde{\omega}-\tilde{\omega}_1}{\omega_{cv}/\alpha-\tilde{\omega}} - \frac{\omega_{cv}-\omega_1}{\omega_{cv}/\alpha-\tilde{\omega}}\frac{\tilde{\omega}_{u1}}{\omega_{u1}}\right)d\tilde{\omega}\,\Gamma_{su,tu}\Bigg\}\,\frac{\delta(\alpha\omega-\omega_{cv})}{|\vec{b}||\vec{d}|}\,d\omega_{cv}.
\end{aligned}$$

Considering the first integral

$$\begin{aligned}
&\int_{\tilde{\omega}_{s'}}^{\tilde{\omega}_{t'}} \left(\frac{\tilde{\omega}-\tilde{\omega}_1}{\omega_{cv}/\alpha-\tilde{\omega}} - \frac{\omega_{cv}-\omega_1}{\omega_{cv}/\alpha-\tilde{\omega}} \frac{\tilde{\omega}_{s1}}{\omega_{s1}} \right) d\tilde{\omega} \\
&\quad = -\Big[\tilde{\omega}\Big]_{\omega_{s'}}^{\omega_{t'}} + \Big[\ln(\tilde{\omega}-\omega_{cv}/\alpha)\Big]_{\omega_{s'}}^{\omega_{t'}} \left(\tilde{\omega}_1 - \omega_{cv}/\alpha + \frac{\tilde{\omega}_{s1}\omega_{cv}}{\omega_{s1}} - \frac{\tilde{\omega}_{s1}\omega_1}{\omega_{s1}} \right) \\
&\quad = \tilde{\omega}_{s'} - \tilde{\omega}_{t'} + \ln\left[\frac{\tilde{\omega}_{t'}-\omega_{cv}/\alpha}{\tilde{\omega}_{s'}-\omega_{cv}/\alpha} \right] \left\{ \tilde{\omega}_1 + \frac{\tilde{\omega}_{s1}}{\omega_{s1}}(\omega_{cv}-\omega_1) - \omega_{cv}/\alpha \right\} \\
&\quad = \ln\left[\frac{\tilde{\omega}_{t'}-\omega_{cv}/\alpha}{\tilde{\omega}_{s'}-\omega_{cv}/\alpha} \right] (\tilde{\omega}_{s'}-\omega_{cv}/\alpha) + \tilde{\omega}_{s't'}.
\end{aligned}$$

where the last equality follows from comparison with eqs. (3.7), it is clear that I_n becomes*

$$\begin{aligned}
I_n(\omega) = &\int \left[\left\{ \ln\left| \frac{\tilde{\omega}_{t'}-\omega_{cv}/\alpha}{\tilde{\omega}_{s'}-\omega_{cv}/\alpha} \right| (\tilde{\omega}_{s'}-\omega_{cv}/\alpha) + \tilde{\omega}_{s't'} \right\} \Gamma_{us,ts} \right. \\
&\left. + \left\{ \ln\left| \frac{\tilde{\omega}_{u'}-\omega_{cv}/\alpha}{\tilde{\omega}_{t'}-\omega_{cv}/\alpha} \right| (\tilde{\omega}_{u'}-\omega_{cv}/\alpha) + \tilde{\omega}_{t'u'} \right\} \Gamma_{su,tu} \right] \frac{\delta(\alpha\omega-\omega_{cv})}{|\vec{d}||\vec{b}|} \, d\omega_{cv} \\
= &\left\{ \ln\left| \frac{\tilde{\omega}_{t'}-\omega}{\tilde{\omega}_{s'}-\omega} \right| (\tilde{\omega}_{s'}-\omega) + \tilde{\omega}_{s't'} \right\} \frac{\Gamma_{us,ts}}{|\vec{b}||\vec{d}|} \\
&+ \left\{ \ln\left| \frac{\tilde{\omega}_{u'}-\omega}{\tilde{\omega}_{t'}-\omega} \right| (\tilde{\omega}_{u'}-\alpha\omega) + \tilde{\omega}_{t'u'} \right\} \frac{\Gamma_{su,tu}}{|\vec{b}||\vec{d}|}
\end{aligned} \tag{3.10}$$

It is recalled that we only considered ω in Region I, why I_n calculated so far shall be denoted $I_n^{(I)}$. From geometry, it is apparent that if the case of ω_{cv} in Region II was calculated using the same expressions as for Region I, S_n would be taken as being too large by an amount $S_n^{(II)\prime}$ (the light green area outside the tetrahedron in Fig. 3.1(a)), hence $I_n^{(II)} = I_n^{(I)} - I_n^{(II)\prime}$ where $I_n^{(II)'}$ is the integral over $S_n^{(II)'}$. In Region III we simply calculate by the same formula as in Region I, although $\vec{k}_4$ is used as reference point in the linear interpolation – hence these expressions follow by making the substitution $1 \leftrightarrow 4$.
To summarize, for ω in Region I: $\omega_1 \leq \omega \leq \omega_2$

$$\begin{aligned}
I_n(\omega) = &6V_n \left\{ \left(\ln\left| \frac{\tilde{\omega}_{t'}-\omega}{\tilde{\omega}_{s'}-\omega} \right| (\tilde{\omega}_{s'}-\omega) + \tilde{\omega}_{s't'} \right) \left| \frac{\omega_{s1}}{\Delta_{su}\Delta_{st}} \right| \right. \\
&\left. + \left[\ln\left| \frac{\tilde{\omega}_{u'}-\omega}{\tilde{\omega}_{t'}-\omega} \right| (\tilde{\omega}_{u'}-\omega) + \tilde{\omega}_{t'u'} \right] \left| \frac{\omega_{u1}}{\Delta_{us}\Delta_{ut}} \right| \right\}
\end{aligned} \tag{3.11}$$

and the set $\{\omega_{s1}, \omega_{t1}, \omega_{u1}\}$ corresponds to the set $\{\omega_{21}, \omega_{31}, \omega_{41}\}$. The exact ordering of this correspondence is found by first calculating $\{\tilde{\omega}_{i'}\}, i \in \{2,3,4\}$ which is then sorted. This set then corresponds to $\{\tilde{\omega}_{s'}, \tilde{\omega}_{t'}, \tilde{\omega}_{u'}\}$ so that $\tilde{\omega}_{s'} \leq \tilde{\omega}_{t'} \leq \tilde{\omega}_{u'}$. Next, $\{\omega_s, \omega_t, \omega_u\}$ is sorted by the ordering of the set $\{\tilde{\omega}_i\}$, making sure that $\vec{k}_s'$ lies on the $\vec{k}_1\vec{k}_s$-line, etc. This assortment allows calculation of all values necessary to find the contribution from Region I.
In Region II: $\omega_2 \leq \omega \leq \omega_3$,

$$I_n(\omega) = I^{(I)} - I^{(II)\prime}, \tag{3.12}$$

where $I^{(I)}$ is calculated in exactly the same way as for Region I. $I_n^{(II)\prime}$ is calculated from the same formula as $I_n^{(I)}$, except the set $\{\omega_{s1}, \omega_{t1}, \omega_{u1}\}$ corresponds to the set $\{\omega_{21}, \omega_{32}, \omega_{42}\}$

*Taking the absolute value of the argument in the log-function amounts to introducing a vanishing broadening when performing the integral over a singularity.

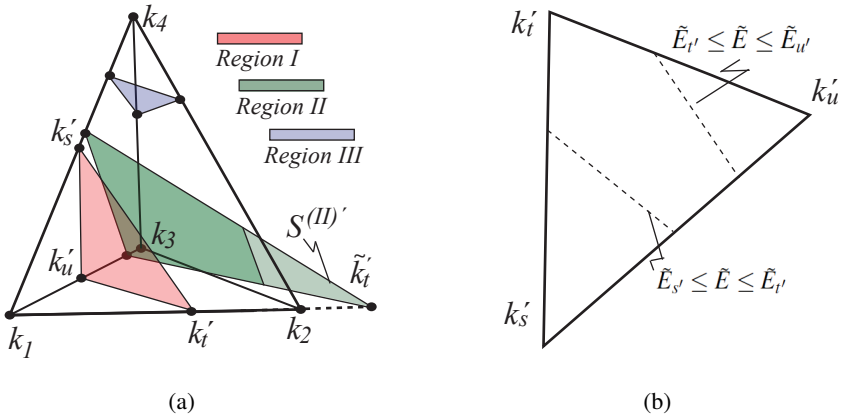


Figure 3.1: (a) Notation for a tetrahedron. For Region I, $\omega_s \sim \omega_4$, $\omega_t \sim \omega_2$, and $\omega_u \sim \omega_3$ where this correspondence was determined by sorting $\omega_1 \leq \omega_2 \leq \omega_3 \leq \omega_4$ and $\omega_{s'} \leq \omega_{t'} \leq \omega_{u'}$. $\omega_{s'}$, $\omega_{t'}$ and $\omega_{u'}$ are all identical, making the surfaces shown first order, isoenergy surfaces. (b) a plane of constant ω_{cv}, S_n, with lines of constant $\tilde{\omega}$, L_n (dashed lines).

using the same ordering procedure as in Region I.
For Region III: $\omega_3 \leq \omega \leq \omega_4$, the same is true, although here $\{\omega_{s1}, \omega_{t1}, \omega_{u1}\}$ corresponds to $\{\omega_{41}, \omega_{42}, \omega_{43}\}$.

To implement this method numerically, the energy interval under investigation is divided into a set of subintervals of constant energy, typically in the range of meV. For each of these subintervals, all N tetrahedra are checked to see if $\omega_1 \leq \omega \leq \omega_4$, if not, the given tetrahedron cannot contribute. If a tetrahedron contributes, it is checked which region ω belongs to followed by application of the analytic integration functions.

3.2 Second harmonic spectra of GaAs and InSb

Materials having an inversion center will in general not display bulk dipole second harmonic generation, which is calculated by (3.1). Using symmetry considerations, this is readily seen; if a system can be described by wave functions of definite parity, then all momentum matrix elements between two odd states and two even states will vanish, respectively, due to the momentum operator being odd. In systems of inversion symmetry, a basis of well defined parity can always be chosen. In addition, composite matrix elements of the kind $p_{\alpha\beta}p_{\beta\gamma}p_{\gamma\alpha}$ where the states $|\alpha\rangle, \ldots$ all represent states of well defined parity, must vanish since the above combinations must contain an even/even or an odd/odd momentum matrix element. Hence, of the semiconductors considered here only GaAs and InSb give any second harmonic contribution when calculated from (3.1); results are shown in Fig. 3.2, neglecting spin-orbit interaction and using the LATM generalized for second harmonic calculations presented above. Below the direct band gap, features similar to what is observed for the linear spectrum is seen at half their corresponding photon energies due to the 2ω-term. The E0 fundamental edge transitions at half band-gap energies display a characteristic $\sqrt{\omega}$ proportionality for GaAs weakly modulated by the singular factor. InSb does not, however, which is due to the transition being modulated strongly by the ω^{-3}-proportionality closer to zero. The E1 transition is apparent in both spectra, as is the E2 transition. At 2.91eV for GaAs and 2.48eV for InSb an extremely powerful an narrow peak

is found (the peak scales outside the figure, and reaches a maximum of 5.00 nm/V for GaAs and 6.51nm/V for InSb with a width of only a single data-point). This feature arises from transitions from the highest valence bands to the lowest and second lowest conduction bands (the latter is doubly degenerate) along the ΓL-line. Here, the joint conditions $2\hbar\omega = E_{cv}$ and $\hbar\omega = E_{lv}$ may be satisfied, making the 2ω-term of (3.1) resonant and singular. The ω-term is similarly resonant, making this a so called double-resonance. Fig. 3.2 is not scaled to show the full height of the peak, since its extremely small width will cause its amplitude to decrease rapidly by including any broadening effects neglected in the simple perturbation theory leading to (3.1). The effect of including 10meV Lorentzian broadening is shown in Fig. 3.3, where the sharp spectral features are indeed seen to vanish. Near 4.5eV for GaAs and 3.8eV for InSb the ω-term displays the E2 peak also observed in the linear spectrum.

Fig 3.4 displays a comparison between experimentally observed SHG spectra from various sources, and the theoretical calculations of this work both neglecting and including spin-orbit effects. No continuous experimental second harmonic spectrum exists over a large energy range, as was the case for the linear susceptibility, why comparison with experimental spectra over a large frequency range will have to be "stitched" together from various sources. This introduces a large number of possible errors - e.g. different substrate may affect the measurements as may differences in equipment, procedure, and experimental setup. This is readily apparent in the experimental results for GaAs, where Bethune *et al.* (1975)[11] and Bergfeld & Daum (2003)[33] disagree on the intensity and minimum positions of $|\chi^{(2)}_{xyz}|$ in the frequency overlap of their respective spectra. It is noted, however, that the spectrum of Bethune *et al.* is measured at 80K, whereas the spectrum of Bergfeld *et al.* is measured at room temperature. Thus, the temperature dependence of the band structure may explain the discrepancy in peak positions.

The agreement between experiment and theory is not terribly convincing, however several spectral features are reproduced in the theoretical spectra; for GaAs, all three peaks in the experimental spectral range are reproduced, although with variations in peak intensity and slight frequency offsets. The weak peak around 1.8eV in the experimental spectrum is believed to be a surface specific feature[33].

The general tendency of the InSb spectrum is also roughly reproduced, although from the small spectral range experimentally available, general agreement cannot be concluded. Including spin-orbit effects seems to improve the agreement between theory and experiment by e.g. removing the double resonance, as might be expected when considering the large impact of spin-orbit coupling on the linear spectra. The peak of the theoretical spectrum near 2.2eV seems to be reproduced experimentally, as do the double-resonance peak near 2.5eV, although considerably damped. The physical justification for the theoretical double resonance peak intensity is questionable, since it arises from a singularity derived using perturbation theory which is lifted by e.g. including spin-orbit effects.

In comparing experimental and theoretical spectra, it should be noted that the equation used to calculate the results for GaAs and InSb (3.1) was derived assuming an infinite, periodic medium. Any real system is not periodic, however, the best approximation to the case being semi-infinite media interfacing each other. At surfaces, the infinite-medium approximation breaks down, allowing e.g. Ge and Si to have a dipole-allowed second harmonic signal even though their bulk phases have an inversion center. These surface contributions are also present in the experimental spectra for GaAs and InSb, and since they are not included in the theoretical spectra disagreements between theory and experiment is expected.

Additionally, (3.1) was derived neglecting the the optical field variation in space, i.e. the optical wave vector $\vec{q}$ was considered much smaller than the width of the Brillouin zone.

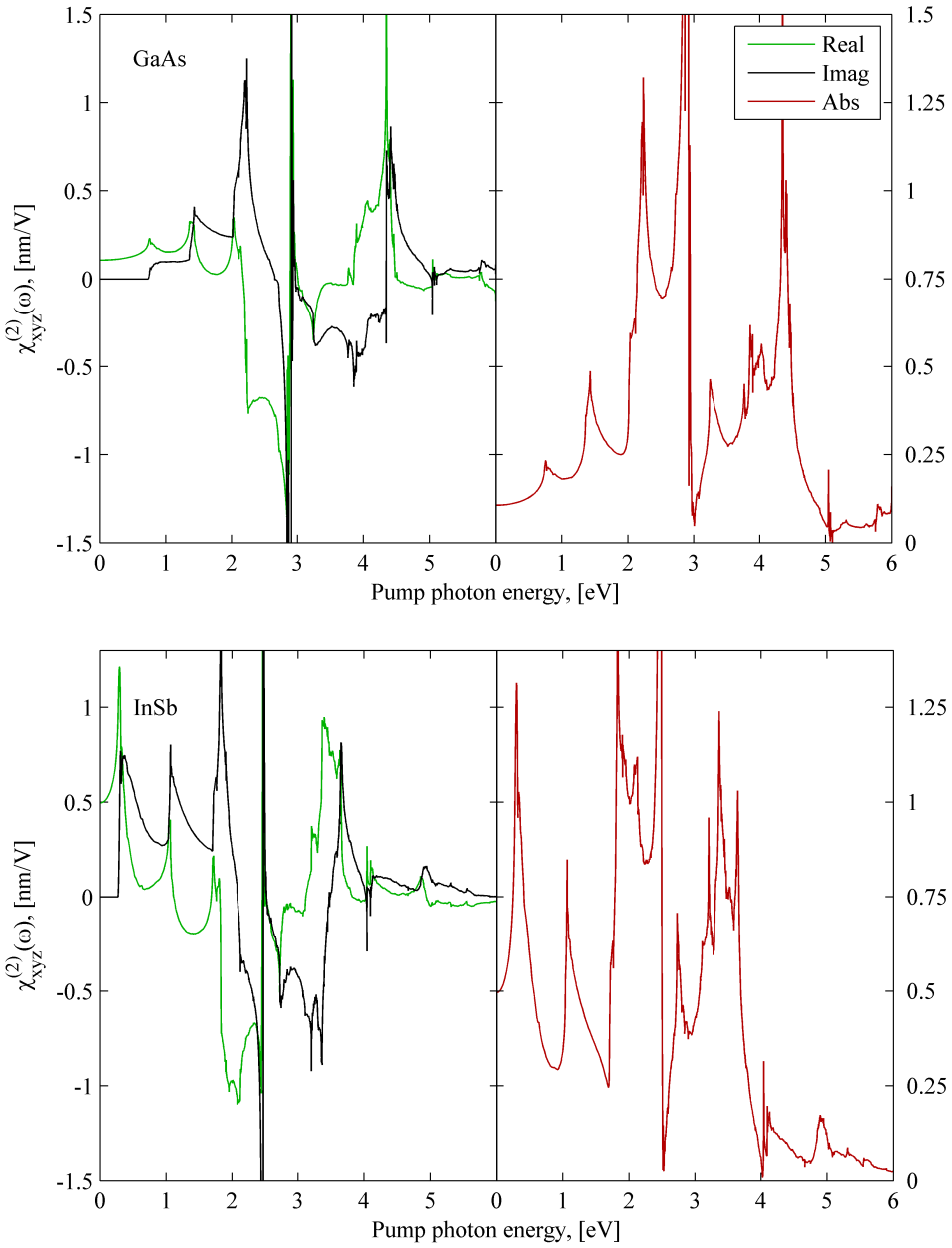


Figure 3.2: Nonlinear spectra calculated from (3.1), neglecting spin-orbit interaction and using the linear-analytic tetrahedron method generalized for second harmonic calculations (3.11). Form factors are from Tab. 2.1. All spectra are calculated using 98.304 tetrahedra in the irreducible zone, energy intervals of 1meV and a basis of 137 plane waves.

An optical, monochromatic wave propagating at free-space wavelength λ and wave vector $\vec{q}$ (in a loss-less material) display the dispersion relation $n2\pi/\lambda = q$ – hence $q \sim 10^8$ at visual frequencies. In comparison, the Brillouin zone are on the scale $G = 2\pi/a \sim 10^{11}$, a being a lattice constant. Hence $q/G \sim 10^{-3}$, making the no q vector approximation rather good at longer wavelengths. Thus, no considerable error is expected in the bulk due to this approximation relative to the dipole signal for GaAs and InSb. Considering Si and Ge, however, whose dipole allowed SHG spectrum vanishes, the q-dependency becomes the dominant bulk term.

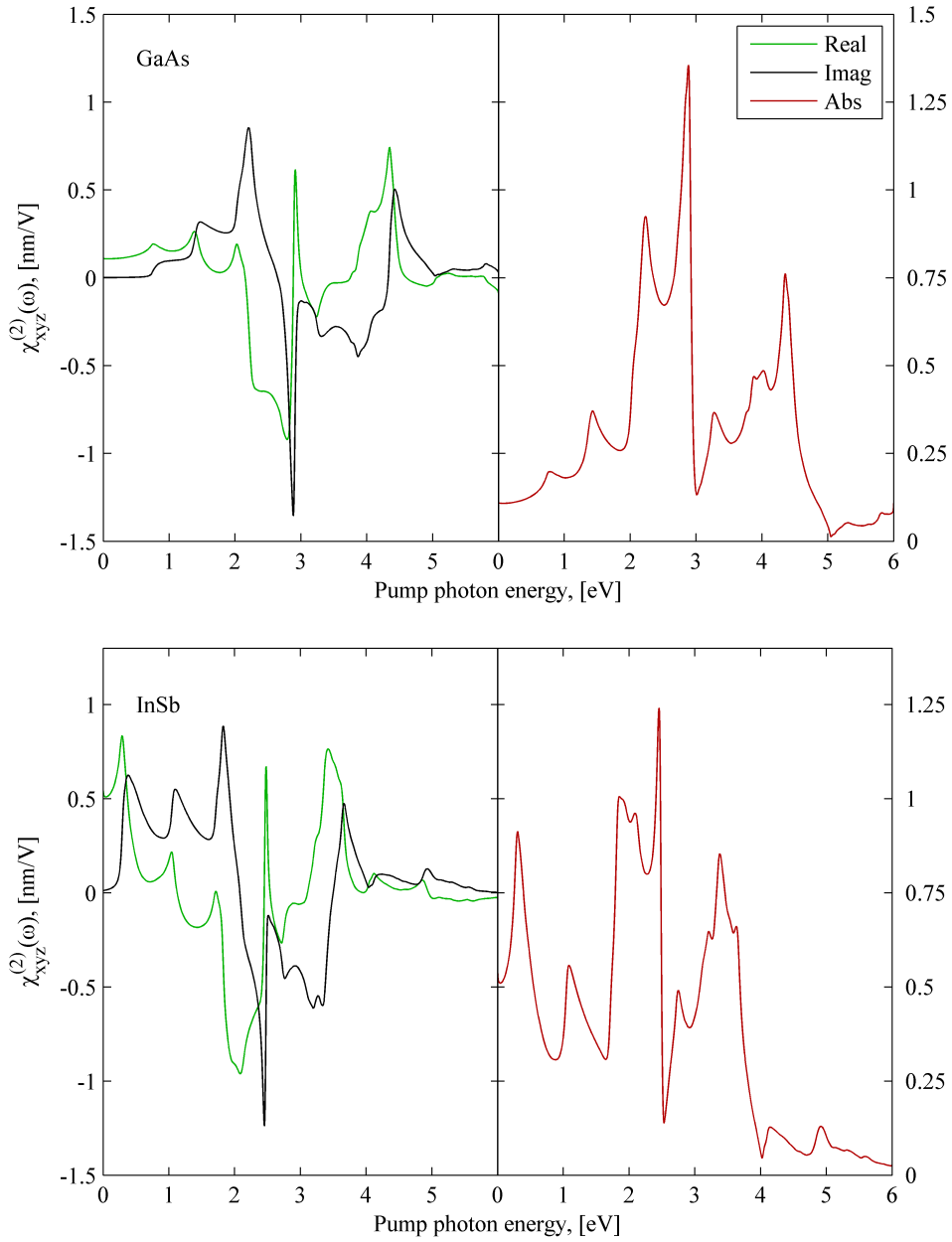


Figure 3.3: Same as Fig. 3.2, but including convolution with a 10meV Lorentzian lineshape to investigate effects of broadening.

3.3 Second harmonic spectra including spatial dispersion: Ge and Si

As was briefly discussed in the previous section, materials of inversion centers do not display bulk SHG in the dipole limit. However, in the presence of a spatially varying optical field this inversion symmetry is broken, although fields varying slowly across the unit-cell are often approximated by constant fields, maintaining the symmetry. This approximation is usually very good, however in the present section the effects of including the optical field variation will be examined. For this purpose, the second harmonic response used previously

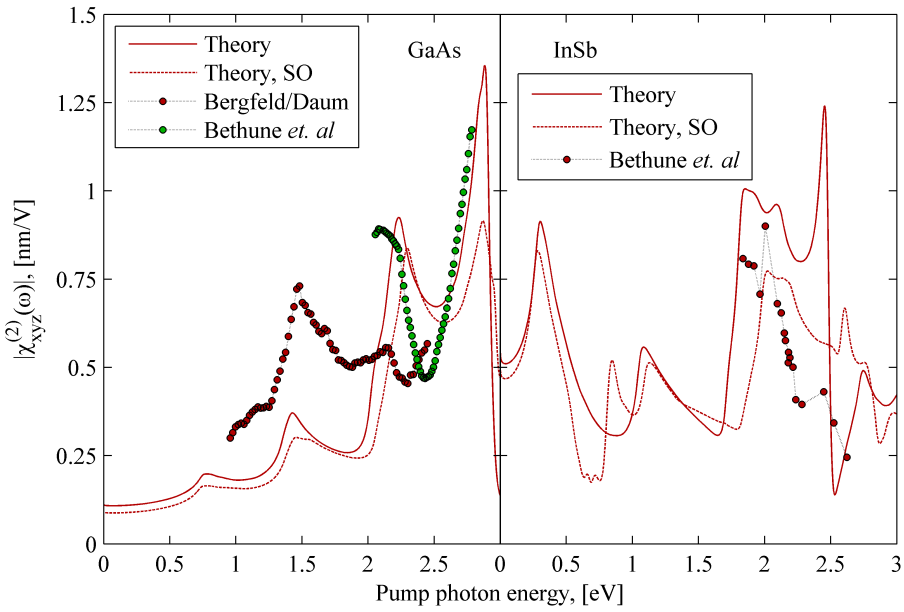


Figure 3.4: Comparison between the theoretical spectra of Fig. 3.2 and experimental results from Bergfeld & Daum[33] and Bethune *et al.*[11]. 25meV Lorentzian broadening have been applied.

(3.1) is generalized to the case of finite optical wave vector $\vec{q}$ (see A.4) to give

$$\tilde{\chi}^{(2)}_{abc}(\omega,\vec{q},\vec{q}') = -\frac{\hbar^3 e^3}{64\pi^3 m^3 \varepsilon_0 \omega^3} \sum_{\alpha,\beta,\gamma} \int G_{\alpha\vec{k}-\vec{q}',\beta\vec{k}+\vec{q},\gamma\vec{k}} I^a_{\alpha\vec{k}-\vec{q}',\beta\vec{k}+\vec{q}} I^b_{\beta\vec{k}+\vec{q},\gamma\vec{k}} I^c_{\gamma\vec{k},\alpha\vec{k}-\vec{q}'} \, d^3k. \quad (3.13)$$

where $G_{\alpha\beta\gamma}$ is given by

$$G_{\alpha\beta\gamma l} = \frac{1}{\hbar^2(2w-\omega_{\beta\alpha})}\left(\frac{f_{\alpha\gamma}}{w-\omega_{\gamma\alpha}} + \frac{f_{\beta\gamma}}{w-\omega_{\beta\gamma}}\right),$$

$I^a_{\alpha-\vec{q}',\beta+\vec{q}}$ is a matrix element

$$I^a_{\beta\vec{k}-\vec{q}',\beta\vec{k}+\vec{q}} = \langle\alpha\vec{k}-\vec{q}'|\hat{I}^a|\beta\vec{k}+\vec{q}\rangle = \langle\alpha\vec{k}-\vec{q}'|\left[e^{-i(\vec{q}+\vec{q}')\cdot\vec{r}}\nabla + \nabla e^{-i(\vec{q}+\vec{q}')\cdot\vec{r}}\right]|\beta\vec{k}+\vec{q}\rangle,$$

$\hat{a}$ is one of the Cartesian unit vectors, w is the complex frequency $\omega + i\Gamma$ and $f_{\alpha\beta}$ is the Fermi energy difference $f_\alpha - f_\beta$. We shall investigate the limit of small optical wave vectors by $\vec{k}\cdot\vec{p}$ perturbation theory; we need states of the type $|\alpha\vec{k}+\vec{q}\rangle$. Inserting into the Schrödinger equation yields

$$\left[\hat{H}_{\vec{k}} + \hat{H}_{\vec{q}}\right]|\alpha\vec{k}+\vec{q}\rangle = \tilde{E}_{\alpha\vec{k},\vec{q}}|\alpha\vec{k}+\vec{q}\rangle$$

where $\hat{H}_{\vec{k}}$ is the usual Hamiltonian and $H_{\vec{q}} = \hbar\vec{q}\cdot\hat{\vec{p}}/m$. $\tilde{E}_{\alpha\vec{k},\vec{q}} = \vec{E}_{\alpha\vec{k}+\vec{q}} + \hbar^2\vec{q}^2/2m + \hbar^2\vec{q}\cdot\vec{k}/m$, where $E_{\alpha\vec{k}+\vec{q}}$ is the energy eigenvalue of the unperturbed Hamiltonian of band α and k-vector $\vec{k}+\vec{q}$. Thus the $\vec{q}$-dependency enter as a perturbation to the Schrödinger equation, yielding by time independent perturbation theory

$$\begin{aligned}|\alpha\vec{k}+\vec{q}\rangle &= |\alpha\vec{k}\rangle + \sum_{\beta\vec{k}'\neq\alpha\vec{k}}|\beta\vec{k}'\rangle\frac{\langle\beta\vec{k}'|\hat{H}_{\vec{q}}|\alpha\vec{k}\rangle}{E_{\alpha\vec{k}}-E_{\beta\vec{k}'}} \\ &= |\alpha\vec{k}\rangle + \frac{\hbar}{m}\sum_{\beta\neq\alpha}|\beta\vec{k}\rangle\frac{\langle\beta\vec{k}|\hat{\vec{p}}|\alpha\vec{k}\rangle}{E_{\alpha\vec{k}}-E_{\beta\vec{k}}}\cdot\vec{q},\end{aligned}$$

for non-degenerate bands.With this, and noting that for "vertical" matrix elements

$$I^a_{\alpha\vec{k},\beta\vec{k}} = \frac{2i}{\hbar} p^a_{\alpha\beta},$$

we get

$$\begin{aligned}
\langle \alpha\vec{k} - \vec{q}' | \hat{p}_a | \beta\vec{k} + \vec{q} \rangle &\approx \sum_{l,l'} \left[\langle \alpha\vec{k}| - \frac{\hbar}{m} \frac{\langle \alpha\vec{k}|\hat{\vec{p}}|l'\vec{k}\rangle \langle l'\vec{k}|}{E_{\alpha\vec{k}} - E_{l'\vec{k}}} \cdot \vec{q}' \right] \hat{p}^a \left[|\beta\vec{k}\rangle + \frac{\hbar}{m} \frac{|l\vec{k}\rangle \langle l\vec{k}|\hat{\vec{p}}|\beta\vec{k}\rangle}{E_{\beta\vec{k}} - E_{l\vec{k}}} \cdot \vec{q} \right], \\
&\approx p^a_{\alpha\beta} + \sum_{l,l'} \frac{\hbar}{m} \left(\frac{p^a_{\alpha l} \vec{p}_{l\beta} \cdot \vec{q}}{E_{\beta l}} - \frac{\vec{p}_{\alpha l'} p^a_{l'\beta} \cdot \vec{q}'}{E_{\alpha l'}} \right), \quad l \neq \beta, l' \neq \alpha \\
\langle \beta\vec{k} + \vec{q} | \hat{p}_b | \gamma\vec{k} \rangle &\approx p^b_{\beta\gamma} + \sum_l \frac{\hbar}{m} \frac{\vec{p}_{\beta l} p^b_{l\gamma} \cdot \vec{q}}{E_{\beta l}}, \quad l \neq \beta \\
\langle \gamma\vec{k} | \hat{p}_c | \alpha\vec{k} - \vec{q}' \rangle &\approx p^c_{\gamma\alpha} - \sum_l \frac{\hbar}{m} \frac{p^c_{\gamma l} \vec{p}_{l\alpha} \cdot \vec{q}'}{E_{\alpha l}}, \quad l \neq \alpha
\end{aligned}$$

where only terms up to second order in q have been retained. Additionally, the wave vector subscript on the energy terms have been dropped. Inserting into (3.13), keeping only terms up to first order in q

$$\begin{aligned}
\tilde{\chi}^{(2)}_{abc}(\omega, \vec{q}, \vec{q}') &= \frac{ie^3}{8\pi^3 m^3 \varepsilon_0 \omega^3} \sum_{\alpha,\beta,\gamma} \int G_{\alpha\beta\gamma} \\
&\times \left(p^a_{\alpha\beta} + \sum_{l,l'} \frac{\hbar}{m} \left(\frac{p^a_{\alpha l} \vec{p}_{l\beta} \cdot \vec{q}}{E_{\beta l}} - \frac{\vec{p}_{\alpha l'} p^a_{l'\beta} \cdot \vec{q}'}{E_{\alpha l'}} \right) \right) \left(p^b_{\beta\gamma} + \sum_l \frac{\hbar}{m} \frac{\vec{p}_{\beta l} p^b_{l\gamma} \cdot \vec{q}}{E_{\beta l}} \right) \\
&\times \left(p^c_{\gamma\alpha} - \sum_l \frac{\hbar}{m} \frac{p^c_{\gamma l} \vec{p}_{l\alpha} \cdot \vec{q}'}{E_{\alpha a'}} \right) d^3k. \\
&\approx \frac{ie^3}{8\pi^3 m^3 \varepsilon_0 \omega^3} \sum_{\alpha,\beta,\gamma} \int d^3k \, G_{\alpha\beta\gamma} \Bigg[p^a_{\alpha\beta} p^b_{\beta\gamma} p^c_{\gamma\alpha} + \frac{\hbar}{m} \sum_l \Bigg(\\
&\frac{p^a_{\alpha\beta} \vec{p}_{\beta l} p^b_{l\gamma} p^c_{\gamma\alpha} \cdot \vec{q}}{E_{\beta l}} - \frac{p^a_{\alpha\beta} p^b_{\beta\gamma} p^c_{\gamma l} \vec{p}_{l\alpha} \cdot \vec{q}'}{E_{\alpha l}} + \frac{p^a_{\alpha l} \vec{p}_{l\beta} p^b_{\beta\gamma} p^c_{\gamma\alpha} \cdot \vec{q}}{E_{\beta l}} - \frac{\vec{p}_{\alpha l} p^a_{l\beta} p^b_{\beta\gamma} p^c_{\gamma\alpha} \cdot \vec{q}'}{E_{\alpha l}} \Bigg) \Bigg],
\end{aligned}$$

Considering systems having an inversion center, the term containing three momentum matrix elements will vanish, as discussed previously. Hence,

$$\begin{aligned}
\tilde{\chi}^{(2)}_{abc}(\omega, \vec{q}, \vec{q}') &= \frac{i\hbar e^3}{8\pi^3 m^4 \varepsilon_0 \omega^3} \sum_{\alpha,\beta,\gamma,l} \int d^3k \, G_{\alpha\beta\gamma} \\
&\left(\frac{p^a_{\alpha\beta} \vec{p}_{\beta l} p^b_{l\gamma} p^c_{\gamma\alpha} \cdot \vec{q}}{E_{\beta l}} - \frac{p^a_{\alpha\beta} p^b_{\beta\gamma} p^c_{\gamma l} \vec{p}_{l\alpha} \cdot \vec{q}'}{E_{\alpha l}} + \frac{p^a_{\alpha l} \vec{p}_{l\beta} p^b_{\beta\gamma} p^c_{\gamma\alpha} \cdot \vec{q}}{E_{\beta l}} - \frac{\vec{p}_{\alpha l} p^a_{l\beta} p^b_{\beta\gamma} p^c_{\gamma\alpha} \cdot \vec{q}'}{E_{\alpha l}} \right).
\end{aligned}$$

Inside the integral, the matrix elements obey the following relation due to the integration limits being symmetric and the relation $-p^*_{ij}(\vec{k}) = p_{ij}(-\vec{k})$

$$\begin{aligned}
p^a_{\alpha\beta} p^b_{\beta i} p^c_{i\gamma} p^d_{\gamma\alpha} &\sim \frac{1}{2} \left(\left[p^a_{\alpha\beta} p^b_{\beta i} p^c_{i\gamma} p^d_{\gamma\alpha} \right] + \left[p^a_{\alpha\beta} p^b_{\beta i} p^c_{i\gamma} p^d_{\gamma\alpha} \right]^* \right) \\
&= \mathrm{Re} \left[p^a_{\alpha\beta} p^b_{\beta i} p^c_{i\gamma} p^d_{\gamma\alpha} \right]. \qquad (3.14)
\end{aligned}$$

Using this, and writing out the components of the q-vectors as q_d, where the sum over d is implicit

$$\tilde{\chi}^{(2)}_{abcd}(\omega,\vec{q},\vec{q}') = \frac{i\hbar e^3}{8\pi^3 m^4 \varepsilon_0 \omega^3} \sum_{\alpha,\beta,\gamma,l} \int G_{\alpha\beta\gamma} \times \left(\frac{\mathrm{Re}[p^a_{\alpha\beta} p^d_{\beta l} p^b_{l\gamma} p^c_{\gamma\alpha}] + \mathrm{Re}[p^a_{\alpha l} p^d_{l\beta} p^b_{\beta\gamma} p^c_{\gamma\alpha}]}{E_{\beta l}} q_d - \frac{\mathrm{Re}[p^a_{\alpha\beta} p^b_{\beta\gamma} p^c_{\gamma l} p^d_{l\alpha}] + \mathrm{Re}[p^d_{\alpha l} p^a_{l\beta} p^b_{\beta\gamma} p^c_{\gamma\alpha}]}{E_{\alpha l}} q'_d \right) d^3k.$$

The similarity between this expression and the calculations of third harmonic generation performed by Moss *et al.* [32] is emphasized. Making the dummy variable substitutions $\beta \leftrightarrow l$ in the second term and $\alpha \leftrightarrow l$ in the fourth

$$\tilde{\chi}^{(2)}_{abcd}(\omega,\vec{q},\vec{q}') = \frac{i\hbar e^3}{8\pi^3 m^4 \varepsilon_0 \omega^3} \sum_{\alpha,\beta,\gamma,l} \int d^3k \times \left(\mathrm{Re}[p^a_{\alpha\beta} p^d_{\beta l} p^b_{l\gamma} p^c_{\gamma\alpha}] \frac{\left(G_{\alpha\beta\gamma} - G_{\alpha l\gamma}\right) q_d}{E_{\beta l}} - \mathrm{Re}[p^a_{\alpha\beta} p^b_{\beta\gamma} p^c_{\gamma l} p^d_{l\alpha}] \frac{\left(G_{\alpha\beta\gamma} - G_{l\beta\gamma}\right) q'_d}{E_{\alpha l}} \right).$$

$G_{\alpha\beta\gamma}$ in the first order $\vec{k}\cdot\vec{p}$-approximation is given by

$$G_{\alpha\beta\gamma} = \frac{1}{2w - \omega_{\beta\alpha} - \delta\omega_{\beta\alpha}} \left(\frac{f_{\alpha\gamma}}{w - \omega_{\gamma\alpha} + \delta\omega_\alpha} + \frac{f_{\beta\gamma}}{w - \omega_{\beta\gamma} - \delta\omega_\beta} \right) \frac{1}{\hbar^2}$$

where $\delta\omega_\alpha$ is the first order energy correction to $|\alpha\rangle$ due to the $\vec{k}\cdot\vec{p}$-perturbation, whereas $\delta\omega_{\alpha\beta} = \delta\omega_\alpha - \delta\omega_\beta$. The summation over α,β,γ and l is seperated into contributions from valence and conductions bands, using the low temperature limits

$$f_{\alpha\beta} = \begin{cases} 1, & \alpha \in c, \beta \in v \\ -1, & \alpha \in v, \beta \in c \\ 0, & \alpha \in c, \beta \in c \\ 0, & \alpha \in v, \beta \in v \end{cases},$$

with c denoting unoccupied conduction bands whereas v indicate occupied valence bands. In Tab. 3.1 all 16 combinations of α, β, γ and l in the sets c and v are shown, where A and B are defined by

$$\begin{aligned} A \equiv \frac{\hbar^3 \left(G_{\alpha\beta\gamma} - G_{\alpha l\gamma}\right)}{\omega_{\beta l}} &= \frac{1}{\omega_{\beta l}(2w - \omega_{\beta\alpha} - \delta\omega_{\beta\alpha})} \left(\frac{f_{\alpha\gamma}}{w - \omega_{\gamma\alpha} + \delta\omega_\alpha} + \frac{f_{\beta\gamma}}{w - \omega_{\beta\gamma} - \delta\omega_\beta} \right) \\ &\quad - \frac{1}{\omega_{\beta l}(2w - \omega_{l\alpha} - \delta\omega_{l\alpha})} \left(\frac{f_{\alpha\gamma}}{w - \omega_{\gamma\alpha} + \delta\omega_\alpha} + \frac{f_{l\gamma}}{w - \omega_{l\gamma} - \delta\omega_l} \right) \\ B \equiv \frac{\hbar^3 \left(G_{\alpha\beta\gamma} - G_{l\beta\gamma}\right)}{\omega_{\alpha l}} &= \frac{1}{\omega_{\alpha l}(2w - \omega_{\beta\alpha} - \delta\omega_{\beta\alpha})} \left(\frac{f_{\alpha\gamma}}{w - \omega_{\gamma\alpha} + \delta\omega_\alpha} + \frac{f_{\beta\gamma}}{w - \omega_{\beta\gamma} - \delta\omega_\beta} \right) \\ &\quad - \frac{1}{\omega_{\alpha l}(2w - \omega_{\beta l} - \delta\omega_{\beta l})} \left(\frac{f_{l\gamma}}{w - \omega_{\gamma l} + \delta\omega_l} + \frac{f_{\beta\gamma}}{w - \omega_{\beta\gamma} - \delta\omega_\beta} \right). \end{aligned}$$

Obviously, when all bands come from the same set, no contribution is found. Whenever β belongs to the same set as l, the first and second line of A becomes equivalent (in the

sense of canceling when the sum is written out) – hence in these cases no contribution from A is found. The same is true for B, although here the terms where α and l belong to the same set cancels – this fills in the zeros of Tab. 3.1. In Eqs (3.15)-(3.22) A and B, where only a single term contributes, is written out. The notation A^{ijko} is understood to mean $A(\alpha \in i, \beta \in j, \gamma \in k, l \in o)$, where $\{ijko\}$ denote band subsets (c or v);

$$A^{vcvv} = \frac{1}{\omega_{\beta l}(2w - \omega_{\beta\alpha} - \delta\omega_{\beta\alpha})(w - \omega_{\beta\gamma} - \delta\omega_{\beta})} \tag{3.15}$$

$$A^{vvvc} = \frac{-1}{\omega_{\beta l}(2w - \omega_{l\alpha} - \delta\omega_{l\alpha})(w - \omega_{l\gamma} - \delta\omega_{l})} \tag{3.16}$$

$$A^{cccv} = \frac{1}{\omega_{\beta l}(2w - \omega_{l\alpha} - \delta\omega_{l\alpha})(w - \omega_{l\gamma} - \delta\omega_{l})} \tag{3.17}$$

$$A^{cvcc} = \frac{-1}{\omega_{\beta l}(2w - \omega_{\beta\alpha} - \delta\omega_{\beta\alpha})(w - \omega_{\beta\gamma} - \delta\omega_{\beta})} \tag{3.18}$$

$$B^{vvvc} = \frac{-1}{\omega_{\alpha l}(2w - \omega_{\beta l} - \delta\omega_{\beta l})(w - \omega_{\gamma l} - \delta\omega_{l})} \tag{3.19}$$

$$B^{cvvv} = \frac{1}{\omega_{\alpha l}(2w - \omega_{\beta\alpha} - \delta\omega_{\beta\alpha})(w - \omega_{\gamma\alpha} - \delta\omega_{\alpha})} \tag{3.20}$$

$$B^{vccc} = \frac{-1}{\omega_{\alpha l}(2w - \omega_{\beta\alpha} - \delta\omega_{\beta\alpha})(w - \omega_{\gamma\alpha} - \delta\omega_{\alpha})} \tag{3.21}$$

$$B^{cccv} = \frac{1}{\omega_{\alpha l}(2w - \omega_{\beta l} - \delta\omega_{\beta l})(w - \omega_{\gamma l} - \delta\omega_{l})} \tag{3.22}$$

The real part of the response is extracted:

$$\mathrm{Re}\left[\tilde{\chi}^{abcd}_{2\omega,\vec{q},\vec{q}'}\right] = \frac{-e^3}{8\hbar^2\omega^3\pi^3 m^4 \varepsilon_0} \sum_{\alpha,\beta,\gamma,l} \int d^3k \\ \times \mathrm{Im}\left[\left(A^{vcvv}\mathrm{Re}[p^{vc}_{\alpha\beta}p^{cv}_{\beta l}p^{vv}_{l\gamma}p^{vv}_{\gamma\alpha}] + ...\right) q_b + \text{B-terms}\right],$$

where all matrix elements have been written in the order *abcd* to make room for the superscripts indicating which band subsets the indices belong to. From this, it is clear that the imaginary part of the A and B terms are needed. Taking the contribution from A^{vcvv} as an example; A^{vcvv} is expanded in partial fractions such that

$$A^{vcvv}w^{-3} = C_1 w^{-3} + C_2 w^{-2} + C_3 w^{-1} + C(w) \tag{3.23}$$

where

$$C_1 = \frac{1}{\omega'_{\beta l}\omega_{\beta\alpha}\omega_{\beta\gamma}},$$

$$C_2 = \frac{2\omega_{\beta\gamma} + \omega_{\beta\alpha}}{\omega'_{\beta l}\omega^2_{\beta\alpha}\omega^2_{\beta\gamma}},$$

$$C_3 = \frac{4\omega^2_{\beta\gamma} + 2\omega_{\beta\alpha}\omega_{\beta\gamma} + \omega^2_{\beta\alpha}}{\omega'_{\beta l}\omega^3_{\beta\alpha}\omega^3_{\beta\gamma}},$$

$$C(\omega) = \frac{-16}{\omega'_{\beta l}(2\omega - \omega_{\beta\alpha})(\omega_{\beta\alpha} - 2\omega_{\beta\gamma})\omega^3_{\beta\alpha}} - \frac{1}{\omega'_{\beta l}(\omega - \omega_{\beta\gamma})(\omega_{\beta\alpha} - 2\omega_{\beta\gamma})\omega^3_{\beta\gamma}},$$

where the first order energy corrections have been included in ω_{ij} and the prime on $\omega_{\beta l}$ indicates that the difference contains no corrections due to $\vec{k}\cdot\vec{p}$-perturbation. In the static

limit $\omega \to 0$ a finite result is expected, why the contributions from the C_1, C_2 and C_3 terms must vanish. In reference [8] very similar terms have indeed been shown to vanish in the optical limit $\{q, q'\} \to 0$, and in in the investigation of third-harmonic generation Moss *et al.*[32] have neglected similar terms using the same argument of finite results when taking ω to zero. Thus it is believed that this is a fair assumption, although no success was made in showing this analytically. Thus, only $C(\omega)$ needs consideration; proceeding to take the imaginary part

$$\mathrm{Im}[A^{vcvv}w^{-3}] = \mathrm{Im}\left[\frac{-16}{\omega'_{\beta l}(2w-\omega_{\beta\alpha})(\omega_{\beta\alpha}-2\omega_{\beta\gamma})\omega_{\beta\alpha}^3}\right] - \mathrm{Im}\left[\frac{1}{\omega'_{\beta l}(w-\omega_{\beta\gamma})(\omega_{\beta\alpha}-2\omega_{\beta\gamma})\omega_{\beta\gamma}^3}\right].$$

Taking the limit of vanishing broadening and investigating the first term

$$\begin{aligned}
\mathrm{Im}[A_1^{vcvv}w^{-3}] &= \lim_{\Gamma\to 0}\mathrm{Im}\left[\frac{-16}{\omega'_{\beta l}\left(2[\omega+i\Gamma]-\omega_{\beta\alpha}\right)\left(\omega_{\beta\alpha}-2\omega_{\beta\gamma}\right)\omega_{\beta\alpha}^3}\right] \\
&= \frac{-16}{\omega'_{\beta l}\left(\omega_{\beta\alpha}-2\omega_{\beta\gamma}\right)\omega_{\beta\alpha}^3}\lim_{\Gamma\to 0}\mathrm{Im}\left[\frac{\left(2\omega-2i\Gamma-\omega_{\beta\alpha}\right)}{\left(2\omega+2i\Gamma-\omega_{\beta\alpha}\right)\left(2\omega-2i\Gamma-\omega_{\beta\alpha}\right)}\right] \\
&= \frac{-16}{\omega'_{\beta l}\left(\omega_{\beta\alpha}-2\omega_{\beta\gamma}\right)\omega_{\beta\alpha}^3}\lim_{\Gamma\to 0}\left[\frac{2\Gamma}{(2\omega-\omega_{\beta\alpha})^2+4\Gamma^2}\right] \\
&= \frac{-16\pi}{\omega'_{\beta l}\left(\omega_{\beta\alpha}-2\omega_{\beta\gamma}\right)\omega_{\beta\alpha}^3}\delta\left(2\omega-\omega_{\beta\alpha}\right),
\end{aligned}$$

where the identity

$$\lim_{\varepsilon\to 0}\frac{\varepsilon}{x^2+\varepsilon^2} = \pi\delta(x) \tag{3.24}$$

was used in the last equality. Similarly for the second term, where (3.24) is used again

$$\begin{aligned}
\mathrm{Im}[A_2^{vcvv}w^{-3}] &= \lim_{\Gamma\to 0}\mathrm{Im}\left[\frac{-1}{\omega'_{\beta l}([\omega+i\Gamma]-\omega_{\beta\gamma})(\omega_{\beta\alpha}-2\omega_{\beta\gamma})\omega_{\beta\gamma}^3}\right] \\
&= \frac{-\pi}{\omega'_{\beta l}(\omega_{\beta\alpha}-2\omega_{\beta\gamma})\omega_{\beta\gamma}^3}\delta\left(\omega-\omega_{\beta\gamma}\right).
\end{aligned} \tag{3.25}$$

We also have contributions where more than one term in A and B contributes, e.g. A^{vccv}

$$\begin{aligned}
A^{vccv} &= \frac{1}{\omega'_{\beta l}}\left[\frac{-1}{(2w-\omega_{\beta\alpha})(w-\omega_{\gamma\alpha})}+\frac{1}{(2w-\omega_{l\alpha})(w-\omega_{\gamma\alpha})}+\frac{1}{(2w-\omega_{l\alpha})(w-\omega_{l\gamma})}\right] \\
&= \frac{1}{\omega'_{\beta l}}\left[\frac{1}{(w-\omega_{\gamma\alpha})(w-\omega_{l\gamma})}-\frac{1}{(2w-\omega_{\beta\alpha})(w-\omega_{\gamma\alpha})}\right],
\end{aligned} \tag{3.26}$$

$$A^{vvcc} = \frac{1}{\omega'_{\beta l}}\left[\frac{1}{(2w-\omega_{l\alpha})(w-\omega_{\gamma\alpha})}-\frac{1}{(w-\omega_{\gamma\alpha})(w-\omega_{\beta\gamma})}\right]. \tag{3.27}$$

Expanding A^{vccv} in partial fractions as before, ignoring all terms divergent for $\omega \to 0$

$$\begin{aligned}
A^{vccv}w^{-3} = \frac{1}{\omega'_{\beta l}}\Bigg[&\frac{16}{\left(2w-\omega_{\beta\alpha}\right)\omega_{\beta\alpha}{}^3\left(2\omega_{\gamma\alpha}-\omega_{\beta\alpha}\right)} \\
&-\frac{1}{\left(w-\omega_{l\gamma}\right)\left(\omega_{l\gamma}-\omega_{\gamma\alpha}\right)\omega_{l\gamma}{}^3}-\frac{\omega_{l\beta}}{\left(w-\omega_{\gamma\alpha}\right)\left(\omega_{l\gamma}-\omega_{\gamma\alpha}\right)\omega_{\gamma\alpha}{}^3\left(2\,\omega_{\gamma\alpha}-\omega_{\beta\alpha}\right)}\Bigg]
\end{aligned}$$

Taking the imaginary part and converting to δ-functions

$$\lim_{\Gamma\to 0}\mathrm{Im}[A^{vccv}w^{-3}] = \frac{\pi}{\omega'_{\beta l}}\left[\frac{16\,\delta\left(2\omega-\omega_{\beta\alpha}\right)}{{\omega_{\beta\alpha}}^3\left(2\omega_{\gamma\alpha}-\omega_{\beta\alpha}\right)}\right.$$
$$\left.-\frac{\omega_{l\beta}\delta\left(\omega-\omega_{\gamma\alpha}\right)}{\left(\omega_{l\gamma}-\omega_{\gamma\alpha}\right)\left(2\omega_{\gamma\alpha}-\omega_{\beta\alpha}\right){\omega_{\gamma\alpha}}^3}-\frac{\delta\left(\omega-\omega_{l\gamma}\right)}{{\omega_{l\gamma}}^3\left(\omega_{l\gamma}-\omega_{\gamma\alpha}\right)}\right]$$

If we first consider energies below the band gap, only the terms containing a δ-function with 2ω as argument contribute. Furthermore, terms containing δ-functions of the kind $\delta(2\omega-\omega_{vc})$ will never contribute, since in a semiconductor $\omega_c > \omega_v$ and only positive frequencies are investigated. Thus only A^{vcvv}, A^{vvvc}, A^{vccv} and A^{vvcc} will contribute, one term each

$$A_2 = \sum_{\alpha\beta\gamma l}\frac{16\pi}{\omega'_{\beta l}}\left[-\frac{\mathrm{Re}[p^{vc}_{\alpha\beta}p^{cv}_{\beta l}p^{vv}_{l\gamma}p^{vv}_{\gamma\alpha}]\delta(2\omega-\omega_{\beta\alpha})}{(\omega_{\beta\alpha}-2\omega_{\beta\gamma})\omega^3_{\beta\alpha}}+\frac{\mathrm{Re}[p^{vv}_{\alpha\beta}p^{vc}_{\beta l}p^{cv}_{l\gamma}p^{vv}_{\gamma\alpha}]\delta(2\omega-\omega_{l\alpha})}{(\omega_{l\alpha}-2\omega_{l\gamma})\omega^3_{l\alpha}}\right.$$
$$\left.-\frac{\mathrm{Re}[p^{vc}_{\alpha\beta}p^{cv}_{\beta l}p^{vc}_{l\gamma}p^{cv}_{\gamma\alpha}]\delta(2\omega-\omega_{\beta\alpha})}{(\omega_{\beta\alpha}-2\omega_{\gamma\alpha})\omega^3_{\beta\alpha}}+\frac{\mathrm{Re}[p^{vv}_{\alpha\beta}p^{vc}_{\beta l}p^{cc}_{l\gamma}p^{cv}_{\gamma\alpha}]\delta(2\omega-\omega_{l\alpha})}{(\omega_{l\alpha}-2\omega_{\gamma\alpha})\omega^3_{l\alpha}}\right].$$

The band indices of the frequency differences in the δ-functions run over the same set for all terms, and are renamed c, v

$$A_2 = 16\pi\sum_{c,v}\left[\sum_{\gamma,l}^{v,v}\frac{-\mathrm{Re}[p_{vc}p_{cl}p_{l\gamma}p_{\gamma v}]}{(\omega_{cv}-2\omega_{c\gamma})\omega^3_{cv}\omega'_{cl}}+\sum_{\beta,\gamma}^{v,v}\frac{\mathrm{Re}[p_{v\beta}p_{\beta c}p_{c\gamma}p_{\gamma v}]}{(\omega-\omega_{c\gamma})\omega^3\omega'_{\beta c}}\right.$$
$$\left.-\sum_{\gamma,l}^{c,v}\frac{\mathrm{Re}[p_{vc}p_{cl}p_{l\gamma}p_{\gamma v}]}{(\omega-\omega_{\gamma v})\omega^3\omega'_{cl}}+\sum_{\beta,\gamma}^{v,c}\frac{\mathrm{Re}[p_{v\beta}p_{\beta c}p_{c\gamma}p_{\gamma v}]}{(\omega-\omega_{\gamma v})\omega^3\omega'_{\beta c}}\right]\delta(2\omega-\omega_{cv}),$$

where the superscript on the sums are used to indicate the summation subsets. Relabeling allows a slightly more convenient form,

$$A_2 = \frac{\pi}{\omega^3}\sum_{c,v}\sum_{l}^{v}\left[\sum_{\gamma}^{v}\frac{\mathrm{Re}[p_{vl}p_{lc}p_{c\gamma}p_{\gamma v}]+\mathrm{Re}[p_{vc}p_{cl}p_{l\gamma}p_{\gamma v}]}{(\omega-\omega_{c\gamma})\omega'_{lc}}\right.$$
$$\left.+\sum_{\gamma}^{c}\frac{\mathrm{Re}[p_{vl}p_{lc}p_{c\gamma}p_{\gamma v}]+\mathrm{Re}[p_{vc}p_{cl}p_{l\gamma}p_{\gamma v}]}{(\omega-\omega_{\gamma v})\omega'_{lc}}\right]\delta(2\omega-\omega_{cv}). \tag{3.28}$$

Similarly B^{vccc}, B^{cccv}, B^{vcvc} and B^{ccvv} contribute with 2ω-terms, again neglecting all terms proportional to $\delta(2\omega-\omega_{vc})$

$$B_2 = \frac{\pi}{\omega^3}\sum_{c,v}\sum_{l}^{c}\left[\sum_{\gamma}^{v}\frac{\mathrm{Re}[p_{lc}p_{c\gamma}p_{\gamma v}p_{vl}]+\mathrm{Re}[p_{vc}p_{c\gamma}p_{\gamma l}p_{lv}]}{(\omega-\omega_{c\gamma})\omega'_{vl}}\right.$$
$$\left.+\sum_{\gamma}^{c}\frac{\mathrm{Re}[p_{lc}p_{c\gamma}p_{\gamma v}p_{vl}]+\mathrm{Re}[p_{vc}p_{c\gamma}p_{\gamma l}p_{lv}]}{(\omega-\omega_{\gamma v})\omega'_{lv}}\right]\delta(2\omega-\omega_{cv}).$$

Denoting $\mathrm{Re}[\tilde{\chi}^{abcd}_{2\omega,\vec{q},\vec{q}'}] = \tilde{\chi}^{abcd\prime}_{2\omega,\vec{q},\vec{q}'}$, the final result becomes

$$\tilde{\chi}^{(2)}_{abcd}(\omega,\vec{q},\vec{q}') = \frac{-e^3}{8\hbar^2\pi^3m^4\varepsilon_0}\int\left[A_2q_b+B_2q'_d\right]\,d^3k. \tag{3.29}$$

If we consider the case where only a single monochromatic wave of frequency ω and wave-vector $\vec{q}$ propagate in a sample, then $\vec{q}' = \vec{q}$. Such a wave will satisfy a dispersion relation $q = n\omega/c$. Using this, we get

$$\tilde{\chi}^{(2)}_{abcd}(\omega,\vec{q},\vec{q}) = -\frac{ne^3\omega}{8\hbar^2\pi^3m^4\varepsilon_0 c}\int\left[A_2^{adcb}+B_2^{abcd}\right]\hat{d}\cdot\hat{q}\,d^3k. \tag{3.30}$$

where the superscript on A_2 and B_2 indicate in which order of the Cartesian indices on the momentum matrix elements appear; i.e. the b and d indices have been interchanged in the A_2 term. Furthermore, the first order energy corrections are neglected since the q-vectors become extremely small compared to the size of the Brillouin zone at energies below the direct gap.

In order to identify which elements should be calculated of the fourth rank tensor (3.30), the phenomenology of second harmonic generation is briefly considered. Sipe *et al.* (1987)[34] presents a general expression for the symmetry of the fourth rank response tensor equivalent to (3.30) for cubic materials of inversion symmetry

$$\Gamma_{ijkl} = a_1\delta_{ijkl} + a_2\delta_{ij}\delta_{kl} + a_3\delta_{ik}\delta_{jl} + a_4\delta_{il}\delta_{jk},$$

where a_i are frequency-dependent parameters and $\delta_{ij}\delta_{kl}$ is understood to mean $\{i=j\}\neq\{k=l\}$. Hence, bulk cubic symmetry dictates that only the *aaaa*, *aabb*, *abab* and *abba* elements needs consideration in order to specify the entire system, where a and b correspond to any two (different) Cartesian indices. Reconsidering the form of the second harmonic polarization response

$$\mathcal{P}^{a\prime}_{2\omega} = \sum_{b,c,d}\varepsilon_0 X^{(2)}_{abcd}(\omega,\vec{q},\vec{q})\mathcal{E}^b_{\omega,q}\mathcal{E}^c_{\omega,q}q_d,$$

where $\tilde{\chi}^{(2)}_{abcd}(\omega,\vec{q},\vec{q}) = X^{(2)}_{abcd}(\omega,\vec{q},\vec{q})q_d$, and writing out the sum suppressing the ω and q arguments, keeping in mind that only the terms allowed by (3.3) contributes

$$\begin{aligned}\mathcal{P}^{a\prime}_{2\omega} = {} & X^{(2)}_{aaaa}q_a\mathcal{E}_a^2 + X^{(2)}_{aabb}(\mathcal{E}_a\mathcal{E}_bq_b+\mathcal{E}_a\mathcal{E}_cq_c) + X^{(2)}_{abab}(\mathcal{E}_a\mathcal{E}_bq_b+\mathcal{E}_a\mathcal{E}_cq_c)\\ & + X^{(2)}_{abba}(\mathcal{E}_b^2q_a+\mathcal{E}_c^2q_a)\end{aligned}$$

where b and c are cartesian indices different from a. Rewriting to a more convenient vector notation

$$\begin{aligned}\mathcal{P}^{a\prime}_{2\omega} &= X^{(2)}_{aaaa}q_a\mathcal{E}_a^2 + (X^{(2)}_{aabb}+X^{(2)}_{abab})(\mathcal{E}_a\vec{\mathcal{E}}\cdot\vec{q}-\mathcal{E}_a^2q_a) + X^{(2)}_{abba}(\vec{\mathcal{E}}\cdot\vec{\mathcal{E}}q_a-\mathcal{E}_a^2q_a)\\ &= (X^{(2)}_{aaaa}-X^{(2)}_{aabb}-X^{(2)}_{abab}-X^{(2)}_{abba})\mathcal{E}_a^2q_a + (X^{(2)}_{aabb}+X^{(2)}_{abab})\mathcal{E}_a\vec{\mathcal{E}}\cdot\vec{q} + X^{(2)}_{abba}\vec{\mathcal{E}}\cdot\vec{\mathcal{E}}q_a\\ &\equiv X_1\mathcal{E}_a^2q_a + X_2\mathcal{E}_a\vec{\mathcal{E}}\cdot\vec{q} + X^{(2)}_{abba}\vec{\mathcal{E}}\cdot\vec{\mathcal{E}}q_a\end{aligned} \tag{3.31}$$

Furthermore; when considering the response to a transversal field, $\hat{q}$ and $\vec{\mathcal{E}}$ are perpendicular – thus the second term above vanishes. According to Sipe *et al.*[34], the first term will not radiate in an infinite medium, however a p-polarized bulk contribution is radiated from this term into a vacuum half-space. Sipe *et al.* also shows the last term of (3.31) to be anisotropic with respect to rotation of the crystal on an axis perpendicular to the surface. This bulk contribution radiates both s- and p-polarized light into a vacuum half-space. Thus, if one measures the radiated second harmonic fields under p-polarization a combination of X_1 and X^{abba} (in addition to surface contributions) is investigated, whereas measuring the response under s-polarization will allow the bulk contribution from X^{abba} to be investigated.

α	β	γ	l	$f_{\alpha\gamma}$	$f_{\beta\gamma}$	$f_{l\gamma}$	A	B
v	v	v	v	0	0	0	0	0
v	c	v	v	0	1	0	(3.15)	0
v	v	c	v	-1	-1	-1	0	0
v	v	v	c	0	0	1	(3.16)	(3.19)
v	c	c	v	-1	0	-1	(3.26)	0
v	v	c	c	-1	-1	0	(3.27)	-
v	c	v	c	0	1	1	0	-
v	c	c	c	-1	0	0	0	(3.21)
c	v	v	v	1	0	0	0	(3.20)
c	c	v	v	1	1	0	-	-
c	c	v	c	1	1	1	0	0
c	v	c	v	0	-1	-1	0	-
c	v	v	c	1	0	1	-	0
c	c	c	v	0	0	-1	(3.17)	(3.22)
c	v	c	c	0	-1	0	(3.18)	0
c	c	c	c	0	0	0	0	0

Table 3.1: All possible band combinations and the Pauli repulsion factors. *A* and *B* are given in the right most columns as either a 0 for vanishing contribution, a reference or a "-" for terms which do contribute, but are not written out.

Moss *et al.*[32] present group theoretical arguments for allowing the integration domain to be reduced to the irreducible Brillouin zone upon using the substitutions

$$p_x p_x p_x p_x \to 16\left[p_x p_x p_x p_x + p_y p_y p_y p_y + p_z p_z p_z p_z\right],$$
$$p_x p_x p_y p_y \to 8\left[p_x p_x p_y p_y + p_x p_x p_z p_z + p_y p_y p_x p_x + p_y p_y p_z p_z + p_z p_z p_x p_x + p_z p_z p_y p_y\right],$$

which can be verified numerically. Similarly, we postulate that

$$p_x p_y p_y p_x \to 8\left[p_x p_y p_y p_x + p_x p_z p_z p_x + p_y p_x p_x p_y + p_y p_z p_z p_y + p_z p_x p_x p_z + p_z p_y p_y p_z\right],$$

verifying numerically. Results calculated using these simplifications and the standard LATM is shown in Fig. 3.5 in units of the material refractive index n. A $\vec{q}$-orientation was chosen so $\hat{q} = 1/\sqrt{2}(\hat{b}+\hat{d})$. It is clear that the neglect of bulk SHG in Si and Ge is a good approximation, since the contribution due to the inclusion of the optical wave vector to first order only constitutes a few percent of the bulk contribution in similar semiconductors without an inversion center (GaAs and InSb).

3.4 Depletion-electric-field-induced SHG

If a bulk material is terminated by a surface, electronic states localized in the first few atomic layers will appear. These states arise e.g. from defects in the surface or adsorbed molecules, but even on perfectly clean surfaces these localized states appear (as will be shown in next chapter). A characteristic of surface states is their appearance in the bulk band gap energy range. A consequence of this, as will be shown shortly, is the build-up of a static electric field perpendicular to the surface, breaking the bulk symmetry of the crystal, enabling non-vanishing tensor elements of $\chi^{(2)}$, which would otherwise be dipole forbidden.

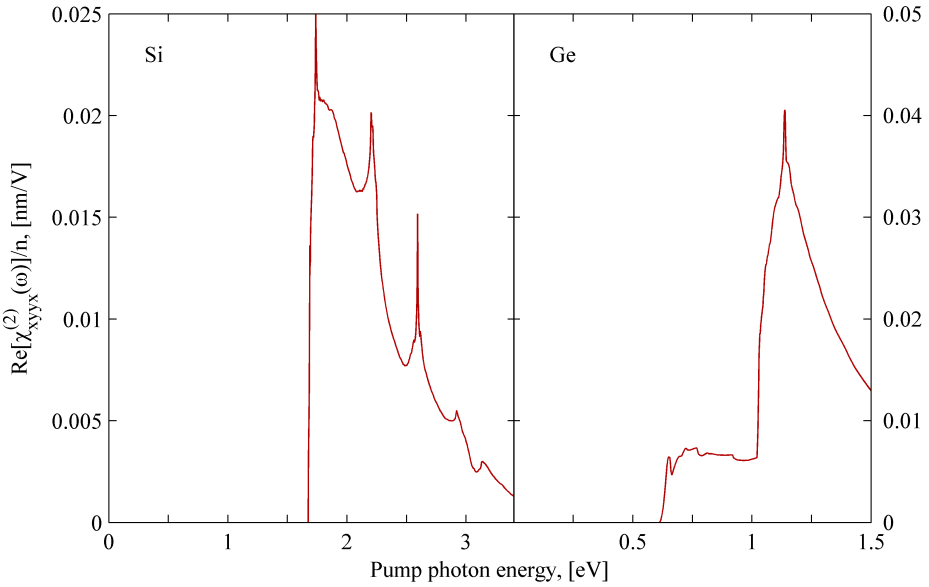


Figure 3.5: The bulk, non-local, second order susceptibility of Silicon and Germanium calculated using the standard LATM (98.304 tetrahedral sub-domains in the irreducible Brillouin zone). Band structures are calculated by the EPM using 137 plane waves as basis. Only the 2ω part of the susceptibility is plotted, why the spectrum of Ge is accurate only below 1.1eV.

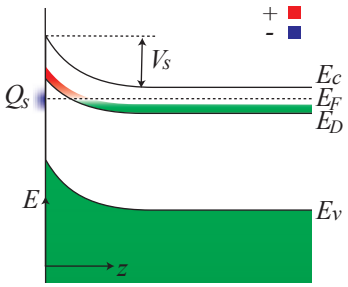


Figure 3.6: A schematic drawing of the surface depletion layer. A charge Q_s is accumulated at the surface due to surface acceptor defects. This causes a band-bending, which depopulates donor ions, causing them to contribute with a positive charge in the depletion layer, as will depletion of the intrinsic electron population. Overall charge conservation demands that the charge on the surface and the total charge of the depletion layer must balance.

In this section, the electrostatic problem of a surface depletion layer will be considered followed by an analysis of its influence on the second harmonic optical properties of zinc-blende crystals. The following exposition is adapted from the book by Lüth[35] and a lecture note on MOSFET transistors by Prof. T. G. Pedersen.

3.4.1 Analysis of electrostatics of surface depletion layer.

A semiconductor surface situated at $z = 0$ is considered. This surface is characterized by defects and surface states in the ”forbidden” region of the band gap, which act as localized electron acceptors of surface density N_s. To investigate how these states affect the sys-

tem, we shall consider an abstraction: A bulk semiconductor is suddenly cleaved, creating a surface with the appropriate surface and defect acceptor states. Right after the surface have been introduced, the electron density is organized exactly as in the bulk; this is not, however, an equilibrium configuration since the electron occupation fails to follow a Fermi-Dirac distribution function in that the surface states are unpopulated. To equilibrate, the surface acceptor states are occupied which could be described by letting the Fermi level vary in space, decreasing towards the surface in a charge conserving fashion. In this way, free electrons are "taken" from the crystal near the surface, since the distance between the conduction (and donor) band edge and the Fermi level increases, and used to occupy the surface states leaving positively charged holes in the region near the surface. The charge build-up on the surface means that an electric field perpendicular to the surface is created; this field will tend to counteract continued diffusion causing an equilibrium to be reached; a situation which corresponds to a vanishing Fermi level gradient, however, the resulting surface electric field will cause the electronic energy bands to bend "upwards" in the direction of the surface; this situation is illustrated schematically in Fig. 3.6. In the following, E_c, E_v and E_D shall denote the energy of the conduction band edge, the valence band edge and the electron donor level of the semiconductor, respectively.

The electric field gradient is related to the local charge density through Poisson's equation

$$\frac{d^2\varphi(z)}{dz^2} = -\frac{e\rho}{\varepsilon\varepsilon_0}, \tag{3.32}$$

where the boundary conditions for the electrostatic potential φ at the surface and at infinity (inside the bulk of the sample) are $\varphi(0) = \varphi_s$ and $\varphi(\infty) = 0$. In the general case, the charge density inside the semiconductor reads

$$\rho = -e(n - p + N)$$

where N is the dopant concentration, respectively negative and positive for n-doped and p-doped materials. In the bulk of the semiconductor, the electric field gradient was taken to zero, why $N = p_0 - n_0$ corresponding to charge conservation , where the subscript 0 refers to the unperturbed bulk case. In equilibrium, the electron and hole densities are governed by Fermi-Dirac statistics, which in the non-degenerate limit translates to

$$n = D_c e^{(E_F - E_c)/k_B T}, \quad p = D_v e^{(E_v - E_F)/k_B T},$$

where D_c and D_v are effective densities of states. We now introduce an explicit parameter specifying the local band bending, i.e. the deviation of the band edges from the bulk case: $-eV(z) = E_c(z) - E_c(\infty) = E_v(z) - E_v(\infty) = E_F - E_i(z)$, where $E_i(z)$ is the intrinsic Fermi level

$$E_i = \frac{1}{2}\big(E_c(z) + E_v(z)\big) - \frac{1}{2}k_B T \ln\left(\frac{D_c}{D_v}\right).$$

With this, the electron and hole densities can be rewritten

$$\begin{aligned} n &= D_c e^{(E_F - E_c(\infty))/k_B T} e^{eV(z)/k_B T} = n_0 e^{eV(z)/k_B T} = n_0 e^{v(z)}, \\ p &= D_v e^{(E_v(\infty) - E_F)/k_B T} e^{-eV(z)/k_B T} = p_0 e^{-eV(z)/k_B T} = p_0 e^{-v(z)} \end{aligned}$$

using the normalized potential $v(z) \equiv eV(z)/k_B T$. Thus, neglecting the explicit notation of the z-dependency, the charge density can be written

$$\rho = -e\left\{n_0(e^v - 1) + p_0(1 - e^{-v})\right\}.$$

Using the the charge density and the normalized potentials above, (3.32) can be written

$$\frac{d^2v}{dz^2} = \frac{e^2}{\varepsilon\varepsilon_0 k_B T}\left\{n_0(e^v-1)+p_0(1-e^{-v})\right\}.$$

It will prove convenient to define a characteristic length L

$$L \equiv \left(\frac{k_B T \varepsilon\varepsilon_0}{e^2(n_0+p_0)}\right)^{1/2},$$

allowing

$$\begin{aligned}\frac{d^2v}{dz^2} &= \frac{1}{L^2}\left(n_0(e^v-1)+p_0(1-e^{-v})\right)\frac{1}{n_0+p_0}\\ &= \frac{1}{L^2}\left(\frac{p_0-n_0}{n_0+p_0}+\frac{n_0e^v-p_0e^{-v}}{n_0+p_0}\right).\end{aligned}$$

Defining a new parameter $u \equiv (E_F - E_i(\infty))/k_BT$ allows the carrier concentrations to be rewritten: $p_0 = p_i\exp(-u)$, $n_0 = n_i\exp(u)$, where n_i and p_i are intrinsic concentrations. Upon noting that $n_ip_i = n_i^2$, this allows the above equation to be rewritten on a very simple form

$$\frac{d^2v}{dz^2} = \frac{1}{L^2}\left(\frac{\sinh(u+v)}{\cosh(u)}-\tanh u\right) \equiv \frac{f(u,v)}{L^2}.$$

Now, proceeding to take the first integral

$$\begin{aligned}\frac{d^2v}{dz^2} &= \frac{1}{2}\frac{d}{dv}\left(\frac{dv}{dz}\right)^2 = \frac{f(v,u)}{L^2}\\ \Rightarrow \frac{dv}{dz} &= \pm\sqrt{2}\frac{\sqrt{F(v,u)+C}}{L}.\end{aligned} \tag{3.33}$$

In the bulk, $v = 0$ and the electric field must vanish; hence $C = -1$. The solution is chosen on physical grounds to make the field point in the same direction as the surface dipole, i.e. if negative charge is build up on the surface, and the space charge region is positive, the minus solution is chosen to make the field point from the positive space charge region to the negative surface. The integral can be solved analytically

$$F(u,v) = \frac{\cosh(u+v)}{\cosh(u)} - \tanh(u)v.$$

One more integration is needed to find $v(z)$ with respect to z

$$\int_0^z \frac{dz}{L} = \frac{z}{L} = \pm\sqrt{2}\int_{v_s}^{v(z)}\frac{dv}{F(v,u)}$$

which will have to be solved numerically for each z to find the appropriate $v(z)$, (or, more conveniently, the other way around). At this point, v_s and u are still unknowns, both depending on the Fermi-level. u depends on the bulk properties of the crystal, and will be taken as a parameter specified by the doping of the crystal (i.e. the location of the Fermi level with respect to the intrinsic Fermi level), whereas v_s is desired in terms of the surface charge density Q_s. To determine Q_s, it is noted that charge conservation requires the space charge layer to exactly cancel the surface charge, hence

$$Q_s = -\int_0^\infty \rho(z)dz = -\int_{v_s}^0 \rho[v(z)]\left(\frac{dv(z)}{dz}\right)^{-1}dv.$$

By (3.33) this becomes

$$\begin{aligned} Q_s &= \mp\frac{eLn_i}{\sqrt{2}}\int_{v_s}^{0}\left\{e^u(e^v-1)+e^{-u}(1-e^{-v})\right\}\left\{F-1\right\}^{-1/2}dv \\ &= \mp Q_0\int_{v_s}^{0}\frac{\sinh(u+v)-\sinh(u)}{\sqrt{F(u,v)-1}}\,dv, \quad Q_0 \equiv \sqrt{2}eLn_i. \end{aligned} \tag{3.34}$$

Again, this is an integral equation which must be solved for v_s numerically given a certain Q_s. This is best in reverse, i.e. a set of Q_s is calculated from a large set of v_s-values, and $v_s(Q_s)$ can then be interpolated from the dataset. The exact value of Q_s for a given system depends on the surface state density and the Fermi-Dirac occupation statistics thereof, however to facilitate a qualitative discussion we shall make the assumption that each surface atom contributes a single surface state energetically positioned near the middle of the bulk band gap, and we shall neglect the possibility of two electrons with opposite spin occupying the same surface state due to Coulomb repulsion. Thus, the mean occupancy for an acceptor type surface state is by Fermi-Dirac statistics estimated to

$$\langle n_s\rangle = \frac{2}{e^{(E_i-E_F)/k_BT}+2} = \frac{2}{e^{-u}+2},$$

making the surface charge density

$$Q_s = \frac{-2eN_{sa}}{e^{-u}+2},$$

where N_{sa} is the surface density of surface acceptor states. Inserting this expressions into (3.34) and solving for N_{sa} allow v_s to be related to N_{sa} Fig. 3.7 shows the surface potential (with $V_s = v_s \times k_BT/e$) as a function of the surface acceptor state density N_{sa} for several dopant concentrations governed by the Fermi level displacement u. It is interesting to note how the surface potential display asymptotically logarithmic proportionality to the number of surface states. Thus, the surface potential saturates at some point N_{sa}, where densities several orders of magnitude larger introduce changes in the meV range for the surface potential. This means that the surface potentials at realistic surface concentrations of ~ 1 surface state pr. surface atom becomes a rather good guess, with large uncertainties in N_{sa} giving only minute changes in v_s. The dopant concentration affects the surface potential strongly, with a linear increase in v_s with increasing u in the saturation range. Fig. 3.8(a) illustrate potential profiles for a few doping levels given by u. It is noted how the potential extends several hundred microns into an intrinsic sample ($u = 0$), whereas the depletion layer is only a few hundred nanometers for heavily doped species. In Fig. 3.8(b) charge density profiles is shown, where intrinsic semiconductors $u = 0$ are seen to have an exponentially decaying charge density in the surface depletion layer as is expected for a weak depletion region governed by the intrinsic free carriers of the sample. With increasing doping, a step-like feature begins to develop. This is due to the positively charged, ionized electron donors in the strongly depleted space charge region. In this limit, a simplified Schottky approximation could have been made from the onset, taking the space charge density as a step-function extending into the sample, yielding a quadratic potential function, which is seen to agree well with the behavior of the potential curves.

3.4.2 Depletion-electric-field-induced SHG

The electric field in the space charge region breaks the symmetry of the bulk crystal, giving rise to additional SHG. To establish a simple model for this, we shall consider the space

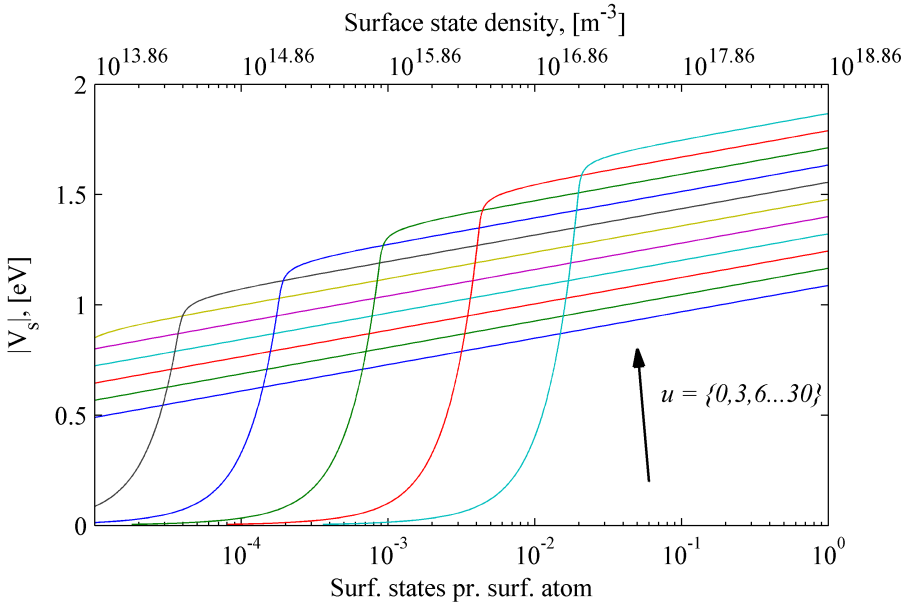
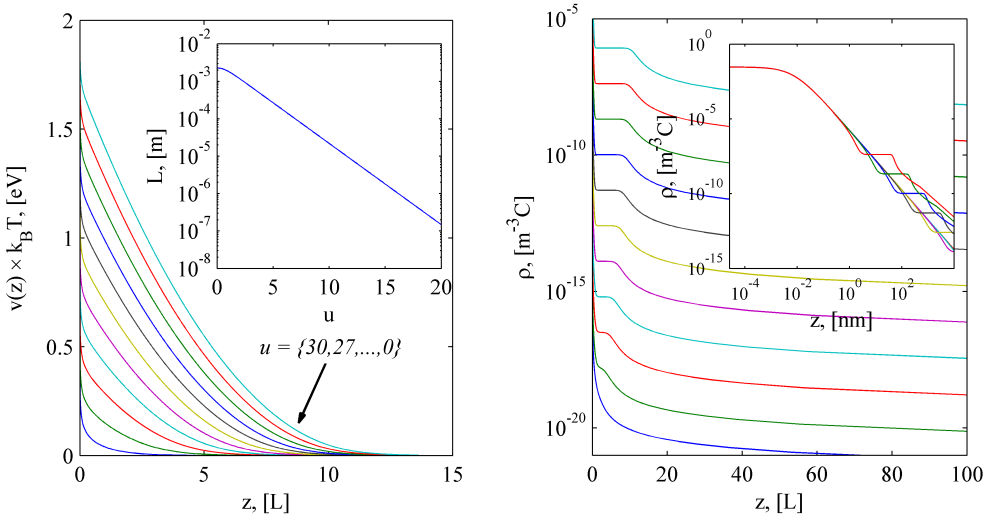


Figure 3.7: Surface potential versus surface state density for various n-doped samples (varying u). Parameters suitable for GaAs was applied: $n_i = 1.79 \times 10^{12}\mathrm{m}^{-3}$, $\varepsilon = 13.1$



(a) (b)

Figure 3.8: (a) Potential profiles for various doping levels u. It is noted how L represents an effective length, and that all potentials approach zero for $L > 10$; the inset, however, illustrates how L varies by several orders of magnitude due to changing doping levels. With a higher doping follows a shorter screening length, as is readily apparent. (b) Charge density profile for the same u as (a); top-most cyan line represents $u = 30$, the next $u = 27$ and so on. Please note the different length scales (nm vs. L) on the inset relative to the main image. Data for GaAs was applied, as in Fig. 3.7.

charge field a small perturbation to atomic potentials, and thus apply time-independent perturbation theory to first order and apply a linear approximation. The surface depletion field shall be described by a simplified potential

$$V_d = e\varphi_s \left(\frac{z}{W} - 1\right) \Theta(W - z),$$

created by surface acceptor states ($\varphi_s < 0$) with Θ denoting a Heaviside step function. It is realized that V_d fails to provide a continuous electric field at $z = W$, however, this is justified by simply considering the nonlinear polarization in the region $W - dz > z > 0$, hence excluding the z-interval $[W - dz; W]$ which is taken to contain a smooth transition of the depletion field to zero. W can for near-intrinsic samples be considered macroscopic in scale, much larger than both wavelengths and unit cells, as is readily apparent from last section. Non-degenerate linear perturbation theory dictates that a state $|\alpha\vec{k}\rangle$ should be replaced by the perturbed state upon introduction of a small perturbing potential

$$|\alpha\vec{k}\rangle \to |\alpha\vec{k}\rangle + \sum_{\beta\vec{k}' \neq \alpha\vec{k}} \frac{|\beta\vec{k}'\rangle\langle\beta\vec{k}'|V_d|\alpha\vec{k}\rangle}{E_{\alpha\vec{k}} - E_{\beta\vec{k}'}}. \tag{3.35}$$

The perturbing matrix element

$$\langle\beta\vec{k}'|V_d|\alpha\vec{k}\rangle = e\varphi_s \left(\frac{\langle\beta\vec{k}'|z|\alpha\vec{k}\rangle}{W} - \langle\beta\vec{k}'|\alpha\vec{k}\rangle\right),$$

where the integration limits of the z-matrix elements only extend from $z = 0$ to $z = W$. Using the orthogonality of the stationary states in the last term and applying the expression[36]

$$\langle\beta\vec{k}'|z|\alpha\vec{k}\rangle = \frac{e\hbar}{imE_{\beta\alpha}}\langle\beta\vec{k}|\hat{p}_z|\alpha\vec{k}\rangle\delta_{\vec{k},\vec{k}'}, \quad \alpha \neq \beta,$$

where $E_{\beta\alpha} = E_{\beta\vec{k}} - E_{\alpha\vec{k}}$, yields

$$|\alpha\vec{k}\rangle \to |\alpha\vec{k}\rangle + D_0 \sum_{\alpha\neq\beta} |\beta\vec{k}\rangle \frac{\langle\beta\vec{k}|\hat{p}_z|\alpha\vec{k}\rangle}{E^2_{\alpha\beta}}, \quad D_0 = \frac{i2e\varphi_s\hbar}{mW}.$$

To estimate the second harmonic polarization of the depletion layer, the perturbed states is substituted into the expression for the nonlinear susceptibility introduced in the beginning of this chapter (3.1). The discussion is restricted to the zzz tensor element, which according to Kikuchi & Tada (1980)[37] is the only independent term of a diamond lattice. This allows the simplification of the triple momentum matrix element of (3.1) $P_{vcl} = -P_{vlc} = P_{clv}$, which can be verified by simply writing out the expressions. Thus, considering P_{vcl}

$$P_{vcl} = \mathrm{Im}\left[p^z_{vc}p^z_{cl}p^z_{lv}\right],$$

the perturbed states is inserted into the respective matrix elements, keeping only terms up to first order in D_0, yielding

$$p^z_{vc} \to p^z_{vc} + D_0 \sum_\beta \left\{\frac{1}{E^2_{v\beta}} + \frac{1}{E^2_{c\beta}}\right\} p^z_{v\beta}p^z_{\beta c},$$

$$p^z_{cl} \to p^z_{cl} + D_0 \sum_\beta \left\{\frac{1}{E^2_{c\beta}} + \frac{1}{E^2_{l\beta}}\right\} p^z_{c\beta}p^z_{\beta l},$$

$$p^z_{lv} \to p^z_{lv} + D_0 \sum_\beta \left\{\frac{1}{E^2_{l\beta}} + \frac{1}{E^2_{v\beta}}\right\} p^z_{l\beta}p^z_{\beta v}.$$

With this, P_{vcl} becomes, again only including terms up to first order in D_0

$$P_{vcl} \to \mathrm{Im}\left[p^z_{vc}p^z_{cl}p^z_{lv}\right] + \sum_{\beta}\frac{2e\hbar}{m}\frac{\varphi_s}{W}\left[\mathrm{Re}\left[p^z_{vc}p^z_{cl}p^z_{l\beta}p^z_{\beta v}\right]\left(\frac{2}{E^2_{l\beta}}+\frac{1}{E^2_{v\beta}}+\frac{1}{E^2_{cl}}\right)\right.$$
$$\left. + \mathrm{Re}\left[p^z_{v\beta}p^z_{\beta c}p^z_{cl}p^z_{lv}\right]\left(\frac{1}{E^2_{c\beta}}+\frac{1}{E^2_{v\beta}}\right)\right]$$
$$\equiv \frac{2e\hbar}{m}\frac{\varphi_s}{W}\tilde{P}_{vcl}.$$

The first triple matrix element will not contribute in zinc-blende crystals, since here the only non-vanishing terms come from the six permutations of *xyz*-elements. Thus, a depletion-electric-field-induced second harmonic susceptibility tensor can be defined according to

$$\vec{\mathcal{P}}_{2\omega} = \sum_{a,b,c}\varepsilon_0\chi^{dc(2)}_{abcz}(\omega)\mathcal{E}^b_\omega\mathcal{E}^c_\omega\mathcal{E}^z_{dc}\hat{a}, \quad \mathcal{E}^z_{dc} = \frac{\varphi_s}{W}.$$

where

$$\chi^{dc(2)\prime\prime}_{zzzz}(\omega) = \frac{e^4}{4\pi^2m^4\hbar\varepsilon_0\omega^3}\sum_{c,v,l,\gamma}{}'\int\tilde{P}_{vcl}\Big[\frac{1}{\omega-\omega_{lv}}\delta\big(2\omega-\omega_{cv}\big)$$
$$+\Big(\frac{1}{\omega+\omega_{cl}}+\frac{1}{\omega+\omega_{lv}}\Big)\delta\big(\omega-\omega_{cv}\big)\Big]d^3k, \quad l\neq v,c \qquad (3.36)$$

where the prime on the summation indicate that terms where the energy fractions of $\tilde{P}_{vcl}$ have the same indices are neglected, as demanded by the summation restriction of (3.35). Additionally, to allow numerical evaluation the energy fractions of the perturbed matrix elements are broadened according to

$$\frac{1}{E^2_{\alpha\beta}} \to \frac{1}{E^2_{\alpha\beta}+\gamma^2}.$$

Results are shown in Figs 3.9 and 3.10; in comparison with bulk SHG from GaAs and InSb, the E0 peak is seen to be intensified relative to the E1 peak; Otherwise the spectra are similar. The inversion symmetric materials, Si and Ge, are seen to display depletion-field-induced second harmonic generation on the same scale as GaAs and InSb. A further complication of this concept is that, at high pump intensities, the system is not in equilibrium, since free electron/hole pairs are continuously generated by light absorption. This means that equilibrium Fermi-Dirac statistics becomes an approximation in any real experimental setup.

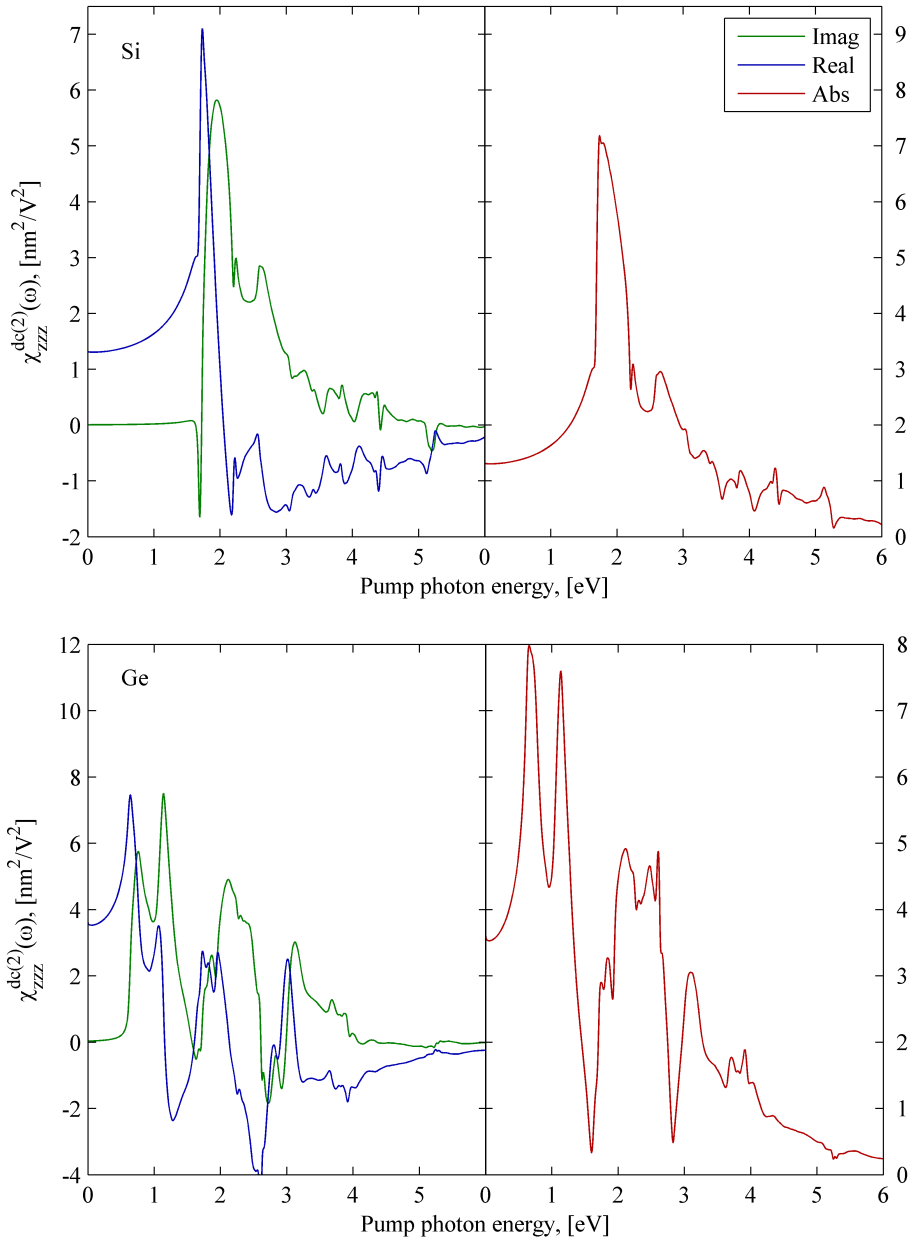


Figure 3.9: Depletion-electric-field-induced second harmonic spectra calculated from (3.36), neglecting spin-orbit interaction and using the linear-analytic tetrahedron method not generalized for second harmonic calculations. Form factors are from Tab. 2.1. All spectra are calculated using 98.304 tetrahedra in the irreducible zone, energy intervals of 1meV and a basis of 137 plane waves.

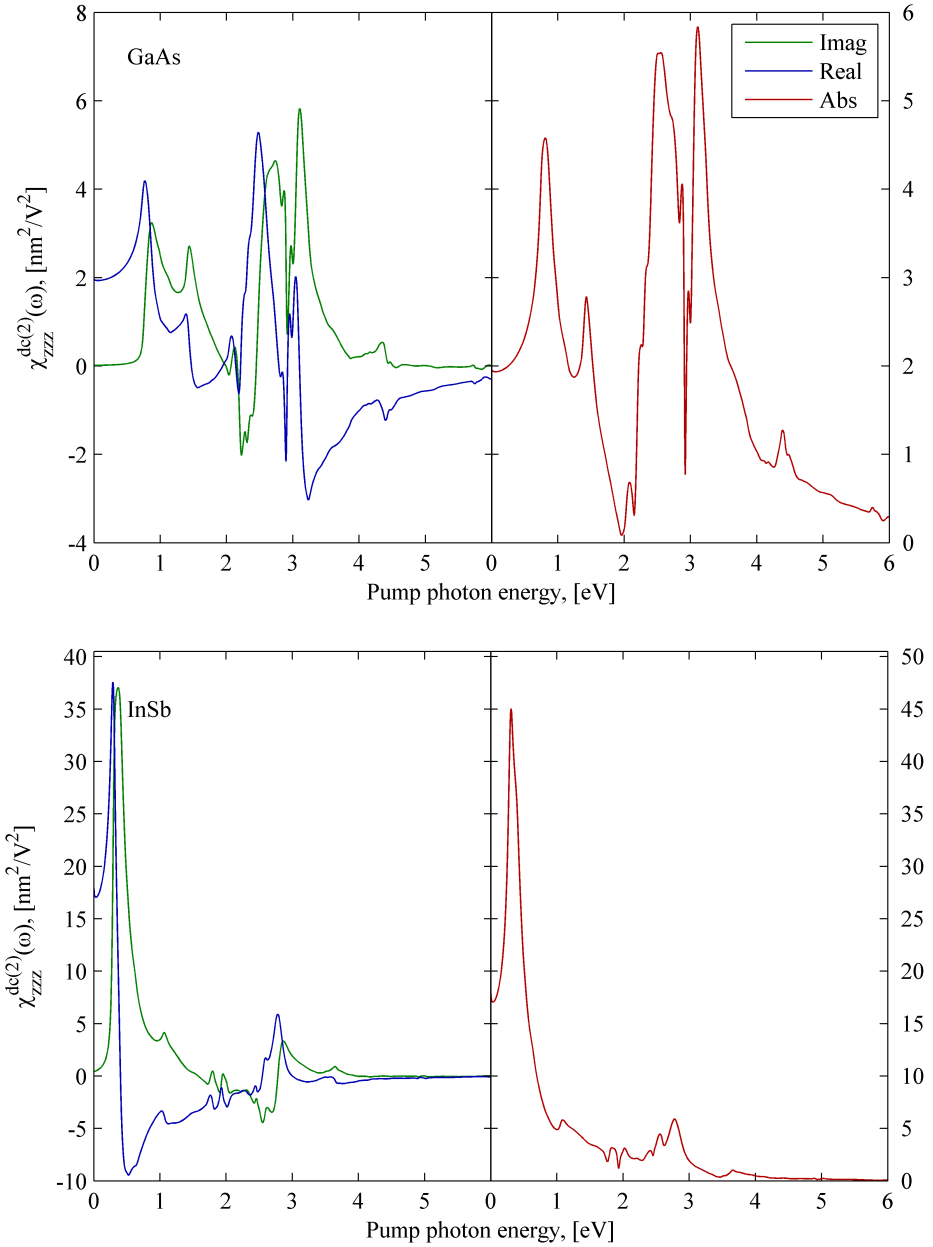


Figure 3.10: Same as Fig. 3.9, but for GaAs and InSb.

4 2D Semiconductors

So far we have been concerned with three-dimensional, bulk semiconductors. In this chapter we shall consider what happens when these are made very thin in one dimension. Additionally, semiconductor slabs shall be applied as a model for the surface region of bulk semiconductors. A plane wave pseudopotential method will still be used, although several problems will have to be addressed; e.g. the description of a finite system using a delocalised basis of sine waves proves a problem. In the method applied here, we shall cheat and introduce a fake periodicity by considering thin-films stacked into a superlattice of alternating semiconductor and vacuum layers. If the vacuum layers become thick enough to decouple the bound electronic states of the semiconductor films, their behavior should mimic that of a single film provided correct normalization is applied. To illustrate this concept on a very simple system, a superlattice defined by the potential

$$\begin{aligned} V(\vec{r}) &= V(x,y) + V(z) \\ V(z) &= \begin{cases} 0, & -w/2 + pd \le z \le w/2 + pd, \\ -V, & \text{otherwise} \end{cases} \end{aligned} \tag{4.1}$$

is considered. Here $V > 0$ and could be the ionization potential of silicon, hence providing a (too) simple model of the coupling between silicon films. w is the film thickness, d is the superlattice constant spanning one film and one spacer layer and p is an integer. To solve this model, we consider the one-electron Schrödinger equation

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r}) \right] \psi_s(\vec{r}) = E_s \psi(\vec{r})$$

and separate it into parallel (xy) and perpendicular (z) components $\psi_s(\vec{r}) = \varphi_l(x,y)\varphi_m(z)$ with $\varphi_m(z)$ satisfying

$$\left[-\frac{\hbar^2}{2m}\frac{\partial}{\partial z} + V(z) \right] \varphi_m(z) = E_m^z \varphi_m(z).$$

A plane wave basis is applied

$$\varphi(z) = \sum_n f_n(k_z) e^{i(k_z + 2\pi n/d)z}$$

and the potential is Fourier transformed

$$V(z) = \sum_n V_n e^{i2\pi n/d}, \quad V_n = -\frac{V}{\pi n} \sin\frac{\pi n w}{d}, \tag{4.2}$$

making the Hamilton matrix elements

$$H_{i,j} = \frac{\hbar^2 (k_z + 2\pi n/d)^2}{2m}\delta_{i,j} - \frac{V}{(i-j)\pi} \sin\frac{2\pi(i-j)}{d},$$

to which energy eigenvalues and -vectors are to be found*. A few examples using $V = 4.9\text{eV}$ (the work function of Si) and the electronic mass† is shown in Fig. 4.1. It is clearly seen

*It is noted that this type of Kronig-Penny model can be solved semi-analytically, however to emphasize the analogy to the full pseudopotential calculation, the same numerical method is used.

†A better model would be to substitute the mass with the effective mass. The effective mass would then be spatially dependent, e.g. the [111] element of the effective mass tensor of silicon inside the wells and the fundamental electronic mass outside.

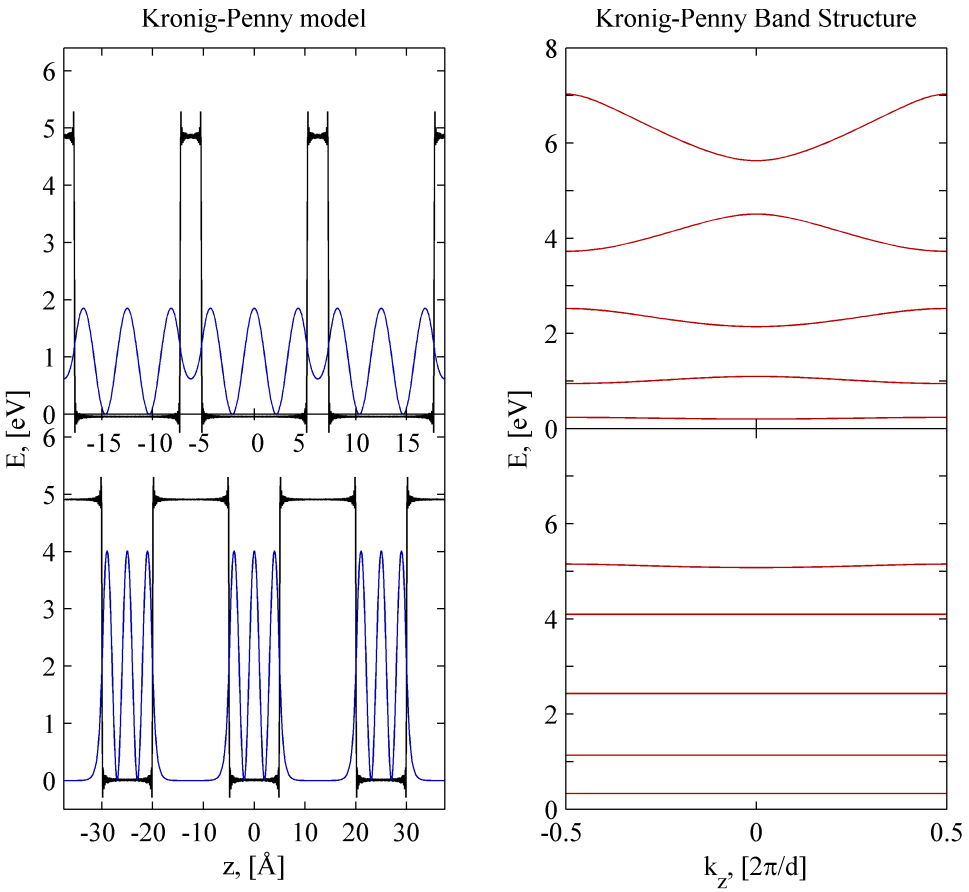


Figure 4.1: Kronig-Penny model. The Fourier transformed potential (black lines) is shown to allow direct evaluation of the quality of the basis (200 plane waves). The blue lines indicate the absolute square of the wave functions of the third band for $k_z = 0$. Red lines display the band structure of the Kronig-Penny model, where it is clearly seen that decoupling the quantum wells in real space means losing dispersion in reciprocal space.

how small spacer layers allow coupling between the respective wells; this corresponds to dispersive bands in the band structure of Fig. 4.1. Decoupled states display no dispersion, and become true quantum-well states confined to the respective layers. Additionally, momentum matrix elements of superlatices with large vacuum layers becomes independent of k_z[38]. Moving beyond the one-electron model, the Hartree and exchange-correlation potentials should be taken into account yielding a self-consistent electronic potential. The treatment of this shall be neglected in the following, however.

In a realistic thin-film, the atomic potentials should be taken into account. This is done by introducing a pseudopotential, as was first done by Schlüter *et al.* (1975)[7], analogously to the bulk case

$$V_{ps}(\vec{r}) = \sum_{\vec{G}} V_{\vec{G}} e^{i\vec{G}\cdot\vec{r}}, \quad V_{\vec{G}} = \sum_{t} S_{\vec{G}}^{(t)} v_{\vec{G}}^{(t)}, \tag{4.3}$$

where t annotates atom type, $v_{\vec{G}}^{(t)}$ is the Fourier transform of the atomic potential normalized

to the number of atoms pr. unit cell volume and $S_{\vec{G}}^{(t)}$ is the structure factor given by

$$S_{\vec{G}}^{(t)} = \frac{1}{N_t} \sum_{j}^{N_t} e^{-i\vec{G}\cdot\vec{\tau}_{t,j}}. \tag{4.4}$$

Here $\vec{\tau}_{t,j}$ represent the basis vector of the j'th basis atom of type t, whereas N_t is the number of atoms of type t in the unit cell. Using $V_{\vec{G}}^{(t)}$ in (2.4) allows the solution of the electronic eigenvalue problem using a superlattice unit cell including vacuum layers by simply specifying the atomic positions through $S_{\vec{G}}^{(t)}$ and the atomic characteristics through $v_{\vec{G}}^{(t)}$.

To solve the Schrödinger equation (2.4), the entire Fourier spectrum of the potential $v_{\vec{G}}^{(t)}$ is needed since, as the film supercell expands, the reciprocal lattice vector describing the supercell periodicity collapses, as in the simple case above for the Kronig-Penny potential (4.2). Wang & Zunger[18] provide a continuous Fourier spectrum for the Si atomic pseudopotential fitted to both empirically determined form factors and DFT-LDA calculations of the charge density of bulk silicon:

$$v_{\vec{G}} = v_G = a_1 \frac{G^2 - a_2}{a_3 \exp(a_4 G^2) - 1} \text{eV}, \tag{4.5}$$

with parameters given in Tab. 2.1, G denoting the length of $\vec{G}$ and wave vectors in units of $2\pi/a$. This pseudopotential yield correct bulk band structures, and produce the correct ionization potential for silicon. A similar pseudopotential was used by Schlüter *et al.*[7], which was simply fitted to the bulk form factors, and thus without any consideration of the electron density, as the input to a self-consistent calculation

$$v_G = a_1 \frac{G^2 - a_2}{\exp\left[a_3(G^2 - a_4)\right] + 1} \text{eV}. \tag{4.6}$$

It was shown that near the surface, the self-consistent potential diverged from the empirical pseudopotential yielding larger values, whereas inside the "bulk" of the film, the two potentials agreed to within a few meV. One of the reasons for this is that v_G is of no consequence in bulk crystals for small G, with the smallest wave vector (apart from zero) having a length of $\sqrt{3} \times 2\pi/a$. Hence, this part of the spectrum is a guess. Furthermore, the movement of a single electron in a many-electron cloud inside the bulk of the crystal is described by the effective potential of the electron due to a charge density of all electrons (the Hartree potential) plus some exchange and correlation potentials including the removal of the electron self-energy in addition to the ionic potential. In the EPM, all these effects are included in an empirical, effective pseudopotential. If an electron is taken out of the film into vacuum, the electron is effectively missing from the crystal, hence reducing the screening of the ions; this should increase the potential barrier at the film surface. This effect is not taken into account in an empirical pseudopotential calculated for the bulk, but is included in a self-consistent model. To allow the description of this additional (relative to the bulk) potential in vacuum, an additional term is introduced to the potential, which in real space is written

$$V_s(z) = \frac{V_s^0}{2} \left\{ \text{erf}\, \alpha(-z - \Delta) + \text{erf}\, \alpha(z - w - \Delta) \right\},$$

where w is the film thickness, V_s^0 is the additional potential difference between bulk and vacuum, α is a broadening factor, and Δ is a parameter allowing adjustment of the position of the barriers. To apply this potential in the EPM, $V_s(z)$ must be introduced to Fourier space

$$V_s(z) = \sum_{\vec{G}_z} V_{G_z}^s e^{i\vec{G}_z z}, \quad V_{G_z}^s = \frac{1}{c_z} \int_0^{c_z} V_s(z) e^{-i\vec{G}_z z} dz,$$

where c_z annotates the z-component (with the z-direction still defined perpendicular to the film) of the superlattice vector. This integral is solved numerically, allowing introduction to the EPM by adding it to the total electronic potential used in (2.4), hence

$$V_{\vec{G}} = V^f_{\vec{G}} + V^s_{\vec{G}},$$

where $V^f_{\vec{G}}$ is the "film" potential of (4.3).

4.1 Fitting pseudopotential

In this section, the method of finding the parameters discussed in the previous section(a_1, a_2, a_3, a_4, Δ and α), specifying the full Fourier spectrum of the pseudopotential is presented. Considering first Ge, the starting point will be a fit of (4.6) to the three bulk form factors given by Cohen & Bergstresser[4] used previously, and the requirement

$$\lim_{G\to\infty} V^f_{\vec{G}} = 0.$$

Bulk band structures produced using the resulting potential, including all Fourier components (i.e. including form factors for $|\vec{G}|^2 > 11$) results in a band structure slightly different from the one produced using just three form factors. Thus, the parameters is used as an input to a numerical optimization procedure, where the the variation in band energy differences between the three-form factor model and the full-spectrum model is minimized. Annotating the energy of the n'th band, at high symmetry point P, E^P_n and denoting the full-spectrum model with a prime and the three-form-factor model without, the problem consists of solving the following non-linear equations simultaneously

$$\begin{aligned}
\left(E^\Gamma_5 - E^\Gamma_4\right) - \left(E^{\Gamma\prime}_5 - E^{\Gamma\prime}_4\right) &= 0 \\
\left(E^\Gamma_4 - E^\Gamma_1\right) - \left(E^{\Gamma\prime}_4 - E^{\Gamma\prime}_1\right) &= 0 \\
\left(E^\Gamma_6 - E^\Gamma_5\right) - \left(E^{\Gamma\prime}_6 - E^{\Gamma\prime}_5\right) &= 0 \\
\left(E^L_5 - E^\Gamma_4\right) - \left(E^{L\prime}_5 - E^{\Gamma\prime}_4\right) &= 0.
\end{aligned}$$

The Matlab function fsolve is particularly useful in this context, allowing a fit of the above energy differences to within $\sim$0.1meV; it was verified that all other band energy differences in the energy region of interest fitted to within 1meV. Fourier spectra of the form factors are shown in Fig. 4.2(b) together with the form factors for Ge by Cohen *et al.* (1966)[4]. Considering the resulting potential for a thin-film plotted in real space (blue line of Fig. 4.2(c)), it is clear at first glance that the fitted potential does not provide an accurate confinement. This was the goal of introducing V^s, why V^0_s must be adjusted to give the correct ionization energy Φ, defined as the energy difference between vacuum and the valence band maximum (VBM)

$$\Phi = V_{\text{VBM}} - V_{\text{vac}}. \tag{4.7}$$

For bulk crystals, the VBM is given from the band structure, whereas the average value of the pseudopotential is simply

$$V^b_{avg} = V_{G=0} = v^S_0 = \frac{1}{2}\left(v^{(1)}_0 + v^{(2)}_0\right),$$

where zinc-blende crystals was assumed in the last equality, and considering a Ge crystal

$$V^b_{avg} = v_0.$$

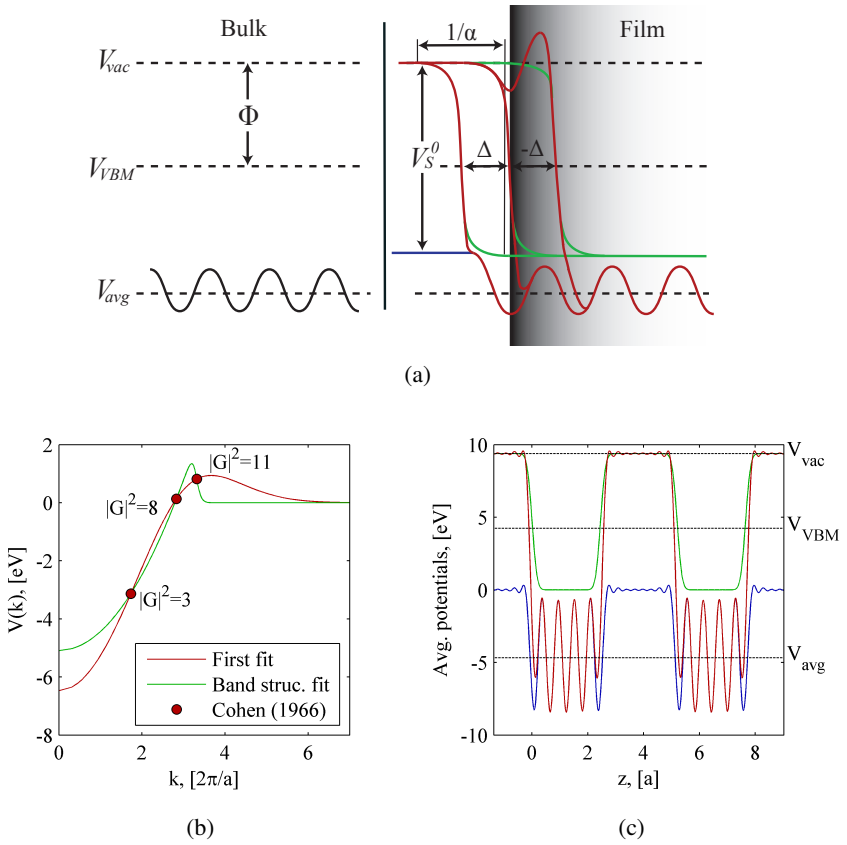


Figure 4.2: (a) Schematic representation of the pseudopotentials used. A few choices of Δ $(-\Delta, 0, \Delta)$ is shown, blue, green and red lines denoting the uncorrected, the surface correction potential and the corrected pseudopotential. (b) The continuous Fourier spectra of the form factors used to model Ge, showing an initial fit (red line) to the data points given by Cohen & Bergestresser[4] (dots). The subsequent fit to band structure energy differences is also shown (green). (c) The uncorrected (blue) and corrected (red) pseudopotentials averaged over a unit cell in the *xy*-plane for a 5 bilayer Ge/5 bilayer vacuum supercell, showing the adjustment due to the surface correction potential (green).

In the thin-film, an averaged "bulk" pseudopotential V^f_{avg} is defined as the pseudopotential averaged in a section of thin film where no contributions from the surface is apparant. The differences $V^B_{vac} - V^B_{avg}$ and $V^f_{vac} - V^f_{avg}$ are now taken to be equal

$$V^B_{vac} - V^B_{avg} = V_{vac} - V_{avg}.$$

The vacuum level in the film must by definition be $V^f_{vac} = V^f_{max} + V^0_S$, V^f_{max} denoting the maximum value of V^f. Thus, by (4.7), the vacuum energy relative to the averaged film pseudopotential is known allowing evaluation of V^0_S as

$$V^0_S = V_{VBM} + V^f_{avg} - V^B_{avg} - \Phi.$$

An illustration of these concepts relative to the film energy is shown in Fig. 4.2(c). The surface specific parameters Δ and α is fitted in order to give a satisfactory band structure, although much uncertainty is connected with these quantities, only surface state is expected to be affected. Increasing Δ moves the barrier of the surface correction potential away from the film, into vacuum, thus reducing the effects of the barriers on the bulk properties of the thin-film, which is readily seen in that the band structure below the band gap makes a better

	a_1	a_2	a_3	a_4	V_s^0, [eV]	Δ/a	$\alpha \times a$
Si*	1.1006	5.842	2.06	0.182	0	-	-
Ge	0.6528	7.7917	2.7871	10.8390	9.879	0.5	2.5

Table 4.1: Parameters used to for the surface band structure. All films have been relaxed by moving the outermost atoms 0.33Å inwards. *) Using (4.5)

fit to the projected bulk band structure. States above the confinement of the crystal potential will, however, change as do surface states. Decreasing Δ have the effect of narrowing a quantum well, hence increasing the energy difference between sub-bands. Surface states are again changed. α essentially controls the effect of the surface potential on the "bulk" states of the crystal with a small α meaning that the surface potential extends well into the thin-film, whereas the surface potential resembles a step-function with larger α. α have a marked effect on surface states, as expected when changing the a surface specific property. Since an unreconstructed surface is not found in nature, there is no experimental results directly applicable to the problem at hand. If, however, the target is to reproduce other theoretical calculations for an unreconstructed Ge surface with weakly dispersive bands in the band gap, Δ and α should be chosen so the surface field refrains from affecting the bulk of the crystal. In addition, the potential at the film boundary should be relatively smooth. This is assured using the parameters given in Tab. 4.1, reproducing the band structure for Ge shown in e.g. the book by H. Lüth[35].

4.2 [111]-surface of zinc-blende crystals

The setup of the geommetry is based on a lecture note by Prof. T. G. Pedersen. Starting from the geometry of the bulk crystal structure, it is convenient to have two lattice vectors in the plane of the film. To determine the in-plane components of these in terms of a cartesian basis (aligned with the z-axis perpendicular to the film surface), it is instructive to view the silicon crystal along the [111]-axis. This is shown in Fig. 4.3(a). Here green atoms indicate atoms in a plane parallel to the paper, closest to the reader followed by planes of blue, red and green atoms. It is apparent that the atoms in the surface plane (green atoms) form a rhombohedral lattice spanned by the unit vectors $\vec{a}_1$ and $\vec{a}_2$; hence simple plane geometry allow determination of $\vec{a}_1$ and $\vec{a}_2$

$$\vec{a}_1 = \frac{a}{\sqrt{8}}\big(\sqrt{3},1,0\big), \quad \vec{a}_2 = \frac{a}{\sqrt{8}}\big(\sqrt{3},-1,0\big)$$

where the length of the vectors is decided by considering the length of a Si-Si bond ($\sqrt{3}a/4$) and consideration of the [110] plane spanned by the z-axis and the $\vec{a}_1$-vector, Fig. 4.3(b). The colors on this figure indicate the same as on Fig. 4.3(a). $\vec{t}$ is the basis vector $\vec{t} = \sqrt{3}a/4\hat{z}$. From this figure $\vec{a}_3$ can also be derived by considering that $\vec{a}_3$ should shift a red atom into a green one in another bilayer in a primitive way

$$\vec{a}_3 = \frac{a}{\sqrt{6}}\big(1,0,\sqrt{2}\big).$$

The corresponding reciprocal lattice vectors are

$$\vec{b}_1 = \frac{2\pi}{a\sqrt{3}}\big(\sqrt{2},\sqrt{6},-1\big), \quad \vec{b}_2 = \frac{2\pi}{a\sqrt{3}}\big(-\sqrt{2},\sqrt{6},1\big), \quad \vec{b}_3 = \frac{2\pi}{a\sqrt{3}}\big(0,0,\sqrt{3}\big).$$

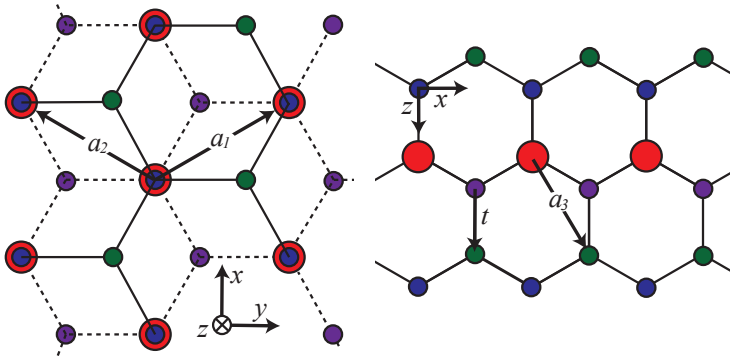


Figure 4.3: A schematic view along the [111] axis of Si crystal. Circles indicate atoms, where green atoms are closest to the viewer followed by blue, red and purple. Red atoms are made slightly larger than the blue to make them visible.

Setting up the eigenvalue problem using the lattice and reciprocal vectors above will simply give the bulk band structure of silicon, although in a coordinate system where the cartesian-axes aligned along the principal crystal axis have been rotated to align the z-axis in the [111]-direction.

It should be noted that atoms at the surface is "missing" a bond on the surface side of the interface, and is thus drawn more tightly into the film. Additional reconstruction will take silicon into a metastable (2x1) or a stable, more complicated (7x7) structure. The simplest case, however, and the starting point of this discussion is the unreconstructed (1x1) case, where relaxation is taken into account by simply displacing the outermost atoms 0.33Å parallel to the z-axis into the film.

4.2.1 Unreconstructed (1×1) surface

To consider the case of a Si thin-film with an unreconstructed [111]-surface, the basis is expanded from two atoms to $2 \times M$ atoms, where M is the number of bilayers perpendicular to the film; hence the superlattice basis is described by

$$\begin{bmatrix} \vec{\tau}_p \\ \vec{\tau}_{p+1} \end{bmatrix} = \begin{bmatrix} j\vec{a}_3 + \vec{t}, \\ (j+1)\vec{a}_3 \end{bmatrix}, \quad j \in \{0..M-1\},$$

spanning the entire thickness of the film. In addition, to introduce N vacuum layers separating the films the lattice vector $\vec{a}_3$ is substituted with a superlattice vector $\vec{c}$ defined by $\vec{c} = (M+N-1)\vec{a}_3$, hence $\vec{b}_3$ must be substituted by $\vec{d} = \vec{b}_3/(M+N-1)$ to maintain the orthonormality condition between real and reciprocal lattice vectors. A few film geometries are shown in Fig. 4.4. The atomic pseudopotential form factors $v(\vec{G})$ must be renormalized to the number of atoms pr. supercell volume. This done by setting $N_t = (M+N-1)$ in (4.4) corresponding to maintaining the volume pr. atom. A 2D band structure of 10 bilayer Si and Ge films is shown in Fig. 4.5, together with the density of states. The 3D band structure projected onto the 2D Brillouin zone is shown as the gray area, allowing easy identification of surface specific bands; for both Si and Ge two is observed in the bulk band gap, degenerate for thick slabs, whereas three more are seen to form near the K-point. It is noted that the degeneracy of the surface bands in the band gap is lifted for thin films, where each side of the film begins to couple into symmetric and anti symmetric modes. It also seen how sub-bands are formed corresponding to quantum confinement in one dimension. The

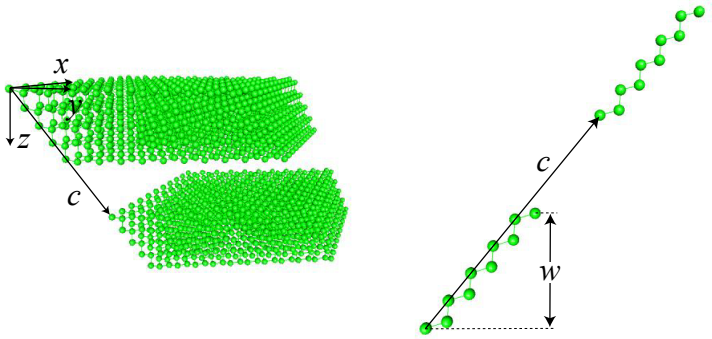


Figure 4.4: (a) Two 5 bilayer films separated by 3 bilayers of vacuum. (b) two 5 bilayer unit cells separated by 3 bilayers of vacuum.

characteristic step-wise behavior in the density of states of 2-dimensional parabolic bands are clearly observed for the first few bands up to -10eV, as are the peaks corresponding to negligible dispersion near -5eV.

To find the Fermi level, it is realized that $2 \times M$ atoms in the unit cell corresponsd to $8 \times M$ occupied states pr. unit cell, hence the density of states integrated up to the Fermi-level E_F must yield $4M$ states/unit cell, excluding spin states

$$\int_{-\infty}^{E_F} D(E)\, dE = 4M \text{ states/unit cell.}$$

This equation will have to be solved numerically for E_F, and the result is shown as the dashed lines in Fig. 4.5. It is noted that the Fermi level crosses the surface band in the bulk band gap, but not the bulk conduction bands. The physical effects of surface bands as displayed here are not observed outside vacuum, since any real crystal outside vacuum will have its surface states passivated by oxygen. In vacuum, the crystal relaxes into a (7×7) surface structure, why this unreconstructed (1×1) model only give qualitative information on surface states. It is also noted that the surface states in the band gap moves into the bulk valence band range upon neglect of surface relaxation. It is noted how the bands of Si and Ge, both bulk indirect gap semiconductors, collapses into indirect and direct gap 2D band structures, respectively, upon confinement along the [111] axis. This is due to the fact that the k-point of the conduction band minimum of Ge is located near the L-point (on the [111]-line from Γ), projected onto the Γ-point in a [111] surface band structure. Conversely, the conduction band minimum of Si is positioned near the X-point (along the [100] line), which is projected onto a point near M in the 2D Brillouin zone.

The (pseudo) charge density can be calculated from the wave functions according to

$$\rho(\vec{r}) = e \sum_{\alpha,\vec{k}} |\alpha\vec{k}\rangle\langle\alpha\vec{k}|,$$

with appropriate normalisation contained in the wave functions. The k-summation can be rewritten to an integral, and using mean-value theorem

$$\rho(\vec{r}) = e \sum_{\alpha} |\alpha\vec{k}_0\rangle\langle\alpha\vec{k}_0|,$$

where k_0 is an appropriate mean-value point which can be approximated using symmetry

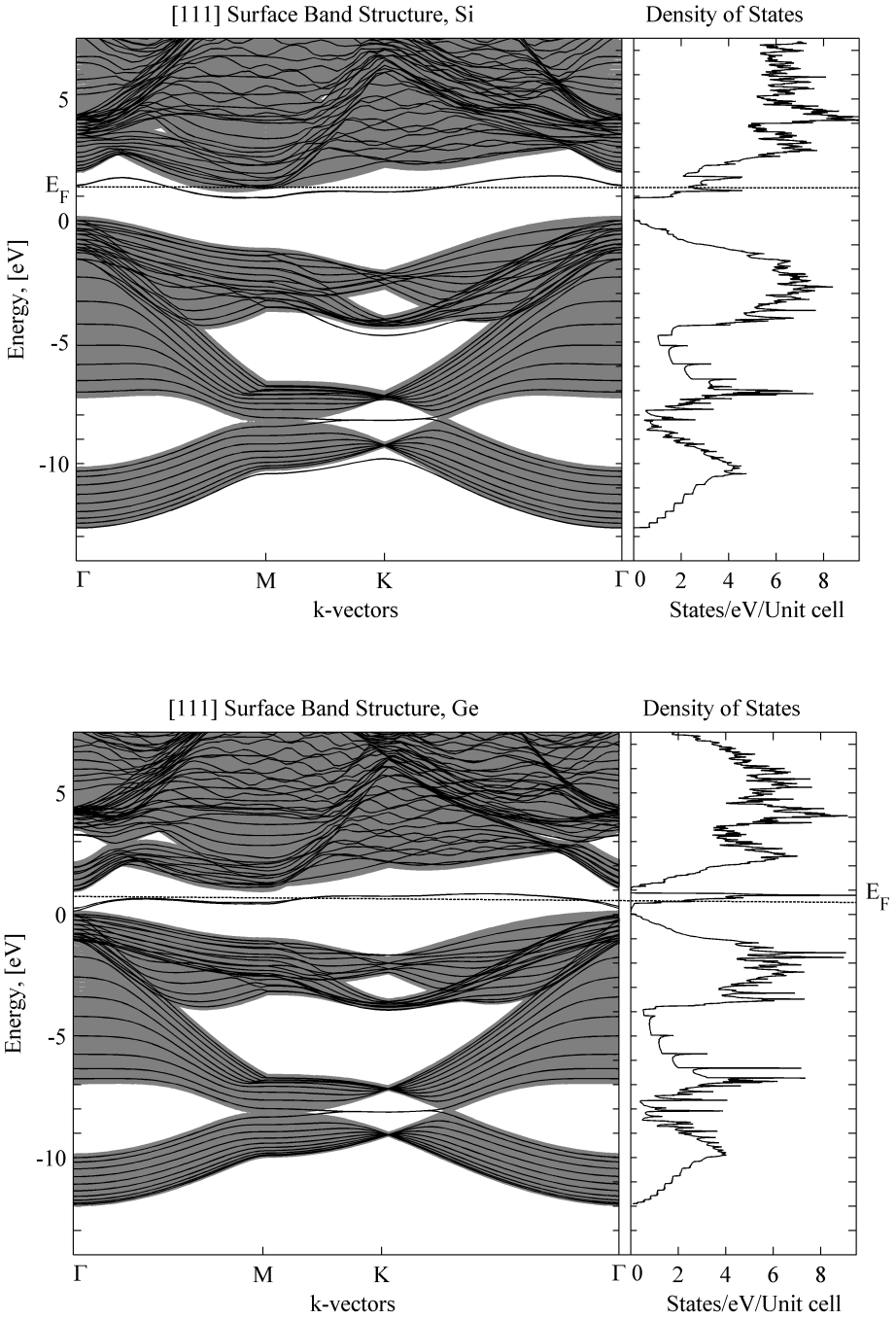


Figure 4.5: Surface band structure and density of states calculated for a 10 bilayer Si thin-film using a cut-off $|\vec{G}| \leq 5 \times 2\pi/a$. The gray patches shows the 3D band structure of silicon projected onto the 2D Brillouin zone allowing easy identification of surface specific bands. The density of states is calculated using the linear-analytic triangle method (4.11) using 1024 triangular subdomains of the irreducible Brillouin zone.

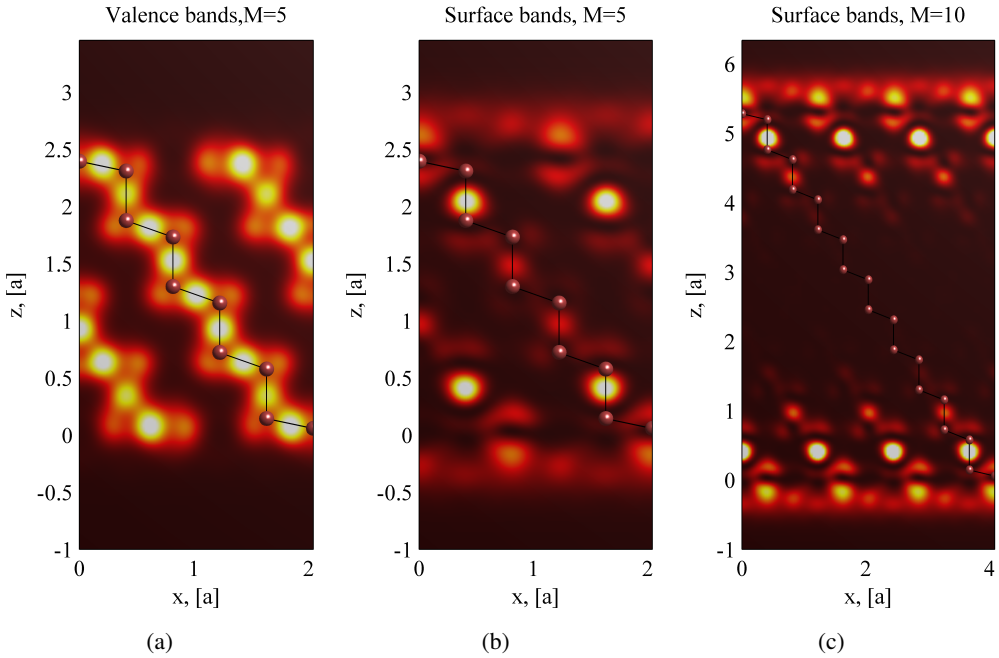


Figure 4.6: (a) Electron densities (with bright indicating large densitiy) for the 39 lowest bands in the bulk valence energy range of a 5-bilayer Ge slab, and (b) the two bands in the bulk band gap. (c) Band gap states for a thicker 10-bilayer slab. A single atomic basis is shown in all figures to indicate atomic positions.

considerations. Thus, following Cunningham (1974)[39] we set

$$\vec{k}_0 = 0.38\left(\frac{\sqrt{3}}{4}, \frac{1}{4}, 0\right)\frac{2\pi}{a},$$

and examples is shown in Fig. 4.6[‡]. Is is seen how valence states appear between atoms in bond-like configurations. Furthermore, the two surface bands are seen to display spatial localisation at the surface bilayers, with a large density at the slightly asymmetric surface back-bonds. p-like, overlapping dangeling bonds are seen to extend into vacuum, which is not expected to be an energetically favourable configuration leading to reconstruction. In a 5 bilayer film, the surface states are not fully decoupled, as is readily seen by the non-vanishing charge density in the interior of the film; this is, however, the case for the 10 bilayer film also displayed.

4.3 Linear-analytic method for 2D band structures

In the present section a 2D analogue to the 3D linear-analytic tetrahedron method shall be developed. In this method, the 2D integration domain will be divided into triangles, the δ-function arguments will be linearized inside each triangle and the necessary linearized

[‡]It is noted that this approximation is rather severe, with a better choice being the special point integration method presented for two-dimensional Brillouin zone integrals in the same article[39]. However, since the charge density is only of qualitative interest in the present, the mean-point method is used instead.

integrals will be performed analytically. Thus, considering first the density of states

$$D(\omega) = \int \delta(\omega - \omega_{cv}) d^2k = \sum_n \int_{S_n} \delta(\omega - \omega_{cv})\, d^2k,$$

where the first integral is performed over the 2-dimensional Brillouin zone and S_n refer to the n'th triagonal subdomain of the Brillouin zone. Rewriting the integral as in Sec. 3.1.2

$$\begin{aligned} D(\omega) &= \int_{\omega_1}^{\omega_3} \int_{L_n} \frac{\delta(\omega - \omega_{cv})}{|\vec{b}_n|}\, dl\, d\omega_{cv}, \quad \vec{b}_n = \nabla_k \omega_{cv} \\ &= \int \frac{L_n \delta(\omega - \omega_{cv})}{|\vec{b}_n|}\, d\omega_{cv} \end{aligned} \tag{4.8}$$

where L_n is a line of constant ω_{cv} inside the n'th triangle whereas the energy integration domain is spanned by the minimum and maximum values of ω_{cv} inside the triangle. ω_{cv} is linearized according to

$$\omega_{cv} = \vec{b}_n \cdot \left(\vec{k} - \vec{k}_1\right) + \omega_1$$

where the notation $\omega_i = \omega(\vec{k}_i)$ is used, $\vec{k}_i$ indicating triangle vertices, and the sortment $\omega_1 \leq \omega_2 \leq \omega_3$ is applied. Thus, L_n becomes a straight line between points on the triangle edges: $L_n = |\vec{k}_s - \vec{k}_t|$. These points are found by linear interpolation, considering first the case of $\omega_1 \leq \omega_{cv} \leq \omega_2$

$$\begin{aligned} \vec{k}_s &= \left(\vec{k}_2 - \vec{k}_1\right) \frac{\omega - \omega_1}{\omega_2 - \omega_1}, \\ \vec{k}_t &= \left(\vec{k}_3 - \vec{k}_1\right) \frac{\omega - \omega_1}{\omega_3 - \omega_1}, \end{aligned}$$

making L_n

$$L_n = \left| \frac{\omega - \omega_1}{\omega_2 - \omega_1} \left(\vec{k}_2 - \vec{k}_1\right) - \frac{\omega - \omega_1}{\omega_3 - \omega_1} \left(\vec{k}_3 - \vec{k}_1\right) \right|.$$

Reintroducing the notation $\omega_{mn} = \omega_m - \omega_n$ and rearranging yields

$$L_n = (\omega - \omega_1) \left| \frac{\vec{k}_{21}\omega_{31} - \vec{k}_{31}\omega_{21}}{\omega_{21}\omega_{31}} \right|. \tag{4.9}$$

$\vec{b}_n$ can be expressed in terms of the vectors

$$\vec{a}_{21} = \frac{\vec{k}_{31} \times \hat{z}}{2A}, \quad \vec{a}_{31} = -\frac{\vec{k}_{21} \times \hat{z}}{2A}$$

satisfying the orthogonality condition $\vec{k}_{\alpha 1} \cdot \vec{a}_{\beta 1} = \delta_{\alpha\beta}$. Hence it is easy to see that $\vec{b}_n$ explicitly becomes

$$\vec{b}_n = \sum_j \omega_{j1} \vec{a}_{j1}, \tag{4.10}$$

by simply checking the relation $\omega_{i1} = \vec{b}_n \cdot \vec{k}_{i1}$. Writing out the sum, $\vec{b}_n$ becomes

$$\vec{b}_n = \frac{1}{2A} \left\{ \omega_{21} \begin{bmatrix} k_{31}^y \\ -k_{31}^x \end{bmatrix} - \omega_{31} \begin{bmatrix} k_{21}^y \\ -k_{21^x} \end{bmatrix} \right\}.$$

Comparing with (4.9), and noting that

$$\vec{x} = (x_1, x_2)\,,\ \vec{y} = (y_1, y_2) : |\vec{x} - \vec{y}| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$
$$\vec{x}' = (x_2, -x_1)\,,\ \vec{y}' = (y_2, -y_1) : |\vec{x}' - \vec{y}'| = |\vec{x} - \vec{y}|,$$

L_n can be written

$$L_n = 2A_n\,(\omega_{cv} - \omega_1)\,\frac{|\vec{b}_n|}{\omega_{21}\omega_{31}}, \quad \omega_1 \leq \omega_{cv} \leq \omega_2.$$

The same formula for L_n when $\omega_2 \leq \omega_{cv} \leq \omega_3$ is used, considering the $\vec{k}_t$ to lie on the $\vec{k}_2\vec{k}_3$-line instead of the $\vec{k}_1\vec{k}_3$-line, yielding

$$L_n = -2A_n|\vec{b}_n|\frac{\omega - \omega_3}{\omega_{23}\omega_{13}}, \quad \omega_2 \leq \omega_{cv} \leq \omega_3.$$

Inserting the expressions for L_n into (4.8) and performing the integral (which amounts to substituting $\omega_{cv} \to \omega$ due to the δ-function) yields

$$D(\omega) = 2A_n \times \begin{cases} \dfrac{\omega - \omega_1}{\omega_{21}\omega_{31}} & \omega_1 \leq \omega \leq \omega_2 \\ -\dfrac{\omega - \omega_3}{\omega_{23}\omega_{13}} & \omega_2 \leq \omega_{cv} \leq \omega_3 \end{cases} \tag{4.11}$$

4.3.1 Triangular subdivision

The linear-analytic triangle method relies on the division of some integration domain into triangular subdomains. Later it will turn out, that the 2D (projected) Brillouin zone of zinc-blende crystals becomes hexagonal, consisting of 12 triangular irreducible zones. Hence, a method of dividing a triangle into subtriangles is needed; considering an arbitrary triangle of vertices $\{\vec{t}_1, \vec{t}_2, \vec{t}_3\}$ and a set of points half-way between these vertices $\vec{t}_{ij} = (\vec{t}_i + \vec{t}_j)/2$ the original triangle can be divided into four new ones

$$\{\vec{t}_1, \vec{t}_2, \vec{t}_3\} \to \begin{cases} \{\vec{t}_1, \vec{t}_{13}, \vec{t}_{12}\} \\ \{\vec{t}_3, \vec{t}_{13}, \vec{t}_{23}\} \\ \{\vec{t}_2, \vec{t}_{23}, \vec{t}_{12}\} \\ \{\vec{t}_{12}, \vec{t}_{13}, \vec{t}_{23}\} \end{cases}.$$

Repeated application of this division scheme, first on the original triangle and then on all the subsequent subtriangles, will generate a 4^{N_d} triangles, where N_d is the number of consecutive subdivisions.

4.4 Linear optical spectra of semiconductor thin-films

To calculate optical spectra of a 2D semiconductor, (2.1) is reconsidered;

$$\chi''_\omega = \frac{e^2}{4\varepsilon_0 m^2 \pi^2 \omega^2} \sum_{c,v} \int |p^z_{cv}|^2 \delta\,(E_{cv} - \hbar\omega)\ d^3k.$$

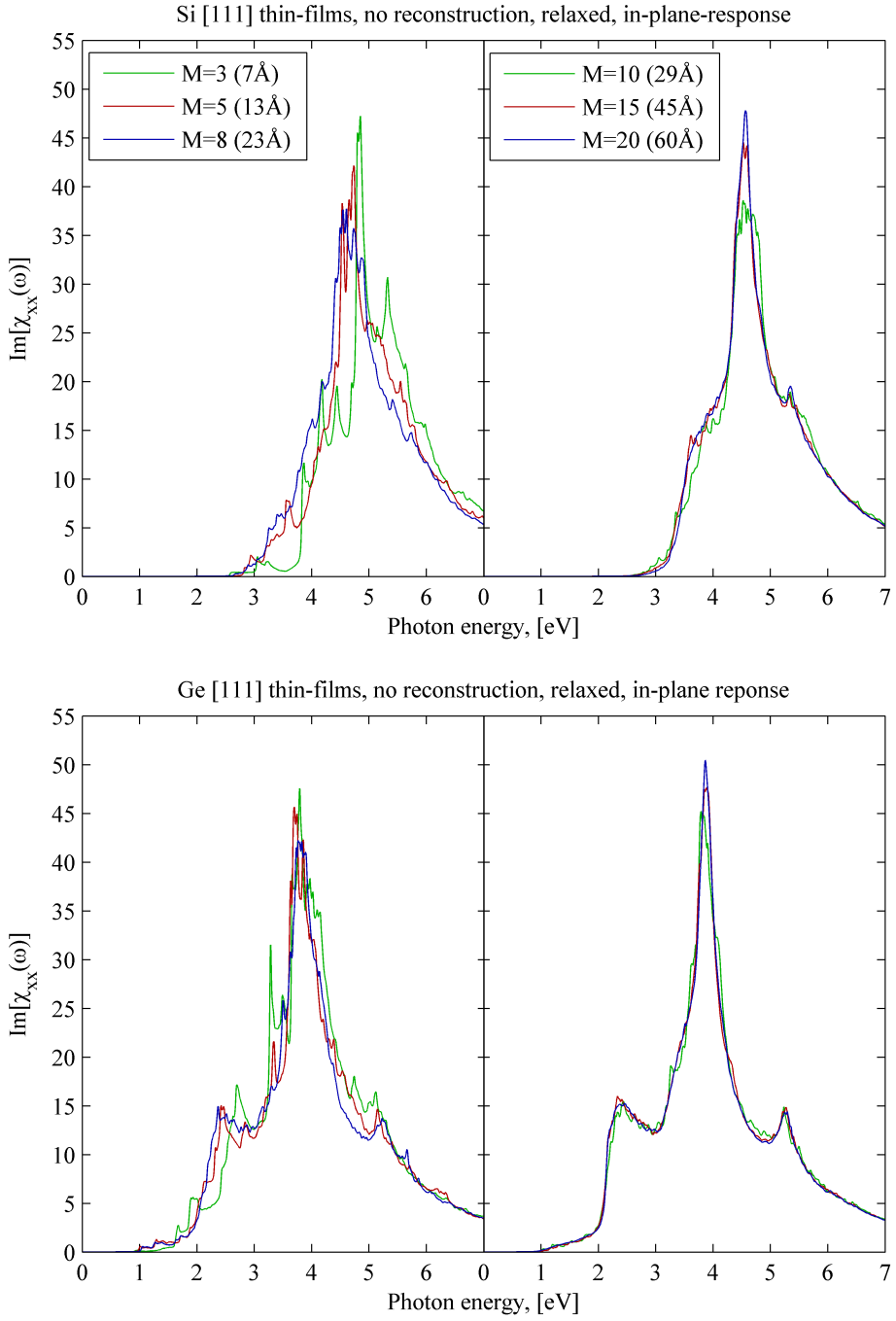


Figure 4.7: The imaginary part of the in-plane susceptibility for various Si and Ge thin-films, neglecting the metallic bands in the band gap. A cut-off $|\vec{G}| \leq 4 \times 2\pi/a$ was used for band structure calculations and 561 triangular subdomains was used for Brillouin zone integration. A Lorentzian broadening of 10meV was applied. No spin-orbit coupling is included.

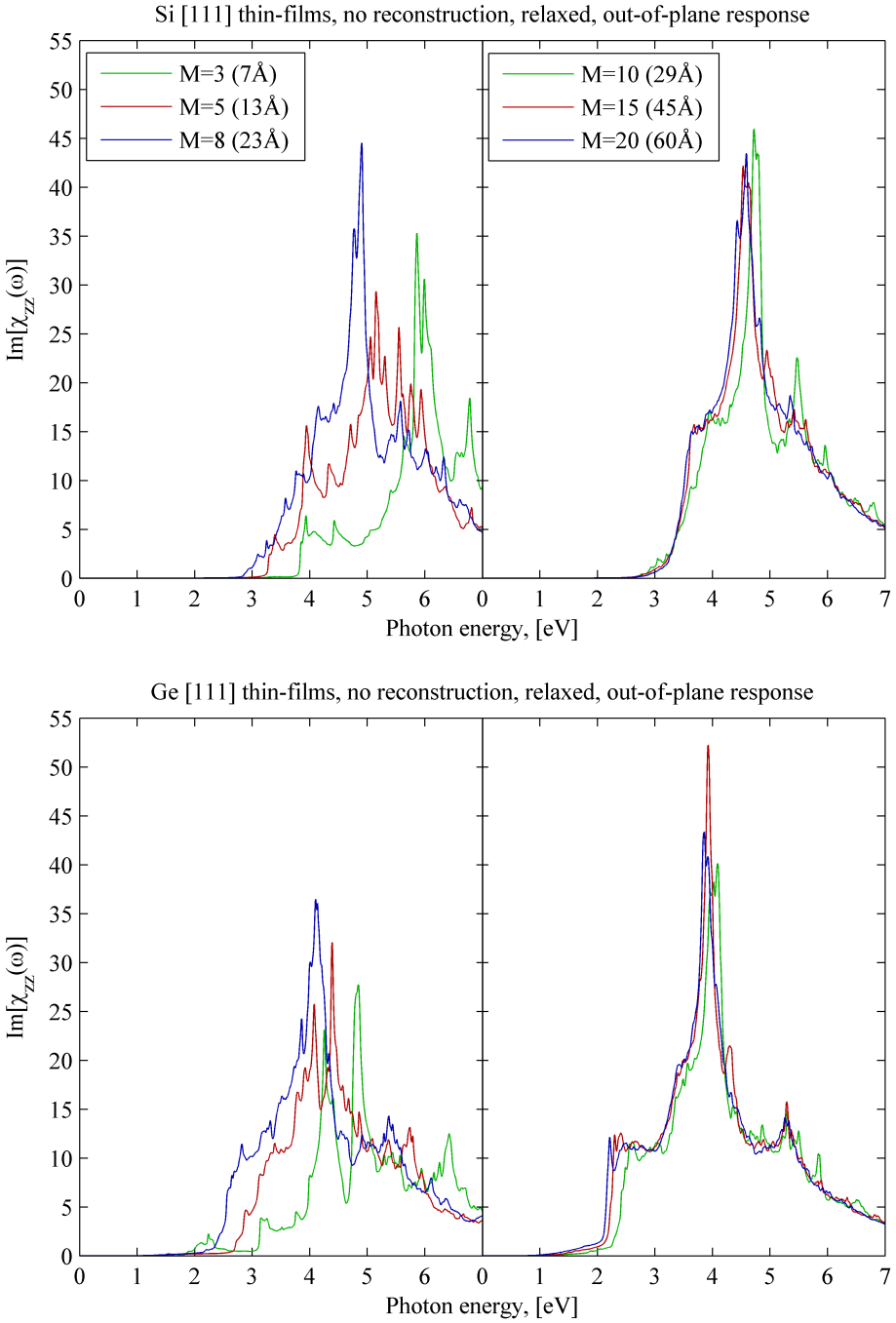


Figure 4.8: The imaginary part of the out-of-plane susceptibility for various Si and Ge thin-films, neglecting the metallic bands in the band gap. A cut-off $|\vec{G}| \leq 4 \times 2\pi/a$ was used for band structure calculations and 561 triangular subdomains was used for Brillouin zone integration. A Lorentzian broadening of 10meV was applied. No spin-orbit coupling is included.

As illustrated by Fig. 4.1 E_{cv} becomes constant in k_z upon decoupling of the films (only modes bound to the respective films are under investigation in the energy range of interest here), allowing

$$\chi''_\omega = \frac{e^2}{4\varepsilon_0 m^2 \pi^2 \omega^2} \sum_{c,v} \int \delta(E_{cv} - \hbar\omega) \int |p^z_{cv}|^2 \, dk_z d^2k. \tag{4.12}$$

A two-dimensional momentum matrix element is defined

$$\frac{2\pi}{w} |\tilde{p}^z_{cv}|^2 \equiv \int |p^z_{cv}|^2 dk_z$$

With these rearrangements, the response becomes the well known expression for a two-dimensional material

$$\chi''_\omega = \frac{e^2}{2w\varepsilon_0 m^2 \pi \omega^2} \sum_{c,v} \int |\tilde{p}^z_{cv}|^2 \delta(E_{cv} - \hbar\omega) \; d^2k. \tag{4.13}$$

The matrix element needs further consideration, however;

$$|\tilde{p}_{cv}|^2 = \frac{w}{2\pi} \int_{-\pi/c_z}^{\pi/c_z} p^z_{cv} p^z_{vc} \, dk_z. \tag{4.14}$$

Renormalising the wave functions to a single film and restricting the z-integral in the momentum matrix elements to a single unit cell still allow the volumes to cancel. The momentum matrix elements are assumed to vary slowly in the z-direction, allowing

$$|\tilde{p}_{cv}|^2 \approx \frac{w}{2\pi} \frac{2\pi}{c_z} p^z_{cv} p^z_{vc} = \frac{w}{c_z} |p^z_{cv}(\vec{k}_\parallel, k_z = 0)|^2. \tag{4.15}$$

where the momentum matrix elements are evaluated at $k_z = 0$ (changing this yielded negligible changes to the spectrum). Upon reducing the dimensions of the crystal, the cubic symmetry causing the linear response tensor to be isotropic is broken. Hence, the case of out-of-plane and in-plane polarization (the zz and the $xx = yy$ elements of the of the linear response tensor, respectively) must now be considered separately. It is noted that the surface states present in the band gap display metallic behavior since it is divided by the Fermi level, hence intraband contributions from this band cannot be ignored. Ignoring this fact for the moment, simply skipping it in the summation of (4.13) and using the linear-analytic triangle method including matrix elements by considering them constant the average of their values at the triangle vertices, the spectra of Fig. 4.7 for the in-plane response and 4.8 for the out-of-plane response can be calculated. It is clear that thin films a few atoms thick display quantum confinement effects, with peaks in the spectrum corresponding to subband transitions. In addition, absorption below the direct band gap is seen for Si, whose bulk band structure projected onto the [111] surface collapses into a direct-gap material (not considering suface bands). With thicker films, this spectrum begins to resemble bulk spectra with no confinement effects due the sub-bands resembling a continuum, and little absorption below the direct band gap. In addition, the out-of-plane response is seen to change dramatically with film thickness for thin films, whereas the results converge towards the in-plane spectrum with increasing film thickness, as is expected when going towards the isotropic bulk case. The metallic band is expected to contribute with a Drude-like behavior, and interband transitions between this and the conduction bands will give rise to absorption below 1eV. However; as noted before, outside vacuum the surface bands are expected to be passivated removing them from the band gap.

4.4.1 Effects of surface bands

In this subsection, the effects of including the surface bands in calculations of the linear optical properties will be discussed. Thus, the surface bands must be included in the summation over both occupied (v) and unoccupied (c) bands in (4.13). For interband terms ($c \neq v$) with v in the valence band range and c a band gap state, a check is made to see if E_c lies above the Fermi-level before including the term. For v in the bulk band gap, and c in the bulk conduction band range, it is similarily checked that $E_v < E_F$. Numerically, the two bands in the band gap were seen not to couple through the momentum matrix element, thus preventing singularities in the optical spectrum arising from degenerate bands. This is readily understood for thick films, where the states localised on each surface have vanishing overlap and hence vanishing momentum matrix element. For thinner films, overlap is found, but the momentum matrix elements still vanishes, although this is less important with respect to singular terms in the optical spectrum, since the two surface bands splits with thinner films. Now, intraband terms $(v = c)$ must also be taken into acount; this gives rise to a metallic Drude contribution to the optical susceptibility

$$\chi_{\omega}^{(D)\prime\prime} = \frac{\omega_p^2}{\omega + i\Gamma},$$

where the microscopic plasma frequency ω_p is given by[36]

$$\omega_p = \frac{2e^2}{\hbar^2 \varepsilon_0 \Omega} \sum_{k,t} \left(\frac{\partial E_t}{\partial k_x} \right)^2 \delta(E_t - E_F),$$

and the partial derivative with respect to k_x is the only independent part of the k-space gradient of the energy bands, since the k_z and k_y derivatives yields zero (see Fig. 4.1) and the same contribution as k_x upon performing the k-space summation, respectively. Using the substitution

$$\left(\frac{\partial E_t}{\partial x} \right)^2 \rightarrow 6 \left(\frac{\partial E_t}{\partial x} \right)^2 + 6 \left(\frac{\partial E_t}{\partial y} \right)^2$$

allows the integration domain to be restricted to the (1/12) irreducible Brillouin zone, which was verified numerically. The plasma frequency contains only terms pertaining to the unperturbed crystal, why it can be calculated for the surface bands once and for all using the linear-analytic trangle method, using the Fermi-energy E_F instead of the photon energy $\hbar\omega$ and the appropriate k-space gradients instead of the momentum matrix elements. Since the k-space gradient of the energy bands are used in the linear triangle method, these can be found from the three triangle vertices using (4.10). Plasma frequencies calculated from the 2D EPM band structures is shown in Fig. 4.9. They are quite large for thin films, scaling as $1/M$. This is due to its dependency on the charge density, hence, a thin film will almost exclusively consist of surface states while an increasing numer of bulk states is found with increasing M while maintaining the number of surface states pr. unit cell.

4.5 Surface second harmonic generation.

In chapter 3, it was mentioned how second harmonic generation in crystals of inversion symmetry was forbidden in the dipole approximation. However, this restriction was shown to be lifted whenever the inversion symmetric crystal potential was perturbed by some external field, be it a spatially varying optical field or a surface depletion field. Near the surface

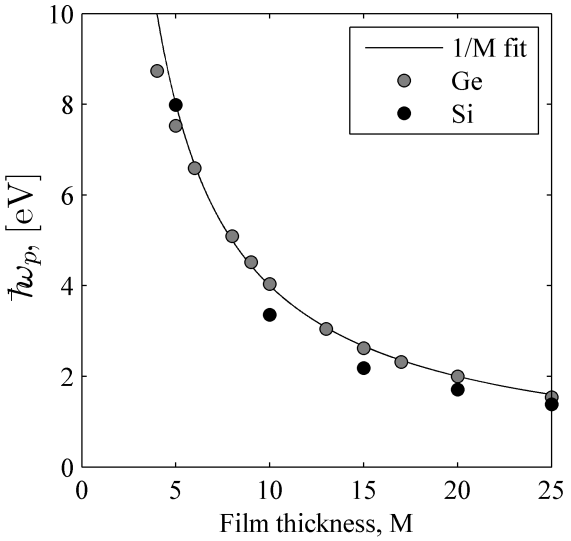


Figure 4.9: Plasma frequencies for Ge and Si thin-fimls.

of any crystal, the bulk symmetry is broken and thus a second harmonic signal from this region is expected. If the second harmonic response tensor of a surface was modeled using the thin semiconductor slabs of the last few sections, the result would still be zero due to the existence of an inversion center in the superlattice. Thus, SHG from a semiconductor surface must be modelled by taking the semi-infinity of the bulk crystal into acount. Here we shall follow the method presented by L. Reining *et al.* (1994)[40] and M.C. Downer *et al.* (2001)[41]; these two sources both introduce the idea of truncating the spatial integral of momentum matrix elements to only include a single surface of a semiconductor slab, hence preventing the two surfaces from interfering destructively. Reining introduce this method on the grounds of taking the fundamental field to zero at one of the interfaces, with most of the field variation occuring in the middle of the sample, i.e. in the "bulk" of the slab where no SHG is expected. Hence, the electric field in the perturbing Hamiltonian is substituted by the electric field times some surface specific function, why two of the three momentum matrix elments needed in calculations of second harmonic spectra pertaining to the perturbation. Downer *et al.* applies a similar argument, although they calculate the polarisation response of the surface layer and hence only the single matrix element pertaining to the response variable sought is modified to exclude one of the surfaces. Here, the latter approach will be applied; both because it is easier to only modify a single matrix element, but also because the physical rationale appears appealing.

Thus, the momentum operator with the superscript a in (3.1) is modified according to

$$\hat{p}_a \rightarrow \tilde{p}_a = \frac{1}{2}\left[S(z)\hat{p}_z + \hat{p}_z S(z)\right], \tag{4.16}$$

where S_z is a surface specific function which is unity at one surface while vanishing at the other. Here $S(z)$ will be taken as a step function

$$S(z) = \begin{cases} 1, & -z_1 + nc_z \leq z \leq z_2 + nc_z \\ 0, & \text{otherwise} \end{cases},$$

where n is an integer. Considering now an explicit matrix element of $\tilde{p}$

$$\tilde{p}_{\alpha\beta} = \frac{1}{2}\langle\alpha\vec{k}|\left[S(z)\hat{p}_z + \hat{p}_z S(z)\right]|\beta\vec{k}\rangle,$$

which, using the resolution of the identity operator $\hat{I}$

$$\hat{I} = \sum_{\gamma} |\gamma\vec{k}\rangle\langle\gamma\vec{k}|,$$

can be rewritten

$$\begin{aligned}\tilde{p}_{\alpha\beta} &= \frac{1}{2}\sum_{\gamma}\left[\langle\alpha\vec{k}|S(z)|\gamma\vec{k}\rangle\langle\gamma\vec{k}|\hat{p}_z|\beta\vec{k}\rangle + \langle\alpha\vec{k}|\hat{p}_z|\gamma\vec{k}\rangle\langle\gamma\vec{k}|S(z)|\beta\vec{k}\rangle\right] \\ &= \frac{1}{2}\sum_{\gamma}\left[S_{\alpha\gamma}p_{\gamma\beta} + p_{\alpha\gamma}S_{\gamma\beta}\right]. \end{aligned} \tag{4.17}$$

Thus, the problem to solve becomes a slightly rewritten version of (3.1)

$$\begin{aligned}\chi^{(2)\prime\prime}_{abc}(\omega) = \frac{e^3}{4\pi m^3\hbar^2\varepsilon_0\omega^3 w}\sum_{c,v,l}\int\Big[&\frac{\tilde{p}^a_{vc}\{p^a_{cl}p^b_{lv}\}}{\omega-\omega_{lv}}\delta\big(2\omega-\omega_{cv}\big) \\ &+\Big(\frac{\tilde{p}^a_{vl}\{p^b_{lc}p^c_{cv}\}}{\omega+\omega_{cl}} + \frac{\tilde{p}^a_{cl}\{p^b_{lv}p^c_{vc}\}}{\omega+\omega_{lv}}\Big)\delta\big(\omega-\omega_{cv}\big)\Big]d^2k, \quad l\neq v,c\end{aligned} \tag{4.18}$$

where the curly brackets denote symmetrization with respect to the last two cartesian indices and the k-space integral have been transformed from 3D to 2D as was done in the linear case. A 2D analouge to the LATM generalized for nonlinear calculations could be derived at this point. This was not done in this project due to time limitations, however simulations were performed by the simpler standard LATM derived in section 4.3, where the full expressions factored onto the respective δ-functions were treated as the momentum matrix elements in the linear case, with the same concerns as for the 3D case. The only thing missing to perform the calculation is the explicit form of $S_{\alpha\beta}$; assuming a plane wave basis, it follows that

$$\begin{aligned}S_{\alpha\beta} &= \frac{1}{\Omega}\sum_{\vec{G},\vec{G}',\vec{G}''} f^{\alpha*}_{\vec{k}-\vec{G}} f^{\beta}_{\vec{k}-\vec{G}'} S_{\vec{G}''}\int e^{i(\vec{G}-\vec{G}'-\vec{G}'')\cdot\vec{r}}d^3r \\ &= \sum_{\vec{G},\vec{G}'} f^{\alpha*}_{\vec{k}-\vec{G}} f^{\beta}_{\vec{k}-\vec{G}'} S_{\vec{G}_z-\vec{G}'_z}\delta_{\vec{G}'_{\parallel},\vec{G}_{\parallel}}.\end{aligned}$$

where $S_{\vec{G}} = S_{G_z}$ is the Fourier transform of $S(z)$ given by

$$S_{G_z} = \begin{cases} \dfrac{1}{ic_zG_z}\left[e^{iz_1G_z} - e^{-iz_2G_z}\right] & G_z \neq 0 \\ \dfrac{1}{z_1+z_2}, & G_z = 0 \end{cases}.$$

With this, SHG from thin surface layers can be calculated using the framework already prestented. A necessary condition for the validity of the above method, is that the contribution to the momentum matrix elements from the bulk-like interior of the thin films vanishes, with a region near the surfaces giving the largest, assymetric contribution. This was tested for a few film thicknesses by evaluating the expression

$$P(\vec{k},z_1,z_2) = \sum_{i=1}^{12}\mathrm{Im}\left[\tilde{p}^x_{ct}p^x_{tv}p^x_{vc}\right](\vec{k})\Big|_{\mathbf{T}_i\cdot\vec{k}}$$

where $\mathbf{T}_i$ takes a k-point of the irreducible Brillouin zone into its i'th equivalent point. The above equation is plotted in 4.10 for a fixed z_1 in the middle of the vacuum layer, $z_1 = (w-c_z)/2$, for varying z_2 . Results are shown for several film thicknesses, where it

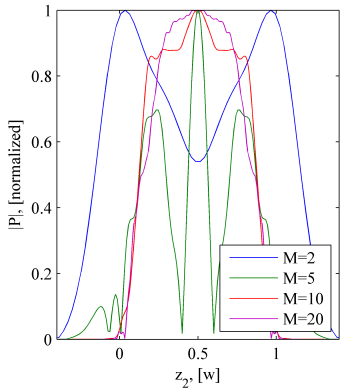


Figure 4.10: Momentum matrix elements for $\vec{k} = (0.41, -0.12, 0)$ (a randomly generated k-point) and $c = 4M+2$, $v = 4M-2$ and $t = c-2$, where the indices are sorted with 1 refering to the state with smallest energy eigenvalue, 2 the second smallest, etc.

should be clear that for thick films, the interior contributes little; Please note that the length scale is normalised to the film thickness. A variety of band indices and random k-points were tested with similar results. Due to symmetry, only four independent tensor elements give finite results, namely $\chi^{(2)}_{yyy}, \chi^{(2)}_{zzz}$, $\chi^{(2)}_{xxz}$ and $\chi^{(2)}_{zxx}$[42]. It was found by numerical testing that the integration domain could be restricted to the irreducible zone upon performing the following momentum matrix element substitutions

$$\begin{aligned} zzz &\to zzz, \\ xxz &\to \frac{1}{2}(zxx + zyy), \\ zxx &\to \frac{1}{2}(xxz + yyz). \end{aligned} \tag{4.19}$$

No similar representation of the xxx element was found, however is was found that the integration could be restricted to half the Brillouin zone without applying any matrix element substitutions. An example of such a calculation for a $M = 10$ Ge film neglecting the states in the band gap with z_1 in the middle of the vacuum layer and $z_2 = w/2$ is shown in Fig. 4.10. Here it is seen that the zzz tensor element dominates with a single peak. Generally, the other matrix elements are seen to be an order of magnitude smaller than the zzz-element. Thus, the tensor elements in the direction of confinement (z-elements) are seen to dominate. The computation time is increased vastly due to the additional band summation and the convolution necessary to the implimentation of $S(z)$ in reciprocal space, combined with a large basis of several thousand plane waves. The computation time of the spectra shown in Fig. 4.11 was $\sim 5-6$ full days.

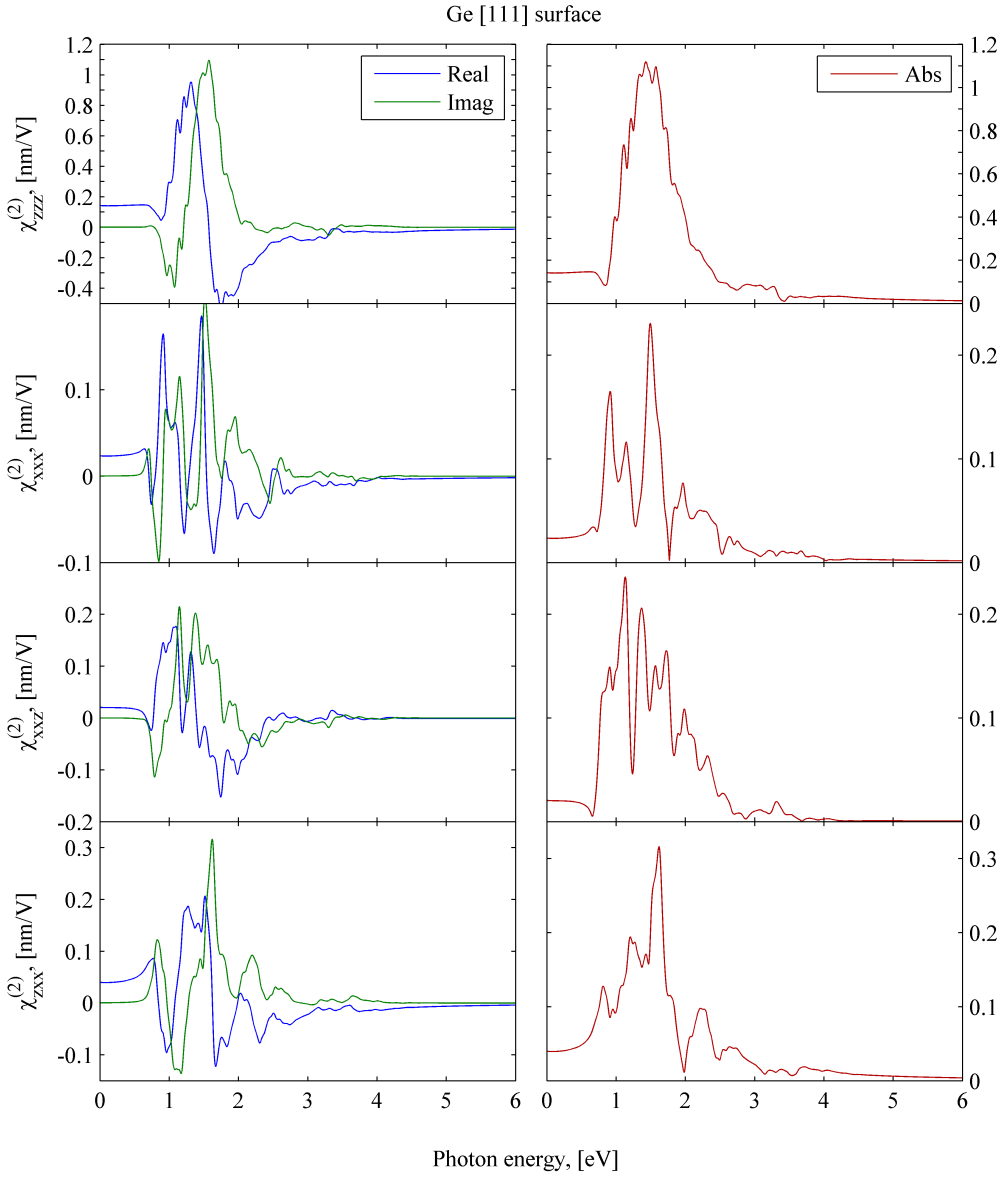


Figure 4.11: Second harmonic generation from Ge [111] surface modelled by an $M = 10$ slab. The EPM with a cut-off $|\vec{G}| <= 4$ was used and the integration was performed over half the Brillouin zone using the standard LATM method using 153 tetrahedra in the irreducible zone. 10meV broadening was applied both in the singular energy fractions and later by convolution with a Lorentzian lineshape. Surface bands in the band-gap is ignored, corresponding to saturation with an optically inactive compound.

5 Experimental Second Harmonic Generation

In this chapter experimental, procedures and theory connected with determining second harmonic spectra will be reviewed. First, an experimental setup for measuring second harmonic generation in a reflection and transmission configuration will be reviewed. Next, the theory needed to interpret the measured signals with respect to the second harmonic tensor will be examined followed by a presentation of the experimental spectra measured in connection with this project.

5.1 Experimental setup

A setup for measuring second harmonic generation in reflection or transmission setup consists, simply put, of a laser pumping the sample at the fundamental wavelength and a detector measuring the radiated second harmonic signal. To this end, a laser capable of supplying a high-intensity output is needed; in this setup a Q-switched Nd:YAG laser operating at a wavelength of 1064nm, 10Hz repetition frequency and a pulse width of 5-6ns pumps two nonlinear crystals allowing first frequency doubling by SHG followed by sum frequency generation to produce three beams of wavelenghts of 1064nm, 532nm and 355nm. An optical parametric oscillator (OPO) allows tuning of the 355nm beam over a large frequency spectrum; a ”signal” beam at wavelenghts 420-700nm and an ”idler” beam at 700-2400nm. The output of the laser/OPO system will primarily be the desired beam gained by adjusting the OPO cavity index, however, traces of the pump beams will still be present, although at intensities several orders of magnitude smaller than the set output frequency. After losses in the wavelength tuning, the beam intensity is in the $\sim$ 20mW range, providing kW-MW peak pulse powers suitable for SHG. Filters are applied before the sample to attenuate all wavelengths appart from the desired fundamental wavelength and after the sample to remove everything but the second harmonic signal (e.g. the pump beam, which can damage the sensitive detector system). A suitable photomultiplier coupled to a lock-in amplifier is used to detect the second harmonic light. Care was taken to only focus the beam where needed, reducing the intensity of the beam and hence harmonic generation in addition to local energy dissipation, damaging the optical components (it was found that focusing the beam at visual frequencies on a polished GaAs wafer produced visible defects in the surface, presumably by evaporating or sputtering the material). To take into acount the frequency dependence of the laser yield, and the optical components of the setup, a spectrum was measured on a quartz sample, providing a reference without prominent spectral features.

When measuring in the pump range 730-1000nm (1000-1400nm), a lowpass RG715 (RG850) filter attenuating below 715nm (850nm) was applied before the sample, a combination of filters providing a band-pass filter in the range 740-1000nm (1000-1400nm) was applied on the PMT whereas a C3C23 (KG3) filter where used after the sample as a 200-700nm (300-650nm) band pass filter allowing second harmonic frequencies to pass, but removing the fundamental beam in addition to higher than second harmonic modes.

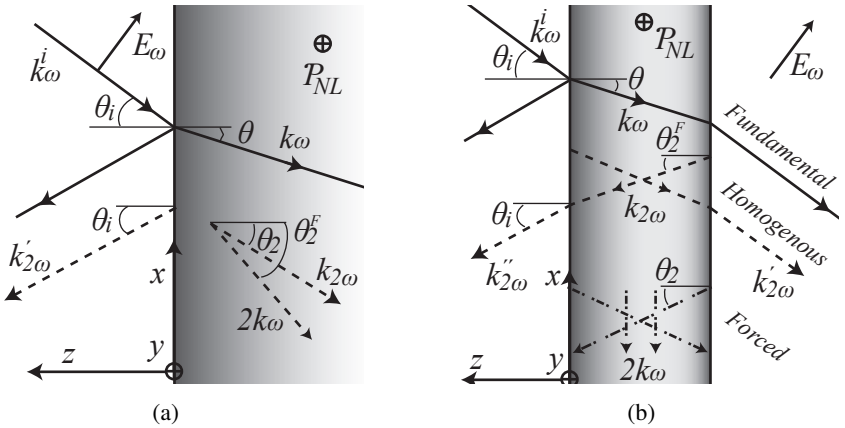


Figure 5.1: Geometry and conventions used for (a) an infinite medium and (b) a slab. Dashed lines indicate second harmonic properties whereas full lines indicate the properties of the fundamental wave.

5.2 SHG from [100] zinc-blende crystals in reflection configuration

In this and the following section the optical theory of SHG from a slab will be reviewed, based on the work of Bloembergen & Pershan (1962)[9]. This case is obviously very relevant for the experimental setup considered above. A p-polarized wave at the fundamental frequency ω is incident from vacuum on a [100]-oriented semiconducting zinc-blende solid. Choosing a coordinate system with the z-axis aligned perpendicular to the surface with positive direction into the vacuum layer, the coordinate axes align with the cubic crystal axes; the geometry is illustrated in Fig. 5.1(a). A plane p-polarized wave is transmitted across the interface, polarizing the material both linearly and non-linearly. Hence the fundamental wave is depleted upon generation of harmonic waves; however, in the context at hand we shall only consider low-efficiency non-linear processes allowing operation within the no-depletion regime. Hence, the fundamental field inside the crystal is given from linear optics

$$\vec{\mathcal{E}}_\omega = \mathcal{E}_0 \left(\sin\theta\hat{x} + \cos\theta\hat{z}\right) e^{i\vec{k}_\omega\cdot\vec{r}}, \quad k_\omega^z < 0. \tag{5.1}$$

The fundamental field induces a second harmonic, non-linear polarization

$$\mathcal{P}^i_{NL} = \varepsilon_0 \sum_{j,k} \chi^{(2)}_{ijk} \mathcal{E}^j_\omega \mathcal{E}^k_\omega.$$

In the bulk of zinc-blende crystals, the only tensor elements different from zero are the six cartesian index permutations of $\chi^{(2)}_{xyz}$, which in turn are all equal[43]. Due to the field polarization, all terms containing y in the indices j and k vanishes, making the only second harmonic polarization component

$$\mathcal{P}^y_{NL} = 2\varepsilon_0 \chi^{(2)}_{xyz} \mathcal{E}^x_\omega \mathcal{E}^z_\omega = 2\varepsilon_0 \chi^{(2)}_{xyz} \mathcal{E}_0^2 \sin\theta\cos\theta e^{i2\vec{k}_\omega\cdot\vec{r}}.$$

The goal is to calculate the second harmonic electric fields generated by the induced, second harmonic polarization $\mathcal{P}_{NL}$. To this end, the inhomogeneous wave equation, with the nonlinear polarization as a source term, is considered

$$\left[\nabla\times(\nabla\times) - k_{2\omega}^2\right]\vec{\mathcal{E}}_{2\omega} = 4\mu_0\omega^2\mathcal{P}_{NL}\hat{y}, \quad k_{2\omega}^2 = \varepsilon_{2\omega}\frac{4\omega^2}{c^2}.$$

As an inhomogeneous differential equation, the general solution is given by the homogeneous solution plus a particular, inhomogeneous solution. Thus, the homogeneous solution is taken on the usual plane-wave form

$$\vec{\mathcal{E}}^0_{2\omega} = \vec{A}e^{i\vec{q}\cdot\vec{r}},$$

where $\vec{q}$ is found by insertion into the homogeneous wave equation

$$\begin{aligned}
\left[\nabla\times\nabla\times -k^2_{2\omega}\right]\vec{\mathcal{E}}^0_{2\omega} &= 0\\
-\vec{q}\times(\vec{q}\times\vec{A}) - \vec{k}^2_{2\omega}\vec{A} &= 0\\
-\vec{q}(\vec{q}\cdot\vec{A}) + \vec{q}^2\vec{A} - \vec{k}^2_{2\omega}\vec{A} &= 0\\
q^2 - k^2_{2\omega} = 0 \Rightarrow \vec{q} &= \pm\vec{k}_{2\omega}.
\end{aligned} \tag{5.2}$$

Transversal fields have been assumed from third to fourth line. Considering the medium infinite corresponds to choosing the solution propagating into the film. Thus, a solution where $k^z_{2\omega} < 0$ is chosen The inhomogeneous "forced" solution is also found by insertion, again using a plane wave

$$\mathcal{E}^I_{2\omega} = \vec{B}e^{i\vec{q}'\cdot\vec{r}}$$

to give

$$\begin{aligned}
\left[\nabla\times(\nabla\times) - k^2_{2\omega}\right]\vec{B}e^{i\vec{q}'\cdot\vec{r}} &= 8\omega^2\varepsilon_0\mu_0\chi^{(2)}_{xyz}\mathcal{E}^x_\omega\mathcal{E}^z_\omega\hat{y} = 8\varepsilon_0\mu_0\omega^2\chi^{(2)}_{xyz}\mathcal{E}^2_0\sin\theta\cos\theta e^{i2\vec{k}_\omega\cdot\vec{r}}\hat{y}\\
\vec{B}\left[q'^2 - k^2_{2\omega}\right]e^{i\vec{q}'\cdot\vec{r}} &= 4\mu_0\omega^2\tilde{\mathcal{P}}_{NL}e^{i2\vec{k}_\omega\cdot\vec{r}}\hat{y}.
\end{aligned}$$

A particular solution to the above equation can be found by first requiring a phase match, $\vec{q}' = 2\vec{k}_\omega$ followed by isolation of $\vec{B}$

$$\vec{B} = \frac{4\omega^2\mu_0\tilde{\mathcal{P}}_{NL}}{4k^2_\omega - k^2_{2\omega}}\hat{y} = \frac{\tilde{\mathcal{P}}_{NL}}{\varepsilon_0(\varepsilon_\omega - \varepsilon_{2\omega})}\,\hat{y}. \tag{5.3}$$

Thus, the general solution inside the medium ($z < 0$) becomes

$$\vec{\mathcal{E}}_{2\omega} = \vec{A}e^{i\vec{k}_{2\omega}\cdot\vec{r}} + \frac{\tilde{\mathcal{P}}_{NL}}{\varepsilon_0(\varepsilon_\omega - \varepsilon_{2\omega})}e^{i2\vec{k}_\omega\cdot\vec{r}}\,\hat{y}, \quad z < 0.$$

Outside of the medium, there is no second harmonic source term, why solutions to the wave equation must correspond to a "reflected" wave generated in the medium, traveling in the $+z$-direction

$$\vec{\mathcal{E}}_{2\omega} = \vec{C}e^{i\vec{k}'_{2\omega}\cdot\vec{r}}, \quad k'^z_{2\omega} > 0. \tag{5.4}$$

Phase matching at the $z = 0$ boundary means that the x-component of all wave vectors must be equal for the respective frequency components. Denoting the wave vector of the incoming fundamental wave $\vec{k}^i_\omega$, the reflected fundamental $\vec{k}'_\omega$ and the transmitted fundamental $\vec{k}_\omega$, the following must hold

$$k^{ix}_\omega = k^x_\omega = k'^x_\omega.$$

Realizing that

$$\begin{aligned}
k^{ix}_\omega &= |\vec{k}^i_\omega|\sin\theta_i = \omega/c\sin\theta_i\\
k^x_\omega &= |\vec{k}_\omega|\sin\theta = n_\omega\omega/c\sin\theta\\
k^{x'}_\omega &= |\vec{k}'_\omega|\sin\theta' = \omega/c\sin\theta'
\end{aligned}$$

leads to Snell's law of refraction for linear media $\sin\theta_i = \sin\theta'$ and $\sin\theta_i = n_\omega \sin\theta$. The second harmonic waves obey a similar relationship; we shall denote the wave vector of the second harmonic "reflected" wave in vacuum $\vec{k}'_{2\omega}$, the wavevector of the homogeneous solution $\vec{k}_{2\omega}$ and the wave vector of the forced solution $\vec{k}^F_{2\omega} = 2\vec{k}_\omega$. Phase matching the second harmonic fields

$$k'^x_{2\omega} = k^x_{2\omega} = k^{Fx}_{2\omega}, \tag{5.5}$$

and by trigonometry

$$\begin{aligned} k'^x_{2\omega} &= |\vec{k}'_{2\omega}| \sin\theta'_2 = 2\omega/c \sin\theta'_2, \\ k^x_{2\omega} &= |\vec{k}_{2\omega}| \sin\theta_2 = 2n_{2\omega}\omega/c \sin\theta_2, \\ k^{Fx}_{2\omega} &= 2|\vec{k}_\omega| \sin\theta = 2n_\omega\omega/c \sin\theta, \end{aligned}$$

giving the refraction and reflection directions, analogous to Snell's law, for the second harmonic fields

$$\begin{aligned} \sin\theta'_2 &= \sin\theta_i, \\ \sin\theta_2 &= n^{-1}_{2\omega} \sin\theta_i, \\ \sin\theta^F_2 &= n^{-1}_\omega \sin\theta_i. \end{aligned}$$

In other words, the radiated second harmonic field propagates in the same direction as the fundamental, reflected wave. The homogeneous, second harmonic wave propagates in general in another direction than the inhomogeneous wave; taking GaAs as an example at a fundamental wavelength of 1600nm. Here $n_\omega = 3.3$ whereas $n_{2\omega} = 3.66$ (plus a small imaginary part, which is neglected), using $\theta_i = 60^\circ$ results in $\theta^F_2 = 15.2^\circ$ and $\theta_2 = 13.6^o$ making the difference 1.6°. Thus, for a thin sample relative to the beam radius the so called "walk-off" effect of the beams separating in space can be neglected.

The coefficients are to be found by boundary conditions. Each frequency component of the $\mathcal{E}$-field must have its tangential parts continuous across the interface; hence

$$A_x + B_x = A_x = C_x.$$

Since the direction of the total second harmonic field inside the nonlinear medium is aligned along $\hat{y}$, $A_x = 0$, and thus $C_x = 0$. For the y-part

$$A_y + B_y = C_y. \tag{5.6}$$

The tangential component of the magnetic intensity $\mathcal{H}^x_{2\omega}$ must also be continuous; the magnetic intensity is given by Maxwell's equation

$$\nabla \times \vec{\mathcal{E}}_{2\omega} = i2\omega\vec{\mathcal{B}}_{2\omega} = i2\omega\mu_0\vec{\mathcal{H}}_{2\omega}.$$

Noting that $\vec{\mathcal{E}}_{2\omega}$ is parallel to the y-axis and writing out the curl, $\mathcal{H}^x_{2\omega}$ can be found

$$i2\omega\mu_0\mathcal{H}^x_{2\omega} = \frac{\partial}{\partial z}\mathcal{E}^y_{2\omega} = \begin{cases} ik^z_{2\omega}A_y e^{i\vec{k}_{2\omega}\cdot\vec{r}} + i2\vec{k}^z_\omega B_y e^{2i\vec{k}_\omega\cdot\vec{r}} & z < 0 \\ -ik'^z_{2\omega}C_y e^{i\vec{k}'_{2\omega}\cdot\vec{r}} & z > 0 \end{cases},$$

where the sign-convention for the z-components of the wave vectors was applied. Equating the expressions for the magnetic intensity above at $z = 0$ leads to

$$k^z_{2\omega}A_y e^{i\vec{k}^x_{2\omega}x} + 2k^z_\omega B_y e^{2ik^x_\omega x} = -k'^z_{2\omega}C_y e^{ik'^x_{2\omega}x},$$

Applying the phase matching conditions allows

$$k^z_{2\omega}A_y + 2k^z_{\omega}B_y = -k'^z_{2\omega}C_y,$$

and expressing A_y through (5.6)

$$B_y\left(2k^z_{\omega} - k^z_{2\omega}\right) = -C_y\left(k'^z_{2\omega} + k^z_{2\omega}\right).$$

Hence, the coefficient for the reflected, second harmonic light is found by isolation and use of (5.3)

$$C_y = -\frac{\tilde{P}_{NL}}{\varepsilon_0(\varepsilon_{\omega} - \varepsilon_{2\omega})}\,\frac{2k^z_{\omega} - k^z_{2\omega}}{k'^z_{2\omega} + k^z_{2\omega}}. \tag{5.7}$$

The remaining wave vector components can be found by consideration of the experimental setup

$$\begin{aligned} k^z_{\omega} &= n_{\omega}\frac{\omega}{c}\cos\theta, \\ k^z_{2\omega} &= n_{2\omega}\frac{2\omega}{c}\cos\theta_2, \\ k^{z'}_{2\omega} &= \frac{2\omega}{c}\cos\theta'_2, \\ k_{\omega} &= n_{\omega}\frac{\omega}{c}, \end{aligned}$$

giving

$$C_y = \frac{\tilde{P}_{NL}}{\varepsilon_0(\varepsilon_{\omega} - \varepsilon_{2\omega})}\,\frac{n_{2\omega}\cos\theta_2 - n_{\omega}\cos\theta}{\cos\theta'_2 + n_{2\omega}\cos\theta_2},$$

or explicitly in terms of the response tensor

$$C_y = \chi^{(2)}_{xyz}\,\frac{2\mathcal{E}_0^2\sin\theta\cos\theta}{(\varepsilon_{\omega} - \varepsilon_{2\omega})}\,\frac{n_{2\omega}\cos\theta_2 - n_{\omega}\cos\theta}{\cos\theta'_2 + n_{2\omega}\cos\theta_2},$$

The squared amplitude of the fundamental field $\mathcal{E}_0^2$ is calculated from the incident field $\mathcal{E}_{inc}^2$ by the Fresnel transmission coefficient

$$\mathcal{E}_0^2 = \tau_p^2\mathcal{E}_{inc}^2, \quad \tau_p = \frac{2\sin\theta\cos\theta_i}{\sin(\theta_i + \theta)\cos(\theta_i - \theta)}.$$

In experiments, measurements are always relative to the intensity of the pump beam; and in second harmonic measurements the result will scale quadratically with the pump intensity. To gain an expression which is independent of the pump intensity, and thus pertains wholly to the physics of the sample, we define a second harmonic reflection coefficient in the p-to-s configuration discussed here:

$$R_{ps} \equiv \frac{I^s_{2\omega}}{(I^p_{\omega})^2}.$$

The intensity (z-component of the Poynting vector) of monochromatic, plane waves are proportional to the absolute square of the field amplitudes; hence

$$I^s_{2\omega} = \frac{n^i_{2\omega r}\varepsilon_0 c}{2}|\mathcal{E}_{2\omega}|^2\cos\theta'_2, \quad I^p_{\omega} = \frac{n^i_{\omega r}\varepsilon_0 c}{2}|\mathcal{E}_{inc}|^2\cos\theta_i,$$

where $n^i_{\omega r}$ and $n^i_{2\omega r}$ are the real parts of the (linear) refractive indices of the medium at frequencies ω and 2ω, respectively; hence they are both equal to one in the case of reflection into vacuum considered here. With this, R_{ps} is found to be (noting that $\theta_i = \theta'_2$)

$$R_{ps} = \left|\chi^{(2)}_{xyz}\right|^2 \ |r_{ps}|^2 \tag{5.8}$$

where $|r_{ps}|^2$ has been defined as the function

$$|r_{ps}|^2 = \frac{8}{\varepsilon_0 c}\left|\frac{\tau_p^2 \sin\theta\cos\theta(n_{2\omega}\cos\theta_2 - n_\omega\cos\theta)}{(\varepsilon_\omega - \varepsilon_{2\omega})(\cos\theta'_2 + n_{2\omega}\cos\theta_2)}\right|^2.$$

5.3 SHG in a transmission configuration

Considering now a slab of non-linear medium of zinc-blende symmetry with vacuum interfaces at $z = 0$ and $z = -w$, illustrated by Fig. 5.1(b). In the space $z > 0$ we have the same fields as in last section, i.e. an incident fundamental and a reflected fundamental wave propagating in the $-z$ and $+z$ directions, respectively. Additionally, a second harmonic "reflected" wave is propagating towards $+z$. The finite thickness of the slab means that reflections from the back side is found; hence inside the slab $z < 0 < -w$ two fundamental waves propagate in the $\pm z$-directions, with the same xy-dispersion; the same is true for the homogeneous, second harmonic wave whereas four inhomogeneous waves are present, two of which propagate parallel to the surface. For $z < -w$, a transmitted fundamental and a transmitted second harmonic wave is found, both propagating towards $-z$.

The fundamental field inside the non-linear medium in the low-depletion regime

$$\vec{\mathcal{E}}_\omega = \left\{\vec{\mathcal{E}}^r_\omega e^{-ik^z_\omega z} + \vec{\mathcal{E}}^l_\omega e^{ik^z_\omega z}\right\} e^{ik^x_\omega x},$$

induces a non-linear polarization

$$\begin{aligned}\vec{\mathcal{P}}_{NL} &= 2\varepsilon_0\chi^{(2)}_{xyz}\mathcal{E}^x_\omega\mathcal{E}^y_\omega\,\hat{y} \\ &= 2\varepsilon_0\chi^{(2)}_{xyz}\left\{\mathcal{E}^{rx}_\omega\mathcal{E}^{ry}_\omega e^{-2ik^z_\omega z} + \mathcal{E}^{lx}_\omega\mathcal{E}^{ly}_\omega e^{2ik^z_\omega z} + \mathcal{E}^{rx}_\omega\mathcal{E}^{ly}_\omega + \mathcal{E}^{lx}_\omega\mathcal{E}^{ry}_\omega\right\} e^{2ik^x_\omega x}\,\hat{y},\end{aligned}$$

where we shall make the low-reflection approximation of retaining only the first term

$$\vec{\mathcal{P}}_{NL} \approx 2\varepsilon_0\chi^{(2)}_{xyz}\mathcal{E}^{rx}_\omega\mathcal{E}^{ry}_\omega e^{2i(k^x_\omega x - k^z_\omega z)}\,\hat{y}.$$

The second harmonic, homogeneous fields become

$$\vec{\mathcal{E}}^0_{2\omega} = \begin{cases}\left[\vec{A}_r e^{-ik^z_{2\omega}z} + \vec{A}_l e^{ik^z_{2\omega}z}\right] e^{i\vec{k}^x_{2\omega}x}, & 0 > z > -w \\ \vec{C}e^{i(k''^z_{2\omega}z + k''^x_{2\omega}x)}, & z > 0 \\ \vec{D}e^{i(-k'^z_{2\omega}z + k'^x_{2\omega}x)}, & z < -w.\end{cases}$$

The particular, inhomogeneous solution can be chosen as in last section (only a single inhomogeneous wave is found in the low-reflection approximation as opposed to the four waves displayed in Fig 5.1(b))

$$\vec{\mathcal{E}}^I_{2\omega} = \vec{B}e^{i2(k^x_\omega x - k^z_\omega z)}, \quad \vec{B} = \frac{\tilde{\mathcal{P}}_{NL}}{\varepsilon_0(\varepsilon_\omega - \varepsilon_{2\omega})}\hat{y}.$$

Boundary conditions for the tangential part of the electric field at $z = 0$

$$C_y = A_r^y + A_l^y + B_y,$$

the magnetic intensity at $z = 0$

$$k_{2\omega}''^z C_y = k_{2\omega}^z (A_l^y - A_r^y) - 2k_\omega^z B_y,$$

the electric field at $z = -w$

$$D_y = A_l^y e^{i\phi} + A_r^y e^{-i\phi} + B_y e^{-i\phi_S},$$

and the magnetic intensity at $z = -w$

$$-k_{2\omega}' D_y = k_{2\omega}^z \left(A_l^y e^{i\phi} - A_r^y e^{-i\phi}\right) - 2k_\omega^z B_y e^{-i\phi_S},$$

using the phase factors $\phi = -k_{2\omega}^z w$ and $\phi_S = -2k_\omega'^z w$. The four boundary conditions are solved for the coefficients A_r^y, A_l^y, C_y and D_y

$$\begin{aligned}
A_l^y &= B_y \frac{(k_{2\omega}^z - k_{2\omega}'^z)(k_{2\omega}''^z + 2k_\omega^z)e^{-i\phi} - (2k_\omega^z + k_{2\omega}'^z)(k_{2\omega}''^z + k_{2\omega}^z)e^{-i\phi_S}}{(k_{2\omega}^z - k_{2\omega}''^z)(k_{2\omega}^z - k_{2\omega}'^z)e^{-i\phi} - (k_{2\omega}^z + k_{2\omega}''^z)(k_{2\omega}^z - k_{2\omega}'^z)e^{i\phi}}, \\
A_r^y &= -B_y \frac{(2k_\omega^z + k_{2\omega}'^z)(k_{2\omega}^z - k_{2\omega}''^z)e^{-i\phi_S} - (2k_\omega^z + k_{2\omega}''^z)(k_{2\omega}^z + k_{2\omega}'^z)e^{i\phi}}{(k_{2\omega}^z - k_{2\omega}''^z)(k_{2\omega}^z - k_{2\omega}'^z)e^{-i\phi} - (k_{2\omega}^z + k_{2\omega}''^z)(k_{2\omega}^z + k_{2\omega}'^z)e^{i\phi}}, \\
C_y &= B_y \frac{(k_{2\omega}^z - k_{2\omega}'^z)(k_{2\omega}^z + 2k_\omega^z)e^{-i\phi} - (k_{2\omega}^z + k_{2\omega}'^z)(k_{2\omega}^z - 2k_\omega^z)e^{i\phi} - k_{2\omega}^z(4k_\omega^z + 2k_{2\omega}'^z)e^{-i\phi_S}}{(k_{2\omega}^z - k_{2\omega}''^z)(k_{2\omega}^z - k_{2\omega}'^z)e^{-i\phi} - (k_{2\omega}^z + k_{2\omega}''^z)(k_{2\omega}^z + k_{2\omega}'^z)e^{i\phi}} \\
&\equiv B_y \tilde{C}, \\
D_y &= -B_y \frac{(k_{2\omega}^z - k_{2\omega}''^z)(k_{2\omega}^z + 2k_\omega^z)e^{-i(\phi_S+\phi)} - (k_{2\omega}^z + k_{2\omega}''^z)(k_{2\omega}^z - 2k_\omega^z)e^{i(\phi-\phi_S)} - k_{2\omega}^z(4k_\omega^z + 2k_{2\omega}''^z)}{(k_{2\omega}^z - k_{2\omega}''^z)(k_{2\omega}'^z - k_{2\omega}^z)e^{-i\phi} - (k_{2\omega}^z + k_{2\omega}''^z)(k_{2\omega}^z + k_{2\omega}'^z)e^{i\phi}} \\
&\equiv B_y \tilde{D}_y.
\end{aligned}$$

The above expressions, although not aesthetically pleasing, can be evaluated easily. The following limits can deduced from the above expressions valid for large w by simply writing out the complex exponentials and realizing that the imaginary part of the k-vectors leads to exponentially increasing and decaying terms:

$$\begin{aligned}
\tilde{C}_y &= \frac{(k_{2\omega}^z + k_{2\omega}'^z)(k_{2\omega}^z - 2k_\omega^z)}{(k_{2\omega}^z + k_{2\omega}''^z)(k_{2\omega}^z + k_{2\omega}'^z)}, \quad n_{2\omega} \in \mathbb{C}, w/\lambda >> 1 \\
\tilde{D}_y &= \frac{(k_{2\omega}^z + k_{2\omega}''^z)(k_{2\omega}^z - 2k_\omega^z)e^{i2k_\omega'^z w}e^{-2k_\omega''^z}}{(k_{2\omega}^z + k_{2\omega}''^z)(k_{2\omega}^z + k_{2\omega}'^z)}, \quad n_{2\omega} \in \mathbb{C}, w/\lambda >> 1.
\end{aligned}$$

where prime and double prime indicate "real part" and "imaginary part", respectively. Thus, the limit of $\tilde{C}_y$ gives the same expression for C_y as was calculated for the semi-infinite sample in the previous section, as might be expected on physical grounds. In a medium where attenuation is found for the fundamental wave, which is the case for semiconductors at fundamental photon energies above the band gap, no SHG can be observed in a transmission configuration due to the dampening arising from the imaginary part of k_ω^z. A reflection coefficient is defined similarly to (5.8) with

$$I_{2\omega}^s = \frac{n_{2\omega r}\varepsilon_0 c}{2} \left| \tilde{C}_y \frac{\tau^2 \mathcal{E}_{inc}^2 \cos\theta \sin\theta}{\varepsilon_0(\varepsilon_\omega - \varepsilon_{2\omega})} \right|^2 |\chi_{xyz}^{(2)}|^2 \cos\theta_i$$

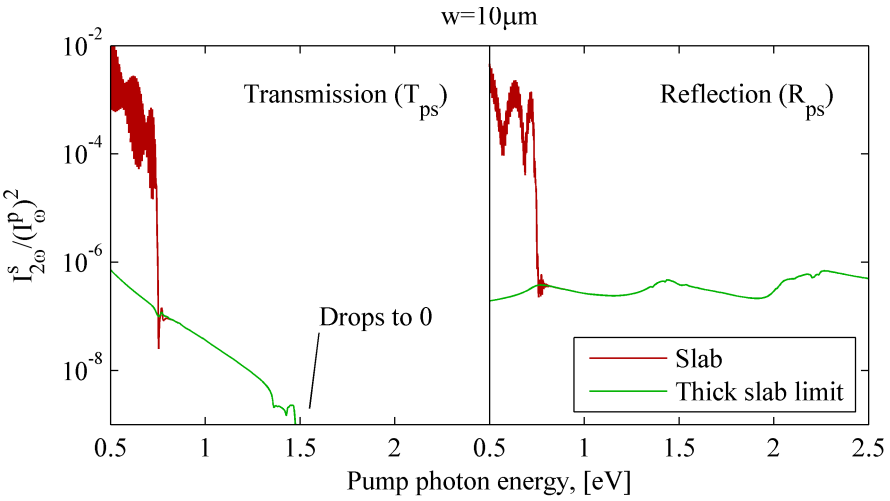


Figure 5.2: Expected second harmonic signal normalized to the square of the incident radiation intensity.

and I_ω^p as before. Thus the reflection coefficient of a slab becomes

$$r_{ps} = \frac{8}{\varepsilon_0 c}\left|\tilde{C}_y \frac{\tau^2 \cos\theta \sin\theta}{(\varepsilon_\omega - \varepsilon_{2\omega})}\right|^2.$$

A transmission coefficient is defined in a similar way, this time noting that the incident and transmitted angles may be different

$$t_{ps} = \frac{8}{\varepsilon_0 c}\left|\tilde{D}_y \frac{\tau^2 \cos\theta \sin\theta}{(\varepsilon_\omega - \varepsilon_{2\omega})}\right|^2 \frac{\cos\theta_2'}{\cos^2\theta_i}.$$

In Fig. 5.2 T_{ps} and R_{ps} is modeled using the optical model presented in this and the previous section and the calculated optical spectra of Figs. 2.4 and 3.2. It is clear that a transmission setup is useless when investigating photon energies above the band gap in thick films, as discussed previously. Furthermore, below half the band gap, the sample becomes transparent to second harmonic radiation. This allows interference phenomena to come into play, as is readily seen by the onset of fluctuations in the spectra. Furthermore, since second harmonic light becomes undamped and these low energies, the entire sample thickness radiates second harmonic fields increasing the overall signal vastly. Another important effect is the phase-matching condition $\varepsilon_\omega = \varepsilon_{2\omega}$, which occurs for low energies when the imaginary part of the linear dielectric constant vanishes and the real part becomes constant. A concern at low energies must be the seemingly large conversion ratio to second harmonic waves, making the no-depletion approximation invalid.

These effects are especially pronounced in Fig. 5.2, since the linear spectrum without broadening was applied, making the absorption edge sharp. However; in comparison with experiments broadening should be applied. The effect of applying 50meV Lorentzian broadening on the linear spectra before applying them to the optical expressions for transmission and reflection is shown in Fig 5.3. The reflection spectra is seen to approach to case for an infinite medium, whereas the transmission signal effectively vanishes due to dampening.

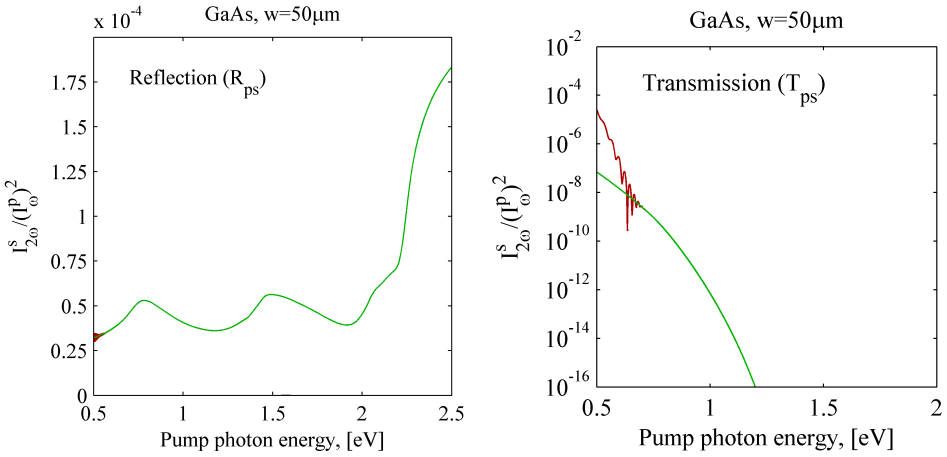


Figure 5.3: Expected second harmonic signal normalized to the square of the incident radiation intensity. 50meV broadening was applied to the linear spectra before application to the optical model.

5.4 Experimental results: GaAs[001]

SHG was from a polished, semi-insulating GaAs(001) wafer was measured in a reflection setup. Upon rotating the sample, using a pump wavelength of 730nm, the second harmonic signal displayed an extreme dependence on the azimuthal angle ϕ, defined as the angle between the incoming wave vector projected onto the [001] surface and the [100]-direction (x-axis), as shown in Fig. 5.4(a) for both P-P and P-S polarization configuration. It is clear that the second harmonic signal is $\pi/2$ periodic in ϕ, in good agreement with the symmetry of the [100] plane of a zinc-blende crystal. This behaviour can easily be understood by revisiting (5.1) and introducing the ϕ-dependency of the field-components explicitly, instead of assuming $\phi = 0$

$$\vec{\mathcal{E}}_\omega = \mathcal{E}_0 \left(\sin\theta\cos\phi\hat{x} + \sin\theta\sin\phi\hat{y} + \cos\theta\hat{z}\right),$$

where ϕ is measured with respect to the x-axis. Thus, the induced second harmonic polarisation becomes

$$\begin{aligned}
\mathcal{P}^x_{NL} &= 2\chi^{(2)}_{xyz}\sin\theta\cos\theta\sin\phi \\
\mathcal{P}^y_{NL} &= 2\chi^{(2)}_{xyz}\sin\theta\cos\theta\cos\phi \\
\mathcal{P}^z_{NL} &= 2\chi^{(2)}_{xyz}\sin\theta\sin\theta\cos\phi\sin\phi,
\end{aligned}$$

and carrying through the ϕ-terms adds $\cos^2(2\phi)$ proportionality to the P-S signal, whereas a P-P signal arises when $\phi \neq 0$ with a $\sin^2(2\phi)$ propertionality. The quartz reference measurement used is shown in Fig. 5.4(b); in order to increase the signal-to-noise ratio, the reference was measured focusing the fundamental beam on the quartz sample and the polarizer following the sample was removed - this is shown to increse the signal by almost a factor of 10. Additionally, the reference is smoothened to give the full lines of 5.4(b), and all spectra shown later is divided by the smoothened reference spectrum. Fig. 5.5 displays the results obtained when investigating a GaAs sample in a reflection setup. In the P-P configuration, no difference between $\phi = 0$ and $\phi = \pi/2$, aside from a general intensity difference, is observed. Results for P-S configuration, where the SHG signal corrected using the optical model of last section (green lines) should be comparable to $|\chi^{(2)}_{xyz}|^2$, display

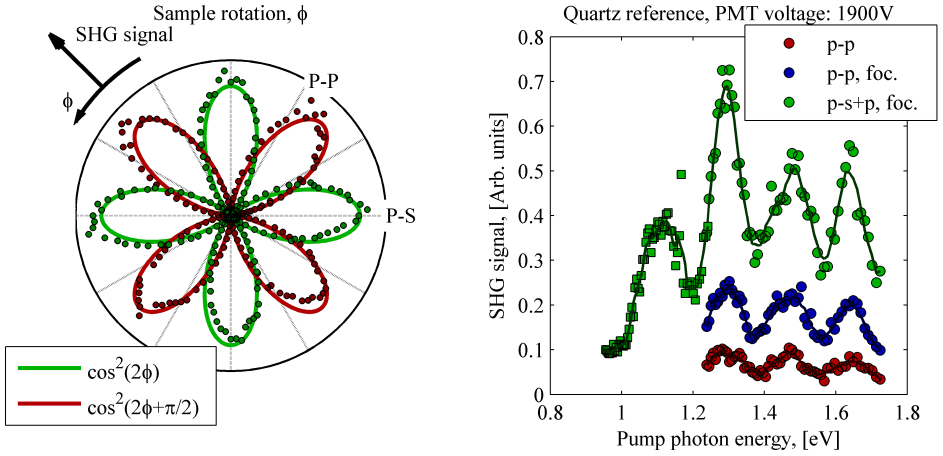


Figure 5.4: (a) Measured second harmonic signal at pump wavelength 730nm as a function of azimuthal angle, ϕ. (b) Quartz reference measurement in P-P polarization configuration. To gain a stronger signal relative to the unfocused beam (red), the beam is focused on the sample (blue) and the polarizer after the sample is removed (green). Full lines indicate smoothed versions of the data points, and it is the smoothed version of the green curve which is used as reference throughout the rest of this section. The outlier near 1.17eV a result of the transmitted OPO pump beam from the Nd:YAG laser operating at 1064nm. The spectra recorded in the intervals 720-1000nm and 1000-1300nm are scaled to give the same result in the overlap.

a rather large discrimination between $\phi = 0$ and $\phi = \pi/2$ for fundamental photon energies above 1.25eV. The agreement between theory and experiment is not striking, although the second harmonic bulk polarisation vanishes with $\phi = \pi/2$ but is maximized for $\phi = 0$ in agreement with the peak near 1.5eV appearing for $\phi = 0$ where the E1 peak is predicted by theory. It is also seen how the lineshape below 1.25eV is very similar for both ϕ, apart from the general intensity difference, which might suggest that this part of the spectrum is due to an effect other than dipole bulk SHG. If this is the case, the measurement for $\phi = \pi/2$ absent of bulk dipole contribution could be used as reference for the $\phi = 0$ measurement. Thus, the quantity

$$\Delta R_{ps} = R_{ps}(\phi = 0) - \Pi R_{ps}(\phi = \pi/2)$$

is investigated, where Π is a normalisation constant taking the SHG signal for $\phi = \pi/2$ at 955meV into the signal for $\phi = 0$. The result is shown in Fig. 5.6(a), which is seen to agree well with the result predicted from theory. The theoretical curve is adjusted to fit the experimental values (the true energy values for the theoretical curve is shown at the top of the figure). A peak position displacement of 30meV between theory and experiment is well within the accepted margin of error in the litterature, and could be explained by e.g. the temperature dependence of the energy band differences. It is noted that even though this result appears convincing, the absolute scale cannot be predicted from the quartz reference since different PMT voltages was appllied. Even though measurements of the SHG intensity for the same setup at varying PMT voltages in the hopes of making a calibration curve was performed, the results were seen not to scale the 920-1000nm and 1000-1300nm data sets correctly (i.e. providing a continous spectrum at the overlap), differing in scale by up to a factor of ~ 2. These spectra were measured hours appart, allowing the laser system ample time to drift in intensity. The lineshape of the experimental peak appears much sharper than predicted by theory, which is possibly due to excitonic effects, as reported By

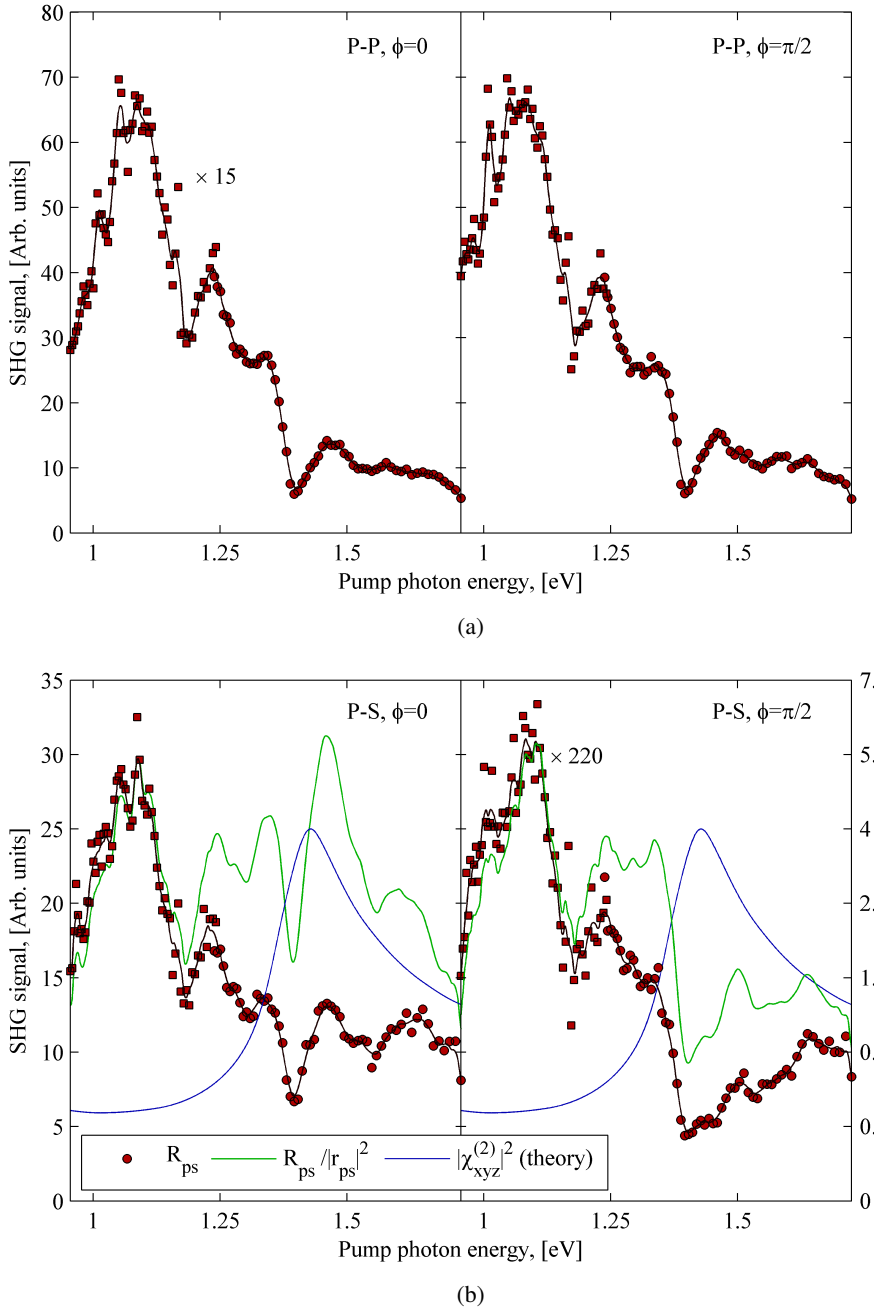
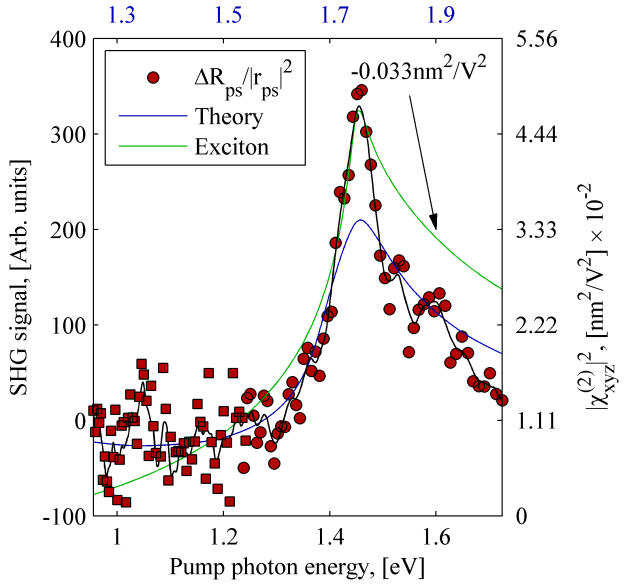


Figure 5.5: Measured SHG from GaAs wafer. (a) P-P polarizer configuration. (b) P-S polarizer configuration.

T.G. Pedersen and H. Cornean[44]. In this work, a three-band model of the tensor element $\chi^{(2)}_{xyz}$ for zinc-blende crystals, including transitions from the highest valence band (v) to the lowest conduction band (c) with the second lowest conduction band (l) as intermediate step, including excitonic effects, is presented and solved analytically. This yields a closed-form* expression:

$$\chi^{(2)}_{xyz} = \chi^{(2)}_0 \Big[S(\Omega_2, \Omega_1) + S(-\Omega_1, -\Omega_2) + S(-\Omega_1, \Omega_1)$$

*in so far the digamma function is taken as a closed-form term



(a)

Figure 5.6: Expected dipole bulk contribution for GaAs.

$$+S(-\Omega_2,-\Omega_1)+S(\Omega_1,\Omega_2)+S(\Omega_1,-\Omega_1)\Big], \tag{5.9}$$

where Ω_n is the complex photon energy $\Omega_n = n\hbar\omega + i\Gamma$, Γ being the usual phenomenological broadening factor. S is given by

$$S(a,b) = -\frac{\sqrt{RyE_0^5}}{3}\left\{\frac{1}{a\Delta(\Delta-b)}g\left(\sqrt{\frac{Ry}{E_g}}\right)+\frac{1}{b\Delta(\Delta+a)}g\left(\sqrt{\frac{Ry}{\tilde{E}_g}}\right)\right.$$
$$\left.-\frac{1}{a(\Delta+a)(\Delta+a-b)}g\left(\sqrt{\frac{Ry}{E_g-a}}\right)-\frac{1}{b(\Delta-b)(\Delta+a-b)}g\left(\sqrt{\frac{Ry}{\tilde{E}_g-b}}\right),\right\}$$

where $g(x) = 2F(x) + 2\ln(x) + x^{-1}$ and F is the digamma function, which can be calculated from the series

$$F(x) = -\gamma + \sum_{n=0}^{\infty}\frac{x-1}{(n+1)(n+x)}, \quad \gamma = 0.57721\ldots \text{ (Euler-Mascheroni constant)}.$$

$\chi_0^{(2)}$ is given by

$$\chi_0^{(2)} = i\frac{\sqrt{18}e^3\mu 3/2}{\pi\varepsilon_0 m^3 E_0^{5/2}\Delta}p_{vc}^x p_{cl}^y p_{lc}^z,$$

where $\mu = m_e m_h/(m_e+m_h)$ is the reduced mass, $E_0^{5/2} = \sqrt{E_g\tilde{E}_g}(\sqrt{E_g}+\sqrt{\tilde{E}_g})^3$, and $\tilde{E}_g = E_g + \Delta$. To model the E1 transitions, which is due to transitions from the highest valence band to the lowest conduction band near L in the Brillouin zone, the parameters $\Delta = 3$eV, $E_g = 2.9$eV and $Ry = 4.2$meV given in the orginal article is used. Furthermore, the effective masses of all three bands are taken to be $m_h = m_e = 0.61$, agreeing well with the EPM band structures2.3. The one-electron momentum matrix elements is also calculated from the EPM yielding $p_{vc}^x p_{cl}^y p_{lc}^z = 0.0275 \times 2\pi\hbar/a$, where a factor of four have been included due to the eight equivalent L-points of the first Brillouin zone, each shared with one

neighbouring zone. The result is also plotted in Fig. 5.6(a), where the lineshape is seen to agree much better with experiment, although a constant difference between the one-electron theoretical results and those calculated by (5.9) is observed. It is noted that the theoretical values calculated here yielded values about half the values reported by W. Daum *et al.* from experiments (see Fig. 3.4) - a discrepency which is apparantly resolved by including excitonic effects. Thus, including excitonic effects seems to be important if good agreement between theory and experiment is to be achieved, as is the case in linear optics. The large signal measured below 1.25eV is not a feature of the electronic joint density of states, as is easily verified by making $\chi^{(2)}$ plots neglecting the momentum matrix elements in the theory of SHG (not shown), presented ealier. The depletion-electric-field-induced spectra of Fig. 3.10 displayed an enhancement of the E0 peak relative to the E1 transition, however, the experimental feature is to far removed from the theoretical curve to fully explain it. Another point worth noting is, that the feature near 1.1eV appear in all experimental measurements. Thus, in future experiements, it should be tested whether this peak scales qruadratically with the applied field, as expected for a second harmonic signal.

5.5 Perspectives in Second Harmonic Generation

Second harmonic generation applied as a spectrocopic tool seems to hold great potential in characterisation of electronic properties of matter, owing to the great sensitivity of the second harmonic spectra on energy differences due to the near-singular energy fractions entering theoretical expressions. This is readily apparant when comparing theoretical results produced from slightly different band structures, which all produce almost the same linear spectra but rather different second harmonic spectra. However, at present, no theory (to the authors knowledge) produce results consistently in quantitative agreement with experimental results, even using advanced DFT methods including quasiparticle effects[33]. Qualitative information can be compared, however, such as SHG peak positions. Additionally, the extreme dependence of second harmonic spectra on the symmetry of crystalline samples provides a convinient method of e.g. monitoring crystalline growth processes. If measurements are performed on substrates absent of bulk second harmonic generation, due to e.g. inversion symmetry, a measured second harmonic signal must originate either from higher order terms, such as the so-called quadrupole effects arising from including spatial dispersion, or from surface regions where the bulk symmetry is broken. Here, however, depletion-electric-field-induced SHG gives contributions as do the first few atomic layers of the surface. In doped GaAs, Germer *et al.*[45] show that the depletion-field-induced SHG actually dominate the bulk signal, and in that reference it is also suggested than photmodulation experiments, where free carriers are injected by optical pumping, should be combined with SHG experiments. This is due to the fact that an increased doping concentration will change the properties of the surface depletion layer, as was also found by the simple model presented previously in this work, allowing control of a paramter which explicitly relates to the depletion-field-induced SHG. All surffface specific quantities will, in general, be affected by e.g. adsorbate molecules or compounds involved in growth processes making the interpretation of surface SHG spectroscopic experiements demanding in theory, however potentially rewarding in the amount of information which can be extracted from a sample. A great advantage of SHG surface characterisation is, that it is a non-destructive method.

6 Conclusion

In this work, the linear optical properties within the one-electron, band theoretical framework have been modeled with results agreeing well with other sources, however with a few discrepancies relative to experiment, most notably due to many particle effects, such as electron-phonon interaction and excitonic effects. Including spin-orbit interaction in the empirical pseudopotential model causes band splitting, which translates to a splitting of the E1 peak in the optical spectra. The inclusion of the optical wave vector in a non-local theory of the linear absorption was modeled, and the effects were shown to have potential ramifications within areas such as nano plasmonics, where localized electric fields couple states differing in k-vector within the Brillouin zone. The linear optical response of free-standing Ge and Si thin-films was modeled using an empirical pseudopotential model, where the introduction of an extra empirical pseudopotential correcting the ionization energies of the thin-films was necessary. Band structures in good agreement with other theoretical calculations was produced. The microscopic theory of second harmonic generation was presented, and spectra both including and neglecting spin orbit effects was calculated. Qualitative agreement between theory and experiment can be concluded for GaAs in the spectral range experimentally available, in that peak positions and general shape is well reproduced, although with large intensity differences. Based on the theory of T.G. Pedersen and H. Cornean, it was realized that excitonic effects may explain this general intensity difference between experiment and theory for some parts of the spectrum. Second harmonic generation was measured on a semi-insulating GaAs sample with reproduction of the E1 peak, although large SHG signals was observed in spectral regions where the bulk theory predict no activity, suggesting surface specific effects. Surface depletions layers of intrinsic semiconductors is indeed shown to extend several hundred microns into the sample, making possible depletion-field-induced second harmonic generation from a spatial region of bulk size. In addition, the effects of the spatial dispersion (i.e. including the optical wave vector) in the calculations was shown to yield small nonlinear susceptibility tensor elements for Si and Ge for fundamental wavelengths below the direct band gap, for wave vectors following plane-wave dispersion. The LATM was found to be an extremely versatile integration method, ensuring fast convergence of δ-function integrals and vastly outperforming random sampling integration techniques.

Bibliography

[1] S. ADLER, *Phys. Rev.* **126**, 413 (1961).

[2] H. EHRENREICH and M. COHEN, *Phys. Rev.* **115**, 786 (1959).

[3] P. NOZIÉRES and D. PINES, *Phys. Rev.* **109**, 741 (1958).

[4] M. COHEN and T. BERGSTRESSER, Phys. Rev. **141**, 789 (1966).

[5] W. SASLOW, T. BERGSTRESSER, C. FONG, and M. COHEN, *Solid State Communications* **5**, 667 (1967).

[6] J. WALTER and M. COHEN, *Phys. Rev. B* **1**, 2661 (1970).

[7] M. SCHLÜTER, J. CHELIKOWSKY, S. LOUIE, and M. COHEN, *Phys. Rev. B* **12**, 4200 (1975).

[8] P. FRANKEN, A. HILL, C. PETERS, and G. WEINREICH, *Phys. Rev. Lett.* **7**, 118 (1961).

[9] N. BLOEMBERGEN and P. PERSHAN, *Phys. Rev.* **128**, 606 (1962).

[10] J. ARMSTRONG, N. BLOEMBERGEN, J. DUCUING, and P. PERSHAN, *Phys. Rev.* **127**, 1918 (1962).

[11] D. BETHUNE, A. SCHMIDT, and Y. SHEN, *Phys. Rev. B* **11**, 3867 (1975).

[12] E. GHARAHMANI, D. MOSS, and J. SIPE, *Phys. Rev. B* , 8990 (1991).

[13] J. HUGHES and J. SIPE, *Phys. Rev. B* **53**, 751 (1996).

[14] M. L. TROLLE, *Technical report, University of Aalborg* (2010), Available from university project library or by request from author (mlt07@student.aau.dk).

[15] A. MACDONALD, S. VOSKO, and P. COLERIDGE, *J. Phys. C: Solid State Phys* **12**, 2991 (1979).

[16] P. BASU, *Theory of Optical Processes in Semiconductors*, Clarendon Press - Oxford, 1997, ISBN: 0-19-852620-2.

[17] M. BALANSKI and R. WALLIS, *Semiconductor Physics and Applications*, Oxford University Press, 2000, ISBN: 978-0-19-851740-5.

[18] L. WANG and A. ZUNGER, J. Phys. Chem. **98**, 2158 (1994).

[19] IBM DAMOCLES, `http://www.research.ibm.com/DAMOCLES/html_files/numerics.html#table1`.

[20] D. ASPNES and A. STUDNA, *Phys. Rev. B* **27**, 985 (1983).

[21] E. PALIK, *Handbook of Optical Constants of Solids*, Academic Press Inc., 1985, ISBN: 0-12-544420-6.

[22] J. CHELIKOWSKY and M. COHEN, *Phys. Rev. B* **14**, 556 (1976).

[23] J. DAHL, *The Quantum World of Atoms and Molecules*, World Scientific, 2001, ISBN: 981-02-4565-3.

[24] W. PRESS, B. FLANNERY, S. TEUKOLSKY, and W. VETTERLING, *Numerical Recipes - The Art of Scientific Computing*, Cambridge University Press, 1989, ISBN: 0-521-38330-7.

[25] I. GORCZYCA, P. PFEFFER, and W. ZAWADZKI, *Semicond. Sci. Technol.* **6**, 963 (1991).

[26] J. BURKE, *Phys. Rev. B* **33**, 5186 (1986).

[27] P. JOHNSON and R. CHRISTY, *Phys. Rev. B* **6**, 4370 (1972).

[28] J. JUNG, M. L. TROLLE, T. SØNDERGAARD, K. PEDERSEN, and T. G. PEDERSEN, *Phys. Rev. Lett.* (2011), To be published.

[29] T. G. PEDERSEN and K. PEDERSEN, *Phys. Rev. B* **79**, 035422 (2009).

[30] D. MOSS, J. SIPE, and H. VAN DRIEL, *Phys. Rev. B* **36**, 1153 (1987).

[31] S. RASHKEEV, W. LAMBRECHT, and B. SEGALL, *Phys. Rev. B* **57**, 3905 (1997).

[32] D. MOSS, E. GHARAHMANI, J. SIPE, and H. VAN DRIEL, *Phys. Rev. B* , 1542 (1990).

[33] S. BERGFELD and W. DAUM, *Phys. Rev. Lett.* **90**, 036801 (2003).

[34] J. SIPE, D. MOSS, and H. VAN DRIEL, *Phys. Rev. B* , 1129 (1987).

[35] H. LÜTH, *Surfaces and Interfaces of Solid Materials*, Springer, 1998, ISBN: 3-540-58576-1.

[36] T. G. PEDERSEN and L. DIEKHÖNER, *Electric, Optical and Magnetic Properties of Nanostructures.*, Aalborg University, 2009.

[37] K. KIKUCHI and K. TADA, *Opt. Quant. Elect* **12**, 199 (1980).

[38] T. G. PEDERSEN, K. PEDERSEN, and T. B. KRISTENSEN, *Phys. Rev. B* **63**, 201101 (2001).

[39] S. L. CUNNINGHAM, *Phys. Rev. B* **4988**, 10 (1974).

[40] L. REINING, R. D. SOLE, M. CINI, and J. PING, *Phys. Rev. B* **50**, 8411 (1994).

[41] M. DOWNER, B. MENDOZA, and V. GAVRILENKO, *Surf. Interface Anal.* **31**, 966 (2001).

[42] P. GUYOT-SIONNEST, W. CHEN, and Y. SHEN, *Phys. Rev. B* **33**, 8254 (1986).

[43] R. BOYD, *Nonlinear Optics*, Elsevier, 2008, ISBN: 978-0-12-369470-6.

[44] T. G. PEDERSEN and H. CORNEAN, *Europhys. Lett.* **78**, 27005 (2007).

[45] T. A. GERMER, K. W. KOLASIN-ACUTESKI, J. C. SPEPHENSOM, and L. J. RICHTER, *Phys. Rev. B* **55**, 10694 (1997).

[46] D. MOSSS, E. GHAHRAMANI, J. SIPE, and H. DRIEL, *Phys. Rev. B* **41**, 1542 (1990).

[47] D. MCQUARRIE, *Mathemathical methods for scientists and engineers*, University science books, 2003, ISBN: 1-891389-24-6.

Optical Response Tensors

In the following, the optical response of semiconductors due to a spatially varying field will be modelled using perturbation theory, arriving at expressions for the q-dependent susceptibility. First, the general expressions of perturbation theory will be derived, describing the first and second order response to an arbitrary applied perturbation.

A.1 Perturbative solution to the dynamic Schrödinger equation

The band structure of a solid is often said to contain all the information necessary to explain most of its electronic and optical properties. Since the band structure represents solutions to the stationary Schrödinger equation, while an optical response is dynamic, this postulate seems paradoxial. However, the interaction of an electron with other electrons and ions of the crystal are often much stronger than its interaction with an applied field – thus the stationary states represent a good starting point for a description of the dynamic system. In any case, the stationary solutions represent a complete set in the mathemathical sense, and considering that they intrinsically capture much of the physical behaviour of the system, it seems reasonable to use them as a basis for solving the time dependent Schrödinger equation. This is the underlying rationale of time-dependent perturbation theory, which will be reviewed in the following based on the book by Thomas Garm Pedersen[36], but moving beyond the dipole approximation of that text and generalising to non-linear theory.

Full knowledge of the stationary system described by

$$\hat{H}_0|n\rangle = E_n|n\rangle$$

is assumed, with $|n\rangle$ and E_n found by e.g. the EPM. For the system under investigation, the stationary Hamiltonian $\hat{H}_0$ is perturbed by some external potential, making the full, time dependent Hamiltonian

$$\hat{H} = \hat{H}_0 + \frac{1}{2}\left[\hat{H}_1 e^{-i\omega t} + c.c.\right].$$

The task is then to solve the dynamic Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\psi(\vec{r},t) = \hat{H}\psi(\vec{r},t). \tag{A.1}$$

The main ansatz of perturbation theory is to assume that the stationary states span the solution set of (A.1), $\psi(\vec{r},t) = \sum_n a_n|n\rangle\ \exp(-i\omega_n t)$, leading to

$$i\hbar\frac{\partial}{\partial t}\sum_n a_n|n\rangle e^{-i\omega_n t} = \sum_n a_n\left[\hat{H}_0 + \frac{1}{2}\left[\hat{H}_1 e^{-i\omega t} + c.c.\right]\right]|n\rangle e^{-i\omega_n t},$$

where a_n is a time dependent expansion coefficient. Performing the time derivative on the left side and noting that $\hat{H}_0|n\rangle = \hbar\omega_n|n\rangle$ on the right

$$i\hbar\sum_n \frac{\partial a_n}{\partial t}|n\rangle e^{-i\omega_n t} = \frac{1}{2}\sum_n a_n\left[\hat{H}_1 e^{-i\omega t} + c.c.\right]|n\rangle e^{-i\omega_n t}.$$

The inner product with $|m\rangle$ is taken

$$i\hbar\sum_n \frac{\partial a_n}{\partial t}\langle m|n\rangle e^{-i\omega_n t} = \frac{1}{2}\sum_n a_n\left[\langle m|\hat{H}_1|n\rangle e^{-i\omega t} + \langle m|\hat{H}_1^\dagger|n\rangle e^{i\omega t}\right]e^{-i\omega_n t},$$

and using the orthonomalisation condition $\langle m|n\rangle = \delta_{m,n}$

$$\frac{\partial a_m}{\partial t} = \frac{1}{i\hbar 2}\sum_n a_n\left[\langle m|\hat{H}_1|n\rangle e^{-i\omega t} + \langle m|\hat{H}_1^\dagger|n\rangle e^{i\omega t}\right]e^{i\omega_{mn} t},$$

where $\omega_{mn} = \omega_m - \omega_n$. The coeffients a_m is now expanded in powers of the perturbation

$$a_m = \sum_{p=0}^{\infty} a_m^{(p)}$$

and terms of like powers in p are set equal to each other, noting that $\hat{H}_1$ and $\hat{H}_1^\dagger$ represent first powers in perturbation (p)

$$\frac{\partial a_m^{(p)}}{\partial t} = \frac{1}{i\hbar 2}\sum_n a_n^{(p-1)}\left[\langle m|\hat{H}_1|n\rangle e^{-i\omega t} + \langle m|\hat{H}_1^\dagger|n\rangle e^{i\omega t}\right]e^{i\omega_{mn} t}.$$

The case of $p = 0$ represents the stationary case, hence

$$\frac{\partial a_m^{(0)}}{\partial t} = 0.$$

Using this, expressions for increasing p's can be itterated

$$\frac{\partial a_m^{(1)}}{\partial t} = \frac{1}{2i\hbar}\sum_n a_n^{(0)}\left[\langle m|\hat{H}_1|n\rangle e^{-i\omega t} + \langle m|\hat{H}_1^\dagger|n\rangle e^{i\omega t}\right]e^{i\omega_{mn} t}.$$

Integrating gives $a_m^{(1)}$

$$a_m^{(1)} = -\frac{1}{2}\sum_n a_n^{(0)}\left[\frac{\langle m|\hat{H}_1|n\rangle e^{-i\omega t}}{E_{mn} - \hbar\omega} + \frac{\langle m|\hat{H}_1^\dagger|n\rangle e^{i\omega t}}{E_{mn} + \hbar\omega}\right]e^{i\omega_{mn} t}. \tag{A.2}$$

For $p = 2$, introducing the notation $H_{mn} = \langle m|\hat{H}_1|n\rangle$

$$\begin{aligned}\frac{\partial a_m^{(2)}}{\partial t} &= \frac{1}{2i\hbar}\sum_n a_n^{(1)}\left[\langle m|\hat{H}_1|n\rangle e^{-i\omega t} + \langle m|\hat{H}_1^\dagger|n\rangle e^{i\omega t}\right]e^{i\omega_{mn} t} \\ &= \frac{1}{4i\hbar}\sum_{n,p} a_p^{(0)}\left[\frac{H_{mn}H_{np}e^{i(\omega_{np}+\omega_{mn}-2\omega)t}}{E_{np} - \hbar\omega} + \frac{H_{mn}H_{np}^* e^{i(\omega_{np}+\omega_{mn})t}}{E_{np} + \hbar\omega}\right. \\ &\quad \left. + \frac{H_{mn}^* H_{np} e^{i(\omega_{np}+\omega_{mn})t}}{E_{np} - \hbar\omega} + \frac{H_{mn}^* H_{np}^* e^{i(\omega_{np}+\omega_{mn}+2\omega)t}}{E_{np} - \hbar\omega}\right]\end{aligned}$$

and integrating

$$\begin{aligned}a_m^{(2)} = \frac{1}{4}\sum_{n,p} a_p^{(0)}\left[\frac{H_{mn}H_{np}e^{-2i\omega t}}{(E_{np} - \hbar\omega)(E_{mp} - 2\hbar\omega)} + \frac{H_{mn}H_{np}^*}{(E_{np} + \hbar\omega)E_{mp}}\right. \\ \left. + \frac{H_{mn}^* H_{np}}{(E_{np} - \hbar\omega)E_{mp}} + \frac{H_{mn}^* H_{np}^* e^{2i\omega t}}{(E_{np} + \hbar\omega)(E_{mp} + 2\hbar\omega)}\right]e^{i\omega_{mp} t}.\end{aligned} \tag{A.3}$$

We are interested in an observabel X, described by the operator $\hat{X}$. Thus the expectation value of X is considered

$$\langle X\rangle = \langle\psi|\hat{X}|\psi\rangle = \sum_{q,r}\left(a_q^{(0)*}+a_q^{(1)*}+a_q^{(2)*}+...\right)\langle q|\hat{X}|r\rangle\left(a_r^{(0)}+a_r^{(1)}+a_r^{(2)}+...\right)e^{i\omega_{qr}t}. \tag{A.4}$$

Considering terms up to second order in perturbation

$$\langle X\rangle = \sum_{q,r}\left(a_q^{(0)*}a_r^{(0)}+a_q^{(1)*}a_r^{(0)}+a_q^{(0)*}a_r^{(1)}+a_q^{(1)*}a_r^{(1)}+a_q^{(2)*}a_r^{(0)}+a_q^{(0)*}a_r^{(2)}\right)\langle q|\hat{X}|r\rangle e^{i\omega_{qr}t}$$

and seperating constant, linear and nonlinear terms

$$\langle X\rangle = \langle X\rangle^{(0)}+\langle X\rangle^{(1)}+\langle X\rangle^{(2)}$$

It can be shown that

$$a_q^{(0)*}a_r^{(0)} = \delta_{q,r}f(E_q),$$

$f(E_q)$ being the Fermi-Dirac distribution function (in the case of electrons)[43].

The first order contribution becomes by (A.2)

$$\langle X\rangle^{(1)} = -\frac{1}{2}\sum_{q,r}f(E_r)X_{rq}\left[\frac{H_{qr}e^{-i\omega t}}{E_{qr}-\hbar\omega}+\frac{H_{qr}^*e^{i\omega t}}{E_{qr}+\hbar\omega}\right]-\frac{1}{2}\sum_{q,r}f(E_q)X_{rq}\left[\frac{H_{qr}^*e^{i\omega t}}{E_{qr}-\hbar\omega}+\frac{H_{qr}e^{-i\omega t}}{E_{qr}+\hbar\omega}\right] \tag{A.5}$$

Considering the second order response by using (A.3)

$$\begin{aligned}\langle X\rangle^{(2)} = \frac{1}{4}\sum_{q,r,n}\Bigg\{&f(E_n)\left[\frac{H_{qn}^*H_{rn}}{(E_{qn}-\hbar\omega)(E_{rn}-\hbar\omega)}+\frac{H_{qn}^*H_{rn}^*e^{2i\omega t}}{(E_{qn}-\hbar\omega)(E_{rn}+\hbar\omega)}\right.\\&\left.+\frac{H_{qn}H_{rn}e^{-2i\omega t}}{(E_{qn}+\hbar\omega)(E_{rn}-\hbar\omega)}+\frac{H_{qn}H_{rn}^*}{(E_{qn}+\hbar\omega)(E_{rn}+\hbar\omega)}\right]\\&+f(E_q)\left[\frac{H_{rn}H_{nq}e^{-2i\omega t}}{(E_{nq}-\hbar\omega)(E_{rq}-2\hbar\omega)}+\frac{H_{rn}^*H_{nq}^*e^{2i\omega t}}{(E_{nq}+\hbar\omega)(E_{rq}+2\hbar\omega)}\right.\\&\left.+\frac{H_{rn}H_{nq}^*}{(E_{nq}+\hbar\omega)E_{rq}}+\frac{H_{rn}^*H_{nq}}{(E_{nq}-\hbar\omega)E_{rq}}\right]\\&+f(E_r)\left[\frac{H_{nr}^*H_{qn}^*e^{2i\omega t}}{(E_{nr}-\hbar\omega)(E_{qr}-2\hbar\omega)}+\frac{H_{qn}H_{nr}e^{-2i\omega t}}{(E_{nr}+\hbar\omega)(E_{qr}+2\hbar\omega)}\right.\\&\left.+\frac{H_{nr}^*H_{qn}}{(E_{nr}-\hbar\omega)E_{qn}}+\frac{H_{nr}H_{qn}^*}{(E_{nr}+\hbar\omega)E_{qn}}\right]\Bigg\}X_{qr}.\end{aligned}$$

Considering only second harmonic terms, i.e. terms proportional to $e^{-i2\omega t}$

$$\begin{aligned}&\frac{1}{4}\sum_{q,r,n}\left\{\frac{f(E_n)}{(E_{qn}+\hbar\omega)(E_{rn}-\hbar\omega)}+\frac{f(E_q)}{(E_{nq}-\hbar\omega)(E_{rq}-2\hbar\omega)}\right.\\&\quad\left.+\frac{f(E_r)}{(E_{nr}+\hbar\omega)(E_{qr}+2\hbar\omega)}\right\}X_{qr}H_{rn}H_{nq}e^{-2i\omega t}\\&=\frac{1}{4}\sum_{q,r,n}\frac{1}{2\hbar\omega-E_{rq}}\left\{\frac{f(E_n)(2\hbar\omega-E_{rq})}{(E_{qn}+\hbar\omega)(E_{rn}-\hbar\omega)}+\frac{f(E_q)}{\hbar\omega-E_{nq}}\right.\\&\quad\left.+\frac{f(E_r)}{E_{nr}+\hbar\omega}\right\}X_{qr}H_{rn}H_{nq}e^{-2i\omega t}\end{aligned} \tag{A.6}$$

Noting that

$$\begin{aligned}\frac{(2\hbar\omega - E_{rq})}{(E_{qn}+\hbar\omega)(E_{rn}-\hbar\omega)} &= -\frac{(\hbar\omega+E_{qn})+(\hbar\omega-E_{rn})}{(E_{qn}+\hbar\omega)(\hbar\omega-E_{rn})} \\ &= -\left(\frac{1}{\hbar\omega-E_{rn}}+\frac{1}{\hbar\omega-E_{nq}}\right),\end{aligned}$$

(A.6) can be simplified to

$$\frac{1}{4}\sum_{q,r,n}\frac{1}{2\hbar\omega-E_{rq}}\left\{\frac{f_{qn}}{\hbar\omega-E_{nq}}+\frac{f_{rn}}{\hbar\omega-E_{rn}}\right\}X_{qr}H_{rn}H_{nq}e^{-2i\omega t},$$

where the notation $f_{mn} = f(E_m) - f(E_n)$ have been used. Thus, by defining a new function G, shifting indices so $(ijk) \sim (qrn)$ to avoid confusion with spatial coordinate $\vec{r}$ and wave vector $\vec{q}$,

$$G_{ijk}(\alpha,\beta) \equiv \frac{1}{\hbar(\omega_\alpha+\omega_\beta)-E_{ji}}\left(\frac{f_{ik}}{\hbar\omega_\beta-E_{ki}}+\frac{f_{jk}}{\hbar\omega_\alpha-E_{jk}}\right),$$

$\langle X\rangle^{(2)}$ can be written on a compressed form

$$\langle X\rangle^{(2)} = \frac{1}{4}\sum_{i,j,k}\sum_{\alpha,\beta}G_{ijk}(\alpha,\beta)X_{ij}H^{(\alpha)}_{jk}H^{(\beta)}_{ki}e^{-i(\omega_\alpha+\omega_\beta)t} \tag{A.7}$$

where the sum over α and β runs over 1 and -1 with $\omega_1 = -\omega_{-1} = \omega$. A superscript of -1 on the Hamiltonian matrix elements are understood to complex conjugate it, while a superscript of 1 does nothing. (A.7) is exactly the expression derived in the appendix of Moss *et al.* (1990)[46], although they used the density matrix formalism. Considering a decomposition of the response into frequency components

$$\langle X\rangle = X_0 + \frac{1}{2}\left[X_\omega e^{-i\omega t}+X^*_\omega e^{i\omega t}\right]+\frac{1}{4}\left[X_{2\omega}e^{-2i\omega t}+X^*_{2\omega}e^{2i\omega t}\right] \tag{A.8}$$

it becomes clear that the response oscillating at ω to a field of frequency ω becomes

$$X_\omega = -\sum_{m,n}f_{nm}\frac{\langle m|\hat{H}_1|n\rangle\langle n|\hat{X}|m\rangle}{E_{mn}-\hbar w} \tag{A.9}$$

where the frequency have been substituted with the complex frequency $w = \omega + i\Gamma$, Γ being a phenomenological dampening factor and the indices have been changed. The response oscillating at 2ω becomes

$$X_{2\omega} = \sum_{i,j,k}G_{ijk}X_{ij}H_{jk}H_{ki}, \tag{A.10}$$

where the same substitution of $\omega \to w$ have been performed in G_{ijk} and where G_{ijk} is understood to be $G_{ijk}(1,1)$.

A.2 Electromagnetic fomalism

The electromagnetic formalism used throughout this appendix is reviewed; a time harmonic electric field and its vector potential can be written as

$$\begin{aligned}\vec{\mathcal{E}}(\vec{r},t) &= \frac{1}{2}\left[\vec{\mathcal{E}}_\omega(\vec{r})e^{-i\omega t}+c.c.\right], \\ \vec{\mathcal{A}}(\vec{r},t) &= \frac{1}{2}\left[\vec{\mathcal{A}}_\omega(\vec{r})e^{-i\omega t}+c.c.\right],\end{aligned}$$

where $c.c.$ denotes "complex conjugate". The spatial dependence of these fields can be Fourier expanded

$$\vec{\mathcal{E}}_\omega(\vec{r}) = \sum_{\vec{q}} \vec{\mathcal{E}}_{\omega,\vec{q}} e^{i\vec{q}\cdot\vec{r}}, \tag{A.11}$$
$$\vec{\mathcal{A}}_\omega(\vec{r}) = \sum_{\vec{q}} \vec{\mathcal{A}}_{\omega,\vec{q}} e^{i\vec{q}\cdot\vec{r}}.$$

The electric and magnetic induction fields are uniquely defined through their relations with the vector potential $\vec{\mathcal{A}}$

$$\vec{\mathcal{E}}(\vec{r},t) = -\frac{\partial\vec{\mathcal{A}}(\vec{r},t)}{\partial t}, \tag{A.12}$$
$$\vec{B}(\vec{r},t) = \nabla\times\vec{\mathcal{A}}(\vec{r},t), \tag{A.13}$$
$$\nabla\cdot\vec{\mathcal{A}}(\vec{r},t) = 0, \tag{A.14}$$

where the first equation means that

$$\vec{\mathcal{E}}_\omega(\vec{r}) = i\omega\vec{\mathcal{A}}_\omega(\vec{r}), \tag{A.15}$$

and the last equation is the Coulomb gauge. The linear current and polarization response is in the frequency domain described by the conductivity $\hat{\sigma}_\omega(\vec{r})$ and the electric susceptibility $\hat{\chi}_\omega(\vec{r})$, respectively

$$\vec{J}_\omega(\vec{r}) = \hat{\sigma}_\omega(\vec{r})\cdot\vec{\mathcal{E}}_\omega(\vec{r}) \tag{A.16}$$
$$\vec{\mathcal{P}}_\omega(\vec{r}) = \varepsilon_0\hat{\chi}_\omega(\vec{r})\cdot\vec{\mathcal{E}}_\omega(\vec{r}). \tag{A.17}$$

The response functions are defined as integral operators

$$\hat{\sigma}_\omega(\vec{r})\cdot\vec{\mathcal{E}}_\omega(\vec{r}) = \int \sigma(\vec{r},\vec{r}')\cdot\vec{\mathcal{E}}(\vec{r}')d^3r'$$
$$\hat{\chi}_\omega(\vec{r})\cdot\vec{\mathcal{E}}_\omega(\vec{r}) = \int \chi(\vec{r},\vec{r}')\cdot\vec{\mathcal{E}}(\vec{r}')d^3r'.$$

and tensors allowing description non-local and anisotropic response, respectively. When the material is source free and non-magnetic, the time rate of change of the polarisation is the only contribution to the current density

$$\vec{J}(\vec{r},t) = \frac{\partial\vec{\mathcal{P}}(\vec{r},t)}{\partial t}$$
$$\Rightarrow\vec{J}_\omega(\vec{r}) = -i\omega\vec{\mathcal{P}}_\omega(\vec{r}). \tag{A.18}$$

A general expression for the time-averaged absorbed power due to a time harmonic field is

$$P_\omega = \frac{\omega}{2}\mathrm{Im}\left[\int \vec{\mathcal{P}}_\omega(\vec{r})\cdot\vec{\mathcal{E}}^*_\omega(\vec{r})d^3r\right] \tag{A.19}$$

With this, the electromagnetic relations for a non-magnetic,linear material is concluded.

A.3 Linear optical response

We are interested in calculating the response functions introduced in the last section within the framwork of linear response theory, hence the full Hamiltonian of the problem is considered

$$\hat{H} = \frac{\left[\hat{\vec{p}}_e + e\vec{\mathcal{A}}(\vec{r}_e,t)\right]^2}{2m} + V(\vec{r}_e) = \hat{H}_0 + \frac{1}{2}\left[\hat{H}_1 e^{-i\omega t} + \hat{H}_1^\dagger e^{i\omega t}\right] + \hat{H}_2,$$

where e as a subscript denotes the electron coordinate but the (positive) fundamental charge quantum otherwise, and

$$\begin{aligned}
\hat{H}_0 &\equiv \frac{\hat{p}_e^2}{2m} + V(\vec{r}_e), \\
\hat{H}_1 &\equiv \frac{e}{2m}\left[\vec{\mathcal{A}}_\omega(\vec{r}_e)\cdot\hat{\vec{p}}_e + \hat{\vec{p}}_e\cdot\vec{\mathcal{A}}_\omega(\vec{r}_e)\right] \\
\hat{H}_2 &\equiv \text{ Terms of second order in } \mathcal{A}_\omega
\end{aligned}$$

The complete solution set $|n\rangle$ to the unperturbed Hamiltonian $\hat{H}_0$ can be found using e.g. the (EPM). The frequency dependent current density response is found by first order perturbation theory (A.9) as

$$\vec{J}_\omega(\vec{r}) = -\sum_{m,n} f_{nm}\frac{\langle m|\hat{H}_1|n\rangle\langle n|\hat{\vec{J}}|m\rangle}{E_{mn} - \hbar w} \tag{A.20}$$

At sufficiently low temperatures (room temperature)

$$f_{nm} \approx 0, \quad \begin{matrix}\{m,n\}\in v \\ \{m,n\}\in c\end{matrix}$$

with c and v denoting states above and below the Fermi level, respectively. With this, (A.20) can be rewritten

$$\vec{J}_\omega(\vec{r}) = -2\sum_{c,v} E_{c,v} f_{v,c}\frac{\langle c|\hat{H}_1|v\rangle\langle v|\hat{\vec{J}}|c\rangle}{E_{c,v}^2 - \hbar w^2} \tag{A.21}$$

The classical current density for an electronic point charge
$\vec{J}(\vec{r},t) = -e\vec{p}(\vec{r},t)\delta\left(\vec{r}-\vec{r}_e\right)/m$ is in quantum mechanics transformed to the operator

$$\hat{\vec{J}}(\vec{r}) = -\frac{e}{2m}\left[\delta(\vec{r}-\vec{r}_e)\hat{\vec{p}}_e + \hat{\vec{p}}_e\delta(\vec{r}-\vec{r}_e)\right].$$

The matrix elements are evaluated using states written on the Bloch form

$$\phi_{\alpha\vec{k}} = |\alpha\vec{k}\rangle = \frac{1}{\sqrt{\Omega}}u_{\alpha\vec{k}}(\vec{r}_e)e^{i\vec{k}\cdot\vec{r}_e} = \frac{1}{\sqrt{\Omega}}\sum_{\vec{G}} f_\alpha(\vec{k}-\vec{G})e^{i(\vec{k}-\vec{G})\cdot\vec{r}_e}. \tag{A.22}$$

c and v shall henceforth index conduction and valence bands, respectively. The current

density matrix element:

$$\begin{aligned}
\langle v\vec{k}'|\hat{\vec{J}}(\vec{r})|c\vec{k}\rangle &\equiv \vec{J}_{v\vec{k}',c\vec{k}}(\vec{r}) \\
&= \frac{i\hbar e}{2m}\int \left[\phi^*_{v\vec{k}'}\delta(\vec{r}-\vec{r}_e)\nabla_e\phi_{c\vec{k}} + \phi^*_{v\vec{k}'}\nabla_e\delta(\vec{r}-\vec{r}_e)\phi_{c\vec{k}}\right] d^3r_e \\
&= \frac{i\hbar e}{2m}\int \left[\delta(\vec{r}-\vec{r}_e)\phi^*_{v\vec{k}'}(\vec{r}_e)\nabla_e\phi_{c\vec{k}}(\vec{r}_e) - \delta(\vec{r}-\vec{r}_e)\phi_{c\vec{k}}(\vec{r}_e)\nabla_e\phi^*_{v\vec{k}'}(\vec{r}_e)\right] d^3r_e \\
&= \frac{i\hbar e}{2m\Omega}\left[iu^*_{v\vec{k}'}(\vec{r})u_{c\vec{k}}(\vec{r})(\vec{k}+\vec{k}') + u^*_{v\vec{k}'}(\vec{r})\nabla u_{c\vec{k}}(\vec{r}) - u_{c\vec{k}}(\vec{r})\nabla u^*_{v\vec{k}'}(\vec{r})\right] e^{i(\vec{k}-\vec{k}')\cdot\vec{r}} \\
&\equiv \frac{i\hbar e}{2m\Omega}\vec{j}_{v\vec{k}',c\vec{k}}(\vec{r})e^{i(\vec{k}-\vec{k}')\cdot\vec{r}} \qquad \text{(A.23)}
\end{aligned}$$

$\vec{j}_{v\vec{k}',c\vec{k}}(\vec{r})$ can be simplified by expanding the lattice periodic parts of the wave functions as pr. (A.22), yielding

$$\begin{aligned}
\vec{j}_{v\vec{k}',c\vec{k}}(\vec{r}) &= \sum_{\vec{G},\vec{G}'} if^*_v(\vec{k}'-\vec{G}')f_c(\vec{k}-\vec{G})\left[\vec{k}+\vec{k}'-\vec{G}'-\vec{G}\right]e^{i(\vec{G}'-\vec{G})\cdot\vec{r}} \\
&\equiv \sum_{\vec{G},\vec{G}'} j_{v\vec{k}',c\vec{k}}(\vec{G},\vec{G}')e^{i(\vec{G}'-\vec{G})\cdot\vec{r}}
\end{aligned}$$

The perturbing Hamiltonian matrix element:

$$\begin{aligned}
\langle c\vec{k}|\hat{H}_1|v\vec{k}'\rangle &= \frac{e}{2m}\int \phi^*_{c\vec{k}}\left[\vec{\mathcal{A}}_\omega\cdot\hat{\vec{p}}_e + \hat{\vec{p}}_e\cdot\vec{\mathcal{A}}_\omega\right]\phi_{v\vec{k}'}d^3r_e \\
&= -\frac{i\hbar e}{2m\Omega}\int u^*_{c\vec{k}}e^{-i\vec{k}\cdot\vec{r}_e}\left[\vec{\mathcal{A}}_\omega\cdot\nabla_e u_{v\vec{k}'}e^{i\vec{k}'\cdot\vec{r}_e} + \nabla_e\cdot\vec{\mathcal{A}}_\omega u_{v\vec{k}'}e^{i\vec{k}'\cdot\vec{r}_e}\right]d^3r_e \\
&= -\frac{i\hbar e}{2m\Omega}\int u^*_{c\vec{k}}e^{-i\vec{k}\cdot\vec{r}_e}\left[\vec{\mathcal{A}}_\omega\cdot(i\vec{k}'u_{v\vec{k}'} + \nabla u_{v\vec{k}'})e^{i\vec{k}'\cdot\vec{r}_e} + \nabla_e u_{v\vec{k}'}e^{i\vec{k}'\cdot\vec{r}_e}\cdot\vec{\mathcal{A}}_\omega + u_{v\vec{k}'}e^{i\vec{k}'\cdot\vec{r}_e}\right]d^3r_e \\
&= -\frac{i\hbar e}{2m\Omega}\int \left[iu^*_{c\vec{k}}u_{v\vec{k}'}(\vec{k}'+\vec{k}) + u^*_{c\vec{k}}\nabla_e u_{v\vec{k}'} - u_{v\vec{k}'}\nabla u^*_{c\vec{k}}\right]\cdot\vec{\mathcal{A}}_\omega e^{i(\vec{k}'-\vec{k})\cdot\vec{r}_e}d^3r_e \\
&= -\int \vec{J}_{c\vec{k},v\vec{k}'}(\vec{r}_e)\cdot\vec{\mathcal{A}}_\omega(\vec{r}_e)\,d^3r_e \qquad \text{(A.24)}
\end{aligned}$$

where the Coulomb gauge was used from second to third line.

Using (A.23) and (A.24) in (A.21), writing $\vec{r}'$ instead of $\vec{r}_e$ and including $\times 2$ for the spin-summation

$$\begin{aligned}
\vec{J}_\omega(\vec{r}) &= 4\sum_{c\vec{k}}\sum_{v\vec{k}'}E_{c\vec{k},v\vec{k}'}f_{v\vec{k}',c\vec{k}}\frac{\int \vec{\mathcal{A}}_\omega(\vec{r}')\cdot\vec{J}_{c\vec{k},v\vec{k}'}(\vec{r}')\vec{J}_{v\vec{k}',c\vec{k}}(\vec{r})d^3r'}{E^2_{c\vec{k},v\vec{k}'} - (\hbar w)^2} \\
&= \frac{4}{i\omega}\sum_{c\vec{k}}\sum_{v\vec{k}'}E_{c\vec{k},v\vec{k}'}f_{v\vec{k}',c\vec{k}}\frac{\int \vec{\mathcal{E}}_\omega(\vec{r}')\cdot\vec{J}_{c\vec{k},v\vec{k}'}(\vec{r}')\vec{J}_{v\vec{k}',c\vec{k}}(\vec{r})d^3r'}{E^2_{c\vec{k},v\vec{k}'} - (\hbar w)^2},
\end{aligned}$$

using (A.15) in the last equality. Thus the conductivity can be recognised by comparison with (A.16)

$$\hat{\sigma}_\omega(\vec{r}) = \frac{4}{i\omega}\int d^3r'\sum_{c\vec{k}}\sum_{v\vec{k}'}E_{c\vec{k},v\vec{k}'}f_{v\vec{k}',c\vec{k}}\frac{\vec{J}_{v\vec{k}',c\vec{k}}(\vec{r})\otimes\vec{J}_{c\vec{k},v\vec{k}'}(\vec{r}')}{E^2_{c\vec{k},v\vec{k}'} - (\hbar w)^2},$$

and the susceptibility can be found using (A.18) to give $\hat{\chi}_\omega(\vec{r}) = i\hat{\sigma}_\omega(\vec{r})/\varepsilon_0\omega$

$$\hat{\chi}_\omega(\vec{r}) = \frac{4}{\omega^2\varepsilon_0}\int d^3r' \sum_{\vec{ck}}\sum_{\vec{vk'}} E_{\vec{ck},\vec{vk'}} f_{\vec{vk'},\vec{ck}} \frac{\vec{J}_{\vec{vk'},\vec{ck}}(\vec{r}) \otimes \vec{J}_{\vec{ck},\vec{vk'}}(\vec{r'})}{E^2_{\vec{ck},\vec{vk'}} - (\hbar w)^2}.$$

Whenever one of the above integral operators acts on a plane wave of wave vector $\vec{q}$ the result becomes proportional to a delta function ensuring conservation of crystal momentum, i.e. $\vec{k} - \vec{k}' = \vec{q}$:

$$\hat{\chi}_\omega(\vec{r}) \cdot \vec{\mathcal{E}}_{\omega,\vec{q}} e^{i\vec{q}\cdot\vec{r}'} = \frac{4}{\omega^2\varepsilon_0}\sum_{\vec{ck}}\sum_{\vec{vk'}} E_{\vec{ck},\vec{vk'}} f_{\vec{vk'},\vec{ck}} \frac{\vec{J}_{\vec{vk'},\vec{ck}}(\vec{r}) \otimes \int \vec{J}_{\vec{ck},\vec{vk'}}(\vec{r'}) e^{i\vec{q}\cdot\vec{r}'} d^3r'}{E^2_{\vec{ck},\vec{vk'}} - (\hbar w)^2} \cdot \vec{\mathcal{E}}_{\omega,\vec{q}}\ ,$$

considering the integral

$$\begin{aligned}
\Omega\vec{I}_{\vec{ck},\vec{vk'}} \equiv \int \vec{j}_{\vec{ck},\vec{vk'}}(\vec{r'}) e^{i(\vec{k}'-\vec{k}+\vec{q})\cdot\vec{r}'} d^3r &= \sum_{\vec{G},\vec{G}'} \vec{j}_{\vec{ck},\vec{vk'}}(\vec{G},\vec{G}') \int e^{i(\vec{k}'-\vec{k}+\vec{q}+\vec{G}'-\vec{G})\cdot r'} d^3r' \\
&= \Omega \sum_{\vec{G},\vec{G}'} \vec{j}_{\vec{ck},\vec{vk'}}(\vec{G},\vec{G}') \delta_{\vec{k}-\vec{k}',\vec{q}} \delta_{G,G'} \\
&= \Omega \sum_{\vec{G}} \vec{j}_{\vec{ck},\vec{vk'}}(\vec{G}) \delta_{\vec{k}-\vec{k}',\vec{q}}, \qquad \text{(A.25)} \\
\vec{j}_{\vec{ck},\vec{vk'}}(\vec{G}) \equiv i f^*_c(\vec{k}-\vec{G}) f_v(\vec{k}'-\vec{G}) &\left[\vec{k}+\vec{k}'-2\vec{G}\right]. \qquad \text{(A.26)}
\end{aligned}$$

The inclusion of the Kronecker delta $\delta_{\vec{G},\vec{G}'}$ is not strictly a mathemathical consequence, but a neglect of Umklapp-type processes. For an arbitrary Fourier expansion of the electric field, such as (A.11), the response becomes a sum of responses to plane waves where each term is weighted with the Fourier coefficient of the field. Thus, the analysis of a single $\vec{q}$-term is continued using the definition

$$\begin{aligned}
\chi_{\omega,\vec{q}}(\vec{r}) &\equiv \hat{\chi}_\omega(\vec{r}) e^{i\vec{q}\cdot\vec{r}} \\
\Rightarrow \vec{\mathcal{P}}_\omega(\vec{r}) &= \sum_{\vec{q}} \chi_{\omega,\vec{q}}(\vec{r}) \cdot \vec{\mathcal{E}}_{\omega,\vec{q}}, \qquad \text{(A.27)}
\end{aligned}$$

and dropping the prime on $\vec{k}'$

$$\chi_{\omega,\vec{q}}(\vec{r}) = -\frac{2i\hbar e}{\omega^2\varepsilon_0 m}\sum_{c,v,\vec{k}} E_{\vec{ck}+\vec{q},\vec{vk}} f_{\vec{vk},\vec{ck}+\vec{q}} \frac{\vec{J}_{\vec{vk},\vec{ck}+\vec{q}}(\vec{r}) \otimes \vec{I}_{\vec{ck}+\vec{q},\vec{vk}}}{E^2_{\vec{ck}+\vec{q},\vec{vk}} - (\hbar w)^2}, \qquad \text{(A.28)}$$

The k-summation is transformed to an integral according to

$$\sum_{\vec{k}} ... = \frac{\Omega}{8\pi^3}\int ...\, d^3k,$$

and the Pauli repulsion factor is taken to be 1 (i.e. the low temperature limit). Only the imaginary part of $\chi_{\omega,\vec{q}}$ is considered (a Kramers-Kronig tranformation can be used to find the real part of the response), and the limit of vanishing Γ is taken using the Dirac identity

$$\mathrm{Im}\left[\lim_{\gamma\to 0}\frac{1}{x-i\gamma}\right] = \pi\delta(x) \qquad \text{(A.29)}$$

to give

$$\begin{aligned}\lim_{\Gamma\to 0}\mathrm{Im}\left[\chi_{\omega,\vec{q}}(\vec{r})\right] &\equiv \chi''_{\omega,\vec{q}}(\vec{r}) \\ &= -\frac{i\hbar e\Omega}{8\omega^2\varepsilon_0 m\pi^2}\sum_{c,v}\int \vec{J}_{v\vec{k},c\vec{k}+\vec{q}}(\vec{r})\otimes\vec{I}_{c\vec{k}+\vec{q},v\vec{k}}\delta(E_{c\vec{k}+\vec{q},v\vec{k}}-\hbar\omega)d^3k \\ &= \frac{\hbar^2e^2}{16\omega^2\varepsilon_0 m^2\pi^2}\sum_{c,v}\int \vec{j}_{v\vec{k},c\vec{k}+\vec{q}}(\vec{r})\otimes\vec{I}_{c\vec{k}+\vec{q},v\vec{k}}e^{i\vec{q}\cdot r}\delta(E_{c\vec{k}+\vec{q},v\vec{k}}-\hbar\omega)d^3k \\ &= -\frac{\hbar^2e^2}{16\omega^2\varepsilon_0 m^2\pi^2}\sum_{c,v}\int \tilde{\mathbf{M}}_{c,v}(\vec{k},\vec{q},\vec{r})\delta(E_{c\vec{k}+\vec{q},v\vec{k}}-\hbar\omega)d^3k\end{aligned}$$

where the matrix elements are contained in $\tilde{\mathbf{M}}_{c,v}(\vec{k},\vec{q},\vec{r})$

$$\tilde{\mathbf{M}}_{c,v}(\vec{k},\vec{q},\vec{r}) = -\vec{j}_{v\vec{k},c\vec{k}+\vec{q}}(\vec{r})\otimes\vec{I}_{c\vec{k}+\vec{q},v\vec{k}}e^{i\vec{q}\cdot r}.$$

It will prove convinient to Fourier transform the $\vec{r}$-dependency in the following way

$$\mathbf{M}_{c,v}(\vec{k},\vec{q})\delta(\vec{q}-\vec{q}')(2\pi)^n\Omega^{n-3} \equiv \int \tilde{\mathbf{M}}_{c,v}(\vec{k},\vec{q},\vec{r})e^{-i\vec{q}'\cdot\vec{r}}d^3r \tag{A.30}$$

$$= -(2\pi)^n\Omega^{(n-3)/3}\vec{I}_{v\vec{k},c\vec{k}+\vec{q}}\otimes\vec{I}_{c\vec{k}+\vec{q},v\vec{k}}\delta(\vec{q}-\vec{q}'). \tag{A.31}$$

Here n denotes the number of dimensions where $\vec{q}$ is continuous, i.e. a wave propagating in the xy plane, decaying in the z direction would have $n=1$ since it is described by a continuous distribution of q_z, but only one q_x and q_y. The matrix elements become

$$\begin{aligned}\mathbf{M}_{c,v}(\vec{k},\vec{q}) = -\vec{I}_{v\vec{k},c\vec{k}+\vec{q}}\otimes\vec{I}_{c\vec{k}+\vec{q},v\vec{k}} = \sum_{\vec{G},\vec{G}'} & f_v^*(\vec{k}-\vec{G}')f_c(\vec{k}+\vec{q}-\vec{G}')f_c^*(\vec{k}+\vec{q}-\vec{G})f_v(\vec{k}-\vec{G}) \\ & \times[2\vec{k}+\vec{q}-2\vec{G}]\otimes[2\vec{k}+\vec{q}-2\vec{G}'].\end{aligned}$$

These can also be written

$$\mathbf{M}_{c,v}(\vec{q},\vec{k}) = \langle v\vec{k}|\left[e^{i\vec{q}\cdot\vec{r}}\nabla+\nabla e^{i\vec{q}\cdot\vec{r}}\right]|c\vec{k}+\vec{q}\rangle\otimes\langle c\vec{k}+\vec{q}|\left[e^{-i\vec{q}\cdot\vec{r}}\nabla+\nabla e^{-i\vec{q}\cdot\vec{r}}\right]|v\vec{k}\rangle.$$

These two forms are used interchangingly throughout the litterature, where the first form is used in[2]. It is noted that Umklapp processes are ignored, i.e. the local-field effects discussed in the introduction are neglected. The following definition is the response in q-space to an electromagnetic wave of frequency ω and wave vector $\vec{q}$:

$$\tilde{\chi}''_{\omega,\vec{q}} \equiv \frac{\hbar^2e^2}{16\omega^2\varepsilon_0 m^2\pi^2}\sum_{c,v}\int \mathbf{M}_{c,v}(\vec{k},\vec{q})\delta(E_{c\vec{k}+\vec{q},v\vec{k}}-\hbar\omega)d^3k. \tag{A.32}$$

To evaluate this expression, one can simply calculate two band structures displaced by $\vec{q}$ relative to each other by e.g. the EPM, and perform the integral by a suitable method. In the optical limit of $\vec{q}\to 0$ for isotropic media, the response reduces to

$$\chi''_\omega = \frac{e^2}{4\varepsilon_0 m^2\pi^2\omega^2}\sum_{c,v}\int |p^z_{cv}|^2\delta\left(E_{cv}-\hbar\omega\right)\,d^3k \tag{A.33}$$

A.4 Second order optical response

In the follwing the response oscillating at 2ω to an applied field of frequency ω will be derived. Considering an optical perturbation as before, neglecting $\hat{H}_2$ for now*

$$\hat{H}_1 = \frac{e}{2m}\left[\hat{\vec{p}}\cdot\vec{\mathcal{A}}_\omega(\vec{r}) + \vec{\mathcal{A}}_\omega(\vec{r})\cdot\hat{\vec{p}}\right],$$

the second harmonic current density response $\vec{J}_{2\omega}(r)$ is found using (A.10) and rewriting the Hamiltonian matrix elements according to (A.24)

$$\vec{J}_{2\omega}(\vec{r}) = \sum_{i,j,k} G_{ijk}\vec{J}_{ij}(\vec{r})\int\int \vec{J}_{jk}(\vec{r}')\cdot\vec{\mathcal{A}}_\omega(\vec{r}')\;\vec{J}_{ki}(\vec{r}'')\cdot\vec{\mathcal{A}}_\omega(\vec{r}'')\;d^3r'\;d^3r''$$

If this expression is written out in cartesian components (annotated as superscript)

$$\begin{aligned} J^l_{2\omega}(\vec{r}) &= \sum_{m,n}\sum_{i,j,k} G_{ijk}J^l_{ij}(\vec{r})\int\int J^m_{jk}(\vec{r}')\mathcal{A}^m_\omega(\vec{r}')\;J^n_{ki}(\vec{r}'')\mathcal{A}^n_\omega(\vec{r}'')\;d^3r'\;d^3r'' \\ &= \sum_{m,n}\sum_{i,j,k} G_{ijk}J^l_{ij}(\vec{r})\int\int J^m_{jk}(\vec{r}')J^n_{ki}(\vec{r}'')\;\mathcal{A}^m_\omega(\vec{r}')\mathcal{A}^n_\omega(\vec{r}'')\;d^3r'\;d^3r'' \end{aligned}$$

Using that $\vec{\mathcal{E}}_\omega = i\omega\vec{\mathcal{A}}_\omega$

$$J^l_{2\omega}(\vec{r}) = -\frac{1}{\omega^2}\sum_{m,n}\sum_{i,j,k} G_{ijk}J^l_{ij}(\vec{r})\int\int J^m_{jk}(\vec{r}')J^n_{ki}(\vec{r}'')\;\mathcal{E}^m_\omega(\vec{r}')\mathcal{E}^n_\omega(\vec{r}'')\;d^3r'\;d^3r''.$$

Defining the non-linear, second order conductivity through

$$\begin{aligned} J^l_{2\omega}(\vec{r}) &= \sum_{m,n}\hat{\sigma}^{(2)}_{lmn}(\omega,\vec{r})\mathcal{E}^m_\omega(\vec{r}')\mathcal{E}^n_\omega(\vec{r}'') \\ \hat{\sigma}^{(2)}_{lmn}(\omega,\vec{r}) &= -\frac{1}{\omega^2}\sum_{ijk}\int\int\; d^3r'd^3r''\; G_{ijk}J^l_{ij}(\vec{r})J^m_{jk}(\vec{r}')J^n_{ki}(\vec{r}'') \end{aligned}$$

Using the standard connection between polarisation and current density (A.18), the optical susceptibility is defined through

$$\begin{aligned} \vec{\mathcal{P}}^l_{2\omega}(\vec{r}) &= \sum_{m,n}\hat{\chi}^{lmn}_{2\omega}(\vec{r})\mathcal{E}^m_\omega(\vec{r}')\mathcal{E}^n_\omega(\vec{r}'') \\ \hat{\chi}^{(2)}_{lmn}(\omega,\vec{r}) &= -\frac{i}{2\varepsilon_0\omega^3}\sum_{ijk}\int\int\; d^3r'd^3r''\; G_{ijk}J^l_{ij}(\vec{r})J^m_{jk}(\vec{r}')J^n_{ki}(\vec{r}''). \end{aligned}$$

Introducing band notation, using band indices α,β,γ making the k-vectors $\vec{k}_\alpha,\vec{k}_\beta$ and $\vec{k}_\gamma$ and including a factor $\times 2$ for spin summation

$$\hat{\chi}^{(2)}_{lmn}(\omega,\vec{r}) = -\frac{i}{\varepsilon_0\omega^3}\sum_{\alpha,\beta,\gamma}\sum_{\vec{k}_\alpha,\vec{k}_\beta,\vec{k}_\gamma}\int\int\; d^3r'd^3r''\; G_{\alpha\vec{k}_\alpha,\beta\vec{k}_\beta,\gamma\vec{k}_\gamma}J^l_{\alpha\vec{k}_\alpha,\beta\vec{k}_\beta}(\vec{r})J^m_{\beta\vec{k}_\beta,\gamma\vec{k}_\gamma}(\vec{r}')J^n_{\gamma\vec{k}_\gamma,\alpha\vec{k}_\alpha}(\vec{r}'').$$

Analogous to the procedure used in the linear case, the response is investigated in Fourier space, i.e. a plane wave spatial dependence is assumed for the fields, and a single plane wave is considered

$$\begin{aligned} \chi^{(2)}_{lmn}(\omega,\vec{q},\vec{q}',\vec{r}) &\equiv \hat{\chi}^{(2)}_{lmn}(\omega,\vec{q},\vec{q}',\vec{r})e^{i\vec{q}\cdot\vec{r}}e^{i\vec{q}'\cdot\vec{r}} \\ \Rightarrow \mathcal{P}^l_{2\omega}(\vec{r}) &= \sum_{\vec{q},\vec{q}'}\sum_{m,n}\chi^{lmn}_{2\omega,\vec{q},\vec{q}'}(\vec{r})\mathcal{E}^m_{\omega,\vec{q}}\mathcal{E}^n_{\omega,\vec{q}'}. \end{aligned} \tag{A.34}$$

*Moss *et al.* (1990)[46] show this contribution to vanish.

(A.23) and (A.25) is used, giving

$$\chi^{(2)}_{lmn}(\omega,\vec{q},\vec{q}') = \frac{i\hbar^2 e^2}{4m^2\varepsilon_0\omega^3}\sum_{\alpha,\beta,\gamma}\sum_{\vec{k}} G_{\alpha\vec{k}-\vec{q}',\beta\vec{k}+\vec{q},\gamma\vec{k}} J^l_{\alpha\vec{k}-\vec{q}',\beta\vec{k}+\vec{q}}(\vec{r}) I^m_{\beta\vec{k}+\vec{q},\gamma\vec{k}} I^n_{\gamma\vec{k},\alpha\vec{k}-\vec{q}'}.$$

Transforming the k-summation to an integral as before and Fourier transforming the last r-dependency as for the linear case

$$\tilde{\chi}^{(2)}_{lmn}(\omega,\vec{q},\vec{q}') = -\frac{\Omega\hbar^3 e^3}{64\pi^3 m^3\varepsilon_0\omega^3}\sum_{\alpha,\beta,\gamma}\int G_{\alpha\vec{k}-\vec{q}',\beta\vec{k}+\vec{q},\gamma\vec{k}} I^l_{\alpha\vec{k}-\vec{q}',\beta\vec{k}+\vec{q}} I^m_{\beta\vec{k}+\vec{q},\gamma\vec{k}} I^n_{\gamma\vec{k},\alpha\vec{k}-\vec{q}'}\, d^3k. \tag{A.35}$$

A.4.1 The optical limit

Here it is sought to verify the consistency of this result with expressions reported elsewhere for the optical limit $\vec{q},\vec{q}' \to 0$. In addition, this case will provide some important physical insight. The following deriviation is aided by a note from Prof. Thomas G. Pedersen. It is relatively easy to show that

$$I^l_{\alpha\vec{k},\beta\vec{k}} = \frac{2i}{\hbar} p^l_{\alpha\vec{k},\beta\vec{k}},$$

which allows $\tilde{\chi}^{lmn}_{2\omega}$ to be rewritten, neglecting the subscript $\vec{k}$ and $\vec{q}$ and introducing the complex frequency $\omega \to w$

$$\tilde{\chi}^{(2)}_{lmn} = i\frac{\Omega e^3}{8\pi^3 m^3\varepsilon_0 w^3}\sum_{\alpha,\beta,\gamma}\int G_{\alpha\beta\gamma} p^l_{\alpha\beta} p^m_{\beta\gamma} p^n_{\gamma\alpha}\, d^3k.$$

$\mathrm{G}_{\alpha,\beta,\gamma}$ is written out

$$\tilde{\chi}^{lmn}_{2\omega} = \frac{\Omega e^3}{8\pi^3 m^3 w^3\hbar^2\varepsilon_0}\sum_{\alpha,\beta,\gamma}\int \frac{i p^l_{\alpha\beta} p^m_{\beta\gamma} p^n_{\gamma\alpha}}{2w-\omega_{\beta\alpha}}\left(\frac{f_{\alpha\gamma}}{w-\omega_{\gamma\alpha}}+\frac{f_{\beta\gamma}}{w-\omega_{\beta\gamma}}\right) d^3k. \tag{A.36}$$

Due to intrinsic permutation symmetry, the order of m and n are completely arbitrary. To reflect this symmetry, the matrix elements are rewritten according to

$$Q_{\alpha\beta\gamma} \equiv p^l_{\alpha\beta} p^m_{\beta\gamma} p^n_{\gamma\alpha} = \frac{p^l_{\alpha\beta}\left(p^m_{\beta\gamma} p^n_{\gamma\alpha} + p^m_{\gamma\alpha} p^n_{\beta\gamma}\right)}{2} = p^l_{\alpha\beta}\left\{p^m_{\beta\gamma} p^n_{\gamma\alpha}\right\}$$

where the curly brackets denote the symmetrisation. Interchanging α and β in the last term of (A.36)

$$\tilde{\chi}^{(2)}_{lmn} = \frac{\Omega e^3}{8\pi^3 m^3 w^3\hbar^2\varepsilon_0}\sum_{\alpha,\beta,\gamma}\int \frac{i}{2w-\omega_{\beta\alpha}}\left(\frac{Q_{\alpha\beta\gamma} f_{\alpha\gamma}}{w-\omega_{\gamma\alpha}}+\frac{Q_{\beta\alpha\gamma} f_{\alpha\gamma}}{w-\omega_{\alpha\gamma}}\right) d^3k.$$

The integrand is seen to be $Q_{\alpha\beta\gamma}$ multiplied by a function $\varsigma(\vec{k})$ whose k-dependency enters exclusively through band energies , why the integral can be written on the form

$$F = \int \varsigma(\vec{k}) Q_{\alpha\beta\gamma}(\vec{k})\, d^3k.$$

Substituting $\vec{k}$ by $-\vec{k}$, noting that the band energies are degenerate in $\vec{k}$ and $-\vec{k}$ and reversing the integration limits give

$$F = \int \varsigma(\vec{k}) Q_{\alpha\beta\gamma}(-\vec{k})\, d^3k.$$

It is easy to show that $p^{x,y,z}_{\alpha\beta}(\vec{k}) = -p^{x,y,z}_{\beta\alpha}(-\vec{k}) = -p^{x,y,z*}_{\alpha\beta}(-\vec{k})$ – hence

$$F = -\int \varsigma(\vec{k}) Q^*_{\alpha\beta\gamma}(\vec{k})\, d^3k.$$

Thus, comparing the first integral to the last, it is clear that $Q_{\alpha\beta\gamma}(\vec{k}) = -Q^*_{\alpha\beta\gamma}(\vec{k})$. Any function which upon complex conjugation returns the negative of itself must be purely imaginary, $Q_{\alpha\beta\gamma} = i\mathrm{Im}\left[Q_{\alpha\beta\gamma}\right] = iK_{\alpha\beta\gamma}$. Thus

$$\tilde{\chi}^{(2)}_{lmn} = -\frac{\Omega e^3}{8\pi^3 m^3 w^3 \hbar^2 \varepsilon_0} \sum_{\alpha,\beta,\gamma} \int \frac{1}{2w - \omega_{\beta\alpha}} \left(\frac{K_{\alpha\beta\gamma} f_{\alpha\gamma}}{w - \omega_{\gamma\alpha}} + \frac{K_{\beta\alpha\gamma} f_{\alpha\gamma}}{w - \omega_{\alpha\gamma}} \right) d^3k.$$

Using that $K_{\beta\alpha\gamma} = \mathrm{Im}\left[\left(p^l_{\alpha\beta} \left\{ p^m_{\beta\gamma} p^n_{\gamma\alpha} \right\} \right)^* \right] = -K_{\alpha\beta\gamma}$, which is easy to see by simply writing out the expressions, we get

$$\begin{aligned}
\tilde{\chi}^{(2)}_{lmn} &= -\frac{\Omega e^3}{8\pi^3 m^3 w^3 \hbar^2 \varepsilon_0} \sum_{\alpha,\beta,\gamma} \int \left(\frac{K_{\alpha\beta\gamma} f_{\alpha\gamma}}{(2w - \omega_{\beta\alpha})(w - \omega_{\gamma\alpha})} - \frac{K_{\alpha\beta\gamma} f_{\alpha\gamma}}{(2w + \omega_{\beta\alpha})(w + \omega_{\gamma\alpha})} \right) d^3k \\
&= -\frac{\Omega e^3}{4\pi^3 m^3 w^2 \hbar^2 \varepsilon_0} \sum_{\alpha,\beta,\gamma} \int \left(\frac{K_{\alpha\beta\gamma} f_{\alpha\gamma} \left(\omega_{\beta\alpha} + 2\omega_{\gamma\alpha} \right)}{\left(4w^2 - \omega^2_{\beta\alpha} \right) \left(w^2 - \omega^2_{\gamma\alpha} \right)} \right) d^3k.
\end{aligned}$$

This expression can be expanded in partial fractions by the following procedure: We wish to write an expression of the form $f(x) = N(x)/D(x)$ as a series of simple terms, and we note that[47] any linear factor in $D(x)$ of the form $(ax+b)^n$ contributes a series

$$\sum_{i=1}^{n} \frac{A_i}{(ax+b)}$$

and any quadratic factor $(ax^2 + bx + c)^n$ contributes a series

$$\sum_{i=1}^{n} \frac{A_i x + B_i}{(ax^2 + bx + c)^n}.$$

The task is then to collect all the partial fractions and equate the numerator to the numerator of the original expression. One then equates like powers in x, and solves for the coefficients. Here we wish to expand the expression

$$F(w) = \frac{\omega_{\beta\alpha} + 2\omega_{\gamma\alpha}}{w^2 \left(4w^2 - \omega^2_{\beta\alpha} \right) \left(w^2 - \omega^2_{\gamma\alpha} \right)}, \tag{A.37}$$

hence from the procedure sketched above

$$F(w) = \frac{Aw + B}{w^2} + \frac{Cw + D}{4w^2 - \omega^2_{\beta\alpha}} + \frac{Ew + F}{w^2 - \omega^2_{\gamma\alpha}}. \tag{A.38}$$

Collecting these partial fractions yields the same denominator as (A.37) and the numerator

$$\begin{aligned}
N(w) =& w^5 \left(4A + C + 4E \right) + w^4 \left(4B + D + 4F \right) + w^3 \left(-A(4\omega^2_{\gamma\alpha} + \omega^2_{\beta\alpha}) - C\omega^2_{\gamma\alpha} - E\omega^2_{\beta\alpha} \right) \\
&+ w^2 \left(-B(4\omega^2_{\gamma\alpha} + B\omega^2_{\gamma\alpha}) - D\omega^2_{\gamma\alpha} - F\omega^2_{\beta\alpha} \right) + w \left(\mathcal{A}\omega^2_{\beta\alpha}\omega^2_{\gamma\alpha} \right) + B\omega^2_{\beta\alpha}\omega^2_{\gamma\alpha}.
\end{aligned}$$

Equating like powers of w from $N(w)$ and the numerator of (A.37) and solving the equation system gives

$$\begin{aligned}
A &= 0, \\
B &= \frac{\omega_{\beta\alpha} + 2\omega_{\gamma\alpha}}{\omega_{\beta\alpha}^2 \omega_{\gamma\alpha}^2}, \\
C &= 0 \\
D &= \frac{16}{\omega_{\beta\alpha}^2 \left(\omega_{\alpha\beta} - 2\omega_{\gamma\alpha}\right)}, \\
E &= 0, \\
F &= \frac{1}{\omega_{\gamma\alpha}^2 \left(\omega_{\beta\alpha} - 2\omega_{\gamma\alpha}\right)}.
\end{aligned}$$

$F(w)$ can thus be expanded as

$$F(w) = \frac{\omega_{\beta\alpha} + 2\omega_{\gamma\alpha}}{\omega_{\beta\alpha}^2 \omega_{\gamma\alpha}^2 w^2} + \frac{16}{\omega_{\beta\alpha}^2 \left(\omega_{\beta\alpha} - 2\omega_{\gamma\alpha}\right)\left(4w^2 - \omega_{\beta\alpha}^2\right)} - \frac{1}{\omega_{\gamma\alpha}^2 \left(\omega_{\beta\alpha} - 2\omega_{\gamma\alpha}\right)\left(w^2 - \omega_{\gamma\alpha}^2\right)}.$$

Upon performing the summation, the first term can be shown not to contribute

$$\tilde{\chi}_{lmn}^{(2)} = -\frac{\Omega e^3}{4\pi^3 m^3 \hbar^2 \varepsilon_0} \sum_{\alpha,\beta,\gamma} \int \frac{K_{\alpha\beta\gamma} f_{\alpha\gamma}}{\omega_{\beta\alpha} - 2\omega_{\gamma\alpha}} \left(\frac{16}{\omega_{\beta\alpha}^2 \left(4w^2 - \omega_{\beta\alpha}^2\right)} - \frac{1}{\omega_{\gamma\alpha}^2 \left(w^2 - \omega_{\gamma\alpha}^2\right)} \right) d^3k.$$

To continue, each of the terms are investigated seperatly; defining

$$Y_{\alpha\beta\gamma} = \frac{16}{\omega_{\beta\alpha}(4w^2 - \omega_{\beta\alpha})(\omega_{\beta\alpha} - 2\omega_{\gamma\alpha})}$$

it is noted that $Y_{\alpha\beta\gamma}$ is antisymmetric upon interchange of the first and second indices, i.e. $Y_{\alpha\beta\gamma} = -Y_{\beta\alpha\gamma}$. The summation can schematically be written

$$\sum_{\alpha,\beta,\gamma} = \sum_{\alpha\in c,\beta,\gamma\in c} + \sum_{\alpha\in v,\beta,\gamma\in v} + \sum_{\alpha\in c,\beta,\gamma\in v} + \sum_{\alpha\in v,\beta,\gamma\in c}.$$

At low temperatures the Pauli repulsion factor causes the first two summations to vanish while $f_{cv} \approx -1$ and $f_{vc} \approx 1$

$$\sum_{\alpha,\beta,\gamma} = \sum_{\alpha\in v,\beta,\gamma\in c} - \sum_{\alpha\in c,\beta,\gamma\in v}.$$

Using the antisymmetry of $Y_{\alpha\beta\gamma}$ to interchange α and β in the last sum

$$\sum_{\alpha,\beta,\gamma} = \sum_{\alpha\in v,\beta,\gamma\in c} + \sum_{\alpha,\beta\in v,\gamma\in c}.$$

The unspecified sum can also seperated to run over occupied and empty states

$$\sum_{\alpha,\beta,\gamma} = \sum_{\alpha\in v,\beta\in v,\gamma\in c} + \sum_{\alpha\in v,\beta\in c,\gamma\in c} + \sum_{\alpha\in v,\beta\in c,\gamma\in v} + \sum_{\alpha\in c,\beta\in c,\gamma\in v}.$$

Since any sum with the two first indices from the same set contains all permutations of these two indices, the total contribution vanishes due to the antisymmetry of $Y_{\alpha\beta\gamma}$. Hence the contribution from Y-term to the total susceptiblility becomes

$$\tilde{\chi}_{lmn}^{A(2)} = -\frac{\Omega e^3}{4\pi^3 m^3 \hbar^2 \varepsilon_0} \sum_{v,c,\gamma} \int \frac{16 K_{vc\gamma}}{\omega_{cv}^2 (4w^2 - \omega_{cv}^2)(\omega_{cv} - 2\omega_{\gamma v})} d^3k.$$

Taking the imaginary part in the limit of vanishing Γ

$$\begin{aligned}
\tilde{\chi}^{A(2)\prime\prime}_{lmn} &= -\frac{\Omega e^3}{4\pi^3 m^3 \hbar^2 \varepsilon_0} \sum_{v,c,\gamma} \int \frac{16 K_{vc\gamma}}{\omega_{cv}^2(\omega_{cv}-2\omega_{\gamma v})} \lim_{\Gamma\to 0} \mathrm{Im}\left[\frac{1}{4w^2-\omega_{cv}^2}\right] d^3k \\
&= -\frac{\Omega e^3}{4\pi^3 m^3 \hbar^2 \varepsilon_0} \sum_{v,c,\gamma} \int \frac{16 K_{vc\gamma}}{\omega_{cv}^2(\omega_{cv}-2\omega_{\gamma v})} \lim_{\Gamma\to 0} \mathrm{Im}\left[\frac{1}{(2w+\omega_{cv})(2w-\omega_{cv})}\right] d^3k \\
&= \frac{\Omega e^3}{4\pi^3 m^3 \hbar^2 \varepsilon_0} \sum_{v,c,\gamma} \int \frac{16 K_{vc\gamma}}{\omega_{cv}^2(\omega_{cv}-2\omega_{\gamma v})(2w+\omega_{cv})} \lim_{\Gamma\to 0} \mathrm{Im}\left[\frac{1}{\omega_{cv}-2w}\right] d^3k
\end{aligned} \tag{A.39}$$

where the last equation follows from the fact that $\omega_{cv} > 0$. Using the Dirac identity (A.29)

$$\begin{aligned}
\tilde{\chi}^{A(2)\prime\prime}_{lmn} &= \frac{\Omega e^3}{4\pi^2 m^3 \hbar^2 \varepsilon_0} \sum_{v,c,\gamma} \int \frac{16 K_{vc\gamma}}{\omega_{cv}^2(\omega_{cv}-2\omega_{\gamma v})(2w+\omega_{cv})} \delta\big(2\omega-\omega_{cv}\big)\, d^3k \\
&= \frac{\Omega e^3}{8\pi^2 m^3 \omega^3 \hbar^2 \varepsilon_0} \sum_{v,c,\gamma} \int \frac{K_{vc\gamma}}{\omega-\omega_{\gamma v}} \delta\big(2\omega-\omega_{cv}\big)\, d^3k.
\end{aligned}$$

Similar considerations leads to the following expression for the second term

$$\tilde{\chi}^{B(2)\prime\prime}_{lmn} = \frac{\Omega e^3}{8\pi^2 m^3 \hbar^2 \varepsilon_0} \sum_{v,\beta,c} \int \left(\frac{K_{v\beta c}}{\omega+\omega_{c\beta}} + \frac{K_{c\beta v}}{\omega+\omega_{\beta v}}\right) \delta\big(\omega-\omega_{cv}\big)\, d^3k$$

The imaginary part of the second harmonic response becomes, neglecting the tilde on χ since it was used to indicate Fourier transformation (which is of no importance in the limit of $q \to 0$)

$$\chi^{(2)\prime\prime}_{lmn} = \frac{\Omega e^3}{8\pi^2 m^3 \hbar^2 \varepsilon_0} \int \left[\sum_{v,c,\gamma} \frac{K_{vc\gamma}}{\omega-\omega_{\gamma v}} \delta\big(2\omega-\omega_{cv}\big) + \left(\frac{K_{v\beta c}}{\omega+\omega_{c\beta}} + \frac{K_{c\beta v}}{\omega+\omega_{\beta v}}\right) \delta\big(\omega-\omega_{cv}\big)\right] d^3k.$$

It should be noted, however, that the dipole allowed bulk second harmonic response of centro-symmetric materials, such as silicon, is exactly zero.

B The Linear-Analytic Tetrahedron Method

This section will provide an overview of the linear-tetrahedron method suitable for calculation of second-order response functions in zinc-blende crystals. Essentially, this section is a review of[30]; the task at hand is to solve integrals of the kind

$$F_{ij}(\omega) = \int_{BZ} f_{ij}(\omega,\vec{k})\, d^3k$$

numerically, where i and j are band indices, ω is frequency and $\vec{k}$ is an electronic wave vector. The integration domain is the Brillouin zone, although often is can be reduced to the irreducible zone by symmetry considerations. The Brillouin zone is divided into a set of tetrahedral microzones, and the integral is divided into contributions from each tetrahedron

$$F_{ij}(\omega) = \sum_{n=1}^{N} \int_{V_n} f_{ij}(\omega,\vec{k})\, d^3k = \sum_{n=1}^{N} g_n(\omega) \tag{B.1}$$

where V_n is the volume of the n'th tetrahedron. g_n can also be written

$$g_n(\omega) = \int \int_{S_n} \frac{f_{ij}(\omega,\vec{k})}{|\nabla_{\vec{k}} E_{ji}(\vec{k})|}\, d^2k\, dE_{ji}.$$

where S_n is the surface of constant energy E_{ji} contained in the n'th tetrahedron. The energy bands are linearized inside each tetrahedron

$$E_{ji}(\vec{k}) = E_{ji}(\vec{k}_1) + \vec{b}\cdot(\vec{k}-\vec{k}_1),$$

where $\vec{k}_m$ denote the tetrahedron vertices. The vertices are sorted according to

$$E_{ji}(\vec{k}_1) \le E_{ji}(\vec{k}_2) \le E_{ji}(\vec{k}_3) \le E_{ji}(\vec{k}_4).$$

The function f is usually, and in this work exclusively, on the form

$$f_{ij}(\omega,\vec{k}) = P_{ij}(\vec{k})\delta\big(\hbar\omega - E_{ji}(\vec{k})\big) \tag{B.2}$$

where P_{ij} is some matrix element. If P_{ij} is considered independent of $\vec{k}$ inside a specific tetrahedron, then the $\vec{k}$-dependence of f_{ij} is contained in energy terms, making f_{ij} invariant on the surface of constant energy S_n, allowing

$$g_n(\omega) = \int A_n(E_{ji}) \frac{f_{ij}(\omega,\vec{k})}{|\nabla_{\vec{k}} E_{ji}(\vec{k})|}\, dE_{ji}. \tag{B.3}$$

where $A_n(E_{ji})$ is the area of constant energy E_{ji} contained in tetrahedron n. To calculate A_n some vectors specifying the specific tetrahedron are necessary.

$$\begin{aligned}\vec{a}_{21} &= \frac{\vec{k}_{31}\times\vec{k}_{41}}{6V_n}\\ \vec{a}_{31} &= \frac{\vec{k}_{41}\times\vec{k}_{21}}{6V_n}\\ \vec{a}_{41} &= \frac{\vec{k}_{21}\times\vec{k}_{31}}{6V_n}\\ V_n &= \frac{\vec{k}_{21}\cdot(\vec{k}_{31}\times\vec{k}_{41})}{6}.\end{aligned} \tag{B.4}$$

These vectors was chosen to have the orthogonality property $\vec{a}_{\alpha 1}\cdot\vec{k}_{\beta,1} = \delta_{\alpha,\beta}$. Letting the subscripts on E denote teteahedron vertices, the energy E is linearized

$$\vec{b}\cdot\vec{k}_{\alpha 1} = E_{\alpha 1}$$

where $E_{\alpha 1} = E_{ij}(\vec{k}_\alpha) - E_{ij}(\vec{k}_1)$. And in terms of the a-vectors

$$\begin{aligned}&\vec{b}\cdot\vec{k}_{\alpha 1} = E_{\alpha 1}\\ \Rightarrow&\sum_\alpha \vec{b}\cdot\vec{k}_{\alpha 1}\vec{a}_{\alpha 1} = \sum_\alpha E_{\alpha 1}a_{\alpha 1}\\ =&\vec{b}\sum_\alpha(\vec{a}_{\alpha 1}\cdot\vec{k}_{\alpha 1}) = \sum_\alpha E_{\alpha 1}a_{\alpha 1}\\ =&\vec{b} = \sum_\alpha E_{\alpha 1}a_{\alpha 1}.\end{aligned}$$

Considering the geommetry of a tetrahedron, it is clear that S_n based on linearly interpolated energies will form either a triangle or a quadrilateral with corners at points on the tetrahedron edges of the same energy. This situation is illustrated in Fig. B.1 The ordering of the energies comes in handy when classifying these cases: If $E_{ji}(\vec{k}_1) < E_{ji}(\vec{k}) < E_{ji}(\vec{k}_2)$, then it follows that $E_{ji}(\vec{k}) < E(\vec{k}_3) < ...$ why it must form a triangle illustrated by in Fig. B.1 by "Region I", similarly if $E_{ji} > E(\vec{k}_3)$ (Region III). The quadrilateral arises from the case when $E(\vec{k}_2) < E_{ji} < E(\vec{k}_3)$ (Region II). The area $S_n^{(I)}$ of Region I is given by the area of a triangle

$$S_n^{(I)}(E_{ji}(\vec{k})) = \frac{1}{2}|(\vec{r}_1-\vec{r}_2)\times(\vec{r}_3-\vec{r}_2)|, \tag{B.5}$$

with the nodes found by linear interpolation along the tetrahedron edges

$$\vec{r}_\alpha = \frac{\vec{k}_{\alpha 1}}{E_{\alpha 1}}\left[E(\vec{k}) - E(\vec{k}_1)\right] = \frac{\vec{k}_{\alpha 0}}{E_{\alpha 1}}\Delta E(\vec{k}).$$

Inserting into (B.5) and reducing

$$A_n = S_n^{(I)} = \frac{3[\Delta E(\vec{k})]^2 V|\vec{b}|}{E_{21}E_{31}E_{41}}. \tag{B.6}$$

To derermine the area $S_n^{(II)}$ of the quadrilateral in Region II, we note that the area of a projected triangle, shown as the the green triangle extended to $\vec{k}_t'$ in Fig. B.1, can be calculated by the same method as $S_n^{(I)}$. This triangle is too large, however, by an amount S_1 – the light green triangle outside the tetrahedron volume

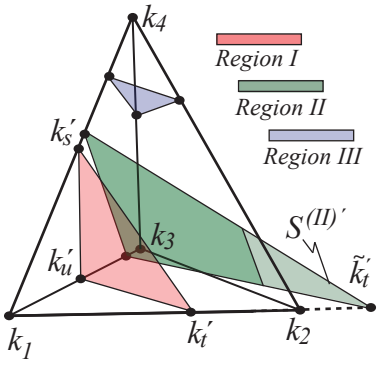


Figure B.1: Notation for a tetrahedron. In Region I $E_1 \leq E > E_2$

$$A_n = S_n^{(II)} = S_n^{(I)} - S_1, \tag{B.7}$$

where S_1 is also a triangle

$$S_1 = \frac{[E(\vec{k}) - E_2]^2}{2E_{21}E_{32}E_{42}} V|\vec{b}|.$$

In Region III $S_n^{(III)}$ is projected from $\vec{k}_4$ instead of $\vec{k}_1$

$$A_n = S_n^{(III)}[\vec{E}(\vec{k})] = \frac{[E(\vec{k}) - E_4]^2}{2E_{41}E_{43}E_{42}} V|\vec{b}|. \tag{B.8}$$

Having defined A_n completes the integration over S_n, and only the integration over E remains. When the $f_{ij}(\vec{k})$ contains a δ-function, the integration is trivial – we simply replace $E(\vec{k})$ by $E = \hbar\omega$. To summerise, by combination of eqs. (B.3), (B.6), (B.7) and (B.8) the integrals become, with f_{ij} on the form (B.2):

$$\begin{aligned}
&\text{If } \mathrm{E_4 < E < E_1}\text{: } g_n = 0 \\
&\text{If } \mathrm{E_1 < E < E_2}\text{: } g_n = 3v\frac{[\hbar\omega - E_1]^2}{E_{21}E_{31}E_{41}}\overline{P}_{ij} \\
&\text{If } \mathrm{E_2 < E < E_3}\text{: } g_n = 3v\left[\frac{[\hbar\omega - E_1]^2}{E_{21}E_{31}E_{41}} - \frac{[\hbar\omega - E_2]^2}{E_{21}E_{32}E_{42}}\right]\overline{P}_{ij} \\
&\text{If } \mathrm{E_3 < E < E_4}\text{: } g_n = 3v\frac{[\hbar\omega - E_4]^2}{vE_{41}E_{43}E_{42}}\overline{P}_{ij}
\end{aligned}$$

where $\overline{P}_{ij} = [P_{ij}(\vec{k}_1) + P_{ij}(\vec{k}_2) + P_{ij}(\vec{k}_3) + P_{ij}(\vec{k}_4)]/4$ is P_{ij} at the matrix element evaluated at the tetrahedron centroid. These expressions are the samed as those derived by MacDonald *et al.* (1979)[15], although they also present expressions including the linear variation of P_{ij}.

C Absorption in Ag thin-film

An electric field inside an Ag thin-film is given by $\vec{\mathcal{E}}_\omega(\vec{r}) = \mathcal{E}_0\left[\mathcal{E}_x\hat{x} + \mathcal{E}_z\hat{z}\right]e^{i\beta x}$ with

$$\mathcal{E}_x = \frac{i}{\omega\varepsilon\varepsilon_0}s_2\left(\sinh s_2 z + \frac{s_1\varepsilon_m}{s_2\varepsilon}\cosh s_2 z\right), \quad h > z > 0, \tag{C.1}$$

$$\mathcal{E}_z = -\frac{\beta}{\omega\varepsilon\varepsilon_0}\cosh s_2 h + \frac{s_1\varepsilon_m}{s_2\varepsilon}\sinh s_2 z, \quad h > z > 0,, \tag{C.2}$$

and the time averaged absorbed power given by the integral

$$P_{abs}(\omega) = \frac{\omega}{2}\mathrm{Im}\left[\int \mathcal{P}_\omega(\vec{r})\cdot\vec{\mathcal{E}}_\omega(\vec{r})^*\, d^3r\right], \tag{C.3}$$

is desired. Absorption in the silver film is found by (C.3)

$$\begin{aligned}\mathcal{P}_{abs}^{(Ag)} &= \frac{\omega|\mathcal{E}_0|^2\varepsilon_0}{2}\mathrm{Im}\left[\int\left(\varepsilon_m - 1\right)\left(|\mathcal{E}_x|^2 + |\mathcal{E}_z|^2\right)d^3r\right] \\ &= \frac{\varepsilon_0\omega|\mathcal{E}_0|^2\Omega^{1/3}\,dx}{2}\mathrm{Im}\left[\left(\varepsilon_m - 1\right)\int\left(|\mathcal{E}_x|^2 + |\mathcal{E}_z|^2\right)dz\right].\end{aligned}$$

Considering $|\mathcal{E}_x|^2$ first using $a \equiv s_1\varepsilon_m/s_2\varepsilon_1$ and (C.1)

$$\begin{aligned}|\mathcal{E}_x|^2 = \left|\frac{s_2}{\omega\varepsilon_0\varepsilon_m}\right|^2 \Big[&|\sinh(s_2 z)|^2 + |a|^2|\cosh(s_2 z)|^2 + a^*\sinh(s_2 z)\cosh^*(s_2 z) \\ &+ a\sinh^*(s_2 z)\cosh(s_2 z)\Big]\end{aligned}$$

The following identities are easy to prove using the exponential definitions of the hyperbolic and trigonometric functions

$$\begin{aligned}|\sinh(s_2 z)|^2 &= \frac{1}{2}\Big[\cosh(2s_2' z) - \cos(2s_2'' z)\Big], \\ |\cosh(s_2 z)|^2 &= \frac{1}{2}\Big[\cosh(2s_2' z) + \cos(2s_2'' z)\Big], \\ \sinh(s_2 z)\cosh^*(s_2 z) &= \frac{1}{2}\Big[\sinh(2s_2' z) + i\sin(2s_2'' z)\Big], \\ \sinh^*(s_2 z)\cosh(s_2 z) &= \frac{1}{2}\Big[\sinh(2s_2' z) - i\sin(2s_2'' z)\Big].\end{aligned}$$

Using these, $|\mathcal{E}_x|^2$ can be written

$$\begin{aligned}|\mathcal{E}_x|^2 = \frac{1}{2}\left|\frac{s_2}{\omega\varepsilon_0\varepsilon_m}\right|^2 \Big[&\cosh(2s_2' z)(1 + |a|^2) + \cos(2s_2'' z)(|a|^2 - 1) \\ &+ 2(a'\sinh(2s_2' z) + a''\sin(2s_2'' z))\Big]\end{aligned}$$

and integrated

$$\begin{aligned}\int_0^h |\mathcal{E}_x|^2 dz &= \left|\frac{s_2}{2\omega\varepsilon_0\varepsilon_m}\right|^2 \Big[\frac{1}{s_2'}\sinh(2s_2'h)(1+|a|^2)+\frac{1}{s_2''}\sin(2s_2''h)(|a|^2-1)\\ &\quad +2\big(\frac{a'}{s_2'}(\cosh(2s_2'h)-1)-\frac{a''}{s_2''}(\cos(2s_2''h)-1)\big)\Big]\\ &= \left|\frac{s_2}{2\omega\varepsilon_0\varepsilon_m}\right|^2 I_x\end{aligned}$$

The same is done for $|\mathcal{E}_z|^2$

$$|\mathcal{E}_z|^2 = \frac{1}{2}\left|\frac{\beta}{\omega\varepsilon_0\varepsilon_m}\right|^2 \Big[\cosh(2s_2'z)\big(1+|a|^2\big)+\cos(2s_2''z)\big(1-|a|^2\big)+2\big(a'\sinh(2s_2'z)-a''\sin(2s_2''z)\big)\Big],$$

and integrating

$$\begin{aligned}\int_0^h |\mathcal{E}_z|^2 dz &= \left|\frac{\beta}{2\omega\varepsilon_0\varepsilon_m}\right|^2 \Big[\frac{1}{s_2'}\sinh(2s_2'h)\big(1+|a|^2\big)+\frac{1}{s_2''}\sin(2s_2''h)\big(1-|a|^2\big)\\ &\quad +2\big(\frac{a'}{s_2'}(\cosh(2s_2'h)-1)+\frac{a''}{s_2''}(\cos(2s_2''h)-1)\big)\Big],\\ &= \left|\frac{\beta}{2\omega\varepsilon_0\varepsilon_m}\right|^2 I_z.\end{aligned}$$

Thus the absorbed power in silver can be written

$$\begin{aligned}\mathcal{P}_{abs}^{(Ag)} &= \frac{\varepsilon_0\omega|\mathcal{E}_0|^2\Omega^{1/3}\,dx}{2}\mathrm{Im}\left[\big(\varepsilon_m-1\big)\left|\frac{s_2}{2\omega\varepsilon_0\varepsilon_m}\right|^2 I_x+\left|\frac{\beta}{2\omega\varepsilon_0\varepsilon_m}\right|^2 I_z\right]\\ &= \frac{|\mathcal{E}_0|^2\Omega^{1/3}\,dx}{8\varepsilon_0\omega}\varepsilon_m''\left[\left|\frac{s_2}{\varepsilon_m}\right|^2 I_x+\left|\frac{\beta}{\varepsilon_m}\right|^2 I_z\right].\end{aligned}$$