INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI
800-521-0600
LINEAR AND NONLINEAR OPTICAL PROPERTIES OF SEMICONDUCTORS: THEORY AND CALCULATIONS

by

James L. P. Hughes

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Physics
University of Toronto

© Copyright by James L. P. Hughes 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

0-612-41441-8
Linear and nonlinear optical properties of semiconductors: Theory and calculations

James L. P. Hughes, Doctor of Philosophy, 1998
Department of Physics, University of Toronto

Abstract

We present calculations of the linear dielectric function, the second harmonic generation (SHG) susceptibility, and the linear electro-optic (LEO) coefficient over a broad frequency range for various semiconductors. The full-potential linearized augmented plane-wave (FLAPW) method has been employed in the calculation of the electronic structure. This first-principles all-electron approach is well suited for accurate investigations of optical response in semiconductors. The explicitly nondivergent formalism of Sipe and Ghahramani forms the basis for our calculations. Indeed, our work is the first comprehensive implementation of this formalism. We have also developed an accurate and numerically efficient method for performing the Brillouin zone integrations necessary to calculate the optical response functions.

The optical response in zinc-blende GaAs and GaP has been investigated. Our results for linear response are in reasonable agreement with other theoretical calculations and experiment. In contrast, a conclusive comparison for second-order response is difficult given the limited availability of experimental data and frequency dependent calculations. We argue, however, that on the basis of our LEO results the calculated value of the SHG susceptibility at zero frequency is reasonable.

For the first time, we have investigated the frequency dependent optical response in the wide band gap wurtzite semiconductors, GaN and AlN. The linear and second-order response has been analyzed in terms of the electronic structure and the predictions of the bond charge model. Various symmetry properties of the second-order susceptibilities have also been investigated.
In order to understand the role that crystal structure plays in optical response, we have calculated the linear dielectric function and the SHG susceptibility in both zinc-blende and wurtzite II-VI semiconductors. This analysis is facilitated by calculating the response in the same coordinate reference frame. We have found that second-order response is strikingly more sensitive to structure than linear response.

Finally, we have developed a new local field corrected optical susceptibility formalism for linear response. This approach has several advantages over existing formalisms. We implement our approach calculating the below band gap linear response in silicon.
To My Wife
Acknowledgements

Although many people have assisted in the successful completion of this thesis, none has done so more directly than my supervisor, John Sipe. I am grateful for having had the opportunity to lean on his expansive knowledge and intuition in facing many challenging physics problems. Thanks to John, I take away from my studies not only a broader and deeper understanding of physics, but perhaps more importantly, an eagerness to continue to learn. On a more personal level, I have always appreciated what I felt was John’s genuine concern for the welfare of his students. John, it has been an honour to have had the opportunity to work with you.

I have also had the chance to work with a terrific group of people in the Department of Physics at the University of Toronto over the past five years. My collaborations with Anatoli Shkrebtii, Yongjiang Wang, and Rosen Atanasov have been very rewarding and thoroughly enjoyable. I have learned a great deal about physics from each of them and am privileged to call them all friends as well as colleagues. Many thanks as well to Andrew Datars, Vatche Deyirmenjian, Claudio Aversa, Alain Hache, and Wen-Chin Wu whose help, guidance, and company made my studies at the University of Toronto a successful and very memorable experience. It was a pleasure to work with you all.

I would also like to acknowledge the Natural Science and Engineering Research Council of Canada, the Ontario Graduate Scholarship program, and the Department of Physics at the University of Toronto for their financial assistance. Without their support this research would not have been possible. The American Physical Society is also acknowledged for authorizing to include here previously published papers. These papers are listed as the last four citations at the end of Chapter 1.

It has been sometimes difficult to explain to my family exactly what it is that I have been doing for the past five years. Discussing the details of my Ph.D. thesis
was often the quickest way to prompt a change in subject! My parents have given me their unquestioning support, love and encouragement, which has always meant a great deal to me. Their confidence in my ability to succeed has been a tremendous source of strength. Thanks Mom and Dad. I would also like to say thank you to my brother and sisters. Mark, Kathy, and Chris, I wish I could convey to you the value of the friendship we have shared over the years.

Finally, I would like to thank my wife, Tania. She has been an endless source of support and love. Over the course of my Ph.D., Tania has helped me face the many challenges, encouraged me through the disappointments, and shared in my successes. Tania, I thank you and will be forever grateful.
# Contents

1 **Introduction** ........................................ 1
   Bibliography ............................................. 11

2 **Linear and Second-Order Optical Response Susceptibilities** 15
   2.1 Introduction ........................................ 16
   2.2 Response Functions ................................... 17
   2.3 Scissors Approximation and Implementation ............. 22
   2.4 Conclusions ......................................... 26
   Bibliography ............................................. 27

3 **Calculational Method** .................................. 29
   3.1 FLAPW Method ....................................... 30
   3.2 Numerical Calculation of Optical Susceptibilities ....... 34
   Bibliography ............................................. 41

4 **Calculation of Second-Order Optical Response in Semiconductors** 43
   4.1 Introduction ......................................... 44
   4.2 Method of Calculation ................................ 46
      4.2.1 Electronic Structure ............................ 46
      4.2.2 Response Functions .............................. 46
   4.3 Results and Discussion ................................ 49
   4.4 Conclusions ......................................... 61
   Bibliography ............................................. 63
# Calculation of Linear and Second-Order Optical Response in Wurtzite GaN and AlN

- **5.1 Introduction** ........................................ 68
- **5.2 Calculational Method** .................................... 69
- **5.3 Results and Discussion** ............................... 71
  - **5.3.1 Band Structures** .................................... 71
  - **5.3.2 Linear Optical Response** ............................ 73
  - **5.3.3 Second-Order Optical Response** ....................... 78
  - **5.3.4 Symmetries of \( \chi^{(2)} \)** ......................... 88
  - **5.3.5 Miller’s Delta** ....................................... 89
- **5.4 Conclusions** ........................................... 91

# Bibliography .................................................. 94

# Comparison of Calculated Optical Response in Cubic and Hexagonal II-VI Semiconductors

- **6.1 Introduction** ........................................... 98
- **6.2 Method** .................................................. 99
- **6.3 Results and Discussion** ............................... 100
  - **6.3.1 Structural Differences and Notation** .................. 100
  - **6.3.2 Band Structures and Linear Response** ............... 103
  - **6.3.3 Second-Order Response** ............................... 109
- **6.4 Conclusions** ........................................... 116

# Bibliography .................................................. 118

# Local Field Effects in Linear Optical Response: A New Approach

- **7.1 Introduction** ........................................... 122
- **7.2 The Local Field** ......................................... 129
- **7.3 Optical Response Formalism** ............................ 131
  - **7.3.1 Single-Particle Wavefunctions** ....................... 131
  - **7.3.2 Hamiltonian** .......................................... 136
  - **7.3.3 Linear Susceptibility** ................................. 137
7.3.4 Self-Consistent Equation ............................ 141
7.4 Calculational Details .................................... 142
7.5 Results and Discussion ................................. 146
7.6 Conclusions ............................................. 152
Bibliography .................................................. 154

8 Conclusions .................................................. 157

Appendices ..................................................... 160

A Susceptibility Conventions ............................... 160

B LEO Susceptibility .......................................... 164

C Local Field .................................................. 166

Bibliography .................................................. 172
Chapter 1

Introduction

Nonlinear optics is the study of the interaction between intense electromagnetic radiation and matter. It is a field that is now more than thirty years old, originating in the discovery of second harmonic generation by Franken et al. in 1961 [1]. This discovery was made shortly after the first demonstration of a laser. Experimental studies of nonlinear optical processes instigate and eventually test theoretical formulations and calculations of these properties. In particular, the recent development of the titanium sapphire laser in addition to the suggested application of synchrotron radiation sources to semiconductors has and will make accurate investigations of the dispersion of nonlinear optical susceptibilities possible. The field has seen a rapid development in both theoretical and experimental knowledge over the past three decades. There remains, however, significant challenges to both the theorist and experimentalist. From a theoretical perspective, advances still need to made in the understanding, formulation, and calculation of the nonlinear optical properties of material systems.

The interest in the nonlinear optics of semiconductors can largely be attributed to two primary motivations. The first concerns the possible exploitation of nonlinear optical processes in technological applications. For example, semiconductors could be used as frequency converters with the purpose of extending the spectral range covered by coherent sources. A second motivation in studying nonlinear optical processes is that they can be used to obtain information about the microscopic properties of materials. Electronic energy levels, surface states, and the microscopic motion of
charge in a crystal are examples of interesting properties that could be explored within the context of nonlinear optical response.

Optical nonlinearity describes the response of a material system that is dependent in a nonlinear way on the strength of the optical field. This statement can be made more definitive by expanding in a schematic way the polarization or dipole moment per unit volume, \( P(t) \), in a power series in the strength of the electromagnetic field \( E(t) \). A scalar equation follows as

\[
P(t) = \chi^{(1)}E(t) + \chi^{(2)}E(t)E(t) + \chi^{(3)}E(t)E(t)E(t) + \ldots \quad (1.1)
\]

Here, \( \chi^{(1)} \), \( \chi^{(2)} \), and \( \chi^{(3)} \) are the linear, second-order, and third-order optical susceptibilities. Developing quantum mechanical expressions for these susceptibilities as well as making the actual calculations still presents a very challenging problem to the theorist. This thesis focuses on the linear and second-order susceptibilities with the aim of expanding the current level of understanding of these optical processes.

Linear optical response has been thoroughly investigated in many empirical and \textit{ab initio} full band structure calculations [2, 3, 4, 5, 6, 7, 8, 9, 10]. There have been, however, very few calculations of second-order response in semiconductors [11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. The reasons for this are clear. The expression for the linear susceptibility, within a set of well understood approximations, is more straightforward than those for the second-order susceptibilities under the same set of approximations. Moreover, the computational requirements for a linear response calculation, while not trivial, have not been prohibitive for some time. The linear response function also tends to be less sensitive to details in the electronic band structure, and so more approximate calculational schemes have been sufficient. Despite the accessibility of these full band structure calculations to theorists for over two decades, the agreement with experimental data remains, in many cases, quite poor. This reflects the difficulty in accurately formulating even linear optical response in that the approximations used in its usual formulation are evidently not well-founded. The exclusion of local field and excitonic effects from standard linear susceptibility expressions are thought to be
the chief reasons for the existing discrepancies. We note in passing that some work has been done to address these effects on the linear optical properties of semiconductors for near band edge response [21]. Our focus here, however, is on broad frequency response, and so a full-band structure formulation is necessitated.

In calculating second-order optical response in semiconductors, these problems are merely compounded. The second-order response functions not only present a substantial increase in numerical expense over those for linear response, but they are also highly sensitive to the details of the calculation itself. This imposes further demands in terms of computational requirements. In addition, the usual quantum mechanical expressions for second-order susceptibilities, at the level of the independent particle approximation, are problematic in that they exhibit unphysical divergences at zero frequency [22]; these divergences have, until recently, only been eliminated with the use of sum rules. Yet, several calculations of these response functions have been undertaken; these have tended to be either semi-empirical in nature or have focused on purely zero frequency response.

An important aspect of nonlinear optical calculations is the electronic structure method that is used to determine the velocity matrix elements and eigenenergies that appear in susceptibility expressions. There is a wide spectrum of approaches that have been applied to the calculation of both linear and nonlinear response. These range in complexity and approximation from simple empirical approaches, to more intermediate pseudopotential plane wave methods, to the most complex first principles all-electron methods.

The empirical tight binding (ETB) method [23, 24, 25, 26] is a common empirical approach used in optical calculations. The essential ingredient in such a scheme is the fitting of various calculational parameters to experimentally determined values. While such a method yields, by construction, a good description of the linear optical properties of semiconductors, its usefulness in nonlinear response calculations is in doubt. As well, from a theoretical perspective, it is more meaningful to have an \textit{ab initio} electronic structure method at the heart of an optical response calculation.

In moving beyond an empirically based electronic structure approach, one pays the
price of increased numerical expense. Computational requirements can be mediated, however, if more approximate \textit{ab initio} methods are used. The pseudopotential plane wave (PPW) approach is such a method, and as a consequence it is very commonly employed in optical response calculations [27, 28]. Results obtained in applications of this method are fairly close to experimental data in below band gap response, but there is essentially only qualitative agreement above the band gap. Moreover, the effects that the kinds of approximations inherent in this method have on optical response has not been clear. In addition, the PPW approach demands a dramatic increase in computational expense to treat systems containing heavier elements. To some extent, this limits the advantages of using the PPW method for such systems.

All-electron methods are well-suited to making accurate calculations of a wide variety of materials [6, 29, 30]. Such methods are truly “first-principles” in nature as they treat both the core and valence electrons in a solid. These methods also have the ability to handle materials that contain elements deep in the periodic table. Because all-electron methods approach the solution of the electronic structure in such a fundamental way, they can be considered the most accurate and refined \textit{ab initio} schemes for calculating optical response in a semiconductor. Of course, commensurate with this accurate treatment is a significant increase in computational time and memory. Consequently, these methods are less flexible in terms of the complexity of the materials that can be investigated.

A primary focus in this thesis has been the desire to make accurate calculations of simpler semiconductors systems rather than calculating more complex systems using more approximate methods. As such, we have chosen to use what is considered an extremely precise all-electron approach in our calculations. We have employed the full-potential linearized augmented plane-wave (FLAPW) method in much of the work presented in subsequent chapters. This approach solves the Kohn-Sham equations in a semiconductor in a self-consistent procedure, allowing for the relaxation of both the core and valence states. Our work is the first to implement the FLAPW method in optical response calculations over a broad frequency range. We believe that such calculations are accurate enough to be benchmarks for computationally
simpler methods.

Recently, we have compared the linear and second-order optical response from these three rather diverse electronic structure approaches [31]. Although there are many different calculations presented in the literature, they each use a distinct numerical approach and formalism. By carrying out this optical response investigation ourselves, we have been able to eliminate differences associated with all calculational details except the electronic structure method itself. We have found that the ETB method is very successful in predicting linear response, but is far from giving even rough qualitative agreement for second-order response. A striking result of this investigation has been the similarity in the results between the PPW and FLAPW methods for both linear and nonlinear response. These results lend support to the justification of using the pseudopotential method in calculations of optical properties.

Modern *ab initio* approaches to calculating optical response have relied almost exclusively upon density functional theory (DFT) within the local density approximation (LDA) as a means of determining the electronic band structure. This involves solving the Kohn-Sham equations to obtain the electronic energies and wavefunctions [32]. Although DFT has been widely successful in predicting the ground state properties of material systems, the validity of its extension to optical susceptibility calculations is somewhat in doubt. Optical expressions are usually formulated in terms of a sum over both the occupied and unoccupied states of the system. As DFT is essentially a ground state theory, the conduction states have little meaning. Attempts to apply DFT without correcting for this limitation have led to the so-called band gap problem; it has been found that optical excitations start at an energy below that observed in experiment [3, 33, 34]. In some cases, this underestimation of the gap is significant.

Problems associated with the use of DFT in optical calculations have been primarily addressed through quasi-particle corrections. These corrections are usually determined from a GW calculation [35, 36]. The essential task in such a calculation is to obtain an approximation for the electron self-energy operator, which can be written in a perturbation series in the fully screened Coulomb interaction. The
GW approach involves taking only the first term in this expansion, and requires the determination of the one-particle Green’s function (G) and the dynamically screened interaction (W). This work has led to a much better description of the optical gap in semiconductors, and as a consequence, linear optical response. The problem with GW calculations is their computational complexity and requirements. This problem essentially precludes investigations of nonlinear optical susceptibilities of complicated semiconductor structures. To overcome these limitations while still addressing the underestimation of the band gap, a simple alternative method has been proposed [7]. This approach uses the so-called scissors correction, which rigidly shifts the conduction states to higher energy while keeping the electronic wavefunctions the same. The underlying motivations for such an approximation are derived from the results of full GW calculations.

Prior to the work presented in this thesis, no implementation of the scissors approximation to nonlinear optical response calculations over a broad frequency range had been undertaken. The scissors correction affects the susceptibility expressions in a way that is different from what might be naively assumed. As a result, this correction cannot be included in an ad hoc way within a susceptibility formalism. Indeed, a careful treatment of the velocity matrix elements is required to achieve a correct analytic expression for the susceptibilities.

In addition to issues surrounding the use of a DFT based electronic structure calculation in the evaluation of the optical response functions, perhaps of more fundamental concern is the susceptibility formalism itself. Quantum mechanical expressions for linear and second-order susceptibilities within the independent particle approximation have been in existence for some time [22]. While the linear response function is well behaved in this formulation, the second-order function is not. Calculations of the nonlinear optical properties have been plagued by unphysical divergences for fields of zero frequency. Aspnes [37] was able to show that these divergences were only apparent in the case of cubic symmetry, and recently this was verified for arbitrary symmetry with the derivation of a new sum rule [17]. This situation is problematic for two reasons. First, the use of a formalism that is manifestly divergent raises questions
as to its veracity, and second, calculations of each new susceptibility could require the development of a new sum rule.

Recently, an independent particle formalism was presented by Sipe and Ghahramani (SG) [38]. Their approach is completely free of any unphysical zero frequency divergences and sets out the prescription for determining the expression of any nonlinear optical response function. This work has not only formed the basis of many of the calculations presented here, but is also being used in the recent nonlinear response calculations of Rashkeev et al. [39, 40]. Much of the work presented in this thesis, however, can be viewed as the first *ab initio* full band structure implementation of this formalism.

For the first order response of a material system to an applied electric field there is only one susceptibility present; this is referred to as the linear susceptibility, and is usually presented in terms of the dielectric function. For second-order response, however, there are several distinct susceptibilities to consider. The most commonly investigated response function is that for second harmonic generation (SHG). This can be viewed as a special case of sum frequency generation where the frequencies of the incident fields are degenerate. It is natural to include the SHG susceptibility in any second-order response calculation as it has been the focus of many experimental studies and theoretical investigations. As a result, comparisons with other calculational approaches and with available experimental data is possible. Another reason to calculate the SHG susceptibility is that its expression is the simplest of any nonlinear response function. This greatly simplifies the coding of this particular susceptibility.

Another second-order response function we have chosen to consider in this thesis is that for the linear electro-optic (LEO) effect within the clamped lattice approximation. While there has been some experimental studies of this effect [41], the calculation of the LEO coefficient is significantly behind that for the SHG susceptibility [42, 43, 44]. The primary reason for this has been the lack of a complete quantum mechanical expression for the LEO coefficient. With the advent of the formalism of SG, however, a complete and nondivergent expression is easily attainable. This thesis constitutes the first serious investigation of this susceptibility over a broad
range of frequencies. In addition, by calculating two different second-order susceptibilities, various symmetry properties of second-order response functions can be more completely tested. This provides for a solid check on our calculation as well as on the predicted values for these susceptibilities. This is both interesting and instructive as such a comparison between calculated second-order response functions at the full band structure level has not been previously possible.

The SG formalism is developed within the independent particle approximation. By this it is meant that excitonic and local field effects have been excluded. Excitonic effects are associated with the interaction between an electron and a hole. It is believed that the inclusion of such effects should dramatically improve the description of the bulk linear optical properties [45], and by extension, the nonlinear properties. Addressing excitonic effects, however, is a very challenging theoretical problem. Rigorous microscopic treatments of these effects for linear response within an ab initio approach have only very recently been proposed [46, 47, 48]. It is not yet clear if such approaches can be extended to second-order response given current computational capabilities.

Local field corrections to the optical response of a semiconductor are also thought to be important in obtaining an accurate description of the optical susceptibilities. Moreover, such corrections are needed as a first step in making any many-body calculation of the electronic structure in a semiconductor. These effects are associated with the treatment of the microscopic electric field in a material system. In usual optical response formalisms, the field in a crystal is simply taken to be the macroscopic Maxwell field. In a semiconductor, however, this approximation leads to non-trivial errors in a susceptibility calculation. Over thirty years ago, Adler [49] and Wiser [50] proposed a method for accounting for local field effects in the linear response of solids. Their approach has been extended to address issues surrounding the use of DFT [51], and has been applied in various investigations [52, 53, 54, 55, 56, 57]. There are several limitations of such an approach which have prompted us to reformulate local field corrections in the optical response of solids. Chief among these are that the Adler and Wiser method is predicated on relating scalar rather than tensor quantities, is
not overly efficient, and its extension to higher-order response is unclear.

Ideally, one would like to incorporate local field corrections within the formalism of SG. This would maintain the advantageous properties of the SG approach, while accounting for a possibly significant correction to the optical properties of semiconductors. In this thesis we have developed a formalism that achieves these goals for linear optical response. Clearly, applying a local field corrected approach to broad frequency response demands a significant increase in computational expense. To ensure the numerical tractability of these calculations we have chosen to use a PPW rather than FLAPW method for the electronic structure calculations. While this is a departure in both spirit and substance from our earlier work, our focus still remains the same: Namely, to make accurate calculations of the optical properties of semiconductors. Ultimately, we are confident that advances in accessible computational power will allow for a more rigorous connection to be made between different aspects of our work.

In summary, the goal of this thesis has been twofold. First, to make state-of-the-art calculations of the linear and second-order optical properties of a range of semiconductors of physical and technological interest. This has been done by using a first principles all-electron electronic structure method in the form of the FLAPW approach [29, 30] and employing the explicitly nondivergent formalism of SG. Second, to extend the analytic formulation of optical response to account for local field corrections. This effort has been restricted to the linear susceptibility, but the formalism presented here lays the groundwork for a future extension to higher-order optical processes.

This thesis is essentially divided into three parts. In the first, the details of the formalism and of the calculational method are presented. The second concerns the application of this method to the calculation of the optical properties of various semiconductors. Finally, the last part addresses the extension of the optical response formalism to account for local field effects.

Chapters 2 and 3 concern the method underlying the calculations of the optical properties of semiconductors presented in later chapters. In Chapter 2 the analytic
expressions for the linear dielectric function, and the SHG and LE0 susceptibilities are given. These are derived from the formalism of SG, and so are free of any unphysical zero frequency divergences. In this chapter we also detail our implementation of the scissors approximation within the SG formalism. In Chapter 3 we address issues related to the actual calculation of the optical susceptibilities. Specifically, aspects of the FLAPW method are discussed as well as the means by which the Brillouin Zone integrations, which appear in the expressions for the response functions, are carried out.

In Chapters 4, 5, and 6, the actual calculations of the optical properties of various semiconductors are presented. In Chapter 4 we focus on the optical response of GaAs and GaP. These two zinc-blende semiconductors have been studied both theoretically and experimentally by various groups, and so allow for a substantive comparative analysis between the approach we pursue here with alternative methods. The optical properties of wurtzite GaN and AlN are presented in Chapter 5. This work constitutes the first calculation of the linear and second-order susceptibilities over a broad range of frequencies in these materials. In Chapter 6 the similarities and differences in optical response between zinc-blende and wurtzite structures are investigated. The II-VI semiconductors are a natural choice for such an analysis as these materials can, for the most part, readily crystallize into both structures.

Finally, in Chapter 7 a new linear response formalism that incorporates local field corrections is detailed. Our approach is an alternative to the usual Adler and Wiser method and uses the formalism of SG as the basis for its development. We present results for the below band gap linear response of silicon as a test case for the method. A pseudopotential plane wave approach has been used to determine the electronic wavefunctions and energies for reasons of computational tractability.

Most of the work in this thesis has either been published or submitted for publication. Specifically, Chapters 2 and 4, with some minor modifications, have been published in a single article [58]. Chapter 5 has been published largely as presented here [59]. We have submitted Chapter 6 [60] and Chapter 7 [61] separately for publication in Physical Review B.
Bibliography


[31] A. I. Shkrebtii, J. L. P. Hughes, and J. E. Sipe, to be published.


Chapter 2

Linear and Second-Order Optical Response Susceptibilities

The optical susceptibility conventions as well as the analytic expressions for the linear dielectric function and two second-order susceptibilities are presented. These expressions are based on the formalism of Sipe and Ghahramani and so are free of any unphysical divergences at zero frequency. We also discuss the scissors approximation as a means of addressing the so-called band gap problem, and show how it is incorporated in a straightforward way within this susceptibility formalism. In subsequent chapters, we will present a series of calculations using the expressions detailed here.
2.1 Introduction

Various expressions for the linear dielectric function and the second harmonic generation (SHG) susceptibility at the level of the independent-particle approximation have been presented in the literature [1, 2, 3]. Yet even at this relatively simplistic level, calculations of the SHG susceptibility for semiconductors have been plagued by unphysical divergences at zero frequency. These divergences, however, have been shown to be only apparent, and can be eliminated from the susceptibility expressions with the use of suitable sum rules [1, 4]. This situation was clearly undesirable from both a theoretical and practical perspective as new response functions to be calculated would require a new sum rule in order to eliminate any unphysical divergence.

These issues were resolved with the development of the formalism presented by Sipe and Ghahramani [5]. Their approach has the advantage of being free from such divergences without the use of sum rules, and so is an appropriate starting point for modern calculations of optical susceptibilities. Furthermore, optical coefficients for which no sum rule has been determined are now easily calculable. This is precisely the case for the linear electro-optic (LEO) susceptibility within the "clamped lattice" approximation for which no prior non-divergent microscopic quantum mechanical expression exists. It should also be noted that the approach of Dal Corso et al. [6] for second-order response is also free of divergences at zero frequency. This approach is based on a Wannier function treatment and an application of the "2n+1" theorem. It is not yet clear if such a method is amenable to above-band gap calculations of optical response.

The evaluation of the optical coefficients presented in this chapter involves a summation over both the occupied valence states and the unoccupied conduction states. If one uses the conduction states associated with the solution of the Kohn-Sham equations, then the so-called band-gap problem arises: an underestimation of the excitation energies. This problem can be addressed by performing a GW calculation as a means of determining the quasi-particle energies and wavefunctions [7]. This approach has the disadvantage of being extremely computationally expensive. An
alternative approach, proposed by Levine and Allan [8], is to use a "scissors" approximation; this has the effect of simply adjusting the conduction state energies upward. The amount of this rigid adjustment is usually determined from either a full GW calculation or from experimental data. The success of this rather crude approximation lies in the fact that the quasi-particle wavefunctions are very similar to the Kohn-Sham wavefunctions associated with the conduction states. A subtle aspect to applying the scissors approximation is that the velocity matrix elements that appear in susceptibility expressions are also modified.

The susceptibility conventions and expressions are presented in Sec. 2.2. In Sec. 2.3 the scissors approximation is developed and discussed in the context of our susceptibility formalism. Finally, our conclusions are presented in Sec. 2.4.

2.2 Response Functions

We begin with results that follow directly from the independent particle approximation; in the following section we describe the modifications that must be made to implement the "scissors" approximation.

To establish our convention for the optical susceptibilities, we define the electric field and the polarization in terms of their frequency components as

\[ E(t) = \sum_n E(\omega_n) e^{-i\omega_n t} \quad (2.1) \]

\[ P(t) = \sum_n P(\omega_n) e^{-i\omega_n t}. \quad (2.2) \]

where the summation extends over positive and negative frequencies \( \omega_n \). We adopt the convention that a zero frequency component is to be included twice in the sums of Eqs. 2.1 and 2.2; then \( 2E(0) \) is the actual value of the DC electric field. For completeness and clarity, we discuss matters of convention and definition in more detail in Appendix A.

At the level of linear response the polarization is given in terms of the electric
field by [5]

\[ P^a(\omega) = \chi^{ab}(-\omega; \omega) E^b(\omega), \]  

(2.3)

where superscripts indicate Cartesian components and are to be summed over if repeated; the linear susceptibility is given by

\[ \chi^{ab}(-\omega; \omega) = \frac{e^2}{\Omega \hbar} \sum_{nmk} f_{nm} \frac{r^a_{nm}(k)r^b_{mn}(k)}{\omega_{mn}(k) - \omega}. \]  

(2.4)

Here and below \( n, m \) etc., label energy bands; \( f_{mn} \equiv f_m - f_n \), with \( f_i \) the Fermi occupation factor which, for the clean, cold semiconductors we study here, is zero or unity. The wavevectors \( k \) range over the Brillouin zone, spaced as required by the normalization volume \( \Omega \). The factor of two to account for spin degeneracy is not included in Eq. 2.4 or any subsequent equation involving a summation over \( k \); spin degeneracy is accounted for in all expressions below where the summation over \( k \) has been converted to an integral. The frequency differences \( \omega_{mn}(k) \equiv \omega_m(k) - \omega_n(k) \), where \( \hbar \omega_n(k) \) is the energy of band \( n \) at wavevector \( k \); frequencies \( \omega \) such as that appearing in the denominator of Eq. 2.4 should be interpreted as \( \omega + i\Omega^+ \), where \( \Omega^+ \) is a small positive quantity which is allowed to vanish at the end of the calculation. Finally, the \( r_{mn} \) are the matrix elements of the position operator, excluding the diagonal part [9]; for \( \omega_{nm} \neq 0 \) we have

\[ r^a_{nm}(k) = \frac{v^a_{nm}(k)}{i\omega_{nm}}, \]  

(2.5)

where \( v^a_{nm}(k) = m^{-1} p^a_{nm}(k) \), \( m \) is the free electron mass, and \( p_{nm} \) is the indicated momentum matrix element.

The dielectric function \( \epsilon^{ab}(\omega) \equiv 1 + 4\pi \chi^{ab}(-\omega; \omega) \), and so the usual expression for the imaginary part of \( \epsilon^{ab}(\omega) \), \( \epsilon_2^{ab}(\omega) \), follows from Eq. 2.4.

\[ \epsilon_2^{ab}(\omega) = \frac{e^2}{\hbar \pi} \sum_{nm} \int d{k} f_{nm} \frac{v^a_{nm}(k)v^b_{mn}(k)}{\omega_{mn}^2} \delta(\omega - \omega_{mn}(k)), \]  

(2.6)

where we have converted to an integral over the Brillouin zone.
For second-order response we generally follow the susceptibility convention of Boyd [10]. Then a nonlinear polarization component is related to the electric field by

\[ P^a(\omega_\beta + \omega_\gamma) = \chi^{abc}(\omega_\beta - \omega_\gamma; \omega_\beta, \omega_\gamma) E^b(\omega_\beta) E^c(\omega_\gamma), \]  

(2.7)

where \( \chi^{abc} \) indicates the second-order susceptibility. Cartesian components are again summed over if repeated, as are frequency components \( \omega_\beta \) and \( \omega_\gamma \), but only such that the sum \( (\omega_\beta + \omega_\gamma) \) is held fixed; the susceptibilities are taken to satisfy intrinsic permutation symmetry \( \chi^{abc}(\omega_\beta - \omega_\gamma; \omega_\beta, \omega_\gamma) = \chi^{acb}(\omega_\beta - \omega_\gamma; \omega_\gamma, \omega_\beta) \). In the specific case of second harmonic generation we have

\[ P^a(2\omega) = \chi^{abc}(2\omega; \omega, \omega) E^b(\omega) E^c(\omega), \]

(2.8)

while for the linear electro-optic susceptibility we have

\[ P^a(\omega) = 2\chi^{abc}(\omega; \omega, 0) E^b(\omega) E_{DC}^c, \]

(2.9)

where \( E_{DC} \) is the actual DC electric field. In expressions 2.8 and 2.9, the only sum is of course over the Cartesian components. We note that

\[ \lim_{\omega \to 0} \chi^{abc}(\omega; \omega, 0) = \lim_{\omega \to 0} \chi^{abc}(\omega; \omega, 0) \]

(2.10)

a statement of the equivalence of the SHG and LEO susceptibilities at zero frequency. This result, which is physically expected and follows from our expression for the susceptibilities (see Appendix A), will be relevant in the context of the discussion of our results in subsequent chapters.

The \( \chi^{abc} \) in Eq. 2.7 can generally be written in the form

\[ \chi^{abc}(\omega_\beta - \omega_\gamma; \omega_\beta, \omega_\gamma) = \chi^{abc}_{II}(\omega_\beta - \omega_\gamma; \omega_\beta, \omega_\gamma) + \eta^{abc}_{II}(\omega_\beta - \omega_\gamma; \omega_\beta, \omega_\gamma) \]

\[ + \frac{i}{(\omega_\beta + \omega_\gamma)} \sigma^{abc}_{II}(\omega_\beta - \omega_\gamma; \omega_\beta, \omega_\gamma). \]

(2.11)

The first term in Eq. 2.11 represents the purely interband contribution that would
result if one thought of the system as only a set of "effective" atoms labelled by their crystal momenta. The second term describes the contribution from the modulation of the linear susceptibility by the intraband motion of the electrons. The third term represents that portion of the susceptibility resulting from the modification of the intraband motion by the polarization energy associated with the interband motion. A more thorough analysis and interpretation of these terms has already been presented in the literature [5].

We turn now to the explicit definitions of the terms in Eq. 2.11 specific to the SHG and LEO susceptibilities. Note that all of the expressions below explicitly satisfy intrinsic permutation symmetry.

For the SHG susceptibility the terms are

\[
\chi_{II}^{abc}(-2\omega; \omega, \omega) = \frac{e^3}{\hbar^2} \sum_{nm} \int \frac{dk}{4\pi^3} \left\{ \frac{2f_{nm}}{(\omega_{mn} - 2\omega)} + \frac{f_{ml}}{(\omega_{ml} - \omega)} + \frac{f_{in}}{(\omega_{ln} - \omega)} \right\}
\]

\[
\eta_{II}^{abc}(-2\omega; \omega, \omega) = \frac{e^3}{\hbar^2} \int \frac{dk}{4\pi^3} \left\{ \sum_{nm} \omega_{mn} \frac{r_{nm}^a r_{nl}^b r_{ln}^c}{\omega_{mn}^2 (\omega_{mn} - \omega)} \omega_{nm}^2 \left( \Delta_{mn}^b \Delta_{mn}^c \right) + 2 \sum_{nm} \frac{f_{nm} r_{nm}^a r_{nl}^b r_{ln}^c (\omega_{ml} - \omega_{in})}{\omega_{mn}^2 (\omega_{mn} - 2\omega)} \right\}
\]

\[
\frac{i}{2\omega} \sigma_{II}^{abc}(-2\omega; \omega, \omega) = \frac{e^3}{2\hbar^2} \int \frac{dk}{4\pi^3} \left\{ \sum_{nm} \omega_{mn} \frac{f_{nm} r_{nm}^a}{\omega_{mn}^2 (\omega_{mn} - \omega)} \omega_{ml} r_{ml}^b r_{nl}^c + i \sum_{nm} \frac{f_{nm} r_{nm}^a}{\omega_{mn}^2 (\omega_{mn} - \omega)} \Delta_{mn}^b \Delta_{mn}^c \right\}
\]

with all symbol definitions and conventions as detailed earlier in this section, and where

\[
\Delta_{nm}^a(k) \equiv v_{nn}^a(k) - v_{mn}^a(k).
\]
For brevity, in the expression for $\chi^{abc}(-2\omega; \omega, \omega)$ we have written the $r_{nm}, \Delta_{nm}$ and $\omega_{nm}$ without the explicit $k$ dependence. The use of the curly brackets $\{\}$ with the matrix elements implies a symmetrization with respect to the Cartesian components, $\{r_{m1}^{b}r_{m1}^{c}\} \equiv \frac{1}{2}(r_{m1}^{b}r_{m1}^{c} + r_{m1}^{c}r_{m1}^{b}).$

For the LEO susceptibility, the constituent expressions are

$$\chi^{abc}_{ll}(-\omega; \omega, 0) = \frac{e^{3}}{2\hbar^{2}} \sum_{nm} \int{d\mathbf{k}} \left\{ \frac{f_{nm}}{(\omega_{mn} - \omega)} \left[ \frac{r_{nm}^{a}r_{m1}^{b}r_{n1}^{c}}{\omega_{mn}} + \frac{r_{nm}^{a}r_{m1}^{c}r_{n1}^{b}}{\omega_{mn}} \right] + \left[ \frac{f_{mt}r_{nm}^{a}r_{m1}^{b}r_{n1}^{c}}{(\omega_{ln} - \omega)\omega_{ml}} + \frac{f_{ml}r_{nm}^{a}r_{m1}^{c}r_{n1}^{b}}{(\omega_{ml} - \omega)\omega_{ln}} \right] \right\}$$

(2.16)

$$\eta^{abc}_{ll}(-\omega; \omega, 0) = \frac{ie^{3}}{2\hbar^{2}} \sum_{nm} \int{d\mathbf{k}} \left\{ \frac{f_{nm}r_{m1}^{a}r_{n1}^{c}}{(\omega_{mn} - \omega)\omega_{mn}} \right\}_{b} + \frac{\partial}{\partial \omega} \left( \frac{f_{nm}r_{m1}^{a}r_{mn}^{b} - r_{nm}^{a}r_{mn}^{b}}{2(\omega_{mn} - \omega)} \right)$$

(2.17)

$$\frac{i}{\omega} \sigma^{abc}_{ll}(-\omega; \omega, 0) = \frac{ie^{3}}{4\hbar^{2}} \sum_{nm} \int{d\mathbf{k}} \left\{ \frac{f_{nm}\Delta_{nm}^{a}(r_{nm}^{b}r_{m1}^{c} - r_{nm}^{c}r_{m1}^{b})}{\omega_{mn}(\omega_{mn} - \omega)^{2}} - \frac{2f_{nm}}{\omega_{mn}(\omega_{mn} - \omega)}r_{nm}^{c}r_{m1}^{b} \right\},$$

(2.18)

with the definitions

$$r_{mn;nc}(k) \equiv \frac{-[r_{mn}(k)\Delta_{mn}^{c}(k) + r_{mn}^{c}(k)\Delta_{mn}^{b}(k)]}{\omega_{mn}(k)} - i \sum_{p} \frac{\omega_{mp}(k)r_{mp}^{b}(k)\Delta_{pn}^{c}(k) - \omega_{pn}(k)r_{mp}^{c}(k)\Delta_{pn}^{b}(k)}{\omega_{mn}(k)}$$

(2.19)

$$\left[ \frac{r_{mn}(k)}{\omega_{mn}(k)} \right]_{;c} \equiv \frac{r_{mn;nc}(k)}{\omega_{mn}(k)} + \frac{\partial}{\partial k_{c}} \left[ \frac{1}{\omega_{mn}(k)} \right].$$

(2.20)

Once again, we have not included the explicit $k$ dependence of the $r_{nm}, \Delta_{nm}$ and $\omega_{nm}$ in the expressions for $\chi^{abc}(-\omega; \omega, 0)$. All symbol definitions are identical to those used in the linear and SHG response functions.
2.3 Scissors Approximation and Implementation

While the Kohn-Sham equations form a fundamental starting point for the determination of ground state properties, the unoccupied conduction bands which are calculated have no direct physical significance. Indeed, if they are used naively as single-particle states in a calculation of optical properties for typical semiconductors, the so-called “band-gap” problem results: the absorption starts at much too low an energy [11]. At a basic level, many-body effects must be included in calculating not only the ground state properties but the response to an applied perturbation [12, 13]; to do this, the GW approximation for the self-energy operator has been employed by a number of workers [7, 14, 15, 16, 17, 18]. Instead of following this route, we take the simpler approach of Levine and Allan [8] and introduce a “scissors” approximation to account for the self-energy effects. Although this technique is at best semi-phenomenological, it rests on the fact that GW calculations often indicate little change in the single-electron Green function even with the inclusion of many-body effects: only the energy gap is significantly modified [14]. Further, it has recently been suggested that the scissors approximation can find justification [19] in the necessity of using a polarization-dependent energy functional [20]. Whatever the final verdict on that proposal, our view is that the approach constitutes a simple extension of a ground state calculation that will allow for relatively easy first investigations of linear and nonlinear optical properties: within this spirit it has been used in a number of calculations [8, 21, 22, 23, 24, 25], although we are among the first to employ it in calculating the optical response over a wide energy range.

A point which requires some degree of care is the appearance of the scissors shift in the expressions for the linear and nonlinear response. Actually, for the formalism we use here the result is straightforward; but for comparison with other work we briefly outline the inclusion of the scissors operator, and detail how from different perspectives rather more complicated expressions can result, and have appeared in the literature [22, 25]. Before the scissors operator is included, the Hamiltonian from
which the response functions are calculated [5] is given by

\[
H = \frac{1}{2m} \left( \mathbf{p} - \frac{e}{c} \mathbf{A}(t) \right)^2 + V(\mathbf{r})
\]  

(2.21)

where \( V(\mathbf{r}) \) is a periodic potential, \( \mathbf{r} \) the position operator, and \( \mathbf{A}(t) \) is the vector potential which is related to the electric field by \( \mathbf{E} = -\frac{\mathbf{A}}{c} \). The simplest and most obvious way to correct for the underestimation of the band gap is to include in this Hamiltonian a \( k \)-independent energy shift "projected" onto the conduction states [8]. Within this approximation the Hamiltonian takes the form

\[
\tilde{H} = H + V_s,
\]

(2.22)

with

\[
V_s = \Delta \sum_{\mathbf{k}c} |\mathbf{c}\rangle \langle \mathbf{c}|,
\]

(2.23)

where the sum in \( V_s \) is over all \( \mathbf{k} \) and conduction bands \( c \); \( \Delta \) is the constant energy shift associated with the correction of the band gap; the \( |\mathbf{c}\rangle \) represent single particle eigenstates of the unperturbed Hamiltonian, \( H_o = p^2/(2m) + V(\mathbf{r}) \).

The effect of the inclusion of the scissors operator within our formalism is to modify the expressions for the response functions in a straightforward way. As an illustration, we consider the linear response coefficient. With \( \tilde{H} \) as the governing Hamiltonian, Eq. 2.4 becomes

\[
\tilde{\chi}^{ab}(-\omega;\omega) = \frac{e^2}{\Omega \hbar} \sum_{\mathbf{k}nm} f_{nm} \frac{r^a_{mn}(\mathbf{k}) r^b_{mn}(\mathbf{k})}{[\omega_{mn}(\mathbf{k}) + \frac{\Delta}{\hbar}(\delta_{mc} - \delta_{nc}) - \omega]}
\]

(2.24)

where the \( c \) in the Kronecker deltas refers to a conduction state and the tilde indicates that this coefficient is derived from \( \tilde{H} \). The form of this function is essentially identical to the previous expression without the scissors operator. The only change has been the modification of the energy difference

\[
\omega_{mn} \rightarrow \omega_{mn} + \frac{\Delta}{\hbar}(\delta_{mc} - \delta_{nc}).
\]

(2.25)
An important point to note is that the matrix elements of the position operator remain unchanged. This follows because implicit in the approximation in Eq. 2.22 is the assumption that the effective wavefunctions do not change when many body effects are included [14]. This analysis has been carried out for second-order response with the same results: the new expressions are identical except for the change in the $\omega_{nm}$ as described above.

In practice, one does not directly calculate the matrix elements of the position operator $r_{nm}$, but rather the velocity matrix elements $v_{nm}$. Consequently, a necessary step in the conversion of the response function expressions to numerically amenable form is to make a correspondence between these two different matrix elements. This correspondence is affected by the inclusion of the scissors operator in our Hamiltonian.

We first consider linear response in the absence of the scissors operator. The $v_{nm}$ are then determined from

$$v = \frac{1}{i\hbar}[r, H],$$

with the result of Eq. 2.5 for the elements of the position operator with $\omega_{nm} \neq 0$. In terms of the $v_{nm}(k)$, Eq. 2.4 can be written in the form

$$\chi_{ab}(-\omega; \omega) = \frac{e^2}{\Omega \hbar} \sum_{nmk} f_{nm} \frac{v_{nm}^* (k)v_{mn}^*(k)}{\omega_{mn}^2 (\omega_{mn}(k) - \omega)}.$$  \hspace{1cm} (2.27)

With the scissors operator included in the Hamiltonian, the velocity operator changes. The equation for the modified velocity becomes

$$\tilde{v} = \frac{1}{i\hbar}[r, \tilde{H}] = \frac{1}{i\hbar}[[r, H] + [r, V_s]],$$

from which we obtain the analog of Eq. 2.5:

$$r_{nm} = \frac{\tilde{v}_{nm}}{i[\omega_{nm} + \frac{2}{\hbar}(\delta_{nc} - \delta_{mc})]}.$$  \hspace{1cm} (2.29)

Since the $r_{nm}$ do not change when the scissors operator (2.23) is included we can
determine the relationship between the velocity matrix elements with the scissors operator, and those without. We find

\[ \tilde{v}_{nm} = v_{nm} \frac{[\omega_{nm} + \frac{A}{A}(\delta_{mc} - \delta_{nc})]}{\omega_{nm}}. \]  

(2.30)

It is now possible to write the linear response coefficient within the scissors approximation in terms of the velocity matrix elements, as was done in Eq. 2.27. We use the correspondence of Eq. 2.29 to write

\[ \tilde{\chi}^{ab}(-\omega; \omega) = \frac{e^2}{\Omega h} \sum_{nmk} f_{nm} \frac{\delta_{nm}^a(k) \delta_{mn}^b(k)}{\omega_{mn}(k) + \frac{A}{A}(\delta_{mc} - \delta_{nc}) - \omega} \frac{1}{[\omega_{mn}(k) + \frac{A}{A}(\delta_{mc} - \delta_{nc})]^2}. \]  

(2.31)

Yet, in order to actually calculate this coefficient we need to have it expressed in terms of the unmodified velocity matrix elements, the matrix elements we obtain from the LDA calculation. To do this we use Eq. 2.30 and write Eq. 2.31 as

\[ \tilde{\chi}^{ab}(-\omega; \omega) = \frac{e^2}{\Omega h} \sum_{nmk} f_{nm} \frac{\delta_{nm}^a(k) \delta_{mn}^b(k)}{\omega_{mn}^2[\omega_{mn}(k) + \frac{A}{A}(\delta_{mc} - \delta_{nc}) - \omega]}. \]  

(2.32)

As can be seen from the above equation, the linear response function within the scissors approximation is no more computationally difficult to calculate than that without the scissors approximation (Eq. 2.27). Note that Eqs. 2.24, 2.31 and 2.32 are all equivalent, and that Eq. 2.32 follows directly from Eq. 2.24 by the use of Eq. 2.5. Furthermore, the approach taken in deriving Eq. 2.32 is easily extended to higher-order response functions; we find that, as for linear response, the extension of a response coefficient expression to the scissors approximation consists of two steps: First the function, written using the matrix elements of the position operator, is modified as indicated by Eq. 2.25. Second, we need only use Eq. 2.5 to write the response function in terms of \( v_{nm}(k) \), producing the analog of Eq. 2.32, to make it suitable for numerical evaluation.

We note that the result in Eq. 2.32 is essentially equivalent to that introduced by Del Sole and Girlanda [26], and Levine and Allan [8, 22, 24]. However, Del Sole and Girlanda restrict their discussion to linear response, where we have extended this
approach to any higher-order response function. Levine and Allan do treat second-order susceptibilities within the scissors approximation and discuss the interpretation of the velocity operator that appears in their expressions. Our approach for the scissors operator, although equivalent, is a general extension of this previous work applied to our formalism. Finally, it is worth mentioning that the results of this section follow in a straightforward way within the length gauge formalism of Aversa and Sipe [27].

2.4 Conclusions

We have presented the expressions for the dielectric function, and the SHG and LEO susceptibilities. These expressions are based on the formalism of Sipe and Ghahramani [5] and consequently are free of any unphysical zero frequency divergences. It is important to note that no assumptions are made with regard to crystal symmetry in developing the susceptibility definitions.

The scissors approximation has been discussed and its incorporation into our susceptibility formalism has been presented. It has been shown that the scissors approximation is a computationally attractive means of accounting for the underestimation of the band gap in calculations based on density functional theory. The scissors approximation not only affects the conduction state energies but also the velocity matrix elements. Neglect of this second effect can result in a substantial error in response function calculations.


Bibliography


Chapter 3

Calculational Method

The basis for any optical response calculation is a determination of the velocity matrix elements and eigenvalues for the system under consideration. We describe a first-principles electronic structure technique in the form of the full-potential linearized augmented plane-wave (FLAPW) method, which has been developed by Krakauer and coworkers. This all-electron approach allows for accurate calculations of eigenvalues and wavefunctions for a variety of semiconductors. We also present a method we have developed for calculating the optical response functions themselves. Here the focus is on a means of performing the Brillouin Zone (BZ) integrations in an efficient and precise way. We employ a tetrahedron-random sampling technique which has the advantage of being highly accurate without excessive computational expense.
3.1 FLAPW Method

There are a variety of methods used in electronic structure calculations of semiconductors. These range from the relatively simple empirical approaches, such as the empirical tight-binding method [1], to the more complex first-principles approaches, such as the all-electron full-potential linearized augmented plane-wave (FLAPW) [2, 3] method. The more complex electronic structure methods involve a commensurate amount of computational complexity and expense, but provide for extremely accurate results in a wide variety of materials. The optical susceptibility calculations presented in Chapters 4, 5, and 6 employ the FLAPW method of Krakauer et al. [4, 5] as a means of determining the velocity matrix elements and eigenvalues for various k points in the BZ. In this section we outline some of the pertinent details of this method.

First-principles electronic structure calculations endeavor to solve the Kohn-Sham equations for the electronic wavefunctions, \( \psi_n(k; x) \) and eigenvalues \( E_n(k) \). These equations are given by

\[
\left( -\frac{\hbar^2}{2m} \nabla^2 + V_{\text{ext}} + V_H + V_{\text{xc}} \right) \psi_n(k; x) = E_n(k) \psi_n(k; x) .
\]  

(3.1)

with

\[
n(x) = \sum_{nk} f_n |\psi_n(k; x)|^2 ,
\]  

(3.2)

where \( V_{\text{ext}} \) is the potential generated by the ions, \( V_H \) is the hartree potential, \( V_{\text{xc}} \) is the exchange and correlation potential, and \( n(x) \) is the electronic number density. Note that both \( V_H \) and \( V_{\text{xc}} \) are functionals of \( n(x) \). There are two aspects of such calculations which become very important. The first involves the development of a self-consistent solution by means of an appropriate iterative procedure. This procedure is well-defined, and for modern computers self-consistency is easily attainable. The second concerns the approximations made for the potential terms \( V_H \) and \( V_{\text{xc}} \), in the Kohn-Sham equations. This point is perhaps more critical in that a variety of approximations are currently in use for various electronic structure methods. The
two main approximations that are made are the form of the exchange-correlation potential, and a shape approximation to the Coulomb potential; these approximations will be addressed in turn below.

The underlying idea behind the FLAPW approach is the partitioning of the crystal into two separate regions; non-overlapping "muffin-tin" (MT) spheres that surround each atomic site, and the remaining interstitial region. Further, the charge density, potential, and basis functions (from which the wavefunctions are deduced) have a dual representation corresponding to each of the crystal regions. In the MT spheres, this representation consists of a spherical harmonic expansion; a plane-wave representation is used in the interstitial region. The recognition that the charge density (or potential) has different characteristics in different regions of a crystal leads naturally to their description in terms of this dual representation. For example, the charge density near the atomic sites is largely composed of core electrons which give rise to an almost spherical charge distribution. A spherical harmonic expansion is clearly suitable for such a situation. Away from the atomic cores, the valence electrons provide the dominant contribution and lead to a charge density that is relatively smooth. This behaviour argues for a plane-wave expansion, as this can easily accommodate the smooth distribution with a limited number of plane-waves.

The electronic wavefunctions in the crystal, \( \psi_n(k; x) \), are expanded in a set of basis functions, \( \phi(k + G; x) \). This expansion can be written as

\[
\psi_n(k; x) = \sum_G c_n(k; G)\phi(k + G; x)
\]

(3.3)

where the \( c_n(k; G) \) are the expansion coefficients determined from a solution of the Kohn-Sham equations. The dual representation of the basis functions in the two distinct crystal regions is given by

\[
\phi(k + G; x) = \begin{cases} 
\frac{1}{\Omega_c}e^{i(k+G)\cdot x} & x \in I \\
\sum_{lm} \{a_{lm}^\alpha u_{l}^\alpha(E_l^\alpha, x_\alpha) + b_{lm}^\alpha \hat{u}_{l}^\alpha(E_l^\alpha, x_\alpha)\} Y_{lm}(x_\alpha) & x \in S_\alpha
\end{cases}
\]

(3.4)

where \( \Omega_c \) is the volume of the primitive unit cell. Here, \( I \) represents the interstitial
region and $S_\alpha$ refers to the MT sphere for atomic site $\alpha$. The $a_{\alpha m}^\rho$ and $b_{\alpha m}^\rho$ are simply coefficients determined from matching the basis functions and their derivatives across the boundary between the interstitial and MT regions. The radial functions, $u_\alpha^\rho(E_\alpha^\rho, x_\alpha)$, are obtained by solving the Dirac equation in the limit of zero spin-orbit coupling; the formulation of Koelling and Harmon [6] is used. The other radial functions, $\dot{u}_\alpha^\rho(E_\alpha^\rho, x_\alpha)$, are the energy derivatives of the $u_\alpha^\rho(E_\alpha^\rho, x_\alpha)$.

The connection between the usual APW method [7] and a linearized APW (LAPW) [8, 9] method can now be made clear. The APW method involves a basis function expansion the same as in Eq. 3.4 but with two important differences. First, the term with the $b_{\alpha m}^\rho$ coefficient is not present, and second, the radial functions are solved using the characteristic crystal energy. The LAPW representation is computationally superior to the APW representation primarily because the basis functions involve an energy parameter $E_\alpha^\rho$ rather than the characteristic energy. The basis functions are then energy-independent and so the secular equations become linear in energy; as a result they can be solved for the eigenvalues and wavefunctions using standard diagonalization techniques [9]. This aspect greatly reduces the computational expense in the self-consistent electronic structure calculation. The significance of using the energy derivative radial function is that to a good approximation the function $u_\alpha^\rho(E, x_\alpha)$ (where $E$ is the characteristic energy) is linear over an energy range centered on $E_\alpha^\rho$: $u_\alpha^\rho(E, x_\alpha) = u_\alpha^\rho(E_\alpha^\rho, x_\alpha) + (E - E_\alpha^\rho)\dot{u}_\alpha^\rho(E_\alpha^\rho, x_\alpha)$ [4]. A further advantage of the LAPW method over the APW method is that the basis functions are both continuous and differentiable over the boundary between the MT spheres and the interstitial region.

An important consideration in using the LAPW rather than the APW method is the range of validity of the linearization assumption. Krakauer et al. [4] have found that the band eigenvalues are very insensitive (over a range of typically 1 Ry) to the choice of the energy parameter used in the linearization. We have also conducted a series of tests in order to address this issue. A "windowing" approach was undertaken in which the energy parameters could be set appropriately for a series of energy regions. In a comparison with the band structure generated from the usual single energy parameter approach, we found very little change in eigenvalues over a
range similar to that found by Krakauer et al. The effects of the linearization on optical response is also very small. This is primarily a result of the fact that the greatest contributions to the optical susceptibilities come from bands close to the Fermi level.

One contribution to the Kohn-Sham effective potential is associated with a Coulomb potential. An approximation that has been employed in both APW and LAPW calculations is that this potential is spherically symmetric in the MT spheres. The validity of such an approximation clearly depends on the type of material being investigated. While this approximation is quite good for close-packed metals, it is suspect for more open structures such as semiconductors. In the FLAPW method there is no shape approximation applied to the potential. The spherically symmetric approximation is removed with the use of a method developed by Weinert [10] for solving Poisson’s equation without shape approximation for a periodic charge distribution. The underlying concept behind this method is that the potential outside a localized charge distribution is only dependent on this charge through its multipole moments. This allows for the replacement of the actual charge density in a given MT sphere by another, more convenient density. This “pseudo-charge density” [11] is chosen on the basis that it has a rapidly convergent Fourier representation while still preserving the same multipole moments as the true charge density. The pseudo-charge density is used to obtain the correct Coulomb potential everywhere on the MT sphere and in the interstitial region. The potential inside the sphere is found by solving the boundary-value problem using the actual charge density in this region. The elimination of any shape approximation to the potential in the FLAPW method greatly expands the kinds of materials that can be effectively investigated. As the focus of this thesis is the optical properties of semiconductors, a spherical shape approximation to the potential inside the sphere would be quite limiting, and indeed, could cause significant errors.

The other contribution to the potential that appears in the Kohn-Sham equations involves an exchange-correlation contribution. In the FLAPW method implemented in the work for this thesis, a local density approximation (LDA) is used for this
potential. Specifically, we have used the Wigner interpolation method as a means of determining the one-electron exchange-correlation potential. We have found, however, that the electronic structure is largely insensitive to the specific choice of the exchange-correlation potential within the LDA. The LDA is widely used in electronic structure calculations and gives good results for many ground state properties. Obviously, extensions beyond the LDA are possible although not pursued in the calculations presented in later chapters.

Finally, we conclude this section by considering the procedure for a self-consistent solution of the Kohn-Sham equations that is employed in our FLAPW calculations. The starting point for these calculations is an input charge density for the crystal. We have taken this charge density as simply the sum of charge densities associated with that for isolated atoms located at each lattice site. This approximation, although admittedly crude, does lead in practice to rapid convergence for the materials considered in later chapters. Using the input charge density, the effective Kohn-Sham potential is calculated everywhere in the crystal. From this, the eigenvalues and wavefunctions are calculated from a solution of the Kohn-Sham equations using the basis functions described in Eq. 3.4. The "new" charge density for the crystal is determined from the calculated wave functions. The iterative procedure follows from mixing this charge density with the initial charge density, and then starting the calculation once again. Self-consistency of the wavefunctions and eigenvalues is assumed when differences between the starting and ending charge densities become relatively small. For the materials considered in subsequent chapters, it has been found that fifteen to thirty iterations are sufficient to ensure a high degree of convergence of the charge density.

3.2 Numerical Calculation of Optical Susceptibilities

The calculation of an optical susceptibility involves both a summation over band indices as well as an integration of a function over the BZ (or more accurately, the irreducible part of the BZ). It is straightforward to determine the number of bands that
should be included in the calculation to obtain an accurate result; one simply monitors the convergence of the numerical result with respect to the number of bands included in the susceptibility summation. Moreover, determining the velocity matrix elements and eigenvalues at an appropriate number of bands to ensure accuracy is usually not computationally prohibitive. The integration over the irreducible BZ (IBZ) presents a more complex problem. The main difficulty arises in developing a numerical method that is both accurate, adaptable to a variety of response functions, and yet not overly demanding from a computational perspective. Limiting computational requirements essentially reduces to limiting the number of k points used in performing the integral. We present in this section details of an integration method that we have developed that has been employed in the calculation of the optical response susceptibilities in Chapters 4, 5, and 6.

We consider the linear dielectric function as the basis for our discussion on numerical integration schemes, with the understanding that this discussion extends in a straightforward way to higher-order response functions. The linear dielectric function is given by

\[ \varepsilon^a_b(\omega) = \frac{\varepsilon^2}{\hbar \pi} \sum_{nm} \int d\mathbf{k} \int_{\text{IBZ}} \frac{\nu^a_{nm}(\mathbf{k})\nu^b_{mn}(\mathbf{k})}{\omega^2_{mn}} \delta(\omega - \omega_{mn}(\mathbf{k})), \]  

where the quantities have been defined in Chapter 2, and where the integral is over the BZ. The reduction of the integral to the the irreducible segment of the BZ (IBZ) is carried out using the crystal symmetry, and is addressed in detail in Chapters 4 and 5. There are a variety of techniques detailed in the literature for performing the integration in Eq. 3.5. These will be outlined so as to set the context for the method that has been developed and used in this work.

A common approach to integrating optical response functions is to use the "special" k point scheme proposed by Chadi and Cohen [12, 13], and later by Monkhorst and Pack [14, 15]. The underlying idea behind this approach is that the integral over \( \mathbf{k} \) can be replaced by a sum over a set of \( \mathbf{k} \) points. These \( \mathbf{k} \) points are chosen within a well defined scheme, and attempt to approximate the integral by satisfying a set of conditions. The more \( \mathbf{k} \) points that are chosen, the more conditions that are satisfied,
and consequently the better the approximation to the integral. Each k point that is chosen has a corresponding weight associated with it. Roughly speaking, the integral is then a sum over the integrand at these special k points multiplied by the appropriate weight. This scheme has the advantage of allowing for some means of measuring its own accuracy, but suffers from disadvantages which make it unattractive for optical response function calculations. One disadvantage of the special k point approach is that the number of k points that are required for a reasonable level of accuracy are large. Obviously, for electronic structure calculations where the determination of eigenvalues and velocity matrix elements are computationally expensive, this poses a very definite problem. A second related disadvantage concerns the delta function in Eq. 3.5. As contributions to the response function are often from localized regions in k space, approximating the integral by the number of k points selected in this scheme can lead to significant errors.

An alternative approach is the so-called linear analytic tetrahedron method (LATM). This was developed by Gilat and Raubenheimer [16] for a rectangular geometry, and Lehmann and Taut [17] for a tetrahedral geometry [18]. The underlying idea behind this approach is to combine both analytic and numerical techniques for integrating response functions, which are only known numerically over the BZ. The general approach is to divide the area of integration into tetrahedra at whose vertices the band energies are calculated. Analytic means are then used to approximate the integral under the assumption that the band energies are linearized within a given tetrahedron. This method has been applied for linear response functions, and has also been extended to the second harmonic generation susceptibility [19]. The approach is advantageous in that spectral functions (such as that given in Eq. 3.5) can be easily integrated. Also, the LATM is not overly expensive from a computational perspective, as accurate results can be obtained with relatively few k points. There are, however, some disadvantages that should be noted. First, implementations of this scheme usually assume that the velocity matrix elements are constant over a tetrahedron. The validity of this approximation clearly depends on the size of the tetrahedra as well as on the form of the integrand itself [20]. A second limitation is that implementing this
scheme in other second-order response functions or in other nonlinear susceptibilities is non-trivial [19]. The analytic approach needs to be worked out in each case.

Finally, we consider a third way of evaluating integrals such as that in Eq. 3.5. The method is to use a random-sampling or Monte-Carlo approach to performing the \( k \) space integration [21]. This method is similar to the special \( k \) point scheme already discussed in that the integral is simply converted to a sum over a set of \( k \) points. The \( k \) points in this method are, however, randomly chosen, and so do not satisfy a rigid set of conditions for the integral. The primary advantage of this approach is the ease with which any response function can be integrated. The form of the integrand is unimportant as it is simply evaluated at each randomly sampled \( k \) point. Also, for spectral functions whose contributions are localized in the integration space, a greater number of \( k \) points can be sampled in such regions to give a better approximation of the integral. This method suffers, however, a serious disadvantage. If the calculation of the eigenvalues and matrix elements that appear in the integrand requires significant computation, then this approach is prohibitive since a large number of randomly sampled points are needed to approximate the integral with any degree of accuracy.

We propose an alternative scheme to approximate the integral in Eq. 3.5, or for that matter, in any optical susceptibility. It is a kind of hybrid approach in that it combines aspects of both the linear analytic tetrahedron and the random-sampling methods. The general idea behind this approach is as follows. The region of integration, which is usually the IBZ, is partitioned into a set of large tetrahedra, whose dimensions are on the order of the IBZ itself. Each of these large tetrahedra are then subpartioned into many smaller tetrahedra [22]. The eigenvalues and velocity matrix elements which appear in the integrand are evaluated at the vertices of all of these smaller tetrahedra. We then linearize the eigenvalue differences and the matrix element product in the numerator separately. This partitioning of \( k \) space and the linearization of quantities is in the spirit of the LATM. The difference lies in what is done next to evaluate the integral. Rather than consider an analytic approach to valuing the integral, we simply randomly sample a large number of \( k \) points within these smaller tetrahedra. At each of these random points we do not know the exact
Figure 3.1: Diagram of a single tetrahedron within the IBZ. The \( k \) values at the vertices are determined and the energy differences \( E_{mn} \) are then calculated explicitly. For each \( k \) point sampled within the tetrahedron, \( E_{mn}(k) \) is calculated using a linearization scheme.

band energies or matrix elements, but we approximate these from the linearization based on the values at the vertices.

This discussion can be made more definitive by describing how one would obtain a given eigenvalue difference \( E_{mn} \) at a particular \( k \) point sampled within a tetrahedron. In Fig. 3.1 we plot one of the many tetrahedra that comprise the IBZ. Within our scheme, the values of the \( k \) points at the vertices are determined, and from these the eigenvalue differences are calculated at each vertex from our electronic structure calculation. To determine the eigenvalue difference \( E_{mn}(k) \) at the sampled \( k \) point we use [19]

\[
E_{mn}(k) = E_{mn}(k_0) + b \cdot (k - k_0),
\]

where

\[
b = (\nabla E_{mn}(k))_{k_0}.
\]

If we introduce the vectors
where $k_{00} = k_\alpha - k_0$, and $V$ is the volume of the tetrahedron, and note that $b \cdot k_{00} = E_{mn}(k_\alpha) - E_{mn}(k_0)$, then we can express the gradient vector as

$$b = \sum_\alpha (E_{mn}(k_\alpha) - E_{mn}(k_0))a_{\alpha 0}.$$  

These equations provide a means of calculating the eigenvalue differences at every $k$ point sampled within a given tetrahedron. By extension, one can calculate any quantity within the tetrahedron as long as its value at each vertex was known.

This approach has several advantages. First, we obtain a very good approximation to the integral that is based on an exact eigenvalue and matrix element determination at relatively few $k$ points. For example, there might be on the order of one thousand $k$ points associated with the tetrahedra vertices, but we might sample a thousand $k$ points in each small tetrahedron. This would translate into approximating the integral with a sum over several million $k$ points, while having only solved the secular equations at a fraction of this number. For complex electronic structure schemes such as the FLAPW method, this point cannot be understated. A second and related advantage is that integrals which contain a delta function are easily handled. Since contributions to the integral may be well localized, the huge number of sampled points ensures that these regions are properly accounted for. A third advantage is that any susceptibility can be integrated in a straightforward way. For higher-order response functions, the integrands become increasingly complex. With our method, this added complexity presents no problem as the integrand is simply evaluated at each sampled $k$ point based on our knowledge of components of the integrand at the tetrahedron vertices.

Obviously, the validity of linearizing quantities over a small tetrahedron depends on the size of the tetrahedron itself, as was discussed with respect to the LATM. We have found, however, that this approximation is very good without having to partition the IBZ into too many tetrahedra. Typically, on the order of five thousand small
tetrahedra in the entire IBZ produces excellent results. This number corresponds to an exact eigenvalue and matrix element calculation at less than one thousand five hundred k points.

In short, the integration method we have discussed here combines the best features of two commonly used approaches. It produces an accurate approximation to linear and higher-order optical susceptibilities without becoming computationally prohibitive or requiring a complex analytic treatment. This method is used in the next three chapters where calculations of the dielectric function and two second-order response functions are carried out for a variety of materials.
Bibliography


[18] Rectangular and tetrahedral refer not to the crystal symmetry, but rather the shape of the cells that partition the integration region.


Chapter 4

Calculation of Second-Order Optical Response in Semiconductors

Abstract

We present a first principles calculation of two second-order optical response functions as well as the dielectric function for GaAs and GaP. Specifically, we evaluate the dielectric function $\epsilon(\omega)$ and the second harmonic generation response coefficient $\chi^{(2)}(-2\omega; \omega, \omega)$ over a large frequency range. The electronic linear electro-optic susceptibility $\chi^{(2)}(-\omega; \omega, 0)$ is also evaluated below the band gap. These results are based on a series of self-consistent LDA calculations using the FLAPW method. Self energy corrections are included at the level of the "scissors" approximation, which corrects for the underestimation of the LDA band gap and produces a change in the velocity matrix elements. The analytic expressions for the second-order response functions are free of the unphysically divergent terms at zero frequency which have previously plagued such calculations. Results for $\chi^{(2)}(-\omega; \omega, 0)$ are in good agreement with experiment below the band gap and those for $\chi^{(2)}(-2\omega; \omega, \omega)$ are compared with experimental data where available. We note that despite the equivalence of both of these second-order response functions at zero frequency, there seems to be some discrepancy between the experimental results for these functions in this regime.
4.1 Introduction

While there have been many empirical and \textit{ab initio} full band structure calculations of linear optical response in semiconductors \cite{1, 2, 3, 4, 5, 6, 7, 8, 9}, there have been very few calculations of the nonlinear response. The understanding and calculation of the linear electro-optic (LEO) susceptibility has lagged behind the experimental studies, with most theoretical calculations based on simple phenomenological models \cite{10, 11, 12}. Of the other second-order susceptibilities, most theoretical calculations have been concerned with second harmonic generation (SHG) \cite{13, 14}, and many of these have been restricted to a determination of the response function at zero frequency \cite{15, 16, 17, 18, 19}. To our knowledge the only attempts at calculating nonlinear response functions over a wide frequency range are the work of Huang and Ching \cite{20, 21, 22} and that of Moss and coworkers \cite{23, 24, 25, 26, 27}. As we discuss below, both of these approaches have certain limitations.

In this chapter we present the results for the SHG and LEO response functions, as well as the dielectric function, for GaAs and GaP over a wide frequency range. Since our interest is in the electronic aspects of optical response, we calculate the LEO response function in the “clamped lattice” approximation; our results are then suitable for comparison with experiments involving “DC-like” fields at low enough frequencies that electronic dispersion associated with the DC-like field can be neglected, but at high enough frequencies that lattice motion can be considered frozen out.

The analytic expressions for the nonlinear response functions are based on the formalism of Sipe and Ghahramani \cite{28}, as extended and developed in the length gauge by Aversa and Sipe \cite{29}; the response calculation is at the level of the independent particle approximation. This approach has the advantage that the response coefficients are inherently free of any unphysical divergences at zero frequency, a consequence of a careful treatment and separation of interband and intraband motion. “Sum rules” are not required to eliminate artificial divergences. The recent work of Dal Corso and Mauri \cite{30, 31}, based on an elegant Wannier function approach, is also free of such divergences. But at finite frequency, in particular for frequencies or
frequency sums across the gap, we feel that any approach easily amenable to numerical analysis will likely involve a k-space focus such as that adopted explicitly in this work, simply because the resonances at any given frequency occur in localized regions of k-space.

The full band structure calculation in this work utilizes the full-potential linearized augmented plane wave (FLAPW) method [32, 33] within the local density approximation (LDA). This method has an advantage over that employed by Moss and coworkers [23, 24, 25, 26, 27] in that it is first principles rather than semi-empirical in nature. We have adopted a "scissors" approximation to correct for the band gap, but account for the change in the velocity matrix elements that appear in the response function expressions. Huang and Ching [20, 21, 22] neglect this modification in the matrix elements; based on previous evidence [18], and the results of our own calculations, this can result in a significant error in the determination of the response functions. We have not included local field effects in this work; as suggested by the work of Levine and Allan [18], we do not expect significant corrections for the materials considered here at the level of second-order response. However, the inclusion of local field effects can be done in a straightforward way within our formalism for the response functions. This will be examined further in Chapter 7.

This chapter is organized in the following way. We outline the calculational technique used in determining the non-linear coefficients in Sec. 4.2. The band structures for the materials considered are also presented in this section. In Sec. 4.3 we give our results for linear response, the SHG coefficient and the LEO function. The comparison with experiment and other theoretical calculations is also investigated and discussed. A conclusion and summary of our results are presented in Sec. 4.4.
4.2 Method of Calculation

4.2.1 Electronic Structure

In order to determine the optical response functions in a full band structure approach, one requires the eigenvalues and velocity matrix elements at many $k$ points in the Brillouin zone (BZ). The velocity matrix elements, in turn, require a knowledge of the electronic wavefunctions. For this purpose, we employ a first principles approach in the form of the FLAPW method [32, 33]. This approach has been described in detail in Chapter 3.

We present in Fig. 4.1 the band structures for GaAs and GaP. In both band structures we have adjusted the band gap to agree with experiment using the scissors approximation, as discussed in Chapter 2. The value of the scissors operator was 1.316 and 1.281 eV for GaAs and GaP respectively.

4.2.2 Response Functions

The evaluation of each of the response functions is carried out in a slightly different way, but there are some general considerations which need to be addressed in all cases. One necessary step in evaluating optical response is the reduction of the analytic expressions for the susceptibilities to ones suitable for computation. The details of this procedure for each of the response functions are too long to present here, but we will outline the general approach.

The expressions for the susceptibilities are first written in terms of the velocity matrix elements rather than the matrix elements of the position operator and converted to an integral over the BZ rather than a sum over $k$. If the response function is sought above the band gap, the imaginary part is extracted, for which there is a delta function present in the integrand; in the case of cubic symmetry some further simplifications can be made to the second-order susceptibilities.

For linear response we evaluate the imaginary part of the dielectric function given in Eq. 2.6, which is already in a form amenable to computation. We obtain the
Figure 4.1: FLAPW electronic band structures for GaAs and GaP. The fundamental band gap has been adjusted within the scissors approximation.
real part of the dielectric function by using the Kramers-Kronig relation. The SHG response function, as given in Chapter 2, requires some manipulation prior to being calculated. Again, since we seek the response above and below the band gap, we evaluate the imaginary part and obtain the real part from the Kramers-Kronig relation for this function [34]. The LEO susceptibility is evaluated below the band gap only, where it is of primary interest, and so we calculate it directly in this energy regime.

We use the symmetry of the crystal to reduce the integration over the BZ to one over the irreducible segment of the BZ (IBZ). This is done by applying the operators $P_R$ of the group elements $R$ of the symmetry group $T_d$ (applicable for the semiconductors considered here), to the expansion dyadics of the response tensor. For linear response, we consider the dielectric tensor

$$
\bar{\epsilon}_2(\omega) = \sum_{ab} \hat{a} \hat{b} \epsilon^{ab}_2. \tag{4.1}
$$

and after applying the operators $P_R$ we obtain

$$
\begin{align*}
\sum_R P_R(\hat{x}\hat{x}) &= \sum_R P_R(\hat{y}\hat{y}) = \sum_R P_R(\hat{z}\hat{z}) = 8 \bar{I}
\sum_R P_R(\hat{a}\hat{b}) &= 0, \ i \neq j. \tag{4.2}
\end{align*}
$$

where $\bar{I}$ is the identity tensor. Thus only the diagonal element survives, as is well known. With the use of these symmetry elements and the inversion symmetry of the BZ, the integration over $k$ can be restricted to the IBZ only.

For second-order optical response we proceed in the same way. We write the SHG susceptibility as

$$
\bar{\chi} (-2\omega; \omega, \omega) = \sum_{abc} \hat{a} \hat{b} \hat{c} \chi^{abc}(-2\omega; \omega, \omega), \tag{4.3}
$$

and applying the operators $P_R$ we find

$$
\begin{align*}
\sum_R P_R(\hat{x}\hat{y}\hat{z}) &= 4(\hat{x}\hat{y}\hat{z} + \hat{y}\hat{z}\hat{x} + \hat{z}\hat{x}\hat{y} + \hat{y}\hat{z}\hat{x} + \hat{z}\hat{x}\hat{y} + \hat{y}\hat{z}\hat{x})
\sum_R P_R(\hat{a}\hat{b}\hat{c}) &= 0 \ a, b, c \text{ not all different}. \tag{4.4}
\end{align*}
$$
It can be seen from Eq. 4.4 that there is only one independent component for the SHG susceptibility which we take to be $\chi^{xyy}$; the same result follows for the LEO susceptibility.

In the evaluation of all the response functions, the essential task becomes the integration of a function over the IBZ. This we do in a "hybrid" random sampling-tetrahedron method as has been discussed in detail in Chapter 3.

In the calculation of all the susceptibilities, we directly evaluate the velocity matrix elements and eigenvalues from the FLAPW calculation at 1365 $k$ points in the IBZ; this corresponds to the partitioning of this region into 5184 tetrahedra. We have further partitioned the region around the $\Gamma$ point into 3993 smaller tetrahedra, which requires the matrix elements and eigenvalues at a further 1092 $k$ points. The reason for this finer mesh of points near the $\Gamma$ point is that we have found that there is some sensitivity in the non-linear response functions to numerical approximations in this region [19, 24]. The susceptibilities calculated from this number of $k$ points are only marginally different from a calculation involving only half this number. On this basis, it is clear that our current calculation should suffer only a limited numerical error using this particular integration scheme.

4.3 Results and Discussion

In Fig. 4.2 we present our results for the imaginary part of the dielectric function, $\varepsilon_2(\omega)$, for GaAs and GaP. The experimental results of Philipp and Ehrenreich [35], and Aspnes and Studna [36] are included for comparison. The main features of the linear response function can be attributed to the same regions of the electronic band structure for both materials. These features are largely governed by the joint density of states, whose structure is associated with those regions in the band structure for which pairs of bands are essentially parallel. In particular, using the terminology specific to zinc-blende semiconductors [37], we can identify the dual peak structure at low energy with the $E_1$ and $E_1 + \Delta_1$ transitions. The main peak and the slight shoulder to the left of this peak are due to the $E_2$ and $E'_0$ transitions respectively. Finally, the
high energy feature in the spectrum is attributed to the \( E'_1 \) optical transition.

It is evident from Fig. 4.2 that our calculation predicts the peak positions higher in energy than those in the experimental results. All \textit{ab initio} calculations share a difficulty in correctly predicting both the band gap and the peak positions in the linear response spectrum. The original first principles work of Wang and Klein [2] employed the LDA and achieved some agreement in peak positions, but underestimated the fundamental band gap. A rigid adjustment of the electronic structure calculated by Wang and Klein, following the scissors approximation adopted here, would result in a similar mispositioning of the location of the peaks in the linear response function. The more recent work of Huang and Ching [21, 20] using the orthogonalized LCAO method for the electronic structure calculation and a scissors correction, has obtained only limited agreement with experiment for the dispersion of the dielectric function. Alouani \textit{et al.} [5] have lately achieved some success in calculating linear response for GaAs, accurately predicting both the band gap and the dispersion in \( \epsilon_2(\omega) \). Their method involves adding in sharply peaked potentials within their LDA framework in such a way that the low band gap is suitably “compensated”. Although arguably no more phenomenological than the simple scissors approximation we employ here, its ultimate justification and extension to a more fundamental level is perhaps less clear.

We make two further comments concerning \( \epsilon_2(\omega) \): First, the intensity of the peaks in the calculated function is overestimated in part due to the exclusion of the effects of a finite relaxation time. Second, the experimental results in Fig. 4.2 are taken at room temperature: experimental work suggests that peak positions shift to higher energy at lower temperatures [38]. A shift of approximately 0.1 eV could be expected for equivalent results at low temperature, which would be more appropriate for comparison with this zero temperature theoretical work. Yet it is clear that the \( k \)-independent rigid shift in the conduction states with the corresponding modification of the response function expression achieves only reasonable accuracy in the dispersion of the dielectric function.

We note that there has been some recent success at the level of linear response in work based on quasi-particle calculations which treat the self-energy corrections more
Figure 4.2: Results for the calculated imaginary part of the dielectric function, $\varepsilon_2(\omega)$ (solid line), for GaAs and GaP. Energy bin size is 0.05 eV. Experimental results (dotted line) are from Philipp and Ehrenreich [35], and Aspnes and Studna [36].
carefully [7]. As our emphasis is on the non-linear optical response, we feel that the scissors approach is an appropriate compromise between computationally amenable calculations and accurate results.

The results for the imaginary part of the SHG susceptibility, $\text{Im}\{\chi^{\text{shg}}(-2\omega;\omega,\omega)\}$, are plotted in Fig. 4.3. Although this part of the response function cannot be directly compared with experiment, it can be more meaningfully related to the band structure than can $|\chi^{\text{shg}}(-2\omega;\omega,\omega)|$, to which experiments are more directly sensitive. The structure in $\text{Im}\{\chi^{\text{shg}}(-2\omega;\omega,\omega)\}$ can be attributed to the same general regions in the band structure for both GaAs and GaP. The onset of the function occurs at the $2\omega$ resonance with the $E_0$ optical transition. The first peak is associated with the $2\omega$ resonance with the $E_1$ and $E_1 + \Delta_1$ optical transitions. The second structure in the function for the most part arises from the $2\omega$ resonance with $E'_2$. The complicated structure in the region between 3 eV and 4 eV is associated with an interference between an $\omega$ resonance with the $E_1$ transition and a $2\omega$ resonance with the $E_2$ and $E'_1$ transitions, while the structure between 5.0 eV and 5.5 eV is due mainly to the $\omega$ resonance with the $E_2$ optical peak.

Our results for the imaginary part of the SHG susceptibility show important differences from those of Huang and Ching [20, 21], although there are some similarities in the shapes of the calculated functions. It is important to note that their failure to adjust the velocity matrix elements after rigidly shifting the conduction states to higher energy can result in an appreciable error in their reported values. Levine has suggested that their results could be underestimated by up to a factor of two [18]; this has been corroborated by our own investigations, comparing calculations with and without the matrix elements appropriately modified. The results of Ghahramani et al. [26, 39] are closer to ours, despite the differences in approach to the calculation of the electronic band structure. They employ a semi-ab initio minimal-basis linear combination of Gaussian orbitals (MLCGO) method.

The absolute value of the SHG susceptibility is plotted in Fig. 4.4 for GaAs and Fig. 4.5 for GaP. Experimental data at energies above the gap is very scarce for the materials considered here. The only data we are aware of is for GaAs, and this is
Figure 4.3: Plot of $\text{Im}\{\chi^{xy}(-2\omega;\omega,\omega)\}$ for GaAs and GaP. The energy bin size is 0.05 eV.
Figure 4.4: Absolute value of the SHG susceptibility, $|\chi^{xyz}(-2\omega; \omega, \omega)|$ (solid line), for GaAs. Experimental results are as follows: Parsons and Chang [40] (dotted line); Bethune et al. [41] (dashed line) and Chang et al., [42] (X).

plotted in Fig. 4.4. The experimental results seem largely contradictory and any serious comparison would remain inconclusive.

There is a considerable amount of data available for the LEO susceptibility and, as the chief interest is in the region below the band gap, we have concentrated our investigation to this regime. Two main issues must be addressed, however, in order to make a suitable comparison of our results to the experimental data. The first is that we require an expression that relates the experimentally measured LEO coefficient $r^{xyz}(\omega)$ to our calculated LEO susceptibility $\chi^{xyz}(-\omega; \omega, 0)$; although straightforward, to avoid confusion this is presented in Appendix B. A second issue in the comparison to experiment is that our calculated quantity is in fact the electronic or "clamped" LEO susceptibility. Some of the experimental data is in terms of the unclamped values and so must be modified to yield the corresponding clamped LEO coefficient.
Figure 4.5: Absolute value of the SHG susceptibility, $|\chi^{xyz}(-2\omega;\omega,\omega)|$ (solid line), for GaP.

values. To do this we use the compilation of experimental coefficients presented by Adachi [37], where estimates of the clamped values are given for those experiments which determined only the unclamped values.

In Fig. 4.6 and Fig. 4.7 we plot $\chi^{xyz}(-\omega;\omega,0)$ for GaAs and GaP in pm/V, as this is the more commonly quoted unit in the literature. We have also plotted $|\chi^{xyz}(-2\omega;\omega,\omega)|$ on the same graph, for a purpose discussed below. The experimental data is presented for comparison. To generate the experimental $\chi^{xyz}(-\omega;\omega,0)$ we have used Eq. B.3 with the experimentally measured LEO coefficient $r^{abc}(\omega)$ and the index of refraction from Palik [43, 44].

The results of our calculation are in good agreement with experiment over a large range of frequencies. This agreement is somewhat better for GaAs than for GaP, but given the complexity of this calculation we feel it is nonetheless quite satisfactory. As we have previously discussed, our calculation is based on a formalism that is well-
Figure 4.6: Plot of second-order optical response in GaAs below the fundamental band gap: LEO susceptibility $\chi^{xyz}(-\omega; \omega, 0)$ (solid line) and the absolute value of the SHG susceptibility $|\chi^{xyz}(-2\omega; \omega, \omega)|$ (dotted line). Experimental results for the LEO effect (solid circles) are from [45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55] as compiled by Adachi [37]. The data at 0.117 eV is the SHG experimental data from Table 4.1.

behaved at low frequencies, so the results are believable in the energy regime presented in Fig. 4.6 and Fig. 4.7. To our knowledge this is the first ab initio calculation of the LEO susceptibility for GaAs and GaP. Ghahramani and Sipe have presented the LEO susceptibility for GaAs based on a semi-empirical MLCGO calculation [58], employing the same nonlinear response formalism given here. Our results are close to this previous work, although we obtain better general agreement with the experimental data for all energies. Various other theoretical calculations of the LEO coefficient have been presented, but these are phenomenological in nature and will not be discussed here [10, 11, 12].

As there is considerable interest in the optical response at low frequencies, we have presented in Table 4.1 our results for the dielectric constant, the SHG susceptibility,
Figure 4.7: Plot of second-order optical response in GaP below the fundamental band gap: LEO susceptibility $\chi^{xyz}(-\omega; \omega, 0)$ (solid line) and the absolute value of the SHG susceptibility $|\chi^{xyz}(-2\omega; \omega, \omega)|$ (dotted line). Experimental results for the LEO effect are from Nelson and Turner [56] (open circles) and Berozashvili et al. [57] (solid circles). The data at 0.117 eV is the SHG experimental data from Table 4.1.

and the LEO (clamped lattice) susceptibility for GaAs and GaP at zero frequency. We have included for comparison other theoretical calculations, and experimental results. For linear response our agreement with experiment is excellent for both materials. We recognize that this may be somewhat fortuitous, given the comparison of our results with experiment over a broad frequency range. A shift of our calculated peaks in $\epsilon_2(\omega)$ to lower energies, towards the experimental peaks, would result in a higher calculated $\epsilon(0)$; this would be partly offset by the inclusion of local field effects, which have been shown to reduce the zero frequency result for the materials considered here [18].

The results for $\chi^{xyz}(0)$ require more discussion, in light of the wide range of values which appear. We first recall, as noted in Chapter 2 (see Eq. 2.10) that the SHG and
Table 4.1: The linear and second-order optical response in GaAs and GaP at zero frequency. The results of the present calculation (FLAPW) are compared with other theoretical calculations and experimental data.

<table>
<thead>
<tr>
<th>Material</th>
<th>Method</th>
<th>$\epsilon(0)$</th>
<th>$\chi^{xyz}(0)$ (pm/V)</th>
<th>LE0</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaAs</td>
<td>FLAPW</td>
<td>10.9</td>
<td>96.5</td>
<td>96.5</td>
</tr>
<tr>
<td></td>
<td>Pseudopotential $^a$</td>
<td>11.0</td>
<td>172.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OLCAO $^b$</td>
<td>11.21</td>
<td>251.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MLCGO</td>
<td></td>
<td>104.3 $^c$</td>
<td>104.3$^d$</td>
</tr>
<tr>
<td></td>
<td>Experiment</td>
<td>10.9 $^e$</td>
<td>162±10 $^f$</td>
<td>99.8$^g$</td>
</tr>
<tr>
<td>GaP</td>
<td>FLAPW</td>
<td>9.0</td>
<td>50.3</td>
<td>50.3</td>
</tr>
<tr>
<td></td>
<td>Pseudopotential $^a$</td>
<td>8.8</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OLCAO $^b$</td>
<td>9.29</td>
<td>134.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MLCGO</td>
<td></td>
<td>43.6 $^c$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experiment</td>
<td>9.1 $^h$</td>
<td>74±4 $^f$</td>
<td>44.3$^i$</td>
</tr>
</tbody>
</table>

$^a$Levine and Allan [19]
$^b$Huang and Ching [21]
$^c$Ghahramani et al. [26, 39]
$^d$Ghahramani and Sipe [58]
$^e$Reference [43]
$^f$Levine and Bethea [59] value at 0.117 eV, as revised by Roberts [60]
$^g$Average of experimental values at 0.117 eV, from Adachi [37]
$^h$Reference [44]
$^i$Average of experimental values at 1.08 eV, from Adachi [37]

LEO (clamped lattice) susceptibilities are equal at zero frequency. This is not only a numerical result of the present calculation: We have shown this analytically within our formalism, and in any case it would be expected on physical grounds. Yet there is significant disagreement between the experimental results for the zero frequency SHG and LEO susceptibilities. This disagreement does not seem to have been fully appreciated.

Our calculated results are in closest agreement with the MLCGO calculation of Ghahramani et al. [26, 39]. We note that their calculation and ours are completely independent and are based on markedly different approaches to the determination of the electronic structure and the velocity matrix elements. The orthogonalized LCAO calculation of Huang and Ching [21] leads to a much higher result for $\chi^{xyz}(0)$.
than the present work. And, as we have previously discussed, we believe that they require an adjustment of their velocity matrix elements in accordance with their scissors shift. With such an adjustment their value for $\chi^{xyz}(0)$ would be raised even higher, exceedingly high in comparison with other theoretical calculations and with experimental values. Levine has presented the most comprehensive work below band gap energies of which we are aware [19]. But his formulation of the SHG susceptibility is drastically different from that used here, and thus it is unfortunately difficult to identify the reasons for the disparity between the results.

Yet we note that our calculated results are in extremely good agreement with the LEO experimental results. Given that various LEO experiments done recently seem to corroborate each other, we would suggest that this gives strong support to the value of $\chi^{xyz}(0)$ we have calculated here. As to its disagreement with the SHG experimental results, we note that the values given in Table 4.1 are from Roberts [60], who revised the original values of Levine and Bethea [59]. Their experiments were conducted more than two decades ago, and it is not clear this data should be given as much weight as the much more recent and consistent work on the LEO susceptibilities in attempting to establish an experimentally determined value of $\chi^{xyz}(0)$. Certainly the discrepancy between the experimental results for the two equivalent susceptibilities indicates that new experimental studies would be helpful in resolving this issue.

It is interesting in the context of the current calculation to assess the validity of "Miller's rule" based on our theoretical results. Miller proposed that the quantity

$$\Delta_M(\omega_1 + \omega_2) = \frac{\chi^{(2)}(\omega_1, \omega_2; \omega_1, \omega_2, \omega)}{\chi^{(1)}(\omega_1 + \omega_2)\chi^{(1)}(\omega_1)\chi^{(1)}(\omega_2)}$$

is approximately constant for a wide range of noncentrosymmetric materials, with little variation in frequency [61]. As we have calculated two second-order susceptibilities we can investigate Miller's $\Delta_M$ in both cases. We have for SHG

$$\Delta_M^{SHG}(2\omega) = \frac{\chi^{xyz}(2\omega; \omega, \omega)}{\chi^{(1)}(2\omega)[\chi^{(1)}(\omega)]^2}$$

a form specific to the zinc-blende materials considered here, given that there is only
Figure 4.8: Plot of Miller’s $\Delta_M$ below half the band gap for GaAs and GaP: $\Delta_{M}^{SHG}$ (dotted line) and $\Delta_{M}^{LEO}$ (solid line).
one independent component for the SHG susceptibility and equal diagonal components for $\chi^{(1)}$. For the LEO effect, the corresponding quantity is

$$\Delta_{M}^{LEO}(\omega) = \frac{\chi^{xyz}(-\omega; \omega, 0)}{[\chi^{(1)}(\omega)]^2 \chi^{(1)}(0)}. \quad (4.7)$$

We plot in Fig. 4.8 both of these functions for GaAs and GaP. Several comments can be made about our results for Miller's $\Delta_{M}$. First we note that $\Delta_{M}^{SHG}$ is equal to $\Delta_{M}^{LEO}$ at zero frequency. This result we would expect given the equivalence of Eqs. 4.6 and 4.7 at zero frequency if the second-order susceptibilities are equivalent in this limit as well. As can be seen in Fig. 4.8 there is a definite variation in frequency for both $\Delta_{M}^{SHG}$ and $\Delta_{M}^{LEO}$. Both functions show an increase with increasing frequency, although $\Delta_{M}^{LEO}$ rises less rapidly than $\Delta_{M}^{SHG}$. The zero frequency value of $\Delta_{M}$ for both materials is remarkably similar; this corroborates Miller's original conjecture, but it is clear that an analysis of a wider range of materials would by necessary to test this further.

We note that our theoretical results for $\Delta_{M}^{SHG}$ are very similar to those of Levine [19]. The magnitudes are somewhat different, due to the differences in our $\chi^{xyz}(-2\omega; \omega, \omega)$ results compared with theirs, but the trend in $\Delta_{M}^{SHG}$ is the same. Huang and Ching's [21] results for $\Delta_{M}^{SHG}$ seems somewhat anomalous; they predict $\Delta_{M}^{SHG}$ to be a flat or slightly decreasing function over the energy range plotted in Fig. 4.8.

### 4.4 Conclusions

We have presented results for linear and second-order optical response in GaAs and GaP based on a first principles FLAPW electronic structure calculation. We have employed a response formalism that is free of any unphysical divergences at zero frequency, providing believable results across the entire energy spectrum for any response function. Within this formalism we have implemented the scissors approximation, and have fully accounted for the modification of the velocity matrix elements that appear more explicitly in other calculation schemes. The response function ex-
pressions within the scissors approximation are straightforward to obtain, and are no less amenable to computation than without the scissors correction.

Our results for the imaginary part of the dielectric function, $\epsilon_2(\omega)$, show only reasonable agreement with experiment across a broad energy range, although we obtain excellent agreement with experiment for $\epsilon(0)$. This illustrates the possible limitations of the scissors approximation and indicates that good zero frequency results do not necessarily imply a good prediction of the dispersion of the dielectric function.

The SHG susceptibility has been presented and shows important differences from other theoretical calculations. The lack of experimental data as well as its contradictory nature prevents any conclusive comparison with experiment over a large energy range. A comparison of zero frequency results for the SHG susceptibility shows that our calculation is closest to Ghahramani et al. [26, 39], but there is large disparity between existing theoretical calculations and experimental data.

We have also calculated the LEO susceptibility below the band gap, and here our results are in excellent agreement with experiment throughout the experimentally studied energy region. Since the SHG and LEO (clamped lattice) susceptibilities are equal at zero frequency, the more recent work on the LEO susceptibility (with which our calculations are in excellent agreement) is thus in contradiction with the much older work on the SHG susceptibility (with which our calculations are in disagreement). This both gives us confidence in our calculated results, and – whatever the status of our calculation – encourages us to urge our experimental colleagues to reinvestigate the SHG susceptibility both at low frequency and over a wide frequency range.
Bibliography


Chapter 5

Calculation of Linear and Second-Order Optical Response in Wurtzite GaN and AlN

Abstract

We calculate the linear and nonlinear optical response of GaN and AlN in the wurztite structure. The dielectric function ε(ω), the second harmonic generation susceptibility χ(2)(ω, ω, ω), and the linear electro-optic susceptibility χ(2)(ω, ω, 0) are all evaluated over a broad frequency range. These results are based on a first principles electronic structure calculation using the FLAPW method within the LDA. Corrections to the underestimation of the band gaps are included at the level of the scissors approximation, which is carefully incorporated within our susceptibility formalism. All independent components for the response functions are calculated; the results for GaN and AlN are very different, and those of AlN are in strong disagreement with predictions of the bond charge model. Results for all response functions at zero frequency are underestimated with respect to the available experimental values. A comparison of the calculated response functions for non-zero frequencies is made with the limited experimental data. We confirm both analytically and numerically that the calculated second-order susceptibilities obey various symmetry constraints below the band gap.
5.1 Introduction

GaN and AlN are considered promising materials for various technological applications. The wide band gaps for these materials makes them candidates for use in the near-ultraviolet and ultraviolet regions. It has been suggested that GaN and AlN could be employed in nonlinear waveguides, short wavelength electroluminescent devices, and in high temperature diodes and transistors [1, 2, 3, 4]; AlN has also been suggested as a frequency doubler for GaAlAs laser diodes [5, 6]. The high thermal conductivity and low compressibility of these materials are attractive mechanical properties for a range of device applications [7].

There have been several experimental studies of the electronic and structural properties of these materials, as well as a number of investigations of the linear and nonlinear optical response [1, 4, 5, 6, 8, 9, 10, 11]. Theoretical efforts, however, have been primarily concerned with ground state properties [2, 12, 13, 14, 15, 16]. While there has been some work on linear optical response [3, 17, 18], the only full band structure calculation of nonlinear response has been restricted to zero frequency [19]. In this context, it would be useful to have a comprehensive analysis of the optical properties of GaN and AlN, as determined from first principles, over a broad range of frequencies.

We have presented results on the linear and nonlinear optical properties of zinc-blende GaAs and GaP in Chapter 4 [20]. The goal of this work is to apply the analytic expressions and computational method employed there to the wurtzite materials GaN and AlN. Our evaluation of optical response is based on a first principles calculation of the electronic structure using the full-potential linearized augmented plane-wave (FLAPW) method [21, 22]. We address the underestimation of the band gap by including self energy corrections at the level of the scissors approximation [20, 23]. This correction rigidly shifts the conduction states upward in energy and produces a corresponding change in the velocity matrix elements; the scissors correction is incorporated in a simple and straightforward way within our susceptibility formalism. The analytic expressions for the optical response functions are derived
from the formalism of Sipe and Ghahramani [24] and Aversa and Sipe [25]. These expressions have the advantage of being inherently free of any unphysical divergences at zero frequency. Thus, we believe this constitutes a good initial approach to the \textit{ab initio} determination of optical properties across a broad frequency range.

The calculational method is briefly presented in Sec. 5.2. In Sec. 5.3 we discuss and present the results of our first principles calculation. The band structures for these materials, as well as the linear and second-order optical response, are given in this section. We also identify the symmetries that the second-order response function satisfy, both analytically and numerically. Finally, in Sec. 5.4 we present our conclusions.

5.2 Calculational Method

In previous chapters [20] we detailed our method for calculating optical response in semiconductors. There we discussed the susceptibility notation and definitions, scissors approximation, electronic structure method, and the techniques employed in the Brillouin zone (BZ) integrations necessary for calculating the response functions. In the present work, our emphasis is on the results for the optical properties of GaN and AlN; we use the notation and definitions that have already been established. We seek the optical response above and below the band gap for these materials. For linear response, we evaluate the imaginary part of the dielectric function, \( \epsilon_2(\omega) \), and then employ the Kramers-Kronig relations to obtain the real part of this function. Similarly, we have chosen to evaluate the imaginary part of the second harmonic generation (SHG) susceptibility, \( \text{Im}\{\chi^{abc}(-2\omega;\omega,\omega)\} \), and again obtain the real part using the Kramers-Kronig relations. This allows for the determination of the absolute value of this susceptibility over an energy spectrum above and below the band gap. We restrict our evaluation of the linear electro-optic (LEO) susceptibility to the region strictly below the band gap. As this function is purely real in this region, we calculate it directly. The LEO susceptibility we consider here is in the clamped lattice approximation.
In the calculation of the susceptibilities it is convenient to reduce the integration over the BZ to one over the irreducible segment of the BZ (IBZ). This is accomplished by applying the operators $P_R$ of the group elements $R$ of the hexagonal symmetry group $6mm$, appropriate for GaN and AlN, to the expansion dyadics of the response tensor. In the case of linear response, the dielectric tensor is given by

$$\bar{\epsilon}(\omega) = \sum_{ab} \hat{a}_b \epsilon^{ab}(\omega),$$

and after applying the operators $P_R$ we obtain

$$\sum_R P_R(\hat{x}\hat{x}) = \sum_R P_R(\hat{y}\hat{y}) = 6 \bar{I},$$
$$\sum_R P_R(\hat{z}\hat{z}) = 12 \bar{I},$$
$$\sum_R P_R(\hat{a}_b) = 0, \ a \neq b,$$

where $\bar{I}$ is the identity tensor. Thus only the diagonal elements survive; contrary to what one obtains for materials with the cubic structure, all of these components are not equal. We can identify two independent components for the linear susceptibility, $\epsilon^{xx} = \epsilon^{yy}$, and $\epsilon^{zz}$.

For second-order optical response we proceed in the same way. We can write the LEO susceptibility as

$$\bar{\chi}(-\omega; \omega, 0) = \sum_{abc} \hat{a}_b \hat{c} \chi^{abc}(-\omega; \omega, 0),$$

and applying the operators $P_R$ we find

$$\sum_R P_R(\hat{z}\hat{z}\hat{z}) = 12(\hat{z}\hat{z}\hat{z}),$$
$$\sum_R P_R(\hat{x}\hat{x}\hat{z}) = 6(\hat{x}\hat{x}\hat{z} + \hat{y}\hat{y}\hat{z}),$$
$$\sum_R P_R(\hat{x}\hat{z}\hat{z}) = 6(\hat{x}\hat{z}\hat{z} + \hat{y}\hat{z}\hat{y}),$$
$$\sum_R P_R(\hat{z}\hat{x}\hat{z}) = 6(\hat{z}\hat{x}\hat{z} + \hat{z}\hat{y}\hat{y}).$$
\[
\sum_R P_R(\hat{a}\hat{b}\hat{c}) = 0, \text{ all other } a, b, c.
\] (5.4)

We can identify four independent components for this second-order tensor, which we take to be \(\chi^{zzz}, \chi^{xxz}, \chi^{zyz},\) and \(\chi^{zzz} \)\(\chi^{xxz}, \chi^{zyz} \)\(\chi^{xxz}, \chi^{zyz} \). The same result follows for the SHG susceptibility with one exception: The SHG tensor possesses intrinsic permutation symmetry, or a symmetry in the permutation of the last two indices. This reduces the number of independent components to three, which for this tensor we take to be \(\chi^{zzz}, \chi^{xxz},\) and \(\chi^{zyz} \).

To perform the integrations over the IBZ in the calculation of the optical response functions, we use the method discussed in Chapter 3. For the calculation of GaN and AlN we have partitioned the IBZ into 5184 tetrahedra, requiring a determination of the eigenvalues and velocity matrix elements at 1365 \(k\) points. In the case of GaN, we have further partitioned the region immediately near the \(\Gamma\) point into 3993 tetrahedra. This demands a further calculation of the eigenvalues and velocity matrix elements at an additional 1092 \(k\) points in this region. This is done for GaN only, as we present results on the anisotropy of the linear optical response function in the proximity of the band edge. To be confident in the accuracy of the calculation in this narrow energy region, the finer mesh of \(k\) points near the \(\Gamma\) point is required.

### 5.3 Results and Discussion

#### 5.3.1 Band Structures

In Fig. 5.1 we present the band structures for GaN and AlN. In both cases the band gaps have been adjusted, via the scissors shift, to coincide with the experimental values. The required scissors corrections were 1.564 and 2.066 eV for GaN and AlN respectively. These band structures are in good agreement with other theoretical results based on various electronic structure methods within the LDA [2, 3, 14, 17]. Agreement is somewhat less striking in a comparison of the eigenvalues at symmetry points from our band structure with existing quasiparticle calculations [2]; differences tend to be smaller than 0.4 eV for bands in proximity to the gap, and larger for bands
Figure 5.1: FLAPW band structures for GaN (upper plot) and AlN (lower plot). The fundamental band gaps have been adjusted to 3.5 and 6.3 eV respectively, within the scissors approximation.
further from the gap. We note that for these calculations, and all those discussed below, we have used the experimental values for the lattice constants [26]; for the crystal coordinate system, we have used the convention of Kobayashi and coworkers [13].

In the all-electron calculation employed in this work, the 3d states in Ga have been explicitly included as valence states. These states encroach on the lowest valence states, as can be seen in the band structure between -10 and -15 eV. The inclusion of these states as part of the fully relaxed valence states in the calculation affects the appearance of these states in the band structure itself, and also affects the shape and position of the upper valence and lower conduction states. On this basis, we believe it is important to include these states in a full band structure calculation of the optical properties of GaN.

5.3.2 Linear Optical Response

The results for linear optical response are presented in Fig. 5.2 and Fig. 5.3, again for GaN and AlN. In both figures we give our results for the two independent components of the imaginary part of the dielectric function, $\varepsilon_2^\prime\prime(\omega)$ and $\varepsilon_2^\prime(\omega)$. The corresponding components of the two materials exhibit a general similarity, although the relative magnitudes of the peaks in the spectra of GaN differ significantly from those of AlN.

The general similarity of the $\varepsilon_2(\omega)$ of these materials reflects the underlying similarity in the shapes of their band structures; the structural features in the $\varepsilon_2(\omega)$ spectra are associated with regions in the band structure for which pairs of bands are nearly parallel and the joint density of states is high. In the column IV semiconductors and the III-V cubic semiconductors, it is usually possible to identify each peak with a parallel joint density of states in a specific band structure region. In contrast, for GaN and AlN we have only managed to construct a preliminary assignment of the different peaks. Similar efforts have previously been presented by other researchers [3, 27].

We begin with GaN. The first peak in both $\varepsilon_2^\prime\prime$ (at 7.2 eV) and $\varepsilon_2^\prime$ (at 7.3 eV) is clearly due to the region between $\Gamma$ and $M$ in the Brillouin zone. The next two peaks
Figure 5.2: Plot of the imaginary part of the dielectric function for GaN. Both independent components are plotted: $\varepsilon_{2\omega}^{\text{xx}}(\omega)$ (upper plot), and $\varepsilon_{2\omega}^{\text{zz}}(\omega)$ (lower plot).
in $\epsilon_2^{\infty}$, one at 8.1 eV and one at 9.5 eV, arise respectively from contributions from A to L and L to M, and from contributions from $\Gamma$ to M and from A to H. In the place of these two peaks there is only one in $\epsilon_2^{\infty}$ (at 9.5 eV), and it seems to arise from different regions in the Brillouin zone: Its main contributions come from the regions from H to K and from K to $\Gamma$. The next peak in each component (at 10.6 eV for $\epsilon_2^{\infty}$ and at 10.8 eV for $\epsilon_2^{\infty}$) arises from contributions from $\Gamma$ to A, from A to H, and from A to L; although obviously not well localized in the Brillouin zone, at least here there appears to be common sources for the strength in $\epsilon_2^{\infty}$ and $\epsilon_2^{\infty}$. The sources of the remaining higher energy peaks are difficult to identify.

In AlN, where there are more peaks in each spectrum than in GaN, the identification of their origins is less clear. Here even the first peaks in the two components - at 8.7 eV in $\epsilon_2^{\infty}$ and at 8.5 eV in $\epsilon_2^{\infty}$ - appear to have somewhat different origins. While the lowest energy peak in $\epsilon_2^{\infty}$ arises from both the region from $\Gamma$ to M that leads to the first peaks in the GaN spectra, and the region L to M, the lowest energy peak in $\epsilon_2^{\infty}$ gains strength as well from the regions from $\Gamma$ to A and A to L. The next four peaks in $\epsilon_2^{\infty}$ (at 9.0 eV, coming from $\Gamma$ to K; at 9.6 eV, coming from $\Gamma$ to A, from A to H, and from L to A; at 9.8 eV coming from $\Gamma$ to M; at 11.2 eV, the source of which we cannot identify) have only one counterpart in $\epsilon_2^{\infty}$, at 9.3 eV, with origins in the regions from $\Gamma$ to K, from H to K, and from $\Gamma$ to M. The two components both display a peak at 12.2 eV; in $\epsilon_2^{\infty}$ this has its origins in the regions from $\Gamma$ to A and from A to H, while in $\epsilon_2^{\infty}$ there is as well a contribution from the region from L to H. We have not been able to identify well-defined regions in the Brillouin zone responsible for the other peaks.

Compared to the column IV elemental semiconductors and the III-V cubic semiconductors, the structure in $\epsilon_2(\omega)$ in GaN and AlN is considerably more complicated. This appears to arise in large part because of the lower symmetry of the wurtzite structure, leading to a more complicated joint density of states than in the cubic materials and a resulting difficulty in assigning peaks to different regions in the Brillouin zone. Support for this comes from our calculation of the II-VI compound semiconductors, many of which can occur in both cubic and wurtzite form. Comparing $\epsilon_2(\omega)$ for the
Figure 5.3: Plot of the imaginary part of the dielectric function for AlN. Both independent components are plotted: $\epsilon^x_2(\omega)$ (upper plot), and $\epsilon^z_2(\omega)$ (lower plot).
two structural forms of the same compound, we consistently find a more complicated spectrum for the wurtzite structure [28].

Despite the complicated spectrum of \( \epsilon_2(\omega) \) for GaN and AlN, our results are in good agreement with those of Christensen and Gorczyca [3] throughout the energy spectrum. Their calculation was based on the linear muffin-tin orbital (LMTO) method within the atomic-sphere approximation (ASA). Although, as they point out, their results are sensitive to the use of the ASA, we found a strong similarity between our results and theirs in the structure of the linear response function as well as in its value at zero frequency. These later results will be subsequently discussed. Solanki et al [18] have also calculated the dielectric function for AlN using the LMTO-ASA method. These results show marked differences from our own. Their work illustrates the sensitivity of the linear response function to the application of the LMTO-ASA approach as well as the number of \( k \) points used in the calculation. Our results differ substantially both qualitatively and quantitatively from those of Xu and Ching [17]: Our structure of \( \epsilon_2(\omega) \) for both GaN and AlN is in disagreement with theirs in respect to both peak positions and relative peak magnitudes. In addition, they found anomalously large values for \( \epsilon_2(\omega) \) for GaN for energies in excess of 20 eV. We found no such increase in the linear response in this region.

The only experimental data for \( \epsilon_2(\omega) \) of which we are aware is the spectroscopic ellipsometry measurements for GaN of Petalas et al [4]. This data does not exhibit the degree of structure that is present in our calculation, but does show general agreement in peak positions.

As there is interest in near band edge optical response, we present in Fig. 5.4 the GaN results for \( \epsilon_2(\omega) \) in the region near the function onset. Both independent components are plotted so as to elucidate the anisotropy of the function in this region. Our calculation indicates a difference in \( \epsilon_2(\omega) \) for the two components of about 0.02 for energies within 0.05 eV of the band gap, increasing to a maximum difference of 0.2 up to about 7 eV; the magnitude of the difference varies quite strongly with energy throughout this range. It should be noted that our calculation does not take into account excitonic effects which are known to affect the shape of the linear response.
Figure 5.4: Results for the calculated imaginary part of the dielectric function in the energy region near the function onset: $\varepsilon_2^{2\pi}(\omega)$ (solid line), and $\varepsilon_2^{2\pi}(\omega)$ (dotted line).

function near the band edge.

5.3.3 Second-Order Optical Response

We present our results for the imaginary part of the SHG susceptibility in Fig. 5.5 for GaN and AlN. We have plotted the three independent components of this function for each material. A striking feature of the results, particularly in comparison with those for the zinc-blende semiconductors [20], is how structured the frequency dependence is for all components of both materials. As in $\varepsilon_2(\omega)$, this appears to be due in large part to the lower wurtzite symmetry of these compounds. It is difficult to interpret the structure in $\chi^{(2)}$ in terms of the band structure of the material - especially so for the
wurtzite structure - because one must consider both one- and two-photon resonances in various regions of the band structure, and the interference between these processes. An initial attempt is made at this below; it is clear, however, that there is virtually no similarity between the results for these two compounds. Particularly considering the similarities in the linear optical response of these two materials, this highlights the greater sensitivity of the second-order response function to details of the band structure and momentum matrix elements of the material.

The imaginary part of the SHG susceptibility for GaN is initially very flat for all components in a region of approximately 1 eV near the half band gap. Beyond this region the structure of the response function, as well as the differences between individual components, becomes dramatically more pronounced. Our results show a strong similarity between the $\chi^{zzz}$ and $\chi^{z^2z}$ components over the entire energy spectrum. To a large degree, peak position and magnitude is approximately the same for both components. The $\chi^{zzz}$ component exhibits a similar general structure, but is opposite in sign and larger in magnitude than the other two components. We will address this relationship between the magnitudes of the components in a subsequent discussion of the predictions of the bond charge approach.

The spectrum of the imaginary part of the SHG susceptibility for AlN bears little resemblance to that of GaN. There appears to be no definite relationship between the independent components for this function. However, in general the $\chi^{zzz}$ component is usually the largest of the three components throughout the energy spectrum.

Turning now to the details of the peaks, for GaN we first consider the $\chi^{zzz}$ component. The broad peak between 3.7 and 4.6 eV seems to be attributable to a two photon resonance associated with the first peak in the $\epsilon_2^{zz}$ spectrum, but with an increasing contribution, at increasing energy, from a two photon resonance associated with the second peak in the $\epsilon_2^{zz}$ spectrum. The peaks at 6.2 and 6.7 eV are most likely due to two photon resonances associated with the fifth and sixth peaks in the $\epsilon_2^{zz}$ spectrum. For the $\chi^{z^2z}$ component, the first main peak at 4.7 eV is due to a two photon resonance associated with the second peak in the $\epsilon_2^{zz}$ response function. The peak at 5.7 eV is most likely from two photon resonances associated with the fifth
Figure 5.5: Imaginary part of the SHG susceptibility $\text{Im}\{\chi^{abc}(-2\omega;\omega,\omega)\}$ for GaN (upper plot) and AlN (lower plot). The three independent components are plotted: $\chi^{zzz}$ (solid line), $\chi^{xzx}$ (dotted line), and $\chi^{zzx}$ (dashed line). Energy bin size is 0.05 eV.
peak in $\epsilon_2^{xx}$ and the fourth peak in $\epsilon_2^{zz}$. Finally, the small peak at 7.2 eV is attributable to one photon resonances associated with the first peaks in both $\epsilon_2^{xx}$ and $\epsilon_2^{zz}$. The $\chi^{zzz}$ component is the most difficult to interpret in terms of the band structure. We can only loosely identify the peak at 4.8 eV with a two photon resonance associated with the third peak in $\epsilon_2^{xx}$, and the peak at 6.7 eV with a two photon resonance associated with the sixth peak in the $\epsilon_2^{zz}$.

For AlN, the $\chi^{zzz}$ component has two peaks for which somewhat definitive statements can be made. The first is the peak at 4.3 eV, which is due to two photon resonances associated with the first peaks in both $\epsilon_2^{xx}$ and $\epsilon_2^{zz}$. The peak at 5.9 eV is probably due to a two photon resonance associated with the fourth peak in $\epsilon_2^{zz}$. We note, however, that we were unable to specify a band structure region contributing to this peak. For the $\chi^{zzz}$ component the peak at 4.7 eV is due to a two photon resonance associated with the second peak in $\epsilon_2^{zz}$. The peak at 5.9 eV again seems associated with the fourth peak in $\epsilon_2^{zz}$, as for the $\chi^{zzz}$ component. Again, the contributions to the $\chi^{zzz}$ component are very hard to determine. We can only suggest that the peak at 4.3 eV is from two photon transitions associated with the first peak in both $\epsilon_2^{xx}$ and $\epsilon_2^{zz}$.

In Fig. 5.6 we plot the absolute value of the SHG susceptibility for both materials. Features of these results which merit noting are as follows: The $\chi^{zzz}$ component dominates in both spectra for virtually all energies. In GaN all three components are of comparable size for energies below the half band gap, whereas in AlN the $\chi^{zzz}$ and $\chi^{xxx}$ components are substantially smaller than the $\chi^{zzz}$ component. GaN exhibits significantly more structure for energies around the half band gap than does AlN.

The results presented in this chapter are, to our knowledge, the first for the SHG susceptibility beyond zero frequency. There are, however, experimental results with which we can compare. Considerable work has been done by Miragliotta et al. on the nonlinear optical response of GaN. Results for $\chi^{xxx}(\omega, \omega, \omega)$ have been presented for energies around the half band gap [11]. We have plotted in Fig. 5.7 these experimental results along with our calculated results for the absolute value of $\chi^{zzz}(\omega, \omega, \omega)$. The magnitude of the experimental data is approximately twice that
Figure 5.6: Absolute value of the SHG susceptibility above and below the band gap for GaN (upper plot) and AlN (lower plot). The independent components are: $\chi^{zzz}$ (solid line), $\chi^{zzx}$ (dotted line), and $\chi^{xxz}$ (dashed line).
of our calculated values. For AlN the experimental data is somewhat less complete. Over a range of energies, the only data of which we are aware is that of Lundquist et al. [5, 6] on radio-frequency sputter deposited AlN thin films. The experimental results exhibit a remarkably similar dispersion to our theoretical predictions, but are about 40% smaller in magnitude. It has been suggested by this experimental group that the value of the measured SHG susceptibility can be expected to increase with improved crystallinity of the AlN film; it is not clear how much larger this response function might become.

The value of the SHG susceptibility at zero frequency was the focus of the only other full band structure work on the nonlinear optical properties of GaN and AlN
Table 5.1: The dielectric function and the SHG susceptibility at zero frequency. Results of the present calculation (FLAPW) are compared with other theoretical calculations and experimental data. The experimental data for SHG is extrapolated to zero frequency following Chen et al. [19]. The values for $\chi^{(2)}(0)$ are in pm/V.

<table>
<thead>
<tr>
<th>Material</th>
<th>Method</th>
<th>$\varepsilon^{xx}(0)$</th>
<th>$\varepsilon^{zz}(0)$</th>
<th>$\chi^{zzzz}(0)$</th>
<th>$\chi^{zzzz}(0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaN</td>
<td>FLAPW</td>
<td>4.82</td>
<td>4.80</td>
<td>6.03</td>
<td>-4.27</td>
</tr>
<tr>
<td></td>
<td>Pseudopot. (no scissors; LDA gap)</td>
<td>5.54 $^a$</td>
<td>5.60 $^a$</td>
<td>10.8 $^b$</td>
<td>-6.4 $^b$</td>
</tr>
<tr>
<td></td>
<td>Pseudopot. (scissors) $^b$</td>
<td>4.75</td>
<td>4.85</td>
<td>7.0</td>
<td>-4.2</td>
</tr>
<tr>
<td></td>
<td>LMTO-ASA $^c$</td>
<td>4.71</td>
<td>4.62</td>
<td>10.8</td>
<td>-6.4</td>
</tr>
<tr>
<td></td>
<td>OLCAO $^d$</td>
<td>8.72</td>
<td>11.16</td>
<td>-10.7 $^f$</td>
<td>5.3 $^f$</td>
</tr>
<tr>
<td></td>
<td>Experiment</td>
<td>5.2 $^e$</td>
<td>-10.7 $^f$</td>
<td>5.3 $^f$</td>
<td></td>
</tr>
<tr>
<td>AlN</td>
<td>FLAPW</td>
<td>3.91</td>
<td>3.97</td>
<td>-3.77</td>
<td>-0.25</td>
</tr>
<tr>
<td></td>
<td>Pseudopot. (no scissors; LDA gap)</td>
<td>4.42 $^a$</td>
<td>4.70 $^a$</td>
<td>-8.4 $^b$</td>
<td>-0.2 $^b$</td>
</tr>
<tr>
<td></td>
<td>Pseudopot. (scissors) $^b$</td>
<td>3.78</td>
<td>3.94</td>
<td>-4.6</td>
<td>-0.2</td>
</tr>
<tr>
<td></td>
<td>LMTO-ASA $^c$</td>
<td>3.91</td>
<td>3.77</td>
<td>-4.6</td>
<td>-0.2</td>
</tr>
<tr>
<td></td>
<td>OLCAO $^d$</td>
<td>3.88</td>
<td>5.06</td>
<td>-4.6</td>
<td>-0.2</td>
</tr>
<tr>
<td></td>
<td>Experiment</td>
<td>4.68 $^g$</td>
<td>-12.6±7 $^h$</td>
<td>≤0.5 $^h$</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Reference [19].
$^b$Reference [29].
$^c$Reference [3].
$^d$Reference [17].
$^e$Reference [36].
$^f$Reference [1].
$^g$Reference [37].
$^h$Reference [9].

For this reason, and the fact that some low frequency experimental data has been presented in the literature, it is important that we place our work within this context. In Table 5.1 we present our results for the dielectric function, and the SHG susceptibility at zero frequency for GaN and AlN. We have also included the values from other theoretical calculations and experiments. We have only included two components for the SHG response function, as the $\chi^{zzz}$ and $\chi^{zzz}$ components are identical in the limit that the frequency approaches zero.

The results of the current calculation for the linear optical response at zero frequency are very close to those of the LMTO-ASA work of Christensen and Gorczyca [3] as well as the pseudopotential calculation of Chen et al [29] (where they have included the scissors approximation). There is some disagreement regarding the
anisotropy of the dielectric function at zero frequency, although this anisotropy is not calculated to be large for either material. Our results as well as the other two calculations predict values for $\varepsilon(0)$ smaller than the available experimental results. This result is striking in that our previous calculation for GaAs and GaP obtained values for the zero frequency linear response that were excellent in comparison with experiment. This may point to a limitation in the scissors approximation or, in the case of the LMTO method, to the use of "corrective" terms to adjust the LDA band gaps. It has been suggested that for wide band gap materials the scissors approach may be inadequate for correctly predicting zero frequency response [30]. The pseudopotential calculation of Chen et al. [19], which does not employ the scissors correction, suffers from the usual LDA band gap problem in that the band gap is underestimated. This naturally leads to higher values for the dielectric function at zero frequency. The only wildly anomalous results are those of Xu and Ching [17]. Their values are exceedingly large for GaN, and predict a strong anisotropy. For AlN, the $\varepsilon_z^\perp(0)$ value is close to other calculations, but again a large anisotropy is predicted leading to a large value for $\varepsilon_z^\perp(0)$

For the SHG susceptibility at zero frequency the theoretical and experimental data is more scarce. We find that our results are very close to those of Chen et al. [29] when they have incorporated the scissors correction. Again, their results without the scissors correction will be higher due to an underestimation of the band gap. The theoretical calculations predict values for the SHG susceptibility lower than those measured experimentally for both GaN and AlN. The experimental data for GaN is opposite in sign to the calculated values, but as Chen et al. [19] point out, the overall sign was not experimentally determined relative to the atomic coordinates.

One of the values of full band structure calculations such as ours is that they provide a check on simpler models for optical response that are often employed in the absence of more detailed calculations. For example, the usual bond charge model for $\chi^{(2)}$ [31] predicts that the $\chi^{zzz}$ component should be twice as large as the $\chi^{zzz}$ component, but of opposite sign. It is based on the assumptions of perfect tetrahedral bonding and an optical response consisting of independent electrons moving only along the
direction of their bonds. Our present work finds that the bond charge prediction for $\chi^{zzz}/\chi^{zzz}$ holds only approximately for GaN; indeed, the bond charge prediction holds to a good approximation not only at zero frequency, where the bond charge model is usually applied, but as well for the ratio of the components of $\text{Im}\{\chi^{(2)}(-2\omega,\omega,\omega)\}$ throughout a large energy range, as can be seen in Fig. 5.5. Yet the bond charge prediction fails completely for AlN. This result has already been found by Chen et al. [19], and is thus independently confirmed here.

The simplest possible reason for the difference between the two materials is that the lattice structure of AlN implies a bonding structure much further than that of GaN from the perfect tetrahedral bonding usually assumed. This, however, is clearly not the whole story: Relaxing only the assumption of perfect tetrahedral bonding and using the actual lattice parameters of GaN and AlN, the so-revised bond charge model would predict a $\chi^{zzz}/\chi^{zzz}$ ratio of -2.053 for GaN, and a ratio of -2.208 for AlN. Clearly the correction for GaN is small and that for AlN is substantial; yet the disagreement of our results for AlN with even this revised bond charge model is poor. Obviously other assumptions of the bond charge model are in error. Already in 1993 the results of Xu and Ching [17] for the ground state charge density contours indicated bonding of considerably more covalent character in GaN than in AlN; our results for density contours confirm this. While what is really of essence is the nature of the charge density response to applied fields, the nature of the ground states at least suggests that the model of directional bonds is in general more appropriate for GaN than AlN. This is clearly a qualitative issue worthy of more study.

We present our results for the clamped-lattice LEO coefficient in Fig. 5.8 for both materials. We have evaluated this function strictly below the band gap. To our knowledge these are the first ab initio calculations of the LEO susceptibility for these materials. We have plotted the four independent components of this response function, but note that for energies below the gap the $\chi^{zzz}$ and $\chi^{zzz}$ components are identical. This response function gradually increases in magnitude as the energy moves towards the band gap, where it experiences a resonance enhancement. For both materials the $\chi^{zzz}$ component is the largest. In GaN the other components are
Figure 5.8: Plot of the LEO susceptibility below the band gap for GaN (upper plot) and AlN (lower plot). The four independent components are plotted: $\chi^{\pi\pi}$ (solid line), $\chi^{\pi\pi}$ and $\chi^{\pi\pi}$ (dotted line), and $\chi^{\pi\pi}$ (dashed line). Note that two of the components are coincident in this energy region.
approximately half as large and opposite in sign. In the case of AlN the remaining components are very small in relation to the $\chi^{zzz}$ component. These results are not surprising given the data we have presented for the SHG susceptibility.

We are aware of only one experimental determination of the LEO coefficient for these materials. Long et al [32] have measured this coefficient for GaN at 1.96 eV and obtained $\chi^{zzz} = 30 \pm 5.5$ pm/V and $\chi^{zxx} = 9 \pm 1.7$ pm/V. They have not determined the absolute sign of this response function. Our calculated results at this energy for comparison are $\chi^{zzz} = 7$ pm/V and $\chi^{zxx} = 6$ pm/V. This comparison is however not strict; we calculated the LEO susceptibility in the clamped lattice approximation while the experimental data is for the total LEO effect.

5.3.4 Symmetries of $\chi^{(2)}$

The nonlinear optical susceptibilities are required to satisfy certain symmetry considerations and constraints [33]. It is the goal of this section to state some of these conditions and demonstrate how our calculation satisfies them.

We first consider intrinsic permutation symmetry in the case of the SHG response function. As has been previously presented [20], our susceptibility expression is written so as to explicitly satisfy this symmetry constraint. We have already discussed this in Sec. 5.2 regarding the reduction of the independent components for the SHG function from four to three, but for completeness mention it here.

For energies at which there is no absorption (below the band gap), full permutation symmetry holds. In terms of the LEO susceptibility this demands that

$$\chi^{zzz}(-\omega; \omega, 0) = \chi^{zxx}(-\omega; \omega, 0). \quad (5.5)$$

As indicated in Fig. 5.8 these two components have been numerically shown to be identical below the band gap. This symmetry condition can also be shown to hold true analytically within our susceptibility formalism.

We next consider Kleinman symmetry [34] which holds for very low frequencies.
This symmetry could be written schematically as

\[
\lim_{\omega \to 0} \chi^{xzx} = \lim_{\omega \to 0} \chi^{zzx} = \lim_{\omega \to 0} \chi^{zxz},
\]

(5.6)
either for the LEO or SHG coefficient, and for them interchangeably. Again Fig. 5.8 illustrates how this symmetry has been numerically satisfied by our calculation; three of the independent components for the LEO susceptibility are equal in the limit of zero frequency. This symmetry constraint can also be shown to hold in Fig. 5.6 for the SHG susceptibility. In this case two of the three independent components are numerically identical in the zero frequency limit; the expressions can also be shown to be equivalent analytically.

Finally, a condition that one would expect on physical grounds is that as the frequency approaches zero, the second-order response functions become equal. This can be written as

\[
\lim_{\omega \to 0} \chi^{abc}(-2\omega; \omega) = \lim_{\omega \to 0} \chi^{abc}(-\omega; \omega; 0).
\]

(5.7)
We have discussed and numerically demonstrated this condition in previous chapters for the case of GaAs and GaP. Here we can illustrate how this condition holds for GaN and AlN in the wurtzite structure by plotting the SHG and LEO susceptibilities on the same graph. In Fig. 5.9 we have plotted the $\chi^{zzz}$ component for each material. This figure clearly shows how both of these susceptibilities are equal for vanishing frequency; again, analytic equivalence can be demonstrated.

The above susceptibility constraints should hold in any calculational approach, and they provide a solid check of not only our formalism but of our numerical method.

5.3.5 Miller's Delta

We have previously calculated Miller's delta for GaAs and GaP [20] (see Chapter 4) and there found reasonable support for some of Miller's original predictions about this function. It is useful, therefore, to calculate Miller's delta for the semiconductors
Figure 5.9: Absolute value of the second-order optical response functions for GaN (upper plot) and AlN (lower plot) below the band gap. The $\chi^{zzz}$ component has been plotted for the SHG susceptibility (solid line), and the LEO susceptibility (dotted line).
Miller's delta for SHG can be written as

\[ \Delta_{ab}^{abc}(2\omega) = \frac{\chi^{abc}(-2\omega; \omega, \omega)}{\chi^{aa}(2\omega)\chi^{bb}(\omega)\chi^{cc}(\omega)}. \]  

We plot in Fig. 5.10 the three independent components of this function for GaN and AlN respectively. Our results indicate that the assumption of frequency independence of this function largely holds over a broad frequency range. We note, however, that there are obvious deviations from Miller's assumption of material independence for the value of a given \( \Delta_M \) function. The values for \( \Delta_M^{zz} \) differ by about 30% over much of the frequency range. The other independent components are very different for different materials. This later result we might expect given that the ratios of the magnitudes of independent components differ dramatically for GaN and AlN as has already been discussed. For comparison we note that the zero frequency values of \( \Delta_M^{yz} \) that we previously obtained for GaAs and GaP are 194 and 198 pm/V [20] respectively. Levine [35] has also calculated \( \Delta_M^{yz} \) for GaAs and GaP and has obtained 347 and 310 pm/V respectively.

### 5.4 Conclusions

We have presented results for the optical response of wurtzite GaN and AlN for a broad range of energies. This work has employed a first principles electronic structure calculation using the FLAPW method, and a completely non-divergent formalism for the optical susceptibilities. The scissors approximation has been incorporated in our calculation in an attempt to correct for the underestimated LDA band gaps.

Our results for linear optical response show striking similarities to those of Christensen and Gorczyca [3]. We are, however, in strong disagreement with other theoretical calculations. Our zero frequency results, in line with those of other workers, underestimate the experimental values. We predict a very small anisotropy at zero frequency as well as a relatively small anisotropy in \( \epsilon_2(\omega) \) for GaN for energies within
Figure 5.10: Plot of Miller's $\Delta_M(2\omega)$ below the half band gap for GaN (upper plot) and AlN (lower plot). The three independent components are plotted: $\Delta^{zzz}$ (solid line), $\Delta^{zzx}$ (dotted line), and $\Delta^{xxz}$ (dashed line).
3 eV of the band gap.

Our work on second-order optical response constitutes the first comprehensive calculation and analysis for GaN and AlN over a broad energy range. We have calculated all independent components for the SHG and LE0 susceptibilities and our results indicate strong dissimilarities between the materials considered in this work. The calculated spectrum of the SHG susceptibility for both materials is extremely rich in structure with the general feature of a dominant $\chi^{zzz}$ component. For energies below the gap, the other independent components are very small by comparison in AlN, but are of comparable magnitude in GaN. The predictions of the bond charge method seem to hold only approximately in GaN and fail completely for AlN. This we can attribute, only in part, to the greater degree of tetrahedral bonding in GaN.

Our calculated values for the second-order susceptibilities are smaller that the available experimental data by approximately a factor of two. We note, however, that very little experimental data exists for a range of energies above and below the band gap, nor for very low energies for which zero frequency results can be equated.

We have also detailed some of the symmetry properties that the second-order susceptibilities must satisfy. These constraints have been demonstrated to be adhered to numerically and are satisfied analytically within our susceptibility formalism.
Bibliography


Chapter 6

Comparison of Calculated Optical Response in Cubic and Hexagonal II-VI Semiconductors

Abstract

We calculate the dielectric function and the second-harmonic generation (SHG) susceptibility of the II-VI semiconductors in both the cubic and hexagonal phases. These results are based on a first principles electronic structure calculation using the full-potential linearized augmented plane-wave (FLAPW) method. A comparison is made between the two phases of the linear and second-order optical response for these semiconductors by examining this response in identical coordinate reference frames. Our calculations indicate that there is little difference in the linear optical response between the two phases of the given material below the band gap, but significant differences above the band gap. For the second-order optical response there are significant differences between the two phases of the II-VI semiconductors in both the real and imaginary parts of this susceptibility. The physics of this behaviour is discussed.
6.1 Introduction

The II-VI semiconductors are of interest not only from the perspective of fundamental solid state physics, but also for their potential use in a variety of technological applications [1, 2, 3, 4, 5, 6]. In addition, they are intriguing in that most can crystallize in either the cubic zinc-blende or hexagonal wurtzite structures; we refer to these as "cubic" and "hexagonal" throughout the rest of this chapter. The structures are quite similar; in fact, they are equivalent up to and including the second-nearest-neighbour positions, in the limit of perfect tetrahedral bonding. From the perspective of linear and nonlinear optics, one is led naturally to inquire how different are the optical properties of the two structures, in what ways they are different, and how the differences can be understood in terms of the underlying electronic structure.

As well, there is current experimental interest in the \( \chi^{(2)} \) component of the second-order susceptibilities of various nonlinear optical materials. This component, present in the hexagonal structure but not in the cubic structure, is particularly attractive for suggested applications in phase matching and quasi-phase matching [7]. It is thus interesting from a theoretical perspective to investigate to what degree the magnitude of this component can be attributed to effects of the lattice structure on the electronic structure. The II-VI semiconductors are an ideal set of materials for such an investigation. By examining the optical response in identical coordinate reference frames, rather than in the usual frames used for characterizing the different structures, we can initiate an examination of the effects of details of the crystal structure on optical response.

There have been a number of studies of the electronic structure of these materials using a variety of calculational methods [8, 9, 10, 11]. The linear optical response has also been examined both theoretically [12, 13, 14], and experimentally [15, 16, 17, 18, 19] for some of these materials. The second-order optical response, however, has not received as much attention. There is a limited amount of experimental data available [20, 21, 22, 23, 24, 25, 26, 27, 28], and most of this is relatively old. Theoretical calculations tend to focus on a subset of these materials and address the optical
response for only one modification [29, 30, 31].

We have presented results for GaAs and GaP in the cubic (zinc-blende) structure in Chapter 4 [32], and GaN and AlN in the hexagonal (wurtzite) structure in Chapter 5 [33]. The calculational method and analytic formalism employed there is applied to the II-VI semiconductors investigated in this work. Specifically, the goal of this work is twofold: First, to examine the similarities and differences between the optical response functions of both structures for a given material, and second, to elucidate the trends among the II-VI semiconductors in this comparison of the optical response.

This chapter focuses on detailing the comparison of optical response between the different phases common to the II-VI semiconductors. To illustrate this we present a subset of our results rather than including all calculated data for the six II-VI semiconductors in both phases.

This chapter is organized as follows. We very briefly discuss some of the pertinent details of our calculational method in Sec. 6.2. In Sec. 6.3 we present our comparison of the optical response between the two modifications for the II-VI semiconductors. Our conclusions are given in Sec. 6.4.

6.2 Method

As much of our method of calculating electronic structure and optical response has been detailed in earlier chapters [32, 33], we here only include a few brief relevant comments.

The basis for the optical response calculation is a well converged electronic charge density and electronic band structure. For this we use a first principles approach in the form of the full-potential linearized augmented plane-wave (FLAPW) method [34, 35], within the local density approximation (LDA); this method was discussed in detail in Chapter 3. This is an all-electron method, and is consequently well suited to a study of II-VI semiconductors where the d-electrons play an important role. We have also included spin-orbit effects, which are significant for the heavier atoms contained in these semiconductors.
For the Brillouin zone (BZ) integrations required for calculating the optical susceptibilities, we use a hybrid tetrahedron-random sampling method. This approach was discussed in Chapter 3 [32], so we only state here that all our calculations are based on an exact determination of the eigenvalues and velocity matrix elements at 1365 k points in the irreducible Brillouin zone (IBZ). From our experience with earlier calculations [32, 33], and on the basis of convergence tests on our response functions with a varying number of k points, we are confident that this number of k points ensures an accurate and well-converged result.

6.3 Results and Discussion

6.3.1 Structural Differences and Notation

In this work we calculate the linear susceptibility and the second-order response coefficient for second harmonic generation (SHG); in standard notation these are denoted by $\chi_{jk}^{*}(-\omega; \omega)$ and $\chi_{jkl}^{*}(-2\omega; \omega, \omega)$ respectively, where the superscripts denote Cartesian components. In place of $\chi_{jk}^{*}(-\omega; \omega)$ we give values for the linear dielectric tensor $\varepsilon_{jk}^{*}(\omega) = \delta_{jk} + 4\pi \chi_{jk}^{*}(-\omega; \omega)$, where $\delta_{jk}$ is the Kronecker delta; we denote the real and imaginary parts of $\varepsilon_{jk}^{*}(\omega)$ by $\varepsilon_{1jk}^{*}(\omega)$ and $\varepsilon_{2jk}^{*}(\omega)$, $\varepsilon_{jk}^{*}(\omega) = \varepsilon_{1jk}^{*}(\omega) + i\varepsilon_{2jk}^{*}(\omega)$. To denote $\chi_{jkl}^{*}(-2\omega; \omega, \omega)$ we use the shorthand expression $\chi_{jkl}^{*}(\omega)$. When we are speaking of the tensors and not their components in a particular reference frame, we use simply $\varepsilon(\omega)$ and $\chi(\omega)$ to denote the tensors having components $\varepsilon_{jk}^{*}(\omega)$ and $\chi_{jkl}^{*}(\omega)$ respectively.

The tensors $\varepsilon(\omega)$ and $\chi(\omega)$ for the cubic phase of a given compound are different than those for the hexagonal phase; we denote by $C\varepsilon(\omega)$ and $C\chi(\omega)$ the tensors for the cubic phase, and by $H\varepsilon(\omega)$ and $H\chi(\omega)$ those for the hexagonal phase. To specify the components of the tensors, one must adopt a reference frame. The conventional reference frames adopted for cubic and hexagonal structures are linked in natural ways to the respective point group symmetries of the structures; if we denote by $\hat{x}^{C}$, $\hat{y}^{C}$, and $\hat{z}^{C}$ the usual unit vectors for the cubic structure, and by $\hat{x}^{H}$, $\hat{y}^{H}$, and $\hat{z}^{H}$ the
Figure 6.1: The lattice structure for the II-VI semiconductors: cubic phase (left diagram), and hexagonal phase (right diagram). Crystals are oriented so that the \( \hat{z} \) direction of the hexagonal coordinate system (\( \hat{z}^H \)) points upwards.

corresponding usual unit vectors for the hexagonal structure, we have

\[
\hat{z}^H = \frac{1}{\sqrt{6}} \left( \hat{z}^C + \hat{z}^C - 2\hat{y}^C \right),
\]

\[
\hat{y}^H = \frac{1}{\sqrt{2}} \left( \hat{z}^C - \hat{z}^C \right),
\]

\[
\hat{z}^H = \frac{1}{\sqrt{3}} \left( \hat{z}^C + \hat{y}^C + \hat{z}^C \right),
\]

if the structures are overlayed so as to emphasize the similarity in their nearest neighbor bonding. This we do in Fig. 6.1, under the assumption of perfect tetrahedral bonding; the crystals are oriented so that \( \hat{z}^H \) points upward in the diagram. The two structures are identical up to and including the second-nearest-neighbor position. In the diagram this results in the top three planes of atoms being identical. It is in the placement of the next plane of atoms that the difference arises.

Values of \( C\epsilon(\omega) \) and \( C\chi(\omega) \) are of course usually quoted in the reference frame specified by \( (\hat{z}^C, \hat{y}^C, \hat{z}^C) \), and those of \( H\epsilon(\omega) \) and \( H\chi(\omega) \) in the reference frame specified by \( (\hat{z}^H, \hat{y}^H, \hat{z}^H) \). But in comparing the physics of the optical response it is more
natural to compare the tensors in the same reference frame. In an admittedly somewhat cumbersome, but unambiguous notation, we denote by $I\epsilon^{jk(J)}(\omega)$ and $I\chi^{jkl(J)}(\omega)$ the components of $I\epsilon(\omega)$ and $I\chi(\omega)$ in the reference frame indicated by $(\hat{z}^J, \hat{y}^J, \hat{z}^J)$, where both $I$ and $J$ can be either $C$ or $H$. Transformations between reference frames are easily effected in the usual way, e.g.

$$
\epsilon^{jk(H)} = \sum_{m,n} (\hat{j}^H \cdot \hat{m}^C)(\hat{k}^H \cdot \hat{n}^C) \epsilon^{mn(C)},
$$

(6.4)

$$
\chi^{jkl(H)} = \sum_{m,n,p} (\hat{j}^H \cdot \hat{m}^C)(\hat{k}^H \cdot \hat{n}^C)(\hat{l}^H \cdot \hat{p}^C) \chi^{mnp(C)},
$$

(6.5)

where the $\epsilon$ and $\chi$ can be either $H\epsilon$ and $H\chi$, or $C\epsilon$ and $C\chi$, and $(j, k, l, m, n, p)$ can all be either $x$, $y$, or $z$.

Since second rank tensors cannot distinguish between cubic and isotropic materials, at each frequency we have $C\epsilon^{zx}(C) = C\epsilon^{yz}(C) = C\epsilon^{zz}(C)$, and $C\epsilon^{jk}(C) = 0$ if $j \neq k$; the same conditions hold for the $C\epsilon^{jk(H)}$, and obviously $C\epsilon^{jk(H)} = C\epsilon^{jk(C)}$. The zincblende cubic structure allows only one independent component of a third rank tensor, and $C\chi^{jkl(C)} = 0$ unless $(j, k, l)$ is a permutation of $(x, y, z)$, in which case $C\chi^{jkl(C)} = C\chi^{xyz(C)}$. Transformed to the hexagonal frame, however, we find four kinds of nonvanishing elements that we can specify by giving $C\chi^{zxx(H)}$, $C\chi^{zzz(H)}$, $C\chi^{xxz(H)}$, and $C\chi^{yyz(H)}$. Thus the usual statement that “$\chi^{zz}$ vanishes for the cubic structure” holds only with respect to the usual cubic reference frame; the statement is nonetheless of practical significance because it is along $\hat{z}^C$ rather than $\hat{z}^H$ that the cubic phase typically grows epitaxially.

For the hexagonal structure we have $H\epsilon^{xx(H)} = H\epsilon^{yy(H)}$, with a generally different value of $H\epsilon^{zz(H)}$, and $H\epsilon^{jk(H)} = 0$ for $j \neq k$; transformed to the cubic reference frame this anisotropy appears in the presence of a nonzero $H\epsilon^{xy(C)} = H\epsilon^{yx(C)}$, but $H\epsilon^{xz(C)} = H\epsilon^{yx(C)} = H\epsilon^{zy(C)}$. In the hexagonal frame there are three independent elements of $H\chi^{jkl(H)}$ that we can specify by giving $H\chi^{zxx(H)}$, $H\chi^{zzz(H)}$, and $H\chi^{xxz(H)}$. Transforming to the cubic frame we find a host of nonvanishing components; there are four different kinds of elements, which we specify by giving $H\chi^{zzz(C)}$, $H\chi^{xxz(C)}$, $H\chi^{zyy(C)}$, and $H\chi^{xyz(C)}$. 

102
In most of our discussions below we find it convenient to use the cubic reference frame when comparing tensors for the cubic and hexagonal phases; hence, when there is no confusion, we omit the superscript \((C)\) when the cubic frame is used.

### 6.3.2 Band Structures and Linear Response

In Fig. 6.2 we plot the band structures for the cubic and hexagonal phases of ZnS. These band structures have been adjusted with the scissors shift to reproduce the experimentally observed band gaps [36]. Spin-orbit effects have been included, but remain quite small due to the light atoms in this material. We note the presence of the energetically flat \(d\)-electron bands in both materials in the region of -6 to -7 eV. The band structures for the other II-VI semiconductors are qualitatively similar to those of ZnS, but differences in band gaps and spin-orbit splitting are seen, as well as differences in higher energy conduction bands. In all our calculations we use the experimental lattice parameters which, in the hexagonal phase, does not correspond to perfect tetrahedral bonding. The departure from perfect tetrahedral bonding is, however, quite small. Such a departure has only small, secondary effects on the optical response; we refer the reader to our previous discussion on this in the context of GaN and AlN [33] in Chapter 5.

Results for the real and imaginary parts of the dielectric tensor, \(\varepsilon_1(\omega)\) and \(\varepsilon_2(\omega)\), are presented for ZnS (Fig. 6.3), ZnTe (Fig. 6.4), and CdSe (Fig. 6.5). These results are representative of those for the other II-VI semiconductors, and so serve to illustrate the various aspects of linear optical response for this group of materials. We have chosen to present the components of the dielectric tensor in the cubic reference frame. In this frame \(C_{\varepsilon^{xx}} = C_{\varepsilon^{yy}} = C_{\varepsilon^{zz}}\), \(H_{\varepsilon^{xx}} = H_{\varepsilon^{yy}} = H_{\varepsilon^{zz}}\), \(C_{\varepsilon^{jk}} = 0\) if \(j \neq k\), and the anisotropy of the hexagonal phase leads to a nonzero \(H_{\varepsilon^{xy}} = H_{\varepsilon^{yx}}\).

We first consider \(\varepsilon_1(\omega)\), which is plotted for all three materials at frequencies below the band gap. Our results indicate that there are no significant differences between \(H_{\varepsilon^{xx}}\) and \(C_{\varepsilon^{xx}}\). Any difference, initially very small at zero frequency, tends to increase only slightly for energies approaching the band gap; for CdSe, however, the difference remains very small over this entire energy region. Such qualitative
Figure 6.2: The band structures for ZnS in the cubic (upper plot) and hexagonal (lower plot) phases.
behaviour is typical of the remaining II-VI semiconductors that we have not shown. The $H \varepsilon^{xy}$ component is found to be extremely small. The magnitude of $H \varepsilon^{xy}$ at zero frequency has been calculated as 0.0005, 0.007, and 0.03 for ZnS, ZnSe, and CdSe respectively. For the remaining II-VI semiconductors, $H \varepsilon^{xy}(0)$ is never greater than 0.03. This indicates that in going from the cubic to the hexagonal structure essentially no significant anisotropy is introduced in the linear response for energies below the band gap.

Our results for $\varepsilon_2(\omega)$ indicate a more marked difference between $H \varepsilon^{xx}$ (hexagonal structure) and $C \varepsilon^{xx}$ (cubic structure) for energies above the band gap. While the two calculated results exhibit a general similarity in magnitude, they are very different functions of frequency. The results for the cubic structure show one dominant peak in the $\varepsilon_2(\omega)$ spectrum, whereas there are two main peaks of roughly equal magnitude for the hexagonal structure. For all of the II-VI semiconductors we find that the peaks in $H \varepsilon_2^{xx}(\omega)$ straddle the main peak in $C \varepsilon_2^{xx}(\omega)$. We find that the hexagonal and cubic phases of the Te-containing compounds show the best agreement in peak positions and relative magnitudes of the $\varepsilon_2(\omega)$ spectrum. Finally, we note that the $H \varepsilon_1^{xy}(\omega)$ component remains relatively small for all energies, as our result for $H \varepsilon_2^{xy}(\omega)$ would suggest: specifically, we find that it is never more than 2. The $H \varepsilon_2^{xy}(\omega)$ component is qualitatively the same for all II-VI semiconductors, showing some peak structure for those energies around the main peaks in the $H \varepsilon_2^{xx}(\omega)$ component.

The contrast of our results for $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ between the two phases is not unexpected, considering how these functions depend on the band structure of a material. For instance, the peaks in $\varepsilon_2(\omega)$ arise largely from regions in the band structure where pairs of bands are parallel. These regions tend to be fairly localized in k-space, and so depend on broader regions of real space in the crystal. This implies that $\varepsilon_2(\omega)$ is in effect "seeing" the structural differences between the two phases. The converse is true for $\varepsilon_1(\omega)$ below the band gap, where there is little peak structure present. The response for these energies is dependent on the entire Brillouin zone, and so on more localized regions of real space. This would imply that $\varepsilon_1(\omega)$ below the band gap depends to a lesser degree on the long-range structure of the material. The reason that
Figure 6.3: The dielectric function for ZnS: $\varepsilon_1(\omega)$ (upper plot), and $\varepsilon_2(\omega)$ (lower plot). Results are plotted in the cubic coordinate reference frame: $\varepsilon^{xx}$ (solid line) for the cubic phase, and $\varepsilon^{zz}$ (dotted line) and $\varepsilon^{xy}$ (dashed line) for the hexagonal phase.
Figure 6.4: The dielectric function for ZnTe: $\varepsilon_1(\omega)$ (upper plot), and $\varepsilon_2(\omega)$ (lower plot). Results are plotted in the cubic coordinate reference frame: $\varepsilon^{xx}$ (solid line) for the cubic phase, and $\varepsilon^{xx}$ (dotted line) and $\varepsilon^{xy}$ (dashed line) for the hexagonal phase.
Figure 6.5: The dielectric function for CdSe: $\epsilon_1(\omega)$ (upper plot), and $\epsilon_2(\omega)$ (lower plot). Results are plotted in the cubic coordinate reference frame: $\epsilon^{xx}$ (solid line) for the cubic phase, and $\epsilon^{xx}$ (dotted line) and $\epsilon^{xy}$ (dashed line) for the hexagonal phase.
such a strong similarity exists in $\epsilon_1(\omega)$ between the two crystal phases is primarily due to the similarity of the crystal structures themselves.

It is possible to gain some physical insight into the peak structure in the linear response function by looking at the sources for the peaks in terms of the band structures. This is reasonably straightforward for cubic materials [32] but is much more difficult for the lower symmetry hexagonal materials [33]. We will endeavor, however, to identify general regions of k-space as sources for the peaks in $\epsilon_2(\omega)$. We restrict our discussion to ZnS, and use this as an example for a similar analysis for the other II-VI materials.

We first look at the cubic phase of ZnS. The main peaks, as seen in Fig. 6.3, are at 6.2, 7.8, and 9.2 eV. For the first two peaks, the strongly contributing regions of k-space are much the same as for the III-V materials [32]. We find that the 6.2 eV peak is primarily due to transitions in the $\Gamma$ to $L$ region; the 7.8 eV peak is associated with transitions in the $\Gamma$ to $X$ and $X$ to $U$ regions. The last peak at 9.2 eV does not seem to be due to the $\Gamma$ to $L$ region as in the III-V materials, but rather from the $\Gamma$ to $X$ and $X$ to $U$ regions.

For the hexagonal structure, there are really just two main peaks, which for ZnS are at 7.4 and 8.9 eV, as can be seen in Fig. 6.3. The sources for the peaks for this structure are much less localized in k-space. We find that the $L$ to $A$ and $A$ to $H$ regions are associated with the first peak at 7.4 eV. The peak at 8.9 eV seems to come from transitions in many regions; the $\Gamma$ to $M$, $\Gamma$ to $A$, and $H$ to $K$ regions seem to be integral to the peak structure.

6.3.3 Second-Order Response

We now turn to a discussion of our calculated results for the second-order response coefficient for second harmonic generation. We have chosen to focus on ZnSe and CdS for our analysis; our results for these semiconductors are typical of those we have found for the entire set of II-VI materials. The comparison of second-order response will initially be made in the cubic coordinate frame, where the cubic phase has only one independent component, $C_\chi^{\text{xyz}}$, and the hexagonal phase has four independent
components, $^H\chi^{xyz}$, $^H\chi^{zzz}$, $^H\chi^{xxx}$, and $^H\chi^{yy}$. We will address our results for the SHG susceptibility in the hexagonal coordinate frame towards the end of this section.

In Fig. 6.6 and Fig. 6.7 we have plotted the real part of the SHG response tensors, $\text{Re}\{^H\chi(\omega)\}$ and $\text{Re}\{^C\chi(\omega)\}$, for ZnSe and CdS respectively. There are two aspects of these results that are in striking contrast to our results for the linear response. First, the $^C\chi^{xyz}$ component for the cubic phase is substantially different than the $^H\chi^{xyz}$ component for the hexagonal phase, even below half-band-gap; we found, in contrast, that the linear response of the two phases is essentially the same below the band gap. Second, the other components of the SHG susceptibility that are not present in
Figure 6.7: The real part of the SHG susceptibility in the cubic coordinate reference frame for both phases of CdS. Plotted are $\chi^{xyz}$ for the cubic phase (solid line), and for the hexagonal phase, $\chi^{xyz}$ (dotted line), $\chi^{zzz}$ (short-dashed line), $\chi^{zxy}$ (long-dashed line), and $\chi^{zyy}$ (dot-dashed line).

The cubic structure, namely $H_1 \chi^{zzz}$, $H_1 \chi^{zzz}$, and $H_1 \chi^{zyy}$, are comparable in magnitude to the $H_1 \chi^{xyz}$ component; we found, in contrast, that the linear response coefficient not present in the cubic phase was very small in the hexagonal phase. These results clearly indicate that the second-order optical response exhibits a stronger sensitivity to the structure of the material than the linear optical response. We return to this point below.

Our results for $\text{Im}\{H_1 \chi(\omega)\}$ and $\text{Im}\{C_1 \chi(\omega)\}$, shown in Fig. 6.8 and Fig. 6.9 for ZnSe and CdS respectively, are what one would expect from the below-half-band-gap behaviour of the real parts. For the $\chi^{xyz}$ component the cubic and hexagonal phases
Figure 6.8: The imaginary part of the SHG susceptibility in the cubic coordinate frame for both phases of ZnSe. Upper plot: $\chi^{xyz}$ for the cubic phase (solid line), and $\chi^{xzy}$ for the hexagonal phase (dotted line). Lower plot: $\chi^{xxz}$ (solid line), $\chi^{zxy}$ (dotted line), and $\chi^{yyz}$ (dashed line) for the hexagonal phase.
Figure 6.9: The imaginary part of the SHG susceptibility in the cubic coordinate frame for both phases of CdS. Upper plot: $\chi^{xyz}$ for the cubic phase (solid line), and $\chi^{yz}$ for the hexagonal phase (dotted line). Lower plot: $\chi^{xx}$ (solid line), $\chi^{xy}$ (dotted line), and $\chi^{yy}$ (dashed line) for the hexagonal phase.
are somewhat similar in magnitude and line shape, but show differences of varying
degree throughout the spectrum. For the hexagonal phase, the imaginary parts of
\( H \chi^{zzz}, H \chi^{zzx}, \) and \( H \chi^{zyz} \) are comparable in magnitude to that of \( H \chi^{xyz} \). As in the
case of linear response, the imaginary part of the response function is more sensitive
to structural differences than is the real part below the half-band-gap.

Of technological interest is the \((zzz)\) component of the nonlinear susceptibility of
the hexagonal phase with respect to the usual hexagonal coordinate system, \( H \chi^{zzz}(H) \). This we plot in Fig. 6.10, along with the corresponding component of the cubic phase
when viewed in this coordinate system, \( C \chi^{zzz}(H) \). Our calculations show that there is
very little difference between the two phases in the below-half-band-gap behaviour of
that component. This is generally true of the other II-VI semiconductors, although
some small differences do arise. They remain fairly small even for energies approaching
the half-band-gap. There are stronger differences between the crystal phases for the
imaginary part of this component, as one would expect based on results that have
already been discussed. But the magnitude and line shape of even the imaginary part
of \( H \chi^{zzz}(H) \) is very similar to that of \( C \chi^{zzz}(H) \).

These results indicate that the \( H \chi^{zzz}(H) \) component is not particularly sensitive to
the hexagonal structure; it is only because the cubic structure does not grow epitaxi-
ally along the \( z^H \) direction that the correspondingly large \( C \chi^{zzz}(H) \) is not of technolog-
ical interest. Extrapolating these results to other materials, in an admittedly cavalier
way, one would not expect a particularly enhanced SHG if a normally cubic material
could be grown in a hexagonal phase, except insofar as the resulting epitaxial growth
would make \( H \chi^{zzz}(H) \) more accessible than the corresponding \( C \chi^{zzz}(H) \). Indeed, one
would guess that a good estimate of the \( H \chi^{zzz}(H) \) could be obtained by calculating
\( C \chi \) for the cubic structure and simply performing the appropriate coordinate trans-
formation.

At least below half-band-gap, this is what one would expect by analogy with the
linear response. In the "nonresonant" region it is more-or-less the local neighbour-
hood of the atoms that determines the response, and this is remarkably similar for
the hexagonal and cubic phases. The surprise is the very large, unexpected difference
Figure 6.10: Plot of $\chi^{zzz}$ in the hexagonal coordinate reference frame: the real part of this susceptibility (upper plot), and the imaginary part (lower plot). For both plots, the solid line is $\chi^{zzz}$ for the hexagonal phase, and the dotted line is $\chi^{zzz}$ for the cubic phase.
between $H \chi_{xyz}^{zzz}(C)$ and $C \chi_{xyz}^{xyz}(C)$ indicated above. This clearly requires further investigation, but a possible reason might lie in the form of the actual component under consideration. While the $\chi^{zzz}$ component is arguably simple physically and relatively straightforward, in that it describes an induced polarization in the $z$ direction due to an electric field in the same direction, the situation is different for $\chi^{xyz}$. Here an induced polarization in the $x$ direction is generated due to electric fields in the $y$ and $z$ directions. It may be that the more complicated $\chi^{xyz}$ component involving all three coordinate directions, and other such components, are more sensitive to the structural details than is the "single coordinate" $\chi^{zzz}$ component. Admittedly $H \chi^{zzz}(C)$ does not vanish and $C \chi^{zzz}(C)$ does, but the latter is strictly forbidden by symmetry.

For linear response, we detailed the origin of much of the peak structure in $\varepsilon_2(\omega)$ in terms of the electronic band structure. Although in principle it is possible to do the same for $\text{Im}\{\chi(\omega)\}$, in practice this task is much more difficult than for linear response; this is especially true in the case of the hexagonal structure where the band structure is much more complicated than that for the cubic phase. Furthermore, we find that it is difficult to generalize the results for one or two of the II-VI semiconductors to the remaining II-VI semiconductors. As a result, we defer such an analysis to a later, more detailed communication; for general analyses of this sort, we refer the reader to previous discussions [32, 33].

6.4 Conclusions

We have presented a comparison of the calculated linear optical response, and second-order optical response for second harmonic generation, for the cubic and hexagonal phases of the II-VI semiconductors. Previous studies of these materials have been restricted to only one of the phases.

We have found that structure plays a limited role for linear response at energies below the band gap. Results for a given material are largely independent of crystal phase in this energy region. Above band gap response, characterized by the imaginary part of the dielectric tensor, shows larger differences between the phases. This is as
one might expect from a simple physical picture of the optical response: Below-band-gap response involves essentially all regions in the Brillouin zone, and thus the local neighbourhood of the atoms in real space. Since the cubic and hexagonal phases are remarkably similar here, no large differences are expected. Above band-gap, the structure in the imaginary part of the dielectric tensor is sensitive to localized regions in k-space and thus to more global aspects of the structure; here greater differences are expected between the phases.

The result for the second-harmonic response coefficient $\chi^{zzz}$ in the hexagonal phase is in accord with this. Below half-band-gap the value differs little from the value of this coefficient in the cubic phase when specified in the same coordinate system. In all the II-VI semiconductors the value of the $\chi^{zzz}$ coefficient in the hexagonal structure is apparently not dependent on those details of that structure that make it specifically hexagonal.

Results for the imaginary part of the second-harmonic response coefficients show considerable difference above half-band-gap, in accord with the physical picture mentioned above. The surprise is in the below-half-band gap values for the $\chi^{yz}$ component of the two phases, as referenced to the cubic coordinate system. When the results for the hexagonal phase are specified in the same coordinate system, the results for the two phases are seen to be very different; as well, components that are not allowed by symmetry in the cubic phase are, in the hexagonal phase, comparable to $\chi^{yz}$. Even below the half-band-gap, where simple physical argument would indicate a sensitivity only to the local environment of the constituent atoms, the two very similar phases exhibit a significantly different response. We conjecture that this is due to the complicated nature of response characterized by the $\chi^{yz}$ component, involving electric fields in two orthogonal directions and a polarization in the third orthogonal direction.

Although we have presented results for only a few of the II-VI semiconductors in this chapter, they are indicative of the semiconductors we did not present. All of the II-VI materials exhibit remarkably similar qualitative behaviour, which is clearly worthy of further study.
Bibliography


[36] The values for the scissors correction used for the II-VI semiconductors are as follows: Cubic - 2.087, 2.083, 1.752, 1.849, 1.812, and 1.518 eV, and Hexagonal - 2.122, 2.110, 1.923, 1.808, 1.698, and 1.585 eV, for ZnS, ZnSe, ZnTe, CdS, CdSe, and CdTe respectively.
Chapter 7

Local Field Effects in Linear Optical Response: A New Approach

Abstract

We present a new formalism for the calculation of the linear optical response of semiconductors, including local field corrections. This approach differs from the usual Adler and Wiser approach in that it is predicated on relating the microscopic polarization to a local electric field that is written as the sum of the macroscopic Maxwell field, a microscopic Coulomb field, and a microscopic field associated with the exchange and correlation potential. This new method of calculating the linear optical response has complete freedom from any symmetry limitations, is numerically efficient, and avoids the need to take the so-called $q \to 0$ limits either analytically or numerically. We have applied this method to the calculation of linear optical response over a broad frequency range for silicon, using a pseudopotential electronic band structure.


7.1 Introduction

Despite several decades of research on the optical properties of semiconductors, theoretical calculations are still far from being able to quantitatively describe the dielectric function in even the simplest case of a bulk semiconductor. Issues involving the use of density functional theory (DFT) at the level of the local density approximation (LDA) for electronic structure calculations, the treatment of so-called quasi-particle corrections, excitonic effects, and finally the inclusion of local field corrections, all contribute to the complexities associated with calculating optical susceptibilities. This chapter focuses on the treatment of local field effects, with the aim of incorporating them within a modern susceptibility formalism [1, 2] in a straightforward way. We begin by briefly outlining the current status of optical response calculations, and the underlying ideas behind our approach, in order to set this present work in context.

Theoretical calculations based on DFT are now considered a standard means of studying the ground state in a variety of materials such as atoms, molecules, and crystals [3, 4]. The most common approach in crystals is to use the LDA in the solution of the Kohn-Sham equations [5]. This has yielded very accurate ground state properties for a variety of crystals. The use of DFT within the LDA has also been extended and employed in many optical susceptibility calculations [6, 7, 8, 9, 10]. The validity of this extension is clearly in question and has been the focus of various investigations [11, 12, 13].

Consider a static perturbation applied to a system due to some external electric field. At least in the case of a finite system, a rigorous DFT perturbation theory exists. It is possible to relate the perturbed charge density, \( \rho(x) \), to the applied external potential, \( \phi_{ext}(x) \), by

\[
\rho(x) = \int \gamma(x, x') \phi_{ext}(x') dx',
\]

where \( \gamma(x, x') \) is the exact density-density response function [14]. The usual approximation that is made in the treatment of such a system, in conjunction with the LDA, is the assumption that the perturbed density can be reproduced by a system
of noninteracting particles in an appropriate single-particle potential, \( \phi_{\text{SCF}}(x) \). That is [14],

\[
\rho(x) = \int \gamma_o(x, x') \phi_{\text{SCF}}(x') dx',
\]

where \( \gamma_o(x, x') \) is the density-density response function of a noninteracting Kohn-Sham ground state corresponding to \( \phi_{\text{SCF}}(x) \); \( \gamma_o(x, x') \) is also referred to as the independent-particle polarizability, and is calculated from usual first-order perturbation theory using the Kohn-Sham wavefunctions. The so-called "self-consistent field" potential appearing in this equation is related to the external potential by

\[
\phi_{\text{SCF}}(x) = \phi_{\text{EXT}}(x) + \phi_h(x) + \phi_{xc}(x),
\]

where

\[
\phi_h(x) = \int \frac{\rho(x')}{|x - x'|} dx'.
\]

and

\[
\phi_{xc}(x) = \left( \frac{dV_{xc}(s)}{ds} \right)_{s=\rho_o(x)} \rho(x)
\]

are the induced Coulomb and exchange-correlation potentials. Here \( V_{xc}(\rho) \) is a local exchange-correlation potential, and \( \rho_o(x) \) is the unperturbed charge density. The approach of Eq. 7.2, which addresses the response of a system to a static external potential, works well for atoms and molecules [14]. It can be reformulated by defining a microscopic polarization, \( \mathbf{p}(x) \), which satisfies

\[
\rho(x) = -\nabla \cdot \mathbf{p}(x).
\]

This definition of \( \mathbf{p}(x) \) can be made unique by, for example, requiring that \( \nabla \times \mathbf{p}(x) = 0 \). Then, at least in principle, a tensor \( \mathbf{X}(x, x') \) could be determined from \( \gamma_o(x, x') \) such that

\[
\mathbf{p}(x) = \int \mathbf{X}(x, x') \cdot \mathbf{e}(x') dx',
\]

with

\[
\mathbf{e}(x) = \mathbf{e}_{\text{EXT}}(x) + \mathbf{e}_h(x) + \mathbf{e}_{xc}(x),
\]
where $e_{ext}(x)$ is the external perturbing field, and where we set

$$e_{\mu}(x) = -\nabla \phi_{\mu}(x)$$  \hspace{1cm} (7.9)

and

$$e_{xc}(x) = -\nabla \phi_{xc}(x).$$  \hspace{1cm} (7.10)

The $\tilde{\chi}(x, x')$ in Eq. 7.7 is a kind of polarization-polarization response function of the noninteracting Kohn-Sham ground state. It is analogous to $\gamma_{\alpha}(x, x')$, as it too can be evaluated in terms of the Kohn-Sham wavefunctions.

This chapter focuses on the response of a periodic crystal perturbed by a field of non-zero frequency. This immediately raises two new issues. A first concern is the applicability of DFT in the context of an infinite system under even a static perturbing field; a second is how is the time dependence of the field addressed within DFT-LDA. Both of these issues involve the validity of the Hohenberg-Kohn theorem under the circumstances of interest.

Gonze et al. [15] have proposed that an infinite periodic solid in the presence of an electric field is a case where the Hohenberg-Kohn theorem does not apply. This has led to the proposition of an alternative to DFT, referred to as density-polarization functional theory (DPFT). Martin and Ortiz [16] have also argued along these lines. We restrict ourselves to DFT in the formalism presented here, but note that our approach could in principle be extended to DPFT once a functional is specified.

The extension to the time domain of the steady-state DFT of Hohenberg, Kohn, and Sham has been investigated and has led to the development of a time-dependent density functional theory (TDDFT) [17, 18]. In a somewhat more restrictive limit, a TDDFT was first developed in the context of the LDA, referred to as time-dependent LDA (TDLDA) [19]. Calculations using the TDLDA approach have been done both prior to [20, 21, 22] and after [23] its more rigorous formulation. The present work is at the level of the TDLDA; consequences and limitations associated with this level of approximation are discussed below in Sec. 7.2

Under the restrictions of DFT and TDLDA, we consider an infinite periodic crystal
under a time-dependent perturbing electric field. We take the applied field to have the form

$$e_{\text{EXT}}(x,t) = e_{\text{EXT}}(x,\omega)e^{-i\omega t} + \text{c.c.},$$  

(7.11)

where $e_{\text{EXT}}(x,\omega)$ is a solution of the vacuum Maxwell equations. A fundamental way to describe the response of the system is in terms of a microscopic current density

$$j(x,t) = j(x,\omega)e^{-i\omega t} + \text{c.c.}$$  

(7.12)

or equivalently in terms of the microscopic polarization, $p(x,t)$, where for harmonic fields there is no ambiguity in the definition of $p(x,t)$ as

$$j(x,t) = \frac{dp(x,t)}{dt}. $$  

(7.13)

Using TDLDA, it is possible to extend Eq. 7.7 as

$$p(x,\omega) = \frac{1}{\Omega} \int \tilde{\chi}(x,x';\omega) \cdot e(x',\omega) dx'. $$  

(7.14)

where

$$e(x,\omega) = e_{\text{EXT}}(x,\omega) + e_{\mu}(x,\omega) + e_{xc}(x,\omega). $$  

(7.15)

and

$$e_{xc}(x,\omega) = -\nabla \phi_{xc}(x,\omega). $$  

(7.16)

with

$$\phi_{xc}(x,\omega) = \left(\frac{dV_{xc}(s)}{ds}\right)_{s=R_0(x)} \rho(x,\omega). $$  

(7.17)

The $\tilde{\chi}(x,x';\omega)$ of Eq. 7.14 is, within the TDLDA, the extension to a time-dependent perturbing field of the $\chi(x,x')$ of Eq. 7.7. It is again a polarization-polarization susceptibility of the noninteracting Kohn-Sham ground state. For later notational convenience, however, we include a normalization volume of the crystal, $\Omega$, in the definition. The field $e_{\mu}(x,\omega)$ is the particular solution of the Maxwell equations, subject to the outward-going radiation condition, with current density $j(x,\omega) = -$.
\[ \nabla \cdot \mathbf{j}(x, \omega), \] as a source. The subtleties of identifying this particular solution in the limit of an infinite crystal have been discussed in the literature in the context of the Ewald-Oseen extinction theorem [24], and are not specific to any approach of calculating the material response; they need not concern us here.

We define the macroscopic electric field, \( \mathbf{E}(x, \omega) \), as the macroscopic average of the microscopic Maxwell field \( \mathbf{e}_M(x, \omega) \) over a distance \( \Delta \) much larger than the lattice spacing \( a \), but much smaller than the wavelength of light, \( \lambda \) [25]. Defining

\[ \mathbf{e}_\varepsilon(x, \omega) \equiv \mathbf{e}_M(x, \omega) - \mathbf{E}(x, \omega) \] (7.18)

and a local correction field as

\[ \mathbf{e}_{LFC}(x, \omega) \equiv \mathbf{e}_\varepsilon(x, \omega) + \mathbf{e}_{x_c}(x, \omega), \] (7.19)

Eq. 7.15 becomes

\[ \mathbf{e}(x, \omega) = \mathbf{E}(x, \omega) + \mathbf{e}_{LFC}(x, \omega). \] (7.20)

Eq. 7.14 can then be rewritten as

\[ \mathbf{p}(x, \omega) = \frac{1}{\Omega} \int \mathbf{X}(x, x'; \omega) \cdot \{ \mathbf{E}(x', \omega) + \mathbf{e}_{LFC}(x', \omega) \} \, dx'. \] (7.21)

The simplest theory of optical response follows from neglecting the spatially dependent microscopic field, \( \mathbf{e}_{LFC}(x, \omega) \). By construction, the macroscopic field \( \mathbf{E}(x, \omega) \) has its main variation over distances on the order of the wavelength of light. Hence, in a region of the crystal extending over many lattice constants but small compared to the wavelength of light we can take \( \mathbf{E}(x, \omega) \) as uniform, \( \mathbf{E}(\omega) \). In such a macroscopically infinitesimal region it follows from Eq. 7.21 and the neglect of \( \mathbf{e}_{LFC}(x, \omega) \) that

\[ \mathbf{P}(\omega) = \mathbf{X}_{NLF}(\omega) \cdot \mathbf{E}(\omega), \] (7.22)
where $P(\omega)$ is the spatial average of $p(x, \omega)$, and

$$\bar{\chi}_{NLF}(\omega) = \frac{1}{\Omega^2} \int_\Omega \bar{\chi}(x, x'; \omega) dx dx', \quad (7.23)$$

where the integrals are over the crystal normalization volume $\Omega$; the spatial average of the current density $j(x, \omega)$ is then given by $J(\omega) = -i\omega P(\omega)$. This approximate approach of treating the perturbing field in the crystal simply as the macroscopic Maxwell field has been employed in many investigations of optical response [7, 8], including the susceptibility formalism of Sipe and Ghahramani (referred to herein as SG) [1], and various calculations in which this formalism has been implemented [10, 26, 27]. Clearly a more accurate calculation of the optical susceptibility in a crystal requires a careful treatment of the spatially varying field, $e_{LFE}(x, \omega)$. We refer to the effects due to the inclusion of $e_{LFE}(x, \omega)$ as local field effects (LFE).

Except for the inclusion of the exchange and correlation contribution, a formalism for the calculation of LFE in cubic crystals was established by Adler [28] and Wiser [29] over thirty years ago. This approach has been applied in many modern calculations for a variety of semiconductors [23, 30, 31, 32, 33, 34]. There are, however, some limitations of this method that make it unattractive from an analytic and computational point of view. First, at least in its usual implementation, the method is limited to cubic crystals; the extension to crystals of lower symmetry has been investigated [23, 35], but is still based on relating scalar rather than vector quantities. Second, the formulation of the susceptibility is in terms of charge densities and potentials rather than electric fields and polarizations; this requires the need to take the so-called $q \to 0$ limits either analytically or numerically. Finally, the convergence properties with respect to the number of conduction states is poor. These limitations can be in some instances quite restrictive.

The goal of this work is to present a new approach to the calculation of LFE in semiconductors, extending the susceptibility formalism of SG. It should be emphasized that this approach does not entail any new fundamental view of LFE, but is rather a more general and transparent representation of these effects, and one that
is readily amenable to efficient computation. Our focus is on linear optical response, although the approach presented in this work forms the basis for an explicitly non-divergent second-order response formalism. We note in passing that the extension of local field corrections to the second harmonic generation susceptibility have been considered in different contexts [36, 37].

Equation 7.14 is essentially the basis of the theory developed in this chapter. The two components of this equation that need to be determined are the expressions for the local electric field and the microscopic linear susceptibility. The local field is treated in Sec. 7.2 where an explicit expression for \( e(x, \omega) \) is developed. The quantum mechanical expression for the polarization-polarization susceptibility, \( \chi(x, x'; \omega) \), is derived in Sec. 7.3. This is essentially done as an extension to the formalism of SG for the case of a spatially varying electric field. The electric field, as written in Eq. 7.21, facilitates and connects the derivation of the optical susceptibility formalism in this work to that of SG.

It is useful to highlight the kinds of approximations that are involved in the formulation of optical response in this chapter, and the aspects of other linear susceptibility calculations that are not included. First, we have not taken into account quasi-particle corrections. At the crudest level these can be included in a straightforward way in terms of a scissors operator, as has been done previous chapters [10, 26]. We have ignored such corrections in this work, primarily for clarity of presentation. We note that a full GW quasi-particle correction could also be developed within this formalism. Second, the formalism presented here neglects excitonic effects. Including such effects greatly increases the computational complexity and requirements and would demand a significant extension from the formalism we have developed. Finally, the use of DFT rather than DPFT, and TDLDA, are approximations employed in this work which were discussed earlier.

The chapter is organized in the following way. In Sec. 7.2 an expression for the total local electric field is presented. The linear susceptibility formalism is developed in Sec. 7.3. We discuss details of our calculation in Sec. 7.4, including the electronic structure method we employ and the approach for evaluating the linear susceptibility.
In Sec. 7.5 we present our results for the linear optical response of silicon at energies below the band gap, and discuss them in the context of previous calculations. Finally, our conclusions are presented in Sec. 7.6.

7.2 The Local Field

The local field correction \( e_{\text{lc}}(x, \omega) \) appearing in Eq. 7.19 consists of two terms. The first, \( e_{E}(x, \omega) \), is purely electrodynamical in nature. We restrict ourselves to a "macroscopically infinitesimal region" of the crystal with size of order \( \Delta \); then, assuming \( a \ll \Delta \ll \lambda \), the macroscopic electric field can be taken as essentially uniform, Eq. 7.18 becomes

\[
e_{M}(x, \omega) = E(\omega) + e_{E}(x, \omega),
\]

and we show in Appendix C that \( e_{E}(x, \omega) \) can be written as the gradient of a potential,

\[
e_{E}(x, \omega) = -\nabla \phi_{E}(x, \omega),
\]

where

\[
\phi_{E}(x, \omega) = \sum_{\mathbf{G}} \phi_{E}(\mathbf{G}, \omega)e^{i\mathbf{G} \cdot \mathbf{x}},
\]

with

\[
\phi_{E}(\mathbf{G}, \omega) = -\frac{4\pi i}{\mathbf{G}^2} p(\mathbf{G}, \omega) \cdot \mathbf{G}, \text{ if } \mathbf{G} \neq 0,
\]

\[
= 0, \text{ if } \mathbf{G} = 0,
\]

and where \( p(\mathbf{G}, \omega) \) is the Fourier component of \( p(x, \omega) \),

\[
p(x, \omega) = \sum_{\mathbf{G}} p(\mathbf{G}, \omega)e^{i\mathbf{G} \cdot \mathbf{x}}.
\]

Here the vectors \( \mathbf{G} \) are reciprocal lattice vectors. The absence of the \( \mathbf{G} = 0 \) term in Eq. 7.27 is clearly consistent with the macroscopic electric field \( E \) being the spatial
average of \( e_M \).

The second contribution to \( e_{LFC}(x, \omega) \), \( e_{XC}(x, \omega) \), is associated with the exchange and correlation potential. If we define

\[
\beta(x) = \left( \frac{dV_{XC}(s)}{ds} \right)_{s=\rho_0(x)}
\]  

(7.29)

and identify \( \rho(x, \omega) = -\nabla \cdot p(x, \omega) \) from the definition 7.13 of \( p(x, \omega) \) and the equation of continuity, we can write Eq. 7.17 as

\[
\phi_{XC}(x, \omega) = -\beta(x) \nabla \cdot p(x, \omega)
\]  

(7.30)

Since \( \beta(x) \) has the periodicity of the lattice we can expand

\[
\beta(x) = \sum_G \beta(G)e^{iG \cdot x}
\]  

(7.31)

and, similarly expanding \( \phi_{XC}(x, \omega) \), we find

\[
\phi_{XC}(G, \omega) = -i \sum_{G'} \beta(G - G')G' \cdot p(G', \omega)
\]  

(7.32)

Note that, unlike \( \phi_E(G, \omega) \), \( \phi_{XC}(G, \omega) \) can have a \( G = 0 \) term. However, it will make no contribution to \( e_{XC}(x, \omega) \).

Returning now to Eq. 7.21, in a macroscopically infinitesimal region of interest we see that there are two effective "driving fields" present. The first is a field that is uniform over the region of interest, the macroscopic Maxwell field. We will describe it by a vector potential \( A(t) \), writing \( E(t) = -c^{-1}\dot{A}(t) \). The second can be described in terms of a scalar potential \( \phi_{LFC}(x, \omega) \equiv \phi_E(x, \omega) + \phi_{XC}(x, \omega) \). It includes both electrodynamical local field corrections and a contribution from the exchange and correlation potential, and must be solved for self-consistently in terms of \( p(x, \omega) \), as we have described above. We turn to the response of the crystal to these two driving fields in the following section.
7.3 Optical Response Formalism

7.3.1 Single-Particle Wavefunctions

In the TDLDA the time-dependent Kohn-Sham equations are solved in the presence of a perturbation. In the problem at hand the perturbation is characterized by a uniform vector potential $A$ describing the macroscopic Maxwell field, and a scalar potential $\phi_{LFC}$ describing the local field corrections.

We begin by considering the Kohn-Sham Hamiltonian appropriate when no external field is present. This Hamiltonian is given by

$$H_o = \frac{1}{2m} \left( \frac{\hbar}{i} \nabla \right)^2 + V(x) ,$$

where $m$ is the electronic mass, and $V(x)$ is the periodic crystal potential, including the exchange and correlation potential. The eigenstates of the Hamiltonian given in Eq. 7.33 satisfy

$$H_o \psi_n(k; x) = \hbar \omega_n(k) \psi_n(k; x) ,$$

where the eigenstates are of the form

$$\langle x|n \rangle = \psi_n(k; x) = \frac{1}{\sqrt{\Omega}} u_n(k; x)e^{i k \cdot x} ,$$

and where $\Omega$ is the normalization volume of the crystal, and the $u_n(k; x)$ are periodic with respect to a direct lattice vector.

Now consider a crystal in the presence of a uniform vector potential describing the macroscopic Maxwell field. The Kohn-Sham Hamiltonian is then given by

$$H_{NLF}(t) = \frac{1}{2m} \left( \frac{\hbar}{i} \nabla + \hbar \mathbf{K}(t) \right)^2 + V(x) ,$$

where for convenience we have defined

$$\mathbf{K}(t) \equiv \frac{e}{\hbar c} \mathbf{A}(t) .$$
The subscript on $H_{NLF}(t)$ is to signify that local field effects (or the spatial variation of the electric field) are not explicitly accounted for in such a Hamiltonian. This Hamiltonian is the starting point of the susceptibility formalism of SG [1] and is related to our discussion following Eq. 7.21. The instantaneous eigenstates of $H_{NLF}(t)$, which have been discussed in detail in SG, satisfy

$$H_{NLF}(t)\overline{\psi}_n(k; x) = \hbar \omega_n(k + K(t))\overline{\psi}_n(k; x), \quad (7.38)$$

and are of the form

$$\langle x|n\overline{k}\rangle = \overline{\psi}_n(k; x) = \frac{1}{\sqrt{\Omega}}u_n(k + K(t); x)e^{iK\cdot x}e^{-i\phi_n(k; t)}.$$

(7.39)

The over-bar in $\overline{\psi}_n(k; x)$ indicates an implicit time dependence for these instantaneous eigenstates. Note that a Berry's phase $\phi_n(k; t)$ appears in the eigenstate definition, and is explicitly defined in equation (2.24) in SG. We refer the reader to SG for a further discussion of these eigenstates.

Finally, we include the local field corrections through the scalar potential $\phi_{LFC}(x, t)$. The complete Kohn-Sham Hamiltonian then becomes

$$H(t) = \frac{1}{2m} \left( \frac{\hbar}{i} \nabla + \hbar K(t) \right)^2 + V(x) + e\phi_{LFC}(x, t). \quad (7.40)$$

We can write Eq. 7.40 equivalently as

$$H(t) = H_{NLF}(t) + e\phi_{LFC}(x, t), \quad (7.41)$$

which makes the connection to the approximations of the SG approach clear. It will be convenient to have Eq. 7.41 expressed in another way. We define

$$n(x, x') = \sum_{nk} \psi_n(k; x)\overline{\psi}^*_n(k; x') = \sum_{nk} \overline{\psi}_n(k; x)\overline{\psi}^*_n(k; x'), \quad (7.42)$$

where completeness guarantees that $n(x, x')$ is the Dirac delta function $\delta(x - x')$, but
extended to be periodic on the scale of the normalization volume, since this is the periodicity of both sets of eigenstates. Equation 7.41 can then be written as

$$H(t) = H_{NFL}(t) + e \int_{\Omega} n(x, x') \phi_{LEC}(x', t) dx'. \quad (7.43)$$

In the following we will require matrix elements of the operators in Eq. 7.43, which we now develop.

Introducing a short-hand notation, we write

$$\bar{\pi}_{nm}(k; x) \equiv \int_{\Omega} \bar{\psi}_n(k; x') n(x, x') \bar{\psi}_m(k; x') dx', \quad (7.44)$$

where the overbar in $$\bar{\pi}_{nm}(k; x)$$ signifies that the matrix elements are with respect to the eigenstates of Eq. 7.39; this will be our convention for the rest of the chapter. Further note that for matrix elements between the eigenstates of Eq. 7.35 the overbar will be omitted. Using this notation, it is possible to write the continuity equation as

$$\frac{i}{\hbar} \int_{\Omega} \bar{\psi}_n(k; x') [H_{NFL}(t), e n(x, x')] \bar{\psi}_m(k; x') dx' + \nabla \cdot \bar{J}_{nm}(k; x, t) = 0, \quad (7.45)$$

where

$$\bar{J}_{nm}(k; x, t) = \int_{\Omega} \bar{\psi}_n(k; x') j(x, x', t) \bar{\psi}_m(k; x') dx'. \quad (7.46)$$

is the microscopic current density operator.

It is straightforward to show that the microscopic current density operator can be written in terms of $$n(x, x')$$ as

$$j(x, x', t) = \frac{e}{2} (\xi(x, t) n(x, x') + n(x, x') \xi(x, t)) , \quad (7.47)$$

where the velocity operator is

$$\xi(x, t) = \frac{1}{m} \left( \frac{\hbar}{i} \nabla + \hbar \mathbf{K} \right). \quad (7.48)$$

The explicit dependence on time for $$\mathbf{K}(t)$$ has been dropped for notational convenience.
in the above equation as it will be throughout the rest of the chapter. The first term on the left hand side of Eq. 7.45 can be rewritten as

\[
\frac{i}{\hbar} \int_{\Omega} \psi_n(\mathbf{k}; \mathbf{x}') [H_{NL_F}(t), e\nu(\mathbf{x}, \mathbf{x}')] \overline{\psi}_m(\mathbf{k}; \mathbf{x}') d\mathbf{x}' = i\omega_{nm}(\mathbf{k} + \mathbf{K}) (e\overline{n}_{nm}(\mathbf{k}; \mathbf{x})) \tag{7.49}
\]

having used Eq. 7.38, and where \(\omega_{nm}(\mathbf{k} + \mathbf{K}) = \omega_n(\mathbf{k} + \mathbf{K}) - \omega_m(\mathbf{k} + \mathbf{K})\). Combining Eqs. 7.45 and 7.49 we obtain for \(n \neq m\)

\[
\overline{n}_{nm}(\mathbf{k}; \mathbf{x}) = \frac{i}{e} \frac{\overline{\mathbf{j}}_{nm}(\mathbf{k}; \mathbf{x}, t)}{\omega_{nm}(\mathbf{k} + \mathbf{K})}. \tag{7.50}
\]

It will be convenient to define

\[
\overline{\nu}_{nm}(\mathbf{k}; \mathbf{x}, t) \equiv \frac{\Omega}{e} \overline{j}_{nm}(\mathbf{k}; \mathbf{x}, t), \tag{7.51}
\]

and for \(n \neq m\) we also define

\[
\overline{\eta}_{nm}(\mathbf{k}; \mathbf{x}, t) \equiv \frac{e}{i} \frac{\overline{\nu}_{nm}(\mathbf{k}; \mathbf{x}, t)}{\omega_{nm}(\mathbf{k} + \mathbf{K})}. \tag{7.52}
\]

We can then write Eq. 7.50 as

\[
e\overline{n}_{nm}(\mathbf{k}; \mathbf{x}) = -\frac{1}{\Omega} \nabla \cdot \overline{\eta}_{nm}(\mathbf{k}; \mathbf{x}, t). \tag{7.53}
\]

The diagonal elements of \(\overline{\eta}_{nm}(\mathbf{k}; \mathbf{x}, t)\) also need to be defined. To do this, we modify Eq. 7.53 and use it as our definition. This equation becomes

\[
e\overline{n}_{nn}(\mathbf{k}; \mathbf{x}) = -\frac{1}{\Omega} \nabla \cdot \overline{\eta}_{nn}(\mathbf{k}; \mathbf{x}, t). \tag{7.54}
\]

We note that this equation does not uniquely define \(\overline{\eta}_{nn}(\mathbf{k}; \mathbf{x}, t)\). It can be made unique by requiring, for example, \(\nabla \times \overline{n}_{nn}(\mathbf{k}; \mathbf{x}, t) = 0\).

For \(n \neq m\), the quantities \(\overline{n}_{nm}(\mathbf{k}; \mathbf{x}, t)\) are related to interband matrix elements
of the position operator. Defining

$$v_{nm}(k; x) = \int_\Omega \psi^*_n(k; x')v(x, x')\psi_m(k; x')dx',$$  \hfill (7.55)

where

$$v(x, x') = \frac{\Omega}{2} (\xi_o(x)n(x, x') + n(x, x')\xi_o(x)),$$  \hfill (7.56)

with

$$\xi_o(x) = \frac{1}{m} \left( \frac{\hbar}{i} \nabla \right),$$  \hfill (7.57)

it then follows from the overbar states that

$$\bar{v}_{nm}(k; x, t) = v_{nm}(k + K; x)e^{i\phi_{nm}(k; t)},$$  \hfill (7.58)

where $\phi_{nm}(k; t) = \phi_n(k; t) - \phi_m(k; t)$. Setting

$$q_{nm}(k; x) \equiv \frac{v_{nm}(k; x)}{i\omega_{nm}}$$  \hfill (7.59)

for $n \neq m$, we then can write

$$\bar{\eta}_{nm}(k; x, t) = \epsilon q_{nm}(k + K; x)e^{i\phi_{nm}(k; t)},$$  \hfill (7.60)

where we have used Eqs. 7.52 and 7.58. It immediately follows that, for $n \neq m$,

$$r_{nm}(k) = \frac{1}{\Omega} \int_\Omega q_{nm}(k; x)dx.$$  \hfill (7.61)

where $r_{nm}(k)$ is the interband matrix element of the position operator [1, 2], and the effective dipole moment matrix element $\bar{\mu}_{nm}(k; t)$ defined in SG is given by

$$\bar{\mu}_{nm}(k; t) = \frac{1}{\Omega} \int_\Omega \bar{\eta}_{nm}(k; x, t)dx.$$  \hfill (7.62)
7.3.2 Hamiltonian

Turning now to the derivation of an expression for the many-particle Hamiltonian, we introduce the field operator in the Heisenberg picture (denoted with an $H$ subscript) as

$$\Psi_H(x, t) = \sum_{nk} a_{nk}(t)\psi_n(k; x),$$  \hspace{1cm} (7.63)

where $a_{nk}(t)$ is the usual fermion destruction operator for an electron in band $n$ with wavevector $k$. Since the instantaneous eigenstates $\psi_n(k; x)$ also form a complete set, we can write an equivalent expression for the field operator as

$$\Psi_H(x, t) = \sum_{nk} \overline{a}_{nk}(t)\overline{\psi}_n(k; x).$$  \hspace{1cm} (7.64)

Note that the fermion destruction operators in Eq. 7.64 are still time dependent, but this time dependence is different from that in Eq. 7.63 since the basis states in Eq. 7.64 are also time dependent. We have denoted these new fermion destruction operators with an over-bar to distinguish them from those in Eq. 7.63. The Hamiltonian of our system is given by

$$\mathcal{H}(t) = \int \Psi_H^\dagger(x, t)H(t)\Psi_H(x, t)dx,$$  \hspace{1cm} (7.65)

or

$$\mathcal{H}(t) = \sum_{nk} \overline{a}_{nk}^\dagger(t)\overline{a}_{nk}(t)\hbar\omega_n(k + \mathbf{K})$$  
$$+ \sum_{nmkk'} \sum_{nk} \overline{a}_{nk}^\dagger(t)\overline{a}_{mk'}(t) \int_{\Omega} e^{i\overline{\psi}_n^*(k; x)n(x, x')\overline{\psi}_m(k'; x)\phi_{LFC}(x', t)}dx'dx,$$  \hspace{1cm} (7.66)

where we have used Eq. 7.64 and the expression for $H(t)$ in Eq. 7.43.

We now simplify the local field correction term in this expression. We begin by noting that the integral in Eq. 7.66 vanishes unless $k = k'$, a result that follows from the periodicity of $\phi_{LFC}(x', t)$ when that integral is broken up into integrals over the primitive unit cells in the normalization volume. Thus, recalling the definition of
\( \tilde{\pi}_{nm}(k; x) \) in Eq. 7.44, we can write Eq. 7.66 as

\[ \mathcal{H}(t) = \sum_{nk} \bar{a}_{nk}^\dagger(t) \bar{a}_{nk}(t) \hbar \omega_n(k + K) + \sum_{nmk} \bar{a}_{nk}^\dagger(t) \bar{a}_{mk}(t) \int_\Omega e_{nm}(k; x') \phi_{LFC}(x', t) dx'. \]  

(7.67)

Next, a partial integration of the integral remaining yields

\[ \int_\Omega e_{nm}(k; x') \phi_{LFC}(x', t) dx' = -\frac{1}{\Omega} \int_\Omega \nabla \cdot \tilde{\eta}_{nm}(k; x', t) \phi_{LFC}(x', t) dx' \]

= \frac{1}{\Omega} \int_\Omega \tilde{\eta}_{nm}(k; x', t) \cdot \nabla \phi_{LFC}(x', t) dx'. \]  

(7.68)

so finally, since \( e_{LFC}(x, t) = -\nabla \phi_{LFC}(x, t) \), we have

\[ \mathcal{H}(t) = \sum_{nk} \bar{a}_{nk}^\dagger(t) \bar{a}_{nk}(t) \hbar \omega_n(k + K) - \frac{1}{\Omega} \sum_{nmk} \bar{a}_{nk}^\dagger(t) \bar{a}_{mk}(t) \int_\Omega \tilde{\eta}_{nm}(k; x', t) \cdot e_{LFC}(x', t) dx'. \]  

(7.69)

### 7.3.3 Linear Susceptibility

We can now derive an expression for the microscopic linear susceptibility, \( \tilde{\chi}(x, x'; \omega) \).

We begin by introducing a many-particle microscopic polarization operator

\[ \hat{p}(x, t) = \frac{1}{\Omega} \sum_{nmk} \bar{a}_{nk}^\dagger(t) \bar{a}_{nk}(t) \tilde{\eta}_{nm}(k; x, t). \]  

(7.70)

Note that this operator is the microscopic analog of that introduced in equation (2.65) of SG; this point will be further developed later in this section. The microscopic polarization is simply the expectation value of this operator [38]

\[ \langle \hat{p}(x, t) \rangle = p(x, t) \]  

(7.71)

and so

\[ p(x, t) = \frac{1}{\Omega} \sum_{nmk} c_{mn}(k; t) \tilde{\eta}_{nm}(k; x, t), \]  

(7.72)
where \( c_{mn}(k; t) \equiv \langle \bar{a}^\dagger_{nk}(t) \bar{a}_{mk}(t) \rangle \). The dynamical equation for \( c_{mn}(k; t) \) follows from the operator equation for the \( \bar{a}_{mk}(t) \). This is easiest recovered by starting with

\[
 i\hbar \frac{\partial}{\partial t} \psi_{H}(x, t) = [\psi_{H}(x, t), \mathcal{H}(t)]
\]  

(7.73)

and using the time dependence of the \( \bar{\psi}_m(k; x) \) \[1\]

\[
 i\hbar \frac{d}{dt} \bar{\psi}_m(k; x) = \sum_q \bar{\mu}_{qm}(k, t) \cdot E(t) \bar{\psi}_q(k; x),
\]  

(7.74)

where \( E(t) = -c^{-1} \dot{A}(t) \); the prime on the summation indicates that the sum is over states \( \bar{\psi}_q(k; x) \) for which \( \omega_q(k + K) \neq \omega_n(k + K) \). We find

\[
 i\hbar \frac{\partial}{\partial t} \bar{a}_{mk}(t) = \hbar \omega_{m}(k + K) \bar{a}_{mk}(t) - \sum_q \bar{\mu}_{mq}(k, t) \cdot E(t) \bar{\psi}_q(k; x) \\
\quad - \frac{1}{\Omega} \sum_q \int_{\Omega} \bar{\eta}_{mq}(k; x', t) \cdot e_{LC}(x', t) dx' \bar{\psi}_q(k; x)
\]

\[
 = \hbar \omega_{m}(k + K) \bar{a}_{mk}(t) \\
\quad - \frac{1}{\Omega} \sum_q \int_{\Omega} \bar{\eta}_{mq}(k; x', t) \cdot e(x', t) dx' \bar{\psi}_q(k; x),
\]  

(7.75)

where we have used Eqs. 7.20 and 7.62. Then

\[
 i\hbar \frac{\partial}{\partial t} c_{mn}(k; t) = \hbar \omega_{mn}(k + K) c_{mn}(k; t) \\
\quad - \frac{1}{\Omega} \sum_q \left\{ \int_{\Omega} \bar{\eta}_{mq}(k; x', t) \cdot e(x', t) dx' c_{qn}(k; t) \\
\quad - \int_{\Omega} \bar{\eta}_{qn}(k; x', t) \cdot e(x', t) dx' c_{mq}(k; t) \right\}
\]  

(7.76)

follows immediately. We consider a perturbation expansion for the \( c_{mn}(k; t) \), which will naturally lead to an expansion of \( p(x, t) \) in orders of the electric field. \( e(x, t) \).

The \( c_{mn}(k; t) \) can be written as

\[
c_{mn}(k; t) = c_{mn}^{(0)}(k) + c_{mn}^{(1)}(k; t).
\]  

(7.77)
with $c^{(2)}_{mn}(k) = f_m(k)\delta_{mn}$, and where $f_m(k)$ is the Fermi occupation factor; for the clean, cold semiconductors considered here it is equal to either 0 or 1 independent of $k$, $f_m(k) = f_m$.

Taking an electric field of the form

$$e(x, t) = e(x, \omega_\beta)e^{-i\omega_\beta t}, \quad (7.78)$$

where a sum over $\omega_\beta$ is implied, we substitute Eqs. 7.77 and 7.78 into Eq. 7.76, and seek a solution of the form

$$c^{(1)}_{mn}(k; t) = B_{mn}^{(2)}(k; t)e^{-i\omega_\beta t} + c^{(2)}_{mn}(k; t), \quad (7.79)$$

where $\beta$ is to be summed over, and $c^{(2)}_{mn}(k; t)$ contains terms at least of second-order in $e(x, t)$. We find Eq. 7.76 can be satisfied with

$$B_{mn}^{(2)}(k; t) = \frac{1}{\hbar \Omega} \frac{f_{nm}}{\omega_{mn}(k + K(t)) - \omega_\beta} \int_\Omega \overline{\eta}_{mn}(k; x', t) \cdot e(x', \omega_\beta) dx' \quad (7.80)$$

and a $c^{(2)}_{mn}(k; t)$ the expression for which we will not give here, since our interest is only in linear response; for clarity we have indicated the time dependence of $K(t)$.

We are now in a position to develop the expression for the microscopic polarization that was introduced in Eq. 7.72. If this polarization is written in a way analogous to Eq. 7.78 as

$$p(x, t) = p(x, \omega_\beta)e^{-i\omega_\beta t}, \quad (7.81)$$

using Eqs. 7.72, 7.79, and 7.80 we have

$$p^a(x, \omega) = \frac{1}{\Omega} \int_{\Omega} \chi^{ab}(x, x'; \omega)e^b(x', \omega) dx', \quad (7.82)$$

where Cartesian components are to be summed over if repeated, and

$$\chi^{ab}(x, x'; \omega) = \frac{1}{\hbar \Omega} \sum_{nmk} \frac{f_{nm}}{\omega_{mn}(k + K(t)) - \omega} \overline{\eta}^a_{nm}(k; x, t)\overline{\eta}^b_{nm}(k; x', t). \quad (7.83)$$
We have dropped the subscript on \( \omega \), as there is no ambiguity for linear response. This expression can be simplified by using results developed in Sec. 7.3.1. Using Eqs. 7.60 and 7.59 the above equation becomes

\[
\chi^{ab}(x, x'; \omega) = \frac{e^2}{\hbar \Omega} \sum_{nmk} \frac{f_{nm}}{(\omega_{mn}(k) - \omega)} \frac{v_{nm}^a(k; x)v_{mn}^b(k; x')}{(\omega_{mn}(k))^2},
\]  

(7.84)

with the \( v_{nm}(k; x) \) defined in Eq. 7.55. The time dependence in (7.83) has vanished (SG): In the case of filled bands, regardless of the value of \( \mathbf{K}(t) \) the sum over \( k \) in the first Brillouin zone involves a set of vectors \( k + \mathbf{K}(t) \) which are equivalent to the set of vectors \( k \). Equation 7.84 is our final expression for the microscopic linear susceptibility. Note that only the total field \( e(x, \omega) \) appears in Eq. 7.82 and not its separate components from the macroscopic electric field (described here by a vector potential \( A(t) \)) and the local correction field (described here by a scalar potential \( \phi_{LFC}(x, \omega) \)). Thus our result exhibits the expected gauge invariance.

The macroscopic polarization, \( P(\omega) \), is given by

\[
P^a(\omega) = \frac{1}{\Omega} \int_{\Omega} p^a(x, \omega) dx
\]

and so using Eq. 7.82 we obtain

\[
P^a(\omega) = \frac{1}{\Omega^2} \int_{\Omega} \chi^{ab}(x, x'; \omega)e^b(x', \omega) dx' dx.
\]

(7.86)

Finally, we note that in the neglect of local field corrections

\[
\chi_{NLF}^{ab}(\omega) = \frac{e^2}{\hbar \Omega} \sum_{nmk} \frac{f_{nm}v_{nm}^a(k)v_{mn}^b(k)}{(\omega_{mn}(k) - \omega)(\omega_{mn}(k))^2},
\]

(7.87)

where we have used Eqs. 7.23, 7.84, 7.59, and 7.61. This agrees with the standard result if local field corrections are neglected [1].
7.3.4 Self-Consistent Equation

To solve for $\chi^{ab}(\omega)$ in the presence of local field corrections we must solve Eq. 7.86 for $P^a(\omega)$ in the presence of a specified macroscopic field $E^b(\omega)$, and identify $\chi^{ab}(\omega)$ from

$$P^a(\omega) = \chi^{ab}(\omega)E^b(\omega). \quad (7.88)$$

We begin by noting that since the microscopic polarization is periodic over a unit cell we can write

$$p^a(x, \omega) = \sum_G p^a(G, \omega)e^{iG \cdot x}, \quad (7.89)$$

as was done in Eq. 7.28, where the $G$ are reciprocal lattice vectors. The Fourier components of the polarization are determined from

$$p^a(G, \omega) = \frac{1}{\Omega_c} \int p^a(x, \omega)e^{-iG \cdot x}d\mathbf{x}, \quad (7.90)$$

where $\Omega_c$ denotes the volume of the primitive unit cell. Using this equation with similar expressions for $e(x, \omega)$ and the $v_{nm}(k; x)$, which are periodic in $x$, we find

$$\chi^{ab}(x, x'; \omega) = \sum_{GG'} \chi^{ab}(G, G'; \omega)e^{iG \cdot x}e^{iG' \cdot x'}. \quad (7.91)$$

where

$$\chi^{ab}(G, G'; \omega) = \frac{e^2}{\hbar \Omega} \sum_{nmk} \frac{f_{nm}}{(\omega_{mn}(k) - \omega)} \frac{v^a_{nm}(k; G)v^b_{mn}(k; G')}{(\omega_{mn}(k))^2}. \quad (7.92)$$

Using the Fourier expansion for $e^b(x, \omega)$, we have

$$e^b(G, \omega) = E^b(\omega)\delta_{G,0} - 4\pi \frac{1}{G^2} \rho^b(G, \omega)G^c(1 - \delta_{G,0}) - G^b \sum_{G'} \beta(G - G')G'^c\rho^c(G', \omega), \quad (7.93)$$

where the $(1 - \delta_{G,0})$ factor is included in the second term on the right hand side as the $G = 0$ term is zero (see Eq. 7.27). Using Eq. 7.89 in Eq. 7.82, we immediately
\[ p^a(G, \omega) = \sum_{G'} \chi^{ab}(G, -G'; \omega) e^b(G', \omega) \]
\[ = \chi^{ab}(G, 0; \omega) E^b(\omega) - 4\pi \sum_{G' \neq 0} \chi^{ab}(G, -G'; \omega) G''^b p^c(G', \omega) G'^c \]
\[ - \sum_{G', G''} \chi^{ab}(G, -G'; \omega) G''^b \beta(G' - G'') p^c(G'', \omega) G'^c. \quad (7.94) \]

This is a self-consistent equation for the Fourier components of the microscopic polarization. The \( p(G, \omega) \) can be determined by successive substitution in the above equation, or alternatively by writing it as a matrix equation and inverting the matrix. The macroscopic polarization is simply the zeroth Fourier component of the microscopic polarization
\[ P^a(\omega) = p^a(0, \omega), \quad (7.95) \]
and using Eq. 7.88 the linear susceptibility \( \chi^{ab}(\omega) \) in the presence of local field corrections is determined.

### 7.4 Calculational Details

In this section we briefly describe the method by which we calculate the linear susceptibility. In previous chapters, we have presented calculations of optical properties excluding LFE for a variety of semiconductors using the full-potential linearized augmented plane-wave (FLAPW) method [10, 26]. Such a method is cumbersome if local field corrections are included, regardless of the formalism or approach adopted. This point can be understood by considering the form of the linear microscopic susceptibility in Eq. 7.92, and its use in the self-consistent equation of Sec. 7.3.4. The \( v_{nm}(k; G) \) of Eq. 7.92 involve a Fourier transform of products of wavefunctions and their derivatives. Ideally, one would like this Fourier expansion to rapidly converge with the number of \( G \) vectors. As the FLAPW method is all-electron in nature, the wavefunctions rapidly vary within the core regions. Fourier expanding products of these wavefunctions and their derivatives requires a large number of reciprocal lattice
vectors.

A simple way to overcome this problem is to employ another electronic structure method that ensures a smooth behaviour of the wavefunctions over all space. Following others [23, 31, 32, 34], we have chosen to use the pseudopotential plane-wave (PPW) method, as indeed the pseudo-wavefunctions smoothly vary over the entire crystal. But three issues immediately arise concerning the use of a PPW method for the calculation of optical properties, especially including local field corrections: (1) The use of pseudo-wavefunctions and the associated pseudo-charge densities in place of the actual valence quantities, (2) The form of the velocity operator in a theory involving nonlocal pseudopotentials, and (3) The form of the (pseudo-) current density operator in such a theory. Of these issues, which we briefly discuss below, the first and third are typically ignored completely.

While the pseudo- and all-electron valence wavefunctions are identical in areas of the crystal outside the core region, inside the core region the pseudo-wavefunction is smoothly varying whereas its all-electron counterpart can exhibit strong oscillations. These oscillations are a result of the orthogonality of the all-electron valence wavefunction to the core wavefunctions. In replacing the "true" valence wavefunction with a pseudo-wavefunction the implicit approximation is that the difference between these two does not have a significant effect on the calculated property of interest. While this approximation has been shown to be reasonable in various ground state calculations, its validity in the case of optical response remains in doubt. In an analysis of the optical response of semiconductors excluding LFE, we have found reasonable agreement between the FLAPW and PPW methods [39]. Although this lends some support to the appropriateness of a PPW based optical response calculation, it is of concern that the rapid variations of the all-electron wavefunction in the core region, which would possibly imply large LFE, are completely omitted in a PPW calculation. We follow others [23, 31, 32, 34] in simply passing over this issue, but a comprehensive analysis of the differences between an all-electron and a PPW approach in optical response calculations is clearly needed.

The other aspects of a PPW calculation that merit attention involve the presence
of a nonlocal potential in the Kohn-Sham Hamiltonian. Rather than Eq. 7.33 one typically has
\[
H_o = \frac{1}{2m} \left( \frac{\hbar}{i} \nabla \right)^2 + V(x, \frac{\hbar}{i} \nabla),
\] (7.96)
where \( V(x, \frac{\hbar}{i} \nabla) \) contains both a local and a nonlocal contribution. Correspondingly, Eq. 7.36 becomes
\[
H_{NL}(t) = \frac{1}{2m} \left( \frac{\hbar}{i} \nabla + \hbar \mathbf{K}(t) \right)^2 + V(x, \frac{\hbar}{i} \nabla + \hbar \mathbf{K}(t)).
\] (7.97)
The nonlocal potential is written as a function of both position and momentum, as described by Starace [40].

The fact that the velocity operator is modified in the presence of a nonlocal potential is well known [32, 41, 42]. This follows directly from the definition of the velocity as the commutator of the position operator and the Hamiltonian. With the Hamiltonian of Eq. 7.96, the velocity operator can be shown to be
\[
\xi_o(x) = \frac{1}{m} \left( \frac{\hbar}{i} \nabla \right) + \frac{\partial}{\partial(\frac{\hbar}{i} \nabla)} V_{NL}(x, \frac{\hbar}{i} \nabla)
\] (7.98)
rather than Eq. 7.57. Furthermore, in the presence of a vector potential, where the Hamiltonian of (7.97) is appropriate, the velocity operator is given by
\[
\xi(x, t) = \frac{1}{m} \left( \frac{\hbar}{i} \nabla + \hbar \mathbf{K} \right) + \frac{\partial}{\partial(\frac{\hbar}{i} \nabla)} V_{NL}(x, \frac{\hbar}{i} \nabla + \hbar \mathbf{K})
\] (7.99)
instead of Eq. 7.48.

Despite the consensus in the literature regarding the expression for the velocity operator in the presence of a nonlocal potential, the expression for the microscopic current density operator has not, to our knowledge, been addressed in any detail. While a current density operator can be found that satisfies the continuity equation (7.45), it no longer has a form similar to that given in Eq. 7.47. It can only be
approximately written as

\[ j(x, x', t) \approx \frac{e}{2} (\xi(x, t)n(x, x') + n(x, x')\xi(x, t)) . \]  

(7.100)

Similarly, in the absence of a vector potential we write

\[ j(x, x') \approx \frac{e}{2} (\xi_0(x)n(x, x') + n(x, x')\xi_0(x)) , \]  

(7.101)

which follows from Eqs. 7.56 and 7.51. Note that these expressions are only approximate even with the modified velocity expressions \((7.98, 7.99)\) At the level of the approximations used to write Eqs. 7.100 and 7.101, the formalism presented in Sec. 7.3 remains unchanged.

Equation 7.101, or equivalently Eq. 7.100, was previously presented by Del Sole and Girlanda \([41]\), although its approximate nature was not made clear at that time. The approximation in writing (7.101) is a neglect of a series of terms involving the nonlocal potential. Indeed, it can be shown that, roughly speaking, these terms involve higher-order derivatives of the nonlocal potential with respect to \(\nabla\frac{\hat{A}}{i}\). The basis for the lack of a strict equality in Eq. 7.101 stems from the complexity of taking commutators of \(n(x, x')\) and \(V_{ NL}(x, \nabla\frac{\hat{A}}{i})\). We have attempted to estimate the error introduced in the linear optical response function in using the approximate microscopic current density operator above (7.101). Although such an estimation is very difficult in a strict sense, it appears that it should be on the order of 5%.

One final point worth noting in the context of this discussion relates to numerical efficiency. Although using a PPW method ensures a rapidly convergent Fourier expansion of the wavefunction (and therefore decreased numerical effort), the end effect of the nonlocality of the pseudopotential is to greatly increase the computational requirements, independent of the approach and formalism used in the optical calculation including LFE. While these effects do not precisely offset each other, they do underscore the point that all-electron calculations, dependent on only local potentials, are perhaps not as computationally prohibitive as it would at first seem.

The PPW method used in this work is an implementation of a modified molecular-
A 15 $\text{Ry}$ energy cutoff has been used, which produces well converged electronic band energies and wavefunctions. The soft-core norm-conserving nonlocal pseudopotential of Hamann [44, 45] forms the basis for this electronic structure calculation. The LDA exchange-correlation potential as parameterized by Perdew and Zunger [46] is used.

The electronic eigenvalues and wavefunctions are used to evaluate $\chi^{ab}(G, G'; \omega)$ in Eq. 7.92. Rather than calculating the $v^{a}_{nm}(k; x)$ first and then using a Fourier expansion to obtain the $v^{a}_{nm}(k; G)$, we have chosen to calculate the latter directly. This is straightforward when using the PPW method, as the wavefunctions are explicitly written as a sum over plane-waves. Having constructed $\chi^{ab}(G, G'; \omega)$, what remains is to evaluate the macroscopic susceptibility. This is done by iteratively solving Eq. 7.94. The procedure is as follows. First, one sets the value of the macroscopic Maxwell field, $E^b(\omega)$, and leaves this value fixed throughout the calculation. Next, a first guess at the microscopic polarization is made by assuming it is due to simply the first term on the right hand side of Eq. 7.94. The iteration then proceeds in the usual way since the second and third terms in Eq. 7.94 contain the microscopic polarization itself. The output polarization is mixed with the input polarization after each iteration to assist in the self-consistent process. One then recovers the linear susceptibility using Eqs. 7.95 and 7.88. The self-consistent result has been found to be robust to the initial guess for the microscopic polarization. The most time consuming aspect of the entire calculation, from a computational perspective, is in the evaluation of $\chi^{ab}(G, G'; \omega)$. The self-consistent procedure requires a relatively small amount of time to perform.

### 7.5 Results and Discussion

We have applied the formalism and approach introduced here to the calculation of the dielectric function, $\varepsilon^{ab}(\omega) = 1 + 4\pi \chi^{ab}(\omega)$, of silicon. We consider this an appropriate test case as it is a semiconductor that has been previously investigated using the traditional Adler and Wiser approach; this will allow for a comparative analysis between
methods. It will be the goal of future work to extend our calculational approach to other semiconductors of interest.

An important consideration in assessing the accuracy and efficiency of our approach is the convergence properties of the calculated values with respect to parameters in the calculation itself. There are three obvious parameters to investigate in the determination of the microscopic linear susceptibility: the number of \( \mathbf{G} \) vectors included in the Fourier expansion, the number of conduction bands, and finally, the the number of \( \mathbf{k} \) points used in the irreducible Brillouin zone (IBZ).

As a means of measuring the accuracy of the Fourier expansion of \( \chi^{ab}(\mathbf{G}, \mathbf{G}'; \omega) \) as well as the self-consistent procedure itself, we have calculated the value of the dielectric function at zero frequency, \( \varepsilon(0) \), with respect to the number of \( \mathbf{G} \) vectors included in Eq. 7.94. The results are presented in Fig. 7.1. Note that the number of \( \mathbf{G} \) vectors that are considered are restricted, as these correspond to complete sets of vectors below a certain maximum \( \mathbf{G} \) value. We have used 110 \( \mathbf{k} \) points in the calculation of \( \chi^{ab}(\mathbf{G}, \mathbf{G}'; \omega) \) in Eq. 7.92. It is clear from the figure that reasonably good convergence in the \( \varepsilon(0) \) value is achieved after including only 65 \( \mathbf{G} \) vectors. The value does exhibit some irregularity prior to this point, but beyond it settles down quite quickly. On this basis, we believe that for non-zero frequency calculations a minimum of 89 \( \mathbf{G} \) points is advisable. We note, however, that for the below band gap response subsequently presented, we have used 169 \( \mathbf{G} \) vectors.

The convergence of the calculated \( \varepsilon(0) \) with respect to the number of bands in the sum in Eq. 7.92 has also been investigated. We have calculated the dielectric function for six and sixteen conduction states in the sum for a total of ten and twenty bands respectively. Our results indicate a remarkable insensitivity to the number of conduction bands used in the two calculations; differences between the \( \varepsilon(0) \) values for all \( \mathbf{G} \) vectors in Fig. 7.1 remain very small, and are typically less than 2%. This result is striking given the observations made in calculations that employ the traditional Adler and Wiser approach [31, 32, 47]. For example, Gavrilenko and Bechstedt [31] found a significant dependence of \( \varepsilon(0) \) on the number of conduction bands used; their calculated value began to converge only after over 75 states had been included, and
they decided to use 176 conduction bands in calculations they presented in their paper. The difference in computational efficiency between these two approaches in light of the observed convergence of $\varepsilon(0)$ with respect to conduction band number is significant. Moreover, the use of states high within the conduction band, at energies far above the band gap, is troublesome; the accurate calculation of these states is clearly in question.

The observed convergence properties with regard to number of conduction states can be understood in terms of the expressions for the susceptibility in the two differing approaches. In the formalism presented here, the form of the linear microscopic susceptibility is very similar to the expression for the macroscopic susceptibility that
excludes local field corrections \([1]\). In Eq. 7.92, one can see that \(\chi^{ab}(G, G'; \omega)\) depends on the band energy differences as \((\omega_{mn})^{-3}\). This implies that contributions from higher conduction bands are small, since the energy differences, \(\omega_{mn}\), are large. This behaviour can be contrasted to the expression for the dielectric matrix that appears in an Adler and Wiser based approach. Here, the so-called "head" of the matrix possesses the same inverse cubic relationship with the eigenstate energy differences as does \(\chi^{ab}(G, G'; \omega)\). The "wings" and "body" of the dielectric matrix, however, vary as the inverse square and the inverse of the energy differences respectively. Clearly there is a greater dependency on the higher energy conduction states for these parts of the matrix as they are a much more slowly decreasing function of interband energy.

Finally, the convergence properties of the calculated \(\epsilon(0)\) values with respect to the number of \(k\) vectors in the IBZ have been found to be similar to those in calculations employing the Adler and Wiser approach \([31]\). We have used a special \(k\) point method to treat the sum in Eq. 7.92. Our results indicate that approximately 110 \(k\) points are needed to achieve reasonably converged values for the dielectric function.

We turn now to the frequency dependent linear response of silicon. At present we have have only calculated the below band gap response of the material. A future communication will address the above-band-gap behaviour. In Fig. 7.2 the dielectric function is plotted for energies below the optical gap. Note that no scissors correction has been applied for these calculations. There are two aspects of this plot which merit attention. The first is the size of the local field correction. LFE clearly play a significant role even in the case of linear optical response. We will address this difference below in the context of previous calculations. A second feature of the below gap response is the relatively featureless effect of the local field correction. This behaviour is similar to that in the results of Gavrilenko and Bechstedt \([31]\).

It is useful to compare the effect of including the local field correction as described by our formalism with calculations that are based on the Adler and Wiser approach. Our calculations indicate a local field correction of approximately 20%; by this we mean both the electrodynamic and exchange-correlation corrections. In the very recent calculation of Gavrilenko and Bechstedt they found that LFE are on the order
Figure 7.2: The linear dielectric function, $\varepsilon(\omega)$, for silicon at energies below the band gap: Without local field correction (solid line), including the local field correction (dotted line).

of 13%. Our results clearly predict a larger correction than found there; other Adler and Wiser based calculations also tend to give a local field correction smaller than that predicted here [23, 34, 47]. Given the significant differences between the new approach presented here and the traditional method of calculating LFE, it is difficult to assess the source of this difference. We believe that the above band gap calculations we have in progress for silicon will shed more light on this discrepancy.

An interesting physical picture of the response in the solid can be generated by investigating the electronic charge density. In Fig. 7.3 we have plotted for reference the unperturbed valence charge density for silicon on the $\mathbf{10\bar{1}}$ plane. The dashed-
Figure 7.3: The charge density contour plot for silicon in the 10\overline{1} plane. Contours are equally spaced and are in units of electrons per unit cell volume. The dashed-dotted line indicates the bond directions.

Dotted lines indicate the bond directions in the solid. The plot clearly indicates the congregation of valence electronic charge between the atoms.

Now consider the crystal under an incident field along the 111 direction (this corresponds to the upward direction in Fig. 7.3). We can calculate the perturbed valence charge density using Eqs. 7.6, 7.28, and 7.94. In Fig. 7.4 we plot this perturbed density on the 10\overline{1} plane (as in Fig. 7.3). As can be seen from the diagram, the perturbed density responds to the electric field as physically expected; the valence charge moves in a direction opposite to the field. This result, although intuitive, is fairly striking given that it is derived from a full band structure calculation using the formalism presented here. It is our hope that investigations of the frequency dependent microscopic response of the charge within a crystal will lead to a deeper
understanding of linear and nonlinear optical response. The formalism and approach we have described in this work should provide the means by which such calculations can be undertaken without excessive computational expense.

7.6 Conclusions

We have presented a new approach to calculating the local field corrected linear optical response of semiconductors. Our approach is based on calculating vector and tensor quantities rather than scalar quantities, as is done in usual calculations of optical
response that ignore local field corrections. By formulating the problem in this way, the connection to standard approaches of addressing optical response is made clear.

The formalism we have presented in this work eliminates some of the theoretical and computational limitations of the Adler and Wiser method. Moreover, our approach sets the foundation on which an explicitly nondivergent nonlinear susceptibility formalism that incorporates local field corrections can be developed. Our results indicate that the approach implemented here has improved convergence properties with respect to the number of conduction bands over previous calculations. This represents a significant increase in numerical efficiency. The convergence of our calculated linear response values with respect to the number of \( G \) vectors and \( k \) points is comparable to previous methods.

Our approach has been applied to the calculation of linear optical response in silicon, where we have used a pseudopotential plane wave approach for the determination of the electronic structure. It has been found that the effect of the local field correction is to reduce the zero frequency dielectric function by approximately 20%. This correction is larger than has been found in previous calculations that use the traditional Adler and Wiser approach. The below band gap frequency dependence of the dielectric function has also be investigated. Our results indicate a fairly constant local field correction in this energy range.
Bibliography


[38] With the definition of $p(x, t)$ as $j(x, t) = \partial p(x, t)/\partial t$, Eq. 7.71 does not hold for nonlinear response. The extra terms that appear are discussed in SG.


Chapter 8

Conclusions

This thesis has focussed on the calculation of linear and second-order optical properties of a variety of semiconductors. The goal of this work has been twofold. First, to make a set of state-of-the-art calculations for the linear dielectric function, the second harmonic generation susceptibility, and the linear electro-optic coefficient for common semiconductors of interest, and second, to extend a modern optical response formalism to account for the effect of local fields.

The results presented in Chapters 4, 5, and 6 have been based on an electronic structure calculation using the highly accurate FLAPW method. The actual calculations of the response functions have been carried out using a hybrid tetrahedron-random sampling method that has been developed for this work. The expressions for the susceptibilities are derived from the nondivergent formalism of Sipe and Ghahramani. Finally, we have used the scissors operator to account for quasi-particle corrections at an approximate level.

We have presented calculations for the optical properties of zinc-blende GaAs and GaP. Although these materials have been investigated previously, our calculation is the first to use the FLAPW method. Furthermore, the calculation of the linear electro-optic coefficient is the first based on a non-divergent formalism and a first-principles electronic structure method. These calculations have and will serve as a bench mark for future calculations that employ either simpler or more complex methods. We have found that our results for linear response are in good agreement with
other theoretical calculations, but despite the use of the scissors approximation, the agreement with experimental data for above band gap response remains only qualitative. Our results for the SHG susceptibility are somewhat inconclusive given the available experimental data; results for the LEO susceptibility, however, have been shown to agree very well with recent experimental investigations. We have argued on the basis of the equivalence of the SHG and LEO susceptibilities at zero frequency, that our results for the SHG response function are reasonable, and that further experimental investigations of this function are required. We have also succeeded in relating the structure in the linear and second-order response functions to the electronic band structure itself. This is useful in that it provides insight into the origins of the optical response within a material.

The wide band gap wurtzite materials, GaN and AlN, have also been examined. This work has constituted the first investigation of the nonlinear optical properties of these materials over a broad frequency range. We have found excellent agreement for linear response between our calculations and previous theoretical calculations. In contradiction to our results for GaAs and GaP, however, the agreement with experimental data for linear response at zero frequency is only rough; this could point to possible limitations of the scissors approximation in wide band gap materials. A comprehensive comparison between our theoretical work and experiment for second-order response still remains due to the scarcity of available experimental data. An interesting aspect of this work has been in a comparison of our results for SHG with those predicted by the bond charge model. We have found good agreement in the case of GaN, but very poor agreement for AlN; this has been attributed, at least in part, to the nature of the bonding found in the two materials. We have also examined various symmetry properties of the optical susceptibilities and have shown that these hold both numerically and theoretically. This provides a concrete check on both our numerical approach and the underlying susceptibility formalism.

Finally, as part of our calculation of the linear and second-order optical properties of semiconductors, we have investigated the II-VI materials in both the zinc-blende and wurtzite phases. Our goal here was to begin to elucidate the underlying physical
processes that account for both the similarities and differences in optical response between the two phases of a given material. This kind of analysis is facilitated by looking at the optical response in the same coordinate reference frame. We have found strong below band gap agreement for linear response between the two modifications of a material and more pronounced differences in the above band gap response. This has highlighted the localized k space nature of the optical response above the band gap. For second-order response, we used the SHG susceptibility as an investigative tool. Here strong differences were found below and above the band gap. This work has clearly demonstrated the greater sensitivity of the higher-order response functions to crystal structure, and is a first step in developing a better understanding of this relationship.

The second aspect of this thesis has been primarily analytic in nature. The aim has been to present a new approach and formalism for the calculation of linear response that incorporates local field corrections. This formalism, similar in development to that of Sipe and Ghahramani, has several advantages over previous approaches. Perhaps the most significant among these is that it is predicated on relating a polarization to an electric field rather than the charge density to a potential (as has been done previously). Our approach connects readily to existing susceptibility formalisms and as a result is a more transparent representation of these effects. We have implemented this new approach in the calculation of linear response in Silicon using a pseudopotential method. These results are in reasonable agreement with existing calculations that use the traditional approach. This work on local field corrected optical response has laid the foundation for further investigations of linear response in both the bulk and on the surface. Moreover, it will be the basis for future work on a second-order optical response formalism that incorporates local field effects, but at the same time maintains the non-divergent properties at zero frequency.
Appendix A

Susceptibility Conventions

In this appendix we establish our convention for the optical susceptibilities and the
definition of frequency components, and for clarity mention other conventions in com-
mon use and how the resulting expressions differ from ours. In Appendix B we detail
the connection between $\chi^{abc}(-\omega; \omega, 0)$ and the usual electro-optic coefficient $r^{abc}(\omega)$.
On these matters various usages exist in the literature, often never explicitly de-
dined. We here endeavor to make our usage, and the way we interpret the quoted
experimental results, as clear as possible.

It is useful to identify three conventions that can be used to introduce the fre-
quency components of the fields. In the first

$$E(t) = \sum_{\omega_i \geq 0} Re[E(\omega_i)e^{-i\omega_i t}]$$

$$= \frac{1}{2} \sum_{\omega_i \neq 0} E(\omega_i)e^{-i\omega_i t} + E(0), \quad (A.1)$$

where in the second expression generally $E(-\omega_i) \equiv E^*(\omega_i)$, and thus $E(0)$ is taken
to be real. Clearly in this first convention the actual DC field, $E_{DC}$, is just $E_{DC} =
E(0)$. The second convention corresponds to introducing $E(\omega_i)$ which are half those
appearing in the first convention; here the same electric field, $E(t)$, is written as

$$E(t) = 2 \sum_{\omega_i \geq 0} Re[E(\omega_i)e^{-i\omega_i t}]$$
In this convention the actual DC field is given by $E_{DC} = 2E(0)$. Finally, one can adopt a third convention that differs from the second only by the treatment of the field component at zero frequency:

$$E(t) = \sum_{\omega_i \neq 0} E(\omega_i) e^{-i\omega_i t} + E(0)$$

$$= \sum_{\omega_i} E(\omega_i) e^{-i\omega_i t}.$$  \hspace{1cm} (A.3)

Here $E_{DC} = E(0)$. We refer to the conventions identified by Eqs. A.1–A.3, and the corresponding expressions for $P(t)$, as conventions (1–3) respectively.

In all conventions the linear response is specified by

$$P^a(\omega) = \chi^{abc}_i(-\omega; \omega) E^b(\omega),$$ \hspace{1cm} (A.4)

where superscripts indicate Cartesian components and are to be summed over if repeated. For nonlinear response we restrict ourselves for the moment to convention (2); we generally follow the susceptibility notation of Boyd [1], adopting however a more common manner of indicating the frequency sum. That is, a nonlinear polarization component is related to the electric field by

$$P^a(\omega_\beta + \omega_\gamma) = \chi^{abc}(\omega_\beta - \omega_\gamma; \omega_\beta, \omega_\gamma) E^b(\omega_\beta) E^c(\omega_\gamma),$$ \hspace{1cm} (A.5)

where $\chi^{abc}$ indicates the second-order susceptibility. Cartesian components are again summed over if repeated, as are frequency components $\omega_\beta$ and $\omega_\gamma$, but only such that the sum ($\omega_\beta + \omega_\gamma$) is held fixed; the susceptibilities are taken to satisfy intrinsic permutation symmetry, $\chi^{abc}(\omega_\beta - \omega_\gamma; \omega_\beta, \omega_\gamma) = \chi^{acb}(\omega_\beta - \omega_\gamma; \omega_\gamma, \omega_\beta)$. We derive our perturbation expressions for $\chi^{abc}(\omega_\beta - \omega_\gamma; \omega_\gamma, \omega_\beta)$ within convention (2), but with all frequency components assumed nonzero; in this case conventions (2) and (3) are identical. A susceptibility involving a zero frequency component, such as
\[ \chi^{abc}(\omega; \omega, 0) \text{ is then obtained from, for example,} \]

\[ \chi^{abc}(\omega; \omega, 0) \equiv \lim_{\omega_0 \to 0} \chi^{abc}(\omega - \omega_0; \omega, \omega_0). \]  \hspace{1cm} (A.6)

Because of the way the frequency components enter the expression A.2 for \( E(t) \) in convention (2), we would physically expect, for example, that

\[ \lim_{\omega \to 0} \chi^{abc}(\omega; \omega, \omega) = \lim_{\omega_0 \to 0} \chi^{abc}(\omega; \omega, 0), \]  \hspace{1cm} (A.7)

and this is indeed found.

Turning now to expressions for the polarization in different nonlinear processes, in convention (2) we have, from Eq. A.5,

\begin{align*}
P^a(2\omega) &= \chi^{abc}(\omega; \omega, \omega) E^b(\omega) E^c(\omega), \\
P^a(\omega + \omega_o) &= 2\chi^{abc}(\omega - \omega_0; \omega, \omega_0) E^b(\omega) E^c(\omega_0), \\
P^a(\omega) &= 4\chi^{abc}(\omega; \omega, 0) E^b(\omega) E^c(0),
\end{align*}  \hspace{1cm} (A.8)

for second harmonic generation, frequency mixing (\( \omega \) with \( \omega_0 \)), and the linear electro-optic effect respectively. Here, and in such expressions below, only the Cartesian components are summed over if repeated. Note that the expressions we derive for \( \chi^{abc} \) can be used within convention (1) by simply replacing in Eq. A.8 \( E(\omega_i) \) by \( E(\omega_i)/2 \), and \( P(\omega_i) \) by \( P(\omega_i)/2 \). The effect is for the prefactors \( (1,2,4) \) appearing respectively in Eq. A.8 to go over to \( (1/2,1,2) \), the usual prefactors introduced in convention (1).

Within convention (3) we have, instead of Eq. A.8,

\begin{align*}
P^a(2\omega) &= \chi^{abc}(\omega; \omega, \omega) E^b(\omega) E^c(\omega), \\
P^a(\omega + \omega_o) &= 2\chi^{abc}(\omega - \omega_0; \omega, \omega_0) E^b(\omega) E^c(\omega_0), \\
P^a(\omega) &= 2\chi^{abc}(\omega; \omega, 0) E^b(\omega) E^c(0).
\end{align*}  \hspace{1cm} (A.9)
However, note that within either convention (2) or (3) we may write

\[ P^a(2\omega) = \chi^{abc}(-2\omega; \omega, \omega)E^b(\omega)E^c(\omega), \]
\[ P^a(\omega + \omega_o) = 2\chi^{abc}(-\omega - \omega_o; \omega, \omega_o)E^b(\omega)E^c(\omega_o), \]
\[ P^a(\omega) = 2\chi^{abc}(-\omega; \omega, 0)E^b(\omega)E^c_{DC}. \]  

That is, when the polarization is written in terms of the actual DC field \( E_{DC} \) the prefactor for the electro-optic effect is the same as the prefactor for frequency mixing. While it is often \( \chi^{abc}(-\omega - \omega_o; \omega, \omega_o) \) that is essentially measured, it is the \( \omega_o \to 0 \) limit of this expression, \( \chi^{abc}(-\omega; \omega, 0) \), that we calculate, with lattice coordinates fixed. For \( \hbar \omega_o \) much less than the electronic energy scales but much larger than phonon energies, these quantities can be expected to be essentially identical. Most often quoted is an experimental result for the electro-optic coefficient \( r^{abc}(\omega) \); in Appendix B we relate \( \chi^{abc}(-\omega; \omega, 0) \) to \( r^{abc}(\omega) \). Although we there assume a DC field \( E_{DC} \) rather than a mixing field \( E(\omega_o) \), the fact that the prefactors in the last two of Eqs. A.10 are identical means that, again for \( \hbar \omega_o \) much less than electronic energy scales but much larger than photon energies, \( E_{DC} \) can be simply replaced by the mixing amplitude \( E(\omega_o) \) in the expressions we derive.
Appendix B

LEO Susceptibility

In this appendix we detail the connection between $\chi^{abc}(\omega; \omega, 0)$ and the usual electro-optic coefficient $r^{abc}(\omega)$. Proceeding with the nonlinear polarization leads, within linear response, to a total displacement vector

$$D^a(\omega) = \sum_b \epsilon^{ab}(\omega)E^b(\omega) + 4\pi P_{NL}^a(\omega)$$

$$= \sum_b \epsilon^{ab}(\omega)E^b(\omega). \quad (B.1)$$

where $P_{NL}(\omega)$ is given by the last of Eqs. A.10, and the effective dielectric function

$$\epsilon^{ab}(\omega) = \epsilon^{ab}(\omega) + 8\pi \sum_c \chi^{abc}(\omega; \omega, 0)E^c_{DC}. \quad (B.2)$$

Defining the permeability tensor $\eta^{ab}(\omega)$ as the inverse of $\epsilon^{ab}(\omega)$,

$$\sum_b \eta^{ab}(\omega)\epsilon^{bc}(\omega) = \delta^{ac}. \quad (B.3)$$

and likewise for $\bar{\eta}^{ab}(\omega)$ and $\bar{\epsilon}^{ab}(\omega)$, the electro-optic coefficient relates the change in $\eta^{ab}$ to the $DC$ field,

$$\bar{\eta}^{ab}(\omega) = \eta^{ab}(\omega) + \sum_c r^{abc}(\omega)E^c_{DC}. \quad (B.4)$$
to first order. For crystal structures where $\epsilon^{ab}$ is diagonal, $\epsilon^{ab}(\omega) = \delta^{ab}\epsilon(\omega)$, Eqs. B.2 and B.4 yield

$$\varepsilon^{ab}(\omega) = \delta^{ab}\epsilon(\omega) + 8\pi \sum \chi^{abc}(-\omega;\omega,0)E_{DC}^c$$

$$\eta^{ab}(\omega) = \delta^{ab}\frac{1}{\epsilon(\omega)} + \sum r^{abc}(\omega)E_{DC}^c. \quad (B.5)$$

The second of Eqs. B.5 seems to be the universally accepted definition of $r^{abc}(\omega)$, in the “cgs” units we are using at the moment. Since $\eta^{ab}(\omega)$ is the inverse of $\epsilon^{ab}(\omega)$ and these equations only hold to lowest order in $E_{DC}$, we recover

$$\chi^{abc}(\omega;\omega,0) = -\frac{\epsilon^2(\omega)r^{abc}(\omega)}{8\pi}. \quad (B.6)$$

the desired relation. We note that the above expression is in “cgs” units. The equivalent expression in “SI” units is

$$\chi^{abc}(\omega;\omega,0) = -\frac{\epsilon^2(\omega)r^{abc}(\omega)}{2}. \quad (B.7)$$

the previously quoted result [2]; $\epsilon_r(\omega)$ is the relative dielectric constant. If we make the approximation $\epsilon_r(\omega) = n^2(\omega)$, then Eq. B.7 becomes

$$\chi^{abc}(\omega;\omega,0) = -\frac{n^4(\omega)r^{abc}(\omega)}{2}. \quad (B.8)$$

where $n(\omega)$ is the index of refraction.
Appendix C

Local Field

Given that $\rho(x, \omega) = -\nabla \cdot p(x, \omega)$, we have $\rho(G, \omega) = -iG \cdot p(G, \omega)$, and Eq. 7.27 follows directly from the work of Adler [3] and Wiser [4]. The goal of this appendix, however, is to connect this result with a much earlier, more general expression for the local field due to Lorentz [5]. The heuristic arguments due to Lorentz (see, e.g. [6]), can be put on a more rigorous foundation in a number of ways. We begin with the result of Sipe and Boyd [7]; they show that, for $a \ll \Delta \ll \lambda$, the microscopic electric field $e_M(x, t)$, assumed to be of the form

$$e_M(x, t) = e_M(x)e^{-i\omega t} + c.c$$  \hspace{1cm} (C.1)

can be written as

$$e_M(x) = E(x) + \frac{4\pi}{3} P(x) + \int \overrightarrow{T}(x - x') \cdot p(x')dx' - \frac{4\pi}{3} p(x)$$  \hspace{1cm} (C.2)

where

$$\overrightarrow{T}(x) = \overrightarrow{T}(x)c(r)$$  \hspace{1cm} (C.3)

and

$$\overrightarrow{T}(x) = \begin{cases} \frac{(3\hat{X} - \hat{U})/r^3}{r > \eta} \\ 0 \quad r < \eta \end{cases}$$  \hspace{1cm} (C.4)

$\eta \rightarrow 0$ after evaluating the integral.
is the static dipole-dipole coupling tensor, and $c(r)$ is a smooth decreasing function of $r$, $c(0) = 1$, $c(r) \to 0$ as $r \to \infty$, with a range on the order of $\Delta$; the expression for $c(r)$ is given by Sipe and Boyd [7]. Here $E(x)$ and $P(x)$ are the macroscopically averaged electric field and polarization respectively, and $p(x)$ is the microscopic polarization, $j(x, t) = \partial p(x, t) / \partial t$. The sum $E(x) + 4\pi P(x)$ is sometimes called the "cavity field". Note that $x$ represents a general position vector, and $r = |x|$ is its magnitude. In this appendix we show how the result (7.18, 7.25-7.27) follows directly from Eq. C.2 in the special case of a periodic structure. Here, for a "macroscopically infinitesimal" region of size of order $\Delta$ we take

$$p(x) = \sum_G p(G) e^{iG \cdot x},$$

where

$$p(G) = \frac{1}{\Omega_c} \int_{\text{cell}} p(x) e^{-iG \cdot x} \, dx,$$

and $\Omega_c$ is the volume of a unit cell. We can identify $p(0)$ with the macroscopic polarization ($P(x)$), taken as essentially uniform over the region.

We begin by considering the term involving $T^c$ in Eq. C.2. denoted as $e_d(x)$ in subsequent equations:

$$e_d(x) \equiv \int \nabla^0 \cdot p(x') \cdot c(|x - x'|) \, dx'$$

$$= -\int \nabla_x \left\{ \frac{(x - x')}{|x - x'|^3} \cdot p(x') \right\} c(|x - x'|) \, dx'.$$

We write this as

$$e_d(x) = -\int \nabla_x \left\{ \frac{(x - x')}{|x - x'|^3} \cdot p(x') \right\} c(|x - x'|) \, dx'$$

$$+ \int \frac{(x - x')}{|x - x'|^3} \cdot p(x') \nabla_x \left\{ c(|x - x'|) \right\} \, dx'.$$
We introduce a potential

$$ \phi_{E}(x) \equiv \int \frac{(x - x')}{|x - x'|^3} \cdot p(x') \, c(|x - x'|) \, dx' \quad . \tag{C.9} $$

and put

$$ \zeta(x) \equiv e_d(x) + \nabla \phi_{E}(x) \quad . \tag{C.10} $$

The determination of $\zeta(x)$ essentially reduces to calculating the difference between taking the gradient of an integrated function and integrating the gradient of the same function. For any differentiable scalar function $F(x, x')$ we can write [8]

$$ \nabla_x \int F(x, x') \, dV' = \int \nabla_x F(x, x') \, dV' - \int F(x, x') \hat{p} \, dS' \quad . \tag{C.11} $$

where $\hat{p}$ is a unit radial vector which points from the center of the integration region to a surface element. The second term on the right hand side of Eq. C.11 is simply a surface integral over the inner surface of radius $\eta$. Again note, as indicated after Eq. C.4, that $\eta \to 0$ after evaluating the integral. Using Eqs. C.8-C.11, the expression for $\zeta(x)$ is given by

$$ \zeta(x) = - \int_{\eta} \frac{(x - x')}{|x - x'|^4} \cdot p(x') \, (x' - x) \, c(|x - x'|) \, (x' - x) \, dS'$$

$$ + \int \frac{(x - x')}{|x - x'|^3} \cdot p(x') \nabla_x \{ c(|x - x'|) \} \, dx' \quad . \tag{C.12} $$

Consider the first term in the above equation. Note that since $\eta$ is very small, $c(|x - x'|) \approx 1$, and $p(x')$ is essentially constant over the surface. From this we deduce

$$ \int_{\eta} \frac{(x - x')}{|x - x'|^4} \cdot p(x') \, (x' - x) \, c(|x - x'|) \, (x' - x) \, dS' = - \frac{4\pi}{3} p(x) \tag{C.13} $$

and so

$$ \zeta(x) = \frac{4\pi}{3} p(x) + \int \frac{(x - x')}{|x - x'|^3} \cdot p(x') \nabla_x \{ c(|x - x'|) \} \, dx' \quad . \tag{C.14} $$

168
It will be convenient to evaluate both Eq. C.9 and the second term on the right hand side of Eq. C.14 using the Fourier expansion of the microscopic polarization. We evaluate the expression for $\phi_E(x)$ first. If we split the expression for $p(x)$ in (C.5) into a term for $G = 0$, and a sum over all $G \neq 0$, we can write Eq. C.9 as

$$\phi_E(x) = \int \frac{(x - x')}{|x - x'|^3} \cdot p(0) c(|x - x'|)dx' + \sum_{G \neq 0} \int \frac{(x - x')}{|x - x'|^3} \cdot p(G) e^{iG \cdot x} c(|x - x'|)dx'. $$  

(C.15)

Given that the cutoff function is spherically symmetric, it is straightforward to show that the first term integrates to zero. The second term requires more work due to the exponential factor that appears in the integrand. It can be shown that after integrating over the angular variables this term is given by

$$\phi_E(x) = -4\pi i \sum_{G \neq 0} \frac{e^{iG \cdot x}}{G^2} p(G) \cdot G \left(1 - \frac{1}{G} \int_0^\infty \frac{\sin(Gs)}{s} c'(s) ds\right), $$  

(C.16)

where $c'(s)$ is the derivative of the cutoff function with respect to its argument $s$. It is possible to simplify this expression further by considering the form of the integrand in the above equation. The cutoff function changes smoothly from a value of 1 to 0 for increasing $s$, with a range of the order of $\Delta \gg a$ (where $a$ is the lattice constant). The region over which this change occurs is also much greater than the lattice spacing. This implies that the derivative of the cutoff function, $c'(s)$, is non-zero for a region that is much greater than $a$. Now consider the trigonometric term appearing in the integrand, and note that its argument $Gs$ is on the order of $\Delta/a$ in the region for which the derivative of the cutoff function is non-negligible. Since $\Delta \gg a$, the trigonometric function varies rapidly with $s$ and so this term approximately integrates to zero. Consequently, we can write Eq. C.16 as

$$\phi_E(x) \simeq -4\pi i \sum_{G \neq 0} \frac{e^{iG \cdot x}}{G^2} p(G) \cdot G. $$  

(C.17)

Turning to the second term on the right hand side of Eq. C.14, we split it into two pieces, one for the $G = 0$ term, and another for the $G \neq 0$ terms associated with
the Fourier expansion of \( p(x) \). We write

\[
\int \frac{(x - x')}{|x - x'|^3} \cdot p(x') \nabla_x \{c(|x - x'|)| \, dx' = e_0(x) + e_G(x),
\]

(C.18)

where

\[
e_0(x) = \int \frac{(x - x')}{|x - x'|^3} \cdot p(0) \nabla_x \{c(|x - x'|)| \, dx'
\]

(C.19)

and

\[
e_G(x) = \sum_{G \neq 0} \int \frac{(x - x')}{|x - x'|^3} \cdot p(G) e^{iG \cdot x} \nabla_x \{c(|x - x'|)| \, dx'.
\]

(C.20)

The evaluation of the first of these two equations is straightforward, given that the cutoff function is spherically symmetric. After performing the integration, Eq. C.19 becomes

\[
e_0(x) = -\frac{4\pi}{3} p(0),
\]

(C.21)

a result similar to that in Eq. C.13. By noting that the crystal is periodic (see the discussion surrounding Eq. C.5), we use Eq. C.6 to rewrite the above equation as

\[
e_0(x) = -\frac{4\pi}{3} P(x),
\]

(C.22)

where we are taking the macroscopic polarization, \( P(x) \), to be the primitive unit cell average of the microscopic polarization.

The second term (Eq. C.20) involves a little more work. We will begin by rewriting it as

\[
e_G(x) = -\sum_{G \neq 0} e^{iG \cdot x} p(G) \cdot \int \frac{s}{s^2} e^{iG \cdot s} c'(s) ds,
\]

(C.23)

where \( s = (x' - x), s = |s|, \) and once again \( c'(s) \) is the derivative of the cutoff function.

The right hand side of this equation can be viewed as a vector dotted to a tensor (for each value of \( G \) in the sum). Consider only the integral part of Eq. C.23, which is itself a tensor. Each component of the tensor can be integrated separately for a given \( G \). For example, if we consider the \( s^x s^x \) component of this tensor we find that after
integrating over the angular variables it has the form

\[ s^x s^x \rightarrow 4\pi \int_0^\infty \frac{1}{(Gs)^3} (\sin(Gs) - G\cos(Gs)) c'(s) ds , \] (C.24)

where the notation on the left hand side simply indicates which components of the tensor we are considering. For such a term, we can invoke an argument similar to that following Eq. C.16 to show that it negligible. Evaluating the remaining eight components of the tensor in Eq. C.23, we find that the result is either zero, or has a form similar to Eq. C.24 in that it involves trigonometric functions with argument \( Gs \). This implies that all of the components of the tensor are either zero or small enough to be neglected, and so we take

\[ \mathbf{e}_G(x) \simeq 0 . \] (C.25)

Combining Eqs. C.14 with C.18, C.22, and C.25 we obtain

\[ \zeta(x) = \frac{4\pi}{3} p(x) - \frac{4\pi}{3} P(x) \] (C.26)

and using Eq. C.10, the expression for \( \mathbf{e}_d(x) \) is given by

\[ \mathbf{e}_d(x) = -\nabla \phi(x) + \frac{4\pi}{3} p(x) - \frac{4\pi}{3} P(x) . \] (C.27)

Finally, we consider the total expression for the microscopic Maxwell field. Recalling Eq. C.2 and the definition of \( \mathbf{e}_d(x) \) in Eq. C.7, we can write

\[ \mathbf{e}_M(x) = \mathbf{E}(x) - \nabla \phi(x) \] (C.28)

which, with Eq. C.17, is the desired result.
Bibliography


