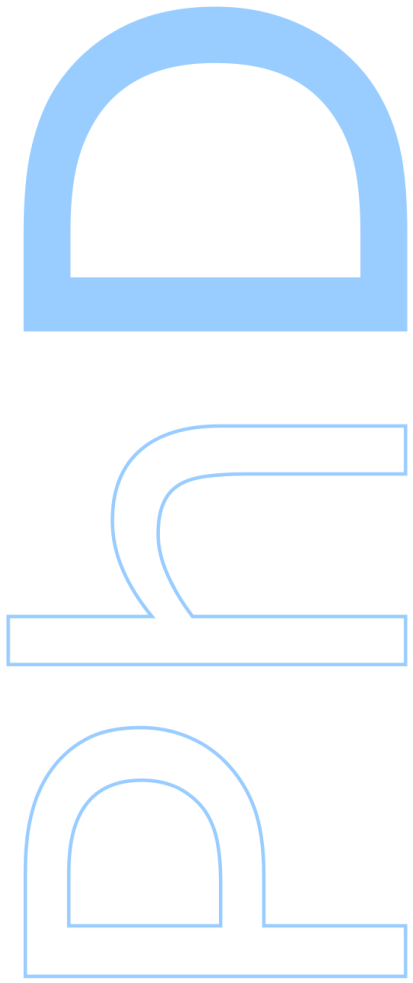
Nonlinear Optical Response of Two-dimensional Crystals

Gonçalo Bastos Ventura
Doutoramento em Física
Departamento de Física e Astronomia
2021

Orientador
João Manuel Borregana Lopes dos Santos, Professor Catedrático,
Departamento de Física e Astronomia
Coorientador
João Manuel Viana Parente Lopes, Professor Auxiliar Convidado,
Departamento de Física e Astronomia



"Every science in the world must possess one and the same passport, without which it is senseless; it must aspire to the truth."

Anton Chekhov, *On the Way*

"Physics would be dull and life most unfulfilling if all physical phenomena around us were linear. Fortunately, we are living in a nonlinear world. While linearization beautifies physics, nonlinearity provides excitement in physics."

Yuen-Ron Shen, *The Principles of Nonlinear Optics*

Acknowledgements

I would like to thank my supervisors, Professors João Lopes dos Santos and João Viana Lopes, for their continued support throughout these five years of work, their inexhaustible patience, and for the time and effort they themselves put into this thesis and its associated works. Though I would lie if I were to say that it has always been a pleasure to discuss things with both of you — whether about physics in general, our work, etc. — I can safely say that it has always been extremely valuable. That, I think, is what matters most. I have also had the benefit of working closely — for the most part of this PhD, on an almost daily basis — with a colleague as talented and as conscientious as Daniel Passos. I think that the diligence and perceptiveness these three people have has somewhat rubbed off on me; if it did not, it was not for their lack of trying.

A second thank you must go to everyone in the research group for Condensed Matter Physics (in both Porto and Minho!). While our work has come together only intermittently, the opportunities to learn from one another did not.

I must also thank my office mates. Though we found ourselves, not infrequently, discussing everything but our work or even physics, I must say that the time we spent together was more than worthwhile, even from the standpoint of what ultimately became this thesis. It helped me set my mind at ease during slumps, and, more importantly, it made me look forward to another day of work.

Finally, I would like to thank the Fundação para a Ciência e Tecnologia for making this possible by funding my research grant, PD/BD/140891/2018, as well as everyone involved in MAP-fis for their support.

FCT Fundação para a Ciência e a Tecnologia

Abstract

This thesis concerns the nonlinear optical response of crystals, specifically that of two-dimensional crystals. Its three central subjects are as follows: a general study of density-matrix calculations in crystals, and of questions involved in the physical equivalence of its two most common gauges (representations of the external electric field) — the length and velocity gauges; the application of a recently developed velocity-gauge formalism to the study of the nonlinear optical response of two two-dimensional materials — the plain graphene and gapped graphene monolayers; a study of the DC divergences in the optical response of a cold semiconductor, and of the role these divergences play in determining the internal structure of its nonlinear optical coefficients. In all three subjects, there is a coming together of something standard and something new.

The optical response of crystals has long been the subject of theoretical studies, and its standard method of calculation, which is based on the density matrix and a length-gauge description of the external electric field, is now over twenty-five years old. The work presented here is, in part, a reformulation of this description, in which the perturbation is expressed in terms of a newly introduced object, the covariant derivative. But the relevance of the covariant derivative goes beyond this reformulation, as it is the key to explaining problems concerning the equivalence between different gauges. Here, it is shown how the equivalence between gauges can be preserved, and why it breaks under certain approximations.

The optical response of two-dimensional materials and its development as a subject on its own are, of course, the result of more recent developments in material science. The optical response of graphene has been the central topic of research, and several theoretical calculations — focused on infrared and low visible part of the spectrum — have been presented over the past ten years. More recently, these have been extended to the high-frequency response of this material, and require, as it will be argued here, the application of a velocity-gauge description of the density matrix. This is followed by a discussion on the use of tight-binding Hamiltonians in calculations of nonlinear optical responses.

Finally, the work presented here is also dedicated to the study of the direct current divergences in the optical response of a cold semiconductor. As before, it starts with a review of two well-known results, namely, the derivation of the second order injection and shift currents. This serves as the basis for a new derivation of the jerk current — which is associated with a new physical interpretation of this effect — as well as for the study of consequences brought about by the existence of the injection current. By means of a general argument, it is shown that a relation between real

parts of the second order conductivity tensor of a cold semiconductor exists. When the injection current is absent, this relation is reduced to an equality involving three different nonlinear optical effects — the sum-frequency generation, the difference-frequency generation and the shift current.
Keywords: nonlinear optics; two-dimensional materials

Resumo

Esta tese diz respeito à resposta óptica não linear de cristais, especificamente de cristais bidimensionais. Os seus três temas centrais são os seguintes: um estudo geral de cálculos de matriz densidade em cristais e das questões associadas à equivalência física das suas duas *gauges* (representações do campo eléctrico externo) mais comuns — as *gauges* apelidadas de *length* e *velocity*; a aplicação de um formalismo para a *velocity gauge*, recentemente derivado, ao estudo da resposta óptica não linear de dois materiais bidimensionais — as monocamadas de grafeno simples e grafeno com um *gap*; um estudo das divergências de corrente contínua na resposta óptica de um semiconductor frio e o papel que estas desempenham no escrita da estrutura interna dos coeficientes ópticos não lineares deste tipo de material. Nestes três temas dá-se a intersecção entre algo que é bem estabelecido com algo que é novo.

A resposta óptica de cristais é, desde há muito tempo, um assunto estudado do ponto de vista teórico; o seu método padrão de cálculo, baseado na matriz densidade e na descrição do campo eléctrico externo na *length gauge*, foi derivado há mais de vinte e cinco anos. O trabalho aqui apresentado é, em parte, uma reformulação desta descrição, em que a perturbação passa a ser expressa em termos de um objecto recentemente derivado, a derivada covariante. Mas a relevância da derivada covariante vai para além desta reformulação, uma vez que é a chave para a compreensão de problemas associados à equivalência entre *gauges* diferentes. Aqui, mostra-se como esta equivalência entre *gauges* pode ser preservada e como esta quebra sob certas aproximações.

A resposta óptica de cristais bidimensionais e o seu desenvolvimento como um assunto em si mesmo são, por sua vez, o resultado de desenvolvimentos recentes na área de ciência de materiais. A resposta óptica do grafeno tem sido o tópico central de pesquisa e vários cálculos teóricos — focados na parte infravermelha e visível (baixa) do espectro — foram apresentados nos últimos dez anos. Mais recentemente, estes foram extendidos para a resposta de frequências altas deste material, cálculos este que requerem, como será aqui argumentado, o uso de um formalismo de matriz densidade para a *velocity gauge*. A este argumento segue-se uma discussão sobre o uso de Hamiltonianos de tight-binding em cálculos de respostas ópticas não lineares.

Por fim, este trabalho é também dedicado ao estudo de divergências de corrente contínua na resposta óptica de um semiconductor frio. Tal como antes, este estudo começa com uma revisão de dois resultados conhecidos, a saber, a derivação de duas correntes de segunda ordem — as correntes de injecção e de *shift*. Esta revisão serve como base para uma nova derivação da corrente de jerk — e que está também associada a uma nova interpretação deste efeito — assim como para o estudo de

consequências que decorrem da existência desta corrente de injecção. Usando um argumento geral, mostra-se a existência de uma relação entre as partes reais da conductividade de segunda ordem de um semiconductor frio. No caso da ausência da corrente de injecção, esta relação reduz-se a uma igualdade entre três efeitos ópticos não lineares — a *sum-frequency generation*, a *difference-frequency generation* e a *shift current*.
Palavras-chave: óptica não linear; materiais bidimensionais

Contents

List of Figures

List of Tables

List of Acronyms

- 1PP: One-photon Processes;
- DM: Density Matrix;
- FBZ: First Brillouin Zone;
- hBN: Hexagonal Boron Nitride;
- HHG: High Harmonic Generation;
- IPA: Independent Particle Approximation;
- NLO: Nonlinear Optical;
- OKE: Optical Kerr Effect;
- PG/GG: Plain/Gapped Graphene;
- SHG: Second Harmonic Generation;
- THG: Third Harmonic Generation;
- TMDs: Transition Metal Dichalcogenides;

Publications by the Author

In this thesis:

- *Gauge covariances and nonlinear optical responses*, **G. B. Ventura**, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres, *Phys. Rev. B* 96 (2017), 035431;

- *A study of the nonlinear optical response of the plain graphene and gapped graphene monolayers beyond the Dirac approximation*, **G. B. Ventura**, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, *J. Phys.: Condens. Matter* **32** 185701;

- *Second order divergence in the third order DC response of a cold semiconductor*, **G. B. Ventura**, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, arXiv:2004.01919;

- *Comment on "Jerk Current: A Novel Bulk Photovoltaic Effect"*, **G. B. Ventura**, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, *Phys. Rev. Lett.* 126 (2021), 259701;

Not in this thesis (and as a second author):

- *Nonlinear optical responses of crystalline systems: Results from a velocity gauge analysis*, D. J. Passos, **G. B. Ventura**, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres, *Phys. Rev. B* 97 (2018), 235446;

- *Optical absorption of single-layer hexagonal boron nitride in the ultraviolet*, J. C. G. Henriques, **G. B. Ventura**, C. D. M. Fernandes, and N. M. R. Peres, *J. Phys.: Condens. Matter* **32**, 025304;

- *Nonlinear optical conductivity of a two-band crystal I*, D. J. Passos, **G. B. Ventura**, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, arXiv:2102.07629;

Chapter 1

Introduction

Nonlinear optics concerns the interaction of light with matter in physical systems whose response — usually expressed either in terms of the polarization, $\mathbf{P}$, or the current, $\mathbf{J}$ — cannot be adequately described using a linear approximation, but that instead must be treated either as an expansion in powers of the electric, $\mathbf{E}$, and magnetic, $\mathbf{B}$, fields, or, in the case in which no such expansion is possible, in terms of general constitutive relations, $\mathbf{P}(\mathbf{E}, \mathbf{B})$ or $\mathbf{J}(\mathbf{E}, \mathbf{B})$. While such descriptions are generally more complicated — sometimes even downright intractable — than their linear counterpart, they are a window into a world that even if it is not as simple and beautiful, can be considered to be more exciting [11]. A description of some nonlinear effects in second order in the external electric field will probably be enough to convince the reader that this is the case, even if the language might seem somewhat out of place.[1] A monochromatic light pulse travelling through a nonlinear (noncentrosymmetric) material will, in second order in the electric field, generate light of *twice* the original frequency (second harmonic generation), a static electric field across the material that can be optically controlled (optical rectification), and, under certain conditions, a ballistic current that can, again, be controlled by an optical field (injection current). If, in addition to the light pulse, a static electric field is applied to the material, it is seen that some of its properties, e.g., the refractive index, can themselves be manipulated so as to make the light pulse propagate differently according to different intensities and directions of the static field (Pockels effect). These effects, already interesting on their own, account only for a small portion of the wealth of physical effects that can be studied in nonlinear optics. Considering this diversity, and the intrinsic complexity involved in the aforementioned descriptions, it is perhaps natural that problems in nonlinear optics are (somewhat) divided into two categories: those that are primarily concerned with studying the propagation of a light pulse in a given material (problems in electrodynamics); those that, on the

[1] We direct the reader interested in a detailed description of the different nonlinear optical processes (involved in the response to an electric field) to the textbooks of Y. R. Shen [11], R. W. Boyd [12], and G. New [13], among others [14–16]. Y. R. Shen's treatment of this subject is particularly thorough (and lengthy); he also discusses several magnetooptical effects. Magnetooptics is something of a subject on its own, which we will not address, and only mention in passing, in this thesis.

other hand, are primarily concerned with quantifying a given material's nonlinear optical (NLO) response by studying how its constitutive elements respond to that pulse (problems concerning charged particles in classical or quantum mechanics).[2] There is, however, no real boundary between the two, nor can there be one: a light pulse induces a nonlinear response in the material in which it travels, and this nonlinear response must then modify the form of the pulse itself, as prescribed by Maxwell's equations [11]. It is primarily due to matters of expediency that these two issues are treated somewhat separately, although, as we will soon discuss, this separation is rather adequate when we are discussing the nonlinear optical response of two-dimensional materials. This thesis is dedicated to this second category of problems.

The description of a material's nonlinear optical response follows, as we have just mentioned, from the study of how its constitutive elements respond to an external electric field. Considering the relative inertia of atomic nuclei, the role of constitutive elements usually falls solely upon the electrons (sometimes only some of the electrons) in the system being considered. Yet, despite this general similarity, electrons in a gas, a metal or a semiconductor respond very differently to the same perturbation. In gases, electrons are considered to be bound to a specific nucleus, and the nonlinear optical response of the gas as a whole is described as the sum of independent contributions of individual atoms; the perturbative treatment of this type of system can thus be formulated in terms of the density matrix (DM) associated with the electronic eigenstates of the individual atom, with the main problem being one of simple bookkeeping (in the sense of accounting for the contributions from all the relevant states) [11, 12]. As we will see in detail later on, there are two immediate descriptions of the interaction of electrons with an external electric field that is considered to be uniform — the length gauge, where $\hat{V}(t) \propto \mathbf{r}$, and the velocity gauge, where $\hat{V}(t) \propto \mathbf{v}$; for electrons in a gas, a density-matrix calculation can be easily formulated in both these gauges. In crystals, which are the focus of this thesis, the description at hand again depends on the properties of the system's electrons (which can no longer be treated as being independently associated with a single atom), but also on the properties of the external electric field. The low-frequency response of a metal can be described in semiclassical terms by means of the Boltzmann equation, in which the *intraband* motion of the electrons is taken into account and, in the relaxation-time approximation, scattering is treated phenomenologically [17]. This treatment, however, does not work in the case of a cold semiconductor, i.e., $k_B T \ll \Delta$ for Δ the electronic band gap, since its bands are considered to be either entirely filled or empty, meaning that its response, according to this description, should be trivial when it is not; the *interband* motion of electrons must then be taken into account (this is also true for the response of metals at high enough frequencies). Doing so requires a quantum treatment

[2] As we have noted above, these can be roughly divided into perturbative and nonperturbative descriptions. The former are usually formulated in terms of nonlinear optical coefficients (susceptibilities, χ, or conductivities, σ) that describe all optical processes at a given order. As an example, the second order susceptibility, $\chi(\omega_1, \omega_2)$, describes all of the aforementioned effects: the second harmonic generation, $\chi(\omega, \omega)$, the optical rectification and the injection current, $\chi(\omega, -\omega)$, and the Pockels effect, $\chi(\omega, 0)$. The latter are formulated directly in terms of the polarization or current (or their expectation values, if that is the case).

of interband transitions, which, as in the case of electrons in a gas, can be formulated in terms of a density matrix associated with the eigenstates of the unperturbed system, i.e., Bloch states.[3] The corresponding functions, unlike those of the electrons in a gas which we have described above, are spread out (more or less evenly) throughout the entire crystal, very much like plane waves in a noninteracting free electron gas — the difference being that Bloch states also carry a modulation that is unit-cell periodic. This means that, as with plane waves, the position operator has to be treated with some care [18, 19], and so does the length gauge [20–23]. As for the velocity gauge, while the aforementioned perturbation can be easily expressed in the Bloch basis, there is a lengthy record of tentative experimentation [24–26] (and, more recently, a comprehensive derivation) that shows that the truncation of the corresponding band space renders this description unphysical. Nonetheless, both these problems can be solved, and a density-matrix description can be formulated in both gauges [2, 3]; these descriptions are one of the subjects of this thesis. Alternatively, the study of these quantum interband transitions can be formulated in terms of the electronic populations derived by Fermi's golden rule arguments [27, 28];[4] while significantly less sophisticated than calculations that use the density matrix formalism, these arguments provide us with relatively simple derivations of some effects, and can be more easily interpreted than their complex counterparts. Over the past four decades, they have been used to good effect to describe, for example, the second [28, 43] and third [28, 44, 45] order injection currents in three-dimensional semiconductors, the intensity dependent refractive index of three-dimensional semiconductors [46, 47] and graphene [48], and, more recently, the charge (and in some cases spin) injection currents in topological insulators [49] and in two-dimensional materials such as graphene and the transition metal dichalcogenides [50, 51]. It is also important to add that these current injection effects, as well as any effect that is associated with current injection in general, can also be derived by considering the *physical* divergences of the nonlinear optical coefficients [52–54].

Despite all its useful applications, the density matrix description of noninteracting electrons is not without certain drawbacks. First, although it can be used in calculations where the electric field is *not* considered to be uniform, e.g., $\mathbf{E}(\mathbf{r},t) \sim \exp(i\mathbf{q}\cdot\mathbf{r} - i\omega t) + \text{c.c.}$, a finite $\mathbf{q}$ treatment of the DM (such as that in ref.[55]) is usually much more cumbersome than the (already cumbersome) DM calculations for a uniform electric field. Even calculations in three-dimensional materials,

[3] However, three differences must be noted: the discrete sum over atomic states is replaced by a discrete sum over band indexes *and* an integral over the first Brillouin zone (FBZ); the standard criterion used to evaluate the validity of the perturbation theory, $\langle\psi_1|\,\hat{V}\,|\psi_2\rangle \ll (\epsilon_1 - \epsilon_2)$, does not seem to hold in the case where contributions from individual states must be integrated over a domain. One can, albeit *a posteriori*, compare the response at different orders, and from there determine if one's perturbation theory is valid; one can have transitions between states whose energy difference is arbitrarily small, since initial and final states can be both in the same band. Intraband transitions play as much a role in the NLO of a cold semiconductor as interband transitions do in the NLO response of a metal. A strict separation of the two processes, especially in a DM description, is therefore impossible [15].

[4] We do not wish to confuse the reader with this either/or. There are other techniques that can be used in the study of the NLO response, e.g., a number of *ab initio* treatments of perturbation theory have also been proposed over the past thirty years [29–39]; more recently, several diagrammatic treatments have been presented that can describe the response in materials that, for whatever reason (disorder, localized impurities, ...), are not periodic or that cannot be treated in the independent particle approximation (IPA) [6, 40–42].

where the electric field is necessarily different in different parts of the material, i.e., at different wavefronts, are usually performed in the long-wavelength limit, as is the case of bulk semiconductors [22, 23]. Secondly, effects associated with phonon and impurity scattering are usually introduced phenomenologically in the equations of motion for electrons in a clean crystal; in some calculations, this procedure involves only the intraband density matrix elements — in which case, they describe a relaxation term driving the system towards its equilibrium distribution — in others, it involves both the intraband and interband density matrix elements — in which case, the nondiagonal elements describe decoherence effects [15]. Considering that a new *sub-gap* photocurrent [56], which can only exist for certain combinations of these phenomenological parameters, has been recently proposed, a careful evaluation of these issues is certainly pertinent. There are also two additional limitations to the aforementioned perturbative treatment of the DM: it cannot easily describe responses to ultrashort light pulses in the time domain, and it cannot describe responses at very high pulse intensities. These difficulties can, however, be surpassed when we consider the full density matrix, although its practical application is very different from that of a perturbative treatment, even if the central concepts are very much the same.[5]

There is one final aspect of these calculations that must be addressed, which concerns the Coulomb interaction between charged particles in a crystal (specifically a semiconductor or an insulator) and the effect that it has on its optical response. When light passes through a semiconductor, it can pump electrons from the valence into the conduction band, leaving a positively charged hole behind. The attractive Coulomb interaction between the two particles can give rise to a bound state — an exciton — which appears in the optical response as a sharp resonance slightly *below* the electronic band gap of the semiconductor [12, 14].[6] While this relatively simple physical picture suggests that the quantitative description is correspondingly simple, that is hardly the case. The direct generalization of the equations of motion of the free-carrier DM gives us the semiconductor Bloch equations [15]; this set of equations, which is coupled in both the matrix elements of the DM and $\mathbf{k}$, and involves quadratic terms in the DM elements — due to the Coulomb interaction — is much more difficult to solve than the free-carrier one. There are, however, simplified descriptions of this problem that are much more tractable: the Wannier equation, which can be derived under the assumption that the Coulomb potential varies little within a unit cell [15], describes a two-body problem of oppositely charged particles with a reduced mass of electron-hole pair, and is equivalent to the Schrödinger equation describing the hydrogen atom. Unsurprisingly, the optical response of

[5] While it is, in principle, possible to calculate the response to ultrashort pulses directly from the nonlinear optical coefficients, e.g., Eq.(4.3), it requires the knowledge of the nonlinear response function over a considerable range of frequencies — sharp pulses in the time domain correspond to wide pulses in the frequency one, and vice-versa. For very high intensities, there is no other sensible alternative to a full DM calculation, i.e., one based on the numerical integration of the DM equations of motion, e.g., Eqs.(3.92) and (3.93). Therefore, this description requires both an integration (of the equations of motion) over the time domain, and an integration (of the polarization or current densities) over the FBZ.

[6] In fact, it is a series of resonances, very much like the Rydberg series of the hydrogen atom. In addition, while binding energies are relatively small in standard 3D semiconductors, $\epsilon_{1s} \approx 10\,\text{meV}$ [15], for a two-dimensional insulator such as the hexagonal boron nitride, they can be of the order of the electron-volt, $\epsilon_{1s} \approx 2\,\text{eV}$ [57].

excitons described by the Wannier equation is similar to that of a hydrogen atom, with a discrete set of states that mimics the Rydberg series. Furthermore, the Coulomb interaction can also change — quite significantly — the form of the optical response for energies above the electronic band gap (which is associated with a continuum of states), as the result of the so-called Coulomb enhancement [14, 15]. The Coulomb interaction is thus something one should keep in mind when studying the optical response of crystal electrons, and this is especially so in the case of two-dimensional crystals.[7]

Two-dimensional Materials and Their NLO Responses

The study of the nonlinear optical response of a two-dimensional sheet of material is, from the standpoint of electrodynamics, the study of a pair of boundary conditions; in two dimensions, source terms generated by an external electric field do not appear directly in the Maxwell's equations, but instead appear in the conditions which ensure that the discontinuity in the magnetic field is related to the current flowing in that sheet, and, since there are no free charges, that the electric field must be continuous. From these two boundary conditions, one can easily derive the coefficients describing the optical phenomena in the sheet of material — the reflection, transmission and absorption coefficients — at the frequency of the incident monochromatic field [58–60]; for processes like the third harmonic generation, $\sigma(\omega, \omega, \omega)$, one needs to think only of the two-dimensional material as a plane in which a current density oscillates at a certain frequency, i.e., 3ω, and from there determine the emitted electric and magnetic fields. This is, therefore, much more tractable than the description of the nonlinear optical response of a three-dimensional material (again from the standpoint of electrodynamics), in which, even in the simplest cases, one must solve a set of coupled-wave equations [11–13]. That the nonlinear optical response of two-dimensional materials has been studied with such interest is not, however, a consequence of the fact that its electrodynamics are simpler than those of their three-dimensional counterparts (though it has probably helped). The optical properties of graphene [61], the first two-dimensional material to be extensively studied, were early recognized as exceptional [62]: its universal conductivity and the step-function shape of its (linear) absorption coefficient, a result of interband transitions of carriers with a linear dispersion relation, were readily identified [63–65]; as for the nonlinear response, its first major prediction also came shortly after this material had been isolated for the first time. In ref.[66, 67], S. A. Mikhailov argued that, because of their linear dispersion relation, the intraband response of carriers in graphene should be intrinsically nonlinear, and proceeded to study this portion of the response,

[7] Note, however, that in this thesis we *always* work in the IPA. While a thorough description of the optical response would certainly have to account for the effect of the Coulomb interaction — we will return to this subject later on — there are problems that can be described in the IPA, such as the optical response of materials like doped graphene, where the presence of additional carriers means that the Coulomb interaction is more effectively screened. Furthermore, when a subject is still relatively unexplored, as is the case with the optical response in many two-dimensional materials or with the study of new nonlinear effects, it is sensible to start with the simplest possible description.

in the time domain, using standard Boltzmann theory. These preliminary studies, as we have already noted, could not fully describe the NLO response of the graphene monolayer. The study of interband transitions, and of the rich interplay between intraband and interband processes, required density-matrix calculations that were somewhat long in the making [59, 68, 69]; these revealed, surprisingly, that the third order conductivity of graphene, $\sigma(\omega_1, \omega_2, \omega_2)$, can be modelled in terms of the constituent elements of the linear one. It was also during this period, but especially after it, that the scope of the studies greatly widened: effects associated with a momentum-dependent field (in the case of nonnormal incidence) were studied in ref.[55, 70]; the existence of an optical bistability was the subject of ref.[71], and is also discussed, among other nonlinear effects (such as the high harmonic generation (HHG)) in ref.[72]; the saturable absorption, and in particular, the existence of anomalous saturation, was studied in ref.[59, 73] and ref.[74]; the nonlinear refractive index and the theoretical description of its measurements (which usually use two different techniques, the Z-scan and the optical heterodyne detection scheme) [75–77] were the subject of ref.[68, 78, 79] and ref.[80]; the perturbative NLO response, specifically the shift current, of a special form of graphene — the twisted bilayer — was recently presented in ref.[81]. These are, of course, but a few of the subjects studied, and a complete review of the literature on the NLO responses of two-dimensional materials is beyond the scope of this introduction and this thesis. It must, however, be emphasized that several time-domain descriptions of the response of graphene, ref.[73, 82–88], have gone beyond the aforementioned Boltzmann treatment of ref.[66, 67], and have been used to study a variety of subjects ranging from the detection of Landau–Zener–Stückelberg interference in graphene [82], to the relative importance of interband and intraband processes in HHG in the monolayer [84], and to the study of HHG in twisted bilayer graphene [88]. Furthermore, a recent study of the nonlinear optical response of undoped graphene [89], which takes into account the role of the Coulomb interaction, has shown that it is enhanced with respect to that of calculations in the IPA; a re-evaluation of the response of doped graphene, so as to see what "role the Coulomb interaction plays", is certainly in order.

The introduction of the Coulomb interaction in the description of the optical response of graphene has, therefore, been somewhat of an afterthought (at least in the case of the NLO response), which is in contrast with what has been done in the study of the optical response of 2D materials like the transition metal dichalcogenides (TMDs) and the hexagonal boron nitride [7, 90–102].[8] It has long been known that the binding energies of excitons confined in two dimensions are larger than those in three dimensions, so that their resonances are further from the gap (and thus

[8] Note that these works used different formulations of the excitonic problem. In ref.[90–92], the calculations were performed using *ab initio* methods. In ref.[7, 93, 94], the authors used a Schrödinger equation with a Keldysh potential, i.e. the Wannier equation, while in ref.[95, 97, 98], they used directly the Bethe-Salpeter equation. (The Wannier equation can be derived from the Bethe-Salpeter equation by first expanding the different terms in the equation around the band minima, and then taking their Fourier transform, so as to obtain an equation (in real space) for the excitonic eigenfunctions [7, 103].) In ref.[100, 104], the authors used the so-called Wannier gauge [105]. In addition, we note that two large-scale studies of the NLO response of three- and two-dimensional semiconductors (in the IPA) have been recently published [106, 107].

more pronounced) [14, 15]. In this case, the Coulomb interaction is not screened in every direction, meaning that the effective attraction between oppositely charged particles is stronger, which results in larger excitonic binding energies. As we have noted, the description of these responses can be more complicated than that of crystal electrons in the IPA, especially in the nonlinear optical response [57, 108], although there are certain treatments — in which the excitons are treated as the eigenstates of the unperturbed Hamiltonian, meaning that the perturbation theory in the external electric field is expressed directly in terms of these states — that are, at least at face value, simpler [15, 99, 103]. The excitonic energies and eigenfunctions are, however, still needed for these calculations, and these are determined using procedures that are very different from those involved in the calculation of NLO responses in the IPA. In order to maintain the scope of this work somewhat consistent, we will, for the rest of this thesis, stick to this approximation. As will become clear in the outline of this thesis, we are not primarily interested in the NLO response of two-dimensional materials, though these certainly play an important part in our work. Instead, our focus is set on the issues concerning two possible descriptions (the length and velocity gauges) that use the density matrix formalism in the IPA, on how the latter gauge can be practically implemented in calculations based on tight-binding models, and, in addition, on certain divergent effects in the DC response of a cold semiconductor. While these different elements can be used in the study of the aforementioned responses (and we do so here, in our study of the second and third order response of the gapped graphene and plain graphene monolayers), there is nothing about them that is *intrinsically* two-dimensional.

Divergences and Weyl Semimetals

As noted above, the nonlinear optical coefficients are divergent for certain frequency combinations (and under certain conditions); these divergences describe physical phenomena, like the injection and shift currents — which are associated with the optical pumping of electrons from the valence into the conduction band at a finite rate — and are not, as could be assumed, instances where perturbation theory is no longer applicable. This subject has been active recently, with several works focusing either on the study of well-understood effects in new materials, which is the case of the works cited above [45, 49–51, 54], or, alternatively, on the isolation of new effects like the jerk current — which can be described as the current resulting from the dragging, by a static electric field, of the carriers generated by optical pumping [53, 109–111]; this effect is one of the subjects of this thesis. With the exception of the shift current, the divergences described in this thesis are (either directly or indirectly) associated with the generation of a current that increases linearly (injection current) or quadratically (jerk current) with time. This does not mean that this relation between divergences and (DC) currents holds in general: the optical Kerr effect, for example, carries a divergence in the one-photon process. One thing that all divergences have in common, however, is that they all play a role in determining the internal structure of the response functions, by modifying

the Kramers-Krönig relations [46–48, 112, 113]. The attention dedicated to these divergences is all the more relevant considering the recent developments in Weyl semimetals and their NLO responses.

Weyl semimetals — three-dimensional materials whose gapless electronic excitations are protected by topology and symmetry [114] — have been the subject of a variety of works, [41, 54, 56, 114–124]; like the aforementioned divergences, several new NLO effects, or new interpretations of old ones, have been proposed [115, 116, 118, 121, 122], most notably the quantization of the difference between circular photocurrents (injection current) of opposite polarizations [116], or the circular difference-frequency generation, [121], which also involves a quantized response (that is, supposedly, simpler to measure). Furthermore, since NLO effects are said to be a direct probe of a material's topological character [41, 115, 122, 123], this interest in the physics of Weyl semimetals is bringing additional attention to the NLO effects themselves. While the subjects in this thesis do not concern this type of materials, and, in fact, they sometimes presuppose the use of symmetries that can be absent in Weyl semimetals, there is no fundamental gap between the treatment of circular photocurrents in ref.[116], and the usual treatment of the divergences of the nonlinear optical coefficients, or the treatment of time-domain responses of semimetals in ref.[120], and those that have been previously used to study the time-domain response of graphene.[9]

Outline and Structure

This thesis is divided into four main chapters and a conclusion. The former can be briefly summarized as follows: Chapter 2 concerns some general properties of linear and nonlinear optical coefficients; Chapter 3 is dedicated to the length- and velocity-gauge descriptions of the density matrix formalism, and to the relations that ensure the equivalence between the two (as well as the conditions in which these hold true); Chapter 4 concerns the application of the description from ref.[3] to the calculation of the nonlinear optical response of the monolayers of plain and gapped graphene. It also includes a discussion of aspects associated with the use of tight-binding Hamiltonians; Chapter 5 is dedicated to the DC divergences in the optical response of a cold semiconductor and to the role these play in determining the structure of nonlinear optical coefficients.

[9] While we are on the subject of discussing possible applications of our work, it is relevant to point out to magnetooptics [125–132]. Although much of what is done in this area is outside the scope of this thesis, i.e., the optical response of crystals subjected to strong magnetic fields [129, 130] (in which case the lattice periodicity of the plain crystal no longer holds true), there instances where magnetic effects, such as those associated with magnetic dopants or with the proximity to a ferromagnetic insulator, can be treated as changes in a crystal's band structure [128]; the calculation of NLO responses of these materials is no more difficult than that of regular crystals, the calculation of the nonlinear equivalent of, e.g., the Kerr rotation, should be rather straightforward.

Chapter 2

General Properties of the Susceptibility Tensor

It is common for textbooks on nonlinear optics to start their introductory chapter with a general description of a physical quantity that describes the system's response to an external electric field, usually, the polarisation, in terms of a power series of the external field [11–13, 15] (while assuming, of course, that such an expansion is possible). This approach, although it might not seem essential for the presentation of the formalism that we have developed (one of the subjects of the next chapter) or to the description of the nonlinear optical response of graphene (one of the subjects of the chapter after next), has the advantage that it can be used to derive, in rather simple and general terms, certain properties that should be retained by, or imposed on, the response functions of a general system: the reality condition, the Kramers-Kronig relations, intrinsic permutation symmetry, etc. These, as we will see, can be key elements in the analysis of both the physics and the expressions of any nonlinear optical response. Furthermore, this general description makes clear some of the approximations employed in the course of our studies of nonlinear optical responses. These two subjects are the focus of the first of the two sections of this chapter. The second section is dedicated to a brief overview of the conditions imposed on the susceptibility tensor of a crystal by its point group symmetries. Readers that are interested in a careful and comprehensive approach to this subject are directed to the references [133, 134].

2.1 General Properties of the Susceptibility Tensor

The susceptibility description of the optical response of a system — where the polarisation is expanded in different orders of the electric field, $\mathbf{E}(\mathbf{r}, t)$, each associated with a given susceptibility — invariably assumes that such an expansion in powers of $\mathbf{E}(\mathbf{r}, t)$ is both possible and useful, i.e., it considers that that series expansion can be reasonably truncated at a given (low) order of the field, usually at second or third order. The validity of this perturbation theory is usually transcribed into

a single criterion — either the intensity of the field, the ratio of two consecutive nonzero terms of that expansion, or even the ratio of two characteristic energies — that can depend on the type of material being studied (whether it is a gas composed of individual atoms or molecules, or a crystal), its physical properties (in the case of a crystal, whether it is metal, a semiconductor or an insulator) and, of course, the intensity and frequency of the external electric field. In this work, we assume this perturbation theory to hold always, and so we will not trouble ourselves with an analysis of the different criteria nor undergo a thorough study of this theory's limitations.

We thus begin by considering the general properties of the susceptibility tensor in the case of the linear response.

2.1.1 Linear Response

The macroscopic polarisation of a system can be generally written, in the case of the linear response [11, 15], as the convolution of the external electric field, $E^{\alpha}(\mathbf{r}', t')$, and the corresponding response function, the linear susceptibility, $\chi^{\beta\alpha}(\mathbf{r}, \mathbf{r}', t, t')$,[1]

$$P^{\beta}(\mathbf{r}, t) = \int_{-\infty}^{t} d\mathbf{r}' dt' \chi^{\beta\alpha}(\mathbf{r}, \mathbf{r}', t, t') E^{\alpha}(\mathbf{r}', t'). \tag{2.1}$$

It is usually assumed that $\chi^{\beta\alpha}$ is a function of both the difference between the two positions, $\mathbf{r}$ and $\mathbf{r}'$, and the difference between the two times, t and t', meaning that it is invariant under both spatial and time translations [11, 135];[2] the polarisation, therefore, reads as,

$$P^{\beta}(\mathbf{r}, t) = \int_{-\infty}^{\infty} d\mathbf{r}' dt' \chi^{\beta\alpha}(\mathbf{r} - \mathbf{r}', t - t') E^{\alpha}(\mathbf{r}', t'). \tag{2.2}$$

Note that we have extended the integration in dt' by imposing that the the field at a time t' cannot affect the response of the system at an earlier time t, as required by causality,

$$\chi^{\beta\alpha}(\mathbf{r} - \mathbf{r}', t - t') \equiv 0 \quad \text{for } t' > t. \tag{2.3}$$

[1] We omit the order of polarisation , i.e., $P^{\beta}_{\text{first order}}(t) \to P^{\beta}(t)$, so as to make our notation as simple as possible; the right-hand side of the expression for $P^{\beta}(t)$ indicates, nevertheless, the order of the response. In addition, note that there is an implicit sum over repeated indexes.

[2] In ref.[135], the authors point out that "in general, the response functions of a crystal to driving fields depend on two spatial variables separately, and not through their difference" as "even an ideally perfect crystal is not rigorously homogeneous on a microscopic unit cell scale". This, therefore, involves an approximation, which they describe as "the simplification of neglecting 'local field effects'". We note that this is not particularly important to our work as we consider the electric field to be uniform throughout the crystal in all our calculations.

This relation between the polarisation, the linear susceptibility and the external electric field is more neatly transcribed when we move to the corresponding momentum, $\mathbf{q}$, and frequency, ω, domains,

$$\mathcal{A}(\mathbf{q},\omega) = \int_{-\infty}^{\infty} d\mathbf{r}dt\mathcal{A}(\mathbf{r},t)e^{-i\mathbf{q}\cdot\mathbf{r}}e^{i\omega t}. \tag{2.4}$$

Multiplying both members of Eq.(2.2) by $\exp\left[-i(\mathbf{q}\cdot\mathbf{r}-\omega t)\right]$ and integrating over the $\mathbf{r}$ and t variables gives us,

$$P^{\beta}(\mathbf{q},\omega) = \int_{-\infty}^{\infty} d\mathbf{r}dt \int_{-\infty}^{\infty} d\mathbf{r}'dt'\chi^{\beta\alpha}(\mathbf{r}-\mathbf{r}',t-t')E^{\alpha}(\mathbf{r}',t')e^{-i\mathbf{q}\cdot\mathbf{r}}e^{i\omega t}, \tag{2.5}$$

$$= \int_{-\infty}^{\infty} d\mathbf{r}dt\chi^{\beta\alpha}(\mathbf{r},t)e^{-i\mathbf{q}\cdot\mathbf{r}'}e^{i\omega t} \int_{-\infty}^{\infty} d\mathbf{r}'dt'E^{\alpha}(\mathbf{r}',t')e^{-i\mathbf{q}\cdot\mathbf{r}'}e^{i\omega t'}, \tag{2.6}$$

and can be reduced to the well-known relation,

$$P^{\beta}(\mathbf{q},\omega) = \chi^{\beta\alpha}(\mathbf{q},\omega)E^{\alpha}(\mathbf{q},\omega), \tag{2.7}$$

which connects the perturbation to the system with its response in a given mode, $(\mathbf{q},\omega)$, by means of the susceptibility associated with that momentum and that frequency. While certain problems in the optical response of crystals (and other physical systems in general) are sometimes predicated on the knowledge of the response function for nontrivial momentum, *e.g.* plasmonics [135, 136], in nonlinear optics one usually makes do with the response function at *zero* momentum, either because the expressions become unwieldy — and thus too difficult to use in practical calculations — or because one can sensibly introduce an approximation. In three-dimensional materials, this tends to be the long wavelength approximation, or dipole approximation, where one considers that the electric field is uniform throughout the material since its characteristic lengths — in the case of a crystal, the dimensions of the unit cell — are much smaller than the wavelength of the external field [11–13, 15, 135]. In two dimensions, however, this somewhat stringent approximation can be reconciled with the fact that most optical experiments are performed with electric fields that shine directly on the material, i.e., the electric field vector is orthogonal to the material's (a plane) normal vector, which means that the field (if it is considered to be a plane wave) is uniform throughout the material. The normal incidence condition thus allows us to introduce the same practical considerations as those of the dipole approximation, without having to consider the length scales associated with the system.[3] In the absence of any spatial dependence, the relation between the polarisation and the external field is expressed as,

$$P^{\beta}(\omega) = \chi^{\beta\alpha}(\omega)E^{\alpha}(\omega), \tag{2.8}$$

[3] We note, however, that there are works on the nonlinear optics of two-dimensional materials that focus on the finite-momentum aspects of the response, e.g., [55, 70].

for,

$$\mathcal{A}(\omega) = \int_{-\infty}^{\infty} dt \mathcal{A}(t) e^{i\omega t}, \tag{2.9}$$

We can now easily derive the reality condition for the susceptibility. First, consider that the polarisation in the time-domain, $P^{\beta}(t)$, is necessarily real, which means that,

$$P^{\beta}(t) = \left[P^{\beta}(t)\right]^{*} \Longrightarrow \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} P^{\beta}(\omega) e^{-i\omega t} = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \left[P^{\beta}(\omega)\right]^{*} e^{i\omega t} \Longrightarrow P^{\beta}(\omega) = \left[P^{\beta}(-\omega)\right]^{*}, \tag{2.10}$$

a statement that is also true in the case of the electric field, $E^{\beta}(t)$,

$$E^{\beta}(\omega) = \left[E^{\beta}(-\omega)\right]^{*}. \tag{2.11}$$

Upon replacing these last two equations in Eq.(2.8), and after some simple manipulations, we obtain the relation between the positive- and negative-frequency components of the susceptibility,

$$\boxed{\chi^{\beta\alpha}(\omega) = \left[\chi^{\beta\alpha}(-\omega)\right]^{*}}. \tag{2.12}$$

This is sometimes called the reality condition [13, 112]. We can present this relation in a somewhat different form by decomposing the susceptibility into its real and imaginary parts, $\chi^{\beta\alpha}(\omega) = \mathrm{Re}\left[\chi^{\beta\alpha}(\omega)\right] + i\mathrm{Im}\left[\chi^{\beta\alpha}(\omega)\right]$,

$$\mathrm{Re}\left[\chi^{\beta\alpha}(\omega)\right] = \mathrm{Re}\left[\chi^{\beta\alpha}(-\omega)\right], \tag{2.13}$$

$$\mathrm{Im}\left[\chi^{\beta\alpha}(\omega)\right] = -\,\mathrm{Im}\left[\chi^{\beta\alpha}(-\omega)\right], \tag{2.14}$$

i.e., the real (imaginary) part of the susceptibility must be an even (odd) function of the frequency. Note that the same must be true of the conductivity, $\sigma^{\beta\alpha}(\omega)$, as it is the object that relates the polarisation current, $J^{\beta}(\omega)$ — which obeys the same property as the polarisation, $P^{\beta}(\omega)$, Eq.(2.10) — and the external electric field. Furthermore, the conductivity and the susceptibility must themselves be related, since the polarisation current in the time domain is the time derivative of the polarisation, $\mathbf{J}(t) = \partial_t \mathbf{P}(t)$. In simple, classical terms, if $\mathbf{P}$, the dipole moment of a pair of charges, changes, then that change is the result of the (relative) motion of those charges, which implies a current.[4] One can easily check that the relation between the linear conductivity and the linear susceptibility must be,

$$\sigma^{\beta\alpha}(\omega) = -i\omega\chi^{\beta\alpha}(\omega). \tag{2.15}$$

[4]This current is sometimes called the polarisation current so as to distinguish it from currents associated with the magnetization of a material (bound currents) and the so-called free currents [58].

Separating the response functions into their real and imaginary parts, we obtain,

$$\mathrm{Re}\big[\sigma^{\beta\alpha}(\omega)\big] = \omega \mathrm{Im}\big[\chi^{\beta\alpha}(\omega)\big], \tag{2.16}$$

$$\mathrm{Im}\big[\sigma^{\beta\alpha}(\omega)\big] = -\omega \mathrm{Re}\big[\chi^{\beta\alpha}(\omega)\big]. \tag{2.17}$$

Later on, in Section 5.4, we will see why it is important to keep the relations between these two response functions in mind; they allow us to formulate the statement of full permutation symmetry in the case of absorptionless crystals.

Let us now consider a second set of relations for the susceptibility, which follows from the consideration that the response must be causal. Causality, as stated in Eq.(2.3), means that the response function must be zero for negative times, $\chi^{\beta\alpha}(t<0)=0$ — it imposes that electric fields in the future must have no bearing on the system's response in the present. When we consider the Fourier transform of this susceptibility,

$$\chi^{\beta\alpha}(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \chi^{\beta\alpha}(\omega) e^{-i\omega t}, \tag{2.18}$$

it is clear that the frequency-domain counterpart of the time-domain susceptibility, $\chi^{\beta\alpha}(\omega)$, must be such that this integral is necessarily zero for all negative times. That is only possible if the susceptibility (which we consider to be nontrivial) is an analytic function of ω in the *upper half* of the complex plane. In that case, we can draw a closed integration contour — by adding a trivial piece to the integral in Eq.(2.18) corresponding to the semicircumference that closes the contour (and whose radius is taken to infinity) — that contains no poles on the inside and that, therefore, must be zero. Note that this additional trivial piece requires t to be negative, so that $\exp(-i\omega t) = \exp(-i\mathrm{Re}[\omega]t)\exp(\mathrm{Im}[\omega]t)$ can decay exponentially, which ensures that the semicircumference portion of the contour integral is vanishing.

So causality implies that the susceptibility in the frequency domain is analytic in the upper half of the complex plane. That being the case, we can construct a contour integral, represented in Fig.2.1, that is necessarily zero,

$$\oint_C dz \frac{\chi^{\beta\alpha}(z)}{z-\omega} = \fint_{-\infty}^{\infty} d\omega' \frac{\chi^{\beta\alpha}(\omega')}{\omega'-\omega} + \int_\gamma dz \frac{\chi^{\beta\alpha}(z)}{z-\omega} + \int_\Gamma dz \frac{\chi^{\beta\alpha}(z)}{z-\omega} = 0, \tag{2.19}$$

as there are no poles in the region bounded by the contour. This integration can be broken up into three different pieces: a Cauchy principal value integral along the real axis (in the variable ω'), a contribution from the small semicircumference, which is labelled γ, and a contribution from the large semicircumference, which is labelled Γ. The latter of these contributions vanishes exactly, provided that the susceptibility goes to zero when $|z| \to \infty$. The second contribution, which involves a clockwise integration around the semicircumference centred in ω, and of vanishing radius r, gives

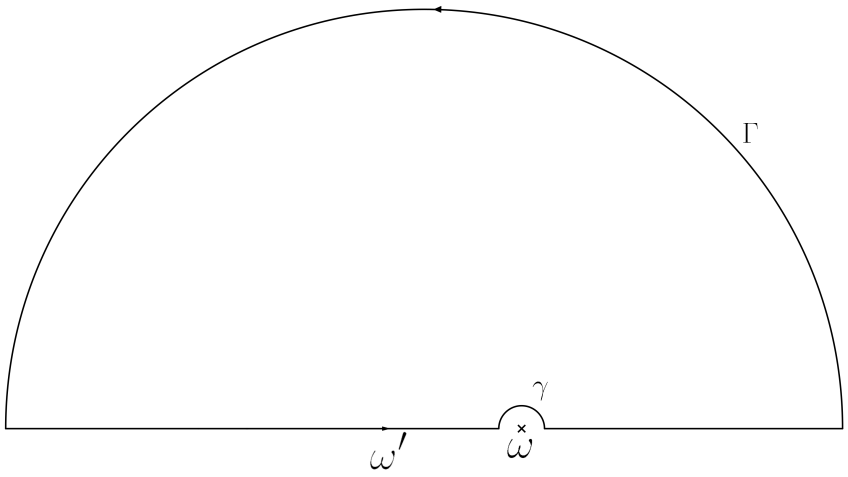


Figure 2.1: The contour used in the integral of Eq.(2.19). The γ portion of the contour represents the semicircumference of vanishing radius, and centred in ω. The Γ portion of the contour represents the semicircumference whose radius is to be taken to infinity.

us,

$$\begin{aligned}\int_\gamma dz \frac{\chi^{\beta\alpha}(z)}{z-\omega} &= \int_\pi^0 d(re^{i\theta}) \frac{\chi^{\beta\alpha}(\omega + re^{i\theta})}{\omega + re^{i\theta} - \omega}, \\ &= i\int_\pi^0 d\theta \chi^{\beta\alpha}(\omega + re^{i\theta}), \\ &\underset{r\to 0}{\Longrightarrow} -i\pi\chi^{\beta\alpha}(\omega). \end{aligned} \tag{2.20}$$

It thus follows that the statement in Eq.(2.19) can be expressed as,

$$\chi^{\beta\alpha}(\omega) = -\frac{i}{\pi} \fint_{-\infty}^{\infty} d\omega' \frac{\chi^{\beta\alpha}(\omega')}{\omega' - \omega}, \tag{2.21}$$

or, alternatively, as the relation between real and imaginary parts of the susceptibility,

$$\mathrm{Re}\big[\chi^{\beta\alpha}(\omega)\big] = -\frac{1}{\pi} \fint_{-\infty}^{\infty} d\omega' \frac{\mathrm{Im}\left[\chi^{\beta\alpha}(\omega')\right]}{\omega' - \omega}, \tag{2.22}$$

$$\mathrm{Im}\big[\chi^{\beta\alpha}(\omega)\big] = \frac{1}{\pi} \fint_{-\infty}^{\infty} d\omega' \frac{\mathrm{Re}\left[\chi^{\beta\alpha}(\omega')\right]}{\omega' - \omega}. \tag{2.23}$$

We note that the first of these relations is sometimes expressed as the relation between the refractive index of a material (which is related to the real part of the susceptibility, by way of the permittivity, ϵ) and its absorption coeffcient (which is related to the imaginary part of the susceptibility), in which case it is called the *dispersion* relation [112]. It is also, as pointed out in [47, 112], the

more useful of the two relations since absorption coefficients are easier to obtain, both theoretically and experimentally, than refractive indexes — the former can sometimes be accurately described by means of simple Fermi's golden rule arguments. There are, however, certain pitfalls to this type of calculation, particularly in the case where one uses effective Hamiltonians (in crystals) whose energy bands are not bounded. This absence of a cutoff energy can lead to an imaginary part of the susceptibility that does not vanish for exceedingly large frequencies, meaning that the Kramers-Kronig relations can no longer be used [10]. As before, the Kramers-Kronig relations, Eqs.(2.19)–(2.23), also apply to the conductivity.

Having derived several important properties of the linear susceptibility, we can now move on to a study of the general properties of the nonlinear susceptibility. While several of these properties lend themselves to easy generalisations, namely, the reality condition, the relation between susceptibilities and conductivities, and the single argument Kramers-Kronig relations, there are certain extensions that are not altogether obvious, *e.g.*, the Kramers-Kronig relations in the case of degenerate arguments, and whose caveats must be considered. There is also a new symmetry, which is absent in the linear susceptibility, that transcribes the fact that the order in which the electric fields appear in the polarisation must be irrelevant. Since it is intrinsic to the definition of a nonlinear polarisation, it is usually called the intrinsic permutation symmetry.

2.1.2 Nonlinear Response

First, let us consider how the nonlinear polarisation itself should be written. For the n-th order polarisation, the response will necessary involve a convolution between the response function, now a function of n different frequencies, and a product of n electric fields, each of which tied to a specific frequency. Given, however, that this polarisation must be a function of a single frequency, one must impose that the sum of running frequencies, i.e., frequencies over which we will be integrating, is equal to the output frequency: that of the polarisation. Here, we will consider the case of the second order polarisation, so as to maintain the notation as simple as possible. It should read,

$$P^{\beta}(\omega) = \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)E^{\alpha_1}(\omega_1)E^{\alpha_2}(\omega_2)(2\pi)\delta(\omega_1+\omega_2-\omega). \tag{2.24}$$

Like its first order counterpart, this polarisation must also satisfy the reality condition, as stated in Eq.(2.10). Since $\left[P^{\beta}(-\omega)\right]^*$ can be written as,

$$\left[P^{\beta}(-\omega)\right]^* = \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\left[\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)\right]^*\left[E^{\alpha_1}(\omega_1)\right]^*\left[E^{\alpha_2}(\omega_2)\right]^*(2\pi)\delta(\omega_1+\omega_2+\omega), \tag{2.25}$$

$$\underset{\omega_i\to-\omega_i}{\Longrightarrow} \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\left[\chi^{\beta\alpha_1\alpha_2}(-\omega_1,-\omega_2)\right]^*\left[E^{\alpha_1}(-\omega_1)\right]^*\left[E^{\alpha_2}(-\omega_2)\right]^*(2\pi)\delta(-\omega_1-\omega_2+\omega), \tag{2.26}$$

and, knowing that the electric fields themselves satisfy the reality condition, Eq.(2.11), we can show that the second order susceptibility must satisfy the following relation between its positive- and negative-frequency contributions,

$$\boxed{\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2) = \left[\chi^{\beta\alpha_1\alpha_2}(-\omega_1,-\omega_2)\right]^*}. \tag{2.27}$$

This is the second order reality condition, which is an immediate generalisation of Eq.(2.12). In addition, the relation between the linear susceptibiltity and the linear conductivity is also easily extended to second order and reads as,

$$\sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2) = -i\omega_{12}\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2), \tag{2.28}$$

for

$$\omega_{12} = \omega_1 + \omega_2, \tag{2.29}$$

which means that, in general, the nonlinear conductivity tensor at order n can be obtained from the corresponding susceptibility by multiplying the latter by a factor of $-i$ times the total output frequency, $\omega_{1\ldots n} = \omega_1 + \ldots + \omega_n$. The third property to be considered here, however, has no correspondence to any of the properties we have seen so far. It is clear that the order of the electric fields in Eq.(2.24) cannot be relevant to the calculation of the second order polarisation — there is only a product of fields, not a first field and a second field — meaning that the simultaneous exchange of Cartesian indexes and their corresponding frequencies, $(\alpha_1,\omega_1) \leftrightarrow (\alpha_2,\omega_2)$, should render us the same (physical) response function; we must impose what is usually called the intrinsic permutation symmetry,

$$\chi^{\beta\alpha_1\alpha_2}_{\text{phys.}}(\omega_1,\omega_2) = \chi^{\beta\alpha_2\alpha_1}_{\text{phys.}}(\omega_2,\omega_1), \tag{2.30}$$

which is necessarily true when we define that object as,

$$\chi^{\beta\alpha_1\alpha_2}_{\text{phys.}}(\omega_1,\omega_2) = \frac{1}{2}\left[\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2) + \chi^{\beta\alpha_2\alpha_1}(\omega_2,\omega_1)\right]. \tag{2.31}$$

Henceforth, when talking about susceptibilities or conductivities, we will always consider the so-called physical — or intrinsically symmetric — response function, as it is the only object that is worth discussing.[5] This point, however trivial it may seem, can have an impact on one's description of nonlinear optical responses, as we will see in our study of the nonlinear optical response of plain (and gapped) graphene in Chapter 4.

Finally, let us consider the nonlinear version of the Kramers-Kronig relations of Eqs.(2.19)–(2.23). The susceptibility is now a function of two arguments; for it to be causal, it must simulta-

[5] We will, however, drop the subscript.

neously satisfy,

$$\chi^{\beta\alpha_1\alpha_2}(t,t') \equiv 0 \quad \text{for } t<0 \quad \text{and} \quad \chi^{\beta\alpha_1\alpha_2}(t,t') \equiv 0 \quad \text{for } t'<0. \tag{2.32}$$

This means the argument we used to show that the first order susceptibility must be analytic in the upper half of the complex plane (and from there derive the Kramers-Kronig relations) is still valid in second order, when we are considering the two frequency arguments to be independent of one another. Therefore, in the case of the single-frequency Kramers-Kronig relations, one frequency is allowed to run as long as the other one remains fixed,

$$\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2) = -\frac{i}{\pi}\fint_{-\infty}^{\infty} d\omega_1' \frac{\chi^{\beta\alpha_1\alpha_2}(\omega_1',\omega_2)}{\omega_1'-\omega_1}, \tag{2.33}$$

$$\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2) = -\frac{i}{\pi}\fint_{-\infty}^{\infty} d\omega_2' \frac{\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2')}{\omega_2'-\omega_2}. \tag{2.34}$$

For Kramers-Kronig relations that carry degenerate frequencies, the argument is somewhat more subtle. Consider the case of the second harmonic generation, $\chi^{\beta\alpha_1\alpha_2}(\omega,\omega)$. We can naïvely define degenerate Kramers-Kronig relations for this susceptibility by simply writing,

$$\chi^{\beta\alpha_1\alpha_2}(\omega,\omega) = -\frac{i}{\pi}\fint_{-\infty}^{\infty} d\omega' \frac{\chi^{\beta\alpha_1\alpha_2}(\omega',\omega')}{\omega'-\omega}. \tag{2.35}$$

This indeed holds true under scrutiny since, when we consider the double Fourier transform of this object — from the time domain to the frequency domain,

$$\chi^{\beta\alpha_1\alpha_2}(\omega,\omega) = \int_{-\infty}^{\infty} dt_1 dt_2 \chi^{\beta\alpha_1\alpha_2}(t_1,t_2) e^{i\omega(t_1+t_2)}, \tag{2.36}$$

causality ensures, for frequencies in the upper half of the complex plane, $\omega = \mathrm{Re}[\omega] + i\mathrm{Im}[\omega]$, that the integral converges,

$$\chi^{\beta\alpha_1\alpha_2}(\omega,\omega) = \int_{-\infty}^{\infty} dt_1 dt_2 \chi^{\beta\alpha_1\alpha_2}(t_1,t_2) e^{i\mathrm{Re}[\omega](t_1+t_2)} e^{-\mathrm{Im}[\omega](t_1+t_2)}. \tag{2.37}$$

This shows that $\chi^{\beta\alpha_1\alpha_2}(\omega,\omega)$ is analytic in the upper half of the complex plane, and we can define degenerate Kramers-Kronig relations for the second harmonic generation, Eq.(2.35). Consider, however, the case of the optical rectification, $\chi^{\beta\alpha_1\alpha_2}(\omega,-\omega)$. The double Fourier transform of this object shows us,

$$\chi^{\beta\alpha_1\alpha_2}(\omega,-\omega) = \int_{-\infty}^{\infty} dt_1 dt_2 \chi^{\beta\alpha_1\alpha_2}(t_1,t_2) e^{i\omega(t_1-t_2)}, \tag{2.38}$$

$$= \int_{-\infty}^{\infty} dt_1 dt_2 \chi^{\beta\alpha_1\alpha_2}(t_1,t_2) e^{i\mathrm{Re}[\omega](t_1-t_2)} e^{-\mathrm{Im}[\omega](t_1-t_2)}, \tag{2.39}$$

that, for $t_1 < t_2$ and $t_1 > 0$, the integral need not converge. This means that $\chi^{\beta\alpha_1\alpha_2}(\omega, -\omega)$ does not have to be analytic in the upper half of the complex plane, in which case there can be no Kramers-Kronig relations for the optical rectification. One can extend this reasoning so as to show that only susceptibilities that can be expressed as $\chi^{\beta\alpha_1\alpha_2}(n_1\omega + \bar{\omega}_1, n_2\omega + \bar{\omega}_2)$ — $n_i \geq 0$ for all i, and at least one nonzero n_i — have degenerate Kramers-Kronig relations [12, 112]. Furthermore, this result can be easily extended to susceptibilities of higher orders [12, 112], meaning that, although we can use the Kramers-Kronig relations to, *e.g.*, obtain the real part of the third harmonic generation, $\chi^{\beta\alpha_1\alpha_2\alpha_3}(\omega, \omega, \omega)$, from its imaginary part, we cannot do so for the photoconductivity, $\chi^{\beta\alpha_1\alpha_2\alpha_3}(\omega, \omega, -\omega)$.

The argument we have presented thus far has also led us to the following conclusion: while there are some pitfalls to the use of degenerate Kramers-Kronig relations, one can safely use the single-frequency ones in the calculation of nonlinear susceptibilities, provided that these susceptibilities are analytic in the upper half of the complex plane. As noted in ref.[112], it is well-known that certain expressions for the susceptibilities — derived by means of simple Fermi's Golden rule arguments or by quantum calculations using the density matrix formalism — do not satisfy this analyticity property, *e.g.*, the case of the first and third order susceptibilities for the steady-state response of the two-level atom [12], so one must tread carefully when using these relations in practical calculations.

This absence of analyticity is, in the case of the nonlinear optical response of crystals, connected with certain physical effects, such as the injection current in the second order response, in which the susceptibility (or the conductivity) must indeed be divergent. Although this means that some of the original appeal of the Kramers-Kronig relations — that of being able to obtain the real part of susceptibility from its imaginary part (and vice-versa) — is lost in this case, a careful consideration of these relations can, even in the case of susceptibilities with poles, prove useful.[6] Therein lies the key to our derivation of the relationship between real parts of the second order conductivity of crystal electrons, which will be the subject of Section 5.5.

2.2 Point Group Symmetries

The second set of general properties that we will discuss here concerns the conditions imposed on the susceptibility tensor of a crystal by its spatial symmetries. As we have previously mentioned, we will not present a comprehensive overview of concepts of group theory in solid state physics — we direct readers interested in this subject to the following works: [133, 134]. We will, however, point out to some of the reasons that make this type of analysis useful in the study of nonlinear optical responses, as well as present elements that will be used in the following chapters, namely the

[6] As we will see in the chapter dedicated to the nonlinear optical response of plain and gapped graphene, Chapter 3, the injection current does not exist for materials with certain point group symmetries, meaning that the second order susceptibilities (or conductivities) are free from poles, in which case we can use the single-frequency Kramers-Kronig relations without an issue.

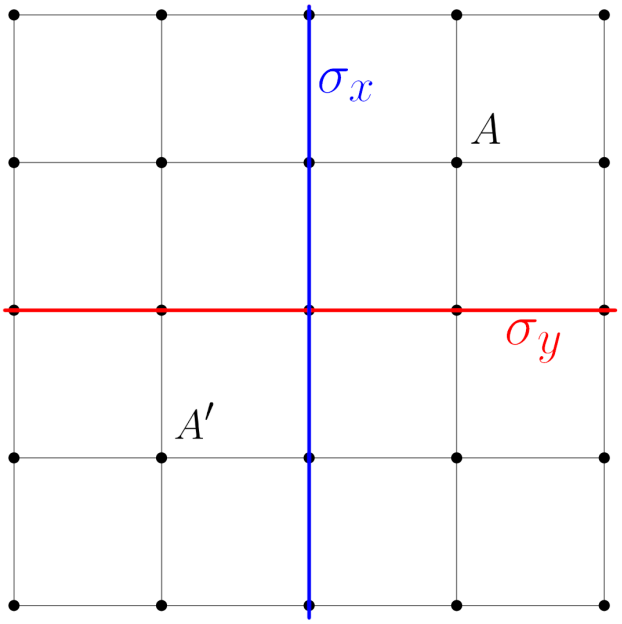


Figure 2.2: The square lattice with one atom per unit cell. The blue and red lines, labelled σ_x and σ_y, represent two reflection planes under which $y \to -y$ and $x \to -x$, respectively. The square lattice is invariant under both these symmetry operations, meaning that it is also invariant under the inversion symmetry operation: the atom at A is transformed into that at A', which is its equivalent (and vice-versa). The point group symmetry of the square lattice is the D_{4h}.

symmetry conditions imposed on the susceptibility tensors describing the second and third order responses of the plain graphene and gapped graphene/hexagonal boron nitride monolayers.

Crystals are objects whose constitutive elements, atoms, are arranged in an orderly fashion and with a certain periodicity. This order — which we can envision more clearly when thinking of a specific crystal, *e.g.* the square lattice with one atom per unit cell, Figure 2.2 — translates itself into a group of symmetry operations under which that crystal is invariant: the reflections along the red and blue lines (these indicate two different reflection planes) of Figure 2.2 are examples of symmetry operations that leave the crystal unchanged. To the complete group of symmetry operations that leave the crystal invariant and that do *not* involve translations, we call the point group of the crystal [133]. An adequate description of the physics of these materials has, of course, these symmetries built into it (we usually think of this in terms of a crystal potential that is invariant under certain symmetry operations), something that is then transcribed onto the entities that describe the dynamics of electrons in a crystal: its energy bands and Bloch functions. The same is true for the tensors describing that crystal's linear and nonlinear optical response.

Consider, once again, the case of the square lattice. The linear response to an electric field polarised along the x direction, $P^x(\omega) = \chi^{xx}(\omega)E^x(\omega)$, must be the same as that to an equivalent electric field (that is, a field of the same amplitude) polarised along the orthogonal direction, y, $P^y(\omega) = \chi^{yy}(\omega)E^y(\omega)$, since these directions are completely equivalent. For electric fields of the same amplitude, i.e., $E^x(\omega) = E^y(\omega)$, the polarisations can only be one and the same if the susceptibilties satisfy the following property, $\chi^{xx}(\omega) = \chi^{yy}(\omega)$.

While arguments such as this one may not be an adequate way of deriving the general form of a tensor for a particular point group, they do tell us that we do not need to know the actual expression for the response function in order to determine the properties that it should satisfy. Knowing these properties before one starts computing (or thinking about computing) a certain linear or nonlinear optical susceptibility, can be extremely useful, since: 1. it saves us the trouble of computing tensor components that are redundant, or exactly zero; 2. we can use them to cross-check our practical calculations; 3. it can tell us if certain effects are present in a material (or not). This first point is clearly illustrated by considering the role played by inversion symmetry in the nonlinear optical response at even orders of the electric field, which, as we will now see, can be quite consequential [12, 13].

In two dimensions, inversion symmetry, or centrosymmetry, involves the symmetry operation that combines the reflections in the σ_x and σ_y planes, so that a given point (x, y) is transformed into $(-x, -y)$. The square lattice with one atom per unit cell is, of course, centrosymmetric — in Figure 2.2, A is transformed into A', a completely equivalent atom, and vice-versa. Now, the polarisation and the electric field vectors, $\mathbf{P}(t)$ and $\mathbf{E}(t)$, both change their signs under this symmetry operation, something that proves itself irrelevant when we are considering the linear response,

$$P^\beta(\omega) = \chi^{\beta\alpha}(\omega)E^\alpha(\omega) \underset{\text{inv.}}{\Longrightarrow} -P^\beta(\omega) = \chi^{\beta\alpha}(\omega)\big[-E^\alpha(\omega)\big] \Longrightarrow P^\beta(\omega) = \chi^{\beta\alpha}(\omega)E^\alpha(\omega). \tag{2.40}$$

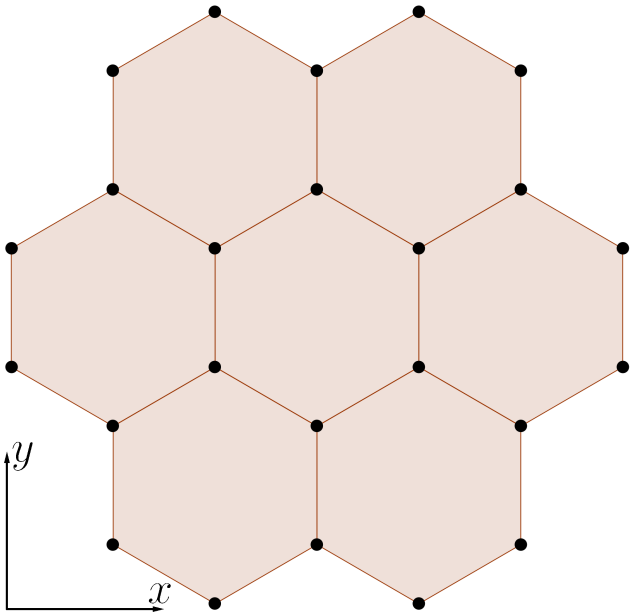


Figure 2.3: The honeycomb lattice. In graphene, the two unit-cell atoms are both carbon atoms and thus entirely equivalent. In gapped graphene or hexagonal boron nitride, however, these two atoms are not equivalent, which means the absence of inversion symmetry. The x and y axes indicate the zig-zag and armchair directions, respectively.

In the case of second order response (or, in general, for any even order response), however, we can see that it implies that all components of the susceptibility tensor must be zero. It follows from Eq.(2.24) that, under this symmetry operation, the polarisation should read,

$$-P^{\beta}(\omega) = \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)\big[-E^{\alpha_1}(\omega_1)\big]\big[-E^{\alpha_2}(\omega_2)\big]\delta(\omega_1+\omega_2-\omega). \tag{2.41}$$

The left-hand sign picks up a minus sign while the minus signs in the right-hand side cancel out, meaning that,

$$P^{\beta}(\omega) = -P^{\beta}(\omega) \rightarrow P^{\beta}(\omega) = 0, \tag{2.42}$$

which is true if and only if the second order susceptibility tensor is the null tensor,

$$\boxed{\chi^{\beta\alpha_1\alpha_2}_{\text{centro.}}(\omega_1,\omega_2) = 0}. \tag{2.43}$$

Centrosymmetric materials are thus materials in which the first nonlinear optical response is a third order response, which is why most commonly used (three-dimensional) crystals in nonlinear optics are noncentrosymmetric [13]. We are, however, interested in the nonlinear optical response of two-dimensional crystals, where graphene [61] — despite its centrosymmetricity — is the most thoroughly studied material, as well as the most extensively used in practical applications. As we

	$\chi^{\beta\alpha}$	$\chi^{\beta\alpha_1\alpha_2}$	$\chi^{\beta\alpha_1\alpha_2\alpha_3}$
Graphene (D_{6h})	$\chi^{xx} = \chi^{yy}$ $\chi^{xy} = \chi^{yx} = 0$	0	$\chi^{xxxx} = \chi^{yyyy} = \chi^{xxyy} + \chi^{xyxy} + \chi^{xyyx}$ $\chi^{xxyy} = \chi^{yyxx}$, $\chi^{xyxy} = \chi^{yxyx}$, $\chi^{xyyx} = \chi^{xyyx}$ All other ele. are zero.
hBN/GG (D_{3d})	As above	$\chi^{xxy} = \chi^{xyx} = \chi^{yxx} = -\chi^{yyy}$ All other ele. are zero.	As above

Table 2.1: The symmetry conditions of the first three susceptiblity tensors for graphene and hBN/GG (hexagonal boron nitride/gapped graphene). The point group associated with each material (or set of materials) is given in the Schoenflies notation. x and y indicate the zig-zag and armchair directions, respectively.

have mentioned in the Introduction, this has to do with the fact that the nonlinear optical properties of graphene were early recognised as being exceptional [66], and that it was the first two-dimensional material to be isolated [137].

Graphene is, of course, a well-known two-dimensional allotrope of carbon, where atoms are arranged in a honeycomb lattice, Figure 2.3. It is, as we have just mentioned, centrosymmetric, as the two atoms in its unit cell are completely equivalent. When this equivalence is absent, i.e., in the cases of gapped graphene and hexagonal boron nitride, the material losses the reflection plane along the zig-zag direction ($y \to -y$ in the figure) and it is, therefore, no longer invariant under the inversion symmetry operation. These materials can have a nontrivial second order response for certain configurations of the electric field: if the electric field is polarised in the zig-zag direction, then there is no second order response (note that the material retains the reflection in the plane parallel to the armchair direction); if the electric field is polarised in the armchair direction, then there is a second order response, also in the armchair direction. For an electric field that has a mixed polarisation, both the zig-zag and armchair components of the polarisation (or the current) are nonzero. In the case of the third order response, and despite the differences between the points groups associated with these materials — D_{6h} for graphene and D_{3d} for gapped graphene and hexagonal boron nitride — there are no differences between the tensor properties of these two groups [134]. The complete set of symmetry properties imposed on the susceptibility tensors of these materials (in the first three orders of the field) is given in Table 2.1.

We can now finish our overview of the general properties of tensors describing nonlinear optical responses. These insights, as the remainder of this work will make clear, can be used in our calculations and can help us understand certain physical effects. In the following chapter, we study how quantum calculations of nonlinear optical responses of electrons in crystals can be performed, and explore the issues raised by the use of different representations of the external electric field.

Chapter 3

Gauge Covariances and Nonlinear Optical Responses

This chapter is based on the following publication:

- *Gauge covariances and nonlinear optical responses*, **G. B. Ventura**, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres, Phys. Rev. B 96 (2017), 035431.

This chapter is dedicated to calculations of nonlinear optical responses of electrons in crystals and to the particular complications one finds therein.[1] It is centred around three different elements: (a) the consequences of gauge invariance, i.e., of the ability to choose different representations of the same electric field, and how these affect calculations of nonlinear optical responses; (b) the issue of expressing the perturbations associated with two specific representations of the electric field, and how this is related to the freedom to set the phases of the eigenfunctions of the unperturbed Hamiltonian — the Bloch functions; (c) the issue of ensuring gauge invariance and the existence of sum rules, and why (and when) these break.

We begin with a short, historical introduction to nonlinear optics in general, and to calculations of nonlinear optical coefficients in particular. This should provide the reader with the necessary context, as well as serve as an overview (even if an extremely abridged one) of the literature on calculations of nonlinear optical coefficients. We then proceed to discuss the ideas of aforementioned work [2], and present them in slightly more pedagogic and illustrative terms than in the paper.

3.1 Introduction

The nonlinear optical properties of crystals have long been a subject of interest to optical and condensed matter physicists. The birth of this field, as is so often emphasized in textbooks on this

[1] For the corresponding problem in atoms, check ref.[138–140].

subject [11–13, 16],[2] is usually ascribed to the 1961 experiment on the second harmonic generation conducted by Peter Franken and his collaborators at the University of Michigan [141]. Franken and his team focused a light beam from a ruby laser (which had just been invented the previous year) on a slab of crystalline quartz, and saw that it could produce ultraviolet radiation at twice the light frequency of the original beam. With this simple experiment, they showed that the harmonic generation of optical waves could — similarly to that of low-frequency electromagnetic waves [11] — be observed in a laboratory and produced, for the very first time, the generation of the second harmonic. The intensity of the laser field that Franken and his team used was, of course, central to their experiment; the efficiency of energy conversion from the fundamental to the second harmonic was so low — about one part in 10^8 [13] — that, without the intensities that a laser can generate, the second harmonic generation could not have been observed in that particular setup. The extremely low efficiency of this experiment, it was discovered several months later [16, 142], was due to the absence of *phase matching*, that is, to the difference between the refractive index of quartz at the two relevant frequencies, n_ω and $n_{2\omega}$, which is the normal behaviour of dispersive media.[3] This mismatch between the accumulated phases of the fundamental wave and the second harmonic meant that the coherence length, $L_{\text{coh.}} = \lambda/4\,|n_{2\omega} - n_\omega|$, for λ the central wavelength of the laser, was extremely small and that only a small portion of the quartz sample contributed to the process of second harmonic generation; the laser had solved this complication by putting just enough energy into the system so as to make this effect measurable, even if only barely so.[4] In addition to this understanding of the problem, Paul Maker and his colleagues also proposed a solution, which serves as the basis for several techniques in nonlinear optics that became standard later on. It involved using the birefringence of a crystal (in their case, a crystal of potassium dihydrogen phosphate, or KDP), as well as an appropriate setup, to align the wave vector of the second harmonic generation field with the axis associated with the smallest refractive index of the material. This process, usually called *birefringent phase matching*, was used to increase the conversion efficiency of the second harmonic generation to values much larger than that of the original experiment — to about the order of one in ten. This made it clear that nonlinear optics could be much more than a mere remarkable curiosity in physics, and that it could be put to use in practical applications [11, 13].

The Franken and Maker experiments opened up the field of nonlinear optics — a field that had, from its inception, crystals as an integral element. It was thus not long before some physicists decided to set their attention on the crystals themselves, so as try to understand why these materials had such interesting nonlinear optical properties, and how these could be quantified: the first

[2] We note that some authors like to emphasize developments in nonlinear optics that came before the Franken experiment of 1961. Geoffrey New starts his book — ref. [13] — with a description, not of the Franken experiment, where the second harmonic generation was observed for the first time, but of the Kerr and Pockels effects. These nonlinear effects, which involve the change of a material's refractive index under the application of a DC electric field, were discovered in the late nineteenth century.

[3] Under normal dispersion conditions, the refractive index is an increasing function of frequency, which further complicates this issue [12].

[4] Geoffrey New notes that the photographic plate presented in Franken's work appears to be totally blank [13].

theoretical descriptions of the polarisation (or, equivalently, the susceptibility) based on a single-particle description began to appear early in 1962 [143, 144]; those based on a density-matrix description, following the Liouville equation, were being presented as a textbook subject by 1965 [16, 145]. Two patterns were already emerging in these proposed descriptions, of which some elements were to become an integral part of later discussions (as well as an integral part of this work). First, the similarities between atomic and crystalline systems were usually emphasized, and the expressions derived for the responses were expressed in terms of sums over generic states; this, while it allowed for a more general formulation, disguised a complication that is found in crystals that is not found in atoms — that of expressing the matrix elements of the position operator in the eigenbasis of the unperturbed system [18, 19].[5] Secondly, the two most immediate representations of the electric field, in the case where it is considered to be uniform — the vector potential, $\mathbf{E}(t) = -\partial\mathbf{A}(t)/\partial t$, and scalar potential, $\mathbf{E}(t) = -\nabla\phi(\mathbf{r}, t)$, representations — were used interchangeably, following the principle that, from a physical standpoint, this choice could not be of any consequence. Gauge invariance ensures that, in theory, responses that are calculated using one representation must be the same as those calculated using the other representation, although it does not tell us which of these can be more easily put in practice; this, as expected, depends on the details of the system at hand — whether it is an atom or a crystal — as well as the method of calculation — whether it is an analytical calculation or a numerical one. Perhaps more importantly, it is also silent as to the rules (and the conditions in which these hold true) that enforce the equivalence between representations, which, once again, can only follow from the details of the system that is being considered.

These matters came to a head in the late eighties and early nineties, when interest in this subject was renewed by several attempts at performing practical calculations of nonlinear optical coefficients in bulk semiconductors [24–26]. The description of choice was, at first, the velocity gauge; in this gauge, the electric field is expressed in terms of the vector potential, and the perturbation to the Hamiltonian of a crystal electron in a continuum description, $\mathcal{H}$, is written as the sum of the dot product of the vector potential and the velocity operator — hence the name velocity gauge — and an extra diamagnetic term,[6]

$$\mathcal{H}_A(t) = \mathcal{H} + |e|\,\mathbf{A}(t)\cdot\mathbf{v} + \frac{e^2}{2m_e}\mathbf{A}(t)\cdot\mathbf{A}(t). \tag{3.1}$$

The expressions for the response, whether in terms of the polarisation or the electric current, derived in this gauge are very similar to those derived for atomic systems, with the differences being the introduction of an integral over all states (that is, over every $\mathbf{k}$) in the first Brillouin zone, and

[5] V. D. Genkin and P. M. Mednis' work of 1968 is somewhat of an exception [144]. Taking the continuum, minimally-coupled, Hamiltonian as their starting point, the authors use a unitary transformation to obtain an effective Hamiltonian that is similar to that usually used in a scalar potential description of the field, but that includes nonlinear terms in addition to the standard dipole interaction. The authors also acknowledge the complications associated with the Bloch matrix elements of the position operator.

[6] Here, $|e|$ and m_e are the absolute value of the charge of the electron and its mass, respectively.

that the sums over atomic quantum numbers are to be replaced by sums over all bands. Despite their simplicity, when these expressions were employed in the calculation of the second harmonic generation in $(\mathrm{Si})_n/(\mathrm{Ge})_n$ superlattices [25], and the third harmonic generation in Si, Ge, and GaAs crystals [26], they showed an *unphysical* behaviour — namely, the susceptibilities of cold semiconductors at zero frequency would diverge, instead of being zero — though this could be, in certain circumstances, corrected by the use of identities, later called sum rules; as an example, the sum rule derived in ref.[26] relates the matrix elements of the velocity operator with its derivatives with respect to $\mathbf{k}$. Nonetheless, since these sum rules are not easily derived (and were not entirely understood), and since the expressions themselves require a lot of nontrivial manipulations, the focus soon turned from the velocity gauge to its counterpart, the length gauge [20, 22, 23]. In this description, the field is expressed in terms of the scalar potential and the corresponding perturbation to the crystal Hamiltonian, for a uniform electric field, is the dipole interaction term,

$$\mathcal{H}_E(t) = \mathcal{H} + |e|\,\mathbf{E}(t)\cdot\mathbf{r}. \tag{3.2}$$

This, although seemingly simple, is not without its issues. Note that the dipole interaction term, unlike the perturbation in the velocity gauge, does not retain the symmetries of the original crystal Hamiltonian, $\mathcal{H}$, since it is not invariant under lattice translations; as a result, while computing the matrix elements of the perturbation is a somewhat trivial task in the velocity gauge, it is a nontrivial one in the length gauge. The calculation of the matrix elements of this operator in the Bloch basis was, nonetheless, a known (and solved) problem by the mid-1990s, having been a subject of E. I. Blount's seminal work of 1962 [18, 19]. The authors of ref.[22, 23] were well-aware of this; in the description presented in ref.[22], they used Blount's ideas on the position operator to derive a perturbation theory in the first two nonlinear orders, based on a density-matrix description, and from there obtain expressions for the second and third order susceptibilities that were free from unphysical divergences.[7] In addition, the authors tried to address the subject of equivalence between the two gauges, which they noted, *had to be true*, provided that calculations in both gauges are done correctly. First, they pointed out the role played by the commutator identity — $\left[r^\alpha, v^\beta\right] = i(\hbar/m_e)\delta^{\alpha\beta}$ — in the first order sum rule, which they used to show the equivalence between gauges in that order. Secondly, they noted that similar sum rules should exist for higher orders, and that these would again be contingent on commutator identities. Finally, they noted that the electronic structure models used in "practical calculations" sometimes contradicted the commutator identities, in which case the sum rules were broken, and the equivalence between gauges was no longer ensured. As we will see in this chapter, these three points offered some extremely

[7] In addition, the authors ref.[22, 23] introduced the idea of a *generalized* derivative, by considering the Bloch matrix elements of the commutator of the intraband portion of the position operator — which, as we will presently see, is its problematic portion — and a given operator. While not entirely relevant for this chapter, since, in our work, the position operator is expressed in terms of the *covariant* derivative, the generalized derivative will be used later on, in Chapters 4 and 5.

valuable insights and helped further the understanding of the equivalence problem between the length and velocity gauges. This work did not, however, solve the problem of turning the velocity gauge into something that could be used *practically,* neither did it clarify the general form of the sum rules and the conditions in which these hold true.[8] Instead, the authors, quite sensibly, advocated the use of the length gauge in practical calculations.

The length gauge was thus held as the standard method of computation for the next two decades. Despite the renewed interest in this subject [59, 66, 68, 69], spurred by the exceptional nonlinear properties of two-dimensional materials like graphene, as we have mentioned in the Introduction, the focus was at first centred around the difficulties concerning the representation of the position operator — which, as we will see in a subsequent section, can only be properly defined in the infinite crystal limit, and leads to equations of motion for the density matrix that are coupled in $\mathbf{k}$ — and, specifically, on attempts to circumvent this problem, while working within the framework of the length gauge [59, 68]. In ref.[68], the authors used a $\mathbf{k}$-space translation to decouple the equations of motion for the density matrix, though this is not without its cost, as the system's response is then expressed both in terms of the vector potential and the electric field, which complicates an order-by-order study of the current response. S. A. Mikhailov, on the other hand, avoided these problems by using a finite wavelength perturbation that satisfies periodic boundary conditions [59],

$$V(\mathbf{r}) = -e\left(V_{\mathbf{q}}e^{i\mathbf{q}\cdot\mathbf{r}} + V_{\mathbf{q}}^{*}e^{-i\mathbf{q}\cdot\mathbf{r}}\right), \tag{3.3}$$

and taking the uniform field limit, $\mathbf{q}\to 0$, directly in the expressions for the linear and third order optical conductivities [59]. This requires a small $\mathbf{q}$ expansion, which is a remarkably cumbersome procedure in the case of the third order response, since several tens of terms have to be taken into consideration. Nevertheless, these two methods can be used to derive nonlinear optical conductivities that, as expected, match with those that follow from the original length-gauge description.

Interest in the equivalence between gauges — as well as in the velocity gauge itself — was therefore renewed with the publication of works written by members of our group [2, 3, 5, 6] in the past four years,[9] the first of which serves as the foundation for this chapter. In the following sections, we recover ideas from ref.[2], and present them in more illustrative terms. We begin with a description of the equivalence between the two gauges in the context of free electrons. This is followed by a section on the physics of electrons in a crystal, and on the issue of writing the position operator in terms of the eigenstates of that system; there, we introduce the covariant derivative in $\mathbf{k}$-space as the position operator in the Bloch representation. The subsequent two sections are dedicated to the density matrix and, once again, to the relation between the two gauges: the end result of the first is a set of general expressions for the density matrices in the length and velocity gauges for a system described by the continuum Hamiltonian; the end result of the second is a set

[8] The work that serves as the basis of this chapter did not do so as well. That problem would be solved the following year, in a work written by my colleague D. J. Passos [3].

[9] At the time of writing, that is, January 2021.

of general sum rules that ensure the equivalence between the length and velocity gauges, and is accompanied by a study of the conditions under which these hold true. The expressions for the linear and third order response of the graphene monolayer — the subject of Section 5 of ref.[2] — are not, however, part of this thesis; a study of the response of plain and gapped monolayer graphene is thus postponed until Chapter 4, where we will make full use of the velocity gauge description presented in another of our group's works [3].

Finally, we take a moment to note that this work has been of some use in recent investigations on a variety of topics, ranging from studies on higher harmonic generation in crystals in the Wannier gauge [100] to studies on the nonlinear conductivity of odd-parity magnetic multipole systems [132]; our description of the length gauge, formulated in terms of a covariant derivative, was either mentioned or used in ref.[100, 120, 124, 132], while our insights on the velocity gauge (in the context of the continuum Hamiltonian) were highlighted in ref.[40, 146]. Furthermore, our comments on the conditions under which gauge invariance is preserved were mentioned in ref.[40–42, 74, 104, 108].

3.2 Equivalence between the Two Gauges (I)

A fundamental aspect of electrodynamics is the ability to express the real, physical fields — the electric, $\mathbf{E}(\mathbf{r},t)$, and magnetic, $\mathbf{B}(\mathbf{r},t)$, fields[10] — in terms of corresponding potentials — the scalar, $\phi(\mathbf{r},t)$, and vector, $\mathbf{A}(\mathbf{r},t)$, potentials — with a certain amount of freedom, which is usually called *gauge* freedom [58]. For the electric field, this means that it can be written as,

$$\mathbf{E}(\mathbf{r},t) = -\nabla\phi(\mathbf{r},t) - \frac{\partial\mathbf{A}(\mathbf{r},t)}{\partial t}, \tag{3.4}$$

which is necessarily invariant under the transformations,

$$\phi(\mathbf{r},t) \to \phi(\mathbf{r},t) - \frac{\partial\lambda(\mathbf{r},t)}{\partial t}, \tag{3.5}$$

$$\mathbf{A}(\mathbf{r},t) \to \mathbf{A}(\mathbf{r},t) + \nabla\lambda(\mathbf{r},t). \tag{3.6}$$

Transformations involving changes in $\phi(\mathbf{r},t)$ and $\mathbf{A}(\mathbf{r},t)$ are, unsurprisingly, called gauge transformations; choosing a specific pair of potentials is usually referred to as gauge fixing. In problems involving uniform electric fields, the gauge is fixed so as to express the electric field in terms of either the scalar potential,

$$\mathbf{E}(t) = -\nabla\phi(\mathbf{r},t), \tag{3.7}$$

[10]The magnetic fields of laser pulses are usually considered to be weak and therefore irrelevant to the description of NLO responses. A laser with a beam intensity of 2.5 kW/cm^2 produces electric and magnetic fields of amplitudes (in vacuum) 1.3 kV/cm^{-1}, and 0.4 mT, respectively. Magnetic effects tend to be considered when the amplitude of the magnetic field is of the order of 1T, e.g. ref.[129].

or the vector potential,

$$\mathbf{E}(t) = -\frac{\partial \mathbf{A}(t)}{\partial t}, \tag{3.8}$$

and the Hamiltonian of a system in an external electric field has therefore two corresponding (and equivalent) formulations.[11] For simplicity, consider a system of free electrons described by the usual continuum Hamiltonian, expressed in terms of second-quantized field operators [147, 148],

$$H_0(t) = \int d^d\mathbf{r}\, \Psi^\dagger(\mathbf{r}) \mathcal{H}_0\left(\mathbf{r}, \frac{\nabla}{i}\right) \Psi(\mathbf{r}), \tag{3.9}$$

for,

$$\mathcal{H}_0\left(\mathbf{r}, \frac{\nabla}{i}\right) = \frac{\hbar^2}{2m_e}\left(\frac{\nabla}{i}\right)^2. \tag{3.10}$$

The fermionic field, $\Psi(\mathbf{r})$, and its Hermitian conjugate, $\Psi^\dagger(\mathbf{r})$, satisfy the usual anti-commutation relations,

$$\left\{\Psi(\mathbf{r}), \Psi(\mathbf{r}')\right\} = 0, \tag{3.11}$$

$$\left\{\Psi^\dagger(\mathbf{r}), \Psi^\dagger(\mathbf{r}')\right\} = 0, \tag{3.12}$$

$$\left\{\Psi(\mathbf{r}), \Psi^\dagger(\mathbf{r}')\right\} = \delta(\mathbf{r} - \mathbf{r}'). \tag{3.13}$$

The interaction with the external electric field is described, in the case of Eq.(3.7), by the dipole interaction term, $-|e|\,\phi(\mathbf{r}, t) \to |e|\,\mathbf{E}(t)\cdot\mathbf{r}$, which means that the full Hamiltonian is,

$$H_E(t) = \int d^d\mathbf{r}\, \Psi^\dagger(\mathbf{r}) \left[\mathcal{H}_0\left(\mathbf{r}, \frac{\nabla}{i}\right) + |e|\,\mathbf{E}(t)\cdot\mathbf{r}\right] \Psi(\mathbf{r}). \tag{3.14}$$

In the case where the field is represented by the vector potential, Eq.(3.8), the interaction enters the Hamiltonian by means of the minimal coupling: the momentum operator, $\mathbf{p}$, similarly to its classical counterpart, is simply shifted by the vector potential (times the absolute value of the electron charge), $|e|\,\mathbf{A}(t)$. For a continuum Hamiltonian in the position representation — where $\mathbf{p}$ is expressed in terms of the gradient operator — the full Hamiltonian can be written as,

$$H_A(t) = \int d^d\mathbf{r}\, \Psi^\dagger(\mathbf{r}) \left[\mathcal{H}_0\left(\mathbf{r}, \frac{\nabla}{i} + \frac{|e|}{\hbar}\mathbf{A}(t)\right)\right] \Psi(\mathbf{r}), \tag{3.15}$$

for,

$$\mathcal{H}_0\left(\mathbf{r}, \frac{\nabla}{i} + \frac{|e|}{\hbar}\mathbf{A}(t)\right) = \frac{\hbar^2}{2m_e}\left(\frac{\nabla}{i}\right)^2 + \frac{\hbar\,|e|}{2m_e}\mathbf{A}(t)\cdot\left(\frac{\nabla}{i}\right) + \frac{e^2}{2m_e}\mathbf{A}(t)\cdot\mathbf{A}(t). \tag{3.16}$$

[11] A careful derivation of these expressions, based on the Lagrangian formulation of classical mechanics, can be found in [4].

Despite their apparent differences, these two Hamiltonians must represent the exact same physical system, as is imposed by gauge invariance. It is, therefore, useful to study the relation between these two descriptions in quantitative terms, that is, by considering the dynamics of an arbitrary state vector in each description; this provides us with certain important insights that clarify the relation between gauges.

The many-body state vector in the scalar potential approach, $|\psi(t)\rangle$, evolves in time according to $H_E(t)$,

$$i\hbar\frac{\partial\,|\psi(t)\rangle}{\partial t} = H_E(t)\,|\psi(t)\rangle\,, \tag{3.17}$$

or, explicitly,

$$i\hbar\frac{\partial\,|\psi(t)\rangle}{\partial t} = H_0(t)\,|\psi(t)\rangle + |e|\,\mathbf{E}(t)\cdot\int d^d\mathbf{r}\,\mathbf{r}\,\rho(\mathbf{r})\,|\psi(t)\rangle\,. \tag{3.18}$$

We have used the definition of the density operator, $\rho(\mathbf{r}) = \Psi^\dagger(\mathbf{r})\Psi(\mathbf{r})$. If we assume that $|\psi(t)\rangle$ is related to some other state vector, $\left|\bar{\psi}(t)\right\rangle$, by means of a time-dependent unitary transformation, $\mathcal{U}(t)\mathcal{U}^\dagger(t) = 1$, whose form is to be determined,

$$|\psi(t)\rangle = \mathcal{U}(t)\left|\bar{\psi}(t)\right\rangle, \tag{3.19}$$

we then can write Eq.(3.18) as,

$$i\hbar\mathcal{U}(t)\frac{\partial\left|\bar{\psi}(t)\right\rangle}{\partial t} + i\hbar\frac{\partial\mathcal{U}(t)}{\partial t}\left|\bar{\psi}(t)\right\rangle = H_0(t)\mathcal{U}(t)\left|\bar{\psi}(t)\right\rangle + |e|\,\mathbf{E}(t)\cdot\int d^d\mathbf{r}\,\mathbf{r}\,\rho(\mathbf{r})\,\mathcal{U}(t)\left|\bar{\psi}(t)\right\rangle, \tag{3.20}$$

which, by acting with $\mathcal{U}^\dagger(t)$ on the left and reordering the terms, can be expressed as,

$$i\hbar\frac{\partial\left|\bar{\psi}(t)\right\rangle}{\partial t} = \left[\mathcal{U}^\dagger(t)H_0(t)\mathcal{U}(t) - i\hbar\mathcal{U}^\dagger(t)\frac{\partial\mathcal{U}(t)}{\partial t}\right]\left|\bar{\psi}(t)\right\rangle + |e|\,\mathbf{E}(t)\cdot\int d^d\mathbf{r}\,\mathbf{r}\,\mathcal{U}^\dagger(t)\rho(\mathbf{r})\,\mathcal{U}(t)\left|\bar{\psi}(t)\right\rangle. \tag{3.21}$$

We can see that Eq.(3.21) describes the time evolution of the state vector $\left|\bar{\psi}(t)\right\rangle$. For it to be a state vector evolving according to the Hamiltonian in the vector potential approach, Eq.(3.15),

$$i\hbar\frac{\partial\left|\bar{\psi}(t)\right\rangle}{\partial t} = H_A(t)\left|\bar{\psi}(t)\right\rangle, \tag{3.22}$$

we must require the two Hamiltonians to be the same,

$$H_A(t) = \left[\mathcal{U}^\dagger(t)H_0(t)\mathcal{U}(t) - i\hbar\mathcal{U}^\dagger(t)\frac{\partial\mathcal{U}(t)}{\partial t}\right] + |e|\,\mathbf{E}(t)\cdot\int d^d\mathbf{r}\,\mathbf{r}\,\mathcal{U}^\dagger(t)\rho(\mathbf{r})\,\mathcal{U}(t). \tag{3.23}$$

The last term in right-hand side of Eq.(3.23) is the dipole interaction term, under the unitary transformation. We can thus write,

$$H_A(t) = \mathcal{U}^\dagger(t) H_E(t) \mathcal{U}(t) - i\hbar \mathcal{U}^\dagger(t) \frac{\partial \mathcal{U}(t)}{\partial t}. \tag{3.24}$$

The first term in the right-hand side of Eq.(3.24) is now the Hamiltonian in the length gauge. To complete this derivation, we need only to determine the form of the unitary transformation. For simplicity, we assume that the generator of this transformation is the density operator itself, $\mathcal{U}(t) = \exp\left[i \int d^d\mathbf{r}\, \theta(\mathbf{r}, t)\, \rho(\mathbf{r})\right]$, for a $\theta(\mathbf{r}, t)$ that ensures that the transformation is unitary. First, note that the second term in the right-hand side of Eq.(3.24) must be expressed in terms of the position, $\mathbf{r}$, so as to cancel out the dipole interaction term. Secondly, since it involves the time derivative of $\mathcal{U}(t)$, $\theta(\mathbf{r}, t)$ cannot be a function of the dot product of $\mathbf{r}$ and the electric field. It must instead be a function of the dot product of $\mathbf{r}$ and the *vector potential*, since its time derivative is, by definition, the electric field (with an additional minus sign), Eq.(3.8). For $\theta(\mathbf{r}, t) = \lambda \mathbf{A}(t) \cdot \mathbf{r}$, we obtain,

$$-i\hbar \mathcal{U}^\dagger(t) \frac{\partial \mathcal{U}(t)}{\partial t} = -\hbar\lambda \mathbf{E}(t) \cdot \int d^d\mathbf{r}\, \mathbf{r}\, \rho(\mathbf{r}), \tag{3.25}$$

which, for $\lambda = |e|/\hbar$, is the dipole interaction term but with the opposite sign. The unitary transformation is now fully defined,

$$\mathcal{U}(t) = \exp\left[i \frac{|e|}{\hbar} \mathbf{A}(t) \cdot \int d^d\mathbf{r}\, \mathbf{r}\, \rho(\mathbf{r})\right]. \tag{3.26}$$

We can now return to Eq.(3.23), and note that the velocity-gauge Hamiltonian is directly related to the unperturbed Hamiltonian by means of this unitary transformation,

$$H_A(t) = \mathcal{U}^\dagger(t) H_0(t) \mathcal{U}(t). \tag{3.27}$$

From this, we can see that the minimal coupling, that is, the momentum shift, $\mathbf{p} \to \mathbf{p} + |e|\, \mathbf{A}(t)$, is indeed a translation in momentum space generated by the position operator, here in expressed in terms of second-quantized field operators; as the momentum operator is the generator of real space translations, the position operator is the generator of translations in momentum space.

Finally, we point out that, since the state vectors in each gauge are related by a unitary transformation, the *observables* in the two gauges must themselves be related, so that their expectation values can be the same,

$$O_A(t) = \mathcal{U}^\dagger(t)\, O_E\, \mathcal{U}(t), \tag{3.28}$$

which, when expanded to first order, becomes,

$$O_A(t) = O_E - i\frac{|e|}{\hbar}\mathbf{A}(t)\cdot\int d^d\mathbf{r}\,\mathbf{r}\left[\rho(\mathbf{r}), O_E\right] + (...). \tag{3.29}$$

Note that there is an exception to the relation between observables: the Hamiltonian. The unitary transformation, $\mathcal{U}(t)$, is itself time dependent, meaning that the generator of time evolutions in the velocity gauge, $H_A(t)$, is not simply $\mathcal{U}^\dagger(t)\,H_E(t)\,\mathcal{U}(t)$; it must include an additional term involving the time derivative of $\mathcal{U}(t)$, Eq.(3.26), as expressed in Eq.(3.24).

The existence of a unitary transformation between the two descriptions of the electric field ensures their complete equivalence [22].

3.3 Elements of Condensed Matter Theory

Having discussed in the previous chapter how the point group symmetries of a crystal impose certain general properties on its nonlinear response functions, we now move on to the study of the physics of crystal electrons, as follows from the standpoint of traditional solid state physics [135, 149]. We begin by considering Bloch's theorem and its implications; from this, we move on to the aforementioned discussion on the representation of the position operator — and the introduction of the covariant derivative — as well as its application in the context of matrices in band-index space. This, as the reader will see, provides us with all the elements we need to formulate the length- and velocity-gauge descriptions in the density matrix formalism.

The single-particle Hamiltonian of an electron in a crystal is an immediate extension of the free-particle Hamiltonian we encountered in the previous section, to which we add the crystal potential, $V(\mathbf{r})$,

$$\mathcal{H} = \frac{\hbar^2}{2m_e}\left(\frac{\nabla}{i}\right)^2 + V(\mathbf{r}). \tag{3.30}$$

This additional potential term, $V(\mathbf{r})$, has the periodicity of the crystal, i.e., $V(\mathbf{r}) = V(\mathbf{r}+\mathbf{R})$, for $\mathbf{R}$ any Bravais lattice vector; for the example from the previous chapter — the square lattice with one atom per unit cell — the corresponding crystal potential has to be such that any translation that connects any two atoms, i.e., a linear combination (with integer coefficients) of the two translations connecting neighbouring atoms along the x and y directions, leaves it unchanged. The eigenfunctions of the crystal Hamiltonian, Eq.(3.30), must therefore carry these symmetry properties. According to Bloch's theorem, these eigenfunctions have the form of a plane wave times a periodic function, $u_{\mathbf{k}s}(\mathbf{r}) = u_{\mathbf{k}s}(\mathbf{r}+\mathbf{R})$,

$$\psi_{\mathbf{k}s}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}s}(\mathbf{r}), \tag{3.31}$$

so that they only pick up a phase factor under lattice translations, $\psi_{\mathbf{k}s}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi_{\mathbf{k}s}(\mathbf{r})$. The

single-particle Schrödinger equation is then expressed as,

$$\mathcal{H}\psi_{\mathbf{k}s}(\mathbf{r}) = \epsilon_{\mathbf{k}s}\psi_{\mathbf{k}s}(\mathbf{r}). \tag{3.32}$$

The solutions to the eigenvalue problem, $\epsilon_{\mathbf{k}s}$, define continuous functions of $\mathbf{k}$ in the FBZ, and are usually called bands. It also follows from Eqs.(3.30) through (3.32) that the eigenvalue problem can be expressed directly in terms of the periodic part of the Bloch functions, $u_{\mathbf{k}s}(\mathbf{r})$,

$$\left[\frac{\hbar^2}{2m_e}\left(\frac{\nabla}{i}\right)^2 + V(\mathbf{r})\right] e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}s}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\left[\frac{\hbar^2}{2m_e}\left(\frac{\nabla}{i} + \mathbf{k}\right)^2 + V(\mathbf{r})\right] u_{\mathbf{k}s}(\mathbf{r}), \tag{3.33}$$

since, by acting with $e^{-i\mathbf{k}\cdot\mathbf{r}}$ on the left, we obtain,

$$\left[\frac{\hbar^2}{2m_e}\left(\frac{\nabla}{i} + \mathbf{k}\right)^2 + V(\mathbf{r})\right] u_{\mathbf{k}s}(\mathbf{r}) = \epsilon_{\mathbf{k}s} u_{\mathbf{k}s}(\mathbf{r}). \tag{3.34}$$

To this new Hamiltonian, which acts *solely* on $u_{\mathbf{k}s}(\mathbf{r})$, we call $\mathcal{H}_{\mathbf{k}}$,

$$\mathcal{H}_{\mathbf{k}} = \left[\frac{\hbar^2}{2m_e}\left(\frac{\nabla}{i} + \mathbf{k}\right)^2 + V(\mathbf{r})\right], \tag{3.35}$$

which allows us to write,

$$\mathcal{H}_{\mathbf{k}} u_{\mathbf{k}s}(\mathbf{r}) = \epsilon_{\mathbf{k}s} u_{\mathbf{k}s}(\mathbf{r}). \tag{3.36}$$

Note that $\mathcal{H}_{\mathbf{k}}$ is related to the original Hamiltonian, Eq.(3.30), by means of the transformation:

$$\mathcal{H}_{\mathbf{k}} := e^{-i\mathbf{k}\cdot\mathbf{r}}\,\mathcal{H}\,e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{3.37}$$

We have thus moved from an eigenvalue/eigenfunction problem that is defined in terms of functions whose domain is the entire crystal to one defined in terms of functions whose domain is the real-space unit cell and that satisfy the boundary condition noted above Eq.(3.31). In this new eigenvalue/eigenfunction problem, each $\mathbf{k}$ point in the first Brillouin zone (FBZ) is associated with a specific Hamiltonian operator, $\mathcal{H}_{\mathbf{k}}$; the eigenfunctions of this Hamiltonian at each $\mathbf{k}$ point, $\{|u_{\mathbf{k}s}\rangle\,,\ s = 0, 1, \dots\}$, form a complete basis of functions defined in the real space unit cell and are, of course, unit-cell periodic. The eigenfunctions for different values of $\mathbf{k}$, which are associated with different Hamiltonians, correspond to different sets of eigenvalues. Moreover, since $\mathbf{k}$ is treated strictly as a parameter — and therefore not subjected to the conditions imposed by boundary conditions set on the crystal[12] — the eigenvalue/eigenfunction problem expressed in Eq.(3.36) is not

[12]For periodic boundary conditions, this would mean restricting the number of points in the FBZ to the total number of cells in the crystal [149]. For the square lattice, these points would be distributed according to the usual formulae: $k_x = \frac{2\pi N_x}{a} n_x$, $n_x = 0, 1..., N_x - 1$, and $k_y = \frac{2\pi N_y}{a} n_y$, $n_y = 0, 1..., N_y - 1$ for N_x and N_y the number of cells

restricted to a specific set of points in the FBZ and can, instead, be solved for any $\mathbf{k}$ in that domain [149]. Taking derivatives with respect to $\mathbf{k}$ of $u_{\mathbf{k}s}(\mathbf{r})$ and $\epsilon_{\mathbf{k}s}$ is, therefore, a legitimate procedure, even when the periodic boundary conditions impose a discretization of the values of $\mathbf{k}$. This, as we will presently see, is different from taking the derivative of the full Bloch function; only in the infinite crystal limit does this difference vanish.

We have noted that the $u_{\mathbf{k}s}(\mathbf{r})$ form a complete basis to a specific Hamiltonian operator, $\mathcal{H}_{\mathbf{k}}$; we can then assume this basis to be orthonormal, with an inner product defined as the integral over the real space unit cell (of volume v_c),

$$\langle u_{\mathbf{k}s}|u_{\mathbf{k}s'}\rangle = \frac{1}{v_c}\int_{uc} d^d\mathbf{r}\, u^*_{\mathbf{k}s}\left(\mathbf{r}\right) u_{\mathbf{k}s'}\left(\mathbf{r}\right) = \delta_{ss'}. \tag{3.38}$$

Unlike the inner product between Bloch functions of different $\mathbf{k}$-points — which is defined over the entire crystal — the inner product between $u_{\mathbf{k}s}(\mathbf{r})$ functions associated with different $\mathbf{k}$ parameters is not trivial. Consider the following inner product between eigenstates associated with different (but, in principle, very similar) eigenvalues from the same band, $\langle u_{\mathbf{k}s}|u_{\mathbf{k}'s}\rangle$, where $\mathbf{k}'$ is assumed to be arbitrarily close to $\mathbf{k}$, $\mathbf{k}' = \mathbf{k} + \Delta\mathbf{k}$,

$$\langle u_{\mathbf{k}s}|u_{\mathbf{k}+\Delta\mathbf{k}s}\rangle = \frac{1}{v_c}\int_{uc} d^d\mathbf{r}\, u^*_{\mathbf{k}s}\left(\mathbf{r}\right) u_{\mathbf{k}+\Delta\mathbf{k}s}\left(\mathbf{r}\right). \tag{3.39}$$

The $u_{\mathbf{k}+\Delta\mathbf{k}s}\left(\mathbf{r}\right)$ function can be readily expanded in powers of $\Delta\mathbf{k}$,

$$u_{\mathbf{k}+\Delta\mathbf{k}s}\left(\mathbf{r}\right) = u_{\mathbf{k}s}\left(\mathbf{r}\right) + \Delta\mathbf{k}\cdot\nabla_{\mathbf{k}}u_{\mathbf{k}s}\left(\mathbf{r}\right) + (...). \tag{3.40}$$

Replacing this in Eq.(3.39), gives us,

$$\langle u_{\mathbf{k}s}|u_{\mathbf{k}+\Delta\mathbf{k}s}\rangle = 1 + \Delta\mathbf{k}\cdot\int_{uc} d^d\mathbf{r}\, u^*_{\mathbf{k}s}\left(\mathbf{r}\right)\nabla_{\mathbf{k}}u_{\mathbf{k}s}\left(\mathbf{r}\right) + (...), \tag{3.41}$$

$$= 1 - i\Delta\mathbf{k}\cdot\boldsymbol{\xi}_{\mathbf{k}ss} + (...), \tag{3.42}$$

and the relation between two arbitrarily close states from the same band, is now expressed in terms of a new object, the Berry connection:

$$\boldsymbol{\xi}_{\mathbf{k}ss'} := i\langle u_{\mathbf{k}s}|\nabla_{\mathbf{k}}u_{\mathbf{k}s'}\rangle, \tag{3.43}$$

$$= i\int_{uc} d^d\mathbf{r}\, u^*_{\mathbf{k}s}\left(\mathbf{r}\right)\nabla_{\mathbf{k}}u_{\mathbf{k}s'}\left(\mathbf{r}\right). \tag{3.44}$$

This connection relates two states that are arbitrarily close (in $\mathbf{k}$-space), with its different components translating the differences between infinitesimal displacements in different (orthogonal)

in the x and y directions, respectively, and a the distance between two neighbouring atoms (usually called the lattice parameter). Note that this affects only the plane wave factor of the Bloch functions.

directions; note that s and s' can be any two band indexes, meaning that the Berry connection can link states in the same band, $s = s'$ — these elements are usually called *intraband* terms — as well as link states from different bands, $s \neq s'$ — these elements are usually called *interband* terms. It is, therefore, a matrix in band-index space. Furthermore, since the set of $u_{\mathbf{k}s}(\mathbf{r})$ functions (with s running through every band) constitutes a complete basis, the knowledge of the Berry connection can, in principle, be used to build the eigenfunctions at another, arbitrary $\mathbf{k}'$, by piecing together the different Berry connection contributions along the way.[13] This, we must add, is not a particularly practical point — we are usually interested in determining the Berry connection from the eigenstates, not the other way around — but it is an interesting one nonetheless. Let us now turn to the matrix elements of the position operator.

3.3.1 The Covariant Derivative

There are certain difficulties that come with determining the Bloch matrix elements of the position operator. These first become evident when we consider a specific crystal, e.g., a one-dimensional lattice with one atom per unit cell; in periodic boundary conditions, this is simply a chain, Figure 3.1. We can see from that picture that, if we assign positions to the atoms of the chain — by taking the appropriate number of increments with respect to the origin, $x = 0$ — once we reach our starting point, the difference between the position of the first atom and the position of the final atom is *not* their actual distance. The position operator is, therefore, ill-defined in periodic boundary conditions.

Consider, however, the infinite-crystal limit of a d-dimensional crystal, i.e., its length in a d number of orthogonal directions is taken to infinity, $L_i \to \infty$.[14] The eigenstates of the Hamiltonian, the Bloch functions, are no longer constrained by conditions imposed on the plane wave portion of its expression, meaning that we can take derivatives of these functions as a whole. It then follows from Eq.(3.31) that,

$$\langle \psi_{\mathbf{k}s} | \, \mathbf{r} \, | \psi_{\mathbf{k}'s'} \rangle = \int d^d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} u^*_{\mathbf{k}s}(\mathbf{r}) \; \mathbf{r} \, e^{i\mathbf{k}'\cdot\mathbf{r}} u_{\mathbf{k}'s'}(\mathbf{r}) \,, \tag{3.45}$$

$$= \int d^d\mathbf{r} \, e^{-i\mathbf{k}\cdot\mathbf{r}} u^*_{\mathbf{k}s}(\mathbf{r}) \left(\nabla_{\mathbf{k}'} e^{i\mathbf{k}'\cdot\mathbf{r}}\right) u_{\mathbf{k}'s'}(\mathbf{r}) \,. \tag{3.46}$$

[13] In quantitative terms, it means that for $\mathbf{k} \approx \mathbf{k}'$, we can write,

$$|u_{\mathbf{k}'s}\rangle = \sum_r |u_{\mathbf{k}r}\rangle \langle u_{\mathbf{k}r}|u_{\mathbf{k}'s}\rangle \to |u_{\mathbf{k}'s}\rangle \approx |u_{\mathbf{k}s}\rangle - i\Delta\mathbf{k} \cdot \sum_r \boldsymbol{\xi}_{\mathbf{k}rs} |u_{\mathbf{k}r}\rangle \,,$$

and $\Delta\mathbf{k} = \mathbf{k}' - \mathbf{k}$. This relation between $u_{\mathbf{k}s}$ states in different points of the FBZ can then be employed iteratively, which allows one to build — at least in principle — the eigenstates for any other value of $\mathbf{k}$; this is, of course, contingent on the knowledge of the Berry connection across the entire FBZ for all possible pairs of band-indexes.

Note that we have used the completeness relation for $u_{\mathbf{k}s}$ states, $\sum_r |u_{\mathbf{k}r}\rangle \langle u_{\mathbf{k}r}| = \hat{1}$.

[14] This entails several consequences, the most significant of which being that sums over the FBZ become integrals over the FBZ. These will be considered at the end of the section.

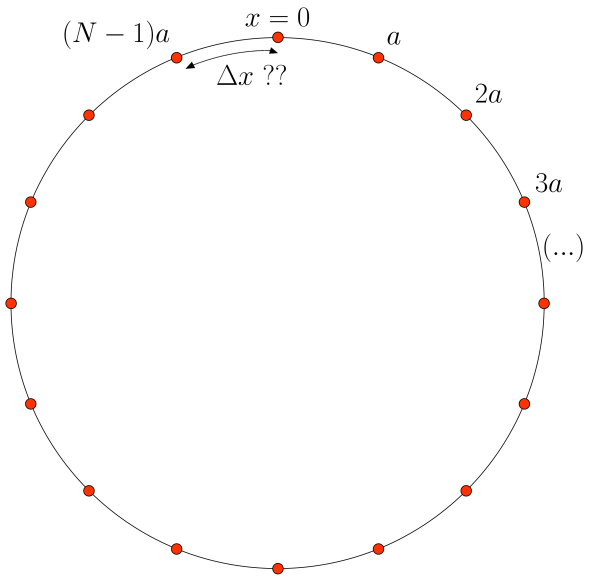


Figure 3.1: The position operator is ill-defined in periodic boundary conditions. When we assign positions to atoms along a chain, we see that, once we reach our starting point, the difference between the positions of the first and final atoms is *not* their actual distance.

After some manipulations — which involve using the chain rule, $\big(\nabla_{\mathbf{k}'} f(\mathbf{k}')\big) g(\mathbf{k}') = \big(\nabla_{\mathbf{k}'} f(\mathbf{k}') g(\mathbf{k}')\big) - f(\mathbf{k}')\big(\nabla_{\mathbf{k}'} g(\mathbf{k}')\big)$, for $f(\mathbf{k}') = e^{i\mathbf{k}'\cdot\mathbf{r}}$ and $g(\mathbf{k}') = u_{\mathbf{k}'s'}(\mathbf{r})$ — and using the definition of the Berry connection, Eq.(3.44), one can show that the Bloch matrix element of the position operator can be expressed as [1, 18, 19],

$$\mathbf{r}_{\mathbf{k}\mathbf{k}'ss'} = \delta_{ss'}(2\pi)^d(-i)\nabla_{\mathbf{k}'}\delta(\mathbf{k}'-\mathbf{k}) + (2\pi)^d\delta(\mathbf{k}'-\mathbf{k})\,\boldsymbol{\xi}_{\mathbf{k}'ss'}. \tag{3.47}$$

The first term is, as expected, equal to the matrix elements of the position operator in a basis of plane waves (in the case of $L_i \to \infty$); the second term is simply a Dirac delta function times the Berry connection. This expression is, however, somewhat opaque, so it is useful to cast this awkward looking expression in a more transparent form. For a continuous non-normalizable basis, the matrix elements of an operator are the kernel of an *integral* transform. Consider the matrix element $\langle\psi_{\mathbf{k}s}|\,\mathbf{r}\,|\Psi\rangle$, for $\Psi(\mathbf{r})$, a general single-particle state that can be expressed as a linear combination of Bloch functions,

$$\Psi(\mathbf{r}) = \sum_s \int \frac{d^d\mathbf{k}}{(2\pi)^d}\Phi_s(\mathbf{k})\,\psi_{\mathbf{k}s}(\mathbf{r}), \tag{3.48}$$

with coefficients $\Phi_s(\mathbf{k}) = \langle\psi_{\mathbf{k}s}|\Psi\rangle$. The matrix element $\langle\psi_{\mathbf{k}s}|\,\mathbf{r}\,|\Psi\rangle$ or, in other words, the state

$\mathbf{r}\,|\Psi\rangle$ in the Bloch representation, is simply,

$$\langle\psi_{\mathbf{k}s}|\,\mathbf{r}\,|\Psi\rangle = \sum_{s'}\int\frac{d^d\mathbf{k}'}{(2\pi)^d}\Phi_{s'}\left(\mathbf{k}'\right)\int d^d\mathbf{r}\psi^*_{\mathbf{k}'s'}(\mathbf{r})\,\mathbf{r}\,\psi_{\mathbf{k}s}(\mathbf{r}), \tag{3.49}$$

$$= \sum_{s'}\int\frac{d^d\mathbf{k}'}{(2\pi)^d}\mathbf{r}_{\mathbf{k}\mathbf{k}',ss'}\Phi_{s'}\left(\mathbf{k}'\right), \tag{3.50}$$

$$=(-i)\int d^d\mathbf{k}'\big(\nabla_{\mathbf{k}'}\delta(\mathbf{k}'-\mathbf{k})\big)\Phi_s\left(\mathbf{k}'\right)+\sum_{s'}\boldsymbol{\xi}_{\mathbf{k}ss'}\Phi_{s'}\left(\mathbf{k}\right), \tag{3.51}$$

The first term now includes an integral over $\mathbf{k}'$. We can therefore integrate the intraband term by parts — since this integration is performed over the FBZ, and given the periodicity of any function of $\mathbf{k}$, $\phi(\mathbf{k})=\phi(\mathbf{k}+\mathbf{G})$, there can be no additional surface terms — so as to take the derivative with respect to $\mathbf{k}'$ from the Dirac delta function to the $\Phi_s\left(\mathbf{k}'\right)$ coefficient.[15] This frees up the Dirac delta function, which can then be integrated over. The Bloch representation of $\mathbf{r}\,|\Psi\rangle$ finally reads as,

$$\langle\psi_{\mathbf{k}s}|\,\mathbf{r}\,|\Psi\rangle = i\sum_{s'}\left(\delta_{ss'}\nabla_{\mathbf{k}}-i\boldsymbol{\xi}_{\mathbf{k}ss'}\right)\Phi_{s'}\left(\mathbf{k}\right). \tag{3.52}$$

We can now define the covariant derivative operator,

$$\boxed{\mathbf{D}_{\mathbf{k}ss'}:=\delta_{ss'}\nabla_{\mathbf{k}}-i\boldsymbol{\xi}_{\mathbf{k}ss'}}, \tag{3.53}$$

such that,

$$\langle\psi_{\mathbf{k}s}|\,\mathbf{r}\,|\Psi\rangle=\sum_{s'}i\mathbf{D}_{\mathbf{k}ss'}\Phi_{s'}\left(\mathbf{k}\right). \tag{3.54}$$

The position operator in the Bloch representation is the covariant derivative (times a factor of i), $i\mathbf{D}_{\mathbf{k}ss'}$ [19]. The designation of *covariant* follows from its behaviour under a local gauge transformation in momentum space, that is, the inclusion of an arbitrary phase factor in the eigenfunctions of the Hamiltonian, Eq.(3.35),

$$u_{\mathbf{k}s}\to e^{i\theta_s(\mathbf{k})}\,u_{\mathbf{k}s}. \tag{3.55}$$

For the covariant derivative, such a transformation,[16]

$$\mathbf{D}_{\mathbf{k}ss'}\to\tilde{\mathbf{D}}_{\mathbf{k}ss'}:=e^{-i\theta_s(\mathbf{k})}\mathbf{D}_{\mathbf{k}ss'}e^{i\theta_{s'}(\mathbf{k})}, \tag{3.56}$$

$$\to\tilde{\mathbf{D}}_{\mathbf{k}ss'}=e^{-i(\theta_s(\mathbf{k})-\theta_{s'}(\mathbf{k}))}\mathbf{D}_{\mathbf{k}ss'}, \tag{3.57}$$

[15] We note that this argument is repeated in footnote 25, where it is presented in quantitative terms.

[16] The title of the paper on which this chapter is based, *Gauge Covariances and Nonlinear Optical Responses*, follows from the existence of this second set of gauge transformations — the first one is, of course, associated with the ability to choose different representations of the same electric field, Eqs.(3.5) and (3.6). Since these additional arbitrary phases cannot, in any way whatsoever, affect the response of the system, the expressions for the nonlinear optical coefficients must be invariant under this type of transformation. We will return to this point later on, when we discuss the second order response of the gapped graphene monolayer.

does not pick up any additional gradient terms — unlike the regular derivative, $\nabla_{\mathbf{k}}$. The term that involves the gradient acting on the phase factor on the right, $-\nabla_{\mathbf{k}}\theta_{s'}(\mathbf{k})$ is cancelled by the term picked up by the transformation of the Berry connection:

$$i\langle u_{\mathbf{k}s}|\nabla_{\mathbf{k}}u_{\mathbf{k}s'}\rangle \to i\langle e^{-i\theta_s(\mathbf{k})}u_{\mathbf{k}s}|\left(\nabla_{\mathbf{k}}e^{i\theta_{s'}(\mathbf{k})}u_{\mathbf{k}s'}\right)\rangle = \delta_{ss'}\nabla_{\mathbf{k}}\theta_s(\mathbf{k}) + ie^{-i\left(\theta_s(\mathbf{k})-\theta_{s'}(\mathbf{k})\right)}\langle u_{\mathbf{k}s}|\nabla_{\mathbf{k}}u_{\mathbf{k}s'}\rangle, \tag{3.58}$$

as follows from its definition, Eq.(3.43). This is, therefore, in complete parallel with the definition of the covariant derivative in gauge theories in real space; the covariant derivative can thus be considered from the perspective of differential geometry, as is done in Appendix A of ref.[40]. Finally, let us emphasize the point that since the covariant derivative carries a derivative with respect to $\mathbf{k}$, it must always act *on* something: in the case of the the arbitrary wave function, it acts on the $\Phi_s(\mathbf{k})$ coefficients; in the case of the equations describing the dynamics of the density matrix in the length-gauge description, as we will see later on, it acts on the elements of the density matrix itself.

3.3.2 Matrices in Band-index Space

At the beginning of this section, we saw that the concept of a basis of periodic functions follows quite naturally from a formulation of the eigenvalue/eigenfunction problem that treats $\mathbf{k}$ simply as a parameter, and that from there, one can also derive the object connecting the periodic functions of different, but nearby, parameters: the Berry connection. The Hamiltonian and the Berry connection are, however, but two examples of matrices that can be expressed in that basis; in principle, any operator whose matrix elements in the Bloch basis are diagonal in $\mathbf{k}$ — that is, they can be expressed in terms of a Dirac delta function times a number (or a vector), akin to the matrix elements of the Hamiltonian (or the interband portion of the matrix elements of the position operator) — can be expressed in similar terms. It is, therefore, useful to take these objects into close consideration.

A matrix in the space generated by the basis of periodic functions, $\left\{|u_{\mathbf{k}s}\rangle\,,\, s=1,2,\dots\right\}$, parametrized by $\mathbf{k}$, can be cast in operator form by writing,

$$\mathcal{O}(\mathbf{k}) = \sum_{s's} |u_{\mathbf{k}s}\rangle\, \mathcal{O}_{\mathbf{k}ss'}\, \langle u_{\mathbf{k}s'}|\,, \tag{3.59}$$

$$\mathcal{O}_{\mathbf{k}ss'} = \langle u_{\mathbf{k}s}|\, \mathcal{O}(\mathbf{k})\, |u_{\mathbf{k}s'}\rangle\,, \tag{3.60}$$

which, for the aforementioned operators, it means that,

$$\mathcal{H}(\mathbf{k}) = \sum_{s's} |u_{\mathbf{k}s}\rangle\, \epsilon_{\mathbf{k}s}\, \langle u_{\mathbf{k}s}|\,, \tag{3.61}$$

$$\boldsymbol{\xi}(\mathbf{k}) = \sum_{s's} |u_{\mathbf{k}s}\rangle\, \boldsymbol{\xi}_{\mathbf{k}ss'}\, \langle u_{\mathbf{k}s'}|\,. \tag{3.62}$$

Just as with their bases, operators associated with different parameters are usually treated inde-

pendently from one another. Whenever we consider products of these operators, or, alternatively, their trace, we think of these as functions of a single parameter, $\mathbf{k}$. An example of this is the electric current, Eqs.(3.94) and (3.95). There is, however, an important caveat to this observation. We have noted that the bases of periodic functions for two neighbouring $\mathbf{k}$ points are connected by means of the Berry connection, $|u_{\mathbf{k}s}\rangle \to |u_{\mathbf{k}'s}\rangle$; as for the matrix elements, we know that they can be seamlessly expanded around a different value of $\mathbf{k}$, $\mathcal{O}_{\mathbf{k}ss'} \to \mathcal{O}_{\mathbf{k}'ss'}$, provided that they are a well-behaved, continuous function. Since the bases — and the matrix elements — of different $\mathbf{k}$ points can be related, it is therefore also possible, at least in principle, to connect *operators* that are associated with different parameters as well, $\mathcal{O}(\mathbf{k}) \to \mathcal{O}(\mathbf{k}')$. This point will be considered in a later section, when we return to the subject of gauge invariance. For now, let us move on to a discussion of two other points concerning operators of a fixed parameter: first, the relation between $\mathbf{k}$-derivatives of an operator and the covariant derivative; secondly, the expression for the velocity operator in the basis of periodic functions.

The matrix elements of the $\mathbf{k}$-derivative of $\mathcal{O}(\mathbf{k})$ can be written as,

$$\left[\nabla_{\mathbf{k}}\mathcal{O}(\mathbf{k})\right]_{ss'} := \langle u_{\mathbf{k}s}| \nabla_{\mathbf{k}}\mathcal{O}(\mathbf{k}) |u_{\mathbf{k}s'}\rangle , \tag{3.63}$$

$$= \nabla_{\mathbf{k}} \langle u_{\mathbf{k}s}| \mathcal{O}(\mathbf{k}) |u_{\mathbf{k}s'}\rangle - \langle \nabla_{\mathbf{k}} u_{\mathbf{k}s}| \mathcal{O}(\mathbf{k}) |u_{\mathbf{k}s'}\rangle - \langle u_{\mathbf{k}s}| \mathcal{O}(\mathbf{k}) |\nabla_{\mathbf{k}} u_{\mathbf{k}s'}\rangle . \tag{3.64}$$

The first term in the right-hand side is the $\mathbf{k}$-gradient of the matrix element, $\nabla_{\mathbf{k}}\mathcal{O}_{\mathbf{k}ss'}$. The other two can be expressed in terms of the Berry connection, Eq.(3.43), by application of the completeness relation,[17]

$$\sum_r |u_{\mathbf{k}r}\rangle \langle u_{\mathbf{k}r}| = \hat{1}, \tag{3.65}$$

the chain rule and the fact that $\nabla_{\mathbf{k}}\langle u_{\mathbf{k}s}|u_{\mathbf{k}s'}\rangle = 0$,

$$\langle \nabla_{\mathbf{k}} u_{\mathbf{k}s}| \mathcal{O}(\mathbf{k}) |u_{\mathbf{k}s'}\rangle = i \sum_r \boldsymbol{\xi}_{\mathbf{k}sr} \mathcal{O}_{\mathbf{k}rs'}, \tag{3.66}$$

$$\langle u_{\mathbf{k}s}| \mathcal{O}(\mathbf{k}) |\nabla_{\mathbf{k}} u_{\mathbf{k}s'}\rangle = -i \sum_r \mathcal{O}_{\mathbf{k}sr} \boldsymbol{\xi}_{\mathbf{k}rs'}. \tag{3.67}$$

Combining these two contributions, we can express them in terms of a commutator in band-index space, since the sum runs over all band indexes,

$$i \sum_r \left[\boldsymbol{\xi}_{\mathbf{k}sr}\mathcal{O}_{\mathbf{k}rs'} - \mathcal{O}_{\mathbf{k}sr}\boldsymbol{\xi}_{\mathbf{k}rs'}\right] = \left[\boldsymbol{\xi}(\mathbf{k}), \mathcal{O}(\mathbf{k})\right]_{ss'}, \tag{3.68}$$

[17] It is sometimes possible to construct a basis of Bloch states (in the sense that they are states which obey Bloch's theorem, not that they are eigenstates of the crystal Hamiltonian) in which the Berry connections are *all* trivial. In that case, matrix elements of the derivative of an operator are reduced to derivatives of the matrix elements themselves, and certain calculations are then made considerably simpler. This is one of the subjects of Chapter 4.

and the collection of all three terms, i.e. the matrix element in Eq.(3.63), is written as,

$$[\nabla_{\mathbf{k}}\mathcal{O}(\mathbf{k})]_{ss'} = \nabla_{\mathbf{k}}\mathcal{O}_{\mathbf{k}ss'} - i\,[\boldsymbol{\xi}, \mathcal{O}(\mathbf{k})]_{ss'}\,. \tag{3.69}$$

Now, consider that instead of the matrix elements of the $\mathbf{k}$-derivative of $\mathcal{O}(\mathbf{k})$, we take the commutator, again in band-index space, of the covariant derivative with $\mathcal{O}(\mathbf{k})$,

$$\big[\mathbf{D}, \mathcal{O}(\mathbf{k})\big]_{ss'}\lambda(\mathbf{k}),$$

for $\lambda(\mathbf{k})$ a given test function. By opening up the commutator, and using the definition of the covariant derivative, Eq.(3.53), we can see that,

$$\begin{aligned}\big[\mathbf{D}, \mathcal{O}(\mathbf{k})\big]_{ss'}\lambda(\mathbf{k}) &= \big(\nabla_{\mathbf{k}}\mathcal{O}_{\mathbf{k}ss'}\lambda(\mathbf{k})\big) - \mathcal{O}_{\mathbf{k}ss'}\big(\nabla_{\mathbf{k}}\lambda(\mathbf{k})\big) - i\,[\boldsymbol{\xi}, \mathcal{O}(\mathbf{k})]_{ss'}\,\lambda(\mathbf{k}), &(3.70)\\ &= \big(\nabla_{\mathbf{k}}\mathcal{O}_{\mathbf{k}ss'}\big)\lambda(\mathbf{k}) - i\,[\boldsymbol{\xi}, \mathcal{O}(\mathbf{k})]_{ss'}\,\lambda(\mathbf{k}), &(3.71)\end{aligned}$$

which is exactly what we have obtained in Eq.(3.69), times the test function. We can therefore conclude that the matrix elements of the derivative of an operator can be expressed as the matrix elements of the commutator of the covariant derivative with that same operator,

$$[\nabla_{\mathbf{k}}\mathcal{O}(\mathbf{k})]_{ss'} = \big[\mathbf{D}, \mathcal{O}(\mathbf{k})\big]_{ss'}, \tag{3.72}$$

or, alternatively, as,

$$\boxed{\nabla_{\mathbf{k}}\mathcal{O}(\mathbf{k}) = \big[\mathbf{D}, \mathcal{O}(\mathbf{k})\big]}. \tag{3.73}$$

The second point of this section concerns the velocity operator, and its expression in the basis of periodic functions. The velocity operator in the unperturbed system — that is, electrons in a crystal without any electric fields — $(i\hbar)^{-1}\,[\mathbf{r}, H_0]$ is given by the Heisenberg equation for the position operator. Unlike the position, this operator is diagonal in $\mathbf{k}$, since it retains the periodicity of the crystal; its matrix elements in the Bloch basis are,

$$\mathbf{v}_{\mathbf{k}\mathbf{k}'ss'} = (2\pi)^d\delta(\mathbf{k} - \mathbf{k}')\,\mathbf{v}_{\mathbf{k}ss'}, \tag{3.74}$$

for $\mathbf{v}_{\mathbf{k}ss'}$ the elements of the velocity matrix in band-index space. These matrix elements can now be determined by using Eq.(3.71), the above definition of the velocity and the correspondence between $\mathbf{r} \to i\mathbf{D}$ and $H \to \mathcal{H}$,

$$\begin{aligned}\mathbf{v}_{\mathbf{k}ss'} &= \frac{1}{\hbar}\,[\mathbf{D}, \mathcal{H}(\mathbf{k})]_{ss'}\,, &(3.75)\\ &= \frac{1}{\hbar}\left[\delta_{ss'}\big(\nabla_{\mathbf{k}}\epsilon_{\mathbf{k}s}\big) - i\Delta\epsilon_{\mathbf{k}s's}\boldsymbol{\xi}_{\mathbf{k}ss'}\right], &(3.76)\end{aligned}$$

for $\Delta\epsilon_{\mathbf{k}s's} = \epsilon_{\mathbf{k}s'} - \epsilon_{\mathbf{k}s}$. The velocity matrix elements carry the usual, *intraband*, term — written

as the gradient of a band energy — as well as an *interband* term — written as the product of the Berry connection between two bands and the energy difference between them. Finally, we must emphasize two aspects here: first, note that we have now written the Bloch matrix elements for the position and velocity operators in the unperturbed system, Eq.(3.53) and Eq.(3.76), and have, therefore, determined the form of the perturbation for both gauges in the case of the continuum Hamiltonian — Eqs.(3.14) and (3.15); secondly, note that we have considered the unperturbed system in the above derivation. The introduction of a perturbation to the Hamiltonan can, in principle, change the form of the velocity itself. Such is the case of the velocity operator in the velocity-gauge description.

3.3.3 Infinite Crystal Limit

Finally, let us add some comments on the changes that the infinite-crystal limit, $L_i \to \infty$ (for i running through a d-number of relevant directions), introduces to our problem. The most significant feature is that momentum sums are replaced by d-dimensional integrals over the FBZ. The many-body crystal Hamiltonian then reads as [1],

$$H_0 = \sum_s \int \frac{d^d\mathbf{k}}{(2\pi)^d} \epsilon_{\mathbf{k}s} c^\dagger_{\mathbf{k}s} c_{\mathbf{k}s}, \tag{3.77}$$

for $c^\dagger_{\mathbf{k}s}$, $c_{\mathbf{k}s}$ the creation and destruction operators of Bloch states, which are expressed in terms of the fermionic fields of Section 3.2 as,

$$c^\dagger_{\mathbf{k}s} = \int d^d\mathbf{r}\, \psi_{\mathbf{k}s}(\mathbf{r})\, \Psi^\dagger(\mathbf{r}). \tag{3.78}$$

The Bloch state orthogonality relation, the anti-commutation relations, and the lattice sum rule are suitably modified,

$$\langle \psi_{\mathbf{k}s} | \psi_{\mathbf{k}'s'} \rangle = (2\pi)^d \delta_{ss'} \delta(\mathbf{k} - \mathbf{k}'), \tag{3.79}$$

$$\left\{ c_{\mathbf{k}s}, c^\dagger_{\mathbf{k}'s'} \right\} = (2\pi)^d \delta_{ss'} \delta(\mathbf{k} - \mathbf{k}'), \tag{3.80}$$

$$\sum_{\mathbf{R}} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}} = \frac{(2\pi)^3}{v_C} \delta(\mathbf{k} - \mathbf{k}'), \tag{3.81}$$

Note that the operator $c^\dagger_{\mathbf{k}s} c_{\mathbf{k}s}$ is a (particle) density in momentum space and not a dimensionless number operator as in the finite volume case.

3.4 Density Matrix Formalism

Over the last two sections, we have been considering aspects of our work that are connected with two elements highlighted in this chapter's preface, namely, the consequences of gauge invariance and

how it is ensured, and the issue of writing the perturbations associated with each gauge. We can now move on to the third: the description of the nonlinear optical response and how this is expressed in terms of the density matrix. In keeping with what we have presented so far, the expressions for the response we derive here follow specifically from the continuum Hamiltonian, Eq.(3.30). This, although inconsequential for the length gauge — which we can safely use even when we consider other Hamiltonians — it is not so for the velocity gauge. As we will see later on, this type of description is susceptible to band-space truncations,[18] i.e., when we consider band structures used in "practical calculations", meaning that the particular description of the velocity gauge we present here (again, associated with the continuum Hamiltonian) is of very little, if any, practical use. As we have mentioned before, practical calculations in the velocity gauge require both the procedure and the ideas of ref.[3]. This description, however, still serves two important purposes: it illustrates the structure of the velocity gauge in the context of a density-matrix formulation; it can be used to show how velocity-gauge descriptions are susceptible to band-space truncations.

3.4.1 Averages of Observables, the Density Matrix and its Equations of Motion

Consider the many-body operator O_α, associated with the matrix elements (in band-index space), $\mathcal{O}^\alpha_{\mathbf{k}ss'}$; the α label indicates the gauge to which this operator is tied, $\alpha = E, A$, following the conclusions of Section 3.2:

$$O_\alpha = \sum_{s's} \int \frac{d^d\mathbf{k}}{(2\pi)^d} c^\dagger_{\mathbf{k}s} \mathcal{O}^\alpha_{\mathbf{k}ss'} c_{\mathbf{k}s'}. \tag{3.82}$$

The average of this observable can be expressed as,

$$\langle O_\alpha(t)\rangle = \sum_{s's} \int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathcal{O}^\alpha_{\mathbf{k}ss'} \langle c^\dagger_{\mathbf{k}s} c_{\mathbf{k}s'}\rangle_\alpha = \sum_{s's} \int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathcal{O}^\alpha_{\mathbf{k}ss'}\, \rho^\alpha_{\mathbf{k}s's}(t) = \int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\big[\mathcal{O}^\alpha(\mathbf{k})\rho^\alpha(\mathbf{k},t)\big], \tag{3.83}$$

by introducing the density matrix, $\rho^\alpha_{\mathbf{k}s's}(t) = \langle c^\dagger_{\mathbf{k}s} c_{\mathbf{k}s'}\rangle_\alpha$ — note how the band indexes are switched — as well as a trace over band-index space,

$$\mathrm{Tr}[\mathcal{A}(\mathbf{k})\mathcal{B}(\mathbf{k})] = \sum_{s's} \mathcal{A}_{ss'}(\mathbf{k})\mathcal{B}_{s's}(\mathbf{k}). \tag{3.84}$$

Like all other matrices in band-index space, the density matrix is itself gauge-dependent so that there is one particular form of this matrix for each gauge. It is also, however, different from all other matrices in band-index space because it alone translates the dynamics of the system of crystal electrons subjected to an external electric field: once we can describe the time-evolution of its matrix elements, i.e., we determine their equations of motion, we will be able to compute (at least in principle) the average of any relevant observable and thus the response of the system. To see that

[18] By band-space truncation we mean the selection of a specific subset of bands — or a specific region of the FBZ — from one's original Hamiltonian.

this is the case, let us study this object more closely. The average in Eq.(3.83) is an average over the different possible configurations of the many-body system so that the density matrix in band-index space involves the trace — in a many-body basis — over the product of the full many-body density matrix, ρ^{α}, and the bilinear operator, $c^{\dagger}_{\mathbf{k}s'}c_{\mathbf{k}s}$,[19]

$$\rho^{\alpha}_{\mathbf{k}ss'}(t) = \mathrm{tr}\big[\rho^{\alpha}(t)c^{\dagger}_{\mathbf{k}s'}c_{\mathbf{k}s}\big], \tag{3.85}$$

for,

$$\rho^{\alpha}(t) = \sum_{n} p_n \left|\psi^{\alpha}_n(t)\right\rangle \left\langle\psi^{\alpha}_n(t)\right|, \tag{3.86}$$

where $\big\{|\psi^{\alpha}_n\rangle\big\}$ is a complete set of many-body state vectors. In the absence of an external field, and at zero temperature, the full-many body density matrix is expressed in terms of a single state — the ground-state of the system, i.e., the Fermi sphere, $|\psi_{FS}\rangle$. For $T \neq 0$ (and again in the absence of an external field), there are different possible static configurations, each of which weighted by the Fermi-Dirac distribution function. All of this, of course, changes when an electric field is coupled to the system. The external perturbation will then drive electronic transitions between different states — in principle, between different bands and/or different values of $\mathbf{k}$ — and the description of the time-evolution of the many-body state vectors becomes nontrivial. In the Schrödinger picture, the time-evolution of the state vectors, $|\psi^{\alpha}_n(t)\rangle$, in a given gauge is given by the Schrödinger equation,

$$i\hbar\frac{\partial\left|\psi^{\alpha}_n(t)\right\rangle}{\partial t} = H_{\alpha}(t)\left|\psi^{\alpha}_n(t)\right\rangle, \tag{3.87}$$

which, naturally, determines the dynamics of the full many-body density matrix, $\rho^{\alpha}(t)$,

$$\frac{\partial\rho^{\alpha}(t)}{\partial t} = \sum_{n} p_n\Big[\frac{\partial\left|\psi^{\alpha}_n(t)\right\rangle}{\partial t}\left\langle\psi^{\alpha}_n(t)\right| + \left|\psi^{\alpha}_n(t)\right\rangle\frac{\partial\left\langle\psi^{\alpha}_n(t)\right|}{\partial t}\Big] = \frac{1}{i\hbar}\big[H_{\alpha}(t),\rho^{\alpha}(t)\big]. \tag{3.88}$$

Furthermore, the time evolution of the $\rho^{\alpha}_{\mathbf{k}ss'}(t)$, the matrix elements of the density matrix in band-index space, is itself expressed in terms of Eq.(3.88), following its definition in terms of the full many-body density matrix. Using the cyclic property of the trace, $\mathrm{tr}\big[ABC\big] = \mathrm{tr}\big[CAB\big] = \mathrm{tr}\big[BCA\big]$, and by carefully manipulating the expressions, we can write the equations of motion for the $\rho^{\alpha}_{\mathbf{k}ss'}(t)$ as,

$$i\hbar\frac{\partial\rho^{\alpha}_{ss'}(\mathbf{k},t)}{\partial t} = \mathrm{tr}\big[\big[H_{\alpha}(t),\rho^{\alpha}(t)\big]c^{\dagger}_{\mathbf{k}s'}c_{\mathbf{k}s}\big] = \big\langle\big[c^{\dagger}_{\mathbf{k}s'}c_{\mathbf{k}s},H_{\alpha}(t)\big]\big\rangle_{\alpha}, \tag{3.89}$$

[19] Some comments on nomenclature: the diagonal — or intraband — elements of the density matrix, $\rho^{\alpha}_{\mathbf{k}ss}(t)$, are usually called *populations*, whereas the nondiagonal — or interband — elements of the density matrix, $\rho^{\alpha}_{\mathbf{k}s\bar{s}}(t)$, are usually called *coherences*. The first describe (in a sense) the probability of finding an electron in the state labelled by $\mathbf{k}$ and s for every possible state, while the second describe (again, in a sense) the probability amplitude between two states that are labelled by $\mathbf{k}$ and band indexes s and $\bar{s}$, for every possible pair of bands.

for $H_\alpha(t)$, $\alpha = E, A$, the many-body equivalents of the single-particle Hamiltonians of Section 3.2. Having determined the Bloch matrix elements of the perturbation in each gauge, Eqs.(3.54) and (3.75), in the previous section, these Hamiltonians can now be written as,

$$H_E(t) = \sum_{ss'} \int \frac{d^d\mathbf{k}}{(2\pi)^d} c^\dagger_{\mathbf{k}s} \left[\delta_{ss'}\epsilon_{\mathbf{k}s} + i\,|e|\,\mathbf{E}(t)\cdot\mathbf{D}_{\mathbf{k}ss'}\right] c_{\mathbf{k}s'}, \tag{3.90}$$

$$H_A(t) = \sum_{ss'} \int \frac{d^d\mathbf{k}}{(2\pi)^d} c^\dagger_{\mathbf{k}s} \left[\delta_{ss'}\left(\epsilon_{\mathbf{k}s} + \frac{e^2A^2(t)}{2m_e}\right) + |e|\,\mathbf{A}(t)\cdot\mathbf{v}_{\mathbf{k}ss'}\right] c_{\mathbf{k}s'}. \tag{3.91}$$

The final step in our derivation of the equations of motion for the $\rho^\alpha_{\mathbf{k}ss'}(t)$ is thus the calculation of the commutators in the last member of Eq.(3.89): this is a lengthy, but not altogether complicated, calculation that we will therefore omit.[20] The final expressions, cast in terms of commutators in band-index space, can be written as follows,

$$\left[i\hbar\frac{\partial}{\partial t} - \Delta\epsilon_{\mathbf{k}ss'}\right]\rho^E_{\mathbf{k}ss'}(t) = i\,|e|\,\mathbf{E}(t)\cdot\left[\mathbf{D}, \rho^E(\mathbf{k},t)\right]_{ss'}, \tag{3.92}$$

$$\left[i\hbar\frac{\partial}{\partial t} - \Delta\epsilon_{\mathbf{k}ss'}\right]\rho^A_{\mathbf{k}ss'}(t) = |e|\,\mathbf{A}(t)\cdot\left[\mathbf{v}, \rho^A(\mathbf{k},t)\right]_{ss'}. \tag{3.93}$$

This length-gauge description of the density matrix is the same as the aforementioned description developed by Sipe [22, 23], with the only difference being that here it is cast in terms of the covariant derivative, $\mathbf{D}_{\mathbf{k}ss'}$. We note that the derivative with respect to $\mathbf{k}$ in Eq.(3.92), which follows from the expression for $\mathbf{D}_{\mathbf{k}ss'}$, will couple the density matrix elements of different values of $\mathbf{k}$. As we have mentioned in the introduction to this chapter, this is a substantial complication — one that becomes patently clear when we consider the nonlinear optical response or when we consider a possible numerical integration of these equations — but one that does not, however, render the length gauge useless. It is, complicated as it may be, our tool of reference for analytical calculations [10]. As for its counterpart description, the velocity gauge, Eq.(3.93), we can see that it is completely decoupled in crystal momentum, $\mathbf{k}$, and can thus be solved independently for each point of the FBZ — similarly to the corresponding equations in atomic systems.[21]

The nonlinear optical response of electrons in a crystal has been traditionally expressed in terms of the expectation values of either the current density operator, $\mathbf{J}_\alpha = -|e|\,\mathbf{v}_\alpha$, or the polarisation density operator, $\mathbf{P}_\alpha = -|e|\,\mathbf{r}_\alpha$, and there is, *a priori*, no reason to choose one over the other; the two objects are related by a time derivative. But now, having studied the representation of both velocity and position operators in the Bloch basis, we know that the covariant derivative is not,

[20] This involves some careful bookkeeping, the use of the commutator relations in Eq.(3.80), and, in the case of the length gauge, an integration by parts. A step-by-step derivation of this can be found in ref.[1].

[21] Note that, while the perturbation to the velocity-gauge Hamiltonian is composed of two terms — one proportional to the velocity operator, the other a diamagnetic term — the equations of motion for the density matrix elements carry only *one* term. The diamagnetic term is a constant, $(2m_e)^{-1}e^2A^2(t)$, which is independent of both $\mathbf{k}$ and s, times the unit matrix in band-index space. It is therefore dropped from the commutator in Eq.(3.89).

strictly speaking, a vector that we can compute (like the Berry connection or the matrix elements of the velocity), and that, because of its derivative term, it must always act on *something* else. In the case of the polarisation density operator — that is, Eq.(3.82) with $\mathcal{O}^{\alpha}_{\mathbf{k}ss'} \to \mathbf{D}_{\mathbf{k}ss'}$ — the derivative must act either on the destruction operator on the right, $c_{\mathbf{k}s}$, or, alternatively, after an integration by parts, on the creation operator on the left, $c^{\dagger}_{\mathbf{k}s}$: this means that its average, unlike that of the velocity operator, cannot be entirely expressed in terms of the density matrix, which is the basis of our description of this system's dynamics. The velocity operator, on the other hand, is not burdened by this issue. One can quite easily write its expectation value in both gauges in terms of their respective density matrices,

$$\big\langle \mathbf{J}_E(t) \big\rangle = -\,|e| \int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\left[\mathbf{v}(\mathbf{k})\,\rho^E(\mathbf{k},t)\right] = -\,|e| \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{s's} \mathbf{v}_{\mathbf{k}ss'}\,\rho^E_{\mathbf{k}s's}(t), \tag{3.94}$$

$$\big\langle \mathbf{J}_A(t) \big\rangle = -\,|e| \int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\Big[\mathbf{v}^A(\mathbf{k},t)\,\rho^A(\mathbf{k},t)\Big] = -\,|e| \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{s's} \mathbf{v}_{\mathbf{k}ss'}\,\rho^A_{\mathbf{k}s's}(t) - \frac{|e|^2}{m_e} n_e \mathbf{A}(t). \tag{3.95}$$

Note that we have used here the definitions of the velocity operators in each gauge. For the length gauge, the velocity operator is simply that of the unperturbed system, since the perturbation to the crystal Hamiltonian — the dipole interaction term — is itself expressed in terms of the position operator, so that $\mathbf{v}_E(t) = (i\hbar)^{-1}\,[\mathbf{r}, H_0]$. For the velocity gauge, however, the perturbation to the crystal Hamiltonian is expressed in terms of the (unperturbed) velocity operator — which, in the case of the continuum Hamiltonian, is essentially equivalent to the momentum operator — so that the velocity operator gets an additional term, proportional to the vector potential,

$$\mathbf{v}_A(t) = \frac{1}{i\hbar}\,[\mathbf{r}, H_A(t)] = \frac{1}{i\hbar}\,[\mathbf{r}, H_0] + \frac{1}{i\hbar m_e}\,[\mathbf{r}, \mathbf{A}(t)\cdot\mathbf{p}] = \mathbf{v} + \frac{1}{m_e}\mathbf{A}(t), \tag{3.96}$$

since it involves the canonical commutation relation. We can proceed to study the order-by-order expansion of the density matrix elements in powers of the electric field; this is the subject of the following subsection. This order-by-order expansion of the matrix elements of the density matrix allows us to derive the expressions for the optical coefficients — in the case of the current density, the conductivities — that describe the system's response at a given order; these expressions will not be presented here. Instead, they will be introduced in the next two chapters.

3.4.2 The Density Matrix and its Equations of Motion: Perturbation Theory

We have just noted that the perturbative treatment of the response requires an order-by-order study of the solutions to the equations of motion for the density matrix. We must, therefore, begin by breaking these matrix elements into their different order terms (in the external field) — $\rho_{\mathbf{k}ss'} = \rho^{(0)}_{\mathbf{k}ss'} + \rho^{(1)}_{\mathbf{k}ss'} + (...)$ — and then proceed to solve, iteratively, the equations of motion at each order. From Eqs.(3.92) and (3.93), we can see that the equations of motion only couple matrix

elements of consecutive orders, meaning that they can be written as,[22]

$$\big[i\hbar\frac{\partial}{\partial t}-\Delta\epsilon_{\mathbf{k}ss'}\big]\rho^{(n),E}_{\mathbf{k}ss'}(t)=i\,|e|\,\mathbf{E}(t)\cdot\big[\mathbf{D},\rho^{(n-1),E}(\mathbf{k},t)\big]_{ss'}, \tag{3.97}$$

$$\big[i\hbar\frac{\partial}{\partial t}-\Delta\epsilon_{\mathbf{k}ss'}\big]\rho^{(n),A}_{\mathbf{k}ss'}(t)=\ |e|\,\mathbf{A}(t)\cdot\big[\mathbf{v},\rho^{(n-1),A}(\mathbf{k},t)\big]_{ss'}. \tag{3.98}$$

By moving into the frequency domain, we can replace the time derivative in the left-hand side of the equations of motion by the corresponding frequency argument of the density matrix element. In taking the Fourier decomposition of these objects, we assume that the perturbation is switched on *adiabatically* ($\eta\to 0^+$), so as to give a well-defined initial condition,

$$\rho^{(n),\alpha}_{\mathbf{k}ss'}(t)=\int\frac{d\omega}{2\pi}\,e^{-i(\omega+i\eta)t}\,\rho^{(n),\alpha}_{\mathbf{k}ss'}(\omega), \tag{3.99}$$

$$\mathbf{E}(t)=\int\frac{d\omega}{2\pi}\,e^{-i(\omega+i\eta)t}\,\mathbf{E}(\omega), \tag{3.100}$$

$$\mathbf{A}(t)=\int\frac{d\omega}{2\pi}\,e^{-i(\omega+i\eta)t}\,\mathbf{A}(\omega). \tag{3.101}$$

After some careful manipulations, we can write the solutions to the equations of motion as,

$$\rho^{(n),E}_{\mathbf{k}ss'}(\omega)=i\,|e|\,\frac{1}{\hbar\omega+i\eta-\Delta\epsilon_{\mathbf{k}ss'}}\int\frac{d\omega_n}{2\pi}\frac{d\omega_{n-1}}{2\pi}\mathbf{E}(\omega_n)\cdot\big[\mathbf{D},\rho^{(n-1),E}(\mathbf{k},\omega_{n-1})\big]_{ss'}(2\pi)\delta(\omega-(\omega_n+\omega_{n-1})), \tag{3.102}$$

$$\rho^{(n),A}_{\mathbf{k}ss'}(\omega)=\ |e|\,\frac{1}{\hbar\omega+i\eta-\Delta\epsilon_{\mathbf{k}ss'}}\int\frac{d\omega_n}{2\pi}\frac{d\omega_{n-1}}{2\pi}\mathbf{A}(\omega_n)\cdot\big[\mathbf{v},\rho^{(n-1),A}(\mathbf{k},\omega_{n-1})\big]_{ss'}(2\pi)\delta(\omega-(\omega_n+\omega_{n-1})), \tag{3.103}$$

which are, admittedly, rather dense expressions. To make this somewhat clearer, let us consider the first nontrivial solutions for each gauge: $\rho^{(1),(E)}_{\mathbf{k}ss'}(\omega)$ and $\rho^{(1),(A)}_{\mathbf{k}ss'}(\omega)$.

$$\rho^{(1),E}_{\mathbf{k}ss'}(\omega)=i\,|e|\,\frac{1}{\hbar\omega+i\eta-\Delta\epsilon_{\mathbf{k}ss'}}\int\frac{d\omega_1}{2\pi}\mathbf{E}(\omega_1)\cdot\big[\mathbf{D},\rho^{(0)}(\mathbf{k})\big]_{ss'}(2\pi)\delta(\omega-\omega_1), \tag{3.104}$$

$$=i\,|e|\,d_{\mathbf{k}ss'}(\omega)\mathbf{E}(\omega)\cdot\big[\mathbf{D},\rho^{(0)}(\mathbf{k})\big]_{ss'}, \tag{3.105}$$

$$=i\,|e|\,d_{\mathbf{k}ss'}(\omega)\mathbf{E}(\omega)\cdot\big(\delta_{ss'}\big(\nabla_{\mathbf{k}}f_{\mathbf{k}s}\big)-i\boldsymbol{\xi}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}\big), \tag{3.106}$$

[22] In a general description of the velocity gauge, one that can be used in practical applications [3], the equation of motion for $\rho^{(n),A}_{\mathbf{k}ss'}(t)$ includes density matrix elements of *all* orders below n. This will be considered later on, in the chapter on the nonlinear optical response of the monolayers of plain and gapped graphene.

and,

$$\rho^{(1),A}_{\mathbf{k}ss'}(\omega) = |e| \frac{1}{\hbar\omega + i\eta - \Delta\epsilon_{\mathbf{k}ss'}} \int \frac{d\omega_n}{2\pi} \mathbf{A}(\omega_1) \cdot \left[\mathbf{v}, \rho^{(0)}(\mathbf{k})\right]_{ss'} (2\pi)\delta(\omega - \omega_1), \tag{3.107}$$

$$= |e|\, d_{\mathbf{k}ss'}(\omega)\mathbf{A}(\omega) \cdot \left[\mathbf{v}, \rho^{(0)}(\mathbf{k})\right]_{ss'}, \tag{3.108}$$

$$= |e|\, d_{\mathbf{k}ss'}(\omega)\mathbf{A}(\omega) \cdot \mathbf{v}_{\mathbf{k}ss'} \Delta f_{\mathbf{k}s's}. \tag{3.109}$$

This follow directly from the observation that $\rho^{(0)}(\mathbf{k})$ is the equilibrium density matrix of electrons in a crystal in the absence of an external electric field, and is therefore expressed in terms of the time-independent Fermi-Dirac distribution function, $\rho^{(0)}_{\mathbf{k}ss'} = \delta_{ss'} f_{\mathbf{k}s}$. We have also introduced two new pieces of notation: $\Delta f_{\mathbf{k}s's}$, which represents the difference of Fermi-Dirac distribution functions, $\Delta f_{\mathbf{k}s's} = f_{\mathbf{k}s'} - f_{\mathbf{k}s}$; $d_{\mathbf{k}ss'}(\omega)$, which represents the following frequency denominator,

$$d_{\mathbf{k}ss'}(\omega) := \frac{1}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'} + i\eta}. \tag{3.110}$$

Higher-order solutions can be determined *iteratively*, i.e., by successively applying the relations in Eqs.(3.102) and (3.103) to lower-order solutions. Introducing two additional pieces of notation, the Hadamard product,

$$(A \circ B)_{ss'} := A_{ss'}\, B_{ss'}, \tag{3.111}$$

and the contracted sum of frequencies,

$$\omega_{[m]} := \sum_{i=1}^{m} \omega_i, \tag{3.112}$$

we can finally write the matrix elements of the density matrix at order n, $\rho^{(n)}_{\mathbf{k}ss'}$, in both gauges, in terms of the equilibrium density matrix,

$$\rho^{(n),E}_{\mathbf{k}ss'}(\omega) = (i\,|e|)^n \left[\prod_{i=1}^{n-1} \int \frac{d\omega_i}{2\pi}\, E^{\alpha_i}(\omega_i)\right] E^{\alpha_n}(\omega - \omega_{[n-1]})$$
$$\times\, d_{\mathbf{k}ss'}(\omega)\left[D^{\alpha_1},\, d(\omega - \omega_1) \circ \left[..., d(\omega - \omega_{[n-1]}) \circ \left[D^{\alpha_n}, \rho^{(0)}(\mathbf{k})\right]...\right]\right]_{ss'}, \tag{3.113}$$

$$\rho^{(n),A}_{\mathbf{k}ss'}(\omega) = |e|^n \left[\prod_{i=1}^{n-1} \int \frac{d\omega_i}{2\pi}\, A^{\alpha_i}(\omega_i)\right] A^{\alpha_n}(\omega - \omega_{[n-1]})$$
$$\times\, d_{\mathbf{k}ss'}(\omega)\left[v^{\alpha_1},\, d(\omega - \omega_1) \circ \left[..., d(\omega - \omega_{[n-1]}) \circ \left[v^{\alpha_n}, \rho^{(0)}(\mathbf{k})\right]...\right]\right]_{ss'}. \tag{3.114}$$

These expressions, although somewhat opaque, translate the dynamics of the system of crystal electrons subjected to an external electric field and, alongside the expressions for the current density, are the key elements in the description of this system's response. They also warrant several additional comments. First, note that we have made use of nested commutators: at order n, the

density matrix elements involves n different commutators, one consecutively inside the other, as in a Russian nesting doll. This makes calculations at higher orders difficult, particularly so in the length-gauge description, where the derivative term of the covariant derivative acts on every term on the inside, Eq.(3.71); as a result, either one computes derivatives of rather complex objects, or one has to break the response into a substantial number of terms and therefore end up doing some tough bookkeeping [10].[23] Secondly, note that these density matrix elements are expressed in terms of three objects: the band energies, $\epsilon_{\mathbf{k}s}$; the Berry connection, $\xi_{\mathbf{k}ss'}$; and the structure of filled/empty bands, $f_{\mathbf{k}s}$. These determine the system's nonlinear optical response, provided, of course, that one can compute the FBZ integrals in Eqs.(3.94) and (3.95). Thirdly, note that it is not altogether clear how these density matrix elements, despite the fact that they represent the same object, are related to one another. The same can be said of the matrix elements $\mathcal{O}^{\alpha}_{\mathbf{k}ss'}$ which, alongside the density matrix elements, are used to write the expectation value of observables in either gauge. We know that these relations follow from a general transformation rule for operators in different gauges, like the one we derived in Section 3.2 for free electrons, but so far we have not seen how this is transcribed to the system of electrons in a crystal. Perhaps more importantly, we have yet to prove — in explicit terms — that the expectation value of an observable is indeed gauge invariant, and to understand how seemingly reasonable approximations can break this invariance. We must, therefore, return to points we presented in the section on the equivalence between gauges, and combine them with results from our section on matrices in band-index space.

3.5 Equivalence between Two Gauges (II) and the Existence of Sum Rules

In Section 3.2, we established that the relation between observables in the length and velocity gauges — in the context of the free electron — is a unitary transformation, whose generator is the position operator, Eq.(3.26), that acts as a translation in *momentum* space. While the continuum Hamiltonian for an electron in a crystal is, of course, different from that of the free electron, the introduction of the crystal potential, $V(\mathbf{r})$, does not, in anyway, affect this set of relations: $V(\mathbf{r})$ commutes with the transformation, meaning that the relations between states and observables in different gauges must be the same as those for the case of the free electron. We can, therefore, take the relation between diagonal observables, Eq.(3.28), and cast it in terms of matrix elements

[23] Note also how this nested commutator structure emphasizes the similarities between the expressions for density matrix elements: exchanging $\mathbf{E}(\omega) \leftrightarrow \mathbf{A}(\omega)$ and $\mathbf{D} \leftrightarrow \mathbf{v}$ takes us from Eq.(3.113) to Eq.(3.114), and vice-versa. This can be, however, somewhat misleading. As we have mentioned above, the covariant derivative acts on every term that is on the inside of the commutator in which it appears, so that it leads to (among other things) higher order poles, e.g.$(\hbar\omega + i\eta - \Delta\epsilon_{\mathbf{k}ss'})^{-n}$, $n \geq 2$. After some careful bookkeeping, the form of the contributions in the length-gauge description is shown to be very different from those in the velocity-gauge.

of Bloch functions,[24]

$$\langle\psi_{\mathbf{k}s}|\,\mathcal{O}^A\left(\frac{\nabla}{i}\right)|\psi_{\mathbf{k}'s'}\rangle = \langle\psi_{\mathbf{k}s}|\,\mathcal{O}^E\left(\frac{\nabla}{i}+\frac{e}{\hbar}\mathbf{A}(t)\right)|\psi_{\mathbf{k}'s'}\rangle\,, \tag{3.115}$$

which, as we have seen in Subsection 3.3.2 (see Eqs.(3.59) and (3.60)), can be expressed as matrix elements of operators in band-index space,

$$(2\pi)^3\delta(\mathbf{k}-\mathbf{k}')\,\langle u_{\mathbf{k}s}|\,\mathcal{O}^A(\mathbf{k})\,|u_{\mathbf{k}s'}\rangle = (2\pi)^3\delta(\mathbf{k}-\mathbf{k}')\,\langle u_{\mathbf{k}s}|\,\mathcal{O}^E\left(\mathbf{k}+\frac{|e|}{\hbar}\mathbf{A}(t)\right)|u_{\mathbf{k}s'}\rangle\,. \tag{3.116}$$

In operator form, this equivalence reads as,

$$\boxed{\mathcal{O}^A(\mathbf{k}) = \mathcal{O}^E\left(\mathbf{k}+\frac{|e|}{\hbar}\mathbf{A}(t)\right)}. \tag{3.117}$$

As we have pointed out previously, these operators are associated with different bases — $|u_{\mathbf{k}s}\rangle$ for the operator in the left-hand side of the equation, and $\left|u_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)s}\right\rangle$ for the operator in the right-hand side — and that these bases are not trivially related. We did establish, however, that an operator centred in a given point of the FBZ can, in principle, be expanded at another $\mathbf{k}$-point, provided that both their bases and their matrix elements are deemed continuous in the FBZ. Expanding the right-hand side of Eq.(3.117) in powers of $|e|\,\hbar^{-1}\mathbf{A}(t)$, so as to write it in terms of operators centred around $\mathbf{k}$, we obtain,

$$\mathcal{O}^A(\mathbf{k},t) = \mathcal{O}^E(\mathbf{k},t) + \frac{|e|}{\hbar}\mathbf{A}(t)\cdot\nabla_{\mathbf{k}}\mathcal{O}^E(\mathbf{k},t) + (....), \tag{3.118}$$

which we can then express in terms of matrix elements in the *same* $\mathbf{k}$-point, by using the relation between the matrix elements of the derivative of an operator and its commutator with the covariant derivative, Eq.(3.73),

$$\mathcal{O}^A_{\mathbf{k}ss'}(t) = \mathcal{O}^E(\mathbf{k},t) + \frac{|e|}{\hbar}\mathbf{A}(t)\cdot\left[\mathbf{D},\mathcal{O}^E(t)\right]_{ss'} + (....), \tag{3.119}$$

Collecting every term in the right-hand side of this last equation, the relation between matrix elements, $\mathcal{O}^\alpha_{\mathbf{k}ss'}$, in different gauges finally reads as,

$$\mathcal{O}^A_{\mathbf{k}ss'}(t) = \sum_{n=0}^{+\infty}\frac{1}{n!}\left(\frac{|e|}{\hbar}\right)^n A^{\alpha_1}(t)\ldots A^{\alpha_n}(t)\left[D^{\alpha_1},\left[...,\left[D^{\alpha_n},\mathcal{O}^E(t)\right]...\right]\right]_{ss'}. \tag{3.120}$$

[24] Note that we are being rather liberal in how we are referring to this unitary transformation. The relation between observables in Eq.(3.28) is, strictly speaking, a relation between many-body operators; the relation between different one-electron operators reads instead as $\mathcal{U}(t) = \exp\left[i\frac{|e|}{\hbar}\mathbf{A}(t)\cdot\mathbf{r}\right]$.

Before we move on to our study of the relation between density matrix elements in the two gauges, let us consider a particular instance of it: the matrix elements of the velocity operator. It follows from Eq.(3.120) that,

$$\mathbf{v}^{A}_{\mathbf{k}ss'}(t) = \sum_{n=0}^{+\infty} \frac{1}{n!}\left(\frac{|e|}{\hbar}\right)^{n} A^{\alpha_1}(t)\ldots A^{\alpha_n}(t)\big[D^{\alpha_1},\big[...,\big[D^{\alpha_n},\mathbf{v}^{E}(t)\big]...\big]\big]_{ss'}, \tag{3.121}$$

$$\rightarrow \mathbf{v}_{\mathbf{k}ss'} + A^{\alpha_1}(t)\big[D^{\alpha_1},\mathbf{v}\big]_{ss'} + \frac{1}{2}A^{\alpha_2}(t)A^{\alpha_1}(t)\big[D^{\alpha_2},\big[D^{\alpha_1},\mathbf{v}\big]\big]_{ss'} + (....), \tag{3.122}$$

as $\mathbf{v}^{E}_{\mathbf{k}ss'}(t) = \mathbf{v}_{\mathbf{k}ss'}$. Note that commutator $\big[D^{\alpha},v^{\beta}\big]_{ss'}$ is but the canonical commutation relation, $\big[\hat{r}^{\alpha},\hat{p}^{\beta}\big] = i\hbar\delta^{\alpha\beta}\hat{1}$, expressed in the Bloch basis [1],

$$\big[D^{\alpha},v^{\beta}\big] = \frac{\hbar}{m_e}\delta^{\alpha\beta}\delta_{ss'}, \tag{3.123}$$

since $\hat{r}^{\alpha} = iD^{\alpha}$ and $v^{\alpha} = p^{\alpha}/m_e$. We can then write the velocity matrix elements in the velocity gauge as,

$$\mathbf{v}^{A}_{\mathbf{k}ss'}(t) = \mathbf{v}_{\mathbf{k}ss'} + \frac{|e|}{m_e}\mathbf{A}(t), \tag{3.124}$$

since the commutator of the covariant derivative with the identity matrix is necessarily zero — terms that are second order in the field (and above) simply vanish. This is exactly the matrix element that enters the expression for the current density in the velocity gauge we derived above, Eq.(3.95).

3.5.1 The Density Matrix

The relation between density matrix elements in the two gauges is tied to a derivation that is somewhat more intricate than that for the matrix elements $\mathcal{O}^{\alpha}_{\mathbf{k}ss'}$. As we have seen in Section 3.3, the density matrix, despite being a matrix in band-index space, has its matrix elements written in terms of a full many-body density matrix, ρ^{α}, and an *ensemble* average, Eq.(3.85). We must, therefore, begin this study by considering the full many-body density matrix,

$$\rho^{A}(t) = \sum_{n} p_n \left|\psi^{A}_{n}(t)\right\rangle\left\langle\psi^{A}_{n}(t)\right|. \tag{3.125}$$

The $\left|\psi^{A}_{n}(t)\right\rangle$ is a many-body state vector whose dynamics are given by the many-body Schrödinger equation, Eq.(3.22). It can be expressed in terms of the state vectors of its counterpart description by means of the unitary transformation we derived at the beginning of this chapter, Eq.(3.19). We can write the relation between operators as,

$$\rho^{A}(t) = \mathcal{U}^{\dagger}(t)\,\rho^{E}(t)\,\mathcal{U}(t), \tag{3.126}$$

and the density matrix elements in the velocity gauge as,

$$\rho^A_{\mathbf{k}ss'}(t) = \mathrm{tr}\big[\rho^A(t) c^\dagger_{\mathbf{k}s'} c_{\mathbf{k}s}\big] = \mathrm{tr}\big[\mathcal{U}^\dagger(t)\,\rho^E\,\mathcal{U}(t) c^\dagger_{\mathbf{k}s'} c_{\mathbf{k}s}\big]. \tag{3.127}$$

Using the cyclic property of the trace, we can move the unitary transformation around, so that it instead acts on the creation and destruction operators of Bloch states,

$$\rho^A_{\mathbf{k}ss'}(t) = \mathrm{tr}\big[\rho^E(t)\,\mathcal{U}(t) c^\dagger_{\mathbf{k}s'}\,\underbrace{\mathcal{U}^\dagger(t)\mathcal{U}(t)}_{\hat{1}}\,c_{\mathbf{k}s}\,\mathcal{U}^\dagger(t)\big] = \mathrm{tr}\big[\rho^E(t)\,\tilde{c}^\dagger_{\mathbf{k}s'}\,\tilde{c}_{\mathbf{k}s}\big]. \tag{3.128}$$

The density matrix elements in the velocity gauge are therefore expressed in terms of an ensemble average of the full many-body density matrix in the *length* gauge, ρ^E, and creation and destruction operators that have been appropriately modified: $\tilde{c}_{\mathbf{k}s}$ and $\tilde{c}^\dagger_{\mathbf{k}s}$. The new creation operator, $\tilde{c}^\dagger_{\mathbf{k}s'}$, is,

$$\tilde{c}^\dagger_{\mathbf{k}s'} = \mathcal{U}(t)\,c^\dagger_{\mathbf{k}s'}\,\mathcal{U}^\dagger(t) = \int d^d\mathbf{r}\,\psi_{\mathbf{k}s'}(\mathbf{r})\,\mathcal{U}(t)\Psi^\dagger(\mathbf{r})\,\mathcal{U}^\dagger(t). \tag{3.129}$$

It follows from the expression for the unitary transformation, Eq.(3.26), that the effect this transformation has on the fermionic field, $\Psi^\dagger(\mathbf{r})$, is,

$$\mathcal{U}(t)\Psi^\dagger(\mathbf{r})\mathcal{U}^\dagger(t) = \Psi^\dagger(\mathbf{r}) + i\frac{|e|}{\hbar}\mathbf{A}(t)\cdot\int d^d\mathbf{r}'\,\mathbf{r}'\,\big[\Psi^\dagger(\mathbf{r}')\Psi(\mathbf{r}'),\Psi^\dagger(\mathbf{r})\big] + (....), \tag{3.130}$$

$$= \Psi^\dagger(\mathbf{r}) + i\frac{|e|}{\hbar}\mathbf{A}(t)\cdot\int d^d\mathbf{r}'\,\mathbf{r}'\,\delta(\mathbf{r}'-\mathbf{r})\Psi^\dagger(\mathbf{r}) + (....), \tag{3.131}$$

that is,

$$\mathcal{U}(t)\Psi^\dagger(\mathbf{r})\mathcal{U}^\dagger(t) = \exp\left[i\frac{|e|}{\hbar}\mathbf{A}(t)\cdot\mathbf{r}\right]\Psi^\dagger(\mathbf{r}). \tag{3.132}$$

Note that we have used the anti-commutation relations for the fermionic fields, Eqs.(3.12) and (3.13). The operators marked with a tilde are not associated with the eigenfunctions of the crystal Hamiltonian, $\psi_{\mathbf{k}s}(\mathbf{r})$,

$$\tilde{c}^\dagger_{\mathbf{k}s'} = \int d^d\mathbf{r}\,\Phi_{\mathbf{k}s'}(\mathbf{r})\,\Psi^\dagger(\mathbf{r}). \tag{3.133}$$

Instead, they are associated with the following functions,

$$\Phi_{\mathbf{k}s'}(\mathbf{r}) := e^{ie\mathbf{r}\cdot\mathbf{A}(t)/\hbar}\psi_{\mathbf{k}s'}(\mathbf{r}) = e^{i\mathbf{r}\cdot\left(\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\right)}u_{\mathbf{k}s'}(\mathbf{r}). \tag{3.134}$$

These new $\Phi_{\mathbf{k}s'}(\mathbf{r})$ functions, while they are Bloch functions of wave vector $\mathbf{q} = \mathbf{k} + |e|\hbar^{-1}\mathbf{A}(t)$ — note that they behave as,

$$\Phi_{\mathbf{k}s'}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{r}\cdot\left(\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\right)}e^{i\mathbf{R}\cdot\left(\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\right)}u_{\mathbf{k}s'}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{R}\cdot\left(\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\right)}\Phi_{\mathbf{k}s'}(\mathbf{r}), \tag{3.135}$$

under a lattice translation — they are *not* eigenfunctions of the Hamiltonian: the $\mathbf{k}$ parameter in the plane wave factor is different from that of the periodic function, $u_{\mathbf{k}s'}(\mathbf{r})$. We can, nevertheless, write these states as a linear combination of shifted Bloch states:

$$|\Phi_{\mathbf{k}s'}\rangle = \sum_{r'} |\psi_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r'}\rangle\langle u_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r'}|u_{\mathbf{k}s'}\rangle. \tag{3.136}$$

Furthermore, the relation between $\Phi_{\mathbf{k}s'}$ states and $\psi_{\mathbf{k}s'}$ states implies the following relation between the creation operator, $\tilde{c}^\dagger_{\mathbf{k}s'}$, and the creation operator of eigenstates of the Hamiltonian, $c^\dagger_{\mathbf{k}s'}$,

$$\tilde{c}^\dagger_{\mathbf{k}s'} = \sum_{r'} \langle u_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r'}|u_{\mathbf{k}s'}\rangle\, c^\dagger_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r'}. \tag{3.137}$$

We can, therefore, express the density matrix elements (in the velocity gauge) in terms of an ensemble average of the full many-body density matrix in the length gauge, ρ^E, and a pair of shifted operators with crystal momentum variable $\mathbf{k}+|e|\,\hbar^{-1}\mathbf{A}(t)$, i.e., in terms of the matrix elements of the density matrix in the *length* gauge, $\rho^E_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,rr'}(t)$, Eq.(3.85),

$$\rho^A_{\mathbf{k}ss'}(t) = \sum_{r'r}\langle u_{\mathbf{k}s}|u_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r}\rangle\rho^E_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,rr'}(t)\,\langle u_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r'}|u_{\mathbf{k}s'}\rangle, \tag{3.138}$$

since,

$$\mathrm{tr}\big[\rho^E(t)c^\dagger_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r'}c_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r}\big] = \langle c^\dagger_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r'}c_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,r}\rangle_E = \rho^E_{\mathbf{k}+|e|\hbar^{-1}\mathbf{A}(t)\,rr'}(t) \tag{3.139}$$

Alternatively, this equality can be written in operator form (see Eqs.(3.59) and (3.60)), so that,

$$\boxed{\rho^A(\mathbf{k},t) = \rho^E\left(\mathbf{k}+\frac{|e|}{\hbar}\mathbf{A}(t),t\right)}. \tag{3.140}$$

This shows that the relation between the density matrices in the two gauges is that of all other matrices in band-index space, Eq.(3.117).

3.5.2 Sum Rules and Their Limitations

We can now return to the expectation value of a given observable in the velocity gauge and see how it is related to its length-gauge counterpart. Consider,

$$\langle\mathcal{O}^A(t)\rangle = \int_{\mathrm{FBZ}}\frac{d^d\mathbf{k}}{(2\pi)^d}\mathrm{Tr}\left[\mathcal{O}^A(\mathbf{k})\rho^A(\mathbf{k},t)\right]. \tag{3.141}$$

It follows from the relations between band-index matrices in the two gauges, Eqs.(3.117) and (3.140), that the matrices in this expectation value are equal to their length-gauge counterparts when their

argument is adequately translated,

$$\langle \mathcal{O}^A(t) \rangle = \int_{\mathrm{FBZ}} \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\left[\mathcal{O}^E\left(\mathbf{k} + \frac{|e|}{\hbar}\mathbf{A}(t) \right) \rho^E\left(\mathbf{k} + \frac{|e|}{\hbar}\mathbf{A}(t), t \right) \right]. \tag{3.142}$$

Note that the argument of the integrand function has been fully shifted by a constant (in the sense that $|e|\,\hbar^{-1}\mathbf{A}(t)$ is the same throughout the first Brillouin zone at any given time), and that this is tantamount to a translation of the domain of integration — where the centre is shifted from the origin, $\mathbf{k} = \mathbf{0}$, to $\mathbf{k} = |e|\,\hbar^{-1}\mathbf{A}(t)$. This new domain of integration (a shifted first Brillouin zone, FBZ′) is entirely equivalent to the original one (the FBZ) because the integrand function, as a trace over a product of band-index space matrices, carries the periodicity of the FBZ. The relation between expectation values in the two gauges reduces to,

$$\langle \mathcal{O}^A(t) \rangle = \int_{\mathrm{FBZ}'} \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\left[\mathcal{O}^E(\mathbf{k}) \rho^E(\mathbf{k}, t) \right] = \int_{\mathrm{FBZ}} \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\left[\mathcal{O}^E(\mathbf{k}) \rho^E(\mathbf{k}, t) \right] = \langle \mathcal{O}^E(t) \rangle. \tag{3.143}$$

This shows that gauge invariance is preserved in actual calculations of nonlinear optical responses of electrons in crystals, provided that certain conditions are fulfilled. Let us return to Eq.(3.142). If we expand the integrand in powers of vector potential, $\mathbf{A}(t)$, by using Eq.(3.120), we can write $\langle \mathcal{O}^A(t) \rangle$ directly in terms of $\langle \mathcal{O}^E(t) \rangle$, as well as some additional contributions that are proportional to the vector potential,

$$\langle \mathcal{O}^A(t) \rangle = \int_{\mathrm{FBZ}} \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\left[\mathcal{O}^E(\mathbf{k})\, \rho^E(\mathbf{k}, t) \right] + \frac{|e|}{\hbar} A^{\alpha_1}(t) \int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\left[\left[D^{\alpha_1}, \mathcal{O}^E(\mathbf{k}, t)\, \rho^E(\mathbf{k}, t) \right] \right] + (....), \tag{3.144}$$

that is,

$$\langle \mathcal{O}^A(\mathbf{k}, t) \rangle = \langle \mathcal{O}^E(\mathbf{k}, t) \rangle + \frac{|e|}{\hbar} A^{\alpha_1}(t) \int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\left[\left[D^{\alpha_1}, \mathcal{O}^E(\mathbf{k}, t)\, \rho^E(\mathbf{k}, t) \right] \right] + (....). \tag{3.145}$$

The principle of gauge invariance — here stated in terms of Eq.(3.143) — dictates that all terms other than $\langle \mathcal{O}^E(t) \rangle$ in the right-hand side of Eq.(3.145) *must* be zero. The integrals associated with the first and second order terms are necessarily zero,

$$\int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\left[\left[D^{\alpha_1}, \mathcal{O}^E(\mathbf{k}, t)\, \rho^E(\mathbf{k}, t) \right] \right] = 0, \tag{3.146}$$

$$\int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\left[\left[D^{\alpha_2}, \left[D^{\alpha_1}, \mathcal{O}^E(\mathbf{k}, t)\, \rho^E(\mathbf{k}, t) \right] \right] \right] = 0. \tag{3.147}$$

These two vanishing integrals — which, we might add, are not patently trivial — are part of what are commonly called *sum rules*; that of Eq.(3.146) is similar to the original sum rule derived in ref.[22], as mentioned in the introduction. They can be expressed in general terms if we introduce

an arbitrary matrix in band-index space, $\mathrm{G}(\mathbf{k},t)$,

$$\int \frac{d^d\mathbf{k}}{(2\pi)^d}\mathrm{Tr}\big[\big[D^{\alpha_1},\mathrm{G}(\mathbf{k},t)\big]\big], \tag{3.148}$$

that can involve any number of nested commutators with the covariant derivative,

$$\mathrm{G}_{\mathbf{k}ss'}(t) := \big[D^{\alpha_2},...,\big[D^{\alpha_n},\big[\mathcal{O}^E(\mathbf{k},t)\,\rho^E(\mathbf{k},t)\big]\,...\big]\big]_{ss'}. \tag{3.149}$$

Using Eq.(3.71), we can decompose the integral in Eq.(3.148) into two terms. The first term carries a derivative of the diagonal matrix elements of $\mathrm{G}(\mathbf{k},t)$. Since $\mathrm{G}(\mathbf{k},t)$ is periodic in reciprocal space, the integral of its derivative over the FBZ is necessarily zero.[25] The second term carries a trace over the commutator of the Berry connection with $\mathrm{G}(\mathbf{k},t)$; using the cyclic invariance of the trace, we can exchange the order of the matrices and show that this commutator is also necessarily zero. We have, therefore, proved that these sum rules exist,

$$\sum_s \int \frac{d^d\mathbf{k}}{(2\pi)^d}\nabla^{\alpha_1}_{\mathbf{k}}\mathrm{G}_{\mathbf{k}ss}(t) - i\int \frac{d^d\mathbf{k}}{(2\pi)^d}\mathrm{Tr}\big[\big[\xi^{\alpha_1},G(\mathbf{k},t)\big]\big] = 0,$$

provided that the integral over $\mathbf{k}$ is performed over the full first Brillouin zone *and* that the trace running through different bands is carried over a complete basis. This establishes the conditions under which the sum rules — for all orders — hold true, i.e., the conditions under which the two gauges are indeed equivalent. If these no longer hold, as is the case when the band-space is truncated — by the selection of a specific region of the FBZ, or a specific subset of bands, or both — this equivalence between gauges is *irretrievably* lost.

Finally, we must append some important comments to this result. While it is clear that the equivalence between gauges can be broken and, perhaps more importantly, why it can be broken, this argument does not tell us exactly which of these gauges can be used when the above conditions are relaxed, and why. In light of our previous comments, the reader can infer, rather confidently at this point, that this equivalence is broken precisely because of what happens to the velocity gauge under a band-space truncation. We finish this section with a discussion of the limitations of the velocity gauge; to do so, we recover the expressions for the density matrix elements we derived in Subsection 3.4.2.

[25] Consider the one-dimensional case. For $g(x+a)=g(x)$, and $g(x)$ a smooth function, with well-defined $g^n(x)$ for all positive integers n, we can write,

$$\int_0^a dx g(x+\lambda) = \int_0^a dx g(x) + \lambda\int_0^a dx\frac{dg(x)}{dx} + \frac{\lambda^2}{2}\int_0^a dx\frac{d^2g(x)}{dx^2} + (...) = \int_0^a dx g(x),$$

since,

$$\int_0^a dx\frac{d^ng}{dx^n} = \frac{d^{(n-1)}g}{dx'^{(n-1)}}\bigg|_{x'=a} - \frac{d^{(n-1)}g}{dx'^{(n-1)}}\bigg|_{x'=0} = 0,$$

for all positive integers n. This recovers the argument presented in the subsection on the covariant derivative.

Figure 3.2: A conceptual picture of the effective Hamiltonian describing the bands in subspace $\mathcal{E}_0$. The Fermi level lies somewhere in that subspace. For bands inside $\mathcal{E}_0$, the energy difference is of order δ, while the energy difference between bands in different subspaces is of order $\Delta \gg \delta$. Figure from ref.[2].

3.5.3 Band-space Truncations and the Limitations of the Velocity Gauge

As we have mentioned throughout this chapter, band-space truncations involve the *a-posteriori* selection of a specific subset of eigenfunctions, and their respective eigenvalues, of the crystal Hamiltonian[26] — i.e., the singling out of some region of the first Brillouin zone or a collection of bands, or both — so that they form a second, effective, Hamiltonian. The reasoning behind these truncations is that certain physical properties — e.g., those directly associated with the density of states — can be accurately described using a specific portion of the energy spectrum. While it is certainly reasonable to expect that calculations of the nonlinear optical response would also fall into this category, it happens that those that follow from a velocity-gauge description *fail to do so*. To show this, let us consider a band structure that can transcribe this concept of an effective Hamiltonian, which is pictorially represented in Fig.3.2. In this picture, (seemingly) relevant bands belong to the subspace labelled $\mathcal{E}_0$, which has a total energy bandwidth δ, and (seemingly) irrelevant bands belonging to two other subspaces that differ in energy from the first subspace by Δ, for $\Delta \gg \delta$. In addition, the Fermi level lies somewhere in $\mathcal{E}_0$ and the frequencies involved in the optical response are such that $\hbar\omega_i \ll \Delta$, and $\hbar\omega_{[i]} \ll \Delta$. This is the system whose optical response we will now consider. For the optical response itself, let us return to the expressions for the density matrix elements — Eqs.(3.113) and (3.114) — derived in Subsection 3.4.2:

$$\rho^{(n),E}_{\mathbf{k}ss'}(\omega) = (i\,|e|)^n \left[\prod_{i=1}^{n-1} \int \frac{d\omega_i}{2\pi}\, E^{\alpha_i}(\omega_i)\right] E^{\alpha_n}(\omega - \omega_{[n-1]}) \\ \times d_{\mathbf{k}ss'}(\omega)\big[D^{\alpha_1},\, d(\omega-\omega_1)\circ\big[...,\, d(\omega-\omega_{[n-1]})\circ\big[D^{\alpha_n},\rho^{(0)}(\mathbf{k})\big]...\big]\big]_{ss'},$$

[26] Note that what is meant by crystal Hamiltonian here is not a specific Hamiltonian — such as the continuum Hamiltonian introduced in Subsection 3.3 — but *any* Hamiltonian that is defined over a full first Brillouin zone and whose eigenstates make up a complete basis, i.e., a Hamiltonian for which the sum rules hold true.

for the length gauge, and

$$\rho^{(n),A}_{\mathbf{k}ss'}(\omega) = |e|^n \left[\prod_{i=1}^{n-1}\int \frac{d\omega_i}{2\pi} A^{\alpha_i}(\omega_i)\right] A^{\alpha_n}(\omega-\omega_{[n-1]}) \\ \times d_{\mathbf{k}ss'}(\omega)\big[v^{\alpha_1},\, d(\omega-\omega_1)\circ\big[...,d(\omega-\omega_{[n-1]})\circ\big[v^{\alpha_n},\rho^{(0)}(\mathbf{k})\big]...\big]\big]_{ss'},$$

for the velocity gauge. Note that the current response, as given by Eqs.(3.94) and (3.95), also carries the matrix elements of the velocity operator in the unperturbed system, Eq.(3.76).[27] In addition, recall that the $d_{\mathbf{k}ss'}(\omega)$ represent different energy denominators,

$$d_{\mathbf{k}ss'}(\omega) = \frac{1}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'} + i\eta}.$$

Our first focus is, by way of juxtaposition, the length gauge. The n-th order contribution to the current in this gauge involves a trace over a product of $n+1$ matrices in band space,

$$\sum_{s's} v^{\beta}_{\mathbf{k}s's} d_{\mathbf{k}ss'}(\omega)\big[D^{\alpha_1},\, d(\omega-\omega_1)\circ\big[...\big[D^{\alpha_n},\rho^{(0)}(\mathbf{k})\big]...\big]\big]_{ss'}, \tag{3.150}$$

and that carries n denominators. These denominators can be divided into two types: intraband denominators, in which the energy difference, $\Delta\epsilon_{\mathbf{k}ss}$, is trivial. These reduce to numerical factors that can be removed from the integrand; interband denominators, in which the energy difference, $\Delta\epsilon_{\mathbf{k}ss'}$, is nontrivial. It follows from Fig.3.2 that the ones describing interband transitions are either associated with transitions between bands inside $\mathcal{E}_0$, in which case,

$$d_{\mathbf{k}ss'}(\omega) = \frac{1}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}},\ \Delta\epsilon_{\mathbf{k}ss'} \sim \delta, \tag{3.151}$$

or, between bands inside $\mathcal{E}_0$ and those outside,

$$d_{\mathbf{k}ss'}(\omega) = \frac{1}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}},\ \Delta\epsilon_{\mathbf{k}ss'} \sim \Delta, \tag{3.152}$$

or even between bands outside $\mathcal{E}_0$,[28]

$$d_{\mathbf{k}ss'}(\omega) = \frac{1}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}},\ \Delta\epsilon_{\mathbf{k}ss'} \sim 2\Delta. \tag{3.153}$$

[27] It also carries an integral over $\mathbf{k}$, a sum over these two band indexes and some additional numerical factors.

[28] Note that this is true only for transitions between bands in which one band is empty — that is, a band inside the subspace (energetically) above $\mathcal{E}_0$ — and the other one is filled — that is, a band inside the subspace (energetically) below $\mathcal{E}_0$ (or vice-versa). Transitions between bands outside $\mathcal{E}_0$ that are in the same subspace — of either completely empty, or completely filled, bands — are not particularly relevant to the argument that we are making here.

As for the other set of objects involved in this integrand — the matrix elements in the numerator of Eq.(3.150) — these too can be separated into intraband and interband contributions, following from the expression for the covariant derivative, Eq.(3.53),

$$\mathbf{D}_{\mathbf{k}ss'} = \delta_{ss'}\nabla_{\mathbf{k}} - i\boldsymbol{\xi}_{\mathbf{k}ss'}.$$

The different terms in the integrand can be divided by the number of derivatives with respect to $\mathbf{k}$ (or intraband contributions) they carry, or by their lack thereof (these would correspond to interband contributions); take the full intraband term, in which we pick the derivative portion of every covariant derivative. This term will involve, for the n-th order response, an n number of derivatives acting on the Fermi-Dirac distribution function — since $\rho^{(0)}_{\mathbf{k}ss'} = \delta_{ss'}f_{\mathbf{k}s}$ — so that it is nonzero only for the band crossed by the Fermi level. Since, by construction, the Fermi level lies somewhere in $\mathcal{E}_0$, bands outside $\mathcal{E}_0$ do not contribute to the integral associated with this full intraband term. Now, note that all other terms in the integrand will carry a difference of occupation factors, and *at least one* interband contribution. For terms carrying a *single* interband contribution — which is to say, those associated with a single pair of running band indexes, s and s' — the relevant transitions are such that s and s' correspond to filled and empty bands, respectively (or vice-versa). This means that there can be three types of terms, which are associated with the aforementioned types of transitions: (a) interband transitions between filled and empty bands inside $\mathcal{E}_0$; (b) interband transitions between filled (empty) bands in $\mathcal{E}_0$ and empty (filled) bands outside $\mathcal{E}_0$; (c) interband transitions between filled and empty bands, both outside $\mathcal{E}_0$.

Consider the product $v^{\beta}_{\mathbf{k}s's}d_{\mathbf{k}ss'}(\omega)$ in Eq.(3.150). For a (b)-type transition — s' belongs to a filled band inside $\mathcal{E}_0$, and s to an empty band outside $\mathcal{E}_0$ (or vice-versa) — we can express this product as,

$$v^{\beta}_{\mathbf{k}s's}d_{\mathbf{k}ss'}(\omega) \sim \frac{i}{\hbar}\frac{\Delta\epsilon_{\mathbf{k}ss'}\xi^{\beta}_{\mathbf{k}s's}}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}} = -\frac{i}{\hbar}\xi^{\beta}_{\mathbf{k}s's},$$

since $\Delta\epsilon_{\mathbf{k}ss'} \sim \Delta \gg \hbar\omega$. For $n \geq 2$,[29] the matrix that is on the right of this product (in Eq.(3.150)) must include *at least one* energy denominator with an energy difference of order Δ, because the last band index has to be the same as the first, s', so as to complete the trace. Terms for this type of transition are, therefore, suppressed when $\Delta \gg \delta$. Furthermore, for a (c)-type transition — s' and s refer to filled and empty bands outside $\mathcal{E}_0$, respectively — the matrix to the right of

[29] Note that for $n = 1$, this integrand would read,

$$\sum_{s's}\xi^{\beta}_{\mathbf{k}s's}\xi^{\alpha_1}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's} = \sum_{s's}[\xi^{\beta}_{\mathbf{k}s's}, \xi^{\alpha_1}_{\mathbf{k}ss'}]f_{\mathbf{k}s's},$$

for $[\xi^{\beta}_{\mathbf{k}s's}, \xi^{\alpha_1}_{\mathbf{k}ss'}] = \xi^{\beta}_{\mathbf{k}s's}\xi^{\alpha_1}_{\mathbf{k}ss'} - \xi^{\alpha_1}_{\mathbf{k}s's}\xi^{\beta}_{\mathbf{k}ss'}$. This object, as we will note later on, is odd under $\mathbf{k} \to -\mathbf{k}$; the corresponding integral is thus trivial.

$v^{\beta}_{\mathbf{k}s's}d_{\mathbf{k}ss'}(\omega)$ must again carry an energy denominator with energy differences of order Δ.[30] Terms for this type of transition are also suppressed when $\Delta \gg \delta$. In other words, terms for transitions of types (b) and (c) must have at least one energy denominator of order $\mathcal{O}(1/\Delta)$, so they must be suppressed when $\Delta \gg \delta$, meaning that their contributions to the response are necessarily negligible. We can, therefore, conclude that in the case of $\Delta \gg \delta$, the dominant terms with a *single* interband contribution come from bands inside $\mathcal{E}_0$, i.e. they are terms connected with transitions of type (a). This reasoning can then be extended to all remaining terms — those carrying an $n \geq 2$ number of interband contributions. Its consequences are clear: only terms carrying transitions of type (a) are relevant to the nonlinear optical response (in addition to the full intraband term). A truncation to this subspace of bands is thus a valid approximation in a *length-gauge* description of the optical response.[31]

Let us now consider this same argument in the velocity gauge, where the integrand reads as,

$$\sum_{s's} v^{\beta}_{\mathbf{k}s's}d_{\mathbf{k}ss'}(\omega)\big[v^{\alpha_1},\, d(\omega-\omega_1)\circ\big[\ldots\big[v^{\alpha_n},\rho^{(0)}(\mathbf{k})\big]\ldots\big]\big]_{ss'}. \tag{3.154}$$

In this case, every energy denominator is associated with a velocity matrix element, Eq.(3.76). This means that a transition involving energy differences of order Δ, $\epsilon_{rr'} \sim \mathcal{O}(\Delta)$ — which will, of course, carry a denominator that reads roughly as $d_{rr'}(\omega) \sim \mathcal{O}(1/\Delta)$ — will also carry an interband contribution of the velocity matrix element, which reads roughly as $v^{\beta}_{r'r} \sim \xi^{\beta}_{rr'}\mathcal{O}(\Delta)$: large energy differences in the denominators are thus associated with large energy differences in the numerators. As a result, terms for transitions of types (b) and (c) are not suppressed, and they can give relevant contributions to the response, regardless of how large Δ is. The velocity gauge does not, therefore, allow for band-space truncations. Calculations in this gauge can only be performed if the Hamiltonian used is defined over a complete FBZ, and if its basis of eigenstates can be deemed complete, i.e., it must be somewhat similar to Hamiltonians that follow from a tight-binding description. Furthermore, we note that the canonical commutation relation does not, in principle, hold for this type of Hamiltonian, which means that the coupling with the external electric field is not described as neatly as in the present description of the velocity gauge. Instead, it follows from the general relation between operators in the two gauges, Eq.(3.120), with the added caveat that for the Hamiltonian — as we have noted in Section 3.2 — this is not a relation between the velocity- and length-gauge operators, *but* between the velocity-gauge operator and its unperturbed counterpart.

[30] Note that if both s' and s refer to filled (or empty) bands outside $\mathcal{E}_0$, then the matrix on the right of $v^{\beta}_{\mathbf{k}s's}d_{\mathbf{k}ss'}(\omega)$ must involve not one, but *two* energy denominators with energy differences of order Δ.

[31] This is why the length gauge is considered to be the most appropriate description for analytical calculations of NLO responses. However, and as pointed out in ref.[3], calculations in this gauge are "only really tractable for simple effective continuum Hamiltonians." We will return to this point in the subsequent chapter.

3.6 Closing Remarks

Now that we have shown that the equivalence between gauges is predicated on the existence of sum rules and that these sum rules, as well as velocity-gauge descriptions, are susceptible to band-space truncations, the natural next step would be to derive expressions for the velocity gauge that can be used in practical calculations. This is the subject of another of our group's work — ref.[3] — and it is not a subject of this thesis. We will, however, review two of its main results — the expressions for the velocity gauge and how these can be put to use — in the following chapter, where we study the nonlinear optical response (in second and third orders) of the plain graphene and gapped graphene monolayers.

Chapter 4

Practical Calculations in the Velocity Gauge

This chapter is based on the following publication:

- *A study of the nonlinear optical response of the plain graphene and gapped graphene monolayers beyond the Dirac approximation*, **G. B. Ventura**, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, *J. Phys.: Condens. Matter* **32** 185701

This chapter is dedicated to the application of the velocity-gauge description of ref.[3] in the calculation of the nonlinear optical response of the monolayers of plain and gapped graphene. It is based on a full-band description of these materials, i.e., a description that allows one to go beyond the usual Dirac approximation and study their high-frequency response. Furthermore, it includes two additional topics, which are relevant to the study of these NLO responses: an overview of the relation between the current response (to a monochromatic electric field) and the nonlinear optical conductivities; a discussion of two important aspects concerning the use of tight-binding Hamiltonians in calculations of nonlinear optical responses. As for the hasty review of the velocity-gauge description introduced in ref.[3], which we have also added to this chapter, it largely follows that work in its content and scope.

As before, we begin this chapter with a short, somewhat historical, introduction to the subject at hand. This should provide the reader with the necessary context, as well as serve as an overview of the literature on calculations of the nonlinear optical response of the monolayers of plain and gapped graphene. We then proceed to discuss the ideas of the aforementioned work [5], and again, present them in slightly more illustrative terms than in the paper.

4.1 Introduction

For what is now well over a decade, the nonlinear optical response of graphene has been the subject of almost unbounded attention from both experimental [75, 76, 78, 82, 86, 150–153], and theoretical [59, 64–70, 72, 73, 80, 83] groups, which have been keen to study it, describe it, and to find practical uses for it: graphene-based nonlinear photonics [154], ultrafast optoelectronic devices [86], ultrathin graphene lenses [155], as well as a wealth of other applications [156], have been mentioned as interesting possible applications of this material's nonlinear optical properties. Such has been the attention devoted to the nonlinear optical response of graphene that it has managed to find itself at the centre of its own controversy [157–159], which, for the moment, appears to have been settled.[1] Like graphene (and, probably, as a result of the attention that it has gathered), the two-dimensional materials that have been isolated since the discovery of this allotrope of carbon — hexagonal boron nitride (hBN), the transition metal dichalcogenides (TMDs), phosphorene, etc. — have also had their nonlinear properties studied and (at least tentatively) described [57, 98, 160–164]. Unlike graphene, however, these materials all have gapped band structures, and are usually characterized as semiconductors (TMDs and phosphorene) or insulators (hBN);[2] this is quite consequential, especially for the study of their optical response. The gapped band structure of these materials, as well as their two-dimensional character — there are no atoms in the out-of plane direction, which means that the screening of the Coulomb interaction is partially suppressed [94, 165] — leads to an optical response that is fundamentally different from that of graphene. The latter can be suitably described by models that neglect the Coulomb interaction between electrons [64, 65, 150, 151]. The optical response of gapped materials in two dimensions requires, on the other hand, a treatment based on excitons, i.e., bound states of negatively charged (conduction band) electrons and positively charged (valence band) holes that are created by the combined action of an external optical field, which pumps valence band electrons into the conduction band, and the Coulomb attraction between oppositely-charged particles [15, 94]; in these circumstances, and for photon energies below the band gap, the optical response has little to do with that of "uncharged" electrons in a crystal. It instead resembles that of an atomic system, with well-defined resonances corresponding to transitions between different excitonic states. Excitonic responses are, however, outside the scope of this thesis; the optical responses studied, described, and in some instances calculated, in this work are those of noninteracting electrons. This, as we have seen in the previous chapter, and as we will again see in the present one, is not without its relevance and usefulness.

We have previously discussed, in extremely brief terms, that the renewed interest in the issue

[1] This dispute concerns whether graphene is an exceptional nonlinear optical material, which is the position held by S. A. Mikhailov. J. B. Khurgin, on the other hand, argues that graphene is a "rather ordinary nonlinear optical", whose practical application will have to do with "considerations of mechanical and thermal robustness, manufacturability, and cost."

[2] The band gap, Δ, i.e., the energy difference between the minimum of the conduction band and the maximum of the valence band, for each of these materials is: $\Delta_{\text{TMDs}} \sim 1 - 2\,\text{eV}$ [95]; $\Delta_{\text{Phos.}} \sim 1 - 1.75\,\text{eV}$ [162, 163, 165]; $\Delta_{\text{hBN}} \sim 6\,\text{eV}$ [57, 161].

of gauge invariance, and how it affected nonlinear optical responses of crystals, stemmed from the discovery of graphene's exceptional optical properties. The first theoretical works on this subject — the "quasi-classical" descriptions of S. A. Mikhailov [66, 67] — focused on the intraband, low-frequency (microwave and low infrared), response of the monolayer of graphene. This was expanded on several years later, when (almost) fully-quantum calculations of the response were advanced that aimed at describing the interband, high-frequency (infrared and low visible), response of this material, as well as the intraband one [59, 68, 69, 73].[3] Using the length-gauge description of the nonlinear optical response, these calculations employed the Dirac model — an effective Hamiltonian that describes the physics of graphene around the $\mathbf{K}$ and $\mathbf{K}'$ points of its first Brillouin zone, which is simple enough so that the NLO responses can be performed analytically[4] — to derive *fully analytical* expressions for the complete third order optical response; this, considering the complexity of the problem involved, that is, the number of contributions that must be taken into account, and the type of integrals that must be performed over $\mathbf{k}$ [59], is an astonishing feat.[5]

There is also, however, a very clear limitation to these calculations, one that is particularly important, considering that there is interest in describing nonlinear phenomena in which the output frequency can be several times larger than the input frequency: the frequency range that these calculations can probe is restricted by the energy bandwidth, $\Delta\epsilon_{cv}^{\max}$, that is appropriately described by the Dirac model (in monolayer graphene, $\Delta\epsilon_{cv}^{\max} \approx 2\,\mathrm{eV}$). In quantitative terms, this means that, for the specific case of the third harmonic generation, where $\omega_{\mathrm{out}} = 3\,\omega_{\mathrm{in}}$, input frequencies should be restricted to about,

$$\nu_{\mathrm{in}}^{\max} = \frac{\Delta\epsilon_{cv}^{\max}}{3h} \rightarrow \nu_{\mathrm{in}}^{\max} \approx 4.85 \times 10^{14}\mathrm{Hz}, \tag{4.1}$$

that is to the near infrared. In order to calculate optical responses for frequencies in the visible or the UV, which is to say, going beyond the Dirac approximation, one must return to a full-band model of graphene; the most immediate description of this is, of course, the nearest-neighbours tight-binding model [61]. It would thus seem, at least at face value, that the problem of determining the high-frequency response of graphene runs *directly* into a second problem, which is that of using the length gauge in a full-band numerical calculation [3].

The first attempt at squaring this particular circle [166], used the length gauge in a full-band calculation but it did not, however, compute physically relevant response functions. Instead, the authors used effective nonlinear conductivities, to which Kleinman's symmetry had been imposed, in their calculations, so as to deal with "spurious divergences". Since this symmetry follows from the condition that the nonlinear susceptibilities (or conductivities) can be deemed *dispersionless* [12], its application lacks any justification in the case of the nonlinear optical response of graphene

[3] By this we mean that the dynamics of the electronic system are described in terms of a quantum density matrix, while the electric field is considered to be classical.

[4] This approximation, it should be said, is extremely accurate for band energies up to 1 eV.

[5] We note that now this task does not look as challenging as it once was. Using the general expressions for the two-band model derived in ref.[10], one can easily derive the third order response associated with the Dirac model; in fact, D. J. Passos demonstrates this in his work.

in the frequency range considered in that work. Furthermore, divergences in one's nonlinear optical response should not be spirited away by using additional, nonessential, symmetries; they either have a clear physical interpretation — e.g., Drude peaks in the low-frequency intraband response, the injection current in the second order DC response of certain noncentrosymmetric materials, etc. — or they are a sign that there is something potentially incorrect in one's calculations.

The solution to this problem, as it had been foreseen in ref.[3], was thus to be found in the velocity gauge, where full-band models *must* be used in the calculations.

Therefore, the present chapter of this thesis is, for the most part, dedicated to our work on the nonlinear optical response of graphene beyond the Dirac approximation [5]. Here, we present the numerical results for the second and third order responses of the plain graphene (PG) and gapped (GG) graphene monolayers to a monochromatic electric field of frequencies (energies) ranging from the microwave ($\hbar\omega \sim 0.005$ eV) to the ultraviolet ($\hbar\omega \sim 6$ eV).[6] We have also found it relevant to include, as we did in the paper, three additional topics: an overview of the current response in second and third orders to a monochromatic field, and their relation with certain nonlinear optical conductivities; a hasty review of the velocity-gauge description introduced in ref.[3]; a discussion of two important aspects concerning the use of tight-binding Hamiltonians in calculations of nonlinear optical responses, namely, the calculation of h coefficients, and the issue of choosing the representation of the position operator in the Wannier basis.

This chapter is organized as follows. The subsequent section is dedicated to the first two (additional) topics that we have mentioned above; Section 4.3 is dedicated to the third. In Section 4.4, we present the results for the nonlinear optical response of the PG and GG monolayers — the second harmonic generation and optical rectification conductivities associated with the GG monolayer, and the harmonic generation and Kerr effect conductivities associated with the PG and GG monolayers — in the aforementioned frequency regime. For the two second order effects, the results are complemented by analytical calculations of the real part of the conductivities that make use of the length-gauge description of Chapter 3.

4.2 Effects and Practical Calculations in the Velocity Gauge

A unifying aspect of this thesis is the use of perturbation theory in the description of a crystal's current response — whether this is expressed as an average of a quantum observable or, as we will see in the following chapter, as the result of a semiclassical description of the electronic populations — which is, for a large part, formulated in terms of nonlinear optical coefficients. A quick look at the relations between the current response for a monochromatic electric field and the nonlinear optical coefficients — for the second and third order responses — is therefore pertinent to our

[6] We note, as we did in the paper, that these results should be taken only as an indication of what the response should look like, as they were calculated in the independent particle approximation and thus do not consider excitonic effects [167, 168].

presentation of the subject at hand, and will certainly prove itself useful to the reader that is not familiar with this topic. Furthermore, we point out that this general analysis should not, by any means, preclude an analysis that takes into account the point group symmetries of specific crystals; the latter should certainly serve as a complement to the former. It is only in this way that we can, for example, understand why effects like the injection current are absent from the second order response of gapped graphene.

4.2.1 Response to a Monochromatic Field

Let us begin by considering the expression for the second order current response in the frequency domain, Eq.(2.24), now written in terms of the electric current and the conductivity,

$$J^{\beta}(\omega) = \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)E^{\alpha_1}(\omega_1)E^{\alpha_2}(\omega_2)(2\pi)\delta(\omega_1+\omega_2-\omega). \tag{4.2}$$

The corresponding response in the time domain is just the Fourier transform of this object,

$$J^{\beta}(t) = \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)E^{\alpha_1}(\omega_1)E^{\alpha_2}(\omega_2)e^{-i(\omega_{12}+2i\eta)t}. \tag{4.3}$$

It involves the convolution of a nonlinear optical coefficient and a product of two external electric fields. These fields, depending on the characteristics of the pulses under consideration, can have frequency profiles that are rather intricate, which can make the analysis of the current response at this order extremely difficult. However, when they are monochromatic, i.e.,

$$E^{\alpha_1}(\omega_1) = E^{\alpha_1}(\omega)\delta(\omega_1-\omega) + E^{\alpha_1}(-\omega)\delta(\omega_1+\omega), \tag{4.4}$$

the regular part of this response becomes quite simple,

$$J^{\beta}_{\text{reg.}}(t) = \sigma^{\beta\alpha_1\alpha_2}(\omega,\omega)E^{\alpha_1}(\omega)E^{\alpha_2}(\omega)e^{-i2\omega t} + \sigma^{\beta\alpha_1\alpha_2}_{\text{shi.}}(\omega,-\omega)E^{\alpha_1}(\omega)E^{\alpha_2}(-\omega) + \text{c.c.}, \tag{4.5}$$

for $\sigma^{\beta\alpha_1\alpha_2}(\omega,\omega)$ and $\sigma^{\beta\alpha_1\alpha_2}_{\text{shi.}}(\omega,-\omega)$, the nonlinear optical coefficients that describe the second harmonic generation and the shift current. As the nomenclature suggests, there is an extra element in the current response that is not well-behaved, which is associated with a finite, and *constant*, time derivative of $J^{\beta}(t)$ — the injection current [22, 28, 43, 52],

$$\dot{J}^{\beta}_{\text{inj.}}(t) = \lim_{\eta\to 0^+} 2\eta\,\sigma^{\beta\alpha_1\alpha_2}_{\text{div.}}(\omega,-\omega)E^{\alpha_1}(\omega)E^{\alpha_2}(-\omega) + \text{c.c.}. \tag{4.6}$$

Later on, we will take a closer look at the physical interpretation of this current — how it follows from the existence of asymmetric carrier (both electrons and holes) populations on *opposite* sides of the FBZ — and why it can only be generated by elliptically polarised light. For now, it suffices to

say that this effect is associated with a divergent contribution to the conductivity (which is why the limit in Eq.(4.6) is not zero), and that it does not exist when the external electric field is linearly polarised; this latter point, as we have noted before, is of particular relevance for gapped graphene. We can also see from the expressions in Eqs.(4.5) and (4.6) that the full second order current response to a monochromatic field is predicated solely on the knowledge of several conductivities, which can be computed using the procedures described in this work. The exact number of coefficients will, of course, depend on the point group symmetries of the crystal of interest; for centrosymmetric materials, as we have seen in our brief study of point group symmetries, all second order coefficients are necessarily trivial. Their lowest order nonlinear optical response is a third order one.

Generalizing the previous expression for the third order current response, Eq.(4.3), $J^{\beta}(t)$, we obtain,

$$J^{\beta}(t) = \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\frac{d\omega_3}{2\pi}\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3)E^{\alpha_1}(\omega_1)E^{\alpha_2}(\omega_2)E^{\alpha_3}(\omega_3)e^{-i(\omega_{123}+3i\eta)t}, \tag{4.7}$$

so that, for a monochromatic field, it reduces to,

$$\begin{aligned} J^{\beta}(t) = {} & \sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega,\omega,\omega)E^{\alpha_1}(\omega)E^{\alpha_2}(\omega)E^{\alpha_3}(\omega)e^{-i3\omega t} \\ & + 3\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega,\omega,-\omega)E^{\alpha_1}(\omega)E^{\alpha_2}(\omega)E^{\alpha_3}(-\omega)e^{-i\omega t} + \text{c.c.}, \end{aligned} \tag{4.8}$$

by using the fact that the conductivities are intrinsically symmetric, Eq.(2.30). Note that there are no DC — or zero frequency — terms in this response, which, nevertheless, does *not* imply that this current is completely free from any divergent terms. In this case, the divergent term in the response (as before, there is a single one) appears as a specific one-photon contribution to what is usually called the Kerr effect, $\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega,\omega,-\omega)$. However, and unlike its second order counterpart, the appearance of this divergence in the optical Kerr effect is contingent solely on the frequency of the external electric field, and not on the way this field is polarised.

Finally, let us emphasize the fact that the perturbative description of the current response to a monochromatic field is formulated in terms of a small number of conductivity tensors: two[7] in second order, and two in third order. We can compute these conductivities either analytically — by using the length-gauge description — or numerically — by using the velocity-gauge description presented in ref.[3][8] — depending on the properties of the model being used, that is, (among other things) whether it satisfies the aforementioned conditions imposed on the velocity gauge, and the properties of the external electric field as well. We thus turn to a quick overview of practical calculations in the velocity gauge — and what these must entail — and will then move on to some particular points concerning the use of tight-binding Hamiltonians in these calculations. With these

[7] Three, if we separate the injection current from its regular counterpart, the shift current.

[8] Note that the length gauge can also be used for numerical calculations. We are simply alluding to the fact that, as we have argued in ref.[3], the velocity gauge "should prove more efficient for numerical computations."

elements in place, we will be able to study the high-frequency nonlinear optical response of the PG and GG monolayers.

4.2.2 Practical Calculations in the Velocity Gauge

In the previous chapter, we showed that the velocity gauge is susceptible to band-space truncations, and noted that the description derived for the continuum Hamiltonian — where the set of bands is necessarily infinite — was of very little, if any, practical use. In order to formulate a description of the velocity gauge that can be used in practical calculations [3], we have to return to the general relation between the Hamiltonian in the velocity gauge and the unperturbed Hamiltonian, which we briefly discussed at the end of Subsection 3.5.3. From there, we will be able to derive the equations of motion for the matrix elements of the density matrix, which is to say, a practical description — in the velocity gauge — of the dynamics of the system of crystal electrons subjected to an external electric field.

If we think of a crystal Hamiltonian in general terms, and by this we mean any Hamiltonian that satisfies the conditions laid out in the previous chapter (the Hamiltonian must be defined over a full FBZ and its eigenstates must form a complete basis), the matrix elements of its velocity-gauge counterpart read, in the basis of periodic functions, as an infinite sum,

$$\mathcal{H}^{A}_{\mathbf{k}ss'}(t) = \sum_{n=0}^{+\infty} \frac{1}{n!} \left(\frac{|e|}{\hbar}\right)^{n} A^{\alpha_1}(t) \dots A^{\alpha_n}(t) \big[D^{\alpha_1}, \big[..., \big[D^{\alpha_n}, \mathcal{H}_0\big]...\big]\big]_{ss'}. \tag{4.9}$$

The first term in this sum is the crystal Hamiltonian, though the notation makes this somewhat opaque; collectively, the remaining terms form the perturbation, which can be expressed as,

$$V^{A}_{\mathbf{k}ss'}(t) = \sum_{n=0}^{+\infty} \frac{1}{n!} |e|^{n} A^{\alpha_1}(t) \dots A^{\alpha_n}(t) h^{\alpha_1 \dots \alpha_n}_{\mathbf{k}ss'}. \tag{4.10}$$

for $h^{\alpha_1 \dots \alpha_n}_{\mathbf{k}ss'}$, the different nested commutators of the covariant derivative, Eq.(3.53), with the unperturbed Hamiltonian [3],

$$h^{\alpha_1 \dots \alpha_n}_{\mathbf{k}ss'} = \frac{1}{\hbar^n} \big[D^{\alpha_1}, \big[..., \big[D^{\alpha_n}, \mathcal{H}_0\big]...\big]\big]_{ss'}. \tag{4.11}$$

As expected, the first term in this collection of coefficients is the velocity matrix element in the unperturbed system, $v^{\alpha_1}_{\mathbf{k}ss'}$. For the continuum Hamiltonian, this term and the term with two indexes are the only two relevant nested commutators. In that specific case, this latter term reads as,

$$h^{\alpha_1 \alpha_2}_{\mathbf{k}ss'} = \frac{1}{\hbar^2} \big[D^{\alpha_1}, \big[D^{\alpha_2}, \mathcal{H}_0\big]\big]_{ss'} = \frac{1}{\hbar} \big[D^{\alpha_1}, v^{\alpha_2}\big]_{ss'} = \frac{1}{m_e} \delta^{\alpha_1 \alpha_2} \delta_{ss'}, \tag{4.12}$$

by using Eq.(3.123). This is, unsurprisingly, consistent with the expression for the velocity matrix elements we derived in Eq.(3.124) for the continuum Hamiltonian. As for the general expression for

the velocity matrix elements, it instead reads as,

$$v^{A,\beta}_{\mathbf{k}ss'}(t) = v^{\beta}_{\mathbf{k}ss'} + \sum_{n=0}^{+\infty} \frac{1}{n!} |e|^n A^{\alpha_1}(t) \dots A^{\alpha_n}(t) h^{\beta\alpha_1\dots\alpha_n}_{\mathbf{k}ss'}. \tag{4.13}$$

Note that here the h coefficients carry an additional index, which is labelled β. This also recovers the expression derived in Eq.(3.122), but is now cast in slightly different notation.

We can now consider the expression for the expectation value of the current density, Eq.(3.95),

$$\big\langle \mathbf{J}_A(t) \big\rangle = -|e| \int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathrm{Tr}\Big[\mathbf{v}^A(\mathbf{k},t)\, \rho^A(\mathbf{k},t)\Big] = -|e| \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{s's} \mathbf{v}^A_{\mathbf{k}s's}(t)\, \rho^A_{\mathbf{k}ss'}(t). \tag{4.14}$$

Both the velocity matrix elements and the density matrix elements depend on the vector potential, and an order-by-order study of the current response must therefore involve an expansion in the vector potential of these two sets of matrix elements [3]. In quantitative terms, it means that the n-th order contribution to the response is the sum of all possible combinations of the product of $\mathbf{v}^{A,(n-p)}_{\mathbf{k}ss'}$ — the $(n-p)$-th order contribution to the velocity matrix element — and $\rho^{A,(p)}_{\mathbf{k}s's}$ — the p-th order contribution to the velocity matrix element, and whose total order in the vector potential is, of course, n,

$$\big\langle \mathbf{J}_A(t) \big\rangle^{(n)} = -|e| \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{p=0}^{n} \sum_{s's} \mathbf{v}^{A,(n-p)}_{\mathbf{k}s's}(t)\, \rho^{A,(p)}_{\mathbf{k}ss'}(t). \tag{4.15}$$

This additional complication follows from the form of the perturbation, Eq.(4.10). Furthermore, it should be recalled that the equations of motion for the (full) density matrix elements also depend on the full perturbation, Eq.(3.89), so that their solutions involve the same sort of accounting that goes into writing the expression for the current density. Writing explicit expressions for the current response in the general velocity gauge is therefore much more difficult than doing so in the corresponding length gauge [3]. There are, nonetheless, several distinct advantages to the use of the velocity gauge: its equations of motion, as we will presently see, remain decoupled in $\mathbf{k}$; as a consequence of this decoupling, there are no higher order poles in the conductivities, e.g. $(\hbar\omega + i\eta - \Delta\epsilon_{\mathbf{k}ss'})^{-n}$, $n \geq 2$, since derivatives do not act directly on denominators. These two aspects of the velocity gauge make it the suitable tool for numerical calculations [3].

The equations of motion for the velocity gauge now read,

$$\Big[i\hbar\frac{\partial}{\partial t} - \Delta\epsilon_{\mathbf{k}ss'}\Big]\rho^A_{\mathbf{k}ss'}(t) = \Big[V(t), \rho^A(\mathbf{k},t)\Big]_{ss'}, \tag{4.16}$$

with the full perturbation, Eq.(4.10), replacing the $\mathbf{A}(t)\cdot\mathbf{v}$ term of Eq.(3.93). Breaking these matrix elements into their different order terms, gives us, for the first two nontrivial orders of the field, the

following equations of motion,

$$\left[i\hbar\frac{\partial}{\partial t}-\Delta\epsilon_{\mathbf{k}ss'}\right]\rho^{(1),A}_{\mathbf{k}ss'}(t) = |e|\,A^{\alpha_1}(t)\left[v^{\alpha_1},\rho^{(0)}(\mathbf{k})\right]_{ss'}, \tag{4.17}$$

$$\begin{aligned}\left[i\hbar\frac{\partial}{\partial t}-\Delta\epsilon_{\mathbf{k}ss'}\right]\rho^{(2),A}_{\mathbf{k}ss'}(t) = {}& |e|\,A^{\alpha_1}(t)\left[v^{\alpha_1},\rho^{(1),A}(\mathbf{k},t)\right]_{ss'}\\ &+\frac{1}{2}e^2A^{\alpha_1}(t)A^{\alpha_2}(t)\left[h^{\alpha_1\alpha_2},\rho^{(0)}(\mathbf{k})\right]_{ss'}.\end{aligned} \tag{4.18}$$

Two points — which we have mentioned above — are made clear here: the matrix elements at order n, $\rho^{(n)}_{\mathbf{k}ss'}$, depend on matrix elements, $\rho^{(m)}_{\mathbf{k}ss'}$, of every order below n ($m<n$); these equations of motion, as those in Eq.(3.93), do not couple different values of $\mathbf{k}$. Now that we have derived the equations of motion, we can repeat the procedure outlined in Subsection 3.4.2 and write their solutions in the frequency domain. From there, we can determine the expressions for the current density, and then write the expressions for the respective conductivities, similarly to what have done in the case of the length gauge. The solutions to Eqs.(4.17) and (4.18) are,

$$\rho^{(1),A}_{\mathbf{k}ss'}(\omega) = |e|\,d_{\mathbf{k}ss'}(\omega)A^{\alpha_1}(\omega)v^{\alpha_1}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}, \tag{4.19}$$

$$\begin{aligned}\rho^{(2),A}_{\mathbf{k}ss'}(\omega) = {}& |e|^2\int\frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}A^{\alpha_1}(\omega_1)A^{\alpha_2}(\omega_2)d_{\mathbf{k}ss'}(\omega_{12})\Big[\left[v^{\alpha_1},d(\omega_2)\circ\left[v^{\alpha_2},\rho^{(0)}(\mathbf{k})\right]\right]_{ss'}\\ &+\frac{1}{2}h^{\alpha_1\alpha_2}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}\Big](2\pi)\delta(\omega-\omega_{12}).\end{aligned} \tag{4.20}$$

The factors of $d_{\mathbf{k}ss'}(\omega)$ represent, once again, energy denominators:

$$d_{\mathbf{k}ss'}(\omega)=\frac{1}{\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'}+i\eta}.$$

We can now combine our knowledge of the velocity matrix elements (for every order) with that of the density matrix elements (for the first two orders), and write the first two terms in the order-by-order expansion of the current density, Eq.(4.15). The first order current density is quite simple,

$$J^{(1),\beta}_A(\omega)=-|e|^2\int\frac{d\omega_1}{2\pi}\int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Big[\frac{1}{\hbar\omega_1-\Delta\epsilon_{\mathbf{k}ss'}+i\eta}v^{\beta}_{\mathbf{k}s's}v^{\alpha_1}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}+h^{\beta\alpha_1}_{\mathbf{k}ss}f_{\mathbf{k}s}\delta_{ss'}\Big]\,A^{\alpha_1}(\omega_1)(2\pi)\delta(\omega-\omega_1).$$

It just involves replacing the diamagnetic term that follows from the canonical commutation relation, Eq.(4.12), with $h^{\beta\alpha_1}_{\mathbf{k}ss}$, the nested commutator of the covariant derivative with the velocity operator. The second order current density, however, is much more complex,

$$J_A^{(2),\beta}(\omega) = -|e|^3 \int \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Big[\frac{1}{\hbar\omega_{12}-\Delta\epsilon_{\mathbf{k}ss'}+2i\eta}v^{\beta}_{\mathbf{k}s's}\big[v^{\alpha_1},\frac{1}{\hbar\omega_{12}-\Delta\epsilon_{\mathbf{k}}+i\eta}\circ\big[v^{\alpha_2},\rho^{(0)}(\mathbf{k})\big]\big]_{ss'}$$
$$+\frac{1}{2}\frac{1}{\hbar\omega_{12}-\Delta\epsilon_{\mathbf{k}ss'}+2i\eta}v^{\beta}_{\mathbf{k}s's}h^{\alpha_1\alpha_2}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}+\frac{1}{\hbar\omega_2-\Delta\epsilon_{\mathbf{k}ss'}+i\eta}h^{\beta\alpha_1}_{\mathbf{k}s's}v^{\alpha_2}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}+\frac{1}{2}h^{\beta\alpha_1\alpha_2}_{\mathbf{k}ss}f_{\mathbf{k}s}\delta_{ss'}\Big]$$
$$\times A^{\alpha_1}(\omega_1)A^{\alpha_2}(\omega_2)(2\pi)\delta(\omega-\omega_1-\omega_2). \tag{4.21}$$

Note that this expression carries, in addition to the usual $\mathbf{v}^{(0)}\rho^{(2)}$ term, two new terms corresponding to $\mathbf{v}^{(1)}\rho^{(1)}$ and $\mathbf{v}^{(2)}\rho^{(0)}$. We have also introduced a new notational element. Sums of frequencies are now expressed as,

$$\omega_{12} := \omega_1+\omega_2, \tag{4.22}$$

so that,

$$\omega_{123} := \omega_1+\omega_2+\omega_3, \tag{4.23}$$

etc.

Finally, we can write the expressions for the first and second order conductivities in the velocity gauge, by recovering two results from Chapter 2, Eqs.(2.7) and (2.24), and by noting that the relation between the vector potential and the electric field, Eq.(3.8), reads, in the frequency domain, as,

$$A^{\alpha_1}(\omega_1) = -\frac{1}{i\omega_1}E^{\alpha_1}(\omega_1). \tag{4.24}$$

The expressions, therefore, read as,

$$\sigma_A^{\beta\alpha_1}(\omega_1) = \frac{|e|^2}{i\omega_1}\int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\frac{1}{\hbar\omega_1-\Delta\epsilon_{\mathbf{k}ss'}+i\eta}v^{\beta}_{\mathbf{k}s's}v^{\alpha_1}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}, \tag{4.25}$$

and,

$$\sigma_A^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2) = \frac{|e|^3}{(i\omega_1)(i\omega_2)}\int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Big[\frac{1}{\hbar\omega_{12}-\Delta\epsilon_{\mathbf{k}ss'}+2i\eta}v^{\beta}_{\mathbf{k}s's}\big[v^{\alpha_1},\frac{1}{\hbar\omega_{12}-\Delta\epsilon_{\mathbf{k}}+i\eta}\circ\big[v^{\alpha_2},\rho^{(0)}(\mathbf{k})\big]\big]_{ss'}$$
$$+\frac{1}{2}\frac{1}{\hbar\omega_{12}-\Delta\epsilon_{\mathbf{k}ss'}+2i\eta}v^{\beta}_{\mathbf{k}s's}h^{\alpha_1\alpha_2}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}+\frac{1}{\hbar\omega_2-\Delta\epsilon_{\mathbf{k}ss'}+i\eta}h^{\beta\alpha_1}_{\mathbf{k}s's}h^{\alpha_2}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}$$
$$+\frac{1}{2}h^{\beta\alpha_1\alpha_2}_{\mathbf{k}ss}f_{\mathbf{k}s}\delta_{ss'}\Big], \tag{4.26}$$

As before, we have introduced an imaginary part to the denominators by means of the adiabatic switching, $\omega_i \to \bar{\omega}_i = \omega_i + i\eta$. This ensures that the equivalence between length- and velocity-gauge descriptions is maintained without having to introduce additional, complicated, terms to the equations of motion of the velocity gauge [3]. In addition, note that this expression for the second order conductivity is not intrinsically symmetric; for simplicity, we will keep it so.

4.3 Velocity Gauge for Tight-Binding Hamiltonians

As we have seen in the previous section, a general description of the response in the velocity gauge at order n is predicated on the knowledge of the h coefficients up to order $n+1$. These coefficients can, in principle, involve rather complicated calculations, as they are expressed in terms of nested commutators with the covariant derivative. In the present section, we show that one can easily compute the h coefficients at any order, without having to use the covariant derivative, by introducing a new basis of Bloch functions.[9] We also show how subtle changes in the definition of the position operator can influence the description of a system's optical response. For simplicity (as well as practical reasons), we will work exclusively with tight-binding Hamiltonians.

4.3.1 Covariant Derivatives in a New Bloch Basis

Tight-binding models are simplified descriptions of the physics of electrons in a lattice, where their motion is characterized by discrete jumps (or hoppings) from one localized orbital to its neighbouring ones, $t_{ij}\left(\mathbf{R}_n, \mathbf{R}_m\right)$ (i and j are orbital indexes), that may have certain, not necessarily equal, on-site energies, ϵ_i. In real space, it can be generally expressed as,

$$\mathcal{H} = \sum_{\mathbf{R}_n,\mathbf{R}_m} \sum_{i,j} \left[t_{ij}(\mathbf{R}_n,\mathbf{R}_m) \left|\phi_{\mathbf{R}_n\, i}\right\rangle\!\left\langle\phi_{\mathbf{R}_m\, j}\right| + \text{H.c.}\right] + \sum_{\mathbf{R}_n} \sum_{i} \epsilon_i(\mathbf{R}_n) \left|\phi_{\mathbf{R}_n\, i}\right\rangle\!\left\langle\phi_{\mathbf{R}_n\, i}\right|. \tag{4.27}$$

Here, the state $\left|\phi_{\mathbf{R}_n\, i}\right\rangle$ represents a Wannier orbital centred at $\mathbf{R}_n + \boldsymbol{\lambda}_i$, with $\boldsymbol{\lambda}_i$ being the vector from that i-orbital site to the unit-cell origin. The eigenvalues of this periodic Hamiltonian are, of course, bands, $\epsilon_{\mathbf{k}s}$, which are labelled by a certain $\mathbf{k}$ (in the FBZ) and a band index, s, and its eigenfunctions are Bloch functions, $\psi_{\mathbf{k}s}(\mathbf{r})$. We now show that by introducing a new basis of Bloch functions, $\tilde{\psi}_{\mathbf{k}i}(\mathbf{r})$ — which are associated with the Wannier orbitals of a specific orbital index i, $\phi_{\mathbf{R}_n\, i}(\mathbf{r})$ — the nested commutator with the covariant derivative can be rendered trivial and, consequently, the calculation of h coefficients can be rendered trivial as well.

These new Bloch states,

$$\left|\tilde{\psi}_{\mathbf{k}i}\right\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_n} e^{i\mathbf{k}\cdot(\mathbf{R}_n+\boldsymbol{\lambda}_i)} \left|\phi_{\mathbf{R}_n\, i}\right\rangle, \tag{4.28}$$

satisfy Bloch's theorem. The inner product of two of these states,

$$\langle\tilde{\psi}_{\mathbf{k}i}|\tilde{\psi}_{\mathbf{k}j}\rangle = \frac{1}{N} \sum_{\mathbf{R}_n,\mathbf{R}_m} e^{-i\mathbf{k}\cdot(\mathbf{R}_n-\mathbf{R}_m+\boldsymbol{\lambda}_i-\boldsymbol{\lambda}_j)} \underbrace{\langle\phi_{\mathbf{R}_n\, i}|\phi_{\mathbf{R}_m\, i}\rangle}_{\delta_{\mathbf{R}_n,\mathbf{R}_m}\delta_{i,j}} = \delta_{i,j}, \tag{4.29}$$

shows that, like the eigenstates of the Hamiltonian and the Wannier states, this set of states also

[9] This is not, of course, without its caveats. Here, we assume that the set of Wannier states forms an orthogonal basis, and that these states are eigenstates of the position operator.

forms a basis. Furthermore, we can see that an eigenfunction of a Hamiltonian like the one in Eq.(4.27) is a linear combination of these functions,

$$|\psi_{\mathbf{k}s}\rangle = \sum_i c_{\mathbf{k}s,i} \big|\tilde{\psi}_{\mathbf{k}i}\big\rangle, \tag{4.30}$$

for $c_{\mathbf{k}s,i}$, the solutions to the eigenvector problem for that particular value of $\mathbf{k}$.

Secondly, note that the Wannier functions are usually considered to be eigenstates of the position operator,

$$\mathbf{r}\big|\phi_{\mathbf{R}_n\, i}\big\rangle = (\mathbf{R}_n + \boldsymbol{\lambda}_i)\, \big|\phi_{\mathbf{R}_n\, i}\big\rangle. \tag{4.31}$$

It follows from this consideration that the periodic factor of the tilded Bloch wavefunctions,

$$\big|\tilde{u}_{\mathbf{k}i}\big\rangle = e^{-i\mathbf{k}\cdot\mathbf{r}}\big|\tilde{\psi}_{\mathbf{k}i}\big\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_n} e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R}_n+\boldsymbol{\lambda}_i)}\big|\phi_{\mathbf{R}_n\, i}\big\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_n} \big|\phi_{\mathbf{R}_n\, i}\big\rangle, \tag{4.32}$$

does not depend on $\mathbf{k}$. This means that the corresponding Berry connections — for every possible set of orbital indexes, i and j — *must be trivial*,

$$\tilde{\boldsymbol{\xi}}_{\mathbf{k}ij} = i\, \langle \tilde{u}_{\mathbf{k}i}|\, \nabla_{\mathbf{k}} \tilde{u}_{\mathbf{k}j}\rangle \to \tilde{\boldsymbol{\xi}}_{\mathbf{k}ij} = 0. \tag{4.33}$$

The covariant derivative $\mathbf{D}_{\mathbf{k}}$, therefore, reduces to the regular derivative $\nabla_{\mathbf{k}}$, and the matrix element of the derivative of an operator, Eq.(3.63), is now just the derivative of matrix element of that operator,

$$\langle \tilde{u}_{\mathbf{k}i}|\, \nabla_{\mathbf{k}} \mathcal{O}(\mathbf{k})\, |\tilde{u}_{\mathbf{k}j}\rangle = [\mathbf{D}, \mathcal{O}(\mathbf{k})]_{ij} = (\nabla_{\mathbf{k}} \mathcal{O}_{\mathbf{k}ij}) \tag{4.34}$$

since the term carrying the commutator of the operator $\mathcal{O}$ with the Berry connection is, for $\xi^{\alpha}_{\mathbf{k}ij} = 0$ — for all i and j — necessarily zero,

$$[\boldsymbol{\xi}, \mathcal{O}(\mathbf{k})]_{ij} = 0. \tag{4.35}$$

The calculation of h coefficients is then fairly simple, as these only involve a certain number of nested commutators of the covariant derivative with the unperturbed Hamiltonian. Using the completeness relation for the basis of tilded Bloch states, we can write Eq.(4.11) — the expression for the h coefficient at order p — as,

$$h^{\alpha_1 \ldots \alpha_p}_{\mathbf{k}ss'} = \sum_{i,j} \langle u_{\mathbf{k}s}|\, \tilde{u}_{\mathbf{k}i}\rangle\, \langle \tilde{u}_{\mathbf{k}i}|\, \nabla^{\alpha_1}_{\mathbf{k}} \ldots \nabla^{\alpha_p}_{\mathbf{k}} H_{\mathbf{k}} \big|\tilde{u}_{\mathbf{k}j}\big\rangle\, \langle \tilde{u}_{\mathbf{k}j}|\, u_{\mathbf{k}s'}\rangle = \sum_{i,j} c_{\mathbf{k}s,i} \left(\nabla^{\alpha_1}_{\mathbf{k}} \ldots \nabla^{\alpha_p}_{\mathbf{k}} H_{\mathbf{k}ij}\right) c^{*}_{\mathbf{k}s',j}\,. \tag{4.36}$$

We can now easily determine all the coefficients involved in the nonlinear optical response, provided, of course, that the $H_{\mathbf{k}ij}$ can be straightforwardly derived, and their derivatives (at whatever order) easily computed. This is exactly the case with Hamiltonians that follow from a tight-binding

description; they can be readily expressed in terms of tilded Bloch states (by means of a Fourier transform), which gives one an analytical expression for the matrix elements, $H_{\mathbf{k}ij}$ — the matrix elements will include either some hopping parameters and their corresponding phase factors, or an on-site energy. Computing the derivatives of these matrix elements is thus trivial.

This procedure for computing the h coefficients can, therefore, have a significant impact on how numerical calculations of nonlinear optical conductivities are performed. Instead of diagonalizing the analytical Hamiltonian, one can solve the eigenvalue/eigenvector problem numerically — knowing, beforehand, all relevant h coefficients at a certain order of the response — at the time that one is evaluating (numerically, of course) all the different matrix elements that appear in the nonlinear optical conductivity. This process is faster and simpler than computing the h coefficients directly from the eigenstates of the Hamiltonian.

Finally, the Berry connection, Eq.(3.43),

$$\boldsymbol{\xi}_{\mathbf{k}ss'} = i \left\langle u_{\mathbf{k}s} \right| \nabla_{\mathbf{k}} \left| u_{\mathbf{k}s'} \right\rangle = i \sum_j c_{\mathbf{k}s,j} \left(\nabla_{\mathbf{k}} c^*_{\mathbf{k}s',j} \right), \tag{4.37}$$

is expressed directly in terms of the solutions of the eigenvector problem, since,

$$\left(\nabla_{\mathbf{k}} \left| u_{\mathbf{k}i} \right\rangle \left\langle u_{\mathbf{k}i} \right| u_{\mathbf{k}s'} \right\rangle) = \left| u_{\mathbf{k}i} \right\rangle \left(\nabla_{\mathbf{k}} \left\langle u_{\mathbf{k}i} \right| u_{\mathbf{k}s'} \right\rangle) = \left| u_{\mathbf{k}i} \right\rangle \left(\nabla_{\mathbf{k}} c_{\mathbf{k}s',i} \right). \tag{4.38}$$

While this does not concern the computation of h coefficients, it does play a relevant role in issue of choosing the representation of one's tight-binding Hamiltonian. This latter point is the subject of the following subsection.

4.3.2 The Representation of the Tight-Binding Hamiltonian

The second subject of this section concerns tight-binding Hamiltonians and their corresponding nonlinear optical response; it does not pertain solely to the velocity gauge, and one can, at least in principle, use this same description (and the insights of this section) for calculations in the length gauge. For simplicity, we present the following discussion in terms of the nearest-neighbours tight-binding model for the plain graphene (PG) monolayer, since it will be used in our description of the nonlinear optical responses of both this material and gapped graphene (GG). The hopping parameter is set to 3 eV in this section, and throughout the rest of this chapter [3, 166].

The tight-binding Hamiltonian in the tilded Bloch basis is usually written in two different ways, i.e., it has two different representations, which can be condensed in a single matrix as,

$$H_{\mathbf{k},(a/\delta)} = \begin{bmatrix} 0 & (-t)\ \phi_{(\delta/a)}(\mathbf{k}) \\ (-t)\ \phi^*_{(\delta/a)}(\mathbf{k}) & 0 \end{bmatrix}. \tag{4.39}$$

Note the subscript in each ϕ. The first of these functions, $\phi_{(\delta)}$, comes directly from the definition

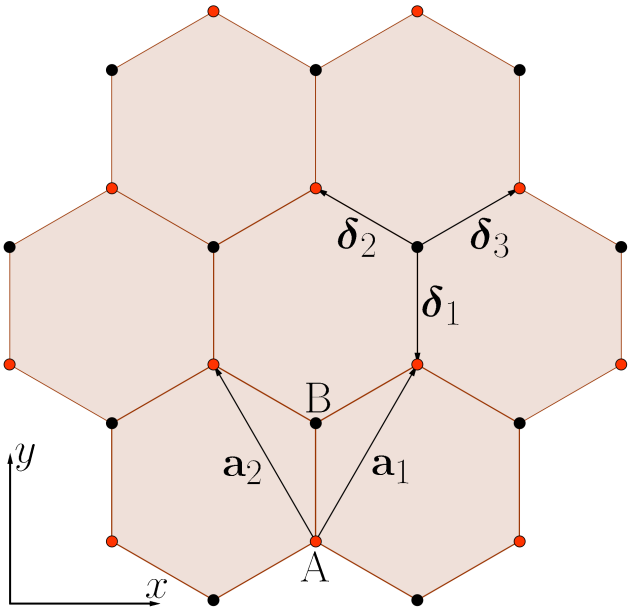


Figure 4.1: The honeycomb lattice of graphene with lattice parameter $|\boldsymbol{\delta}_1| = a_0$ and the armchair in the vertical ($\hat{y}$) direction. We have represented both the lattice vectors — $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ — and the vectors connecting a B atom to its nearest-neighbours — $\boldsymbol{\delta}_1$, $\boldsymbol{\delta}_2$, $\boldsymbol{\delta}_3$. In plain graphene (PG), the A and B atoms are equivalent; in gapped graphene (GG), they are not.

of the second Bloch basis as that in Eq.(4.28),

$$\phi_{(\delta)}(\mathbf{k}) = e^{-i\mathbf{k}\cdot\boldsymbol{\delta}_1} + e^{-i\mathbf{k}\cdot\boldsymbol{\delta}_2} + e^{-i\mathbf{k}\cdot\boldsymbol{\delta}_3}, \tag{4.40}$$

and is expressed in terms of the vectors connecting a B atom to its three nearest-neighbours, Figure 4.1. A second way of writing this Hamiltonian, now in terms of the $\phi_{(a)}$ functions, is associated with a second Bloch basis that has its states phase-shifted with respect to those of Eq.(4.28),

$$\left|\bar{\psi}_{\mathbf{k}i}\right\rangle = \sum_{\mathbf{R}_n} e^{i\mathbf{k}\cdot\mathbf{R}_n} \left|\phi_{\mathbf{R}_n\, i}\right\rangle. \tag{4.41}$$

The phase factors that enter the matrix elements are now written in terms of the lattice vectors, $\mathbf{a}_1$ and $\mathbf{a}_2$ [3, 59, 68],

$$\phi_{(a)}(\mathbf{k}) = 1 + e^{i\mathbf{k}\cdot(\mathbf{a}_2 - \mathbf{a}_1)} + e^{i\mathbf{k}\cdot\mathbf{a}_2}. \tag{4.42}$$

Note that the ϕ functions are related by a simple phase factor,

$$\phi_{(a)}(\mathbf{k}) = e^{i\mathbf{k}\cdot\boldsymbol{\delta}_2}\, \phi_{(\delta)}(\mathbf{k}), \tag{4.43}$$

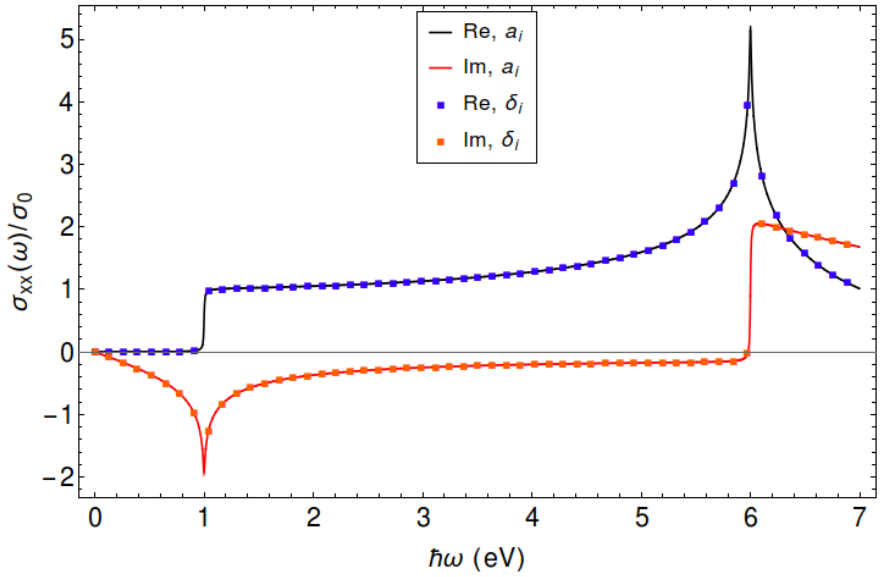


Figure 4.2: The real and imaginary parts of the interband portion of the linear conductivity, $\sigma_{xx}(\omega)$, for tight-binding Hamiltonians written in the lattice-vectors (a_i) and nearest-neighbours (δ_i) representations. The relevant parameters here are $\mu = 0.5$ eV, $\gamma = 0.005$ eV and $T = 1$ K. The conductivity is normalized with respect to $\sigma_0 = e^2/4\hbar$. Plot from ref.[5].

which means that the two representations (the nearest-neighbours and lattice-vectors representations) have the exact same eigenvalues. There is, however, an important subtlety: when we use Eq.(4.36) to compute the h coefficients, or Eq.(4.37) to compute the Berry connections, we obtain different results for each representation. Let us consider this more closely.

We begin by noting that it could appear that the condition for a trivial Berry connection in the entire FBZ, Eq.(4.33), which follows from the definition of the position operator of Eq.(4.31), is *not* satisfied in the $\bar{\psi}_{\mathbf{k}i}$ basis, that is, the basis associated with the Hamiltonian expressed in terms of the lattice vectors $H_{\mathbf{k},(a)}$, since,

$$\big|\bar{u}_{\mathbf{k}i}\big\rangle = e^{-i\mathbf{k}\cdot\mathbf{r}}\big|\bar{\psi}_{\mathbf{k}i}\big\rangle = \frac{1}{\sqrt{N}}\sum_{\mathbf{R}_n} e^{-i\mathbf{k}\cdot\boldsymbol{\delta}_i}\big|\phi_{\mathbf{R}_n\, i}\big\rangle = e^{-i\mathbf{k}\cdot\boldsymbol{\delta}_i}\big|\tilde{u}_{\mathbf{k}i}\big\rangle, \tag{4.44}$$

carries a phase factor that depends on $\mathbf{k}$. However, if we redefine the position operator, so as to effectively neglect distances inside the unit cell, $|\boldsymbol{\lambda}_i|$,

$$\mathbf{r}\big|\phi_{\mathbf{R}_n\, i}\big\rangle = \mathbf{R}_n\big|\phi_{\mathbf{R}_n\, i}\big\rangle, \tag{4.45}$$

we obtain barred periodic states that are independent of $\mathbf{k}$. As expected, the Berry connections in the bases of tilded and barred periodic states can be made trivial only when two different, and necessarily incompatible, choices of representation of the position operator are used; this, as we will

now see, is not without its consequences.[10]

The linear optical responses associated with each representation — for both the zig-zag and armchair directions of graphene, whose lattice is represented in Figure 4.1 — are represented in Figures 4.2 and 4.3.

Consider the linear response along the zig-zag (or $\hat{x}$) direction, $\sigma_{xx}(\omega)$, represented in Figure 4.2.[11] In this case, the two responses are exactly the same. Both $\tilde{\xi}^{x}_{\mathbf{k}ij}$ and $\bar{\xi}^{x}_{\mathbf{k}ij}$ are zero, since $\boldsymbol{\delta}_2$ — the vector connecting a B atom to an A atom inside the same unit cell — points in the armchair (or $\hat{y}$) direction. It does not, therefore, influence in any way the response along the zig-zag direction. For the armchair direction, however, the results that follow from these two representations are indeed different, Figure 4.3, as this additional phase factor comes into play. Moreover, if we consider the responses in the lattice-vectors representation along the two relevant directions, Figure 4.3(a), we can see that they are *not* the same, meaning that the definition of the position operator introduced in Eq.(4.45) cannot adequately translate the symmetry properties of the PG monolayer, Table 2.1, especially at high frequencies [12]. It cannot, therefore, be sensibly used in a description of the nonlinear optical response of this type of material. (Note how we are effectively using the symmetry properties of the tensor at hand in order to cross-check our calculations.)

As for the nearest-neighbours representation, we note that that property is indeed retained. It is this representation, along with the definition of the position operator introduced in Eq.(4.31), that we used to study the nonlinear optical response of the PG and GG monolayers [5].

4.4 Results

Now that we have clarified some aspects involved in the numerical calculation of nonlinear optical responses in the velocity gauge, we can move on to the results presented in ref.[5]: the second harmonic generation and optical rectification conductivities of the GG monolayer; the third harmonic generation and Kerr effect conductivities of the PG and GG monolayers. We note that we considered different values for the gap and chemical potential, Δ and μ, as well as different values for the relaxation rate, γ, in these calculations; we did not consider, however, different values for the temperature, T, as its effect is similar to that of γ, which is to broaden the features. T is thus henceforth set to 1K. In addition, considering that all nonlinear optical conductivities are monotonically decreasing — the exception being the regions around processes at the gap (or twice the chemical potential) and around one of the van Hove singularities — these were represented in these two frequency regions separately, so as to make their features more visible. Finally, we note that we imposed intrinsic permutation symmetry on all following conductivities.

[10] The exception to this is, of course, the case where there is a single atom in the unit cell.
[11] The conductivities with superscripts are henceforth used interchangeably with those with subscripts.

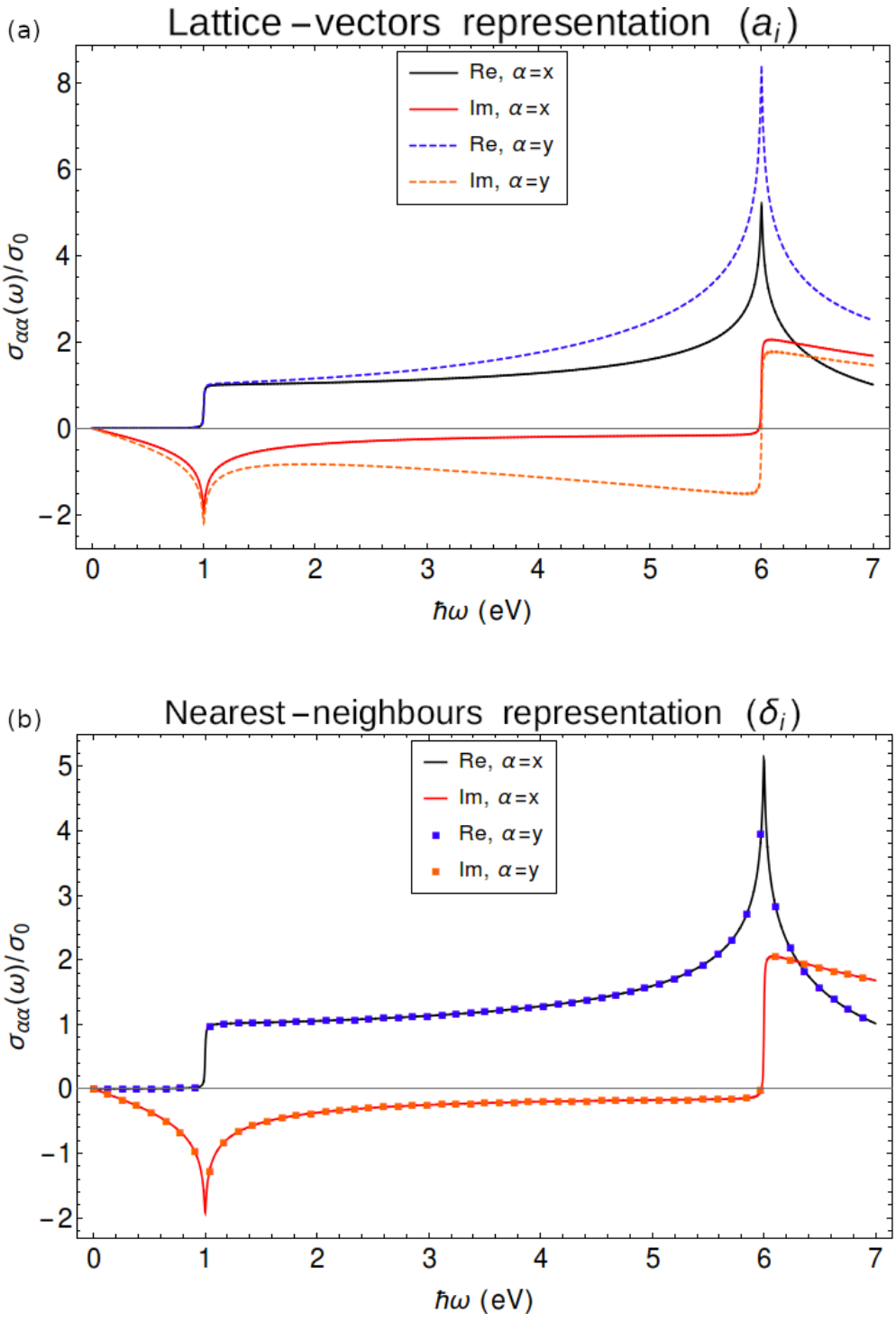


Figure 4.3: The real and imaginary parts of the interband portion of the linear conductivities, $\sigma_{xx}(\omega)$ and $\sigma_{yy}(\omega)$, for the tight-binding Hamiltonians written in the lattice-vectors (a_i) — plot (a) — and nearest-neighbours representation (δ_i) — plot (b). The chemical potential is again fixed to 0.5 eV. Plots from ref.[5].

4.4.1 Second Order Response of the Gapped Graphene Monolayer

The tight-binding Hamiltonian used to describe gapped graphene is obtained from its counterpart for plain graphene, which we introduced in Eq.(4.39), by adding to it a term that breaks the equivalence between the A and B atoms, $\mathrm{diag}(\Delta/2, -\Delta/2)$. This, of course, breaks the centrosymmetricity of the plain graphene monolayer and thus enables the generation of optical responses in second order, Fig.4.1. As we have seen in the section on the point group symmetries of crystals, there is a single independent element in this conductivity tensor, σ_{yyy}; there, we also noted that the relation between this component and the remaining nontrivial components reads as,

$$\sigma_{xxy} = \sigma_{xyx} = \sigma_{yxx} = -\sigma_{yyy}. \tag{4.46}$$

The following results were normalized with respect to $\sigma_2 = e^3 a_0/4t\hbar = 2.87 \times 10^{-15}$ S·m/V [161].

Second Harmonic Generation (SHG)

First, consider the one-photon ($\hbar\omega \sim \Delta$) and two-photon ($2\hbar\omega \sim \Delta$) processes at the gap for values of $\Delta = 30, 300$ meV [166], and two different values of the relaxation rate, $\gamma = 0.005, 0.001$ eV. These are represented in Figure 4.4. As expected, the shape of the features depends on the relative size of the gap and the relaxation parameter, Δ and γ. For larger relaxation rates and smaller gaps, the one-photon and two-photon peaks tend to overlap, Figure 4.4 (a); for larger gaps, however, the two peaks are clearly distinct, Figure 4.4 (b).

The results at finite γ, which were computed numerically in the velocity gauge, are supplemented by $\gamma \to 0$ results, which were computed analytically in the length gauge using an effective Hamiltonian. These are represented by thicker green curves, and follow from the expression for the second order conductivity derived in Section A.1, Eq.(A.8). The real part of the two-photon process in the second harmonic generation can be written in terms of the shift current coefficient derived by Aversa and Sipe in ref.[22, 23],[12]

$$\frac{1}{\sigma_2}\mathrm{Re}\left[\sigma_{yyy}(\omega,\omega)\right] = -\frac{8it}{\pi}\int d^2\mathbf{k}\,\xi^y_{\mathbf{k}vc}\left(\xi^y_{\mathbf{k}cv}\right)_{;y}\delta(2\hbar\omega - \Delta\epsilon_{\mathbf{k}cv}), \tag{4.47}$$

which also describes the other second order effect (the optical rectification) we consider here. This expression is valid for frequencies below the threshold of one-photon processes (1PP), i.e., $\hbar\omega \leq \Delta$. For frequencies above the 1PP threshold, the conductivity gets a second contribution, which reads

[12] D. J. Passos has shown that the real part of the second order conductivity of a time-reversal symmetric cold semiconductor — such as gapped graphene — can be reduced to a linear combination of integrals associated with the shift and injection currents [10]. As there is no injection current in graphene, Eq.(4.46), the real part of the second order response is described by a single type of integral, the shift current coefficient, Eq.(4.47). Furthermore, we can show that this relation between the diagonal components of the second harmonic generation, $\sigma^{\beta\beta\beta}(\omega,\omega)$, and those of the optical rectification, $\sigma^{\beta\beta\beta}(\omega,-\omega)$, is connected with a more general relation between the real parts of three different second order conductivities. This latter point will be the subject of Section 5.5.

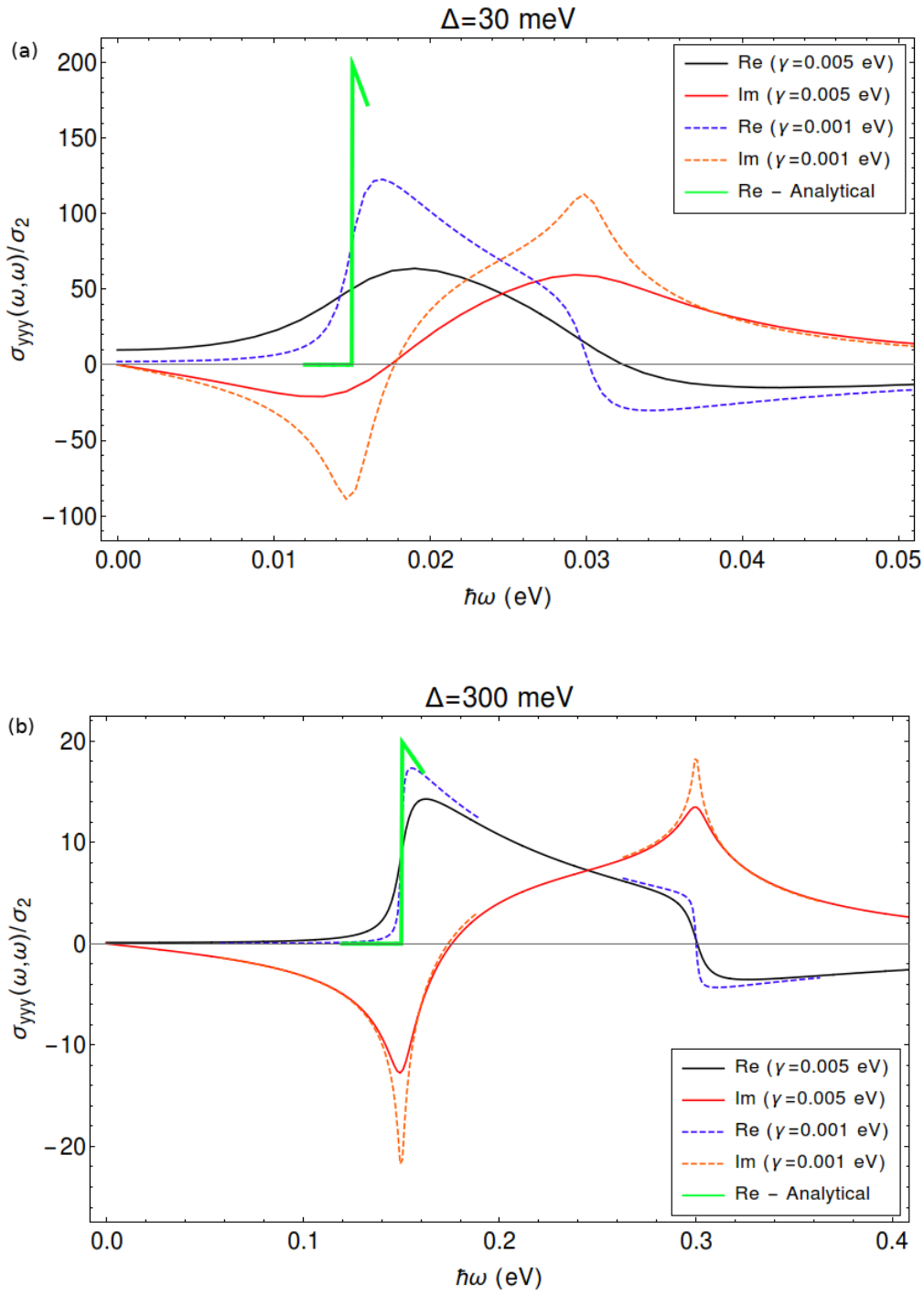


Figure 4.4: The real and imaginary parts of the second harmonic generation, $\sigma_{yyy}(\omega,\omega)$, close to the one-photon and two-photon processes at the gap: 30 meV in the top plot, (a), and $\Delta = 300$ meV in the bottom plot, (b), for different values of the relaxation rate. The green curve represents the analytical ($\gamma \to 0$) result that follows from Eq.(4.47). The chemical potential and the temperature were set to $\mu = 0$ eV, $T = 1$K. Plots from ref.[5].

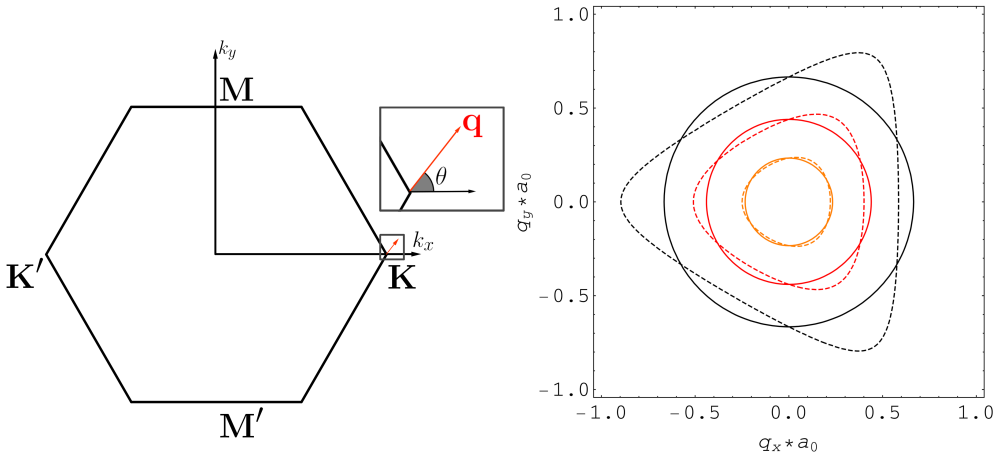


Figure 4.5: [Left] The first Brillouin zone associated with the honeycomb lattice. The labelled edges of the FBZ are the **K** and **K**′ points (the other four unlabelled vertexes are their equivalents), **K**, **K**′ $= (\pm 4\pi/3\sqrt{3}a_0, 0)$, which correspond to band minima, Eq.(4.49); the labelled midpoints of the FBZ's sides are the **M** and **M**′ points (the other four unlabelled midpoints are their equivalents), **M**, **M**′ $= (0, \pm 2\pi/3a_0)$, which correspond to a van Hove singularity, Eq.(4.56). The inset represents the small region around **K** where the expansion in **q** of Eq.(4.50) is performed. [Right] Three sets of curves of constant energy ($\Delta\epsilon_{\mathbf{q}cv}/t =$ cons.) for a hopping function, $|\phi_\delta(\mathbf{k} = \mathbf{K} + \mathbf{q})|$, that is either linear (full curves), or quadratic (dashed curves), in $|\mathbf{q}|$. The origin of this plot is the **K** point on the picture on the left.

exactly like the one in Eq.(4.47), but with $\hbar\omega$ in the argument of the Dirac delta function (and some additional numerical factors) [10]. Furthermore, we have introduced the standard notation of the generalized derivative [22], so that the generalized derivative of the Berry connection reads as,

$$\left(\xi^{\alpha_2}_{\mathbf{k}ss'}\right)_{;\alpha_1} = (\nabla^{\alpha_1}_{\mathbf{k}}\xi^{\alpha_2}_{\mathbf{k}ss'}) - i(\xi^{\alpha_1}_{\mathbf{k}ss} - \xi^{\alpha_1}_{\mathbf{k}s's'})\xi^{\alpha_2}_{ss'}. \tag{4.48}$$

This is equal to the commutator of the diagonal (in band-index space) part of the covariant derivative, Eq.(3.53), $D^{\alpha_1}_{ss',\text{intra.}} = \delta_{ss'}\left(\nabla^{\alpha_1}_{\mathbf{k}} - i\xi^{\alpha_1}_{\mathbf{k}ss'}\right)$, with the Berry connection. Note that this object is gauge covariant.

To perform this calculation, we must first note that the Dirac delta function in the integrand of Eq.(4.47) restricts the relevant contributions to that integral to a equi-energetic surface in the Brillouin zone. For two-photon processes at the gap, the relevant contributions come from the two regions of the FBZ centred around the band minima, the Dirac points in the PG monolayer, **K**, **K**′ $= (\pm 4\pi/3\sqrt{3}a_0, 0)$, Figure 4.5. Furthermore, since the Dirac delta function fixes $\Delta\epsilon_{\mathbf{k}cv}$ directly to twice the photon energy, it is a suitable integration variable, provided, of course, that we can change integration variables (from the momentum **k** to this energy difference). The relation between $\Delta\epsilon_{\mathbf{k}cv}$ and **k** reads as,

$$\Delta\epsilon^2_{\mathbf{k}cv} = \Delta^2 + 4t^2\left|\phi_\delta(\mathbf{k})\right|^2. \tag{4.49}$$

This follows from the eigenvalue problem for the tight-binding Hamiltonian for the GG monolayer, i.e., Eq.(4.39) with an extra diag$(\Delta/2, -\Delta/2)$ term. Expanding the hopping function, ϕ_δ, for small momenta, $\mathbf{q}$, around one of the band minima, we obtain an effective descripition of the band-structure. The absolute value of the hopping function then reads as,

$$|\phi_\delta(\mathbf{k} = \mathbf{K} + \mathbf{q})| = \frac{3a_0\,|\mathbf{q}|}{2} - \frac{3a_0^2\,|\mathbf{q}|^2}{8}\cos(3\theta) + \mathcal{O}(|\mathbf{q}|^3), \tag{4.50}$$

for $|\mathbf{q}|$ and θ, the radial and polar coordinates associated with $\mathbf{q}$. Rewriting Eq.(4.49) with the help of Eq.(4.50), we obtain,[13]

$$\frac{1}{t}\sqrt{\Delta\epsilon_{\mathbf{q}cv}^2 - \Delta^2} = \frac{3a_0\,|\mathbf{q}|}{2} - \frac{3a_0^2\,|\mathbf{q}|^2}{8}\cos(3\theta) + \mathcal{O}(|\mathbf{q}|^3). \tag{4.51}$$

This relates one of our our integration variables, $|\mathbf{q}|$, with a new variable, $\lambda(\Delta\epsilon_{\mathbf{q}cv})$,

$$\lambda = \frac{1}{t}\sqrt{\Delta\epsilon_{\mathbf{q}cv}^2 - \Delta^2}, \tag{4.52}$$

that can be deemed small, provided that we consider energy differences close to the band gap. Inverting the series in Eq.(4.51), we can write the modulus of the momentum $|\mathbf{q}|$ in terms of λ,

$$|\mathbf{q}| = \frac{\lambda}{3a_0} + \frac{\lambda^2}{36a_0^2}\cos(3\theta) + \mathcal{O}(\lambda^3), \tag{4.53}$$

thus mapping the original integration variables, $\mathbf{k}$ (or $\mathbf{q}$), into λ and θ. With this map between variables, we can now express the integrand in a form that can be easily integrated over, as the integral over λ is trivial,

$$\delta(2\hbar\omega - \Delta\epsilon_{\mathbf{q}cv}) = \frac{2\hbar\omega}{t\sqrt{(2\hbar\omega)^2 - \Delta^2}}\delta(\lambda - t^{-1}\sqrt{(2\hbar\omega)^2 - \Delta^2}), \tag{4.54}$$

while the integral over θ is made simple by the fact that it involves only simple trigonometric functions. The end result, after some careful manipulations, is an expansion in powers of $\sqrt{(2\hbar\omega)^2 - \Delta^2}$, and reads for the first few terms as,

$$\frac{1}{\sigma_2}\mathrm{Re}\left[\sigma_{yyy}(\omega,\omega)\right] = \Theta(2\hbar\omega - \Delta)\left[\frac{2t}{\Delta} + \left(\frac{t}{9\Delta} - \frac{2t^3}{\Delta^3}\right)\frac{1}{t^2}((2\hbar\omega)^2 - \Delta^2) + (...)\right]. \tag{4.55}$$

As we have mentioned, this analytical result is represented in Figure 4.4 by the thicker green curves. Finally, and before we move on to the high frequency response, we must add a few comments on this calculation. First, note that, while an expansion in $\mathbf{q}$ is justified by the Dirac delta function

[13] The Berry connection, and its generalized derivative, can be easily determined by means of Eq.(4.37). They can then be written in terms of λ and θ, using the map between $\mathbf{k}$ (or $\mathbf{q}$) and λ and θ mentioned below.

in the integrand, it does not, by itself, guarantee that the end result will be correct: if we had carried only the linear term in $|\mathbf{q}|$ in Eq.(4.51), the end result, $\mathrm{Re}\left[\sigma_{yyy}(\omega,\omega)\right]$, would have been *exactly* zero. This happens for the same reason that the second order response in the monolayer of plain graphene, when it is described by the Dirac Hamiltonian, is trivial; the Berry connections associated with that Hamiltonian, $\xi^{\alpha}_{\mathbf{q}ss'}$, are odd under $\mathbf{q}\to-\mathbf{q}$, meaning that the integrand (which involves either three Berry connections or two Berry connections and a derivative with respect to $\mathbf{q}$, Eq.(4.48)) is odd, and thus the integral necessarily vanishes.[14] Secondly, the introduction of higher order terms in Eq.(4.51) changes the shape of the band, which is no longer isotropic — check the plot on the right in Figure (4.5) — so that the integral (and thus the SHG) is no longer zero. However, this also brings about an added difficulty. As we have seen in this derivation, the exchange of the momentum variable, $|\mathbf{q}|$, by the difference of band energies, $\Delta\epsilon_{\mathbf{q}cv}$, is not something that can be trivially done by hand, and instead requires a series inversion. Furthermore, the integrand, which we have purposefully omitted, is now a complicated function of λ and θ that requires a significant amount of manipulation.

The high frequency results, i.e., those for the two- ($\hbar\omega\sim t$) and one- ($\hbar\omega\sim 2t$) photon processes at one of the van Hove singularities, are represented in Figure 4.6. We can see that the features — for different values of the gap, Δ — are centred around slightly different energies, as,

$$\Delta\epsilon^{2}_{\mathrm{vHs}}=\Delta^{2}+4t^{2}\left|\phi_{\delta}(\mathbf{M},\mathbf{M}')\right|^{2}, \tag{4.56}$$

for $\mathbf{M}$, $\mathbf{M}'=(0,\pm 2\pi/3a_0)$. Note also that the absolute value of these conductivities scales with Δ — the opposite behavior to what we found for the response at the gap. Another, quite surprising, point concerns the features of the real and imaginary parts of these conductivities; they are switched with respect to the real and imaginary parts of the conductivities at the gap, Figure 4.4. It is now the real part that has the shape of a logarithmic-like divergence while the step-like behavior is present in the imaginary part.

Optical Rectification

The other second order process that can be observed in the response to a monochromatic field is the generation of a DC current, described by the optical rectification conductivity: $\sigma_{yyy}(\omega,-\omega)$, Figures 4.7 and 4.8. Figure 4.7 represents, for two values of Δ, the one-photon process at the gap. We can see that this tensor component is always finite (even in the relaxation-free limit), meaning that there is indeed no injection current, as prescribed by the symmetry properties of the GG monolayer, Eq.(4.46). The remaining portion of this response is associated with the shift current, whose features are similar to those of the second harmonic generation at the gap, Figure 4.4(b). There are, however, two noteworthy differences: there are no two-photon processes, so the response

[14] This linearized Hamiltonian, which is usually called the gapped Dirac Hamiltonian, is centrosymmetric. Its second order response must, therefore, be exactly zero.

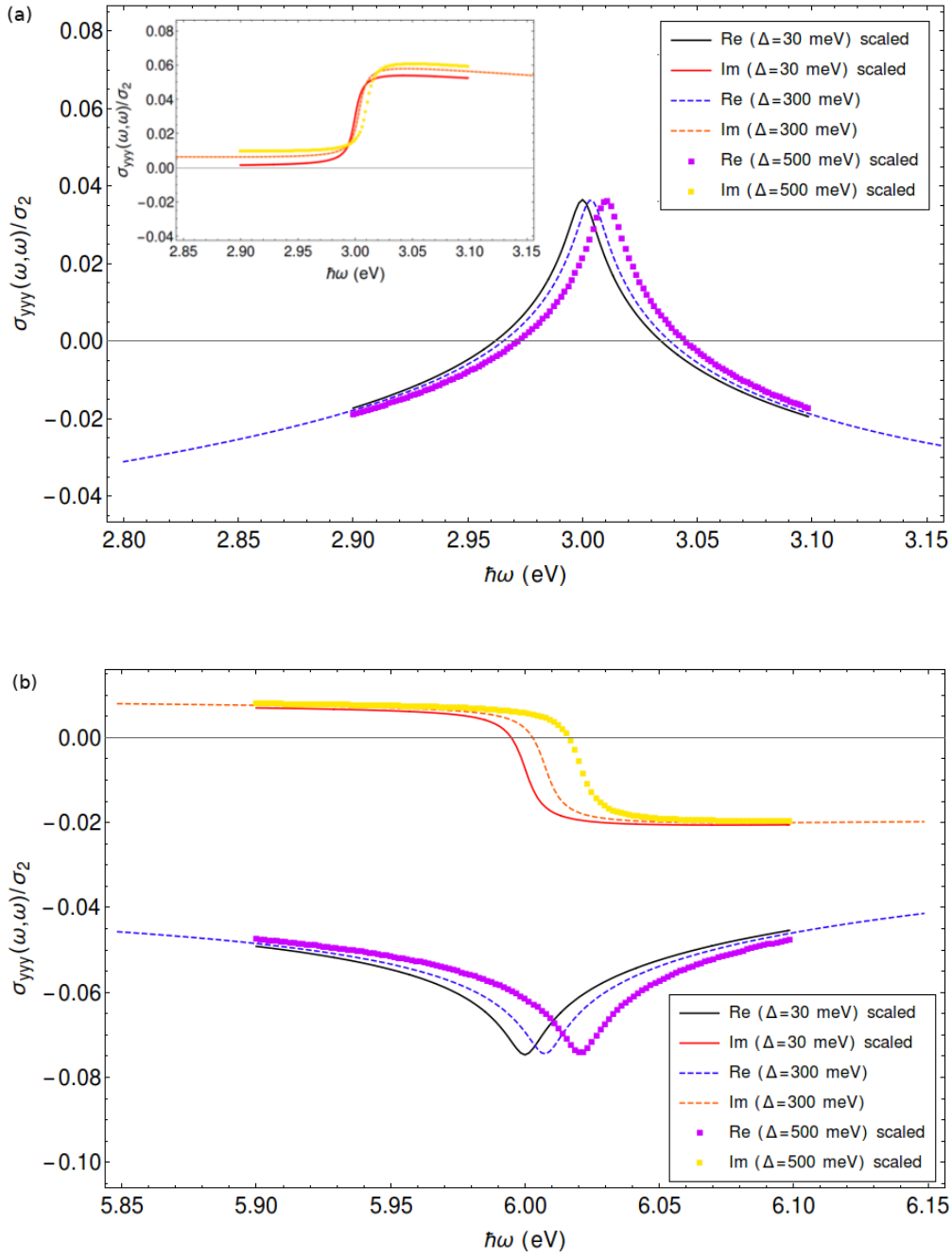


Figure 4.6: The real and imaginary parts of the SHG in GG, $\sigma_{yyy}(\omega,\omega)$, for frequencies around the two- ($\hbar\omega \sim t$) (a) and one- ($\hbar\omega \sim 2t$) (b) photon processes at one of the van Hove singularities, Eq.(4.56). The imaginary parts for the two photon processes are represented in the inset. Curves labelled as scaled have been divided by a factor of $\Delta/300$ meV. The relaxation parameter for these plots is $\gamma = 0.005$ eV. The chemical potential and the temperature were set to $\mu = 0$ eV, $T = 1$K. Plots from ref.[5].

has a single set of features at low frequencies, which are associated with the aforementioned one-photon processes at the gap, not two; the imaginary part of this conductivity is *exactly* zero.[15] As we have noted, the relaxation-free limit of this conductivity gives us a single contribution, the shift current [23],

$$\frac{1}{\sigma_2}\sigma_{yyy}(\omega,-\omega)=-\frac{4it}{\pi}\int d^2\mathbf{k}\,\xi^{y}_{\mathbf{k}vc}\left(\xi^{y}_{\mathbf{k}cv}\right)_{;y}\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}cv}), \tag{4.57}$$

which is, apart from a factor of two (and a different argument in the Dirac delta function), the exact same expression as that for the two-photon process in the second harmonic generation, Eq.(4.55). We can thus write,

$$\frac{1}{\sigma_2}\mathrm{Re}\left[\sigma_{yyy}(\omega,-\omega)\right]=\Theta(\hbar\omega-\Delta)\Big[\frac{t}{\Delta}+\Big(\frac{t}{18\Delta}-\frac{t^3}{\Delta^3}\Big)\frac{1}{t^2}((\hbar\omega)^2-\Delta^2)+(...)\Big]. \tag{4.58}$$

This similarity between the optical rectification conductivity and the second harmonic generation is, naturally, also present at higher frequencies, $\hbar\omega\sim 2t$, Figure 4.8. Apart from a sign switch and the absence of an imaginary part — as we have noted, the diagonal components of an intrinsically symmetrized optical rectification conductivity are necessarily real — this result is very similar to that of Figure 4.6(b).

4.4.2 Third Order Response of the Plain Graphene and Gapped Graphene Monolayers

We now consider the third order response to a monochromatic field of plain and gapped graphene monolayers: the third harmonic generation and the optical Kerr effect. As we have noted, these effects are finite even in the presence of inversion symmetry [59, 68, 69, 73]. We stated in Section 2.2 that, although associated with different point group symmetries, the components of their third order conductivities satisfy the same relations,

$$\begin{gathered}\sigma_{yyyy}=\sigma_{xxxx}=\sigma_{xxyy}+\sigma_{xyxy}+\sigma_{xyyx},\\ \sigma_{xxyy}=\sigma_{yyxx},\ \sigma_{xyxy}=\sigma_{yxyx},\ \sigma_{xyyx}=\sigma_{yxxy}.\end{gathered} \tag{4.59}$$

[15] This follows from the fact that the physical conductivity must obey the reality condition, Eq.(2.27), and be intrinsically symmetric, Eq.(2.30). Its only relevant element, $\sigma_{yyy}(\omega,-\omega)$, must therefore simultaneously fulfill both,

$$\sigma_{yyy}(\omega,-\omega)=\left[\sigma_{yyy}(-\omega,\omega)\right]^*,$$

and

$$\sigma_{yyy}(\omega,-\omega)=\sigma_{yyy}(-\omega,\omega),$$

so it must be real.

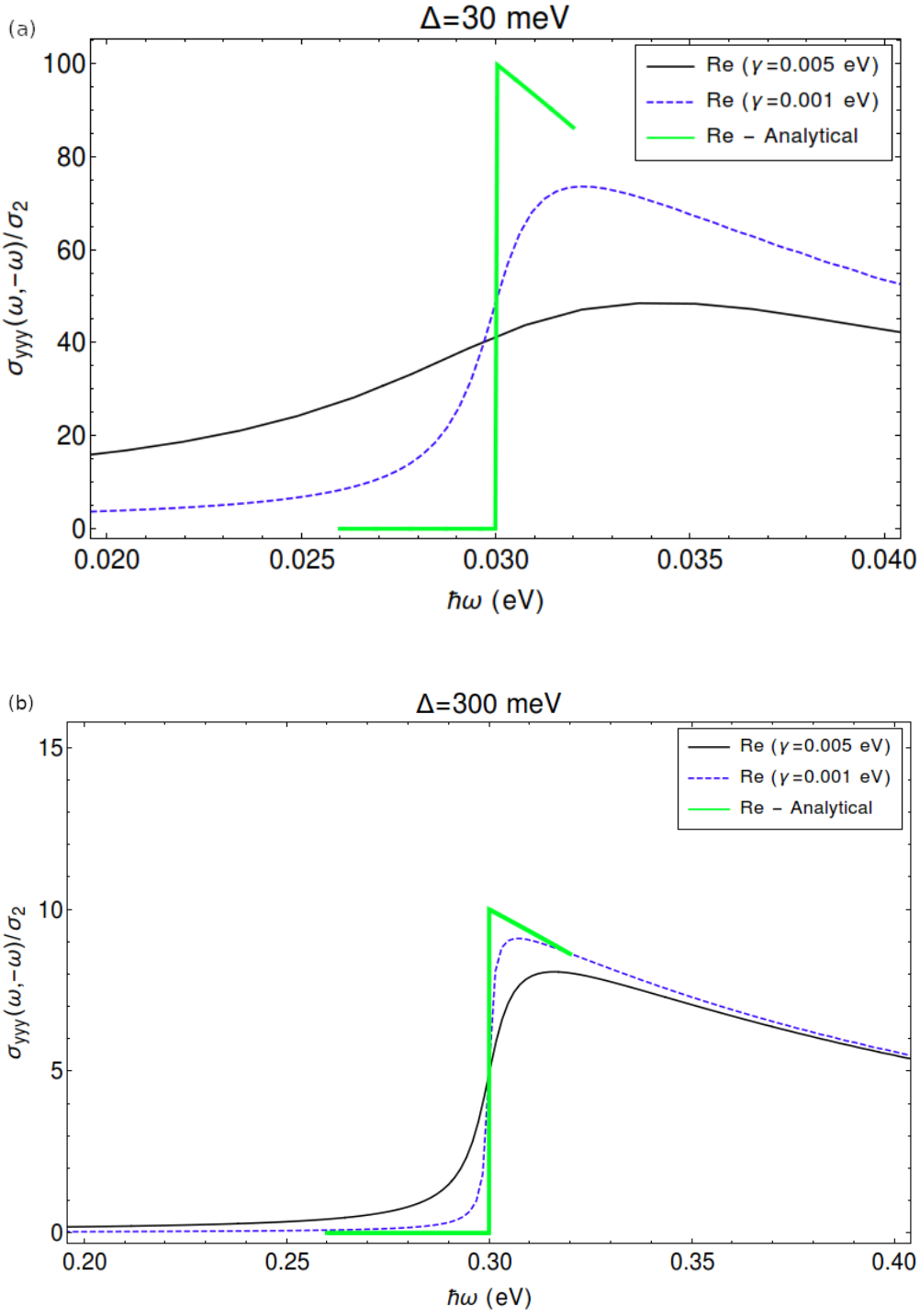


Figure 4.7: The optical rectification conductivity, $\sigma_{yyy}(\omega, -\omega)$, in GG for frequencies close to the gap: $\Delta = 30$ meV in the top plot (a) and $\Delta = 300$ meV in the bottom plot (b) for different values of the relaxation rate. The green curve represents the analytical ($\gamma \to 0$) result that follows from Eq.(4.58). Plots from ref.[5].

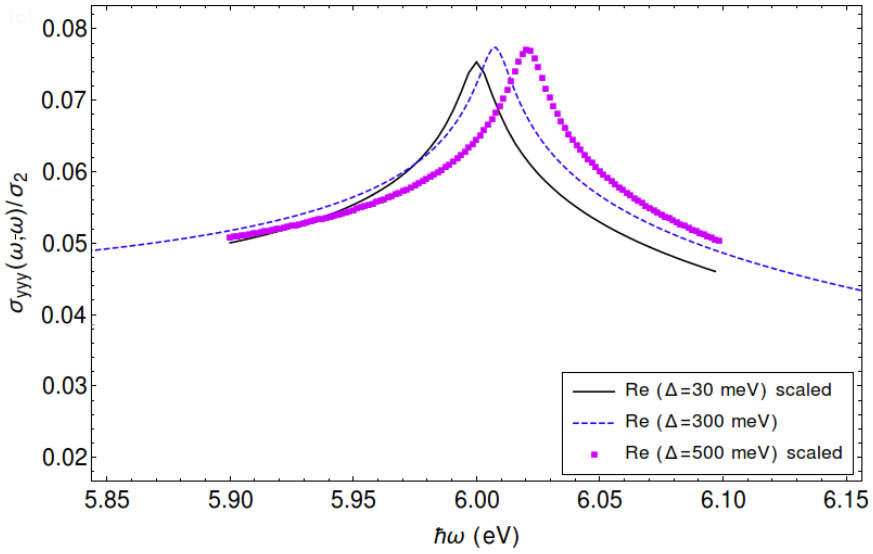


Figure 4.8: The optical rectification conductivity, $\sigma_{yyy}(\omega,-\omega)$, in GG for frequencies around the one photon process at one of the van Hove singularities, Eq.(4.56). Curves labelled as scaled have been divided by a factor of $\Delta/300$ meV. The relaxation parameter in this plot is $\gamma = 0.005$ eV. Plot from ref.[5].

Since we are also imposing intrinsic permutation symmetry, in the case of the third harmonic generation (THG) all components of the tensor can be expressed in terms of a single one,

$$\begin{aligned}\sigma_{xxyy}(\omega,\omega,\omega) &= \sigma_{xyxy}(\omega,\omega,\omega) = \sigma_{xyyx}(\omega,\omega,\omega),\\ \sigma_{xxyy}(\omega,\omega,\omega) &= \tfrac{1}{3}\sigma_{xxxx}(\omega,\omega,\omega) = \tfrac{1}{3}\sigma_{yyyy}(\omega,\omega,\omega).\end{aligned} \tag{4.60}$$

Therefore, in our study of the THG, we present only the σ_{yyyy} component. For the optical Kerr effect, we consider both the σ_{yyyy} and σ_{yxxy} components. The following results were normalized with respect to $\sigma_3 = e^4 a_0^2/8\hbar t^2 = 6.84\times10^{-26}$ S·m^2/V^2.

Third Harmonic Generation (THG)

In the previous section, we discussed how the relative size of Δ and γ affects the features of optical conductivities. For smaller gaps and larger relaxation rates, one sees a broadening, possibly even a merging, of the features associated with different processes at the gap. There is, therefore, not much to be gained from repeating this discussion in our study of the third order response; the relaxation rate is henceforth fixed to $\gamma = 0.005$ eV. We also consider that the energy associated with the excluded, Pauli-blocked, region of the bands of the plain graphene monolayer, 2μ, is equal to the band gap of the gapped graphene monolayer, Δ. We set these values to 300 meV, Figures 4.9 and 4.10.

We begin by studying the response of the three types of photon processes, $n\hbar\omega = \Delta$, 2μ for $n =$

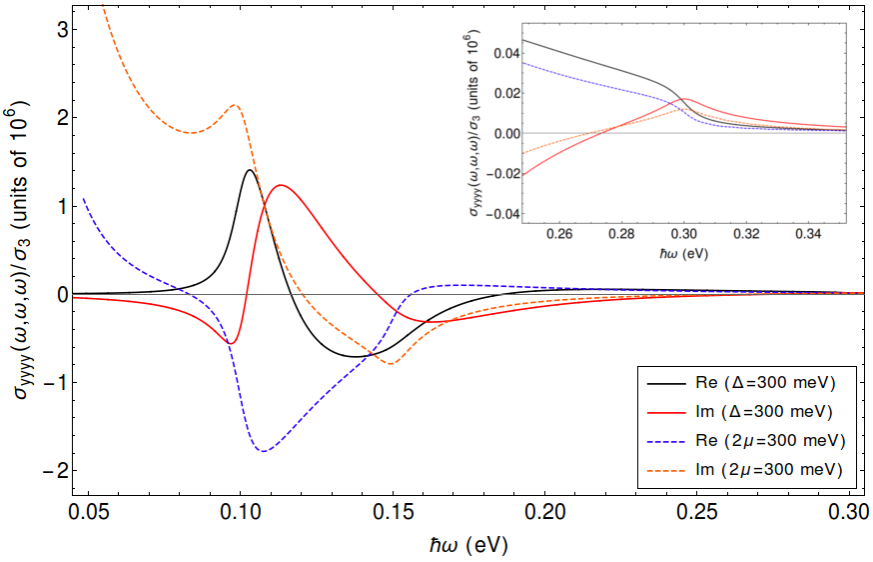


Figure 4.9: The real and imaginary parts of the third harmonic generation, $\sigma_{yyyy}(\omega,\omega,\omega)$, in the gapped graphene, $\Delta = 300$ meV, and plain graphene, $2\mu = 300$ meV, monolayers. The frequency range covers the different — one-, two-, and three-photon — processes at the gap / twice the value of the chemical potential. Note that the vertical scale is in units of 10^6. The inset represents a zoom-in on the region of the one photon process, $\hbar\omega \sim \Delta$, 2μ. Plot from ref.[5].

1, 2, 3, Figure 4.9. As expected, the conductivities of the plain and gapped graphene monolayers are very different: for frequencies below the three-photon resonance at the gap, we see that the response for the plain graphene monolayer diverges — due to the existence of Drude-type contributions [10] — whereas that of the gapped graphene monolayer goes to zero — due to the absence of said contributions, as should be the case in the response of a cold semiconductor; for the three-photon resonance itself, there are prominent features in both sets of curves, which have opposite signs; for the two-photon resonance, there are no clear features in the GG monolayer, whereas in the monolayer of PG, one finds a shoulder and a local minimum in both the real and imaginary parts; finally, for the one-photon resonance at the gap, represented in the inset to Figure 4.9, the responses are similar.

For higher frequencies, i.e., those associated with the different processes around one of the van Hove singularities, Figure 4.10, the conductivities of the PG and GG monolayers are again similar. For these energies, the band structures are almost identical (as $\Delta \ll t$), and the chemical potential of the PG monolayer is irrelevant. We see that the only distinction between the responses of the two monolayers comes from the fact that the van Hove singularity at the $\mathbf{M}$ and $\mathbf{M}'$ points is placed at different energies, being slightly higher in the GG monolayer than in the PG one, Eq.(4.56).

Optical Kerr Effect (OKE)

Finally, we turn our attention to the optical Kerr effect (OKE). As mentioned before, and unlike the third harmonic generation, not all nonzero components of the conductivity tensor for the OKE can be related to one of its diagonal components. To show the differences between components, we present both $\sigma_{yyyy}(\omega, \omega, -\omega)$ and $\sigma_{yxxy}(\omega, \omega, -\omega)$ in the low-frequency range of the response, i.e., around the one- and two photon- processes at the gap (or twice the chemical potential for PG), Figures 4.11(a) and 4.11(b), as well as the response around the one-photon process at one of the van Hove singularities, Figure 4.12.

Figure 4.11(a) shows that the two conductivity components of the GG monolayer have opposite signs — in both the real and the imaginary parts — and that they are similar, but not exactly equal, in modulus. In the PG monolayer, it is only the height of the features that is different, being less pronounced in the off-diagonal component. For the two-photon processes at the gap (twice the chemical potential), Figure 4.12(b), the GG and PG conductivities show differences between the two tensor components, and the property we observed in the one-photon response seems to appear in reverse: here it is the response of the GG monolayer that shows less pronounced features in its off-diagonal component; as for the conductivities of the PG monolayer, they now have opposite signs and are rather similar, in modulus, across the frequency range considered.

For the high frequency response, i.e., for one-photon processes at one of the van Hove singularities, Figure 4.12, we see that the features of both components are essentially the same, and that these differ only by an overall factor of three. As in the case of the third harmonic generation, the only distinction between the responses of the two monolayers comes from the fact that the van Hove singularity at the $\mathbf{M}$ and $\mathbf{M}'$ points is placed at different energies, being slightly higher in the GG monolayer than in the PG one, Eq.(4.56).

A second point of interest in the OKE concerns the existence of a *divergence* in the real part of the conductivity for frequencies above the one-photon process at the gap (twice the chemical potential) in the relaxation-free limit [22].[16] As expected, this divergence is present in both the gapped graphene and plain graphene monolayers, and has been previously seen in an analytical calculation of the OKE in plain graphene [68]. While we cannot probe this singularity directly — in the sense that the relaxation parameter is necessarily finite in numerical calculations — it is clear that such a divergence exists in both materials. In Figure 4.13, we represent the real and imaginary parts of the OKE conductivity for frequencies around the two-photon (a) and one-photon (b) processes at the gap (twice the chemical potential) for different values of the relaxation parameter, γ. For the former, Figure 4.13(a), we can see that smaller values of γ are associated with sharper features in a small region around the absorption threshold, which then merge as one moves away from the energy threshold, $2\hbar\omega \sim \Delta$, 2μ. This is similar to what we observed in Figures

[16] The divergence is due to the acceleration of electron-hole pairs — produced in one-photon *absorption* processes — by a static, nonlinear, electric field.

4.4(b) and 4.7(b), and is indeed the expected behavior for features in a *regular* conductivity. For the one-photon process at the gap, this no longer holds true (for the real part of the OKE conductivity). This real part increases in absolute value as γ is reduced, while the curves for different values of γ run parallel to one another. Instead of a well-localized feature, we see the existence of a *divergence*.

4.5 Closing Remarks

This chapter has been devoted to the study of the second and third order responses of the plain graphene and gapped graphene monolayers, which we originally presented in ref.[5], and to certain aspects that follow, quite naturally, from the calculation of these conductivities. While it is, primarily, an application of a velocity-gauge formalism derived elsewhere [3], we have seen that even this is not without its subtleties, and brings to the fore certain important aspects of its own. Furthermore, it is important to note that, contrary to what had been reported in the literature, the high-frequency response of these materials can be studied within the standard framework of calculations of nonlinear optical responses in two-dimensional materials and does not require the inclusion of any additional symmetries. The results we have derived are in agreement with what little intuition we have on this subject. Finally, we note that this chapter (and its corresponding work) includes our first treatment, even if rather tentatively, of a divergent contribution to an NLO response. The study of these divergences, specifically, the jerk current in third order DC response, $\sigma(\omega, -\omega, 0)$, is one of the subjects of the following chapter.

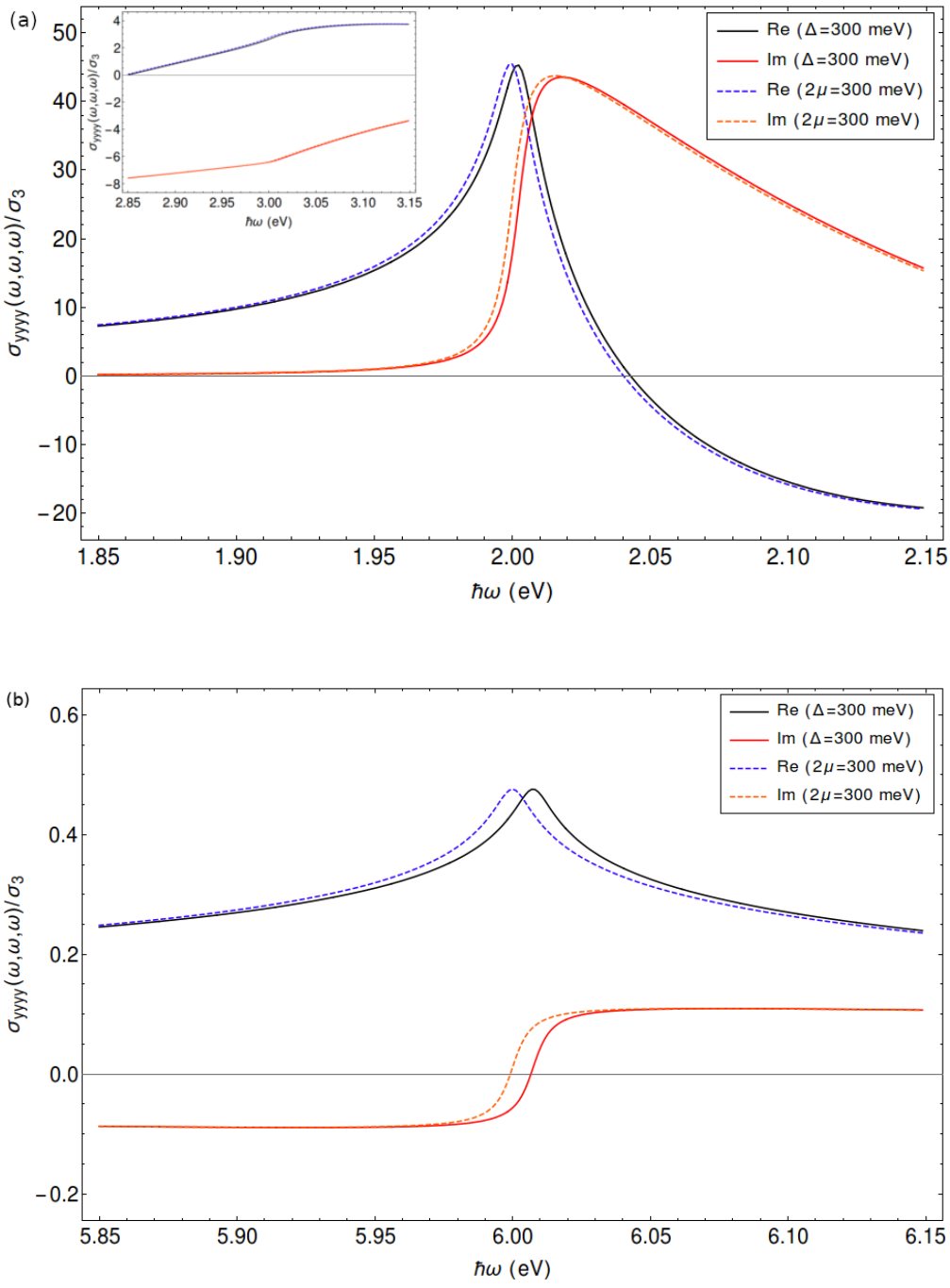


Figure 4.10: The real and imaginary parts of the third harmonic generation, $\sigma_{yyyy}(\omega,\omega,\omega)$, in the gapped graphene, $\Delta = 300$ meV, and plain graphene, $2\mu = 300$ meV, monolayers, for frequencies around the three-photon ($3\hbar\omega \sim 2t$), (a), two-photon ($\hbar\omega \sim t$), inset of (a), and one-photon ($\hbar\omega \sim 2t$), (b), processes at one of their van Hove singularities. Plots from ref.[5].

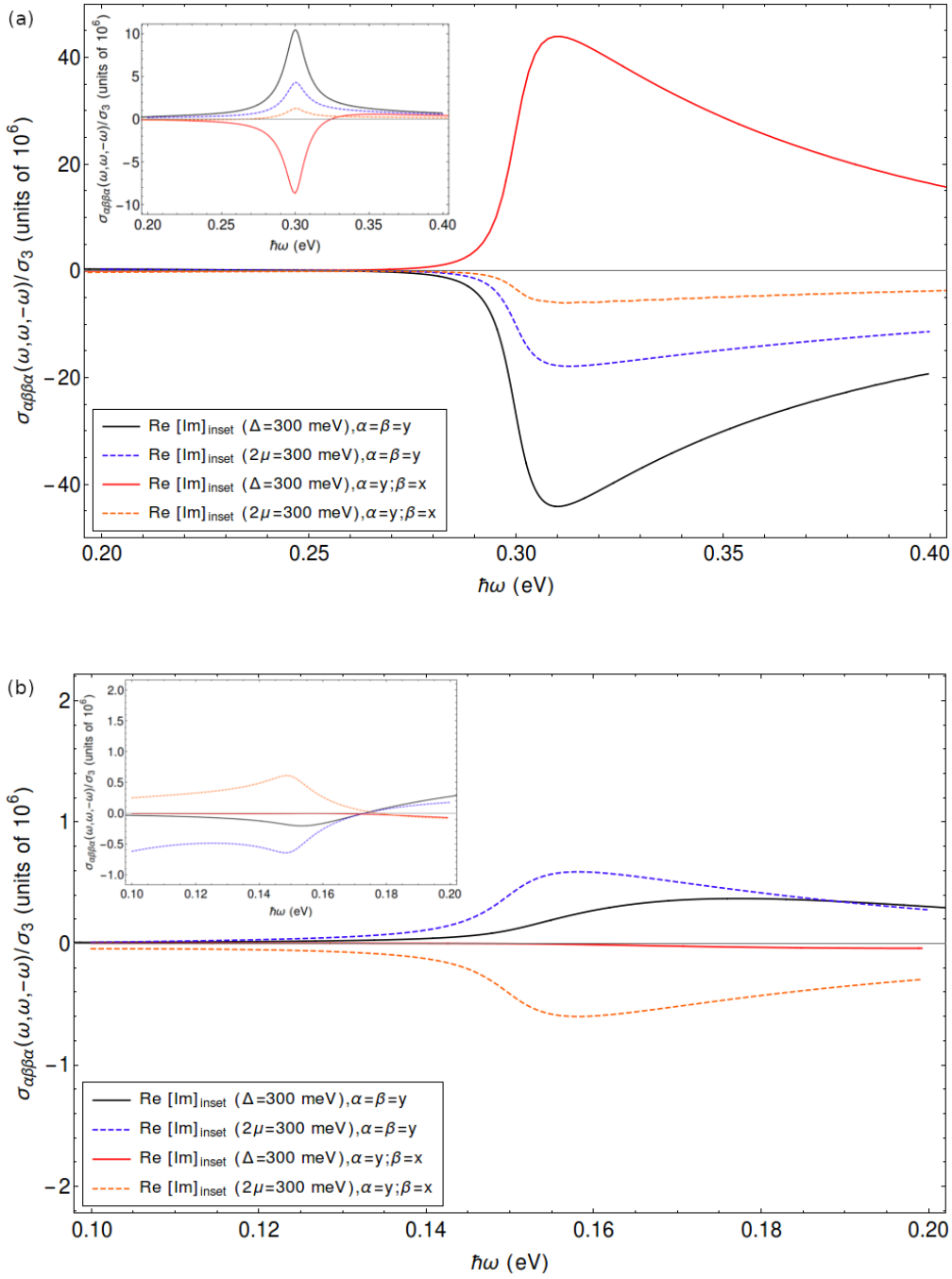


Figure 4.11: The real and imaginary parts (the latter are represented in the insets) of the optical Kerr effect, for the tensor components $\sigma_{yyyy}(\omega,\omega,-\omega)$ and $\sigma_{yxxy}(\omega,\omega,-\omega)$, in the gapped graphene, $\Delta = 300$ meV, and plain graphene, $2\mu = 300$ meV, monolayers. Plot (a) represents the one-photon process at the gap / twice the chemical potential, while plot (b) represents the two-photon process. Note that the vertical scale in both figures is in units of 10^6. Plots from ref.[5].

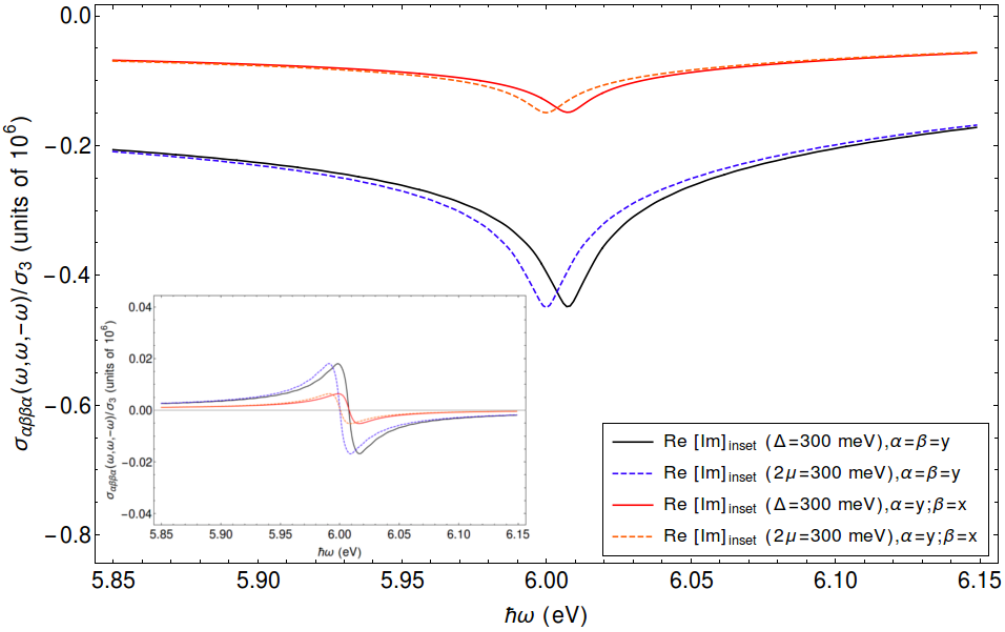


Figure 4.12: The real and imaginary part (the latter is represented in the inset) of the optical Kerr effect, for the tensor components $\sigma_{yyyy}(\omega,\omega,-\omega)$ and $\sigma_{yxxy}(\omega,\omega,-\omega)$, in the gapped graphene, $\Delta = 300$ meV, and plain graphene, $2\mu = 300$ meV, monolayers, for frequencies around one of the van Hove singularities, Eq.(4.56). Note that the vertical scale is in units of 10^6. Plot from ref.[5].

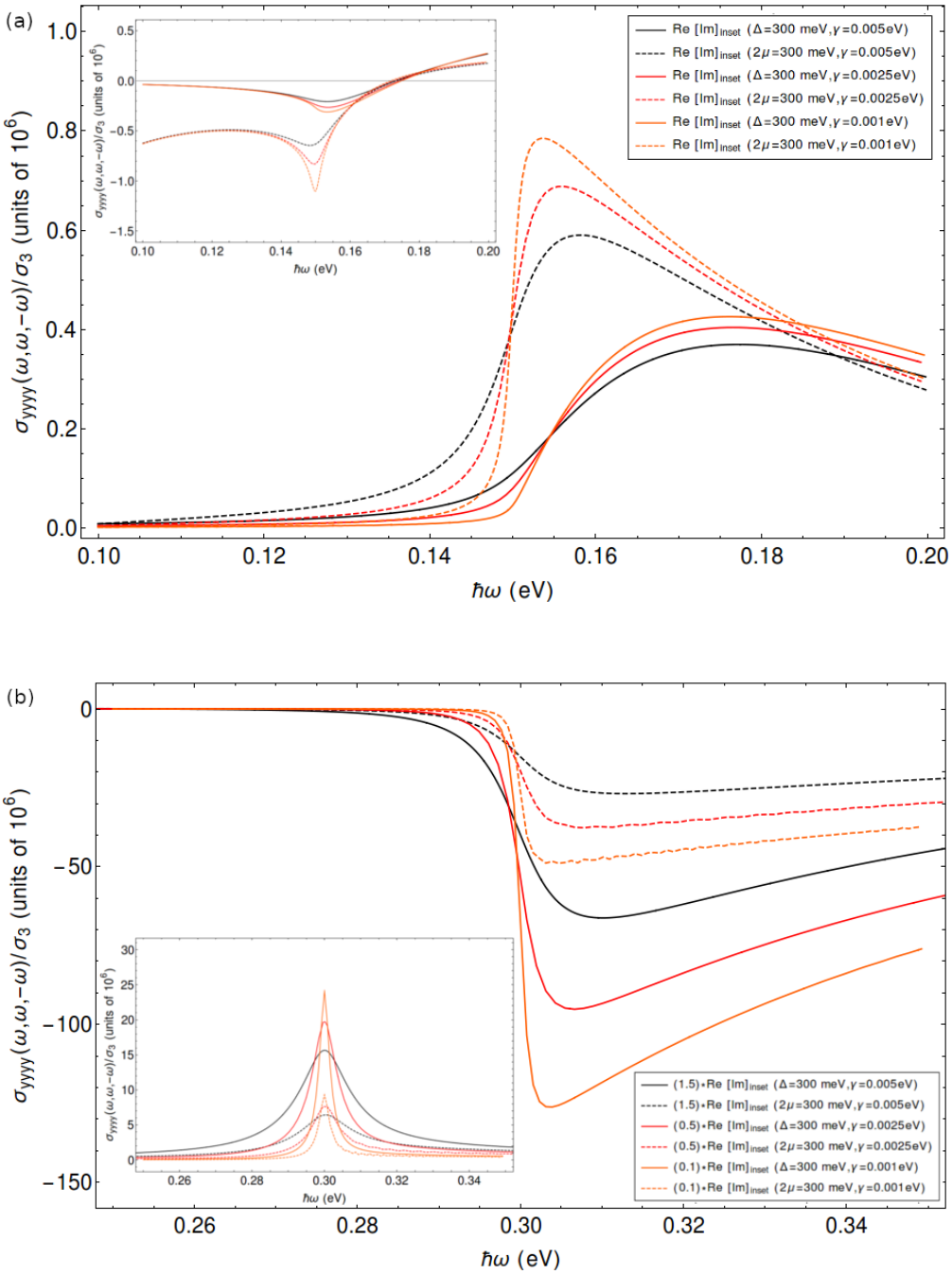


Figure 4.13: The real and imaginary (the latter represented in the insets) parts of the Kerr effect, $\sigma_{yyyy}(\omega, \omega, -\omega)$, in the gapped graphene, $\Delta = 300$ meV, and plain graphene, $2\mu = 300$ meV, monolayers, for frequencies around the two-photon process at the gap, (a), and around the one-photon process at the gap, (b), and for different values of the relaxation parameter: $\gamma = 0.005$ eV (black), $\gamma = 0.0025$ eV (red) and $\gamma = 0.001$ eV (orange). Note that in (b), the conductivities for different γ have been scaled by different factors. As before, the vertical scale in both figures is in units of 10^6. Plots from ref.[5].

Chapter 5

DC Divergences in the NLO Response of a Cold Semiconductor

This chapter is partially based on the following works:

- *Second order divergence in the third order DC response of a cold semiconductor*, **G. B. Ventura**, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, arXiv:2004.01919;

- *Comment on "Jerk Current: A Novel Bulk Photovoltaic Effect"*, **G. B. Ventura**, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, *Phys. Rev. Lett.* 126 (2021), 259701;

This chapter is dedicated to the DC divergences in the optical response of a cold semiconductor, and to the role these play in determining the structure of nonlinear optical coefficients. Unlike the previous two chapters, which followed closely their corresponding papers, this chapter begins and ends with subjects that are not directly associated with our published work; it begins with the derivation of two well-known second order divergences — the injection and shift currents — by means of a simple energy-balance argument, and it ends with the derivation of a new relation between different real parts of the second order conductivity. The middle third of this chapter reproduces the works highlighted above [8, 9], and concerns the third divergence addressed in this chapter, the *jerk* current.[1] In addition, it involves the intersection of the (somewhat) disparate procedures used in rest of the chapter, since both a simple physical argument and a technical, conductivity-based, derivation are used in the description of this effect.

As before, this chapter begins with an introduction to the necessary context, and serves as an overview (even if an extremely abridged one) of the literature on this subject. The rest of the chapter is then roughly divided among the three parts we have just outlined.

[1] As with our previous use of the term *divergence*, this is also somewhat of a misnomer. The jerk current is but one the possible probes to the second order divergence in the third order DC (or quasi-DC) response of a cold semiconductor.

5.1 Introduction

The current response of crystal electrons is known to be, for certain materials and under certain conditions, divergent. The simplest, most obvious, example of this is the DC response of a crystal whose bands are neither entirely filled nor empty, and whose electrons are not subjected to any type of scattering mechanism (either with impurities, or phonons, etc.). In this case, the static electric field accelerates electrons indefinitely, which leads to the existence of asymmetric populations on opposite sides of the FBZ, and thus to the creation of an electric current that (in first order) increases linearly with time, $J^{\beta}_{\mathrm{DC}} \propto t$. When a relaxation mechanism is introduced, say, by adding a term to the Boltzmann equation that causes the electronic population to relax to its equilibrium distribution, the electric current no longer diverges and instead carries a so-called Drude peak or feature, $(\omega + i\gamma)^{-1}$. This type of divergence, which involves only low-frequency electric fields, arises from the *intraband* dynamics of the crystal electrons and is, unsurprisingly, absent in the current response of materials whose bands are entirely filled or empty, as is the case with cold semiconductors. The current response of these materials can, nonetheless, be divergent as well, since the mechanism we have just described — the creation of asymmetric populations — can also be prompted by the existence of *interband* transitions of crystal electrons. The cases of the injection current (in second order) and the jerk current (in third order), which we will study in this chapter, are instances in which this mechanism is at play. The physics of these effects is, therefore, different from that which is involved in low-frequency divergences — this is emphasized by the fact that these interband divergences exist in a frequency *range* (instead of in a frequency limit) — and so are their corresponding descriptions.

The interband transitions involved in these effects, and the effects themselves, can be described in different ways. The first derivation of the injection current [43], which the authors called the circular photogalvanic effect, involved a Fermi's golden rule (FGR) calculation of rate at which valence-band electrons are pumped intro the conduction band [28, 43, 44]. In addition to its simplicity, this type of calculation also shows that certain divergences, like the injection current, are strictly quantum interference effects [28].[2] This requires, however, that one has an *a priori* understanding of the physics of the effect at hand, which is not always easily attained (especially when one considers effects at higher orders). Derivations based on the expressions for the nonlinear optical coefficients, Eqs.(A.8) and (A.10), circumvent this problem by using the fact that these coefficients describe everything there is to known about the crystal's optical response at a given order; a divergent effect must, therefore, correspond to the existence of a divergence, for certain frequency combinations, in a nonlinear optical coefficient [52]. The injection and shift currents were thus readily described as physical divergences of the second order susceptibility in the first comprehensive study of the nonlinear optical susceptibilities derived in the length gauge [22]. In the following years, several

[2] We can also use a slightly different argument, based on the study of the energy balance of this system, to derive the injection and shift currents. Its reasoning, nonetheless, runs along the same lines as that of the FGR calculation.

other divergences, in either the susceptibility (e.g., the current-induced second harmonic generation, ref.[169]) or the conductivity (e.g., the two-colour injection current, ref.[52]), were identified using this type of procedure, which attests to its usefulness. There is thus no reason to choose one of these descriptions over the other, as they can be used in a complementary manner; the parallel development of the two can give us a fuller understanding of these effects.[3]

Divergences in the nonlinear optical response have been, therefore, a well-established subject of study for many years, and, like many other topics in nonlinear optics, have benefited from the recent developments in the areas of two-dimensional materials [51, 66, 68, 82, 86, 101, 102, 151, 152] and Weyl semimetals [41, 54, 56, 114–124]. Divergences like the jerk current, the cross-phase modulation and the degenerate four-wave mixing [53, 109–111], have been derived only recently using nothing but the well-known expressions for the nonlinear optical coefficients. It has been emphasized that such a systematic study of these divergences is important for nonlinear optics, because these describe, in principle, effects that should be easily detectable, and for which there may be useful applications [53]. Of these new divergences, the jerk current can be picked out as being of particular interest, since it corresponds to a *second* order divergence (in the relaxation rate, η) in the third order (strictly) DC conductivity, $\sigma(\omega, -\omega, 0)$, and thus to a finite second time derivative of the current response in the clean limit, $\eta \to 0$, ref.[109]. The study of this effect can, however, be subsumed into a general analysis of the second order divergence of the third order DC (or quasi-DC) conductivity that can be probed, as we have shown in ref.[8], using different setups of the electric field. The jerk current is but one possible probe — the one associated with an electric field that has both a static and a monochromatic component — for which $\sigma \propto \eta^{-2}$. Two additional probes can be considered: a dichromatic probe, associated with a *dichromatic* field of frequencies ω and $\delta \ll \omega$, and for which $\sigma \propto \eta^{-1}\delta^{-1}$; a trichromatic probe, associated with a *trichromatic* field of frequencies, $\omega + \delta_1 \sim \omega$, ω and $\delta_2 \ll \omega$, and for which $\sigma \propto \delta_1^{-1}(\delta_1 + \delta_2)^{-1}$. The three different setups involve output frequencies that are either zero, $\omega_{123} = 0$, or are small, $\omega_{123} = \delta \ll \omega$, $\omega_{123} = \delta_1 + \delta_2 \ll \omega$, and therefore fall under the scope of a DC (or quasi-DC) response. In this chapter, we recover the results from ref.[8] and show that, regardless of the setup, the divergent response always involves the *same* coefficient. This feature of the divergence had not been noted before. In addition, the expression we derived for this effect and our physical interpretation of jerk current are different from those original proposed by Fregoso *et al.* in ref.[109]. It is important, therefore, that we study the jerk current using complementary descriptions, whose results can cross-check one another.

Divergences like the injection and jerk currents, besides being relevant in themselves, also play a singular role in determining the structure of the nonlinear optical coefficients with which they are associated. As we have noted above, effects like the injection and jerk currents are connected with certain divergent portions of the response functions, which is to say, they are connected with

[3] The original descriptions of the two-colour injection current, $\sigma(\omega, \omega, -2\omega)$, are a good example of this [44, 52]. The FGR formulation of this effect has more recently been used to describe the two-colour current injection in topological insulators [49], graphene (monolayer and bilayer) [50], and transition metal dichalcogenides [51].

specific regions of our n-dimensional frequency space (n being the order of the response) in which the response function is nonanalytic. Whenever these divergences appear, the Kramers-Kronig relations must be reassessed, so as to take the these effects into account. It is important, therefore, that these new relations, and the reasoning that should go into them, are written out clearly.

Finally, we recover a comment Aversa and Sipe made in their original work on the length gauge [22], which concerns the study of these divergences in the NLO response. A quantitative description of current generation in crystals, they argued, most likely requires the use of treatments that go beyond the scope of perturbation theory, and that can account for "effects such as saturation, space charge accumulation, and dephasing processes". This means moving away from much of what has been discussed, or is yet to be discussed, in this thesis. There are, nonetheless, aspects that were covered here that can be used — and, in fact, are being used — in the development of a nonperturbative, time-domain, description of the current response. We hope that this proves to be a fruitful and interesting area of study.

The contents of the remaining sections of this chapter can be summed up as follows. In the subsequent section, we derive the second order injection and shift currents by studying the energy balance of a cold semiconductor subjected to an external electric field, and applying some physical reasoning. This section, although mostly a description of two well-known and well-understood effects, introduces many of the ideas used in our semiclassical description of the jerk current. Furthermore, it also introduces the injection current, which is relevant for our study of the structure of the second order conductivity; in Section 5.3, we return to a conductivity-based description of the current and isolate the portion of the third order conductivity carrying the second order divergence in the DC (and quasi-DC) response. We discuss how different electric field setups probe this divergence in different ways. In addition to this formal derivation, we present a numerical check of this result, performed in the velocity gauge and using the gapped graphene monolayer as its subject model, as well as a semiclassical description of the jerk current which, far from being a simple cross-check, provides us with its physical interpretation. These three results are in agreement, meaning that the jerk current can now be deemed to be well described; the final two sections of the chapter are then connected with the internal structure of second order conductivities. In Section 5.4, we show that the imaginary part of the second order conductivity of a cold semiconductor has, for frequencies below the electronic band gap, full permutation symmetry. This property and the expression for the injection current derived in Section 5.2 are then used to derive a new relation between real parts of the second order conductivity tensor. This relation reveals some interesting, previously unknown, aspects of the internal structure of these conductivities.

5.2 Divergences in the Second Order Response of a Cold Semiconductor

Cold semiconductors are crystals that have a gapped band structure, in which the valence and conduction bands are considered to be either completely full or completely empty, i.e., $\Delta\epsilon^{\text{min.}}_{\mathbf{k}cv} =$

$\Delta > 0$, and $f_{\mathbf{k}v(c)} \approx 1(0)$, for $k_B T \ll \Delta$. The description of the optical response of these materials is, of course, different from that of metals or doped semiconductors, since neither the full intraband term, nor any other term carrying a derivative (or derivatives) of the Fermi-Dirac distribution function, contribute to the response. From a technical standpoint, the cold semiconductor is a system in which about half the terms of an optical conductivity are zero and, hence, simpler to study; more importantly, and from a physical standpoint, the absence of these terms dictates that there can be no relevant low-frequency effects in these materials. Herein lies the reason why the first divergences in the optical response of a cold semiconductor appear only in *second* order. The first order optical response has a single, low-frequency, divergence — the Drude peak — that is expressed in terms of the first derivative of the Fermi-Dirac distribution function. It is, therefore, necessarily absent in the response of a cold semiconductor; electrons in a filled band do not contribute to the current response (at low frequencies).[4]

The origin of the divergences in the second order response of a cold semiconductor is thus somewhat more complicated.[5] It is also, as we will presently see, not connected with a specific frequency limit (as in the case of Drude peaks), but instead with a specific frequency *range*. In order to derive the injection and shift currents, we will use a practical argument based on the energy balance of this physical system (cold semiconductor + external electric field) in conjunction with some other physical considerations. We must add that, while these effects can be derived directly from the expression for the relevant (in this case, a second order) conductivity — as will be the case, later on, with the jerk current — that procedure is not particularly enlightening, and therefore we do not present it here, especially since the shift and injection currents are well-understood and comprehensively studied effects [22, 52]. The jerk current and the remaining divergences in the third order response (other than the two-colour injection current) are, however, newly identified effects, and thus require a careful approach based on the conductivity; this is supplemented by a more physical and intuitive description of the effect. Furthermore, the present derivation of the injection and shift currents is more than just a first step into the subject of divergences in the optical response of cold semiconductors. Divergences, as we will see in a later section, also play a singular role in determining the structure of these response functions.

The dynamics of a crystal electron subjected to a uniform external electric field can be intuitively decomposed into two types of process: an intraband process, in which an electron is accelerated

[4] The low-frequency, intraband, current response of crystal electrons can be appropriately described by the Boltzmann equation. In the absence of scattering, the population distribution at a given point of the FBZ, $\mathbf{k}$, and at an instant t, reads simply as $n_s(\mathbf{k}, t) = f_s(\mathbf{k} + |e|\,\hbar^{-1}\mathbf{A}(t))$, which is *always* either one or zero for a cold semiconductor. In this type of material, and for the fields being considered, states on opposite sides of the FBZ are always equally populated. This means that for filled bands, the contributions associated with opposite crystal momenta $\mathbf{k}$, $-|e|\,v^{\beta}_{\mathbf{k}ss}$ and $-|e|\,v^{\beta}_{-\mathbf{k}ss} = +|e|\,v^{\beta}_{\mathbf{k}ss}$, cancel out exactly. As for the empty bands, their contributions are necessarily zero, as the states are unoccupied. Low-frequency currents do not exist in cold semiconductors.

[5] We note that calling the shift current a divergence in the second order response can be somewhat of a misnomer, since it corresponds to a divergence in the second order susceptibility, not the conductivity. Furthermore, the effects we describe in this section are, in principle, present in any type of crystal (be it a metal, a semiconductor or an insulator).

across the first Brillouin zone, while remaining in its original band, $(\mathbf{k}, s) \to (\mathbf{k}', s)$; and an interband process, in which an electron is pumped into a different band, while retaining its crystal momentum, $(\mathbf{k}, s) \to (\mathbf{k}, s')$. For a cold semiconductor, as we have already noted, the first type of process cannot, by itself, generate an electric current.[6] That leaves us, in the lowest order of the response, only with the interband processes and their corresponding contributions (See Eq.(A.6)),

$$\sigma^{\alpha_1\alpha_2}(\omega) = -\frac{e^2}{\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Delta\epsilon_{\mathbf{k}s's}\Delta f_{\mathbf{k}s's}\xi^{\alpha_1}_{\mathbf{k}s's}\xi^{\alpha_2}_{\mathbf{k}ss'}\frac{1}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'} + i\eta}. \tag{5.1}$$

Let us extend this argument further, by considering this system's energy balance. For a clean cold semiconductor, that is, a system in which there can be no ohmic dissipation, energy can only be transferred to the crystal electrons by pumping them from a valence into a conduction band; energy cannot be expended in the acceleration of electrons, since intraband currents are necessarily zero. This means, in quantitative terms, that the average over a few periods $(m \times T)$ of the rate at which the energy of the system changes, which reads as,

$$\dot{\epsilon}(t) = \int \frac{d^d\mathbf{k}}{(2\pi)^d}\mathrm{Tr}\left[\dot{\rho}(t)\,\mathcal{H}\right], \tag{5.2}$$

must be equal to the number of electrons pumped into the conduction (per unit time and area), $\dot{n}_1(t)$, times the energy associated with each transition, $\hbar\omega$,

$$\hbar\omega\dot{n}(t) = \langle\dot{\epsilon}(t)\rangle_{mT}. \tag{5.3}$$

Note that this holds if and only if the population density increases *linearly* with time. As for the right-hand side of Eq.(5.2), remember that, in the length gauge, the time evolution of the density matrix is described by Eq.(3.92), so we can write $\dot{\epsilon}(t)$ as,

$$\dot{\epsilon}(t) = \frac{|e|}{\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\mathbf{E}(t)\cdot\mathrm{Tr}\left[\left[\mathbf{D}, \rho(t)\right]\mathcal{H}\right] = -|e|\int \frac{d^d\mathbf{k}}{(2\pi)^d}\mathbf{E}(t)\cdot\mathrm{Tr}\left[\mathbf{v}\,\rho(t)\right], \tag{5.4}$$

where we have used the cyclic property of the trace and the definition of the velocity operator in band-index space, Eq.(3.75). This relates the energy transfer rate with the dot product of the electric current density and the external electric field; the time average of the former quantity can then be expressed as,

$$\hbar\omega\dot{n}(t) = \langle\mathbf{E}(t)\cdot\mathbf{J}(t)\rangle_{mT}, \tag{5.5}$$

which shows that the population density of injected electrons is related to the linear conductivity.

[6] It can, however, when it is combined with interband processes, play a role in the generation of an electric current.

For a monochromatic electric field,

$$E^{\alpha_1}(t) = E^{\alpha_1}(\omega)e^{-i\omega t} + \text{c.c.}, \tag{5.6}$$

the linear current density reads as,

$$J^{\alpha_1}(t) = \sigma^{\alpha_1\alpha_2}(\omega)E^{\alpha_2}(\omega)e^{-i\omega t} + \text{c.c.}, \tag{5.7}$$

so that by performing the average over a few periods, $m \times T = 2\pi m/\omega$, of $\mathbf{E}(t) \cdot \mathbf{J}(t)$, we obtain,[7]

$$\hbar\omega\dot{n}(t) = 2\text{Re}\left[\sigma^{\alpha_1\alpha_2}(\omega)E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\right]. \tag{5.8}$$

Finally, we can express the population density rate of injected electrons in terms of the real and imaginary parts of the conductivity. After some additional manipulations, we obtain an expression for the population density rate, Eq.(B.17),

$$\dot{n}(t) = \frac{\pi e^2}{\hbar}\text{Re}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\right]\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Delta f_{\mathbf{k}s's}\left\{\xi^{\alpha_1}_{\mathbf{k}s's}, \xi^{\alpha_2}_{\mathbf{k}ss'}\right\}\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}),$$

or, alternatively, an expression for the population density rate at a given point of the first Brillouin zone,

$$\dot{n}(\mathbf{k},t) = \frac{2\pi e^2}{\hbar}\sum_{s's}\Delta f_{\mathbf{k}s's}\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'})\text{Re}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\xi^{\alpha_1}_{\mathbf{k}s's}\xi^{\alpha_2}_{\mathbf{k}ss'}\right], \tag{5.9}$$

which is, unsurprisingly, equal to the probability transition rate that follows from a Fermi's golden rule calculation. For simplicity, we consider from now on a single pair of conduction and valence bands in our calculations,

$$\dot{n}(\mathbf{k},t) = \frac{2\pi e^2}{\hbar}\Delta f_{\mathbf{k}vc}\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}cv})\text{Re}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\xi^{\alpha_1}_{\mathbf{k}vc}\xi^{\alpha_2}_{\mathbf{k}cv}\right]. \tag{5.10}$$

This simple energy argument has given us the expression for the rate at which electrons are pumped into the conduction band.[8] We proceed to derive the currents associated with these electrons.

[7] The time average carries two contributions. One is time-independent and can be trivially integrated, so as to give, $\sigma^{\alpha_1\alpha_2}(\omega)E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega) + \text{c.c.}$; as for the other contribution, it involves the following integral,

$$\frac{1}{mT}\int_0^{mT} dt\, e^{-i2\omega t} \approx 0,$$

which can be neglected when we average over a reasonable number of periods [16]. This second contribution is thus dropped from Eq.(5.5).

[8] This simply energy argument does not, however, tell us everything there is to be known about these effects. A calculation of a quantum probability rate, such as the one in ref.[28], shows that the injection current is a quantum interference effect, in which different possible transition paths are mixed in such a way as to generate an imbalance between populations on opposite sides of the FBZ. As for the shift current, while it is not strictly an interference effect, it can also include this type of contribution.

5.2.1 The Injection Current

As the reader will note, what we have been presenting in this section is a simplified description of electric currents in crystals, in which the elements of the response — rigorously speaking, the matrix elements of both the appropriate observable and the density matrix — are treated in a rather heuristic fashion. The end result of our previous argument, Eq.(5.10), describes the latter of these two elements, and so we must now consider the former. The simplest possible description of an electric current, which we have briefly discussed as a side note, involves the diagonal elements of the velocity operator. These matrix elements describe the traditional velocities of carriers in a crystal, $v^{\beta}_{\mathbf{k}ss} = \hbar^{-1}\nabla^{\beta}_{\mathbf{k}}\epsilon_{\mathbf{k}s}$, and are expressed in terms of derivatives of the band energies. They are, therefore, odd functions of $\mathbf{k}$, so that the relevant portion of the population density rate is that which is also odd,

$$\Delta\dot{n}(\mathbf{k},t) = \frac{\dot{n}(\mathbf{k},t) - \dot{n}(-\mathbf{k},t)}{2}. \tag{5.11}$$

Furthermore, note that we are dealing with a finite population density *rate*, which means that this current we are presently considering is not, strictly speaking, a static (or direct) current — it corresponds to a finite, and constant, time derivative of a current, which is usually called the *injection* current. In the case of the two-band model, the injection current reads as,

$$\dot{J}^{\beta}_{\text{inj.}}(t) = \int \frac{d^d\mathbf{k}}{(2\pi)^d}\Delta\dot{n}(\mathbf{k},t) \times (-|e|)\big(v^{\beta}_{\mathbf{k}cc} - v^{\beta}_{\mathbf{k}vv}\big). \tag{5.12}$$

Its physical interpretation is very clear: electron-hole pairs are created when valence-band electrons are pumped into the conduction band, leaving a hole behind. These electron-hole pairs can then generate a current response — which, like the population density, increases linearly with time — provided that the population densities on opposite sides of the FBZ do not cancel each other, i.e., that $\Delta\dot{n}(\mathbf{k},t) \neq 0$.

Let us now consider the expression for the population density rate, Eq.(5.10), so as to write $\Delta\dot{n}$ as,

$$\Delta\dot{n}(\mathbf{k},t) = \frac{\pi e^2}{\hbar}\Delta f_{\mathbf{k}vc}\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}cv})\text{Re}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\left[\xi^{\alpha_1}_{\mathbf{k}vc}, \xi^{\alpha_2}_{\mathbf{k}cv}\right]\right], \tag{5.13}$$

by using the parity of the Berry connections. In addition, conservation of probability imposes a second property on these objects, $\left(\xi^{\alpha_1}_{\mathbf{k}vc}\right)^* = \xi^{\alpha_1}_{\mathbf{k}cv}$. The real part in this equation is then expressed as,

$$i\text{Im}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\right]\left[\xi^{\alpha_1}_{\mathbf{k}vc}, \xi^{\alpha_2}_{\mathbf{k}cv}\right], \tag{5.14}$$

since the commutator of two Berry connections is strictly imaginary. Replacing the expression for $\Delta\dot{n}(\mathbf{k},t)$ in Eq.(5.12), we finally obtain,

$$\boxed{\dot{J}^{\beta}_{\text{inj.}}(t) = \frac{i\pi|e|^3}{\hbar^2}\text{Im}\left[E^{\alpha_1}(\omega)E^{\alpha_2}(-\omega)\right]\int \frac{d^d\mathbf{k}}{(2\pi)^d}\left[\xi^{\alpha_1}_{\mathbf{k}vc}, \xi^{\alpha_2}_{\mathbf{k}cv}\right]\left(\nabla^{\beta}_{\mathbf{k}}\Delta\epsilon_{\mathbf{k}cv}\right)\Delta f_{\mathbf{k}vc}\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}cv})}, \tag{5.15}$$

the expression for the injection current [52].[9] For all its simplicity, this result is also not without some subtleties that require some additional comments. First, note that this effect does not exist in materials with certain point group symmetries, even if they are not centrosymmetric; for materials in which the second order conductivity tensor has elements only along the diagonal, or alternatively, in which the remaining nondiagonal elements can be reduced to the diagonal ones, as is the case in the two-dimensional material that we have been considering here, the injection current is necessarily absent. Secondly, the generation of this current requires a particular type of electric field. For linearly polarised electric fields, the injection current is, again, absent, as their components are real: $\big[E^{x(y)}(\omega)\big]^* = E^{x(y)}(\omega)$; this effect requires a phase difference between components of electric field, which means that its polarisation must be either elliptical or circular (which is why the authors of ref.[43] called it the circular photogalvanic effect). Thirdly, and as we have noted in Subsection 4.2.1, this injection current increases linearly with time and it is, therefore, associated with a divergence in the second order conductivity. Note, however, that unlike the divergences associated with Drude peaks, where the conductivity goes to infinity as its frequency arguments go to zero, this divergence exists in a frequency range, specifically, for all frequencies above the electronic band gap. This attests to the fact that the physical origins of these effects are fundamentally different.

Finally, let us end this subsection on the injection current by recovering a result derived in Subsection 4.2.1. Note that, by separating the fields from the rest of the expression in Eq.(5.15), we can obtain the expression for the divergent portion of the conductivity we wrote in Eq.(4.6),

$$\lim_{\eta\to 0^+} 4\eta\,\mathrm{Im}\big[\sigma^{\beta\alpha_1\alpha_2}_{\mathrm{div.}}(\omega,-\omega)\big] = -\frac{i\pi|e|^3}{\hbar^2}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\big[\xi^{\alpha_1}_{\mathbf{k}vc},\xi^{\alpha_2}_{\mathbf{k}cv}\big]\big(\nabla^{\beta}_{\mathbf{k}}\Delta\epsilon_{\mathbf{k}cv}\big)\Delta f_{\mathbf{k}vc}\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}cv}). \quad (5.16)$$

This is exactly the result that follows from the expression for the second order conductivity, Eq.(A.8).

5.2.2 The Shift Current

There is a second effect associated with the second order DC conductivity: the shift current. As with the injection current, the shift current can be understood in terms of the same interband processes, in which electron-hole pairs are created by optical pumping. In this case, however, the response is truly a direct current, meaning that the corresponding $J^{\beta}_{\mathrm{shi.}}(t)$ is finite and constant: our description of this effect as a divergence in the second order response is somewhat of a misnomer; the divergence does indeed exist, but it is present in the second order susceptibility, not the conductivity. Furthermore, its physical interpretation is somewhat more intricate, and the derivation that accompanies it is not as sound as that of the injection current. The former is as follows.

We begin by considering the charge distributions of valence-band and conduction-band electrons. The Bloch functions associated with these electrons are, as we know very well, orthogonal, which

[9] Note that we have exchanged the Cartesian indexes in the imaginary part of the product of electric fields, which gives us an additional minus sign.

means that valence-band electrons should have a charge distribution in real space, $-|e|\ |\psi_{\mathbf{k}v}(\mathbf{r})|^2$, that is different from that of conduction-band electrons, $-|e|\ |\psi_{\mathbf{k}c}(\mathbf{r})|^2$. We can see this in the case of the monolayer of gapped graphene, where the probability distribution of valence-band electrons is slightly more pronounced around the lowest-energy sites, the B atoms, than around the highest-energy sites, the A atoms; the opposite is true for the probability distribution of conduction-band electrons. When an electron is pumped from the valence band into the conduction band, there must be, in addition to this interband transition, the creation of an electric dipole moment between the electron in the conduction band and the hole in the valence band [170]. The sum total of all these dipole contributions (over the first Brillouin zone) gives us the macroscopic polarisation associated with the continuous creation of these electron-hole pairs. Note that, since the population of electron-hole pairs, Eq.(5.10), increases linearly with time, the polarisation must also increase linearly with time, $\mathbf{P}(t) \propto t$. The corresponding electric electric current, the polarisation current, is therefore constant; this DC contribution to the current response is the so-called shift current.

Let us now formulate this in quantitative terms. Following this argument, the shift current reads as,[10]

$$J^{\beta}_{\text{shi.}}(t) = \int \frac{d^d\mathbf{k}}{(2\pi)^d} \dot{n}(\mathbf{k},t) \times (-|e|)(\xi^{\beta}_{\mathbf{k}cc} - \xi^{\beta}_{\mathbf{k}vv}). \tag{5.17}$$

For the systems we are considering, in which time-reversal symmetry is present, the intraband elements of the Berry connection are even functions of $\mathbf{k}$, meaning that, $\xi^{\beta}_{\mathbf{k}ss} = \xi^{\beta}_{-\mathbf{k}ss}$. Like the injection current, the shift current gets a portion of $\dot{n}(\mathbf{k},t)$ that has a well-defined parity, which in this case is the *average* of the contributions from opposite sides of the FBZ,

$$\dot{n}_{\text{ave.}}(\mathbf{k},t) = \frac{\dot{n}(\mathbf{k},t) + \dot{n}(-\mathbf{k},t)}{2}. \tag{5.18}$$

Note, however, that unlike the injection current, the shift current is a cumulative effect. Taking the expression for the time derivative of Eq.(5.10), we can now write $\dot{n}_{\text{ave.}}(\mathbf{k},t)$ as,

$$\dot{n}_{\text{ave.}}(\mathbf{k},t) = \frac{\pi e^2}{\hbar} \Delta f_{\mathbf{k}vc} \delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}cv}) \text{Re}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\left\{\xi^{\alpha_1}_{\mathbf{k}vc}, \xi^{\alpha_2}_{\mathbf{k}cv}\right\}\right], \tag{5.19}$$

where we have once again used the parity of the Berry connections. The real part of this equation is then expressed as,

$$\text{Re}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\right]\left\{\xi^{\alpha_1}_{\mathbf{k}vc}, \xi^{\alpha_2}_{\mathbf{k}cv}\right\}, \tag{5.20}$$

since the anticommutator of two Berry connections is strictly real. Replacing the expression for

[10] We must add two comments here: first, note that the density-matrix calculation of the shift current *does* involve the interband matrix elements. This helps explain why the following derivation is not entirely sound; secondly, the intraband Berry connections are not the intraband matrix elements of the position operator. The latter, as we have seen, involve a derivative with respect to $\mathbf{k}$, which means that they must always act on something. In this case, however, it is not entirely clear on what this derivative should act, as $\dot{n}(\mathbf{k},t)$ is periodic in reciprocal space. We discussed this issue, if somewhat briefly, in Subsection 3.4.1.

$\dot{n}_{\text{ave.}}(\mathbf{k},t)$ in Eq.(5.17), we finally obtain,

$$\boxed{J^{\beta}_{\text{shi.}}(t) = -\frac{i\pi|e|^3}{\hbar^2}\text{Re}\left[E^{\alpha_1}(\omega)E^{\alpha_2}(-\omega)\right]\int \frac{d^d\mathbf{k}}{(2\pi)^d}\left[\xi^{\alpha_1}_{\mathbf{k}vc},(\xi^{\alpha_2}_{\mathbf{k}cv})_{;\beta}\right]\Delta f_{\mathbf{k}vc}\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}cv})}, \tag{5.21}$$

by imposing the fact that its integrand *must* be gauge invariant. This requires some additional manipulations, which we present in Appendix B.2.

Finally, and as we have done with the injection current, we can relate the shift current with a specific portion of the second order conductivity. We know, from the expression for the second order response to a monochromatic field, Eq.(4.5), that,

$$J^{\beta}_{\text{shi.}}(t) = 2\sigma^{\beta\alpha_1\alpha_2}_{\text{reg.}}(\omega,-\omega)\text{Re}\left[E^{\alpha_1}(\omega)E^{\alpha_2}(-\omega)\right]. \tag{5.22}$$

Separating the fields from the rest of the expression in Eq.(5.21), we obtain,

$$\sigma^{\beta\alpha_1\alpha_2}_{\text{reg.}}(\omega,-\omega) = -\frac{i\pi|e|^3}{2\hbar^2}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\left[\xi^{\alpha_1}_{\mathbf{k}vc},(\xi^{\alpha_2}_{\mathbf{k}cv})_{;\beta}\right]\Delta f_{\mathbf{k}vc}\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}cv}). \tag{5.23}$$

This is exactly the result that follows from the expression for the second order conductivity, Eq.(A.8).

5.3 The Second Order Divergence in the Third Order DC Response of a Cold Semiconductor

The physical arguments we have presented in the previous section for the injection and shift currents provide some physical intuition as to what is behind these effects. Later on, we will see how similar ideas ideas can be used to derive the jerk current and, more importantly, to obtain a clear physical interpretation of this effect.[11] It is, however, necessary at this point, to recover the conductivity-based description of these divergences, as we move on to study the second order divergence in the third order response of a cold semiconductor. There has been some discussion on the expression describing this effect [8, 9, 109, 110] — as well as on the underlying derivation itself — and so it is sensible to proceed with some caution. As mentioned in the introduction to this chapter, the results from this conductivity-based derivation are cross-checked with results that follow both from a physical argument, presented in subsection 5.3.2, and from a numerical calculation of these responses in gapped graphene monolayer. In addition, the procedure we describe here can be used in the treatment of certain (seemingly) divergent terms in the integrand function which appear when

[11] We will not, however, use an energy-balance argument to derive the expression for the relevant population density of conduction-band electrons at this (third) order. This sort of argument can still be used — one simply has to relate the second time derivative of this population density with the electrooptic effect, $\sigma^{\beta\alpha_1\alpha_2}(\omega,0)$, much in the same way that the relation between the time derivative of the population density and the linear conductivity was derived — but it involves a rather dense, and unnecessarily lengthy, calculation.

studying other nonlinear optical effects [21].

Begin by considering, once again, the expression for the third order response in the time domain, Eq.(4.7),

$$J^{\beta}(t) = \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\frac{d\omega_3}{2\pi}\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3)E^{\alpha_1}(\omega_1)E^{\alpha_2}(\omega_2)E^{\alpha_3}(\omega_3)e^{-i(\omega_{123}+3i\eta)t}.$$

The divergence we are interested in is the highest order divergence (at this order) of the conductivity associated with this type of response — DC or nearly DC. Just as in the case of the injection and shift currents, this divergence is connected with responses whose output frequency is either zero or much smaller than the frequency of the optical field, i.e., $\omega_{123} = 0$ or $\omega_{123} = \delta \ll \omega$. To see that this is the case, we will work directly with the third conductivity we derived in Appendix A.1, which, still unsymmetrized, reads as,[12]

$$\begin{aligned}\frac{1}{ie^4}\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's} v^{\beta}_{\mathbf{k}s's}\frac{1}{\hbar\bar{\omega}_{123}-\Delta\epsilon_{\mathbf{k}ss'}} \\ \times\left[D^{\alpha_1},\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon}\circ\left[D^{\alpha_2},\frac{1}{\hbar\bar{\omega}_{3}-\Delta\epsilon}\circ\left[D^{\alpha_3},\rho^{(0)}(\mathbf{k})\right]\right]\right]_{ss'},\end{aligned} \tag{5.24}$$

for $\bar{\omega}_i = \omega_i + i\eta$. First, we exchange the band labels in the denominator outside the nested commutator,

$$\frac{1}{\hbar\bar{\omega}_{123}-\Delta\epsilon_{\mathbf{k}ss'}} \to \frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon_{\mathbf{k}s's}}, \tag{5.25}$$

so that this term can be combined with the velocity matrix element as,

$$v^{\beta}_{\mathbf{k}s's}\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon_{\mathbf{k}s's}} = \left(v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\right)_{\mathbf{k}s's}, \tag{5.26}$$

using the Hadamard product of two matrices, which we introduced in Subsection 3.4.2. This allows us to express Eq.(5.24) as,

$$\frac{1}{ie^4}\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \int \frac{d^d\mathbf{k}}{(2\pi)^d}\mathsf{Tr}\left[v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\left[D^{\alpha_1},\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon}\circ\left[D^{\alpha_2},\frac{1}{\hbar\bar{\omega}_{3}-\Delta\epsilon}\circ\left[D^{\alpha_3},\rho^{(0)}(\mathbf{k})\right]\right]\right]\right]. \tag{5.27}$$

Using the cyclic invariance of the trace — $\mathsf{Tr}\left[A\left[B,C\right]\right] = -\mathsf{Tr}\left[\left[B,A\right]C\right]$ — and integrating by parts over $\mathbf{k}$, we can move the covariant derivative, $D^{\alpha_1}_{\mathbf{k}}$, from one side to the other, so as to act on the terms on its left. The conductivity then reads,

$$\frac{1}{ie^4}\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \int \frac{d^d\mathbf{k}}{(2\pi)^d}\mathsf{Tr}\left[\left[D^{\alpha_1},v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\right]\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon}\circ\left[D^{\alpha_2},\frac{1}{\hbar\bar{\omega}_{3}-\Delta\epsilon}\circ\left[D^{\alpha_3},\rho^{(0)}(\mathbf{k})\right]\right]\right]. \tag{5.28}$$

[12] Note that the expression for the conductivity that we are using here is different from that of ref.[8], so as to be consistent with the rest of this thesis.

We can now separate this conductivity into its interband, $s' \neq s$, and intraband, $s' = s$, portions, so as to isolate $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}$, the element in the response function carrying the highest order divergence,

$$\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \Gamma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) + (....), \tag{5.29}$$

the ellipsis represents the remaining contributions to the conductivity. We show in Appendix B.3, after a rather lengthy argument, that the interband portion of Eq.(5.28) cannot contribute to $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}$, since interband denominators cannot be directly associated with divergent contributions. The only relevant contributions to $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}$ are, therefore, and exactly as in the injection current (in the second order response), those associated with the intraband portion of the response function. This means that for there to be a second order divergence, two of the three denominators in Eq.(5.28) must have trivial energy differences, $\Delta\epsilon_{\mathbf{k}ss} := 0$, while the remaining one must involve a nontrivial interband transition. Let us consider these terms.

For $s' = s$, the integrand in Eq.(5.28) reads as,

$$-\frac{1}{\hbar\bar{\omega}_{23}}\sum_s \big[D^{\alpha_1}, v^\beta \circ \frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\big]_{ss}\big[D^{\alpha_2}, \frac{(-i)}{\hbar\bar{\omega}_3-\Delta\epsilon}\circ\big[\xi^{\alpha_3},\rho^{(0)}(\mathbf{k})\big]\big]_{ss}. \tag{5.30}$$

We consider separately the two commutators, so as to make this derivation clearer. The commutator carrying the velocity matrix reads as,

$$\big[D^{\alpha_1}, v^\beta \circ \frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\big]_{ss} = \frac{1}{\hbar\bar{\omega}_{123}}\big(\nabla^{\alpha_1} v^\beta_{\mathbf{k}ss}\big) - i\big[\xi^{\alpha_1}, v^\beta \circ \frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\big]_{ss}. \tag{5.31}$$

We can identify the first term in this equation as the relevant contribution to $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}$; the terms in the commutator either involve interband transitions, and thus cannot be directly associated with a divergent contribution, or they are exactly zero. As for the nested commutator carrying the Fermi-Dirac distribution function,

$$\big[D^{\alpha_2}, \frac{(-i)}{\hbar\bar{\omega}_3-\Delta\epsilon}\circ\big[\xi^{\alpha_3},\rho^{(0)}(\mathbf{k})\big]\big]_{ss} = (-i)^2\big[\xi^{\alpha_2}, \frac{1}{\hbar\bar{\omega}_3-\Delta\epsilon}\circ\big[\xi^{\alpha_3},\rho^{(0)}(\mathbf{k})\big]\big]_{ss}, \tag{5.32}$$

there is also a single relevant term, as $\Delta f_{\mathbf{k}ss} := 0$. Gathering the relevant contributions from Eqs.(5.31) and (5.32), as well as the factor of $-(\hbar\bar{\omega}_{23})^{-1}$, we obtain,

$$\frac{1}{\hbar\bar{\omega}_{123}}\frac{1}{\hbar\bar{\omega}_{23}}\sum_s\big(\nabla^{\alpha_1}v^\beta_{\mathbf{k}ss}\big)\big[\xi^{\alpha_2}, \frac{1}{\hbar\bar{\omega}_3-\Delta\epsilon}\circ\big[\xi^{\alpha_3},\rho^{(0)}(\mathbf{k})\big]\big]_{ss}, \tag{5.33}$$

which can then be written as,

$$-\frac{1}{\hbar\bar{\omega}_{123}}\frac{1}{\hbar\bar{\omega}_{23}}\frac{1}{\hbar}\sum_{r\neq s}\big(\nabla^\beta\nabla^{\alpha_1}\epsilon_{\mathbf{k}s}\big)\left[\frac{\xi^{\alpha_3}_{\mathbf{k}sr}\xi^{\alpha_2}_{\mathbf{k}rs}\Delta f_{\mathbf{k}rs}}{\hbar\bar{\omega}_3-\Delta\epsilon_{\mathbf{k}sr}} - (s\leftrightarrow r)\right], \tag{5.34}$$

by using the definition of the velocity matrix elements, Eq.(3.75). Finally, we can manipulate the band index sums so as to cast the terms inside the square brackets as a single contribution; relabelling $r \to s'$, we obtain,

$$-\frac{1}{\hbar\bar{\omega}_{123}}\frac{1}{\hbar\bar{\omega}_{23}}\frac{1}{\hbar}\sum_{s\neq s'}\left(\nabla^{\beta}\nabla^{\alpha_1}\Delta\epsilon_{\mathbf{k}ss'}\right)\frac{\xi^{\alpha_3}_{\mathbf{k}ss'}\xi^{\alpha_2}_{\mathbf{k}s's}\Delta f_{\mathbf{k}s's}}{\hbar\bar{\omega}_3-\Delta\epsilon_{\mathbf{k}ss'}}. \tag{5.35}$$

Note that the derivatives with respect to $\mathbf{k}$ now involve the difference of band energies, $\Delta\epsilon_{\mathbf{k}ss'}$. We have thus shown that the second order divergence in the third order DC response can be described by a single integrand function with a single nontrivial energy denominator. This expression makes it clear how different choices of ω_1, ω_2 and ω_3, that is, different setups of the electric field, correspond to different ways of probing the second order divergence in the DC (or quasi-DC) response [53]:

- For $\omega_{123}=\omega_{23}=0$, i.e., $\omega_3=-\omega_2$ and $\omega_1=0$,

$$\frac{1}{\hbar\bar{\omega}_{123}}\frac{1}{\hbar\bar{\omega}_{23}}\to\frac{1}{\hbar^2(3i\eta)(2i\eta)}=\frac{-1}{6\hbar^2}\eta^{-2}. \tag{5.36}$$

  This corresponds to the *jerk* current: $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}(0,-\omega,\omega)$.

- For $\omega_{123}=\delta$, $\omega_{23}=0$, i.e., $\omega_3=-\omega_2=\omega$ and $\omega_1=\delta$, with $\eta\ll\delta\ll\omega$,

$$\frac{1}{\hbar\bar{\omega}_{123}}\frac{1}{\hbar\bar{\omega}_{23}}\to\frac{1}{\hbar^2(\delta)(2i\eta)}=\frac{-i}{2\hbar^2}\eta^{-1}\delta^{-1}. \tag{5.37}$$

  This corresponds to the *dichromatic* setup probe: $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}(\delta,-\omega,\omega)$.

- For $\omega_{123}=\delta_1+\delta_2$, $\omega_{23}=\delta_2$, *e.g.*, $\omega_3=\omega+\delta_2$, $\omega_2=-\omega$ and $\omega_1=\delta_1$, with $\eta\ll\delta_1,\delta_2\ll\omega$,

$$\frac{1}{\hbar\bar{\omega}_{123}}\frac{1}{\hbar\bar{\omega}_{23}}\to\frac{1}{\hbar^2(\delta_1+\delta_2)(\delta_2)}=\frac{1}{\hbar^2}(\delta_1+\delta_2)^{-1}\delta_2^{-1}. \tag{5.38}$$

  This corresponds to the *trichromatic* setup probe: $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}(\delta_1,-\omega,\omega+\delta_2)$.

The interplay between the different frequency parameters — the relaxation rate and the electric field frequencies — plays a determinant role in the study of this divergence. We must emphasize, however, that the coefficient describing the response as a function of optical frequency is always the same, regardless of the specific type of probe being used; this coefficient, we will now show, is expressed in terms of a Dirac delta function, and can thus be considered as a specific Fermi's golden rule contribution to the third order response [10].

Consider the denominator in Eq.(5.35). Now that there are no other (possibly overlapping) features in the integrand, it can be safely expressed in terms of its real and imaginary parts,

$$\frac{1}{\hbar\bar{\omega}_3-\Delta\epsilon_{\mathbf{k}ss'}}\to\frac{\mathrm{P}}{\hbar\omega_3-\Delta\epsilon_{\mathbf{k}ss'}}-i\pi\delta(\hbar\omega_3-\Delta\epsilon_{\mathbf{k}ss'}), \tag{5.39}$$

provided that η is small enough that the rest of the integrand can be deemed constant in a finite energy strip around $\Delta\epsilon_{\mathbf{k}ss'} = \hbar\omega_3$. This means that we can write the second order divergence, $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}$, as,

$$\frac{1}{ie^4}\Gamma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = -\frac{1}{\hbar\bar{\omega}_{123}}\frac{1}{\hbar\bar{\omega}_{23}}\int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s\neq s'}\left(\nabla^\beta\nabla^{\alpha_1}\Delta\epsilon_{\mathbf{k}ss'}\right)\xi^{\alpha_3}_{\mathbf{k}ss'}\xi^{\alpha_2}_{\mathbf{k}s's} \times\Delta f_{\mathbf{k}s's}\Big[\frac{\mathrm{P}}{\hbar\omega_3-\Delta\epsilon_{\mathbf{k}ss'}} - i\pi\delta(\hbar\omega_3-\Delta\epsilon_{\mathbf{k}ss'})\Big], \tag{5.40}$$

which we have not yet symmetrized. After another lengthy calculation, which we again leave to an appendix, one obtains a symmetrized Γ, $\tilde{\Gamma}^{\beta\alpha_1\alpha_2\alpha_3}$, that reads as follows,

$$-\frac{3\hbar^2}{\pi e^4}\tilde{\Gamma}^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \frac{1}{\bar{\omega}_{123}}\frac{1}{\bar{\omega}_{23}}\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega_3) + \frac{1}{\bar{\omega}_{123}}\frac{1}{\bar{\omega}_{13}}\iota^{\beta\alpha_2\alpha_3\alpha_1}(\omega_1) + \frac{1}{\bar{\omega}_{123}}\frac{1}{\bar{\omega}_{12}}\iota^{\beta\alpha_3\alpha_1\alpha_2}(\omega_2) \tag{5.41}$$

for a coefficient, $\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega_3)$, that is expressed in terms of a single integral involving a Dirac delta function,[13]

$$\boxed{\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega) = \int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s\neq s'}\left(\nabla^\beta\nabla^{\alpha_1}\Delta\epsilon_{\mathbf{k}ss'}\right)\left\{\xi^{\alpha_2}_{\mathbf{k}ss'},\xi^{\alpha_3}_{\mathbf{k}s's}\right\}\Delta f_{\mathbf{k}s's}\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'})}. \tag{5.42}$$

The parallel between this coefficient and that of the injection current is discussed in Appendix B.4. Finally, we end this section on the conductivity-based derivation of the second order divergence by adding some comments on the subject of the numerical pre-factors in the expression for $\tilde{\Gamma}^{\beta\alpha_1\alpha_2\alpha_3}$, Eq.(5.41). For the jerk current and the dichromatic setup probe to the divergence, different phenomenologies — i.e., different parametrizations of the relaxation rate — are associated with different numerical factors. Consider, as an example, the case where the relaxation rate is introduced via equations of motion. The equations describing the dynamics of the density matrix in the length gauge, Eq.(3.92), must now carry an additional term, $-\eta(\rho_{\mathbf{k}ss'}(t)-\rho^{(0)}_{\mathbf{k}ss'})$, which induces the system to relax to the Fermi-Dirac distribution function. As a result, the imaginary parts of the frequencies are written as $\bar{\omega}_1 = \omega_1 + i\eta$, $\bar{\omega}_{12} = \omega_{12} + i\eta$, and $\bar{\omega}_{123} = \omega_{123} + i\eta$, meaning that the factors of 1/6 and 1/2 in Eq.(5.36) and Eq.(5.37) are absent. This study of the second order divergence in the third order DC response requires us to consider two different phenomenologies: the adiabatic switching, which, as we have mentioned in Chapter 4, is the most appropriate choice of relaxation rates for numerical calculations in the velocity gauge; and a parametrization in which the static component of the field has no associated relaxation — so as to be consistent with our semiclassical description of the jerk current. In this case, we have, for one of the contributions in Eq.(5.41), $\bar{\omega}_2$, $\bar{\omega}_3 = \pm\omega + i\eta$ (for $\omega_{23} = 0$) and $\bar{\omega}_1 = \omega_1 = 0$. The former phenomenology is used in the following section, where we show that the results from this analytical expression agree with

[13] Note that for the second order divergence, this coefficient will be a function of the main optical frequency, ω.

numerical results for the full third order conductivity, $\sigma^{\beta\alpha_1\alpha_2\alpha_3}$.

5.3.1 The Second Order Divergence in the DC Third Order Response of Gapped Graphene

As noted in the introduction to this chapter, the jerk current was originally presented in ref.[109], where its analytical expression was derived by two different procedures. The first procedure involves studying the highest order divergence in the third order response function; the second procedure, on the other hand, employs a semiclassical argument to determine the relevant contributions to the electronic populations, $n_s(\mathbf{k}, t)$, at this order in the field, and from there an expression for the jerk current. Now, since this is also true for our description — we present our phenomenological argument in the following subsection — the results of which, nonetheless, differ from those of Fregoso *et al,* it seems appropriate to add some further proof to these analytical studies. This subsection is thus dedicated to numerical calculations of the full third order response, which are performed in the velocity gauge. This third set of evidence, as we will presently see, shows that our analytical expression is in fact correct, and backs our interpretation of this effect.

We begin by considering the description used in Chapter 4 in the calculation of the nonlinear optical response of the gapped and plain graphene monolayers (to a monochromatic field). It is not necessary, at this point, to add another model or a different material to serve as the basis of this calculation; we are only interested in ensuring that Eq.(5.41) is the appropriate description of this second order divergence, and this can be adequately shown using the aforementioned tight-binding Hamiltonian, for the numerical calculations, and its $\mathbf{K}$ point expansion, the gapped Dirac Hamiltonian, for the analytical ones. Unlike the injection current, which is suppressed in gapped graphene due to its point group symmetries, the jerk current should be present even in the case of fields linearly polarised along one of the natural directions, as we can see from the expression for the ι coefficient, Eq.(5.42). For numerical calculations, we therefore recover the nearest-neighbours tight-binding model for graphene, with parameters $\Delta = 300\,\text{meV}$ and $t = 3\,\text{eV}$, Eq.(4.39). For analytical calculations, we instead consider an expansion of this tight-binding Hamiltonian around the band minima, $\mathbf{k} = \mathbf{K}(\mathbf{K}') + \mathbf{q}$, Figure 5.1, which renders us, to the lowest order, the familiar gapped Dirac Hamiltonian,

$$H_\lambda(\mathbf{q}) = \begin{bmatrix} \Delta/2 & \hbar v_F(\lambda q_x - i q_y) \\ \hbar v_F(\lambda q_x + i q_y) & -\Delta/2 \end{bmatrix}, \tag{5.43}$$

where $\lambda = \pm 1$, for $\mathbf{K} = 4\pi/3\sqrt{3}a_0\,\hat{k}_x$ and $\mathbf{K}' = -4\pi/3\sqrt{3}a_0\,\hat{k}_x$, and $\hbar v_F = 3ta_0/2$, for $a_0 = 1.42\text{Å}$, the distance between two neighbouring atoms. This description, while it fails to describe appropriately the second order response of the gapped graphene monolayer, can be used to describe its third order response, as this is unaffected by the redundant centrosymmetricity of this Hamiltonian. The corresponding analytical calculations are also much simpler than those for the second order response; these can be done using a straightforward change of variables, instead of having to invert

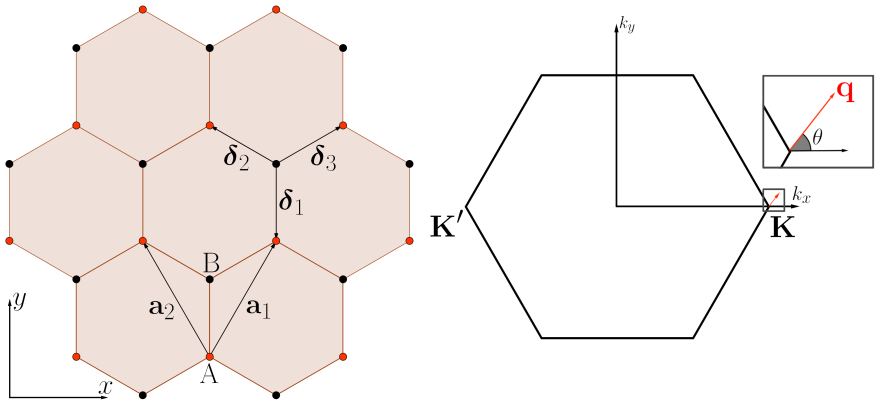


Figure 5.1: The honeycomb lattice of the gapped graphene monolayer and its associated FBZ. [Left] The lattice structure of the gapped graphene monolayer. Note that the A and B sites are not equivalent and have different on-site energies, $\epsilon_{\mathrm{A}} - \epsilon_{\mathrm{B}} = \Delta$. [Right] The first Brillouin zone associated with the honeycomb lattice. The labelled edges of the FBZ are the $\mathbf{K}$ and $\mathbf{K}'$ points (the other four unlabelled vertexes are their equivalents), $\mathbf{K}$, $\mathbf{K}' = (\pm 4\pi/3\sqrt{3}a_0, 0)$, which correspond to band minima, Eq.(4.49). The inset represents the small region around $\mathbf{K}$ where the expansion in $\mathbf{q}$ of the tight-binding Hamiltonian is performed.

a series.

For our numerical calculations, as we have already noted, we use the the tight-binding Hamiltonian of Eq.(4.39). The object that is to be computed is not, however, the coefficient that we determined in the previous section; it is, instead, the full third order conductivity, $\tilde{\sigma}^{\beta\alpha_1\alpha_2\alpha_3}$. In this case, we must consider that the relaxation rate η is necessarily finite for all three probes, and that the frequency offsets, δ_1 and δ_2, are nonzero for two of these three: δ_1 for the dichromatic field setup probe; δ_1 and δ_2 for the trichromatic one. The result that follow from a numerical calculation is, therefore, a numerical ι coefficient, $\iota^{\beta\alpha_1\alpha_2\alpha_3}_{\mathrm{num.}}(\omega_3)$, which is obtained from the full third order conductivity by inverting the relation between the two objects, Eq.(5.29),

$$\iota^{\beta\alpha_1\alpha_2\alpha_3}_{\mathrm{num.}}(\omega_3) \approx -\frac{3\hbar^3}{e^4}\bar{\omega}_{123}\bar{\omega}_{23}\,\tilde{\sigma}^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3), \tag{5.44}$$

and replacing $\bar{\omega}_{123}$, $\bar{\omega}_{12}$ by the expressions given in Eqs.(5.36) to (5.38); note how this relation is not, strictly, an equality, because the conductivity carries terms other than the highest order divergence. In addition, we must emphasize that these $\tilde{\sigma}^{\beta\alpha_1\alpha_2\alpha_3}$ are computed in the *velocity-gauge* description, not the length-gauge one which we have used in the previous subsection. The equivalence between these two is, of course, ensured by gauge invariance. In order to use the method of calculation outlined in the previous chapter, we once again introduce the relaxation rate by way of adiabatic switching.

The results for the three different probes to the divergence, Eqs.(5.36)–(5.38), are presented in the three plots of Figure 5.2. These three plots show that the analytical results of Eq.(5.42) and the

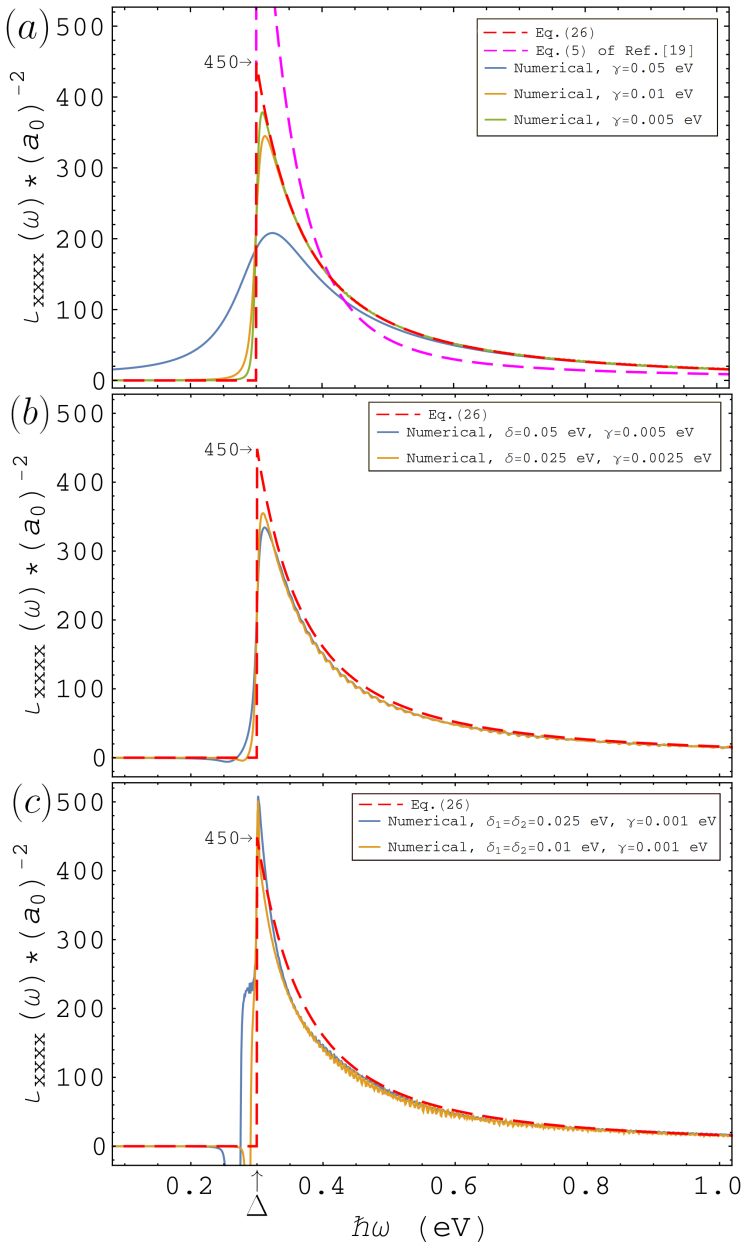


Figure 5.2: A comparison between the analytical result that follows from Eq.(5.42) — represented by the red dashed curve — and numerical results from a velocity gauge calculation of the conductivity [3, 5] — represented by the full curves — in gapped graphene, $\Delta = 300\,\text{meV}$ and $t = 3\,\text{eV}$, for the response along the zig-zag direction, $\beta = \alpha_i = x$, $i = 1, 2, 3$. The (a), (b) and (c) plots represent the second order divergence in the jerk current and the dichromatic and trichromatic field setup probes, respectively. In (a), we have also represented the analytical solution from Fregoso *et al.* — represented by the magenta dashed curve. This, we note, does not match with our numerical result. Note also that value of ι at the gap, given by Eq.(5.45), reads as $\iota_{xxxx}(\Delta)a_0^{-2} = 450$ for the aforementioned parameters. Note that these plots were extracted from ref.[8], where the notation used is different from that of this thesis: γ is here used in place of the relaxation parameter, η; the ι coefficient is written with subscripts instead of superscripts.

numerical results of a velocity gauge calculation of the *full* conductivity are, for frequencies above the gap, in good agreement, which corroborates our description of the second order divergence. These results warrant two additional comments. First, we note that there are discrepancies between the analytical and numerical results for frequencies below the gap. These follow from the fact that our numerical calculations carry contributions other than $\tilde{\Gamma}^{\beta\alpha_1\alpha_2\alpha_3}$ — the terms represented by the ellipsis in Eq.(5.29) — which necessarily contribute to the response. That the differences between results are more noticeable in plots (b) and (c) of Figure 5.2 is due to the existence, in the full conductivity, of resonances at frequencies $\omega \pm \delta$ and $\omega \pm \delta$, $\omega + 2\delta$. Secondly, the analytical calculation gives us an expression for $\iota^{xxxx}(\Delta)$ that reads as,

$$\frac{1}{a_0^2}\iota^{xxxx}(\Delta) = \frac{9t^2}{2\Delta^2}. \tag{5.45}$$

The second order divergence in gapped graphene should be more pronounced when the band gap is smaller; this is consistent with the results of ref.[53] and similar to what we obtained for the second harmonic generation and the optical rectification effects at the gap in this material. We can also estimate the amplitude of the jerk current along one of its natural directions, so as to give us an idea of how pronounced this effect should be. For the aforementioned values of the hopping parameter and band gap, and for $\hbar\omega \sim \Delta$, $\tau = 1/\eta \sim 100\,\text{fs}$, $E^x_\omega = 10^7\,\text{V/m}$ and $E^x_0 = 10^6\,\text{V/m}$, the jerk current along the zig-zag direction in gapped graphene, J^x_{jerk}, is estimated to be $J^x_{\text{jerk}} \approx 12\,\text{A/m}$. This, as was noted in the work of Fregoso *et al,* should be within experimental reach.

5.3.2 The Jerk Current: A Semiclassical Description and its Physical Interpretation

Having derived an analytical expression for the second order divergence — and having shown that it is consistent with a numerical calculation of the full third order conductivity — we can now move on to our semiclassical description of the jerk current. This description, despite a certain lack of technical rigour, helps us understand the physics behind this effect in simple terms, with which the reader will probably be familiar. Therefore, it works best as a complement to the rigorous, albeit more complicated, conductivity description that we have presented in this section; this somewhat parallel development of different descriptions can give us a fuller understanding of this second order divergence in particular, and of nonlinear optical responses of crystals in general.

As we have seen in Chapter 3, a crystal's current response, $J^\beta(t)$, is expressed in terms of the velocity matrix elements, $v^\beta_{\mathbf{k}s's}$, and the matrix elements of the density matrix, $\rho_{\mathbf{k}ss'}(t)$,

$$J^\beta(t) = -|e| \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{s's} v^\beta_{\mathbf{k}s's}\rho_{\mathbf{k}ss'}(t).$$

The matrix elements of the density matrix encode the dynamics (at a quantum level) of a system of crystal electrons under the effect of an external electric field: the diagonal elements of the matrix

correspond to the population density of a given band, s, at a given $\mathbf{k}$ and instant t, $\rho_{\mathbf{k}ss}(t)$, while the nondiagonal elements, $\rho_{\mathbf{k}s's}(t)$, for $s' \neq s$, describe vertical transitions between different bands. The current response is not, in general, complete if we neglect one of these types of contributions, but there are certain effects, *e.g.*, the injection and shift currents, that can be more or less appropriately described when we consider the intraband elements, the populations.[14] The jerk current is another of these effects. We have seen in Subsection 5.3 that the terms responsible for the second order divergence in the third order DC conductivity follow from the intraband terms of the density matrix, which indicates that the jerk current — the finite second time derivative of $J^{\beta}(t)$ — can be described by a physical argument that is similar to those of injection and shift currents; indeed, we will now see that the populations, $n_s(\mathbf{k},t)$, can be determined by separating the carriers' dynamics into two processes: the pumping of electrons into the conduction band, mediated by the optical component of the electric field; their subsequent acceleration by the static component of the electric field. The argument is as follows.

The electric current density due to electrons in a crystal, at an instant t, reads — when we neglect the interband matrix elements — as the integral over the first Brillouin zone (FBZ) of all different $\mathbf{k}$-state contributions (in a given band s) to the current density, $-|e|v_s^{\beta}(\mathbf{k})$, each of which weighted by its corresponding population, $n_s(\mathbf{k},t)$,

$$J^{\beta}_{\text{pheno.}}(t) = -\frac{|e|}{\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_s n_s(\mathbf{k},t)\big(\nabla^{\beta}_{\mathbf{k}}\epsilon_{\mathbf{k}s}\big), \tag{5.46}$$

summed over all bands. To determine the dynamics of $n_s(\mathbf{k},t)$, we must consider the role played by each component of the electric field. First, the optical component of the field, $E^{\alpha_1}(\omega)$, pumps electrons from a valence into a conduction band at a rate $\partial|\gamma_{cv}|^2(\mathbf{k},t)/\partial t$, which is described, in first approximation, by Fermi's golden rule, or equivalently by Eq.(5.10),

$$\frac{\partial|\gamma_{cv}|^2(\mathbf{k},t)}{\partial t} = \frac{2\pi e^2}{\hbar}\text{Re}\big[E^{\alpha_1}(\omega)E^{\alpha_2}(-\omega)\xi^{\alpha_1}_{\mathbf{k}vc}\,\xi^{\alpha_2}_{\mathbf{k}cv}\big]\,\Delta f_{\mathbf{k}vc}\,\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}cv}). \tag{5.47}$$

Secondly, the static component, $E^{\alpha_3}(0)$, accelerates electrons across the FBZ at the rate, $\dot{\mathbf{k}}(t) = -|e|\,E^{\alpha_3}(0)\hbar^{-1}$, in much the same way as it would accelerate a classical particle with momentum $\mathbf{p}=\hbar\mathbf{k}$ and charge $-|e|$. We note, however, that despite the apparent simplicity of this description, this motion's effect on the balance equations is not altogether trivial and could easily lead one to an incorrect result. Consider the electronic population density in a given conduction band, $n_c(\mathbf{k},t)$, at a given point of the FBZ, $\mathbf{k}$, at an instant t, Figure 5.3. Let us also, for the sake of simplicitly, forget (momentarily) about the role that pumping plays in this population's dynamics. As indicated in that figure, an electric field that points to the left ($-\hat{x}$ direction) accelerates electrons in the

[14] There is an important caveat here. As we have noted in Subsection 5.2.2, the density-matrix calculation of the shift current *does* involve the interband matrix elements. This helps explain why that physical argument for the shift current is not altogether rigorous.

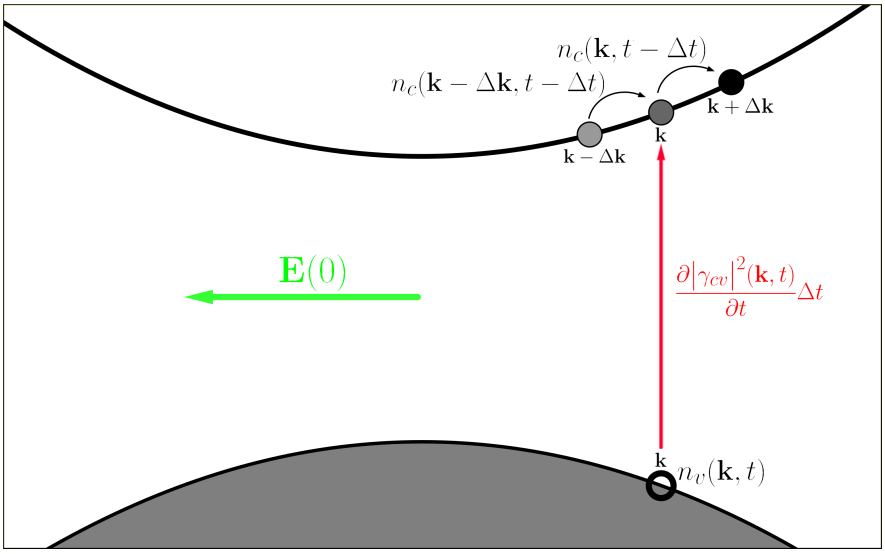


Figure 5.3: A schematic representation of two processes contributing to the change of the population density in the conduction band, $n_c(\mathbf{k}, t)$, for fixed $\mathbf{k}$. The optical component of the electric field pumps electrons into a conduction band at a rate $\partial\left|\gamma_{cv}\right|^2(\mathbf{k}, t)/\partial t$, while the static component drags them $\Delta\mathbf{k} = -\left|e\right|\hbar^{-1}\mathbf{E}(0)\Delta t$ over a interval Δt. The second term in Eq.(5.49) describes the difference between the number of electrons that are moving *into* that $\mathbf{k}$ — which at a time $t - \Delta t$ were in $\mathbf{k} - \Delta\mathbf{k}$ — and those that are moving *out* of it, per interval Δt. Figure from ref.[9].

opposite direction, that is, to the right ($\hat{x}$ direction); for a time interval Δt, these electrons travel $\Delta\mathbf{k} = -\left|e\right| E^{\alpha_3}(0)\hbar^{-1}\Delta t$ to the right in the first Brillouin zone. The electronic population density at a given point of the FBZ, $\mathbf{k}$, and an instant t, will thus *increase* with the population density at $\mathbf{k} - \Delta\mathbf{k}$, and an earlier instant $t - \Delta t$ — i.e., via electrons that come from the left (into $\mathbf{k}$) — and *decrease* with that at $\mathbf{k}$ and that same earlier instant — i.e., via electrons that go to the right (out of $\mathbf{k}$). In somewhat quantitative terms, this can be expressed as,

$$n_{c(v)}(\mathbf{k}, t) \propto \left(n_{c(v)}(\mathbf{k} - \Delta\mathbf{k}, t - \Delta t) - n_{c(v)}(\mathbf{k}, t - \Delta t)\right), \tag{5.48}$$

which means that the dragging term in the balance equation will carry the same sign as that of the electric field, which is, of course, *opposite* to that of the classical acceleration. After some careful manipulations, we can write the rate of change of the population density, for a fixed $\mathbf{k}$-point and at an instant t, in the conduction (and valence) bands as,

$$\frac{\partial n_{c(v)}(\mathbf{k}, t)}{\partial t} = \underbrace{\pm\frac{\partial\left|\gamma_{cv}\right|^2(\mathbf{k}, t)}{\partial t}}_{\text{pumping}} \underbrace{+\frac{|e|}{\hbar}E_0^{\alpha_3}\left(\nabla_{\mathbf{k}}^{\alpha_3} n_{c(v)}(\mathbf{k}, t)\right)}_{\text{motion along the FBZ}}. \tag{5.49}$$

An expansion of $n_{c(v)}$ in its first two orders of the static field, $n_{c(v)} = n^{(0)}_{c(v)} + n^{(1)}_{c(v)} + (...)$, gives us,

$$\frac{\partial n^{(0)}_{c(v)}(\mathbf{k},t)}{\partial t} = \pm \frac{\partial \left|\gamma_{cv}\right|^2(\mathbf{k},t)}{\partial t}, \tag{5.50}$$

$$\frac{\partial n^{(1)}_{c(v)}(\mathbf{k},t)}{\partial t} = \frac{|e|}{\hbar} E^{\alpha_3}(0)\big(\nabla^{\alpha_3}_{\mathbf{k}} n^{(0)}_{c(v)}(\mathbf{k},t)\big), \tag{5.51}$$

so that, by taking the the second time derivative of Eq.(5.51), we obtain,[15]

$$\frac{\partial^2 n^{(1)}_{c(v)}(\mathbf{k},t)}{\partial t^2} = \pm \frac{|e|}{\hbar} E^{\alpha_3}(0) \left(\nabla^{\alpha_3}_{\mathbf{k}} \frac{\partial \left|\gamma_{cv}\right|^2(\mathbf{k},t)}{\partial t}\right). \tag{5.52}$$

Neglecting contributions that are quadratic in the static component of the field — the jerk current is a third order effect — the finite contribution to the second time derivative Eq.(5.46) reads as,

$$\partial_t^2 J^{\beta}_{\text{jerk}}(t) = -\frac{|e|}{\hbar} \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_s \frac{\partial^2 n^{(1)}_s(\mathbf{k},t)}{\partial t^2} \big(\nabla^{\beta}_{\mathbf{k}} \epsilon_{\mathbf{k}s}\big), \tag{5.53}$$

$$= \frac{e^2}{\hbar^2} E^{\alpha_3}(0) \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{cv} \frac{\partial \left|\gamma_{cv}\right|^2(\mathbf{k},t)}{\partial t} \big(\nabla^{\alpha_3}_{\mathbf{k}} \nabla^{\beta}_{\mathbf{k}} \Delta\epsilon_{\mathbf{k}cv}\big). \tag{5.54}$$

We can now replace the transition rate given by Fermi's Golden rule, Eq.(5.47), and, by imposing time-reversal symmetry, we obtain,

$$\partial_t^2 J^{\beta}_{\text{jerk}}(t) = \frac{2\pi e^4}{\hbar^3} \text{Re}\big[E^{\alpha_1}(\omega) E^{\alpha_2}(-\omega)\big] E^{\alpha_3}(0) \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{cv} \big(\nabla^{\beta}_{\mathbf{k}} \nabla^{\alpha_3}_{\mathbf{k}} \Delta\epsilon_{\mathbf{k}cv}\big) \big\{\xi^{\alpha_1}_{\mathbf{k}cv}, \xi^{\alpha_2}_{\mathbf{k}vc}\big\} \, \Delta f_{\mathbf{k}vc} \, \delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}cv}). \tag{5.55}$$

This, as we will now show, is exactly the same result as that which follows from the fully-quantum treatment we presented in this section. First, we need to choose the appropriate parametrization of the relaxation rate. In the physical argument we have just presented, the static component of the field is not adiabatically switched, which means that $\bar{\omega}_3 = \omega_3 = 0$. The optical component,

[15] There are two additional points we can make on the jerk current, which again follow from a comparison between this effect and the second order injection and shift currents. First, note that jerk current is, like the injection current, the result of an imbalance between populations on opposite sides of the FBZ: time-reversal symmetry dictates that only $2\Delta n_s(\mathbf{k},t) = n_s(\mathbf{k},t) - n_s(-\mathbf{k},t)$ can contribute to Eq.(5.46). However, unlike the injection current (and similarly to the shift current), this population imbalance is proportional to the average *injected* population on both sides of the FBZ, $\tau\dot{n}(\mathbf{k},t)$, Eq.(5.19). It follows from Eq.(5.52) that,

$$\frac{\partial^2 \Delta n^{(1)}_c(\mathbf{k},t)}{\partial t^2} = \frac{1}{2}\frac{|e|}{\hbar} E^{\alpha_3}(0) \left[\left(\nabla^{\alpha_3}_{\mathbf{k}} \frac{\partial \left|\gamma_{cv}\right|^2(\mathbf{k},t)}{\partial t}\right) - \left(\nabla^{\alpha_3}_{-\mathbf{k}} \frac{\partial \left|\gamma_{cv}\right|^2(-\mathbf{k},t)}{\partial t}\right)\right] = \frac{|e|}{\hbar} E^{\alpha_3}(0) \left(\nabla^{\alpha_3}_{\mathbf{k}} \dot{n}_{\text{ave.}}(\mathbf{k},t)\right),$$

since $\partial \left|\gamma_{cv}\right|^2(\mathbf{k},t)/\partial t = \dot{n}(\mathbf{k},t)$, for $\dot{n}(\mathbf{k},t)$, the injected population we derived in Section 5.2, Eq.(5.10). We can, therefore, interpret the jerk current both as a subtracting effect — like the injection current — and as a cumulative effect — like the shift current.

however, does carry a factor of $i\eta$ — $\bar{\omega}_1, \bar{\omega}_2 = \pm\omega + i\eta$ — that is implicitly present in the derivation of the transition rate, $\partial\left|\gamma_{cv}\right|^2(\mathbf{k}, t)/\partial t$, and is associated with the adiabatic switching of the optical component of the field. Therefore, our fully-quantum treatment is analogous to the semiclassical description of the jerk current when our parametrization of the relaxation rate reads as,

$$\bar{\omega}_1 = \omega_1 + i\eta, \tag{5.56}$$

$$\bar{\omega}_2 = \omega_2 + i\eta, \tag{5.57}$$

$$\bar{\omega}_3 = \omega_3, \tag{5.58}$$

for the real frequencies we specified. It then follows from the expression for the time-domain third order current density, Eq.(4.7), that the jerk current can be written as,

$$\partial_t^2 J^\beta(t) = \int \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\frac{d\omega_3}{2\pi}(-i\bar{\omega}_{123})^2\,\tilde{\Gamma}^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3)E^{\alpha_1}(\omega_1)E^{\alpha_2}(\omega_2)E^{\alpha_3}(\omega_3)e^{-i\bar{\omega}_{123}t}. \tag{5.59}$$

Replacing the expression for the electric field,

$$E^\lambda(\omega_1) = \delta^{\lambda\alpha_1}E^{\alpha_1}(\omega)\delta(\omega_1-\omega) + \delta^{\lambda\alpha_2}E^{\alpha_2}(-\omega)\delta(\omega_1+\omega) + \delta^{\lambda\alpha_3}E^{\alpha_3}(0)\delta(\omega_1), \tag{5.60}$$

and the expression for $\tilde{\Gamma}$, Eq.(5.41), in the expression for $\partial_t^2 J^\beta(t)$, gives us, after some manipulations,

$$\begin{aligned}\partial_t^2 J^\beta(t) = \frac{\pi e^4}{3\hbar^3}\big[&\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega_3)E^{\alpha_1}(0)E^{\alpha_2}(-\omega)E^{\alpha_3}(\omega)\\ &+ \iota^{\beta\alpha_2\alpha_3\alpha_1}(\omega)E^{\alpha_1}(\omega)E^{\alpha_2}(0)E^{\alpha_3}(-\omega)\\ &+ \iota^{\beta\alpha_3\alpha_1\alpha_2}(\omega)E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)E^{\alpha_3}(0) + \text{c.c.}\big].\end{aligned} \tag{5.61}$$

Exchanging the Cartesian indexes in the first and second terms finally gives us,

$$\partial_t^2 J^\beta(t) = \frac{2\pi e^4}{\hbar^3}\text{Re}\big[E^{\alpha_1}(\omega)E^{\alpha_2}(-\omega)\big]E^{\alpha_3}(0)\iota^{\beta\alpha_3\alpha_1\alpha_2}(\omega). \tag{5.62}$$

This is the same result as that in Eq.(5.55).

5.4 Energy Absorption in a Cold Semiconductor

The effects we considered in the last two sections, and which we described as divergences in the DC response of a cold semiconductor, are connected with processes of energy absorption by its electrons: these are either pumped from a valence band into a conduction band, as in the case of the injection and shift currents (both of which are second order effects), or they are pumped from a valence band into a conduction band and then subsequently accelerated by the static component of the electric field, as in the case of the jerk current (a third order effect). In either case, energy absorption in a

cold semiconductor is predicated on the existence of an electric field whose frequency is above the electronic band gap,[16] and its absence must, therefore, mean that there can be no energy absorption. This, as we will now show, implies that the response functions must possess (for certain frequencies) full permutation symmetry [12]. The knowledge of this, combined with the expressions we derived for the divergences in the second order DC response, will enable us to derive a general relation between different real parts of the second order conductivity, which is the subject of the next, and final, section of this chapter.

We begin by recovering the expression for the rate at which energy is transferred to (and from) the cold semiconductor, Eq.(5.4),

$$\dot{\epsilon}(t) = -\big|e\big| \int \frac{d^d\mathbf{k}}{(2\pi)^d} \mathbf{E}(t) \cdot \mathrm{Tr}\left[\mathbf{v}\, \rho(t)\right] = \mathbf{E}(t) \cdot \mathbf{J}(t). \tag{5.63}$$

We are now interested in studying the case where this system does not absorb any energy from the electric field. This means that the total energy transferred to the cold semiconductor, which reads simply as the integral (from $t = -\infty$ to $t = \infty$) of the quantity in the left-hand side of Eq.(5.63),

$$\epsilon(\infty) - \epsilon(-\infty),$$

must be exactly zero, since the cold semiconductor must retain the energy that it had before it was subjected to the external electric field. It then follows from the relation between $\dot{\epsilon}(t)$ and $\mathbf{J}(t)$ that,

$$\int_{-\infty}^{\infty} dt \mathbf{E}(t) \cdot \mathbf{J}(t) = 0, \tag{5.64}$$

or, by writing the field and the current in terms of their Fourier transforms,

$$\int_{-\infty}^{\infty} d\omega \mathbf{E}(-\omega) \cdot \mathbf{J}(\omega) = 0. \tag{5.65}$$

Let us first consider the simplest possible case, that of the linear response. The current then reads as, Eq.(2.8),

$$J^\beta(\omega) = \sigma^{\beta\alpha}(\omega) E^\alpha(\omega), \tag{5.66}$$

so that the integral in Eq.(5.65) is expressed as,

$$\int_{-\infty}^{\infty} d\omega \sigma^{\beta\alpha}(\omega) E^\beta(-\omega) E^\alpha(\omega) = \frac{1}{2} \int_{-\infty}^{\infty} d\omega \left[\sigma^{\beta\alpha}(\omega) + \sigma^{\alpha\beta}(-\omega)\right] E^\beta(-\omega) E^\alpha(\omega). \tag{5.67}$$

This vanishes if and only if the conductivity satisfies the following property,

$$\sigma^{\beta\alpha}(\omega) = -\sigma^{\alpha\beta}(-\omega) \tag{5.68}$$

[16] Or, in the case of an electric field that is not monochromatic, at least one of its frequencies.

which, for its diagonal elements reads as,

$$\sigma^{\alpha\alpha}(\omega) = -\sigma^{\alpha\alpha}(-\omega) = -\left[\sigma^{\alpha\alpha}(\omega)\right]^*, \tag{5.69}$$

i.e., the conductivity must be strictly imaginary for photon energies below the electronic band gap, $\hbar\omega < \Delta$, exactly as we have mentioned in Appendix B.1. Note that we have not considered the actual form of the conductivity. We have simply stated that there is a certain frequency threshold for energy absorption, below which the energy of the cold semiconductor must remain constant, $\epsilon(\infty) = \epsilon(-\infty)$. Let us move on to the second order response. In this case, the electric current reads as, Eq.(4.2),

$$J^{\beta}(\omega) = \int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\, \sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)E^{\alpha_1}(\omega_1)E^{\alpha_2}(\omega_2)\,(2\pi)\delta(\omega-\omega_1-\omega_2),$$

and the integral in Eq.(5.65) is therefore expressed in terms of the second order conductivity,

$$\int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\, \sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)E^{\beta}(-\omega_{12})E^{\alpha_1}(\omega_1)E^{\alpha_2}(\omega_2). \tag{5.70}$$

As before, we can use changes of frequency variables and Cartesian indexes to manipulate the integrand. We can thus write Eq.(5.70) as,

$$\frac{1}{3}\int_{-\infty}^{\infty} \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\left[\sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)+\sigma^{\alpha_1\beta\alpha_2}(-\omega_{12},\omega_2)+\sigma^{\alpha_2\alpha_1\beta}(\omega_1,-\omega_{12})\right]E^{\beta}(-\omega_{12})E^{\alpha_1}(\omega_1)E^{\alpha_2}(\omega_2), \tag{5.71}$$

which means that for this object to be zero, the three conductivites in the integrand must cancel each other out,

$$\sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)+\sigma^{\alpha_1\beta\alpha_2}(-\omega_{12},\omega_2)+\sigma^{\alpha_2\alpha_1\beta}(\omega_1,-\omega_{12}) = 0. \tag{5.72}$$

Note that we can alternatively express this in terms of the susceptibility, using the relation derived in Chapter 2, Eq.(2.28),

$$\omega_{12}\chi^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)-\omega_1\chi^{\alpha_1\beta\alpha_2}(-\omega_{12},\omega_2)-\omega_2\chi^{\alpha_2\alpha_1\beta}(\omega_1,-\omega_{12}) = 0. \tag{5.73}$$

For this to hold true for all values of ω_1 and ω_2, the susceptibilities must possess intrinsic permutation symmetry [16, 171],

$$\chi^{\beta\alpha_1\alpha_2}(-\omega_{12};\omega_1,\omega_2) = \chi^{\alpha_1\beta\alpha_2}(\omega_1;-\omega_{12},\omega_2) = \chi^{\alpha_2\alpha_1\beta}(\omega_2;\omega_1,-\omega_{12}), \tag{5.74}$$

that is, the total output frequency and index can be (simultaneously) exchanged with those of one of the electric fields. The extra piece of notation we have just introduced, wherein the output frequency is made evident, is not something that we will use throughout the rest of this work; we

use it momentarily so as to make this property clearer. As for the frequencies under consideration, remember that we are imposing that the energy of the cold semiconductor remains constant, meaning that all frequencies — ω_1, ω_2 and ω_{12} — correspond to photon energies below the electronic band gap. When this is true, $\chi^{\beta\alpha_1\alpha_2}$ is strictly real and $\sigma^{\beta\alpha_1\alpha_2}$ is strictly imaginary.[17] If we now return to the conductivities, by again using Eq.(2.28), we obtain,

$$\mathrm{Im}\big[\sigma^{\beta\alpha_1\alpha_2}(-\omega_{12};\omega_1,\omega_2)\big] = -\frac{\omega_{12}}{\omega_1}\mathrm{Im}\big[\sigma^{\alpha_1\beta\alpha_2}(\omega_1;-\omega_{12},\omega_2)\big], \tag{5.75}$$

$$\mathrm{Im}\big[\sigma^{\beta\alpha_1\alpha_2}(-\omega_{12};\omega_1,\omega_2)\big] = -\frac{\omega_{12}}{\omega_2}\mathrm{Im}\big[\sigma^{\alpha_2\alpha_1\beta}(\omega_2;\omega_1,-\omega_{12})\big], \tag{5.76}$$

for $\hbar\omega_1$, $\hbar\omega_2$, $\hbar\omega_{12} < \Delta$. The absence of energy absorption below the gap implies that the imaginary part of the conductivity tensor (times some additional frequency factors) possesses, for frequencies below the gap, full permutation symmetry.[18] Furthermore, this property does in fact hold for frequencies outside this domain. One can show, using the general expressions for the response of the two-band model [10], that the imaginary part of the conductivity of the two-band model preserves this property, even for photon energies *above* the gap, when we consider only its regular portion — i.e., that which is not associated with the divergences in the second order response [172].

Finally, let us consider these divergences. As we have mentioned in Subsection 5.3, the shift current — while it appears in the susceptibility as a divergent contribution — corresponds to a finite, well-behaved, contribution to the real part of $\sigma^{\beta\alpha_1\alpha_2}(\omega,-\omega)$, Eq.(5.23). The injection current does, however, correspond to a divergence in the conductivity and contributes to the response with a finite $\dot{J}^{\beta}_{\mathrm{inj.}}(t)$, Eq.(5.15). We can now return to Eq.(5.16), to derive the exact form of the divergent terms in the conductivity. Defining $I_1^{\alpha_1\alpha_2\beta}(\omega)$, i.e., the injection-current coefficient,

$$I_1^{\alpha_1\alpha_2\beta}(\omega) = \int \frac{d^d\mathbf{k}}{(2\pi)^d}\frac{1}{2}\big[\xi^{\alpha_1}_{\mathbf{k}vc},\xi^{\alpha_2}_{\mathbf{k}cv}\big]\big(\nabla^{\beta}_{\mathbf{k}}\Delta\epsilon_{\mathbf{k}cv}\big)\Delta f_{\mathbf{k}vc}\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}cv}), \tag{5.77}$$

we can show that $\sigma^{\beta\alpha_1\alpha_2}_{\mathrm{div.}}(\omega_1,\omega_2)$ must be written as [10],

$$\frac{2\hbar^2}{i\pi|e|^3}\sigma^{\beta\alpha_1\alpha_2}_{\mathrm{div.}}(\omega_1,\omega_2) = \frac{1}{\bar{\omega}_{12}}I_1^{\alpha_1\alpha_2\beta}(\omega_1)+\frac{1}{\bar{\omega}_{12}}I_1^{\alpha_2\alpha_1\beta}(\omega_2). \tag{5.78}$$

This ensures that Eq.(5.16) is satisfied; as before, $\bar{\omega}_{12} = \omega_{12} + 2i\eta$. Note that this portion of the conductivity possesses intrinsic permutation symmetry, Eq.(2.30), and that, as was noted in ref.[10], it is strictly real when we exclude the frequencies in (or around) the line associated with the injection current, $\omega_2 = -\omega_1$. In addition, note that the injection-current coefficient satisfies the following

[17] One can show this either by means of an energy-balance argument or by direct inspection of the expressions [10]. Note that this is only true in the case of systems that are time-reversal symmetric.

[18] For simplicity's sake, we will treat the properties in Eqs.(5.75) and (5.76) as if they were a statement of full permutation symmetry, even though this is somewhat of a misnomer. The property associated with full permutation symmetry is stated in Eq.(5.74).

properties,

$$I_1^{\alpha_1\alpha_2\beta}(\omega) = -\left(I_1^{\alpha_1\alpha_2\beta}(\omega)\right)^*, \tag{5.79}$$

$$I_1^{\alpha_1\alpha_2\beta}(\omega) = -\, I_1^{\alpha_1\alpha_2\beta}(-\omega), \tag{5.80}$$

$$I_1^{\alpha_1\alpha_2\beta}(\omega) = -\, I_1^{\alpha_2\alpha_1\beta}(\omega). \tag{5.81}$$

5.5 A New Relation between Real Parts of the Second Order Conductivity Tensor of a Cold Semiconductor

In the present, and final, section of this chapter, we consider the consequences brought about by the existence of divergent terms in the nonlinear optical coefficients, specifically in the second order conductivity of a cold semiconductor. We show that, by using properties, Eqs.(5.75) and (5.76), and expressions, Eqs.(5.77) and (5.78), derived in the previous section, the well-known Kramers-Kronig relations presented in Section 2.1 must be reassessed in order to account for the existence of the injection current. We also derive, by using a general argument, a relation between real parts of the second order conductivity tensor of a cold semiconductor that is (as we understand it) entirely new, and that can account for the existence of the injection current. A special case of this relation shows that three important effects in nonlinear optics — the sum-frequency generation, the difference-frequency generation and the optical rectification — are tied together. Furthermore, while this reasoning can, in principle, be extended to the study of the third order response, the assortment of divergences present in that order makes this a much more difficult task than the one undertaken here (which, although not terribly complicated, involves a rather long-winded derivation). Finally, and as an aside, we note that there is a somewhat remarkable coincidence involved here; we began this thesis by thinking, in general terms, about these nonlinear optical coefficients and their general properties. We now finish this thesis by adding another new, possibly important, element to this study.

5.5.1 Reassessing the Second Order Kramers-Kronig Relations

In the chapter on the general properties of nonlinear optical coefficients, we discussed, if briefly, the existence of relations connecting the real part of a response function with its imaginary part, and vice-versa. There, we showed that these relations — the Kramers-Kronig relations — are predicated on the fact that response in the time domain is causal, or equivalently, that the response function in the frequency domain is analytic in the upper half of the complex plane. The appearance of physical poles (associated with physical divergences) close to, but not crossing, the real axis thus forces us to reassess this description, although as we will see in the following subsections, these physical poles do not render these relations invalid. For the moment, let us focus on the immediate consequences of having a divergent contribution in the conductivity, that is, let us consider the

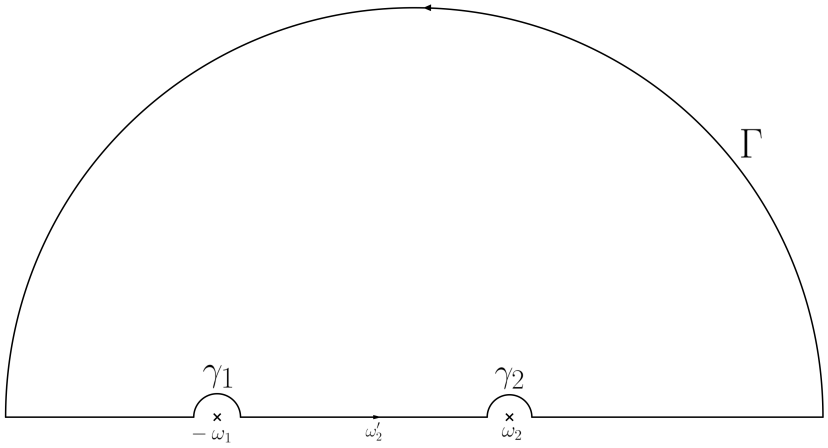


Figure 5.4: The contour used in the integral of Eq.(5.83). Note that it goes above the real axis for both $\omega_2' = \omega_2$ and $\omega_2' = -\omega_1$; the γ_1 and γ_2 portions of the contour represents the semicircumference of vanishing radius centred around each of these points. As before, the Γ portion of the contour represents the semicircumference whose radius is to be taken to infinity.

changes we must make to the contour-integral description used to derive the usual second order (nondegenerate) Kramers-Kronig relations, Eq.(2.34).

First, let us consider our integrand,

$$\frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1, z)}{z - \omega_2}. \tag{5.82}$$

It now carries not one, but two poles: one in ω_2, as made evident by the denominator; the other, which is hidden inside the expression for the second order conductivity, in $-\omega_1$, Eq.(5.78). So, in order to obtain an integral that is once again trivial, we must deform the contour of Fig.2.1, so as to go around the new, physical, pole. This new contour is represented in Fig.5.4.

$$\oint_C dz \frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1, z)}{z - \omega_2}, \tag{5.83}$$

As before, no poles lie inside the domain defined by this contour, and so the integral in Eq.(5.83) must be exactly zero. We can now break up the integral into its different contributions, like we did in Section 2.1, and obtain,

$$⨍ d\omega_2' \frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1, \omega_2')}{\omega_2' - \omega_2} + \int_{\gamma_1} dz \frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1, z)}{z - \omega_2} + \int_{\gamma_2} dz \frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1, z)}{z - \omega_2} = 0, \tag{5.84}$$

as there are no contributions coming from the external semicircumference, Γ. The first integral in Eq.(5.84) is a principal value integral, while the second and third integrals correspond to the integrations around the (vanishing) semicircumferences centred in points, $-\omega_1$ and ω_2. The latter

integral is evaluated exactly as before. For $z = \omega_2 - \delta e^{-i\theta}$, for $\theta \in \left[0, \pi\right]$ and $\delta \to 0$, we obtain,

$$\int_{\gamma_2} dz \frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1, z)}{z - \omega_2} = -i\pi\sigma^{\beta\alpha_1\alpha_2}(\omega_1, \omega_2). \tag{5.85}$$

It then follows from Eq.(5.84) that,

$$\sigma^{\beta\alpha_1\alpha_2}(\omega_1, \omega_2) = \frac{1}{i\pi} \fint d\omega_2' \frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1, \omega_2')}{\omega_2' - \omega_2} + \frac{1}{i\pi} \int_{\gamma_1} dz \frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1, z)}{z - \omega_2}, \tag{5.86}$$

which, if we consider solely the real part of both hand sides, reads as,

$$\mathrm{Re}\left[\sigma^{\beta\alpha_1\alpha_2}(\omega_1, \omega_2)\right] = \frac{1}{\pi} \fint d\omega_2' \frac{\mathrm{Im}\left[\sigma^{\beta\alpha_1\alpha_2}(\omega_1, \omega_2')\right]}{\omega_2' - \omega_2} + \left[\frac{1}{2i\pi} \int_{\gamma_1} dz \frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1, z)}{z - \omega_2} + \text{c.c.}\right]. \tag{5.87}$$

Note that when the conductivity does not carry a divergence for $\omega_2 = -\omega_1$, the contributions from this last couple of terms are automatically zero, meaning that the usual Kramers-Kronig relations are indeed satisfied.[19] This is, of course, the case of the cold semiconductor when we consider the response along its diagonal elements: for $\beta = \alpha_1 = \alpha_2$, the injection-current coefficient, $I_1^{\beta\beta\beta}(\omega)$, is zero, Eq.(5.81), and the contour in Fig.2.1 does not have to be deformed. If, however, $\sigma_{\text{div.}}^{\beta\alpha_1\alpha_2}(\omega_1, \omega_2)$ is *not* zero, as is generally the case in the cold semiconductor, these relations have to be reassessed, Eq.(5.87). Before we return to the particular case of the cold semiconductor, so as to to compute these contributions, let us first consider what happens when we use the fact that the imaginary part of the conductivity possesses full permutation symmetry.

5.5.2 A New Relation between Real Parts

Let us consider the principal value integral in Eq.(5.87),

$$\frac{1}{\pi} \fint d\omega_2' \frac{\mathrm{Im}\left[\sigma^{\beta\alpha_1\alpha_2}(\omega_1, \omega_2')\right]}{\omega_2' - \omega_2}. \tag{5.88}$$

As previously noted, the imaginary part of the conductivity has full-permutation symmetry, meaning that it is symmetric under the simultaneous exchange of its first and last Cartesian indexes and their corresponding frequency arguments, Eq.(5.76). We can use write this property to write the integrand in Eq.(5.88) as,

$$-\frac{1}{\pi} \fint d\omega_2' \frac{\omega_{12'}}{\omega_2'} \frac{\mathrm{Im}\left[\sigma^{\alpha_2\alpha_1\beta}(\omega_1, -\omega_{12'})\right]}{\omega_2' - \omega_2}. \tag{5.89}$$

[19] By this we mean the Kramers-Kronig relations as written in Subsection 2.1.

Note that we now have two denominators expressed in terms of the integration variable, ω_2'. Decomposing this product in simple fractions,

$$\frac{\omega_{12'}}{\omega_2'}\frac{1}{\omega_2'-\omega_2}=\frac{\omega_1}{\omega_2}\frac{1}{\omega_2'}-\frac{\omega_{12}}{\omega_2}\frac{1}{\omega_2'-\omega_2}, \tag{5.90}$$

we obtain two different contributions, which, after changing their integration variables, are simply,

$$-\frac{\omega_1}{\omega_2}\frac{1}{\pi}\fint d\omega_2'\frac{\mathrm{Im}\big[\sigma^{\alpha_2\alpha_1\beta}(\omega_1,\omega_{2'})\big]}{\omega_2'+\omega_1}, \tag{5.91}$$

and,

$$\frac{\omega_{12}}{\omega_2}\frac{1}{\pi}\fint d\omega_2'\frac{\mathrm{Im}\big[\sigma^{\alpha_2\alpha_1\beta}(\omega_1,\omega_{2'})\big]}{\omega_2'+\omega_{12}}. \tag{5.92}$$

We can write thus Eq.(5.89), our first principal value integral, as the sum of these two contributions,

$$\frac{1}{\pi}\fint d\omega_2'\frac{\mathrm{Im}\big[\sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2')\big]}{\omega_2'-\omega_2}=\frac{\omega_{12}}{\omega_2}\frac{1}{\pi}\fint d\omega_2'\frac{\mathrm{Im}\big[\sigma^{\alpha_2\alpha_1\beta}(\omega_1,\omega_{2'})\big]}{\omega_2'+\omega_{12}}-\frac{\omega_1}{\omega_2}\frac{1}{\pi}\fint d\omega_2'\frac{\mathrm{Im}\big[\sigma^{\alpha_2\alpha_1\beta}(\omega_1,\omega_{2'})\big]}{\omega_2'+\omega_1}. \tag{5.93}$$

If we now return to our generalization of the Kramers-Kronig relations, Eq.(5.87), and use it so as to express these imaginary parts in terms of their real counterparts, this relation becomes the following,

$$\mathrm{Re}\big[\sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)\big]=\frac{\omega_{12}}{\omega_2}\mathrm{Re}\big[\sigma^{\alpha_2\alpha_1\beta}(\omega_1,-\omega_{12})\big]-\frac{\omega_1}{\omega_2}\mathrm{Re}\big[\sigma^{\alpha_2\alpha_1\beta}(\omega_1,-\omega_1)\big]+(\gamma_1\text{ contri.}), \tag{5.94}$$

for,

$$(\gamma_1\text{ contri.})=\frac{1}{2i\pi}\int_{\gamma_1}dz\frac{\sigma^{\beta\alpha_1\alpha_2}(\omega_1,z)}{z-\omega_2}-\frac{1}{2i\pi}\frac{\omega_{12}}{\omega_2}\int_{\gamma_1}dz\frac{\sigma^{\alpha_2\alpha_1\beta}(\omega_1,z)}{z+\omega_{12}}+\frac{1}{2i\pi}\frac{\omega_1}{\omega_2}\int_{\gamma_1}dz\frac{\sigma^{\alpha_2\alpha_1\beta}(\omega_1,z)}{z+\omega_1}+\text{ c.c.} \tag{5.95}$$

This result, we note, is completely general, as is the result for the set of slightly modified Kramers-Kronig relations (of which we have written explicitly only Eq.(5.87)); the only condition we have imposed throughout these two subsections is that there is a single additional pole, centred in $-\omega_1$. This reasoning can thus be easily extended to the case where the response functions have more two or more divergences and, therefore, two or more poles.[20] Finally, let us return to the cold semiconductor — and its particular divergence, the injection current — and derive the actual form of Eq.(5.94) for this type of system.

[20] An example of this is found in the third order response. There, we have to account for the jerk current, the divergence in the one-photon process of the Kerr effect, as well as several other effects. Furthermore, since these effects do not vanish when we consider the diagonal components of the conductivity, even these elements must carry the contributions from the γ_1 integrals. This, as we have already noted, is different from what happens in second order.

5.5.3 A New Relation between Real Parts of the Second Order Conductivity Tensor of a Cold Semiconductor

We begin by recovering the expression for $\sigma^{\beta\alpha_1\alpha_2}_{\text{div.}}$, Eq.(5.78),

$$\frac{1}{\sigma_2}\sigma^{\beta\alpha_1\alpha_2}_{\text{div.}}(\omega_1,\omega_2) = \frac{1}{\bar{\omega}_{12}} I_1^{\alpha_1\alpha_2\beta}(\omega_1) + \frac{1}{\bar{\omega}_{12}} I_1^{\alpha_2\alpha_1\beta}(\omega_2),$$

where we have defined σ_2, as $\sigma_2 = i\pi\left|e\right|^3/2\hbar^2$. If we consider the frequencies that define the contour, that is, $\omega_2 = -\omega_1 - \delta e^{-i\theta}$, for δ the radius of the γ_1 semicircumference, and expand $\sigma^{\beta\alpha_1\alpha_2}_{\text{div.}}$ in powers of δ, we obtain,

$$\frac{1}{\sigma_2}\sigma^{\beta\alpha_1\alpha_2}_{\text{div.}}(\omega_1,-\omega_1-\delta e^{-i\theta}) \approx -\frac{2e^{i\theta}}{\delta} I_1^{\alpha_1\alpha_2\beta}(\omega_1) - \left(\partial_{\omega_1} I_1^{\alpha_1\alpha_2\beta}(\omega_1)\right). \tag{5.96}$$

Let us evaluate the contribution associated with each of the γ_1 integrals in Eq.(5.95). The first γ_1 integral reads simply as [172],

$$\frac{1}{2\pi i}\int_{\gamma_1} dz \frac{\sigma^{\beta\alpha_1\alpha_2}_{\text{div.}}(\omega_1,z)}{z-\omega_2} = -\frac{1}{2\pi i}\frac{1}{\omega_{12}}\int_0^\pi d\theta (i\delta e^{-i\theta})\sigma^{\beta\alpha_1\alpha_2}_{\text{div.}}(\omega_1,-\omega_1-\delta e^{-i\theta}), \tag{5.97}$$

which, using Eq.(5.96), gives us,

$$\frac{1}{2\pi i}\int_{\gamma_1} dz \frac{\sigma^{\beta\alpha_1\alpha_2}_{\text{div.}}(\omega_1,z)}{z-\omega_2} \approx \sigma_2 \frac{1}{\omega_{12}} I_1^{\alpha_1\alpha_2\beta}(\omega_1). \tag{5.98}$$

The second γ_1 integral, which carries a conductivity where the first and last Cartesian indexes are exchanged with respect to that of Eq.(5.96), $\sigma^{\alpha_2\alpha_1\beta}_{\text{div.}}$, gives us a similar result,

$$-\frac{1}{2\pi i}\frac{\omega_{12}}{\omega_2^2}\int_{\gamma_1} dz \frac{\sigma^{\alpha_2\alpha_1\beta}_{\text{div.}}(\omega_1,z)}{z-\omega_2} \approx \sigma_2 \frac{\omega_{12}}{\omega_2^2} I_1^{\alpha_1\beta\alpha_2}(\omega_1), \tag{5.99}$$

since the integral along the semicircumference is exactly the same. Finally, let us consider the third γ_1 integral. This is different from the other two integrals, because its denominator is not just a frequency factor that can be factored out. It instead reads as,

$$\frac{1}{z+\omega_1} \to \frac{1}{(-\delta e^{-i\theta})}, \tag{5.100}$$

so that the integral (in the right-hand side) does not carry a factor of δ in the numerator,

$$\frac{1}{2\pi i}\frac{\omega_1}{\omega_2}\int_{\gamma_1} dz \frac{\sigma^{\alpha_2\alpha_1\beta}_{\text{div.}}(\omega_1,z)}{z+\omega_1} = \frac{1}{2\pi i}\frac{\omega_1}{\omega_2}\int_0^\pi (-id\theta)\sigma^{\alpha_2\alpha_1\beta}_{\text{div.}}(\omega_1,-\omega_1-\delta e^{-i\theta}), \tag{5.101}$$

meaning that it thus carries an additional term. Replacing the expression for $\sigma^{\alpha_2\alpha_1\beta}_{\text{div.}}$, the integral can be expressed as,

$$\frac{1}{2\pi i}\frac{\omega_1}{\omega_2}\int_{\gamma_1} dz \frac{\sigma^{\alpha_2\alpha_1\beta}_{\text{div.}}(\omega_1, z)}{z+\omega_1} \approx \frac{\sigma_2}{2\pi i}\frac{\omega_1}{\omega_2}\left[-\frac{4}{\delta} I_1^{\alpha_1\alpha_2\beta}(\omega_1) + i\pi\big(\partial_{\omega_1} I_1^{\alpha_1\beta\alpha_2}(\omega_1)\big)\right]. \tag{5.102}$$

Let us now sum each of these contributions with its respective complex conjugate, so as to obtain what we described in Eq.(5.94) as the γ_1 contributions. For the first two integrals, Eqs.(5.98) and (5.99), this is trivial. We obtain,

$$\sigma_2 \frac{1}{\omega_{12}} I_1^{\alpha_1\alpha_2\beta}(\omega_1) + (-\sigma_2)\frac{1}{\omega_{12}}(-I_1^{\alpha_1\alpha_2\beta}(\omega_1)) = \frac{2\sigma_2}{\omega_{12}} I_1^{\alpha_1\alpha_2\beta}(\omega_1), \tag{5.103}$$

and,

$$\sigma_2 \frac{\omega_{12}}{\omega_2^2} I_1^{\alpha_1\beta\alpha_2}(\omega_1) + (-\sigma_2)\frac{\omega_{12}}{\omega_2^2}(-I_1^{\alpha_1\beta\alpha_2}(\omega_1)) = 2\sigma_2 \frac{\omega_{12}}{\omega_2^2} I_1^{\alpha_1\beta\alpha_2}(\omega_1), \tag{5.104}$$

since $(\sigma_2)^* = -i\pi|e|^3/2\hbar^2 = -\sigma_2$ and, as stated in Eq.(5.79), $\big(I_1^{\alpha_1\alpha_2\beta}(\omega_1)\big)^* = -I_1^{\alpha_1\alpha_2\beta}(\omega_1)$. For the third integral, Eq.(5.102), however, we see that the divergent term (as $\delta \to 0$) changes sign under complex conjugation — since it carries an additional factor of $(2\pi i)^{-1}$ — meaning that it cancels out when it is summed with its complex conjugate. Therefore, we are left only with,

$$\frac{\sigma_2}{2}\frac{\omega_1}{\omega_2}\big(\partial_{\omega_1} I_1^{\alpha_1\beta\alpha_2}(\omega_1)\big) + \frac{(-\sigma_2)}{2}\frac{\omega_1}{\omega_2}\big(-\partial_{\omega_1} I_1^{\alpha_1\beta\alpha_2}(\omega_1)\big) = \sigma_2\frac{\omega_1}{\omega_2}\big(\partial_{\omega_1} I_1^{\alpha_1\beta\alpha_2}(\omega_1)\big). \tag{5.105}$$

This can then be rewritten in terms of the shift-current coefficients, Eq.(B.53), by using Eq.(B.54),[21]

$$\sigma_2 \frac{\hbar\omega_1}{\omega_2}\big[I_2^{\alpha_1\beta\alpha_2}(\omega_1) - I_2^{\beta\alpha_1\alpha_2}(\omega_1)\big]. \tag{5.106}$$

Summing the three γ_1 contributions, we can finally write the relation between different real parts of $\sigma^{\beta\alpha_1\alpha_2}$ we derived in the previous section, Eq.(5.94) — in the case of the cold (two-band) semiconductor — as,[22]

$$\begin{aligned}\sigma_2^{-1}\text{Re}\big[\sigma^{\beta\alpha_1\alpha_2}(\omega_1,\omega_2)\big] = {}& \frac{\omega_{12}}{\omega_2}\sigma_2^{-1}\text{Re}\big[\sigma^{\alpha_2\alpha_1\beta}(\omega_1,-\omega_{12})\big] - \frac{\omega_1}{\omega_2}\sigma_2^{-1}\text{Re}\big[\sigma^{\alpha_2\alpha_1\beta}(\omega_1,-\omega_1)\big] \\ & + \frac{2}{\omega_{12}} I_1^{\alpha_1\alpha_2\beta}(\omega_1) + \frac{2\omega_{12}}{\omega_2^2} I_1^{\alpha_1\beta\alpha_2}(\omega_1) + \frac{\hbar\omega_1}{\omega_2}\big[I_2^{\alpha_1\beta\alpha_2}(\omega_1) - I_2^{\beta\alpha_1\alpha_2}(\omega_1)\big].\end{aligned} \tag{5.107}$$

[21] We derive the relation between these two coefficients in Appendix B.5.

[22] Note that we can also show this explicitly [171], by using the expressions for the second order conductivity of the two-band model derived by D. J. Passos [10].

Furthermore, we can also see that the relation between diagonal elements,

$$\mathrm{Re}\big[\sigma^{\beta\beta\beta}(\omega_1,\omega_2)\big]=\frac{\omega_{12}}{\omega_2}\mathrm{Re}\big[\sigma^{\beta\beta\beta}(\omega_1,-\omega_{12})\big]-\frac{\omega_1}{\omega_2}\mathrm{Re}\big[\sigma^{\beta\beta\beta}(\omega_1,-\omega_1)\big], \tag{5.108}$$

is exactly what we have previously described. We must emphasize the relevance of this result: Eq.(5.108) states that the processes involved in the sum-frequency generation, for frequencies ω_1 and ω_2, can be described as the sum (weighted by appropriate frequency factors) of the processes involved in the difference-frequency generation, for frequencies ω_1 and $-\omega_{12}$, and the shift current; three seemingly different processes are therefore tied together in a relatively simple manner [171].

The existence of relations between different portions of the conductivity, whether they are between different real parts, as in Eqs.(5.107) and (5.108), or between different imaginary parts, as in Eqs.(5.75) and (5.76), shows how intricate a nonlinear response function can in principle be, and how interconnected its different constitutive elements actually are. While these results do not seem to have any immediate practical application, they add a new element to the study of the general properties of nonlinear optical coefficients.

5.5.4 A Numerical Check of this New Relation

We close this section, as well as this chapter, with a semi-numerical check of Eq.(5.107). It has been our policy throughout our work to cross-check our results, either directly or indirectly, by using different frameworks (the velocity and length gauges) and different Hamiltonians (full-band and effective descriptions). It is, therefore, fitting that we validate these new results much in the same manner as before. There is, however, something of an obstacle: for the two physical models we have considered throughout this work, the second order response is either trivial (the case of the plain graphene monolayer) or it is simple enough that it inhibits the study of the relation written in Eq.(5.107) (the case of the gapped graphene monolayer, where all second order conductivities can be reduced to a single diagonal element). Thus, we must introduce a new model, which we constructed so as to give out a simple enough description that is, nonetheless, nontrivial.[23] It is represented in Figure 5.5. As for its tight-binding Hamiltonian, it reads as,

$$H_{\mathbf{k}}=\begin{bmatrix}\Delta/2 & (-t)\left(e^{-ik_xa}+e^{-ik_ya}\right)\\ (-t)\left(e^{ik_xa}+e^{ik_ya}\right) & -\Delta/2\end{bmatrix}, \tag{5.109}$$

which has the following eigenvalues,

$$\epsilon_{\mathbf{k}c(v)}=\pm\frac{1}{2}\sqrt{\Delta^2+8t^2+8t^2\cos\left(k_xa-k_ya\right)}. \tag{5.110}$$

[23] The particulars of this toy model need not concern us much, as it is something that we will use strictly to study three special cases of Eq.(5.107).

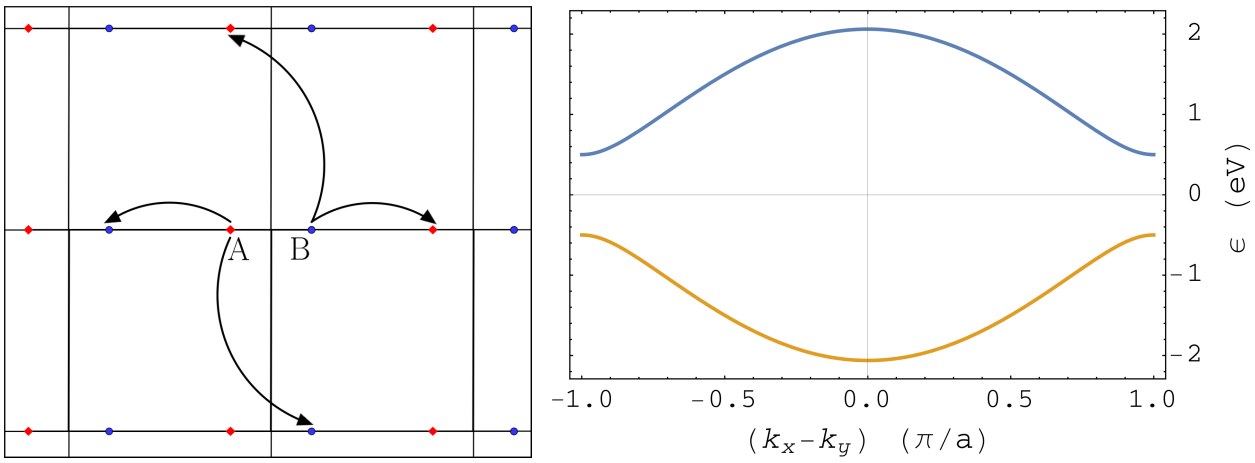


Figure 5.5: [Left] A pictorial representation of the tight-binding model of Eq.(5.109). It involves a square lattice (of lattice parameter a) with a basis of two different atoms, represented by the red squares (A) and the blue circles (B), which have different on-site energies, $\epsilon_{\rm A} - \epsilon_{\rm B} = \Delta$; inversion symmetry is therefore absent. The black arrows represent the relevant hoppings $(-t)$ to nearest neighbours. Note the two atoms inside the same unit cell are not connected by a hopping matrix element, and that we neglect the distances between atoms inside the same unit cell in the tight-binding model. The atoms have been represented slightly apart so as to make this picture clearer. [Right] The conduction and valence bands of this two-dimensional toy model, Eq.(5.110). Note that these are a function of the difference between the crystal momenta along the x and y directions, $k_x - k_y$. While this model is two-dimensional, its analytical description can be made to be one-dimensional.

These are also represented in Figure 5.5 for the parameters, $\Delta = t = 1$ eV. The other relevant elements for calculations in the length gauge, the Berry connections, are written explicitly in Appendix B.6. We can now consider the following special cases of the relation between different real parts we have just derived, Eq.(5.107),

$$\mathrm{Re}\big[\sigma^{\alpha\alpha\beta}(\omega,\omega)\big] = 2\mathrm{Re}\big[\sigma^{\beta\alpha\alpha}(\omega,-2\omega)\big] - \mathrm{Re}\big[\sigma^{\beta\alpha\alpha}(\omega,-\omega)\big] + \frac{\sigma_2}{\omega} I_1^{\alpha\beta\alpha}(\omega), \tag{5.111}$$

$$\mathrm{Re}\big[\sigma^{\beta\alpha\alpha}(\omega,\omega)\big] = 2\mathrm{Re}\big[\sigma^{\alpha\alpha\beta}(\omega,-2\omega)\big] - \mathrm{Re}\big[\sigma^{\alpha\alpha\beta}(\omega,-\omega)\big] + \frac{4\sigma_2}{\omega} I_1^{\alpha\beta\alpha}(\omega) + \hbar\sigma_2\big[I_2^{\alpha\beta\alpha}(\omega) - I_2^{\beta\alpha\alpha}(\omega)\big], \tag{5.112}$$

$$\mathrm{Re}\big[\sigma^{\beta\alpha\alpha}(\omega,-\omega)\big] = \mathrm{Re}\big[\sigma^{\alpha\alpha\beta}(\omega,-\omega)\big] - \hbar\sigma_2\big[I_2^{\alpha\beta\alpha}(\omega) - I_2^{\beta\alpha\alpha}(\omega)\big]. \tag{5.113}$$

The first two relations connect the real part of the second harmonic generation with the real parts of the difference-frequency generation and the optical rectification, while the third relation connects two different components of this latter effect. As for the procedure involved in these checks, it is on the one hand, numerical, as the conductivities are calculated using the velocity gauge description introduced in Chapter 4, and, on the other hand, analytical, as the FBZ integrals for the I_1 and I_2 coefficients, Eqs.(5.77) and (B.53), can be computed explicitly for the entire band. These semi-numerical tests are presented in Figure 5.6, where we plot the right- and left-hand sides of these relations for the toy model described by Eq.(5.109) and the aforementioned parameters. Furthermore, note that we have only considered frequencies above the one-photon process at the gap, as otherwise the coefficients in Eqs.(5.111)–(5.113) would be exactly zero.

The clear agreement between the two sets of results validates the relations between real parts in Eqs.(5.111)–(5.113). This, therefore, indicates that Eq.(5.107) can indeed be treated as an integral property of the nonlinear optical coefficients of cold semiconductors.

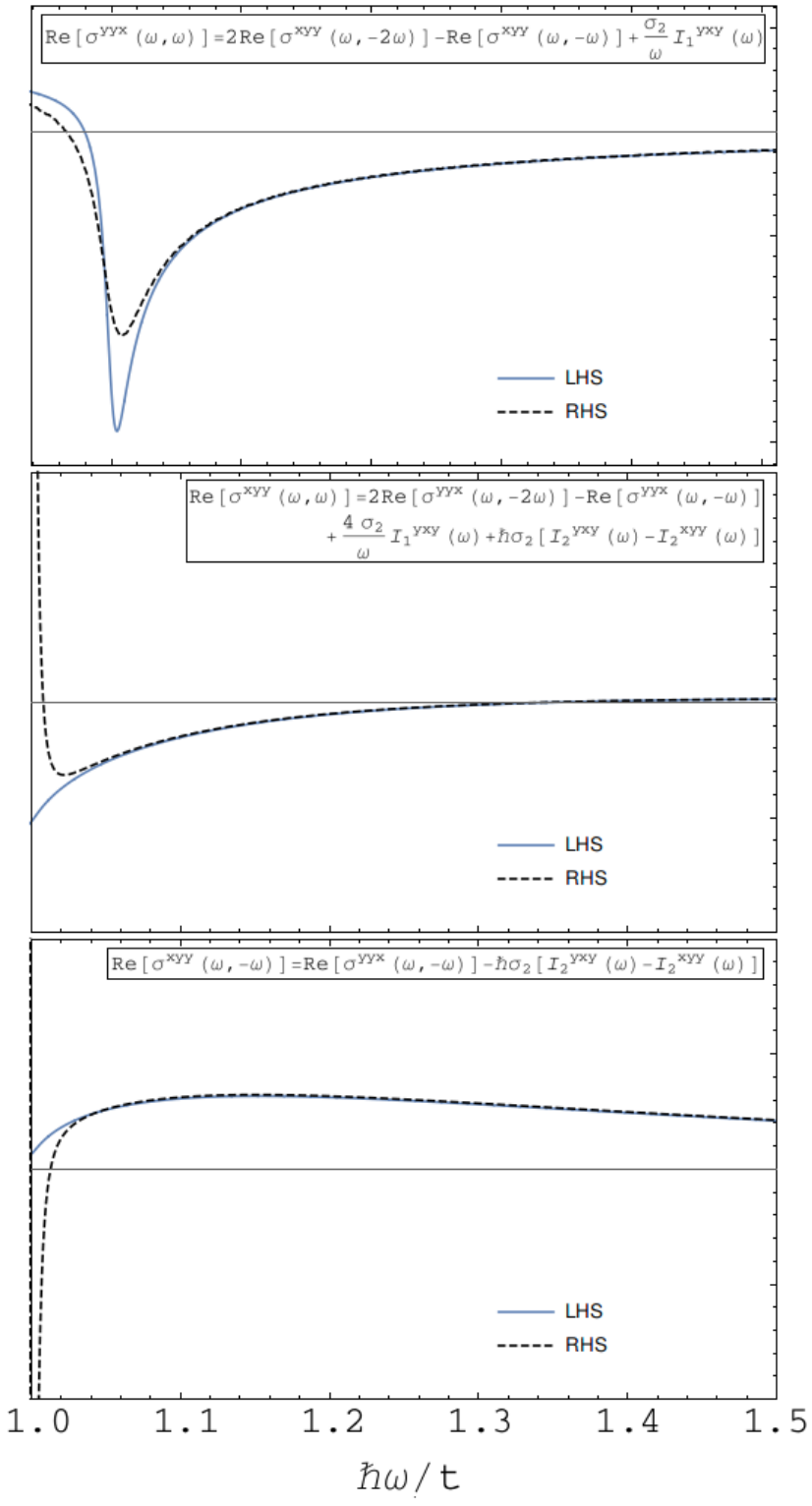


Figure 5.6: Plots of the semi-numerical tests of the relations in Eqs.(5.111)–(5.113) for the two-dimensional toy model described by Eq.(5.109), and using the aforementioned parameters. In the numerical calculations, the relaxation parameter was fixed to $\eta = 0.005t$.

Chapter 6

Conclusion

This thesis was dedicated to three subjects in the nonlinear optical response of crystals: a general study of density-matrix calculations, and of questions involved in the physical equivalence of its two most common gauges; the application of the velocity-gauge formalism that was formulated in ref.[3], and a review of some practical considerations that appear therein; a study of the DC divergences in the optical response of a cold semiconductor, and of the role these divergences play in determining the internal structure of its nonlinear optical coefficients. Considering that the reader has had to read through a somewhat lengthy work (and that the contents of the chapters are briefly described in their respective preambles), we will spare them from a final comprehensive overview. Instead, we think it is useful to add some remarks on how these three subjects can be best used in the study of NLO responses, as well as point out to some possibly interesting applications.

The density matrix formalism allows one to study the full nonlinear optical response of crystals; if treated perturbatively, it provides one with expressions for the nonlinear optical coefficients which can, in principle, describe every effect in that order; if treated nonperturbatively, it can describe the complete current response of these materials. The latter procedure, which was not carefully discussed in this work, has been used to study the response to pulses that are extremely short or extremely intense, but it can also be used to study effects that are associated with divergences in the NLO response, such as the shift, injection, and jerk currents, and how these are affected by, e.g., saturation and decoherence phenomena. These effects, as we have seen in this thesis, can be studied (in the context of perturbation theory) both from the standpoint of the nonlinear optical coefficients and from the standpoint of Fermi's golden rule arguments; while the former is more rigorous and can be used in a thorough analysis of every divergence at a given order, the latter is simpler and more intuitive. These two procedures can thus complement each other, and endow one with a clearer understanding of this unavoidable element in a crystal's NLO response. Furthermore, while these divergences are relevant on their own, they also play an important role in determining the internal structure of the nonlinear optical coefficients, as we have seen here in the case of the second order response of a cold semiconductor. An extension of this result to a more general type

of crystal, as well as to the response in third order, is certainly in order.

Finally, some comments on the methods themselves. As we have mentioned in this thesis, the length and velocity gauges are probably best suited for analytical and numerical calculations, respectively; they can also be used, as was the case in our work on the jerk current, to validate results obtained from different procedures. When we consider their implementation, these two descriptions are subjected to different problems and obstacles. These issues can, nevertheless, be mitigated by using some of the insights of this work: the h coefficients, in the context of a tight-binding Hamiltonian, can be trivially computed when a new basis of Bloch functions is introduced; problems concerning nontrivial band structures, such as that of gapped graphene with trigonal warping terms, can be overcome when the series relating the difference in band energies with the crystal momentum is inverted. As with the above-described complementary use of density matrix and Fermi's golden rule calculations, the future study of NLO responses will most likely benefit from the use (when possible) of both analytical and numerical methods.

Appendix A

Appendices Pertaining to Chapter 3

A.1 The First Three Nonlinear Optical Conductivities in the Length Gauge

Consider the expression for the current density in the length gauge,

$$\big\langle \mathbf{J}_E(t)\big\rangle = -\,|e|\int \frac{d^d\mathbf{k}}{(2\pi)^d}\mathrm{Tr}\left[\mathbf{v}(\mathbf{k})\,\rho^E(\mathbf{k},t)\right] = -\,|e|\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\mathbf{v}_{\mathbf{k}ss'}\,\rho^E_{\mathbf{k}s's}(t). \tag{A.1}$$

We have seen in Chapter 3 that the solutions to the density matrix elements can be written as,

$$\begin{aligned}\rho^{(n),E}_{\mathbf{k}ss'}(\omega) = \;& (i\,|e|)^n\left[\prod_{i=1}^{n-1}\int\frac{d\omega_i}{2\pi}\,E^{\alpha_i}(\omega_i)\right]E^{\alpha_n}(\omega-\omega_{[n-1]})\\ &\times d_{\mathbf{k}ss'}(\omega)\big[D^{\alpha_1},\,d(\omega-\omega_1)\circ\big[...,d(\omega-\omega_{[n-1]})\circ\big[D^{\alpha_n},\rho^{(0)}(\mathbf{k})\big]...\big]\big]_{ss'},\end{aligned} \tag{A.2}$$

which means that we can easily derive the expressions for the nonlinear optical coefficients that describe the current density, the conductivities. It only involves replacing the density matrix elements of Eq.(A.2) in Eq.(A.1), and carrying on some menial manipulations.

The first order current density in the frequency domain thus reads as,

$$J^{(1),\beta}_E(\omega) = -\,|e|\int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}v^{\beta}_{\mathbf{k}s's}\,\rho^{(1),E}_{\mathbf{k}s's}(\omega), \tag{A.3}$$

$$= -\,ie^2\int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}v^{\beta}_{\mathbf{k}s's}\frac{1}{\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'}+i\eta}\big[D^{\alpha_1},\rho^{(0)}(\mathbf{k})\big]_{ss'}\,E^{\alpha_1}(\omega). \tag{A.4}$$

It follows from Eq.(2.8) that,

$$J^{(1),\beta}_E(\omega) = \sigma^{\beta\alpha_1}_E(\omega)E^{\alpha_1}(\omega), \tag{A.5}$$

so that the conductivity is written as,

$$\sigma_E^{\beta\alpha_1}(\omega) = -ie^2 \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{s's} v^{\beta}_{\mathbf{k}s's} \frac{1}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'} + i\eta} \left[D^{\alpha_1}, \rho^{(0)}(\mathbf{k})\right]_{ss'}. \tag{A.6}$$

For higher order conductivities, however, these manipulations are slightly more complicated. First, note that we can introduce an additional frequency integral (over ω_n) in the expression for the density matrix elements, Eq.(A.2), by including a Dirac delta function, $\delta(\omega - \omega_{[n]})$, in the integrand. In the case of $J_E^{(2),\beta}(\omega)$, this means using a delta function that reads as $\delta(\omega - \omega_{[2]})$, or, equivalently, $\delta(\omega - \omega_{12})$, inside the (now double) frequency integral,[1]

$$\begin{aligned} J_E^{(2),\beta}(\omega) = |e|^3 \int \frac{d\omega_1}{2\pi} \int \frac{d\omega_2}{2\pi} \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{s's} v^{\beta}_{\mathbf{k}s's} \frac{1}{\hbar\omega_{12} - \Delta\epsilon_{\mathbf{k}ss'} + 2i\eta} \left[D^{\alpha_1}, \frac{1}{\hbar\omega_2 - \Delta\epsilon + i\eta} \circ \left[D^{\alpha_2}, \rho^{(0)}(\mathbf{k})\right]\right]_{ss'} \\ \times E^{\alpha_1}(\omega_1) E^{\alpha_2}(\omega_2) (2\pi)\delta(\omega - \omega_{12}). \end{aligned} \tag{A.7}$$

The second order conductivity thus reads as, Eq.(2.24),

$$\sigma_E^{\beta\alpha_1\alpha_2}(\omega_1, \omega_2) = |e|^3 \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{s's} v^{\beta}_{\mathbf{k}s's} \frac{1}{\hbar\omega_{12} - \Delta\epsilon_{\mathbf{k}ss'} + 2i\eta} \left[D^{\alpha_1}, \frac{1}{\hbar\omega_2 - \Delta\epsilon + i\eta} \circ \left[D^{\alpha_2}, \rho^{(0)}(\mathbf{k})\right]\right]_{ss'}. \tag{A.8}$$

Note that this conductivity is not intrinsically symmetric.

Finally, the third order current density reads as,

$$\begin{aligned} J_E^{(3),\beta}(\omega) = ie^4 \int \frac{d\omega_1}{2\pi} \int \frac{d\omega_2}{2\pi} \int \frac{d\omega_3}{2\pi} \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{s's} v^{\beta}_{\mathbf{k}s's} \frac{1}{\hbar\omega_{123} - \Delta\epsilon_{\mathbf{k}ss'} + 3i\eta} \\ \times \left[D^{\alpha_1}, \frac{1}{\hbar\omega_{23} - \Delta\epsilon + 2i\eta} \circ \left[D^{\alpha_2}, \frac{1}{\hbar\omega_3 - \Delta\epsilon + i\eta} \circ \left[D^{\alpha_3}, \rho^{(0)}(\mathbf{k})\right]\right]\right]_{ss'} \\ \times E^{\alpha_1}(\omega_1) E^{\alpha_2}(\omega_2) E^{\alpha_3}(\omega_3) (2\pi)\delta(\omega - \omega_{123}), \end{aligned} \tag{A.9}$$

which, following the same reasoning as before, gives us,

$$\begin{aligned} \sigma_E^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1, \omega_2, \omega_3) = ie^4 \int \frac{d^d\mathbf{k}}{(2\pi)^d} \sum_{s's} v^{\beta}_{\mathbf{k}s's} \frac{1}{\hbar\omega_{123} - \Delta\epsilon_{\mathbf{k}ss'} + 3i\eta} \\ \times \left[D^{\alpha_1}, \frac{1}{\hbar\omega_{23} - \Delta\epsilon + 2i\eta} \circ \left[D^{\alpha_2}, \frac{1}{\hbar\omega_3 - \Delta\epsilon + i\eta} \circ \left[D^{\alpha_3}, \rho^{(0)}(\mathbf{k})\right]\right]\right]_{ss'}. \end{aligned} \tag{A.10}$$

Once again, note that this conductivity is not intrinsically symmetric. For the corresponding expressions in the velocity gauge (in first and second orders), check Subsection 4.2.2.

[1] This follows from the notation introduced in Subsection 4.2.2.

Appendix B

Appendices Pertaining to Chapter 5

B.1 The Final Steps in the Derivation of the Population Density Rate of Injected Electrons

The population density rate of injected electrons reads as, Eq.(5.8),

$$\dot{n}(t) = \frac{2}{\hbar\omega}\mathrm{Re}\left[\sigma^{\alpha_1\alpha_2}(\omega)E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\right]. \tag{B.1}$$

Separating the real and imaginary parts of the conductivity and the product of electric fields, we obtain,

$$\dot{n}(t) = \frac{2}{\hbar\omega}\mathrm{Re}\left[\sigma^{\alpha_1\alpha_2}(\omega)\right]\mathrm{Re}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\right] - \frac{2}{\hbar\omega}\mathrm{Im}\left[\sigma^{\alpha_1\alpha_2}(\omega)\right]\mathrm{Im}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\right]. \tag{B.2}$$

Let us then consider the conductivity. In the clean limit, $\eta \to 0$, we can write it as the sum of two different contributions,

$$\sigma^{\alpha_1\alpha_2}(\omega) = \frac{ie^2}{\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Delta\epsilon_{\mathbf{k}s's}\Delta f_{\mathbf{k}s's}\xi^{\alpha_1}_{\mathbf{k}s's}\xi^{\alpha_2}_{\mathbf{k}ss'}\left[\frac{\mathrm{P}}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}} - i\pi\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'})\right]. \tag{B.3}$$

One involves a principal value integral, which can be decomposed in its real and imaginary parts,

$$\mathrm{Re}\left[\sigma^{\alpha_1\alpha_2}_{\mathrm{P}}(\omega)\right] = \frac{ie^2}{2\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Delta\epsilon_{\mathbf{k}s's}\Delta f_{\mathbf{k}s's}\xi^{\alpha_1}_{\mathbf{k}s's}\xi^{\alpha_2}_{\mathbf{k}ss'}\left[\frac{\mathrm{P}}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}} - \frac{\mathrm{P}}{\hbar\omega + \Delta\epsilon_{\mathbf{k}ss'}}\right], \tag{B.4}$$

$$\mathrm{Im}\left[\sigma^{\alpha_1\alpha_2}_{\mathrm{P}}(\omega)\right] = \frac{e^2}{2\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Delta\epsilon_{\mathbf{k}s's}\Delta f_{\mathbf{k}s's}\xi^{\alpha_1}_{\mathbf{k}s's}\xi^{\alpha_2}_{\mathbf{k}ss'}\left[\frac{\mathrm{P}}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}} + \frac{\mathrm{P}}{\hbar\omega + \Delta\epsilon_{\mathbf{k}ss'}}\right]. \tag{B.5}$$

Note that we have used Eqs.(2.13) and (2.14). The other involves a Dirac delta function. In this case, the real and imaginary parts read as,

$$\mathrm{Re}\left[\sigma_{\delta}^{\alpha_1\alpha_2}(\omega)\right] = \frac{\pi e^2}{2\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Delta\epsilon_{\mathbf{k}s's}\Delta f_{\mathbf{k}s's}\xi^{\alpha_1}_{\mathbf{k}s's}\xi^{\alpha_2}_{\mathbf{k}ss'}\left[\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'})+\delta(\hbar\omega+\Delta\epsilon_{\mathbf{k}ss'})\right], \quad \text{(B.6)}$$

$$\mathrm{Im}\left[\sigma_{\delta}^{\alpha_1\alpha_2}(\omega)\right] = \frac{\pi e^2}{2i\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Delta\epsilon_{\mathbf{k}s's}\Delta f_{\mathbf{k}s's}\xi^{\alpha_1}_{\mathbf{k}s's}\xi^{\alpha_2}_{\mathbf{k}ss'}\left[\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'})-\delta(\hbar\omega+\Delta\epsilon_{\mathbf{k}ss'})\right]. \quad \text{(B.7)}$$

Imposing time-reversal symmetry — which implies that $\epsilon_{-\mathbf{k}s}=\epsilon_{\mathbf{k}s}$ and that the Berry connections can be chosen such that $\xi^{\alpha_1}_{-\mathbf{k}ss'}=\xi^{\alpha_1}_{\mathbf{k}s's}$ — we can write the product of Berry connections as,

$$\xi^{\alpha_1}_{\mathbf{k}s's}\xi^{\alpha_2}_{\mathbf{k}ss'} = \frac{1}{2}\left\{\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right\}+\frac{1}{2}\left[\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right], \quad \text{(B.8)}$$

the sum of contributions that have well-defined parities (under $\mathbf{k}\to-\mathbf{k}$). The anticommutator of Berry connections is an even function of $\mathbf{k}$,

$$\left\{\xi^{\alpha_1}_{-\mathbf{k}s's},\xi^{\alpha_2}_{-\mathbf{k}ss'}\right\} = \left\{\xi^{\alpha_1}_{\mathbf{k}ss'},\xi^{\alpha_2}_{\mathbf{k}s's}\right\} = \left\{\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right\}, \quad \text{(B.9)}$$

while the commutator is odd,[1]

$$\left[\xi^{\alpha_1}_{-\mathbf{k}s's},\xi^{\alpha_2}_{-\mathbf{k}ss'}\right] = \left[\xi^{\alpha_1}_{\mathbf{k}ss'},\xi^{\alpha_2}_{\mathbf{k}s's}\right] = -\left[\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right]. \quad \text{(B.12)}$$

We can, therefore, replace the product of Berry connections in Eqs.(B.4)–(B.7) by the anticommutator, $\left\{\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right\}$, as the terms that carry commutators necessarily vanish under an integration over the full FBZ. Note that the anticommutator is invariant under the exchange of band indexes, $s\leftrightarrow s'$, as is the product $\Delta\epsilon_{\mathbf{k}s's}\Delta f_{s's}$. Manipulating the band-index sums, we can show that two of the contributions in Eqs.(B.4)–(B.7) are automatically zero. Take, for example, Eq.(B.4),

$$\mathrm{Re}\left[\sigma_{\mathrm{P}}^{\alpha_1\alpha_2}(\omega)\right] = \frac{ie^2}{4\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Delta\epsilon_{\mathbf{k}s's}\Delta f_{\mathbf{k}s's}\left\{\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right\}\left[\frac{\mathrm{P}}{\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'}}-\frac{\mathrm{P}}{\hbar\omega+\Delta\epsilon_{\mathbf{k}ss'}}\right]. \quad \text{(B.13)}$$

[1] Conservation of probability imposes an additional property on the Berry connection: it must exchange indexes under complex conjugation, $\left(\xi^{\alpha_1}_{\mathbf{k}s's}\right)^* = \xi^{\alpha_1}_{\mathbf{k}ss'}$. We can then conclude that the anticommutator of Berry connections is strictly *real*,

$$\left(\left\{\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right\}\right)^* = \left\{\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right\}, \quad \text{(B.10)}$$

and that the commutator of Berry connections is strictly *imaginary*,

$$\left(\left[\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right]\right)^* = -\left[\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right]. \quad \text{(B.11)}$$

While the terms outside the brackets are invariant under an exchange of band indexes, those inside the brackets,

$$\frac{\mathrm{P}}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}} - \frac{\mathrm{P}}{\hbar\omega + \Delta\epsilon_{\mathbf{k}ss'}} \to \frac{\mathrm{P}}{\hbar\omega - \Delta\epsilon_{\mathbf{k}s's}} - \frac{\mathrm{P}}{\hbar\omega + \Delta\epsilon_{\mathbf{k}s's}}, \tag{B.14}$$

$$= \frac{\mathrm{P}}{\hbar\omega + \Delta\epsilon_{\mathbf{k}ss'}} - \frac{\mathrm{P}}{\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}}, \tag{B.15}$$

change their sign. The real part of $\sigma_{\mathrm{P}}^{\alpha_1\alpha_2}(\omega)$ cannot, therefore, contribute to the linear conductivity; the same is true for the imaginary part of $\sigma_{\delta}^{\alpha_1\alpha_2}(\omega)$.

Finally, note that $\mathrm{Im}\left[\sigma_{\mathrm{P}}^{\alpha_1\alpha_2}(\omega)\right]$ is symmetric under the exchange of Cartesian indexes — it involves an anticommutator of Berry connections — while its accompanying term in Eq.(B.2), i.e., the imaginary part of the product of field, is antisymmetric under the exchange of those indexes; their product must, therefore, be zero. The imaginary part of the linear conductivity does not not play a role in the energy absorption of a time-reversal symmetric cold semiconductor.[2] We can thus write the population density rate as,

$$\dot{n}(t) = \frac{2}{\hbar\omega}\mathrm{Re}\left[\sigma^{\alpha_1\alpha_2}(\omega)\right]\mathrm{Re}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\right]. \tag{B.16}$$

or, using Eq.(B.4), as,

$$\dot{n}(t) = \frac{\pi e^2}{\hbar}\mathrm{Re}\left[E^{\alpha_1}(-\omega)E^{\alpha_2}(\omega)\right]\int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s's}\Delta f_{\mathbf{k}s's}\left\{\xi^{\alpha_1}_{\mathbf{k}s's},\xi^{\alpha_2}_{\mathbf{k}ss'}\right\}\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}ss'}). \tag{B.17}$$

B.2 The Final Steps in the Derivation of the Shift Current

The integrand in Eq.(5.17), which was derived by means of a physical argument, is not gauge invariant, meaning that the corresponding current, as is, is not very physically consistent. In the subsection on the covariant derivative, we noted that the intraband Berry connection picks up a gradient term under a phase transformation, Eq.(3.58); we must, therefore, covariantize the expression for our integrand.

Consider the product of Berry connections in Eq.(5.17),

$$(\xi^{\beta}_{\mathbf{k}cc} - \xi^{\beta}_{\mathbf{k}vv})\left\{\xi^{\alpha_1}_{\mathbf{k}vc},\xi^{\alpha_2}_{\mathbf{k}cv}\right\}. \tag{B.18}$$

This can also be expressed as,

$$i\xi^{\alpha_1}_{\mathbf{k}vc}\xi^{\alpha_2}_{\mathbf{k}cv}(-i)(\xi^{\beta}_{\mathbf{k}cc} - \xi^{\beta}_{\mathbf{k}vv}) - i\xi^{\alpha_1}_{\mathbf{k}cv}\xi^{\alpha_2}_{\mathbf{k}vc}(-i)(\xi^{\beta}_{\mathbf{k}vv} - \xi^{\beta}_{\mathbf{k}cc}), \tag{B.19}$$

[2] Note, on the other hand, that the real part of the conductivity is *necessarily* zero when there is no energy absorption by the system, i.e., when $\dot{n}(t) = 0$, for $\hbar\omega < \Delta$.

by simply using $i(-i)=1$. Note that this is already quite similar to an expression that involves the generalized derivative. This object,

$$(\xi^{\alpha_2}_{\mathbf{k}cv})_{;\beta} = \big(\nabla^{\beta}_{\mathbf{k}}\xi^{\alpha_2}_{\mathbf{k}cv}\big) - i\big(\xi^{\beta}_{\mathbf{k}cc} - \xi^{\beta}_{\mathbf{k}vv}\big)\xi^{\alpha_2}_{\mathbf{k}cv}, \tag{B.20}$$

as we have noted in Subsection 4.4.1, is gauge covariant,

$$(\xi^{\alpha_2}_{\mathbf{k}cv})_{;\beta} \to e^{-i(\theta_c(\mathbf{k})-\theta_v(\mathbf{k}))}(\xi^{\alpha_2}_{\mathbf{k}cv})_{;\beta}. \tag{B.21}$$

Adding two derivative terms to Eq.(B.19),

$$i\xi^{\alpha_1}_{\mathbf{k}vc}(\nabla^{\beta}_{\mathbf{k}} + (-i)(\xi^{\beta}_{\mathbf{k}cc} - \xi^{\beta}_{\mathbf{k}vv}))\xi^{\alpha_2}_{\mathbf{k}cv} - i\xi^{\alpha_1}_{\mathbf{k}cv}(\nabla^{\beta}_{\mathbf{k}} + (-i)(\xi^{\beta}_{\mathbf{k}vv} - \xi^{\beta}_{\mathbf{k}cc}))\xi^{\alpha_2}_{\mathbf{k}vc}, \tag{B.22}$$

we obtain,

$$i\xi^{\alpha_1}_{\mathbf{k}vc}(\xi^{\alpha_2}_{\mathbf{k}cv})_{;\beta} - i\xi^{\alpha_1}_{\mathbf{k}cv}(\xi^{\alpha_2}_{\mathbf{k}vc})_{;\beta} = i\big[\xi^{\alpha_1}_{\mathbf{k}vc}, (\xi^{\alpha_2}_{\mathbf{k}cv})_{;\beta}\big], \tag{B.23}$$

a covariantized version of our integrand. The corresponding current is now physically consistent.

B.3 The Interband Contributions to the Third Order Conductivity of a Cold Semiconductor

The integrand function associated with the interband portion of Eq.(5.28) reads as follows,

$$-\sum_{s'\neq s}\big[D^{\alpha_1}, v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\big]_{s's}\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}\big[D^{\alpha_2}, \frac{(-i)}{\hbar\bar{\omega}_{3}-\Delta\epsilon}\circ\big[\xi^{\alpha_3}, \rho^{(0)}(\mathbf{k})\big]\big]_{ss'}.$$

Using the expression for the commutator of the covariant derivative with an arbitrary matrix in band-index space, Eq.(3.71), we can separate this interband portion into two contributions,

$$\begin{aligned} &i\big[D^{\alpha_1}, v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\big]_{s's}\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}\Delta f_{\mathbf{k}s's}\big(\nabla^{\alpha_2}_{\mathbf{k}}\frac{1}{\hbar\bar{\omega}_{3}-\Delta\epsilon_{\mathbf{k}ss'}}\xi^{\alpha_3}_{\mathbf{k}ss'}\big)\\ &+\big[D^{\alpha_1}, v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\big]_{s's}\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}\big[\xi^{\alpha_2}, \frac{1}{\hbar\bar{\omega}_{3}-\Delta\epsilon}\circ\big[\xi^{\alpha_3}, \rho^{(0)}(\mathbf{k})\big]\big]_{ss'}, \end{aligned} \tag{B.24}$$

since we are considering the response of a cold semiconductor, $\rho^{(0)}_{\mathbf{k}ss'} = \delta_{ss'}f_{\mathbf{k}s}$, for $f_{\mathbf{k}s} = 1(0)$ for s a valence (conduction) band. Let us take the first term in Eq.(B.24) and manipulate it further, so that it reads as,

$$\begin{aligned} &i\big[D^{\alpha_1}, v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\big]_{s's}\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}\frac{1}{(\hbar\bar{\omega}_{3}-\Delta\epsilon_{\mathbf{k}ss'})^2}\xi^{\alpha_3}_{\mathbf{k}ss'}\Delta f_{\mathbf{k}s's}\big(\nabla^{\alpha_2}_{\mathbf{k}}\Delta\epsilon_{\mathbf{k}ss'}\big)\\ &+i\big[D^{\alpha_1}, v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\big]_{s's}\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}\frac{1}{\hbar\bar{\omega}_{3}-\Delta\epsilon_{\mathbf{k}ss'}}\Delta f_{\mathbf{k}s's}\big(\nabla^{\alpha_2}_{\mathbf{k}}\xi^{\alpha_3}_{\mathbf{k}ss'}\big). \end{aligned} \tag{B.25}$$

The first term in Eq.(B.25) can then be expressed as,

$$
\begin{aligned}
& i\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon_{\mathbf{k}s's}}\frac{1}{(\hbar\bar{\omega}_{3}-\Delta\epsilon_{\mathbf{k}ss'})^2}\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}\left(\nabla_{\mathbf{k}}^{\alpha_1}v_{\mathbf{k}s's}^{\beta}\right)\left(\nabla_{\mathbf{k}}^{\alpha_2}\Delta\epsilon_{\mathbf{k}ss'}\right)\xi_{\mathbf{k}ss'}^{\alpha_3}\Delta f_{\mathbf{k}s's} \\
-i\frac{1}{(\hbar\bar{\omega}_{123}+\Delta\epsilon_{\mathbf{k}s's})^2}\frac{1}{(\hbar\bar{\omega}_{3}-\Delta\epsilon_{\mathbf{k}ss'})^2}\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}&\left(\nabla_{\mathbf{k}}^{\alpha_1}\Delta\epsilon_{\mathbf{k}s's}\right)v_{\mathbf{k}s's}^{\beta}\left(\nabla_{\mathbf{k}}^{\alpha_2}\Delta\epsilon_{\mathbf{k}ss'}\right)\xi_{\mathbf{k}ss'}^{\alpha_3}\Delta f_{\mathbf{k}s's} \\
&+\left[\xi^{\alpha_1},v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\right]_{s's}\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}\frac{1}{(\hbar\bar{\omega}_{3}-\Delta\epsilon_{\mathbf{k}ss'})^2}\left(\nabla_{\mathbf{k}}^{\alpha_2}\Delta\epsilon_{\mathbf{k}ss'}\right)\xi_{\mathbf{k}ss'}^{\alpha_3}\Delta f_{\mathbf{k}s's},
\end{aligned}
\tag{B.26}
$$

by using the fact that,

$$
\left[D^{\alpha_1},v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\right]_{s's}=\left(\nabla_{\mathbf{k}}^{\alpha_1}v_{\mathbf{k}s's}^{\beta}\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon_{\mathbf{k}s's}}\right)-i\left[\xi^{\alpha_1},v^{\beta}\circ\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon}\right]_{ss'}. \tag{B.27}
$$

Note that we have only isolated some of the contributions to Eq.(5.28), and that an analysis of the entire integrand function would involve many more terms. The terms that we did isolate are, nonetheless, representative of these other terms,[3] so that by proving that the terms in Eq.(B.26) do not contribute to $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}$, we will also show that the remaining terms — which we have not written explicitly — do not contribute as well.

Let us then turn to that first statement — that the terms in Eq.(B.26) do not contribute to $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}$. Consider the denominator structure of the first term in that expression,

$$
\frac{1}{\hbar\bar{\omega}_{123}+\Delta\epsilon_{\mathbf{k}s's}}\frac{1}{(\hbar\bar{\omega}_{3}-\Delta\epsilon_{\mathbf{k}ss'})^2}\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}. \tag{B.28}
$$

It corresponds to a contribution that, like all other contributions from Eq.(B.25), vanishes when $s=s'$. This means that none of the expressions in the denominators of Eq.(B.28) can go to zero when its frequencies, in whatever combination, go to zero; for the cold semiconductor, there is always a lower finite bound to the minimum energy, $\Delta\epsilon_{\mathbf{k}ss'}^{\text{min.}}=\Delta$. So, if there is a divergence associated with these terms, it must follow from the possible interplay between contributions from different denominators, when certain photon energies are *above* the gap. One could assume, at face value, that since these denominators, when treated independently, can (in the limit of $\eta\to 0$) be written as,

$$
\frac{1}{\hbar\bar{\omega}_{23}-\Delta\epsilon_{\mathbf{k}ss'}}\to\frac{1}{\hbar\omega_{23}-\Delta\epsilon_{\mathbf{k}ss'}}-i\pi\delta(\hbar\omega_{23}-\Delta\epsilon_{\mathbf{k}ss'}) \tag{B.29}
$$

that Eq.(B.28) would therefore contain contributions such as,

$$
\frac{1}{\hbar\omega_{3}-\Delta\epsilon_{\mathbf{k}ss'}}\frac{1}{\hbar\omega_{23}-\Delta\epsilon_{\mathbf{k}ss'}}(-i\pi)\delta(\hbar\omega_{3}-\Delta\epsilon_{\mathbf{k}ss'}), \tag{B.30}
$$

[3] By this we mean that the denominator structure in these terms is rather similar to that of those which have not been explicitly written. It is this structure that dictates whether or not a term is divergent and if so, if it contributes to the second order divergence, i.e., to $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}$.

that diverge when $\omega_2 \to 0$ and $\hbar\omega_3 \geq \Delta$. This argument is, however, wrong, for the simple reason that the denominators in Eq.(B.28) can only be treated independently if their different features do *not* overlap in any way. The $\eta \to 0$ limit, in which one separates the real and imaginary parts of an energy denominator, presupposes that the remaining terms in the integrand can be deemed relatively constant in a small energy region around the feature, i.e., around $\hbar\omega_3 = \Delta\epsilon_{\mathbf{k}ss'}$, in this case. Degenerate denominators have to be treated with this in mind. Consider the $\omega_2 \to 0$ limit of this denominator structure. It reads simply as,

$$\frac{1}{\hbar\bar{\omega}_{123} + \Delta\epsilon_{\mathbf{k}s's}} \frac{1}{(\hbar\bar{\omega}_3 - \Delta\epsilon_{\mathbf{k}ss'})^3}, \tag{B.31}$$

which can then be written in terms of,

$$-\frac{1}{2\hbar^2} \frac{1}{\hbar\bar{\omega}_{13} + \Delta\epsilon_{\mathbf{k}s's}} \big(\partial^2_{\omega_3} \frac{1}{\hbar\bar{\omega}_3 - \Delta\epsilon_{\mathbf{k}ss'}}\big). \tag{B.32}$$

This does not involve a divergence of any kind. In the case where $\omega_{13} \neq \omega_3$, we can write this denominator structure, in the $\eta \to 0$ limit, as the sum of three different terms,

$$\begin{aligned} &-\frac{1}{2\hbar^2} \frac{1}{\hbar\omega_{13} + \Delta\epsilon_{\mathbf{k}s's}} \big(\partial^2_{\omega_3} \frac{1}{\hbar\omega_3 - \Delta\epsilon_{\mathbf{k}ss'}}\big) \\ &+\frac{i\pi}{2\hbar^2} \delta(\hbar\omega_{13} + \Delta\epsilon_{\mathbf{k}s's}) \big(\partial^2_{\omega_3} \frac{1}{\hbar\omega_3 - \Delta\epsilon_{\mathbf{k}ss'}}\big) \\ &+\frac{i\pi}{2\hbar^2} \frac{1}{\hbar\omega_{13} + \Delta\epsilon_{\mathbf{k}s's}} \big(\partial^2_{\omega_3} \delta(\hbar\omega_3 - \Delta\epsilon_{\mathbf{k}ss'})\big). \end{aligned} \tag{B.33}$$

Note that term involving two Dirac delta functions vanishes automatically. If ω_1 also goes to zero, then the denominator structure in Eq.(B.28) would be expressed in terms of a third derivative with respect to ω_3. We can thus conclude that multiple interband denominators cannot give rise to divergences [21], meaning that the interband portion of Eq.(5.28) cannot contribute to $\Gamma^{\beta\alpha_1\alpha_2\alpha_3}$, and is thus irrelevant for our studies. It can, however, contribute to the first order divergence, and indeed it does.

B.4 Symmetrizing the Third Order DC Conductivity in the Context of a Divergent Response

As we have noted in the chapter on the general properties of nonlinear optical coefficients, the relevant physical object in a conductivity description of the response must satisfy intrinsic permutation

symmetry:

$$\tilde{\sigma}^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \frac{1}{3!}\big[\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) + \sigma^{\beta\alpha_2\alpha_1\alpha_3}(\omega_2,\omega_1,\omega_3) + \sigma^{\beta\alpha_1\alpha_3\alpha_2}(\omega_1,\omega_3,\omega_2) + \sigma^{\beta\alpha_3\alpha_2\alpha_1}(\omega_3,\omega_2,\omega_1) + \sigma^{\beta\alpha_3\alpha_1\alpha_2}(\omega_3,\omega_1,\omega_2) + \sigma^{\beta\alpha_2\alpha_3\alpha_1}(\omega_2,\omega_3,\omega_1)\big]. \tag{B.34}$$

As we are interested in singling out the contributions that carry a second order divergence, Eqs.(5.36)–(5.38), the unsymmetrized conductivities in the right-hand side (RHS) of Eq.(B.34) should be separated into three different types of terms, which are associated with three different frequency combinations:

1. when both ω_{123} and ω_{23} go to zero (or are much smaller than ω), the relevant contributions follow from the first and third terms on the RHS of Eq.(B.34), $\sigma^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3)$ and $\sigma^{\beta\alpha_1\alpha_3\alpha_2}(\omega_1,\omega_3,\omega_2)$;

2. when both ω_{123} and ω_{13} go to zero (or are much smaller than ω), the relevant contributions follow from the second and sixth terms on the RHS of Eq.(B.34), $\sigma^{\beta\alpha_2\alpha_1\alpha_3}(\omega_2,\omega_1,\omega_3)$ and $\sigma^{\beta\alpha_2\alpha_3\alpha_1}(\omega_2,\omega_3,\omega_1)$;

3. when both ω_{123} and ω_{12} go to zero (or are much smaller than ω), the relevant contributions follow from the fourth and fifth terms on the RHS of Eq.(B.34), $\sigma^{\beta\alpha_3\alpha_2\alpha_1}(\omega_3,\omega_2,\omega_1)$ and $\sigma^{\beta\alpha_3\alpha_1\alpha_2}(\omega_3,\omega_1,\omega_2)$;

This symmetrization procedure is identical for all three groups of terms, and so we can present that of the first group only. We note, however, that the end result of our calculation, Eq.(5.41), does contain the relevant contributions from all three groups, so as to satisfy intrinsic permutation symmetry.

Combining the relevant contributions from the conductivities in the first group, one obtains the following integrand function,

$$\sum_{s\neq s'}\frac{1}{6}\big(\nabla^{\beta}\nabla^{\alpha_1}\Delta\epsilon_{\mathbf{k}ss'}\big)\Delta f_{\mathbf{k}s's}\left(\xi^{\alpha_3}_{\mathbf{k}ss'}\xi^{\alpha_2}_{\mathbf{k}s's}\left[\frac{1}{\hbar\omega_3-\Delta\epsilon_{\mathbf{k}ss'}}-i\pi\delta(\hbar\omega_3-\Delta\epsilon_{\mathbf{k}ss'})\right] + \xi^{\alpha_2}_{\mathbf{k}ss'}\xi^{\alpha_3}_{\mathbf{k}s's}\left[\frac{1}{\hbar\omega_2-\Delta\epsilon_{\mathbf{k}ss'}}-i\pi\delta(\hbar\omega_2-\Delta\epsilon_{\mathbf{k}ss'})\right]\right). \tag{B.35}$$

Note that the integration associated with the denominators is a principal value one. Exchanging

$s' \leftrightarrow s$ in the terms with ω_2 gives us,

$$\sum_{s\neq s'} \frac{1}{6}\big(\nabla^{\beta}\nabla^{\alpha_1}\Delta\epsilon_{\mathbf{k}ss'}\big)\Delta f_{\mathbf{k}s's}\xi^{\alpha_3}_{\mathbf{k}ss'}\xi^{\alpha_2}_{\mathbf{k}s's} \\ \times\left[-i\pi\delta(\hbar\omega_3-\Delta\epsilon_{\mathbf{k}ss'})-i\pi\delta(-\hbar\omega_2-\Delta\epsilon_{\mathbf{k}ss'})+\frac{1}{\hbar\omega_3-\Delta\epsilon_{\mathbf{k}ss'}}+\frac{1}{\hbar\omega_2+\Delta\epsilon_{\mathbf{k}ss'}}\right] \tag{B.36}$$

Let us now consider the different cases in which we are interested. For the jerk current, Eq.(5.36) — $\omega_3=-\omega_2=\omega$, $\omega_1=0$ — we can see that the terms that carry Dirac delta functions combine,

$$-i\pi\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'})-i\pi\delta(-\hbar(-\omega)-\Delta\epsilon_{\mathbf{k}ss'}) = -2i\pi\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'}), \tag{B.37}$$

while those associated with principal value integrals cancel out,

$$\frac{1}{\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'}}+\frac{1}{-\hbar\omega+\Delta\epsilon_{\mathbf{k}ss'}}=0. \tag{B.38}$$

The same is true in the case of the dichromatic setup, Eq.(5.37) — $\omega_3=-\omega_2=\omega$ and $\omega_1=\delta$ — since only ω_1 changes, a frequency that is not included in the integrand function of Eq.(B.36). For the trichromatic setup, Eq.(5.38) — $\omega_3=\omega+\delta_1$, $\omega_2=-\omega$ and $\omega_3=\delta_2$ — the resonances do carry an additional term,

$$-i\pi\delta(\hbar(\omega+\delta_1)-\Delta\epsilon_{\mathbf{k}ss'})-i\pi\delta(-\hbar(-\omega)-\Delta\epsilon_{\mathbf{k}ss'})+\frac{1}{\hbar(\omega+\delta_1)-\Delta\epsilon_{\mathbf{k}ss'}}+\frac{1}{\hbar(-\omega)+\Delta\epsilon_{\mathbf{k}ss'}}\rightarrow \\ -2i\pi\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'})+\delta_1\left(\partial_\omega\frac{1}{\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'}}\right), \tag{B.39}$$

that does not, however, contribute to the second order divergence. From these considerations, one concludes that the contribution to the second order divergence from the first and third terms on the RHS of Eq.(B.34) should read (up to a normalization factor) as,

$$\frac{i\pi}{3}\frac{1}{\hbar\bar{\omega}_{123}}\frac{1}{\hbar\bar{\omega}_{23}}\int\frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s\neq s'}\big(\nabla^{\beta}\nabla^{\alpha_1}\Delta\epsilon_{\mathbf{k}ss'}\big)\xi^{\alpha_3}_{\mathbf{k}ss'}\xi^{\alpha_2}_{\mathbf{k}s's}\Delta f_{\mathbf{k}s's}\delta(\hbar\omega_3-\Delta\epsilon_{\mathbf{k}ss'}), \tag{B.40}$$

and carry a single, localized, contribution. Now, let us consider, as we did in the case of the injection and shift currents, Section 5.2, the consequences of imposing time-reversal symmetry. The product of Berry connections, Eq.(B.8), can then be separated into two different contributions with well-defined parities,

$$\xi^{\alpha_3}_{\mathbf{k}ss'}\xi^{\alpha_2}_{\mathbf{k}s's}=\frac{1}{2}\big\{\xi^{\alpha_3}_{\mathbf{k}ss'},\xi^{\alpha_2}_{\mathbf{k}s's}\big\}+\frac{1}{2}\big[\xi^{\alpha_3}_{\mathbf{k}ss'},\xi^{\alpha_2}_{\mathbf{k}s's}\big].$$

The anticommutator of Berry connections is even under $\mathbf{k}\rightarrow-\mathbf{k}$, Eq.(B.9), while the commutator is odd, Eq.(B.12). Note that, since time-reversal symmetry also imposes that $\epsilon_{-\mathbf{k}s}=\epsilon_{\mathbf{k}s}$, the

contribution to Eq.(B.40) carrying the commutator of Berry connections must necessarily vanish under an integration over the FBZ, as that integrand is *odd* under $\mathbf{k} \to -\mathbf{k}$. We are left with the anticommutator of Berry connections, which means that Eq.(B.40), i.e., the contributions to the second order divergence that come from the first group of terms, can be defined as a new κ coefficient,

$$\kappa^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \frac{i\pi}{3}\frac{1}{\hbar\bar{\omega}_{123}}\frac{1}{\hbar\bar{\omega}_{23}}\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega_3), \tag{B.41}$$

which is itself expressed in terms of a new object, $\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega)$,

$$\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega) = \int \frac{d^d\mathbf{k}}{(2\pi)^d}\sum_{s\neq s'}\left(\nabla^\beta\nabla^{\alpha_1}\Delta\epsilon_{\mathbf{k}ss'}\right)\left\{\xi^{\alpha_2}_{\mathbf{k}ss'},\xi^{\alpha_3}_{\mathbf{k}s's}\right\}\Delta f_{\mathbf{k}s's}\delta(\hbar\omega-\Delta\epsilon_{\mathbf{k}ss'}), \tag{B.42}$$

which, as we have already mentioned, carries a single, localized, contribution to the response. Note that this coefficient is rather similar to the one describing the injection current, but its integrand has a second derivative of the band energies — instead of a first derivative — so as to have (overall) four Cartesian indexes, and it carries an anti-commutator of Berry connections — instead of a commutator — so has to be even under $\mathbf{k} \to -\mathbf{k}$. This is an indication that the form of the response functions at a given order, n, can be, to a certain extent, guessed on the basis of the form of the response functions at order $n-1$, at least in the case where $\mathbf{k}$-parity arguments can be employed.[4] In addition, note that $\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega)$ satisfies the following properties,[5]

$$\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega) = (\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega))^*, \tag{B.43}$$
$$\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega) = \iota^{\beta\alpha_1\alpha_2\alpha_3}(-\omega), \tag{B.44}$$
$$\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega) = \iota^{\alpha_1\beta\alpha_2\alpha_3}(\omega), \tag{B.45}$$
$$\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega) = \iota^{\beta\alpha_1\alpha_3\alpha_2}(\omega), \tag{B.46}$$

whereas $\kappa^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3)$, Eq.(B.41), satisfies,

$$\kappa^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \kappa^{\alpha_1\beta\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3), \tag{B.47}$$
$$\kappa^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \kappa^{\beta\alpha_1\alpha_3\alpha_2}(\omega_1,\omega_2,\omega_3), \tag{B.48}$$
$$\kappa^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \kappa^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_3,\omega_2). \tag{B.49}$$

The first two properties of this coefficient follow directly from Eqs.(B.45) and (B.46). The third

[4] Note that this comment is of little practical use when we are thinking about deriving complete expressions for the nonlinear optical coefficients — as was done in ref.[10] in the case of the two-band model. The third order response is far too complex to be guessed from the form of the second order one. This derivation can, however, benefit from an *a priori* analysis of the different types of integrals that should appear in the third order response, which can be constructed out of those from the second order response.

[5] Of these four properties, the last three can be easily checked using the definition of ι. The first property, i.e., that the ι coefficient is real, can be checked using the fact that the anticommutator of Berry connections is necessarily real, Eq.(B.10).

one, on the other hand, follows from the condition that κ must carry a second order divergence — in which $\omega_3 \approx -\omega_2$ — and from the fact that the ι coefficient is an even function of the frequency, Eq.(B.44).

Repeating this procedure for the terms from the second and third groups, we finally obtain a symmetrized Γ coefficient, $\tilde{\Gamma}^{\beta\alpha_1\alpha_2\alpha_3}$, that reads as,

$$\tilde{\Gamma}^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \kappa^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) + \kappa^{\beta\alpha_2\alpha_3\alpha_1}(\omega_2,\omega_3,\omega_1) + \kappa^{\beta\alpha_3\alpha_1\alpha_2}(\omega_3,\omega_1,\omega_2), \tag{B.50}$$

$$\begin{aligned} &= \frac{1}{2}\big[\kappa^{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) + \kappa^{\beta\alpha_1\alpha_3\alpha_2}(\omega_1,\omega_3,\omega_2) + \kappa^{\beta\alpha_2\alpha_3\alpha_1}(\omega_2,\omega_3,\omega_1) \\ &\quad + \kappa^{\beta\alpha_2\alpha_1\alpha_3}(\omega_2,\omega_1,\omega_3) + \kappa^{\beta\alpha_3\alpha_1\alpha_2}(\omega_3,\omega_1,\omega_2) + \kappa^{\beta\alpha_3\alpha_2\alpha_1}(\omega_3,\omega_2,\omega_1)\big]. \end{aligned} \tag{B.51}$$

We emphasize the fact that the frequency conditions noted above ensure that only *one* contribution in Eq.(B.50) can be nontrivial for a given group of frequencies. Nonetheless, this expression is useful, since the current response in the time domain follows from an integration over all three frequencies, Eq.(4.7), and thus captures all three contributions. Alternatively, we can also write Eq.(B.50) as,

$$-\frac{3\hbar^2}{\pi e^4}\tilde{\Gamma}_{\beta\alpha_1\alpha_2\alpha_3}(\omega_1,\omega_2,\omega_3) = \frac{1}{\bar{\omega}_{123}}\frac{1}{\bar{\omega}_{23}}\iota^{\beta\alpha_1\alpha_2\alpha_3}(\omega_3) + \frac{1}{\bar{\omega}_{123}}\frac{1}{\bar{\omega}_{13}}\iota^{\beta\alpha_2\alpha_3\alpha_1}(\omega_1) + \frac{1}{\bar{\omega}_{123}}\frac{1}{\bar{\omega}_{12}}\iota^{\beta\alpha_3\alpha_1\alpha_2}(\omega_2). \tag{B.52}$$

B.5 The Shift-Current Coefficient and its Relation with the Injection-Current Coefficient

The injection-current coefficient, which we introduced in Subsection 5.4, Eq.(5.77), is connected with its shift-current counterpart, $I_2^{\alpha_1\alpha_2\beta}(\omega)$,

$$I_2^{\alpha_1\alpha_2\beta}(\omega) = \int \frac{d^d\mathbf{k}}{(2\pi)^d}\frac{1}{2}\big[\xi^{\alpha_1}_{\mathbf{k}vc},(\xi^{\alpha_2}_{\mathbf{k}cv})_{;\beta}\big]\Delta f_{\mathbf{k}vc}\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}cv}), \tag{B.53}$$

by means of the following relation,

$$\boxed{\hbar^{-1}(\partial_\omega I_1^{\alpha_1\alpha_2\beta}(\omega)) = I_2^{\alpha_1\alpha_2\beta}(\omega) - I_2^{\alpha_2\alpha_1\beta}(\omega)}. \tag{B.54}$$

Let us show that this is the case. Begin by considering the difference of commutators,

$$\big[\xi^{\alpha_1}_{\mathbf{k}vc},(\xi^{\alpha_2}_{\mathbf{k}cv})_{;\beta}\big] - \big[\xi^{\alpha_2}_{\mathbf{k}vc},(\xi^{\alpha_1}_{\mathbf{k}cv})_{;\beta}\big], \tag{B.55}$$

which we can write as a gradient of the commutator of Berry connections,

$$\xi^{\alpha_1}_{\mathbf{k}vc}(\nabla^\beta_{\mathbf{k}}\xi^{\alpha_2}_{\mathbf{k}cv}) + \xi^{\alpha_2}_{\mathbf{k}cv}(\nabla^\beta_{\mathbf{k}}\xi^{\alpha_1}_{\mathbf{k}vc}) - \xi^{\alpha_1}_{\mathbf{k}cv}(\nabla^\beta_{\mathbf{k}}\xi^{\alpha_2}_{\mathbf{k}vc}) - \xi^{\alpha_2}_{\mathbf{k}vc}(\nabla^\beta_{\mathbf{k}}\xi^{\alpha_1}_{\mathbf{k}cv}) = (\nabla^\beta_{\mathbf{k}}\big[\xi^{\alpha_1}_{\mathbf{k}vc},\xi^{\alpha_2}_{\mathbf{k}cv}\big]), \tag{B.56}$$

by using Eq.(B.20). Integrating by parts, we obtain, for the case of a cold semiconductor — $f_{\mathbf{k}v(c)} = 1(0)$,

$$I_2^{\alpha_1\alpha_2\beta}(\omega) - I_2^{\alpha_2\alpha_1\beta}(\omega) = -\int \frac{d^d\mathbf{k}}{(2\pi)^d}\frac{1}{2}\left[\xi^{\alpha_1}_{\mathbf{k}vc}, \xi^{\alpha_2}_{\mathbf{k}cv}\right]\Delta f_{\mathbf{k}vc}\left(\nabla^\beta_{\mathbf{k}}\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}cv})\right). \tag{B.57}$$

As the gradient of this Dirac delta function can be expressed in terms of a derivative with respect to the difference of band energies, or, alternatively, in terms of a derivative with respect to ω,

$$\left(\nabla^\beta_{\mathbf{k}}\delta(\omega - \Delta\epsilon_{\mathbf{k}cv})\right) = (\nabla^\beta_{\mathbf{k}}\Delta\epsilon_{\mathbf{k}cv})\left(\partial_{\Delta\epsilon_{\mathbf{k}cv}}\delta(\omega - \Delta\epsilon_{\mathbf{k}cv})\right) = -\frac{1}{\hbar}(\nabla^\beta_{\mathbf{k}}\Delta\epsilon_{\mathbf{k}cv})\left(\partial_\omega\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}cv})\right), \tag{B.58}$$

we can write the difference of shift-current coefficients as,

$$I_2^{\alpha_1\alpha_2\beta}(\omega) - I_2^{\alpha_2\alpha_1\beta}(\omega) = \frac{1}{\hbar}\int \frac{d^d\mathbf{k}}{(2\pi)^d}\frac{1}{2}\left[\xi^{\alpha_1}_{\mathbf{k}vc}, \xi^{\alpha_2}_{\mathbf{k}cv}\right](\nabla^\beta_{\mathbf{k}}\Delta\epsilon_{\mathbf{k}cv})\Delta f_{\mathbf{k}vc}\left(\partial_\omega\delta(\hbar\omega - \Delta\epsilon_{\mathbf{k}cv})\right), \tag{B.59}$$

as the derivative with respect to ω of the injection-current coefficient, Eq.(5.77). This coefficient, together with the shift-current coefficient, can be used to describe the entire real part of the (regular) second order conductivity tensor [10].[6] In addition, the latter satisfies the following properties,

$$I_2^{\beta\alpha_1\alpha_2}(\omega) = I_2^{\beta\alpha_2\alpha_1}(\omega), \tag{B.60}$$

$$I_2^{\beta\alpha_1\alpha_2}(-\omega) = I_2^{\beta\alpha_1\alpha_2}(\omega). \tag{B.61}$$

B.6 The 2D Toy Model

The Berry connections for this two-dimensional toy model are as follows,

$$\begin{aligned} a^{-1}\xi^y_{\mathbf{k}cv} = a^{-1}\xi^x_{\mathbf{k}vc} = {} & \frac{4\sqrt{2}(1+\cos(k_xa - k_ya))}{\bar{\Delta}^2\sqrt{8+\bar{\Delta}^2+8\cos(k_xa-k_ya)}} - \frac{1}{\sqrt{2}\bar{\Delta}^2}\sqrt{1+\cos(k_xa-k_ya)}\sqrt{8+\bar{\Delta}^2+8\cos(k_xa-k_ya)} \\ & - \frac{i}{\sqrt{2}\bar{\Delta}}\frac{\sin(k_xa-k_ya)}{\sqrt{1+\cos(k_xa-k_ya)}} + i\frac{4\sqrt{2}}{\bar{\Delta}}\frac{\sqrt{1+\cos(k_xa-k_ya)}\sin(k_xa-k_ya)}{8+\bar{\Delta}^2+8\cos(k_xa-k_ya)}, \end{aligned} \tag{B.62}$$

and, $\xi^y_{\mathbf{k}vc} = \xi^x_{\mathbf{k}cv} = (\xi^y_{\mathbf{k}cv})^*$, for $\bar{\Delta} = \Delta/t$. These objects, like the bands, Eq.(5.110), are functions of the difference between the crystal momenta along the x and y directions, $k_x - k_y$, which means that the integrands of the injection- and shift-current coefficients are also expressed in terms of $k_x - k_y$.

[6] Furthermore, the imaginary part of the (regular) second order conductivity is obtained by the Hilbert transforms of these coefficients [10].

We can then write any integral over the FBZ as,

$$\int_{-\pi/a}^{\pi/a}\frac{dk_y}{2\pi}\int_{-\pi/a}^{\pi/a}\frac{dk_x}{2\pi}g(k_x-k_y)\to\int_{-\pi/a}^{\pi/a}\frac{dk_y}{2\pi}\int_{-\pi/a-k_y}^{\pi/a-k_y}\frac{d\bar{k}_x}{2\pi}g(\bar{k}_x), \tag{B.63}$$

which, since $g(\bar{k}_x)$ is periodic, reduces to,

$$\int_{-\pi/a}^{\pi/a}\frac{dk_y}{2\pi}\int_{-\pi/a-k_y}^{\pi/a-k_y}\frac{d\bar{k}_x}{2\pi}g(\bar{k}_x)=\int_{-\pi/a}^{\pi/a}\frac{dk_y}{2\pi}\int_{-\pi/a}^{\pi/a}\frac{dk_x}{2\pi}g(\bar{k}_x)=\int_{-\pi/a}^{\pi/a}\frac{d\bar{k}_x}{2\pi}g(\bar{k}_x), \tag{B.64}$$

i.e., despite the two-dimensional character of this model, it allows for a one-dimensional description. Furthermore, by combining this fact with this model's rather simple band structure, we can show that the energy bands are easily inverted, and that we can solve this problem for whatever value of $\bar{k}_x$ (that is, for the full band),

$$\delta(\hbar\omega-\Delta\epsilon_{cv}(\bar{k}_x))=\frac{1}{4at}\frac{\hbar\bar{\omega}}{\sqrt{1-\left(\frac{(\hbar\bar{\omega})^2-\bar{\Delta}^2-8}{8}\right)^2}}\left[\delta(\bar{k}_x-\bar{k}_x^0)+\delta(\bar{k}_x+\bar{k}_x^0)\right], \tag{B.65}$$

for $\hbar\bar{\omega}=\hbar\omega/t$ and,

$$\bar{k}_x^0=\frac{1}{a}\arccos\left(\frac{(\hbar\bar{\omega})^2-\bar{\Delta}^2-8}{8}\right). \tag{B.66}$$

This allows one to compute the analytical expressions for the injection- and shift-current coefficients, Eqs.(5.77) and (B.53), rather easily [172].

Bibliography

[1] Gonçalo Ventura. Gauge invariance and nonlinear optical coefficients. Master's thesis, Faculdade de Ciências da Universidade do Porto, 2016.

[2] G. B. Ventura, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres. Gauge covariances and nonlinear optical responses. *Phys. Rev. B*, 96:035431, Jul 2017.

[3] D. J. Passos, G. B. Ventura, J. M. Viana Parente Lopes, J. M. B. Lopes dos Santos, and N. M. R. Peres. Nonlinear optical responses of crystalline systems: Results from a velocity gauge analysis. *Phys. Rev. B*, 97:235446, Jun 2018.

[4] Simão João. Non-linear optical response with the kernel polynomial method. Master's thesis, Faculdade de Ciências da Universidade do Porto, 2017.

[5] G. B. Ventura, D. J. Passos, J. M. Viana Parente Lopes, and J. M. B. Lopes dos Santos. A study of the nonlinear optical response of the plain graphene and gapped graphene monolayers beyond the dirac approximation. *Journal of Physics: Condensed Matter*, 2019.

[6] S M João and J M Viana Parente Lopes. Basis-independent spectral methods for non-linear optical response in arbitrary tight-binding models. *Journal of Physics: Condensed Matter*, 32(12):125901, dec 2019.

[7] J C G Henriques, G B Ventura, C D M Fernandes, and N M R Peres. Optical absorption of single-layer hexagonal boron nitride in the ultraviolet. *Journal of Physics: Condensed Matter*, 32(2):025304, oct 2019.

[8] G. B. Ventura, D. J. Passos, J. M. Viana Parente Lopes, and J. M. B. Lopes dos Santos. Second order divergence in the third order dc response of a cold semiconductor, 2020.

[9] G. B. Ventura, D. J. Passos, J. M. Viana Parente Lopes, and J. M. B. Lopes dos Santos. Comment on "jerk current: A novel bulk photovoltaic effect". *Phys. Rev. Lett.*, 126:259701, 2021.

[10] D. J. Passos, G. B. Ventura, J. M. Viana Parente Lopes, and J. M. B. Lopes dos Santos. Nonlinear optical conductivity of a two-band crystal i. *arXiv:2102.07629*, 2021.

[11] Y. R. Shen. *The Principles of Nonlinear Optics*. Wiley Classics Library, 2002.

[12] Robert W. Boyd. *Nonlinear Optics, Third Edition*. Academic Press, 3rd edition, 2008.

[13] Geoffrey New. *Introduction to Nonlinear Optics*. Cambridge University Press, 2011.

[14] Nasser Peyghambarian, Stephan W. Koch, and Andre Mysyrowicz. *Introduction to Semiconductor Optics*. Prentice-Hall, Inc., USA, 1994.

[15] Hartmut Haug and Stephan W Koch. *Quantum Theory of the Optical and Electronic Properties of Semiconductors*. WORLD SCIENTIFIC, 5th edition, 2009.

[16] N. Bloembergen. *Nonlinear Optics*. W. A. Benjamin, 1965.

[17] GIUSEPPE GROSSO and GIUSEPPE PASTORI PARRAVICINI. Xi - optical and transport properties in metals. In GIUSEPPE GROSSO and GIUSEPPE PASTORI PARRAVICINI, editors, *Solid State Physics*, pages 389 – 424. Academic Press, London, 2000.

[18] E. I. Blount. *Solid State Physics: Advances in Research and Applications*, volume 13. Academic, New York, 1962.

[19] L. D. Landau, L. P. Pitaevskii, and E. M. Lifshitz. *Statistical Physics, Part 2. Vol.9*. Butterworth-Heinemann, 1980.

[20] J. E. Sipe and E. Ghahramani. Nonlinear optical response of semiconductors in the independent-particle approximation. *Phys. Rev. B*, 48(16):11705–11722, 1993.

[21] Claudio Aversa, J. E. Sipe, M. Sheik-Bahae, and E. W. Van Stryland. Third-order optical nonlinearities in semiconductors: The two-band model. *Phys. Rev. B*, 50:18073–18082, Dec 1994.

[22] C. Aversa and J. E. Sipe. Nonlinear optical susceptibilities. *Phys. Rev. B*, 52(20):14636–14645, 1995.

[23] J. Sipe and A. Shkrebtii. Second-order optical response in semiconductors. *Phys. Rev. B*, 61(8):5337–5352, 2000.

[24] C. Y. Fong and Y. R. Shen. Theoretical studies on the dispersion of the nonlinear optical susceptibilities in gaas, inas, and insb. *Phys. Rev. B*, 12:2325–2335, Sep 1975.

[25] D. J. Moss, E. Ghahramani, J. E. Sipe, and H. M. Van Driel. Band-structure calculation of dispersion and anisotropy in $\chi \rightarrow^{(3)}$ for third-harmonic generation in si, ge, and gaas. *Phys. Rev. B*, 41(3):15–1990, 1989.

[26] E. Ghahramani, D. J. Moss, and J. E. Sipe. Full-band-structure calculation of second-harmonic generation in odd-period strained $(si)_n/(ge)_n$ superlattices. *Phys. Rev. B*, 43:8990–9002, Apr 1991.

[27] GIUSEPPE GROSSO and GIUSEPPE PASTORI PARRAVICINI. Xii - optical properties of semiconductors and insulators. In GIUSEPPE GROSSO and GIUSEPPE PASTORI PARRAVICINI, editors, *Solid State Physics*, pages 425–472. Academic Press, London, 2000.

[28] Henry M. van Driel and John E. Sipe. *Coherence Control of Photocurrents in Semiconductors*, pages 261–306. Springer New York, New York, NY, 2001.

[29] Zachary H. Levine and Douglas C. Allan. Calculation of the nonlinear susceptibility for optical second-harmonic generation in iii-v semiconductors. *Phys. Rev. Lett.*, 66:41–44, Jan 1991.

[30] W. Y. Ching and Ming-Zhu Huang. Calculation of optical excitations in cubic semiconductors. iii. third-harmonic generation. *Phys. Rev. B*, 47:9479–9491, Apr 1993.

[31] M. Murayama and T. Nakayama. Two-photon-absorption spectra originating from higher-energy transitions. *Phys. Rev. B*, 49:5737–5740, Feb 1994.

[32] Misao Murayama and Takashi Nakayama. Ab initio calculations of two-photon absorption spectra in semiconductors. *Phys. Rev. B*, 52:4986–4997, Aug 1995.

[33] Liang Z. Tan and Andrew M. Rappe. Enhancement of the bulk photovoltaic effect in topological insulators. *Phys. Rev. Lett.*, 116:237402, Jun 2016.

[34] Suman Raj Panday and Benjamin M. Fregoso. Strong second harmonic generation in two-dimensional ferroelectric iv-monochalcogenides. *Journal of Physics: Condensed Matter*, 29(43):43LT01, September 2017.

[35] Tonatiuh Rangel, Benjamin M. Fregoso, Bernardo S. Mendoza, Takahiro Morimoto, Joel E. Moore, and Jeffrey B. Neaton. Large bulk photovoltaic effect and spontaneous polarization of single-layer monochalcogenides. *Phys. Rev. Lett.*, 119:067402, Aug 2017.

[36] Chong Wang, Xiaoyu Liu, Lei Kang, Bing-Lin Gu, Yong Xu, and Wenhui Duan. First-principles calculation of nonlinear optical responses by wannier interpolation. *Phys. Rev. B*, 96:115147, Sep 2017.

[37] Julen Ibañez Azpiroz, Stepan S. Tsirkin, and Ivo Souza. Ab initio calculation of the shift photocurrent by wannier interpolation. *Phys. Rev. B*, 97:245143, Jun 2018.

[38] Julen Ibañez Azpiroz, Ivo Souza, and Fernando de Juan. Directional shift current in mirror-symmetric bc_2N. *Phys. Rev. Research*, 2:013263, Mar 2020.

[39] Jiewen Xiao and Binghai Yan. First principles calculations for topological quantum materials. *arXiv:2010.08371*, 2021.

[40] Daniel E. Parker, Takahiro Morimoto, Joseph Orenstein, and Joel E. Moore. Diagrammatic approach to nonlinear optical response with application to weyl semimetals. *Phys. Rev. B*, 99:045121, Jan 2019.

[41] Tobias Holder, Daniel Kaplan, and Binghai Yan. Consequences of time-reversal-symmetry breaking in the light-matter interaction: Berry curvature, quantum metric, and diabatic motion. *Phys. Rev. Research*, 2:033100, Jul 2020.

[42] Yoshihiro Michishita and Robert Peters. Correlation effect on nonlinear responses in strongly-correlated electron systems. *arXiv:2012.10603v2*, 2020.

[43] V.M. Asnin, A.A. Bakun, A.M. Danishevskii, E.L. Ivchenko, G.E. Pikus, and A.A. Rogachev. Circular photogalvanic effect in optically active crystals. *Solid State Communications*, 30(9):565–570, 1979.

[44] R. Atanasov, A. Haché, J. L. P. Hughes, H. M. van Driel, and J. E. Sipe. Coherent control of photocurrent generation in bulk semiconductors. *Phys. Rev. Lett.*, 76:1703–1706, Mar 1996.

[45] Rodrigo A. Muniz, Cuauhtémoc Salazar, Kai Wang, S. T. Cundiff, and J. E. Sipe. Quantum interference control of carriers and currents in zinc blende semiconductors based on nonlinear absorption processes. *Phys. Rev. B*, 100:075202, Aug 2019.

[46] D. A. B. Miller, C. T. Seaton, M. E. Prise, and S. D. Smith. Band-gap—resonant nonlinear refraction in iii-v semiconductors. *Phys. Rev. Lett.*, 47:197–200, Jul 1981.

[47] M. Sheik-Bahae, D. J. Hagan, and E. W. Van Stryland. Dispersion and band-gap scaling of the electronic kerr effect in solids associated with two-photon absorption. *Phys. Rev. Lett.*, 65:96–99, Jul 1990.

[48] Daniel B. S. Soh, Ryan Hamerly, and Hideo Mabuchi. Comprehensive analysis of the optical kerr coefficient of graphene. *Phys. Rev. A*, 94:023845, Aug 2016.

[49] Rodrigo A. Muniz and J. E. Sipe. Coherent control of optical injection of spin and currents in topological insulators. *Phys. Rev. B*, 89:205113, May 2014.

[50] J. Rioux, Guido Burkard, and J. E. Sipe. Current injection by coherent one- and two-photon excitation in graphene and its bilayer. *Phys. Rev. B*, 83:195406, May 2011.

[51] Rodrigo A. Muniz and J. E. Sipe. All-optical injection of charge, spin, and valley currents in monolayer transition-metal dichalcogenides. *Phys. Rev. B*, 91:085404, Feb 2015.

[52] C. Aversa and J. E. Sipe. Coherent current control in semiconductors: a susceptibility perspective. *IEEE Journal of Quantum Electronics*, 32(9):1570–1573, Sep. 1996.

[53] J. L. Cheng, J. E. Sipe, S. W. Wu, and Chunlei Guo. Intraband divergences in third order optical response of 2d systems. *APL Photonics*, 4(3):034201, 2019.

[54] J. L. Cheng, J. E. Sipe, and S. W. Wu. Third-order optical nonlinearity of three-dimensional massless dirac fermions. *ACS Photonics*, 7(9):2515–2526, September 2020.

[55] J. L. Cheng, N. Vermeulen, and J. E. Sipe. Second order optical nonlinearity of graphene due to electric quadrupole and magnetic dipole effects. *Scientific Reports*, 7(1):43843, March 2017.

[56] Daniel Kaplan, Tobias Holder, and Binghai Yan. Nonvanishing subgap photocurrent as a probe of lifetime effects. *Phys. Rev. Lett.*, 125:227401, Nov 2020.

[57] Thomas Garm Pedersen. Intraband effects in excitonic second-harmonic generation. *Phys. Rev. B*, 92:235432, Dec 2015.

[58] David J. Griffiths. *Introduction to Electrodynamics*. Prentice Hall, 1999.

[59] S. A. Mikhailov. Quantum theory of the third-order nonlinear electrodynamic effects of graphene. *Phys. Rev. B*, 93:085403, Feb 2016.

[60] G. B. Ventura. Studying the optical response of a sheet of graphene. Personal Notes, January 2019.

[61] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim. The electronic properties of graphene. *Rev. Mod. Phys.*, 81:109–162, Jan 2009.

[62] F. Bonaccorso, Z. Sun, T. Hasan, and A. C. Ferrari. Graphene photonics and optoelectronics. *Nature Photonics*, 4(9):611–622, sep 2010.

[63] N. M. R. Peres, F. Guinea, and A. H. Castro Neto. Electronic properties of disordered two-dimensional carbon. *Phys. Rev. B*, 73:125411, Mar 2006.

[64] L. A. Falkovsky and A. A. Varlamov. Space-time dispersion of graphene conductivity. *The European Physical Journal B*, 56(4):281–284, April 2007.

[65] L. A. Falkovsky. Optical properties of graphene. *Journal of Physics: Conference Series*, 129:012004, October 2008.

[66] S. A. Mikhailov. Non-linear electromagnetic response of graphene. *EPL (Europhysics Letters)*, 79(2):27002, 2007.

[67] S A Mikhailov and K Ziegler. Nonlinear electromagnetic response of graphene: frequency multiplication and the self-consistent-field effects. *Journal of Physics: Condensed Matter*, 20(38):384204, 2008.

[68] J L Cheng, N Vermeulen, and J E Sipe. Third order optical nonlinearity of graphene. *New Journal of Physics*, 16(5):053014, 2014.

[69] J. L. Cheng, N. Vermeulen, and J. E. Sipe. Third-order nonlinearity of graphene: Effects of phenomenological relaxation and finite temperature. *Phys. Rev. B*, 91:235320, Jun 2015.

[70] M. M. Glazov and S. D. Ganichev. High frequency electric field induced nonlinear effects in graphene. *Physics Reports*, 535(3):101–138, 2014.

[71] N. M. R. Peres, Yu. V. Bludov, Jaime E. Santos, Antti-Pekka Jauho, and M. I. Vasilevskiy. Optical bistability of graphene in the terahertz range. *Phys. Rev. B*, 90:125425, Sep 2014.

[72] S. A. Mikhailov. Nonperturbative quasiclassical theory of the nonlinear electrodynamic response of graphene. *Phys. Rev. B*, 95:085432, Feb 2017.

[73] J. L. Cheng, N. Vermeulen, and J. E. Sipe. Numerical study of the optical nonlinearity of doped and gapped graphene: From weak to strong field excitation. *Phys. Rev. B*, 92:235307, Dec 2015.

[74] Behrooz Semnani, Roland Jago, Safieddin Safavi-Naeini, Hamed Majedi, Ermin Malic, and Philippe Tassin. Anomalous optical saturation of low-energy dirac states in graphene and its implication for nonlinear optics. *2D Materials*, 6(3):031003, may 2019.

[75] Nathalie Vermeulen, David Castelló-Lurbe, JinLuo Cheng, Iwona Pasternak, Aleksandra Krajewska, Tymoteusz Ciuk, Wlodek Strupinski, Hugo Thienpont, and Jürgen Van Erps. Negative kerr nonlinearity of graphene as seen via chirped-pulse-pumped self-phase modulation. *Phys. Rev. Applied*, 6:044006, Oct 2016.

[76] Evdokia Dremetsika, Bruno Dlubak, Simon-Pierre Gorza, Charles Ciret, Marie-Blandine Martin, Stephan Hofmann, Pierre Seneor, Daniel Dolfi, Serge Massar, Philippe Emplit, and Pascal Kockaert. Measuring the nonlinear refractive index of graphene using the optical kerr effect method. *Opt. Lett.*, 41(14):3281–3284, Jul 2016.

[77] Evdokia Dremetsika and Pascal Kockaert. Enhanced optical kerr effect method for a detailed characterization of the third-order nonlinearity of two-dimensional materials applied to graphene. *Phys. Rev. B*, 96:235422, Dec 2017.

[78] Nathalie Vermeulen, David Castelló-Lurbe, Mulham Khoder, Iwona Pasternak, Aleksandra Krajewska, Tymoteusz Ciuk, Wlodek Strupinski, JinLuo Cheng, Hugo Thienpont, and Jürgen

Van Erps. Graphene's nonlinear-optical physics revealed through exponentially growing self-phase modulation. *Nature Communications*, 9(1):2675, July 2018.

[79] S. A. Mikhailov. Equations of macroscopic electrodynamics for two-dimensional crystals. *APL Photonics*, 4(3):034501, March 2019.

[80] N. A. Savostianova and S. A. Mikhailov. Optical kerr effect in graphene: Theoretical analysis of the optical heterodyne detection technique. *Phys. Rev. B*, 97:165424, Apr 2018.

[81] Daniel Kaplan, Tobias Holder, and Binghai Yan. Momentum shift current at terahertz frequencies in twisted bilayer graphene. *arXiv:2101.07539*, 2021.

[82] Takuya Higuchi, Christian Heide, Konrad Ullmann, Heiko B. Weber, and Peter Hommelhoff. Light-field-driven currents in graphene. *Nature*, 550:224, September 2017.

[83] Ibraheem Al-Naib, J. E. Sipe, and Marc M. Dignam. High harmonic generation in undoped graphene: Interplay of inter- and intraband dynamics. *Phys. Rev. B*, 90:245423, Dec 2014.

[84] Larisa A. Chizhova, Florian Libisch, and Joachim Burgdörfer. High-harmonic generation in graphene: Interband response and the harmonic cutoff. *Phys. Rev. B*, 95:085436, Feb 2017.

[85] Darko Dimitrovski, Lars Bojer Madsen, and Thomas Garm Pedersen. High-order harmonic generation from gapped graphene: Perturbative response and transition to nonperturbative regime. *Phys. Rev. B*, 95:035405, Jan 2017.

[86] Matthias Baudisch, Andrea Marini, Joel D. Cox, Tony Zhu, Francisco Silva, Stephan Teichmann, Mathieu Massicotte, Frank Koppens, Leonid S. Levitov, F. Javier Garcia de Abajo, and Jens Biegert. Ultrafast nonlinear optical response of dirac fermions in graphene. *Nature Communications*, 9(1):1018, March 2018.

[87] Andrea Tomadin, Sam M. Hornett, Hai I. Wang, Evgeny M. Alexeev, Andrea Candini, Camilla Coletti, Dmitry Turchinovich, Mathias Kläui, Mischa Bonn, Frank H. L. Koppens, Euan Hendry, Marco Polini, and Klaas-Jan Tielrooij. The ultrafast dynamics and conductivity of photoexcited graphene at different fermi energies. *Science Advances*, 4(5), 2018.

[88] Tatsuhiko N. Ikeda. High-order nonlinear optical response of a twisted bilayer graphene. *Phys. Rev. Research*, 2:032015, Jul 2020.

[89] J. L. Cheng, J. E. Sipe, and Chunlei Guo. Third harmonic generation of undoped graphene in hartree-fock approximation. *Phys. Rev. B*, 100:245433, Dec 2019.

[90] M. Grüning and C. Attaccalite. Second harmonic generation in h-bn and mos_2 monolayers: Role of electron-hole interaction. *Phys. Rev. B*, 89:081102, Feb 2014.

[91] C. Attaccalite and M. Grüning. Nonlinear optics from an ab initio approach by means of the dynamical berry phase: Application to second- and third-harmonic generation in semiconductors. *Phys. Rev. B*, 88:235113, Dec 2013.

[92] Claudio Attaccalite, Myrta Grüning, Hakim Amara, Sylvain Latil, and Fran çois Ducastelle. Two-photon absorption in two-dimensional materials: The case of hexagonal boron nitride. *Phys. Rev. B*, 98:165126, Oct 2018.

[93] Timothy C. Berkelbach, Mark S. Hybertsen, and David R. Reichman. Theory of neutral and charged excitons in monolayer transition metal dichalcogenides. *Phys. Rev. B*, 88:045318, Jul 2013.

[94] Alexey Chernikov, Timothy C. Berkelbach, Heather M. Hill, Albert Rigosi, Yilei Li, Ozgur Burak Aslan, David R. Reichman, Mark S. Hybertsen, and Tony F. Heinz. Exciton binding energy and nonhydrogenic rydberg series in monolayer ws_2. *Phys. Rev. Lett.*, 113:076802, Aug 2014.

[95] A J Chaves, R M Ribeiro, T Frederico, and N M R Peres. Excitonic effects in the optical properties of 2d materials:an equation of motion approach. *2D Materials*, 4(2):025086, apr 2017.

[96] Hanzhe Liu, Yilei Li, Yong Sing You, Shambhu Ghimire, Tony F. Heinz, and David A. Reis. High-harmonic generation from an atomically thin semiconductor. *Nature Physics*, 13(3):262–265, March 2017.

[97] Emilia Ridolfi, Caio H. Lewenkopf, and Vitor M. Pereira. Excitonic structure of the optical conductivity in mos_2 monolayers. *Phys. Rev. B*, 97:205409, May 2018.

[98] F. Ferreira, A. J. Chaves, N. M. R. Peres, and R. M. Ribeiro. Excitons in hexagonal boron nitride single-layer: a new platform for polaritonics in the ultraviolet. *J. Opt. Soc. Am. B*, 36(3):674–683, Mar 2019.

[99] Daniel B. S. Soh, Christopher Rogers, Dodd J. Gray, Eric Chatterjee, and Hideo Mabuchi. Optical nonlinearities of excitons in monolayer mos_2. *Phys. Rev. B*, 97:165111, Apr 2018.

[100] R. E. F. Silva, F. Martín, and M. Ivanov. High harmonic generation in crystals using maximally localized wannier functions. *Phys. Rev. B*, 100:195201, Nov 2019.

[101] K. S. Novoselov, A. Mishchenko, A. Carvalho, and A. H. Castro Neto. 2d materials and van der waals heterostructures. *Science*, 353(6298), 2016.

[102] Sajedeh Manzeli, Dmitry Ovchinnikov, Diego Pasquier, Oleg V. Yazyev, and Andras Kis. 2d transition metal dichalcogenides. *Nature Reviews Materials*, 2:15. 17033, 2017.

[103] G. B. Ventura. Exctions. Personal Notes, August 2019.

[104] Lun Yue and Mette B. Gaarde. Structure gauges and laser gauges for the semiconductor bloch equations in high-order harmonic generation in solids. *Phys. Rev. A*, 101:053411, May 2020.

[105] Nicola Marzari, Arash A. Mostofi, Jonathan R. Yates, Ivo Souza, and David Vanderbilt. Maximally localized wannier functions: Theory and applications. *Rev. Mod. Phys.*, 84:1419–1475, Oct 2012.

[106] Liang Z. Tan and Andrew M. Rappe. Upper limit on shift current generation in extended systems. *Phys. Rev. B*, 100:085102, Aug 2019.

[107] Alireza Taghizadeh, Kristian S. Thygesen, and Thomas G. Pedersen. Two-dimensional materials with giant optical nonlinearities near the theoretical upper limit. *ACS Nano*, March 2021.

[108] Alireza Taghizadeh and T. G. Pedersen. Gauge invariance of excitonic linear and nonlinear optical response. *Phys. Rev. B*, 97:205432, May 2018.

[109] Benjamin M. Fregoso, Rodrigo A. Muniz, and J. E. Sipe. Jerk current: A novel bulk photovoltaic effect. *Phys. Rev. Lett.*, 121:176604, Oct 2018.

[110] Benjamin M. Fregoso. Bulk photovoltaic effects in the presence of a static electric field. *Phys. Rev. B*, 100:064301, Aug 2019.

[111] Bernardo S. Mendoza and Benjamin M. Fregoso. Terahertz radiation of jerk photocurrent. *Phys. Rev. B*, 102:195410, Nov 2020.

[112] D. C. Hutchings, M. Sheik-Bahae, D. J. Hagan, and E. W. Van Stryland. Kramers-kronig relations in nonlinear optics. *Optical and Quantum Electronics*, 24(1):1–30, January 1992.

[113] F. L. Ridener and R. H. Good. Dispersion relations for third-degree nonlinear systems. *Phys. Rev. B*, 10:4980–4987, Dec 1974.

[114] N. P. Armitage, E. J. Mele, and Ashvin Vishwanath. Weyl and dirac semimetals in three-dimensional solids. *Rev. Mod. Phys.*, 90:015001, Jan 2018.

[115] Takahiro Morimoto and Naoto Nagaosa. Topological nature of nonlinear optical effects in solids. *Science Advances*, 2(5), 2016.

[116] Fernando de Juan, Adolfo G. Grushin, Takahiro Morimoto, and Joel E. Moore. Quantized circular photogalvanic effect in weyl semimetals. *Nature Communications*, 8(1):15995, July 2017.

[117] Shreyas Patankar, Liang Wu, Baozhu Lu, Manita Rai, Jason D. Tran, T. Morimoto, Daniel E. Parker, Adolfo G. Grushin, N. L. Nair, J. G. Analytis, J. E. Moore, J. Orenstein, and D. H.

Torchinsky. Resonance-enhanced optical nonlinearity in the weyl semimetal taas. *Phys. Rev. B*, 98:165113, Oct 2018.

[118] Habib Rostami and Marco Polini. Nonlinear anomalous photocurrents in weyl semimetals. *Phys. Rev. B*, 97:195151, May 2018.

[119] O. Matsyshyn and I. Sodemann. Nonlinear hall acceleration and the quantum rectification sum rule. *Phys. Rev. Lett.*, 123:246602, Dec 2019.

[120] Zachariah Addison and E. J. Mele. Nonlinear time-domain spectroscopy near a band inversion. *Phys. Rev. B*, 102:115206, Sep 2020.

[121] F. de Juan, Y. Zhang, T. Morimoto, Y. Sun, J. E. Moore, and A. G. Grushin. Difference frequency generation in topological semimetals. *Phys. Rev. Research*, 2:012017, Jan 2020.

[122] Junyeong Ahn, Guang-Yu Guo, and Naoto Nagaosa. Low-frequency divergence and quantum geometry of the bulk photovoltaic effect in topological semimetals. *Phys. Rev. X*, 10:041041, Nov 2020.

[123] J. Orenstein, J.E. Moore, T. Morimoto, D.H. Torchinsky, J.W. Harter, and D. Hsieh. Topology and symmetry of quantum materials via nonlinear optical responses. *Annual Review of Condensed Matter Physics*, 12(1):247–272, 2021.

[124] Hikaru Watanabe and Youichi Yanase. Chiral photocurrent in parity-violating magnet and enhanced response in topological antiferromagnet. *Phys. Rev. X*, 11:011001, Jan 2021.

[125] Iris Crassee, Julien Levallois, Andrew L. Walter, Markus Ostler, Aaron Bostwick, Eli Rotenberg, Thomas Seyller, Dirk van der Marel, and Alexey B. Kuzmenko. Giant faraday rotation in single- and multilayer graphene. *Nature Physics*, 7(1):48–51, January 2011.

[126] R. Shimano, G. Yumoto, J. Y. Yoo, R. Matsunaga, S. Tanabe, H. Hibino, T. Morimoto, and H. Aoki. Quantum faraday and kerr rotations in graphene. *Nature Communications*, 4(1):1841, May 2013.

[127] G. Catarina, J. Have, J. Fernández-Rossier, and N. M. R. Peres. Optical orientation with linearly polarized light in transition metal dichalcogenides. *Phys. Rev. B*, 99:125405, Mar 2019.

[128] G Catarina, N M R Peres, and J Fernández-Rossier. Magneto-optical kerr effect in spin split two-dimensional massive dirac materials. *2D Materials*, 7(2):025011, jan 2020.

[129] J. L. Cheng and C. Guo. Nonperturbative model for optical response under intense periodic fields with application to graphene in a strong perpendicular magnetic field. *Phys. Rev. B*, 97:205406, May 2018.

[130] L. E. Golub and E. L. Ivchenko. Circular and magnetoinduced photocurrents in weyl semimetals. *Phys. Rev. B*, 98:075305, Aug 2018.

[131] Pankaj Bhalla, Allan H. MacDonald, and Dimitrie Culcer. Resonant photovoltaic effect in doped magnetic semiconductors. *Phys. Rev. Lett.*, 124:087402, Feb 2020.

[132] Hikaru Watanabe and Youichi Yanase. Nonlinear electric transport in odd-parity magnetic multipole systems: Application to mn-based compounds. *Phys. Rev. Research*, 2:043081, Oct 2020.

[133] Michael Tinkham. *Group Theory and Quantum Mechanics*. Dover Publications, 2003.

[134] Mildred S. Dresselhaus, Gene Dresselhaus, and Ado Jorio. *Group Theory: Application to the Physics of Condensed Matter*. Springer Science, 2008.

[135] Giuseppe Grosso and Giuseppe Parravicini. *Solid State Physics*. Academic Press, 2013.

[136] P A D Gonçalves and N M R Peres. *An Introduction to Graphene Plasmonics*. WORLD SCIENTIFIC, 2016.

[137] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov. Electric field effect in atomically thin carbon films. *Science*, 306(5696):666–669, 2004.

[138] Willis E. Lamb. Fine structure of the hydrogen atom. iii. *Phys. Rev.*, 85:259–276, Jan 1952.

[139] Willis E. Lamb, Rainer R. Schlicher, and Marlan O. Scully. Matter-field interaction in atomic physics and quantum optics. *Phys. Rev. A*, 36:2763–2772, Sep 1987.

[140] K. Rzążewski and R. W. Boyd. Equivalence of interaction hamiltonians in the electric dipole approximation. *Journal of Modern Optics*, 51(8):1137, May 2004.

[141] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich. Generation of optical harmonics. *Phys. Rev. Lett.*, 7:118–119, Aug 1961.

[142] P. D. Maker, R. W. Terhune, M. Nisenoff, and C. M. Savage. Effects of dispersion and focusing on the production of optical harmonics. *Phys. Rev. Lett.*, 8:21–22, Jan 1962.

[143] J. A. Armstrong, N. Bloembergen, J. Ducuing, and P. S. Pershan. Interactions between light waves in a nonlinear dielectric. *Phys. Rev.*, 127:1918–1939, Sep 1962.

[144] V. N. Genkin and M. P. Mednis. Contribution to the Theory of Nonlinear Effects in Crystals with Account Taken of Partially Filled Bands. *Sov. Phys. JETP*, 27(4):609, 1968.

[145] Hung Cheng and P. B. Miller. Nonlinear optical theory in solids. *Phys. Rev.*, 134:A683–A687, May 1964.

[146] Lingyuan Gao, Zachariah Addison, E. J. Mele, and Andrew M. Rappe. Intrinsic fermi surface contribution to the circular photogalvanic effect. *arXiv:2011.06542*, 2020.

[147] Alexander Altland and Ben D. Simons. *Condensed Matter Field Theory*. Cambridge University Press, 2 edition, 2010.

[148] Radi A. Jishi. *Feynman Diagram Techniques in Condensed Matter Physics*. Cambridge University Press, 2013.

[149] N.W. Ashcroft and N.D. Mermin. *Solid State Physics*. HRW International Editions. Holt, Rinehart and Winston, 1976.

[150] E. Hendry, P. J. Hale, J. Moger, A. K. Savchenko, and S. A. Mikhailov. Coherent nonlinear optical response of graphene. *Phys. Rev. Lett.*, 105:097401, Aug 2010.

[151] Tao Jiang, Di Huang, Jinluo Cheng, Xiaodong Fan, Zhihong Zhang, Yuwei Shan, Yangfan Yi, Yunyun Dai, Lei Shi, Kaihui Liu, Changgan Zeng, Jian Zi, J. E. Sipe, Yuen-Ron Shen, Wei-Tao Liu, and Shiwei Wu. Gate-tunable third-order nonlinear optical response of massless dirac fermions in graphene. *Nature Photonics*, 12(7):430–436, July 2018.

[152] Nardeep Kumar, Jatinder Kumar, Chris Gerstenkorn, Rui Wang, Hsin-Ying Chiu, Arthur L. Smirl, and Hui Zhao. Third harmonic generation in graphene and few-layer graphite films. *Phys. Rev. B*, 87:121406, Mar 2013.

[153] Sung-Young Hong, Jerry I. Dadap, Nicholas Petrone, Po-Chun Yeh, James Hone, and Richard M. Osgood. Optical third-harmonic generation in graphene. *Phys. Rev. X*, 3:021014, Jun 2013.

[154] Han Zhang, Stéphane Virally, Qiaoliang Bao, Loh Kian Ping, Serge Massar, Nicolas Godbout, and Pascal Kockaert. Z-scan measurement of the nonlinear refractive index of graphene. *Opt. Lett.*, 37(11):1856–1858, Jun 2012.

[155] Xiaorui Zheng, Baohua Jia, Han Lin, Ling Qiu, Dan Li, and Min Gu. Highly efficient and ultra-broadband graphene oxide ultrathin lenses with three-dimensional subwavelength focusing. *Nature Communications*, 6(1):8433, September 2015.

[156] Anshuman Kumar, Andrei Nemilentsau, Kin Hung Fung, George Hanson, Nicholas X. Fang, and Tony Low. Chiral plasmon in gapped dirac systems. *Phys. Rev. B*, 93:041413, Jan 2016.

[157] J. B. Khurgin. Graphene: A rather ordinary nonlinear optical material. *Applied Physics Letters*, 104(16):161116, 2014.

[158] S. A. Mikhailov. Comment on "graphene: A rather ordinary nonlinear optical material". *Applied Physics Letters*, 111(10):106101, 2017.

[159] Jacob B. Khurgin. Response to "comment on 'graphene: A rather ordinary nonlinear optical material'". *Applied Physics Letters*, 111(10):106102, 2017.

[160] Corey Janisch, Yuanxi Wang, Ding Ma, Nikhil Mehta, Ana Laura Elias, Nestor Perea-Lopez, Mauricio Terrones, Vincent Crespi, and Zhiwen Liu. Extraordinary second harmonic generation in tungsten disulfide monolayers. *Nature: Scientific Reports*, 4:5530, July 2014.

[161] F. Hipolito, Thomas G. Pedersen, and Vitor M. Pereira. Nonlinear photocurrents in two-dimensional systems based on graphene and boron nitride. *Phys. Rev. B*, 94:045434, Jul 2016.

[162] Thomas Garm Pedersen. Nonlinear optical response of relativistic energy bands: Application to phosphorene. *Phys. Rev. B*, 95:235419, Jun 2017.

[163] F. Hipolito and T. G. Pedersen. Optical third harmonic generation in black phosphorus. *Phys. Rev. B*, 97:035431, Jan 2018.

[164] Nathan Youngblood, Ruoming Peng, Andrei Nemilentsau, Tony Low, and Mo Li. Layer-tunable third-harmonic generation in multilayer black phosphorus. *ACS Photonics*, 4(1):8–14, 2017.

[165] Alexandra Carvalho, Min Wang, Xi Zhu, Aleksandr S. Rodin, Haibin Su, and Antonio H. Castro Neto. Phosphorene: from theory to applications. *Nature Reviews Materials*, 1(11):16061, Nov 2016.

[166] F. Hipolito, Alireza Taghizadeh, and T. G. Pedersen. Nonlinear optical response of doped monolayer and bilayer graphene: Length gauge tight-binding model. *Phys. Rev. B*, 98:205420, Nov 2018.

[167] Dong-Hun Chae, Tobias Utikal, Siegfried Weisenburger, Harald Giessen, Klaus v. Klitzing, Markus Lippitz, and Jurgen Smet. Excitonic fano resonance in free-standing graphene. *Nano Letters*, 11(3):1379–1382, 2011. PMID: 21322607.

[168] Kin Fai Mak, Felipe H. da Jornada, Keliang He, Jack Deslippe, Nicholas Petrone, James Hone, Jie Shan, Steven G. Louie, and Tony F. Heinz. Tuning many-body interactions in graphene: The effects of doping on excitons and carrier lifetimes. *Phys. Rev. Lett.*, 112:207401, May 2014.

[169] Jacob B. Khurgin. Current induced second harmonic generation in semiconductors. *Applied Physics Letters*, 67(8):1113–1115, 1995.

[170] V.I. Belinicher, E.L. Ivchenko, and B.I. Sturman. Kinetic theory of the displacement photovoltaic effect in piezoelectric. *Sov. Phys. JETP*, 56(2):359, 1982.

[171] D. J. Passos. On a generalization of full permutation symmetry. Personal Notes, January 2020.

[172] G. B. Ventura. Full permutation symmetry and a relation between real parts of the second order conductivity tensor. Personal Notes, January 2020.

[173] D. J. Passos, G. B. Ventura, J. M. Viana Parente Lopes, J. M. B. Lopes dos Santos, and N. M. R. Peres. Tbd. *TBD*, 2018.

[174] F. Hipolito and T.G. Pedersen. Optical third harmonic generation in black phosphorus. *arXiv:1708.06553*, 2017.

[175] Behrooz Semnani, Amir Hamed Majedi, and Safieddin Safavi-Naeini. Nonlinear quantum optical properties of graphene. *Journal of Optics*, 18(3):035402, 2016.

[176] N. A. Savostianova and S. A. Mikhailov. Third harmonic generation from graphene lying on different substrates: optical-phonon resonances and interference effects. *Opt. Express*, 25(4):3268–3285, Feb 2017.

[177] F. Hipolito, A. Taghizadeh, and T. G. Pedersen. Nonlinear optical response of doped mono- and bilayer graphene: length gauge tight-binding model. *ArXiv e-prints*, February 2018.

[178] Jesse J. Dean and Henry M. van Driel. Graphene and few-layer graphite probed by second-harmonic generation: Theory and experiment. *Phys. Rev. B*, 82:125411, Sep 2010.

[179] Richard Ciesielski, Alberto Comin, Matthias Handloser, Kevin Donkers, Giovanni Piredda, Antonio Lombardo, Andrea C. Ferrari, and Achim Hartschuh. Graphene near-degenerate four-wave mixing for phase characterization of broadband pulses in ultrafast microscopy. *Nano Letters*, 15(8):4968–4972, 2015. PMID: 26121487.

[180] S. M. João and J. M. V. P. Lopes. Non-linear optical response of non-interacting systems with spectral methods. *arXiv e-prints*, page arXiv:1810.03732, October 2018.

[181] Pierluigi Cudazzo, Ilya V. Tokatly, and Angel Rubio. Dielectric screening in two-dimensional insulators: Implications for excitonic and impurity states in graphane. *Phys. Rev. B*, 84:085406, Aug 2011.

[182] Jacob B. Khurgin. Current induced second harmonic generation in semiconductors. *Applied Physics Letters*, 67(8):1113–1115, 1995.

[183] Benjamin M. Fregoso, Rodrigo A. Muniz, and J. E. Sipe. Erratum: The jerk current: A novel bulk photovoltaic effect, 2020.

[184] Tatsuya Kaneko, Zhiyuan Sun, Yuta Murakami, Denis GoleÅŸ, and Andrew J. Millis. Bulk photovoltaic effect driven by collective excitations in a correlated insulator, 2020.

[185] Haowei Xu, Hua Wang, Jian Zhou, and Ju Li. Pure spin photocurrent in non-centrosymmetric crystals: Bulk spin photovoltaic effect, 2020.

[186] S. A. Mikhailov and K. Ziegler. New electromagnetic mode in graphene. *Phys. Rev. Lett.*, 99:016803, Jul 2007.

[187] S. A. Mikhailov. Theory of the giant plasmon-enhanced second-harmonic generation in graphene and semiconductor two-dimensional electron systems. *Phys. Rev. B*, 84:045432, Jul 2011.

[188] S.A. Mikhailov. Theory of the nonlinear optical frequency mixing effect in graphene. *Physica E: Low-dimensional Systems and Nanostructures*, 44(6):924–927, 2012. The proceedings of the European Materials Research Symposium on Science and Technology of Nanotubes, Nanowires and Graphene.

[189] Ashutosh Singh, Kirill I. Bolotin, Saikat Ghosh, and Amit Agarwal. Nonlinear optical conductivity of a generic two-band system with application to doped and gapped graphene. *Phys. Rev. B*, 95:155421, Apr 2017.

[190] Aaron M. Burger, Radhe Agarwal, Alexey Aprelev, Edward Schruba, Alejandro Gutierrez-Perez, Vladimir M. Fridkin, and Jonathan E. Spanier. Direct observation of shift and ballistic photovoltaic currents. *Science Advances*, 5(1), 2019.

[191] Lei Chen, Jack Zuber, Zhongshui Ma, and Chao Zhang. Nonlinear optical response of the $\alpha-T_3$ model due to the nontrivial topology of the band dispersion. *Phys. Rev. B*, 100:035440, Jul 2019.

[192] Anton Autere, Henri Jussila, Yunyun Dai, Yadong Wang, Harri Lipsanen, and Zhipei Sun. Nonlinear optics with 2d layered materials. *Advanced Materials*, 30(24):1705963, 2018.

[193] Hassan A. Hafez, Sergey Kovalev, Klaas-Jan Tielrooij, Mischa Bonn, Michael Gensch, and Dmitry Turchinovich. Terahertz nonlinear optics of graphene: From saturable absorption to high-harmonics generation. *Advanced Optical Materials*, 8(3):1900771, 2020.

[194] Vl.A. Margulis, E.E. Muryumin, and E.A. Gaiduk. Electric-field-induced optical second-harmonic generation in doped graphene. *Solid State Communications*, 246:76–81, 2016.

[195] Vl A Margulis, E E Muryumin, and E A Gaiduk. Quadratic electro-optic kerr effect in doped graphene. *Journal of Optics*, 19(6):065505, may 2017.

[196] Martin Mittendorff, Torben Winzer, Ermin Malic, Andreas Knorr, Claire Berger, Walter A. de Heer, Harald Schneider, Manfred Helm, and Stephan Winnerl. Anisotropy of excitation and relaxation of photogenerated charge carriers in graphene. *Nano Lett.*, 14(3):1504–1507, March 2014.

[197] N. M. R. Peres, R. M. Ribeiro, and A. H. Castro Neto. Excitonic effects in the optical conductivity of gated graphene. *Phys. Rev. Lett.*, 105:055501, Jul 2010.

[198] Gang Wang, Alexey Chernikov, Mikhail M. Glazov, Tony F. Heinz, Xavier Marie, Thierry Amand, and Bernhard Urbaszek. Colloquium: Excitons in atomically thin transition metal dichalcogenides. *Rev. Mod. Phys.*, 90:021001, Apr 2018.

[199] Habib Rostami, Mikhail I. Katsnelson, Giovanni Vignale, and Marco Polini. Gauge invariance and ward identities in nonlinear response theory. *arXiv:2102.04425*, 2021.

[200] Ashutosh Singh, Saikat Ghosh, and Amit Agarwal. Nonlinear and anisotropic polarization rotation in two-dimensional dirac materials. *Phys. Rev. B*, 97:205420, May 2018.

[201] Benjamin M. Fregoso, Takahiro Morimoto, and Joel E. Moore. Quantitative relationship between polarization differences and the zone-averaged shift photocurrent. *Phys. Rev. B*, 96:075421, Aug 2017.