PhD-FSTC-2019-61
The Faculty of Sciences, Technology, and Communication

The Institute of Condensed Matter and Nanosciences

DISSERTATION

Defense held on 30/08/2019 in Luxembourg
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN PHYSIQUE

AND

DOCTEUR DE L'UNIVERSITÉ CATHOLIQUE DE LOUVAIN EN SCIENCE DE L'INGÉNIEUR

by
Francesco NACCARATO
Born on 10 July 1989 in Paola (Italy)

FIRST-PRINCIPLES HIGH-THROUGHPUT STUDY OF LINEAR AND NONLINEAR OPTICAL MATERIALS

Dissertation defence committee

Dr Ludger Wirtz, dissertation supervisor
Professor, Université du Luxembourg

Dr Gian-Marco Rignanese, dissertation supervisor
Professor, Université Catholique de Louvain

Dr Alexandre Tkatchenko, Chairman
Professor, Université du Luxembourg

Dr Xavier Gonze, Vice Chairman
Professor, Université Catholique de Louvain

Francesco Naccarato
First-principles high-throughput study of linear and nonlinear optical materials
Ph. D. Dissertation to obtain the degree of
Docteur en Science de l'Ingénieur
Université Catholique de Louvain
and
Docteur en Physique
University of Luxembourg
Chairperson: Prof. Alexandre Tkatchenko, University of Luxembourg
Secretary: Prof. Xavier Gonze, Université Catholique de Louvain
Reviewers: Prof. Silvana Botti, Friedrich-Schiller-Universität Jena
and Prof. Venkatraman Gopalan, Pennsylvania State University

Supervisors:

Gian-Marco Rignanese
Université Catholique de Louvain
Institute of Condensed Matter and Nanosciences
Modelling Unit

Ludger Wirtz
University of Luxembourg
Faculty of Science, Technology and Communication
Physics and Materials Science Research Unit

EJD FUNMAT
European Joint Doctorate in Functional Materials
Marie Sklodowska Curie Actions - Innovative Training Networks

EJD-FunMat is the European Joint Doctorate for Multifunctional Materials. EJD-FunMat is a European Project funded by the Horizon 2020 program under Marie Sklodowska Curie Actions, Innovative Training Networks. EJD-FunMat has started in June 2015 and will operate for 4 years, until May 2019. At the core of EJD-FunMat are 15 PhD projects, each of them co-supervised by two universities from different countries. The Ph.D.'s are grouped in 6 thematic clusters. Each cluster includes 3 or 4 academic, and 2 or 3 industry partners. EJD-FunMat offers joint training schools and workshops, covering both scientific topics and transferable skills.

Abstract

Nonlinear optical (NLO) processes, such as second harmonic generation (SHG), play an important role in modern optics, especially in laser-related science and technology. They are at the core of a wide variety of applications ranging from optoelectronics to medicine. Among the various NLO materials, insulators are particularly important for second-order NLO properties. In particular, only crystals which are non-centrosymmetric can display a non-zero second-order NLO susceptibility. However, given the large number of requirements that a material needs to meet in order to be a good nonlinear optical material, the choice of compounds is drastically limited. Indeed, despite recent progress, a systematic approach to design NLO materials is still lacking. In this work, we conduct a first-principles high-throughput study on a large set of semiconductors for which we computed the linear and nonlinear susceptibility using Density Functional Perturbation Theory. For the linear optical properties, our calculations confirm the general trend that the refractive index is roughly inversely proportional to the band gap. In order to explain the large spread in the data distribution, we have found that two descriptors successfully describe materials with relatively high refraction index: (i) a narrow distribution in energy of the optical transitions which brings the average optical gap close to the direct band gap (ii) a large number of transitions around the band edge and/or high dipole matrix elements. For non-centrosymmetric crystals, we perform the calculation of the efficiency of SHG. We observe some materials with particularly high SHG, much stronger than the general relation with the linear refraction index through Miller's rule predicts. We relate the value of Miller's coefficient to geometric factors, i.e., how strongly the crystal deviates from a centrosymmetric one. We also identified interesting materials that show high optical responses for which it would be worth performing further analysis.

Activities

Journal articles

- Francesco Naccarato, Francesco Ricci, Jin Suntivich, Geoffory Hautier, Ludger Wirtz, and Gian-Marco Rignanese, Physical Review Materials 3, 044602, *Searching for materials with high refractive index and wide band gap: A first-principles high-throughput study* (Published 5 April 2019)

Oral presentations

- EJD-FunMat training school, Bordeaux (France), 13-18/03/2016
- EJD-FunMat training school, Aveiro (Portugal), 23-31/03/2017
- 8^{th} Abinit developer workshop, Fréjus (France), 9-5/05/2017
- 14^{th} ETSF young researcher meeting, Tarragona (Spain), 4-9/06/2017
- APS March meeting 2018, Los Angeles (USA), 5-9/03/2018
- EJD-FunMat training school, Luxembourg (Luxembourg), 18-23/03/2018
- IMCN Ph.D. Day, Louvain la Neuve (Belgium), 25/05/2018
- 9^{th} Abinit developer workshop, Louvain-la-Neuve (Belgium), 20-22/05/2019

Poster presentations

- 13th ETSF Young Researcher Meeting, London (UK), 6-10/06/2016
- CAMD summer school, Copenhagen (Denmark), 14-19/08/2016
- EJD-FunMat training school, Aveiro (Portugal), 23-31/03/2017
- Advanced computing of excited state properties in solids and nanostructures with Yambo, Lausanne (Switzerland),23-28/04/2017
- IMCN Ph.D. Day, Louvain la Neuve (Belgium), 24/05/2017
- EJD-FunMat training school, Luxembourg (Luxembourg), 18-23/03/2018
- Nomad summer: A hands-on course on tools for novel-materials discovery, Lausanne (Switzerland), 24-27/09/2018
- IMCN Ph.D. Day, Louvain-la-Neuve (Belgium), 29/05/2019

Other contributions and activities

- QuantumWise research internship, 05/08/2016-10/09/2016
- Abinit: `www.abinit.org`
- Abipy: `github.com/abinit/abipy`
- Abiflows: `github.com/abinit/abiflows`

Acknowledgements

After four very intense years I have come to the end of a journey that has profoundly marked my life, and now is the time to gather thoughts and thank those who have been close to me over the years and helped me to become the person I am today.

First of all, I have to thank my Dad Antonio, my Mom Antonella, my amazing Sister Ilenia, and my little Brother Andrea for all the support they always give me in every decision I make and for everything they have taught me in my life. I hope the experience I had could be of great inspiration for Andrea. A special thank goes to Amedeo as well, because somehow he took my place when I had to leave to follow my dreams and career. Another special thank goes to my grandparents: Nonno Peppe, Nonna Adelina, Nonna Esterina and Nonno Fiore. I have always felt you at my side.

I will always be attached to my roots and grateful to my friends from Amantea: Mario, Flavio, Puli, Congi, Arturo, Marian, Peppe, Danny, Botti, Matteo, Fabio, Serena, Gaja, Rito, Twister, Pasky, Marco, Claudio, Benzo . . . Guys, you are really a lot, and it is almost impossible to mention everybody. Just let me say thanks to all my friends from Pagliarone.

I cannot forget how all of this started, and a special thank goes to Sara, because I am not sure I would be here without her, and to Ambra, for the support she gave me during this experience, even though our roads separated now.

Going now to this four years I spent between Belgium and Luxembourg, I met a lot of people that I would like to thank, and once again it is almost impossible to mention everybody. First of all, let me just start thanking my supervisors Gian-Marco and Ludger that gave me the possibility of doing this amazing experience called

Ph.D. The discussions we had and all the tips you gave me made me grow a lot. Today I feel myself a new person compared to the guy that came for the first time in 2015 to visit Luxembourg and Louvain-la-Neuve, and you had a played an important role in this. Another special thank goes to Geoffroy, Xavier, Alexandre, Silvana, and Venkat for all the scientific discussions we had that gave me an important professional grow.

A special thank goes to all my colleagues and all the staff from the University of Luxembourg and from Université Catholique de Louvain for the scientific discussions, help, fun, and beers: Fulvio, Engin, Henrique, Matteo B., Sven , Thomas, Alejandro, Sandeep, Amir, Fred, Stéphanie, Vinciane, Francesco R., Aurélie, Guido, Samuel, Benoit, Yannick, Guillaume, Matteo G., Vishank, David, Viêt-Anh, David, Sadia, Jean-Michel, Georgios, Martha, Yasmine, Nicolas, Simon, Carina, Janine, Tushar, Volodia, Jan, Aurélien, Maksim, Sergio.

The second-year I spent in Luxembourg gave me the occasion to met amazing people that I want to thank for all the fun we had together: Maria, Ace, Momo, Andrés, Kobi, Jennifer, Andrea, Filippo, and Emiliano. Belval is not so boring if you find the right people. Of course I cannot forget my amazing flatmate Claudia and all the crazy experiences we had in this last year.

I would also like to say thank you to the guys from Greenta: Giacomo, Pietro, Valerio, Alessandro, Vincent, Ernani, Ivano, and Fred. You made me rediscover my passion for football and being in this team is one of the best things happened in this last year.

Last but not least, I would like to spend a last word thanking teh people with whom I shared the EJD-FunMat experience as we have all faced tough and peculiar challenges: Bruno, Danila, Catarina, Mirasbek, Kenny, Sara, Getnet, Prya, Philip, Yannick, Raphaël,

Guillaume, and Cong.
Finally, I would also like to thank the EJD-FunMat network (European Unions Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement N° 641640) for funding this project, and all the computational facilities provided by the Université Catholique de Louvain (CISM/UCL), the University of Luxembourg, and by the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI).

Contents

List of Figures

List of Tables

1

Introduction

The beginning of the 21^{st} century has been characterized by new opportunities and challenges for material scientists. One of the main steps in this direction was the discovery of graphene in 2004, that was worth the Physics Nobel prize in 2010 to Andre Geim and Konstantin Novoselov. The great potential of this material opened a new highway for materials science and raised new questions to the scientific community that somehow can be summarized with the following: are the materials currently used in technological applications the best we can have? The question is, of course, much deeper than it may seem. Materials are indeed technology enablers, and the discovery and commercialization of advanced materials are crucial to solving major challenges in technological innovation, economic growth, and the environment.

In the attempt to address this question, in 2011 the Materials Genome Initiative (MGI) was announced. This is an effort to introduce new materials into commerce more quickly and at a lower cost than it could be achieved by a purely experimental trial and error approach [14–16]. The MGI approach requires contributions in three critical areas: computational tools, experimental tools, and digital data. Focusing on the first critical area, the growing computational power at our disposition made it possible to combine first-principles calculations with the high-throughout (HT) approach.

The combination of these two methods enabled the creation of large databases of material properties that would be prohibitive (in time and cost) for experimental measurements. By screening those databases, new materials, targeting specific applications, can be identified. Successful discoveries (i.e. predictions confirmed in the lab) include materials for batteries, hydrogen production and storage, thermoelectrics and photovoltaics (see, e.g., Refs. [17] and [18]). Databases can also be analyzed using data mining techniques, aiming at identifying trends that can give a further insight into the comprehension of the materials properties, or even make predictions for unknown compounds through Machine Learning (see, for example, Refs. [19] and [20]). As an obvious effect, the MGI has made significant progress in predicting the structures and properties of new functional materials through computational simulation and modelling, since its inception [21].

We are just in the first decade of this scientific era that can bring interesting developments in material science. Indeed, critical technologies such as energy production and utilization, microelectronics, and catalysis await immediate materials solutions through the discovery and development of higher performance photovoltaic (PV), thermoelectric, energy storage, fuel cell, semiconductor, catalytic materials and optical materials. The latter are the main focus of this work. Important technology applications take benefit from such materials, lead to innovations, for instance, in lasers, LCD, LED technologies. Somehow, we are quite often dealing with optical materials in our daily lives, sometimes without even realizing it. As a matter of fact, the development of technologies to which optical materials are associated has permanently changed communications and human interactions.

Further, to highlight the growing interest of scientists in optical materials, it is worth to name a few recent events. Indeed, UNESCO designated the year 2015 as the international Year of Light and Light-Based Technologies. More recently (October 2018), Gérard Mourou, Arthur Ashkin, and Donna Strickland, were awarded the Physics Nobel prize for their groundbreaking inventions in the field of laser physics, for the optical tweezers, their application to biological systems, and for their method of generating high-intensity, ultra-short optical pulses.

The interest in high-performing optical materials continues to increase in many sectors, going from the industry up to the academic context. Indeed, despite the availability of high-performance computer and advanced synthetic methods to control the structure of the materials at the atomic structure, there are still several key obstacles to overcome. One of the main reasons is that the optimization of new materials is quite complex and requires careful evaluation of several factors. Among these factors, we have for instance mechanical properties, thermal stability, and potential for large-scale fabrication. Obviously, the cost of this process is also another key factor to be evaluated.

However, as we have already stated, much progress has been done lately to overcome the above mentioned limiting factors, pushing forward the materials research. As a matter of fact, first-principles approaches as Density Functional Theory (DFT) are being widely used nowadays. This has the additional benefit of accessing many fundamental materials properties that can be difficult to obtain experimentally. However, DFT is a theory built for the prediction of the ground-state properties of the materials, and the calculation of the excited state properties in crystals continues to be an active area of research. One of the main problems is due to the limitations that the current methods for the calculation of the optical response of the materials suffer, both in the linear and in the nonlinear regime. Indeed, at the present moment, methods that go beyond DFT, such as many-body perturbation theory, for the calculation of the optical response still need a large availability of computational resources. This makes it prohibitive to calculate the optical properties for a large set of materials. Furthermore, so far these methods have been mostly used to deepen the experimental results rather than being used to make predictions.

Another important role in this field is given by the interaction through the exchange of data between computational predictions and experimental results. In fact, knowing that certainly, the experimental validation is essential as a final confirmation of the quality of the optical materials, reliable and robust computational approaches can significantly improve and speed up the search of novel materials, alleviating some of the key limiting factors.

Given all these considerations, contributions that can help material design and accelerate the material discovery are really sought. To this purpose, in this thesis, we show an high-throughput first-principles study of optical materials. In this work, most of the optical properties of materials are computed in the framework of Density Functional Perturbation Theory (DFPT). Indeed, by merging DFT within perturbation theory, new quantities beyond the ground-state properties of the materials can be accessed. Considering thus an electric field as the external perturbation, we can compute the response of the materials to such perturbation, in the linear and nonlinear regime, at the static limit ($\omega = 0$ eV). As already stated, the success of DFT and DFPT stems from their reliability and low computational cost. We then combined these methodologies with the HT approach, calculating the optical responses for a large set of semiconductors. The analysis performed on these materials are basically focused on two aspects: (i) the search for novel optical materials, by screening, among the list of materials we have computed, the property requirements that a materials should possess to be a good candidate, among the list of materials we have computed and (ii) the analysis of the data trend through data mining approaches that, giving a description of the data distribution, can also give us further physical insight into the comprehension of the investigated phenomena.

We then compute the linear optical properties, such as the refractive index, for more than 4000 semiconductors. Our calculated data confirms the global inverse trend between this quantity and the band gap of the materials. This inverse relationship has been highlighted also in similar works [22, 23]. In their analysis, Petousis *et al.* tried to fit various quantitative relationships between the refractive index and the band gap to describe the data distribution. However, due to the spread in the values of the refractive index for a given band gap, it was not possible to establish such a quantitative relation between these quantities. Furthermore, it is worth mentioning that our calculated data constitute the largest dataset of linear optical properties to date, to the best of our knowledge.

Aiming at finding outlier materials that combine a high refractive index with a wide band gap, we found well known materials (TiO_2, $LiNbO_3$, ...), already widely used for optical applications, and other materials, not yet considered for such applications (Ti_3PbO_7, $LiSi_2N_3$, BeS, ...). By investigating the data trend of the materials we found two descriptors: the average optical gap and the effective frequency. The former can be determined directly from the electronic structure of the compounds, but the latter cannot. This calls for further analysis in order to obtain a predictive model. Nonetheless, it turns out that the negative effect of a large band gap on the refractive index can counterbalanced in two ways: (i) by limiting the difference between the direct band gap and the average optical gap which can be realized by a narrow distribution in energy of the optical transitions and (ii) by increasing the effective frequency which can be achieved through either a high number of transitions from the top of the valence band to the bottom of the conduction or a high average probability for these transitions. Focusing on oxides, we use our data to investigate how the chemistry influences this inverse relationship and rationalize why certain classes of materials would perform better. This work has been published in Physical Review Materials [24].

We select the non-centrosymmetric materials present in the linear optical materials database, to start the screening of the nonlinear optical materials aiming at computing their second order susceptibility. At this stage, we have computed the nonlinear properties for more than 400 semiconductors, which have been inserted in a new database. This number is intended to increase in the future and the database will be released for public consultation at a later stage. We then investigate the relationship between the nonlinear and linear optical properties of the materials. As suggested by the empirical Miller's rule [10], compounds with high nonlinear optical properties also show high response in the linear regime. However, having a high response in the linear regime is not a sufficient condition to have a high nonlinear response as well. From our analysis of the data distribution, it turns out that the Miller's coefficient, that is the coefficient of proportionality between the nonlinear and the linear susceptibility, is material dependent and that a particular atomic distribution can favor the nonlinear process. These findings will be further analyzed in future works.

Finally, we look at the distribution of materials considering the nonlinear optical coefficient vs. the band gap. As in the case of the linear optical properties, there is an inverse relationship between these two quantities. Among the outliers that combine a wide band gap with high nonlinear optical coefficients, we have found some of the already mentioned nonlinear optical materials such as $LiNbO_3$, $LiTaO_3$, and other well-known as $KTiPO_5$ and $TiPbO_3$. In contrast, to the best of our knowledge, some of these outliers have not yet been considered as nonlinear optical materials (as for instance PON, B_2O_3, PbO, Sb_2WO_6, ..).

This manuscript is organized as follows:

- In the second chapter, the scientific and theoretical background is given. An overview of optical materials is given, focusing then on the criteria that a material should have to be a good optical material, followed by a theoretical overview of these aspects.

- In the third chapter the methods we used to compute the structural, electronic, and optical properties are given. These are the already mentioned DFT and DFPT. In the last part of this section, an overview on high-throughput calculations is given, focused mainly on the description of how DFT and DFPT can be merged in such an approach.

- In the fourth chapter our results are shown and analyzed. Using the HT-DFPT approaches explained in the third chapter, we initially built up a database containing more than 4000 materials, for which we stored the optical properties in the linear regime. After this analysis, non-centrosymmetric materials have been extracted from the initial list of candidates, to build a new set of data, for which we have computed the nonlinear optical coefficients for more than 400 semiconductors.

- In chapter 5, we summarize our work and give an outlook to future work in this field.

2

Linear and nonlinear optics: Theoretical background

Optical materials are at the core of many technologies (e.g. lasers, liquid-crystal displays, light-emitting diodes, ...) with applications in many sectors (telecommunications, medicine, energy, transistors, microelectronics, ...). Light-matter interaction at the atomic scale, is a fundamental aspect for this class of materials. Improvement and further developments of these technologies requires a thorough comprehension of the underlying physical processes and how optical properties are linked to the structure and the electronic structure of the compounds. The possible range of phenomena can be generally grouped into reflection, propagation, and transmission, that raise of course different responses and different fields of application. The improvement of the efficiency of available devices, or the development of new technologies, has to go through the comprehension of these aspects. The aim is that, if one can more readily accelerate the rate at which basic condensed matter discoveries are made, then translation of such materials research into new technologies may be more efficient.

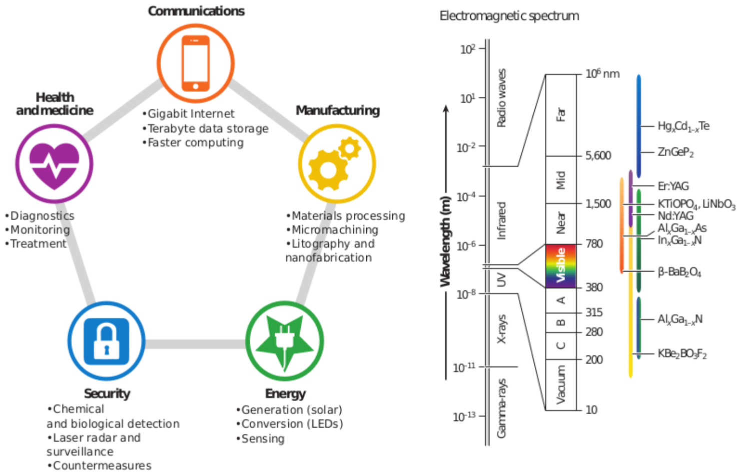
In this chapter, all the fundamental aspects of optical materials are given starting from a general overview. Then, the main ingredients to describe optical materials are shown, starting from a classical picture and going to a quantum mechanical description of the phenomena, considering both the linear and nonlinear regime. Finally a method to exploit phase-matching condition in bi-axial crystals is shown.

2.1 Optical materials: An overview

Optical materials are all those materials used in the construction of technological devices whose functioning is given by the alteration of electromagnetic radiation. Interacting with these materials, photons can be reflected, they can propagate through the material, or eventually they can be transmitted if the photons go across of the material. In the case of propagation, many phenomena can occur since photons can be either absorbed and converted to other forms of energy, they can be refracted by changing direction and velocity of propagation, or they can be simply randomly scattered in the case of an opaque material.

Optical materials are essential for economic growth, enabling advanced manufacturing and many next generation innovative technologies. As shown in Fig. 2.1, taken from Ref. [1], optical materials are used in a variety of industries and engineering fields. In communications, optical materials are critical for storing, transferring, and computing information, whereas in manufacturing they enable cutting, welding, and drilling of materials as well as precise photolithography with features less than 200 nm. Optical materials continue today to be critical for a balanced energy portfolio, in which they serve as the active elements in generation and conversion technologies. These materials provide security and defense applications as high-power direct energy weapons and laser guidance; in health and medicine,they enable cancer detection, needle less testing, and vision correction.

Figure 2.1: Fields of usage for technological applications of optical materials (Taken from Ref. [1]).



To improve the performance of optoelectronic devices, the optical properties of the given materials have to be investigated. A compilation of articles about recent developments of optoelectronic devices can be found in Ref. [25]. The linear optical properties are involved for example in solar cells, where the absorption of sunlight is the first step in the energy conversion mechanism, and in LED, where the radiative recombination leads to light emission.

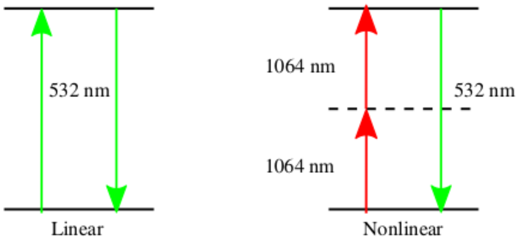
Effect	Input	Output
Second Harmonic Generation (SHG)	ω	2ω
Sum Frequency Generation (SFG)	ω_1, ω_2	$\omega_1 + \omega_2$
Difference Frequency Generation	ω_1, ω_2	$\omega_1 - \omega_2$
Optical rectification	ω	0
Down conversion	ω	ω_1, ω_2
Linear electro-optic effect	ω, 0	ω

Table 2.1: Second order nonlinear optical phenomena. The input column indicates the frequency of the light beam incident on the nonlinear crystal for each effect. The output column represents the frequency of the output beam. A frequency of zero indicates a DC electric field.

As light with sufficient intensity propagates through a material, the light-matter interaction results in nonlinear optical (NLO) effects that find utility in various frequency conversion technologies. Table 2.1 illustrates several nonlinear processes that involve the interaction of photons within a nonlinear crystal.

In this manuscript, we will focus on the phenomena of Second Harmonic Generation (SHG). As it is well known, in a linear process, a photon is absorbed at a frequency ω and is then re-emitted at the same frequency. In SHG, new frequencies are created. Indeed, after the absorption of a first photon, a virtual process happens in the nonlinear material that is capable of absorbing another photon at the same frequency, emitting then at double the frequency. These processes are naively depicted in Fig. 2.2. The situation depicted for SHG, where two photons with wavelength of 1064 nm, that is in the infrared region of the spectrum, are absorbed, and one photon is emitted at a wavelength of 532 nm, in the visible green region of the spectrum is barely the mechanism that produces the green light in the green laser pointers nowadays used in conferences. The nonlinear material typically used to achieve this effect is the potassium dihydrogenphosphate or KDP (KH_2PO_4).

Figure 2.2: Example of a linear and nonlinear (SHG) processes.



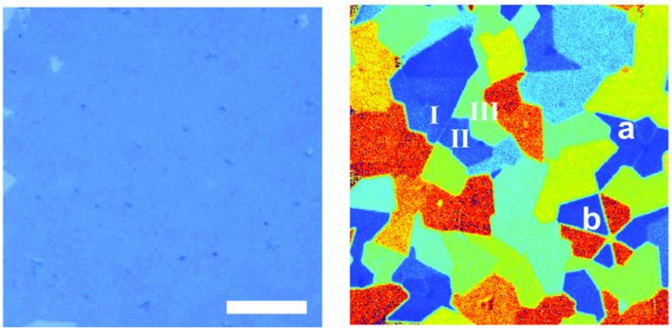
Nonlinear effects need a strong laser source to be detected. Hence, it is not a coincidence that the first observation came only in 1961 [26], just after the laser invention (1960) [27]. After these discoveries, and with the rapid growth of available laser sources, nonlinear optics became a powerful tool in material research. Modern laser relies nowadays on nonlinear optics. This is a sensitive tool to probe inversion symmetry in materials, since the SHG mechanism is active only in materials without an inversion symmetry (non-centrosymmetric materials).

Because of this sensitivity, nonlinear response is an essential tool to characterize and explore electronic and structural properties of materials in condensed matter. As a straightforward example SHG can be used to detect phase changing from centrosymmetric to non-centrosymmetric phases of piezoelectrics, pyroelectrics, and ferroelectrics. For instance, by looking at the SHG signal ferroelectric $BaTiO_3$ as a function of the temperature, it is possible to detect a change of phase at around 120° C, through the disappearance of the

SHG signal at higher temepratures. At that temperature, inversion symmetry is restored in $BaTiO_3$ crystal and it loses its ferroelectric properties [28]. Further details on how to probe ferroelectrics tools and related experimental techniques can be found in Ref. [29].

The general sensitivity of SHG has also recently been used to characterize surfaces and interfaces. Indeed, SHG it is not only sensitive to the symmetries of the sample, but also to the lattice orientation. This idea has been recently applied by X. Yin *et al.* [2]. They developed a nonlinear optical imaging technique to determine the crystal orientations in 2D materials at a large scale. Their result is reported in Fig. 2.3. The left side is the image of a single layer of MoS_2 investigated within linear optics, and on the right side the same material is studied considering SHG techniques. It is clearly evident that the different flakes and orientation are visible in the SHG image. Nonlinear optics applications can range from opto-electronics to biology. Just to cite an example, SHG microscopy can be used to probe protein dynamics, as shown by Förderer [30]. This is done basically bounding a nonlinear nanocrystal to proteins. SHG nanocrystals modify the light color when they interact with

Figure 2.3: On the left a linear optical image of layered MoS_2, on the right side the SHG image of the same material (Readapted from Ref. [2])



a strong light source, and thus they can be imaged by means of the two-photon microscopy. By contrast, biological tissue do not present any nonlinear response, thus the nonlinear signal is only given by the nanocrystal. This allows scientists to visualize the dynamic of proteins.

As reported in Ref. [1], inorganic compounds are the most useful NLO materials for SHG. Their performance depends on

1. Lack of inversion symmetry;
2. Large linear and nonlinear optical coefficients;
3. Appropriate absorption edges;
4. Phase matchability;
5. Mechanical strength, high laser-damage threshold, optical homogeneity, and extent of facile growth.

Due to all the requirements that a material has to meet to be a good nonlinear optical material, very few compounds have been used so far in this field. It would be important to identify new possible candidates. Of course, it has to be mentioned as well that, while it is possible to compute some of these properties from a first-principles point of view, others can currently be evaluated only after the material is synthesized. High laser-damage threshold, optical homogeneity, and extent of facile growth are examples of properties that cannot be checked by first-principles making it, of course, difficult to predict and characterize materials. However, a computational approach that could be used to compute ascertain some of these requirements would already help and speed up materials research and discovery.

The lack of NLO materials has been recently discussed by different authors. In fact, looking at the infrared and ultraviolet region of the electromagnetic spectrum, few materials that somehow satisfy all the requirements are known nowadays. IR materials are particularly sought for organic and inorganic molecular sensing and for medical surgery because of their small band gap and large nonlinear optical coefficients. However, even if smaller band gaps should bring to a large nonlinear optical coefficients, the price to pay is a narrower transparency window and a lower damage threshold. Few materials beyond the standard chalcopyrites $AgGaS_2$ and $ZnGeP_2$ find widespread use owing to low laser damage thresholds and two-photon absorption (see Ref. [31] and references therein for further details). The same situation is found on the other side of the electromagnetic spectrum, in the ultraviolet region. Technologically, these materials are used in semiconducting manufacturing, photolitography, laser systems, atto second pulse generation, and advanced instrument development. Really few materials satisfy the required conditions besides the well known $\beta-BaB_2O_4$. It is also worth to mention that most of the materials found so far for applications in the UV are beryllium-based, and given the toxicity of the latter, new chemically benign compounds are highly sought [32]. In addition to the limitations already mentioned, here there is a problem related to the optical coefficients. Indeed, UV materials have a large band gap ($\simeq 6\ eV$) and this is a huge limitation because of the inverse relationship that exists between the susceptibility, at every order, and the band gap. Thus, a large band gap means small optical coefficients.

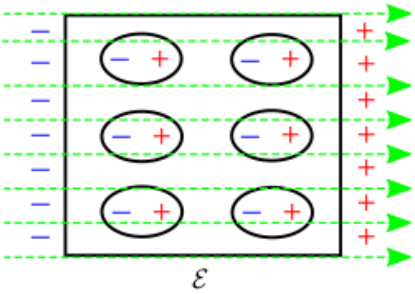
From these observations, it is clear that models and approaches that could help to circumvent these limitations are actively sought. This is one of the key aspects of this thesis: given the range of usability of optical materials, it is nowadays fundamental to find a way to accelerate materials discoveries and improve the existing devices, or either develop new ones more efficient. From our side this can be done by collecting materials properties from first-principles, developing models and analyzing trends that can improve the comprehension of the behaviour of materials in a given circumstance. An example of how the analysis of trends and models can help the material research can be found in the field of catalysis. The latter has relied for a long time on computing adsorbate energies that are not complicated quantities to compute within first-principles approaches. Thanks to models we could understand what inherent features in the electronic structure drive these adsorption energies. Following that spirit, it is now known that an important descriptor for the adsorption energies in catalysis is the position of the d bands, leading to great chemical and structural insight.

After this brief overview on optical materials, their usability, and the necessity of finding new materials, we move now to the ingredients we need to get a good optical material, following the required criteria. The guidelines for the extraction of physical optical properties as the susceptibility χ at the different orders is given, within different approaches. After that, the phase-matching problem will also be discussed. As already stated, this can only be the starting point for searching for new optical materials, since other criteria that the material must satisfy have to be investigated after the material is synthesized.

2.2 Linear and nonlinear optical susceptibility

When a semiconductor interacts with an external electric field, a series of oriented dipoles are created internally, generating an internal microscopic response of the material, that is the origin of the macroscopic polarization. This situation is naively depicted in Fig. 2.4. The polarization, also defined by the dipole moment per unit volume, describes indeed how the system is affected by the external electric field. Considering an homogeneous and isotropic dielectric medium, the polarization is aligned with and proportional to the electric field $\mathcal{E}$:

Figure 2.4: Creation of oriented dipoles inside a material that interacts with an external electric field.



$$\mathcal{P} = \chi\mathcal{E} \tag{2.1}$$

where the proportionality coefficient χ is known as the electric susceptibility of the material. Assuming that the amplitude of the incident electric field is much weaker than the atomic electric field strength, we can rewrite the previous equation considering a Taylor expansion of the polarization:

$$\mathcal{P}_i = \mathcal{P}_i^{(0)} + \sum_j \chi_{ij}^{(1)} \mathcal{E}_j + \sum_{jk} \chi_{ijk}^{(2)} \mathcal{E}_j \mathcal{E}_k + ... \tag{2.2}$$

where, $\mathcal{P}_i^0$ is the spontaneous polarization of the material, typically present in ferro-electric materials for example, and the next orders give rise to the linear contribution (proportional to the electric field), and the nonlinear contributions, for which the polarization and the electric field are related by a nonlinear relationship. In the rest of this thesis, we will generally refer to nonlinear processes considering the first nonlinear term present in Eq. (2.2), referring then to the second order of the expansion. This term is at the origin of phenomena as SHG. The knowledge of the material susceptibility is the first step that we carry out in our screening procedure, since this is the quantity related to the refractive index in the linear order, and the SHG nonlinear coefficient considering the nonlinear term. This is then a key quantity to investigate the optical coefficients of the materials at the different orders.

In the next sections, a brief description of how to extract the susceptibility at linear and nonlinear order, considering a classical and quantum mechanical approach, is given, focusing on some physical aspects that can be extracted from these treatments. Extended discussions can be found in the literature [33–36].

2.2.1 Classical picture: The anharmonic oscillator

The interaction of an external electric field with a material produces an oscillation of the electrons. This oscillation around its equilibrium position gives rise to a dipole moment $p(t)$ that will change in time with an amplitude proportional to the motion of the electrons. In classical mechanics, we can depict this oscillation considering the anharmonic oscillator, described by the following equation of motion

$$\ddot{x} + 2\gamma\dot{x} + \omega_0^2 x + ax^2 + bx^3 + ... = -\mathcal{E}(t)/m \tag{2.3}$$

where ω_0 is the resonant frequency of the electron, $2\gamma\dot{x}$ is the damping force, and $\mathcal{E}(t)$ is the applied electric field. The Taylor expansion represents the different orders that contribute to the restoring force. For the electric field, we consider a linearly polarized plane-wave in the x direction that propagates along the z axis with a wave vector k. This writes as

$$\begin{aligned} \mathcal{E}(z,t) &= \mathcal{A}(\omega)e^{-i(\omega t - kz)} + c.c. \\ &= \tilde{\mathcal{E}}(\omega)e^{-i\omega t} + c.c. \end{aligned} \tag{2.4}$$

where $\mathcal{A}$ is the spatially slowly varying field amplitude, with $\tilde{\mathcal{E}} = \mathcal{A}e^{-ikz}$, and c.c. indicates the complex conjugate ($\tilde{\mathcal{E}}(\omega)^* = \tilde{\mathcal{E}}(-\omega)$).

At the first order of the perturbation, we have a steady state solution of the type:

$$x^{(1)}(t) = c_1(\omega)e^{-i(\omega t - kz)} + c.c. \tag{2.5}$$

where:

$$c_1(\omega) = \frac{-\mathcal{A}(\omega)}{(\omega_0^2 - \omega^2 - i\gamma\omega)} \tag{2.6}$$

As already stated, the macroscopic polarization can be defined as the dipole moment per unit volume. Thus, if we have N atoms per unit volume, the resonant polarization can be written as $\mathcal{P} = Np$, giving as a result at the first order of the expansion:

$$\begin{aligned} \mathcal{P}^{(1)}(z,t) &= -Nx^{(1)}(t) \\ &= \frac{N\mathcal{A}(\omega)}{\omega_0^2 - \omega^2 - 2i\gamma\omega} e^{-i(\omega t - kz)} + c.c. \end{aligned} \tag{2.7}$$

Comparing this equation to Eq. (2.2) we can finally write for the first order susceptibility:

$$\chi^{(1)}(\omega) = \frac{N}{(\omega_0^2 - \omega^2 - i\gamma\omega)} \tag{2.8}$$

In linear optics, a more familiar concept to describe the response of materials with respect to an external electric field, is the dielectric function $\varepsilon(\omega)$, that is related to the susceptibility in the following way:

$$\varepsilon(\omega) = 1 + 4\pi\chi^{(1)}(\omega) \tag{2.9}$$

The dielectric function is a complex quantity. Its real and imaginary parts can be connected to the optical response of the material whether it is transparent, refractive or conductive. For example, we have for the calculation of the refractive index that this is related to the dielectric function through the relationship

$$n(\omega) = \frac{1}{\sqrt{2}}\sqrt{\varepsilon_1(\omega) + \sqrt{\varepsilon_1(\omega)^2 + \varepsilon_2(\omega)^2}} \tag{2.10}$$

where $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ are the real and imaginary parts of $\varepsilon(\omega)$, respectively.

At the second order, considering the case of SHG (see Fig. 2.2), the steady state solution is given by

$$x^{(2)}(t) = c_2(2\omega)e^{-2i\omega t} \tag{2.11}$$

with

$$c_2(2\omega) = \frac{-a\mathcal{A}}{(\omega_0^2 - 4\omega^2 - 4i\omega\gamma)(\omega_0^2 - \omega^2 - 2i\omega\gamma)^2} \tag{2.12}$$

In a similar way, the polarization at the second order is given by

$$\mathcal{P}^{(2)}(2\omega) = -Nx^{(2)}(2\omega) \tag{2.13}$$

extracting a nonlinear susceptibility of the form:

$$\chi^{(2)}(2\omega, \omega, \omega) = \frac{Na}{(\omega_0^2 - 4\omega^2 - 4i\omega\gamma)(\omega_0^2 - \omega^2 - 2i\omega\gamma)^2} \tag{2.14}$$

At this point, it should be reminded that this model represents the physical situation in which an atom that possesses a single resonance frequency ω_0 is immersed in an electric field at frequency ω. For many aspects, such a model is not ideal to reproduce the behaviour of a solid immersed into an electric field. Indeed, first of all, in a real system we have multiple resonance frequencies. Further, to simplify the treatment, we considered the case of an isotropic dielectric medium immersed in an monochromatic field. We could then avoid the tensor notation. Further, within a classical approach, a fundamental quantity such as the optical matrix elements of the transition are not taken into account, since this is purely a quantum mechanical quantity. However, this simplified model gives us a good idea of the phenomena, since it corresponds to the physical situation of electrons in a real material, because the actual potential well that the electrons feel is not perfectly parabolic, as in this case.

Nonetheless, within this theoretical model, it is easy to show the relationship between the nonlinear susceptibility and the linear one.

The direct proportionality between these two quantities was previously deduced by Miller from empirical observations [10]. Actually, for a long time, the anharmonic oscillator has been the only way to prove this relationship from a theoretical point of view. This is basically the reason why one can think that a material that possesses high linear optical properties, should also have high nonlinear optical properties. Following Ref. [33] we can see that it is possible to rewrite Eq. (2.14), as a function of the linear one (Eq. (2.8)):

$$\chi^{(2)}(2\omega,\omega,\omega) = \Delta\chi^{(1)}(2\omega)\left[\chi^{(1)}(\omega)\right]^2 \tag{2.15}$$

where Δ is the Miller's coefficient. In the case of the anharmonic oscillator, this is given by $\Delta = \frac{a}{N}^2$, where the parameter $a = \frac{\omega_0^2}{d}$, and d is the lattice constant. Since the majority of the materials initially taken in consideration to investigate this relationship show some peculiarity (e.g. GaSb, GaAs, InSb, GaP, ...), Miller's coefficient was thought to be roughly a constant for different materials, but this has been demonstrated not to be exactly correct [37]. This aspect will be further investigated in the next sections of this thesis.

In order to go beyond this model, we need to include quantum effects. In the next sections we will then extract the susceptibility from quantum theory. As already stated, the physical conclusions will be drawn in a similar way to the classical one, but we will have the natural inclusions of new contributions that can give us further insights into the comprehension of the phenomena.

2.2.2 Quantum picture

In quantum mechanics, all the properties of the atomic system can be described in terms of the atomic wave function $\Psi(\mathbf{r}, t)$ that is the solution of the time dependent Schrödinger equation:

$$i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t}=\hat{H}\Psi(\mathbf{r},t) \tag{2.16}$$

where $\hat{H}$ is the Hamiltonian describing the light-matter interaction and can be written as

$$\hat{H}=\hat{H}_0+\hat{V}(t) \tag{2.17}$$

where $\hat{H}_0$ is the free-atom Hamiltonian and $\hat{V}(t)$ is the external interacting potential. Including the proper form of the potential that describes the interaction between the system and the external perturbation, all the properties of the system can be accessed. However, to obtain the susceptibility, we have to follow a different approach than the previous one. Indeed, as we will also see in the next chapter, only a few quantum mechanical problems have an exact analytical solution. We then have to apply perturbation theory to retrieve useful quantities in this case [38].

First of all, we consider the static case, in which no perturbations are applied to the system. Since we deal with periodic systems, we use Bloch states to represent energy eigenstates that assume the following form:

$$\Psi_n(\mathbf{r},t)=u_n(\mathbf{r})e^{-i\omega_n t} \tag{2.18}$$

where n is the band index. These eigenstates have to satisfy the time-independent Schrödinger equation

$$\hat{H}_0u_n(\mathbf{r})=E_nu_n(\mathbf{r}) \tag{2.19}$$

where $E_n=\hbar\omega_n$ are the stationary eigenvalues, and have to be chosen in such a way that they constitute a complete, orthonormal set satisfying the condition

$$\int u_m^\star u_n \, d\mathbf{r} = \delta_{mn} \tag{2.20}$$

The perturbed case can be extracted from the static one by expanding these basis functions.

In the perturbative approach, we rewrite the Hamiltonian (2.17) with

$$\hat{H} = \hat{H}_0 + \lambda \hat{V}(t) \tag{2.21}$$

where $\lambda \in [0, 1]$ characterize the strength of the perturbation. The solution will be now expressed in terms of a power series in λ:

$$\psi(\mathbf{r}, t) = \psi^{(0)}(\mathbf{r}, t) + \lambda\psi^{(1)}(\mathbf{r}, t) + \lambda^2\psi^{(2)}(\mathbf{r}, t) + ... \tag{2.22}$$

The set of equations to be solved can be now written as

$$i\hbar \frac{\partial \psi^{(0)}}{\partial t} = \hat{H}_0 \psi^{(0)}, \tag{2.23a}$$

$$i\hbar \frac{\partial \psi^{(N)}}{\partial t} = \hat{H}_0 \psi^{(N)} + \hat{V}\psi^{(N-1)}, \quad N = 1, 2, 3, ... \tag{2.23b}$$

At the zeroth order, the system is in its ground-state and the solution takes the form

$$\psi^{(0)}(\mathbf{r}, t) = u_g(\mathbf{r}) e^{-iE_g t/\hbar} \tag{2.24}$$

Where g labels indeed the ground-state of the system. The N^{th}-order solution can be written by expanding the ground-state one:

$$\psi^{(N)}(\mathbf{r}, t) = \sum_l a_l^{(N)}(t) u_l(\mathbf{r}) e^{-i\omega_l t} \tag{2.25}$$

where $a_l^{(N)}(t)$ is the probability amplitude that the atom is in an energy eigenstate l at the time t, to the N^{th}-order of the perturbation. Substituting this solution in Eq. (2.23b), we have that the probability amplitude is given by a recursive system of equations

$$i\hbar \sum_l \dot{a}_l^{(N)} u_l(\mathbf{r}) e^{-i\omega_l t} = \sum_l a_l^{(N-1)} \hat{V} u_l(\mathbf{r}) e^{-i\omega_l t} \tag{2.26}$$

that can be further rewritten as

$$a_m^{(N)}(t) = \frac{1}{i\hbar} \sum_l \int_{-\infty}^{t} V_{ml}(t') a_l^{(N-1)}(t') e^{i\omega_{ml} t'} \tag{2.27}$$

where $\omega_{ml} = \omega_m - \omega_l$ is the energy difference between the m and l electronic state, and $V_{ml} = \langle u_m | \hat{V} | u_l \rangle = \int u_m^{\star} \hat{V} u_l \, d\mathbf{r}$ is the matrix element of the perturbed Hamiltonian. From Eq. (2.27), the recursive nature of our set of equations is even more evident.

Finally, whatever form the perturbation external potential takes, at the zeroth order case the probability amplitude is given by:

$$a_l^{(0)} = \delta_{lg} \tag{2.28}$$

This means of course that, without any perturbing term, the system is in its ground-state. Switching on the perturbation with respect to external field the potential can be written as

$$\hat{V} = -\hat{\boldsymbol{\mu}} \cdot \mathcal{E}(t) \tag{2.29}$$

where $\hat{\boldsymbol{\mu}}$ is the electric dipole moment operator of the interaction that can be written considering the position operator:

$$\hat{\boldsymbol{\mu}} = -\hat{\mathbf{r}}(t) \tag{2.30}$$

Assuming the solution in Eq. (2.28), and solving the integral in Eq. (2.27) we then have at the first order

$$a_m^{(1)}(t) = \frac{\boldsymbol{\mu}_{mg} \cdot \tilde{\mathcal{E}}(\omega)}{\omega_{mg} - \omega - i\gamma} e^{i(\omega_{mg}-\omega)t} \tag{2.31}$$

where the positive broadening $\gamma \to 0^+$ accounts for the causal behavior of the response. At the second order we have

$$a_n^{(2)}(t) = \sum_m \frac{[\boldsymbol{\mu}_{nm} \cdot \tilde{\mathcal{E}}(\omega)][\boldsymbol{\mu}_{mg} \cdot \tilde{\mathcal{E}}(\omega)]}{(\omega_{ng} - 2\omega - i\gamma)(\omega_{mg} - \omega - i\gamma)} e^{i(\omega_{ng}-2\omega)t} \tag{2.32}$$

where we have introduced the dipole transition moment, defined as

$$\boldsymbol{\mu}_{ml} = \int u_m^\star \hat{\boldsymbol{\mu}} u_l \, d\mathbf{r} \tag{2.33}$$

This quantity gives the probability that an electronic transition from a state l to a state n will happen.

At this point, we can compute the susceptibility at the different orders by considering the expectation value of the electric dipole moment

$$\langle \tilde{\mathbf{p}} \rangle = \langle \psi | \hat{\boldsymbol{\mu}} | \psi \rangle \tag{2.34}$$

where the wave functions are given by the perturbation expansion in Eq. (2.22) with $\lambda = 1$. The linear contribution to the expectation value is given by

$$\langle \tilde{\mathbf{p}}^{(1)} \rangle = \left\langle \psi^{(0)} \middle| \hat{\boldsymbol{\mu}} \middle| \psi^{(1)} \right\rangle + \left\langle \psi^{(1)} \middle| \hat{\boldsymbol{\mu}} \middle| \psi^{(0)} \right\rangle \tag{2.35}$$

Using Eqs. (2.24) and (2.26) we have:

$$\langle \tilde{\mathbf{p}}^{(1)} \rangle = \sum_m \left(\frac{\boldsymbol{\mu}_{gm} [\boldsymbol{\mu}_{mg} \cdot \tilde{\mathcal{E}}(\omega)]}{\omega_{mg} - \omega} e^{-i\omega t} \right) + c.c. \tag{2.36}$$

Considering the linear polarization as $\tilde{\mathcal{P}}^{(1)} = N\langle\tilde{\mathbf{p}}^{(1)}\rangle$, where N is the density of electrons, we have then for the linear susceptibility

$$\chi_{ij}^{(1)}(\omega,\omega) = N\sum_m \left(\frac{\mu_{gm}^i \mu_{mg}^j}{\omega_{mg}-\omega-i\gamma}+\right) + c.c. \tag{2.37}$$

The same procedure can be followed to get the second order nonlinear susceptibility, starting from the second order contribution to the induced dipole moment per atom

$$\langle\tilde{\mathbf{p}}^{(1)}\rangle = \left\langle\psi^{(0)}\middle|\hat{\boldsymbol{\mu}}\middle|\psi^{(2)}\right\rangle + \left\langle\psi^{(1)}\middle|\hat{\boldsymbol{\mu}}\middle|\psi^{(1)}\right\rangle + \left\langle\psi^{(2)}\middle|\hat{\boldsymbol{\mu}}\middle|\psi^{(0)}\right\rangle \tag{2.38}$$

The second order susceptibility takes then the form

$$\begin{aligned}\chi_{ijk}^{(2)}(2\omega,\omega,\omega) =& N\sum_{mn}\left(\frac{\mu_{gn}^i \mu_{nm}^j \mu_{mg}^k}{(\omega_{ng}-2\omega-i\gamma)(\omega_{mg}-\omega-i\gamma)}+\right.\\ &\frac{\mu_{gn}^j \mu_{nm}^i \mu_{mg}^k}{(\omega_{ng}+\omega+i\gamma)(\omega_{mg}-\omega-i\gamma)}+\\ &\left.\frac{\mu_{gn}^j \mu_{nm}^k \mu_{mg}^i}{(\omega_{ng}+\omega+i\gamma)(\omega_{mg}+2\omega+i\gamma)}\right)\end{aligned} \tag{2.39}$$

Finally, it is interesting to look at the similarities between the second-order susceptibility derived within the two approaches. In Eq. (2.14) the overall set of electronic transition is replaced by a single frequency at ω_0. Furthermore, all the electronic transitions from ground state to an excited state of the system are in principle possible. In Eq. (2.39), we have likely the same behaviour at the denominator, where now multiple non-resonant frequencies are taken in consideration. The main difference lies on the denominator, in which the optical matrix elements of the transition naturally appears in Eq. (2.39). As already stated, this is a fundamental contribution to the susceptibility, both at the linear and nonlinear regime, since, for example, transitions that are forbidden by symmetry are suppressed by having a null value of the matrix elements.

For the sake of the readability of this thesis, we limit our discussion to the approach just proposed. The density matrix approach, that is more general and can include also resonant processes such as relaxation processes, is indeed more complicated from a mathematical point of view. More details on these theories can be found for example in the literature cited at the beginning of this section.

2.2.3 More on Miller's rule

We have already introduced the Miller's rule that relates the nonlinear optical susceptibility with the linear one for the case of homogeneous and isotropic dielectric media. More generally, this relationship writes [10]

$$\chi^{(2)}_{ijk}(2\omega,\omega,\omega) = \Delta_{ijk}\chi^{(1)}_{ii}(2\omega)\chi^{(1)}_{jj}(\omega)\chi^{(1)}_{kk}(\omega) \tag{2.40}$$

where Δ_{ijk} is the Miller's coefficient already seen in Eq. (2.15). From many points of view, having a confirmation on the validity of this empirical rule could be of fundamental importance. Of course, accessing the nonlinear susceptibility from the knowledge of the linear one would be a great advantage both from an experimental and computational point of view, because linear optical properties are much easier to calculate and analyze. The main message we have from this empirical rule at this stage is that, by having a material with a high linear coefficient, the chance to have high nonlinear properties is quite high. It is anyway important to mention that this is valid only if certain assumptions on the Miller's coefficient are verified. Indeed, it is still debated if this quantity can be assumed as a constant, and thus material independent. Of course the comprehension of the main ingredients that constitute the Miller's coefficient would open the way to new comprehension of the phenomena, and to new predictive power. This will be further discussed in the results section.

An approach based on the quantum form of $\chi^{(2)}$ was proposed by Scandolo and Bassani [39]. They first derived a series of sum rules that apply to the nonlinear susceptibility [40]

$$\chi^{(2)}(0,0) = \frac{2}{\pi}\int_0^\infty \frac{\mathrm{Im}\,\chi^{(2)}(\omega',\omega')}{\omega'}\,d\omega', \tag{2.41}$$

$$\int_0^\infty \omega^n\,\mathrm{Re}\,\chi^{(2)}(\omega,\omega)\,d\omega = 0 \tag{2.42}$$

with $n = 0, 2, 4$;

$$\int_0^\infty \omega^m\,\mathrm{Im}\,\chi^{(2)}(\omega,\omega)\,d\omega = 0 \tag{2.43}$$

with $m = 1, 3$; and finally

$$\int_0^\infty \omega^5\,\mathrm{Im}\,\chi^{(2)}_{ijk}(\omega,\omega)\,d\omega = \frac{N\pi}{16}\left\langle \frac{\partial^3 V}{\partial x_i \partial x_j \partial x_k}\right\rangle_0 \tag{2.44}$$

where N is the density of electrons, $V(\mathbf{x})$ is the external potential experienced by the electrons, and the average, indicated by the angle bracket, is performed in the ground-state of the system. Taking as a starting point Eq. (2.39), they considered a simplified model in which there is a single resonant frequency that approximates the overall set of electronic transitions. Thus, using the sum rules in

Eq. (2.41) to Eq. (2.44), they were able to show that

$$\chi^{(2)}_{ijk}(\omega,\omega) = -\frac{\left\langle \frac{\partial^3 V}{\partial x_i \partial x_j \partial x_k} \right\rangle_0}{2N^2} \chi^{(1)}_{ii}(2\omega)\chi^{(1)}_{jj}(\omega)\chi^{(1)}_{kk}(\omega) \tag{2.45}$$

Even more important for the sake of this work, they showed the theoretical validity of the Miller's rule also in the static limit. Taking indeed the limit $\omega \to 0$ we have

$$\chi^{(2)}_{ijk}(0,0) = -\frac{\left\langle \frac{\partial^3 V}{\partial x_i \partial x_j \partial x_k} \right\rangle_0}{2N^2} \chi^{(1)}_{ii}(0)\chi^{(1)}_{jj}(0)\chi^{(1)}_{kk}(0) \tag{2.46}$$

It is evident anyway that one of the main results of their works is the explicit expression of the Miller's coefficient. As already stated, this aspect will be further investigated in the results section, since we believe that a better comprehension of such an empirical rule can open new ways in the study of nonlinear optical properties of semiconductors.

2.2.4 Symmetry properties of the nonlinear susceptibility

To conclude the discussion on the second order susceptibility it is worth discussing the symmetry properties of this quantity that is a third rank tensor.

First of all, SHG effects are not possible in centrosymmetric crystals for which all the elements of the tensor are identically null. Following Ref. [33], this can be shown in an easy way. A centrosymmetric crystal is basically a system where a centre of inversion exist. If we generate a nonlinear polarization in a centrosymmetric crystal considering a single applied field $\mathcal{E}$, the components of the polarization are:

$$\mathcal{P}_i^{(2)}(\mathcal{E}) = \sum_{j,k} \chi_{ijk}^{(2)} \mathcal{E}_j \mathcal{E}_k \tag{2.47}$$

Reversing the direction of the field, we have:

$$\begin{aligned} \mathcal{P}_i^{(2)}(-\mathcal{E}) &= \sum_{j,k} \chi_{ijk}^{(2)} (-\mathcal{E}_j)(-\mathcal{E}_k) \\ &= \sum_{j,k} \chi_{ijk}^{(2)} \mathcal{E}_j \mathcal{E}_j = P_i^{(2)}(\mathcal{E}) \end{aligned} \tag{2.48}$$

Because of the Neumann's principle [41], that states that the symmetry of a macroscopic property tensor of a material must at least possess the symmetry of the point group of the material, since the crystal has an inversion symmetry, we must obtain the same physical result by keeping the field in the original direction and rotating by 180 degree the crystal. In terms of coordinates axes of the inverted crystal, all the components of $\mathcal{E}$ and $\mathcal{P}^{(2)}$ change sign, and therefore we have:

$$-\mathcal{P}_i^{(2)} = \sum_{j,k} \chi_{ijk}^{(2)} (-\mathcal{E}_j)(-\mathcal{E}_k) \tag{2.49}$$

It is obvious that, the only solution that satisfies simultaneously both equations is $\chi_{ijk}^{(2)} = 0 \ \forall \ i, j, k$. This shows that SHG, or more generally, all the optical nonlinear phenomena that involve the second order term of the susceptibility are forbidden in centrosymmetric materials.

It has to be mentioned also that in literature the SHG tensor is typically expressed by:

$$d_{ijk} = \frac{1}{2} \chi_{ijk}^{(2)} \tag{2.50}$$

having then

$$\mathcal{P}_i(2\omega) = \sum_{jk}\sum_{mn} 2d_{ijk}\mathcal{E}_j(\omega)\mathcal{E}_k(\omega) \tag{2.51}$$

Symmetries are important especially because the number of independent elements that describe the nonlinear interaction can be drastically reduced. Indeed, if Kleinman's symmetry condition is valid, at small frequencies ($\omega \to 0$) the d_{ijk} tensor is symmetric in its indices. This is always valid in the case of SHG (see Ref. [42]). We can then contract the tensor following the notation:

jk:	11	22	33	23,32	31,13	12,21
l:	1	2	3	4	5	6

This reduces the number of independent elements from 27 to 18. If we further look for the independent elements, then it is simple to notice that not all them are independent. Indeed, just to give an example

$$d_{12} = d_{122} = d_{212} = d_{26} \tag{2.52}$$

Thus, in the case $\omega \to 0$, the independent elements of the SHG susceptibility can be reduced to a maximum of 10 using the symmetry properties of the tensor. This is generally valid, but a further reduction of the independent elements is still possible including the symmetry of the crystal. All these considerations can be used to express the response of a material through a scalar relationship that writes

$$\mathcal{P}(2\omega) = 2d_{\text{eff}}\mathcal{E}(\omega)^2 \tag{2.53}$$

In each case, d_{eff} is obtained by evaluating the summation in Eq. (2.47). Further details can be found in Refs. [34, 43]. In our case, the effective nonlinear coefficient is calculated averaging the response along the crystal axes and exploiting the point group symmetry of the system. This is done following the work of Kurtz [44], where the general formula that holds for all the non-centrosymmetric crystals is written as:

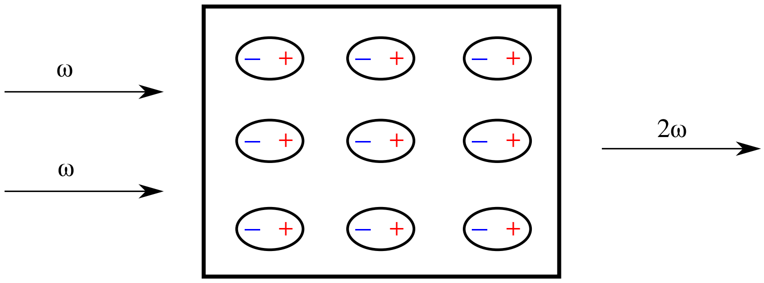
$$\begin{aligned} d_{\text{eff}}^2 = &\frac{19}{105}\sum_i (d_{iii})^2 + \frac{13}{105}\sum_{i\neq j} d_{iii}d_{ijj} \\ &+ \frac{14}{105}\sum_{i\neq j}(d_{iij})^2 + \frac{13}{105}\sum_{ijk,cyclic} d_{iij}d_{jkk} + \frac{5}{7}(d_{ijk})^2 \end{aligned} \tag{2.54}$$

2.3 Phase matching properties

Introducing the optical materials, we mentioned the requirements that a material needs to meet in order to be a good nonlinear optical material. In the previous section, the first condition, that is the derivation of the d_{ij} tensor has been discussed. Of course, this represents a necessary but not sufficient condition. The next step is the concept of phase-matching.

Nonlinear effects are typically small and we need a strong laser intensity to study these phenomena. This will of course affect the nonlinear energy conversion efficiency. Such a limitation can be circumvented either having a long sample or by requiring that the phases of the nonlinear waves generated throughout the whole crystal are all the same so that the microscopical fields add together coherently to enhance the efficiency, as naively depicted in Fig 2.6. When this is achieved, we are in the so-called phase-matching conditions. If the crystal is phase matched the power generated is large, if not, only very little power is obtained. This means that we can obtain large d_{ij} coefficients for a given material, but if this is not phase matchable the SHG signal in output will be too small. For

Figure 2.5: Sample of material containing N atomic dipoles. Each dipole oscillates with a phase that is determined by the phase of the incident field. The output beam is well defined if the field radiated by each dipole add constructively.



this purpose, we need to consider Maxwell equations as a starting point, in order to derive the coupled wave equation that describes the nonlinear process of SHG. The full derivation can be found for Example in Refs. [33, 34]. For the sake of the readability of this manuscript, here we just report the final quantity that we need to discuss the phase-matching properties. This quantity is the intensity of the generated wave at 2ω from the fundamental one at ω and it can be written as:

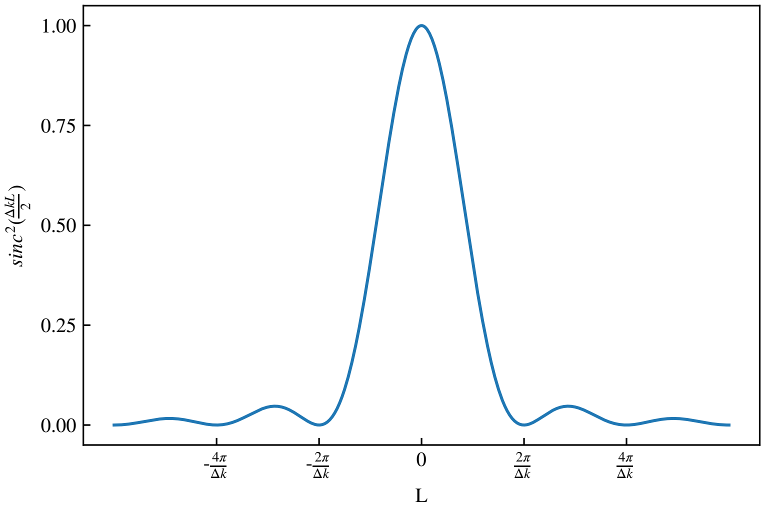
$$I^{2\omega} = \frac{2\omega^2 d_{\text{eff}}^2 \left(I^{\omega}\right)^2 L^2}{n(\omega)^2 n(2\omega)^2} sinc^2\left(\frac{\Delta k L}{2}\right) \tag{2.55}$$

Here we are assuming that the fundamental light at ω has traveled a distance L through the nonlinear material. The quantity Δk is called the momentum mismatch and is given by

$$\Delta k = k_{2\omega} - 2k_{\omega} \tag{2.56}$$

From Eq. (2.55), it is evident that the amplitude of the field is proportional to the length of the sample, and Δk. If we consider all the remaining quantities as a constant, then the maximum intensity is achieved when $\Delta k = 0$, as shown in Fig 2.6. This is known as the *perfect phase-matching condition*. When this condition is fulfilled, the individual atomic dipoles that constitute the material system are properly phased so that the field emitted by each dipole adds coherently in the forward direction. The total power radiated by the ensemble of atomic dipoles thus scales as the square of the number of atoms that participate. In Fig. 2.6, the dependency of the intensity of the generated wave at 2ω is shown as a function of the distance traveled in the material. The maximum is reached at $L = 0$, and this function rapidly drops down to 0 at $l = 2\pi/\Delta k$. This is know as the coherence length l_c and correspond to the distance where the SHG intensity drops to zero after propagating in the material.

Figure 2.6: Representation of the *sinc*^{2} function.



However, a perfect phase-matching condition is in general impossible to achieve in the case of SHG. Indeed, in such a case, the perfect phase-matching condition translates into:

$$n(\omega) = n(2\omega) \tag{2.57}$$

This basically means that to achieve the phase-matching, the refractive index value at the fundamental frequency needs to be the same as the refractive index value at second harmonic frequency. This is, of course, impossible to achieve because the refractive index of loss-less materials shows an effect known as normal dispersion in the range $\omega \rightarrow 2\omega$, in which the refractive index is a growing function of the frequency.

However, it is still possible to achieve phase-matching, playing with the variation of the refractive index. One can indeed produce a change of the refractive index, if the crystal is birefringent, varying the temperature or rotating the sample, for example. The main focus will be on the so-called angular phase-matching. This technique involves the variation of the refractive index following a rotation of the sample. This is one of the simplest technique to assess phase-matching from a computational point of view. Other techniques, such as temperature phase-matching or quasi-phase-matching are also possible. These can be either used independently or combined all together to overcome the different drawbacks that one can encounter. However, these other techniques would require more sophisticated calculations that are not in the interest of this project at the moment. In the following section an overview on the angular phase-matching is given. Further details about this method and other above mentioned ones can be found in Refs. [33, 34].

2.3.1 Angular phase-matching

One technique to achieve the phase-matching condition is the angular phase-matching. This methodology takes advantage of the birefringent properties of uniaxial and bi-axial crystals, in which the light can experience different refractive indices depending on the propagation direction with respect to the optic axis of the crystal.

Considering the case of bi-axial crystals, in which there are two optic axis, we can check if a material is phase matchable by looking for a solution of the Fresnel equations for the fundamental wave and the second harmonic one. As can be seen in Ref [45], this set of equations can be written as:

$$\begin{aligned} &\frac{k_x^2}{(n^{-2}(\omega) - n_x^{-2}(\omega))} + \frac{k_y^2}{(n^{-2}(\omega) - n_y^{-2}(\omega))} \\ &+ \frac{k_z^2}{(n^{-2}(\omega) - n_z^{-2}(\omega))} = 0 \end{aligned} \tag{2.58}$$

$$\begin{aligned} &\frac{k_x^2}{(n^{-2}(2\omega) - n_x^{-2}(2\omega))} + \frac{k_y^2}{(n^{-2}(2\omega) - n_y^{-2}(2\omega))} \\ &+ \frac{k_z^2}{(n^{-2}(2\omega) - n_z^{-2}(2\omega))} = 0 \end{aligned} \tag{2.59}$$

where $k_x = sin\theta cos\phi$, $k_y = sin\theta sin\phi$, and $k_x = cos\theta$. The angle θ is the angle between the wave normal and the z-axis, and ϕ is the angle from the x axis in the x-y plane. $n_x(\omega)$, $n_y(\omega)$, $n_z(\omega)$, and $n_x(2\omega)$, $n_y(2\omega)$, $n_z(2\omega)$ are the three principal refractive indices of fundamental and harmonic waves, respectively. We can then use this set of equations to look for the angles (θ_m, ϕ_m), the so-called phase match angles, that satisfy the phase-matching condition Eq. (2.57).

3

Method: High-throughput ab-initio materials properties

In the previous chapter we have introduced the theoretical framework for the calculation of optical properties of materials. Accurate and quantitative analysis of such properties is essential for the comprehension of the observed behaviour and optimization. From this point of view, first-principles calculations have proven to be a very powerful tool to explore the electronic and optical properties of materials. Density Functional Theory (DFT) [46, 47] provides a good description of the electronic structure. Density Functional Perturbation Theory (DFPT) [48, 49] is widely used to predict the linear and nonlinear response (and related physical quantities) of periodic systems when they are submitted to an external perturbation. For instance, when considering the effect of a homogeneous electric field, DFPT allows one to compute the macroscopic dielectric function and the second harmonic susceptibility in the static limit

($\omega = 0\ eV$). The success of DFT and DFPT stems from their reliability and low computational cost. As a result, first-principles calculations have recently been combined with a high-throughput (HT) approach [17, 50] targeting the discovery of new materials. Indeed, the combination of these two methods enables the creation of large databases of materials properties that would be prohibitive (in time and cost) for experimental measurements.

In this chapter the general aspects of DFT and DFPT will be given, focusing on some relevant details for the purpose of this thesis. Detailed discussions can be found in literature. In the last section the HT approach combined with DFT and DFPT is then discussed.

3.1 Density Functional Theory: A ground-state theory

The description of a system in Quantum mechanics is associated with the resolution of the Schrödinger equation, here reported in the most general form :

$$\hat{H}\Psi = E\Psi \tag{3.1}$$

where $\hat{H}$ represents the Hamiltonian operator, Ψ is the many-body wave function of the quantum system, and E the energy level of the system. The solution of such an equation gives information on the ground-state of the system. However, it is known that Schrödinger equation has an analytic solution only for a few rather simple cases (harmonic oscillator, hydrogen atom, ...). Alternative resolutions have been proposed for the problem of N interacting electrons. One of these approaches is based on a variational principle on the energy for which, given a wave function Φ, that satisfies the appropriate boundary conditions, then the expectation value represents an upper bound to the exact ground-state

$$E[\Phi] \leq \frac{\langle\Phi| \hat{H} |\Phi\rangle}{\langle\Phi|\Phi\rangle} \tag{3.2}$$

The equality holds only if Φ is the exact ground-state of the system. Indeed, by definition, the ground-state energy of the system is the smallest eigenvalue. In summary, to calculate the ground-state energy E, we can minimize the functional $E[\Phi]$ with respect to all states Φ. The value of this functional gives an upper bound to the value of E, and even a relatively poor estimate of the ground-state wave function gives a relatively good estimate of E. This variational approach is at the base of the theories as the Hartree-Fock (HF) theory [51] and the DFT [52]. These are the widest used methods in computational physics and chemistry communities. The difference between the two relies in the fact that while HF tries to solve this problem reconstructing a trial many-body wave function via the use of Slater determinants, in DFT this problem is reformulated in such a way that the electronic energy is a functional of the density, reducing significantly the degrees of freedom of the solution and the complexity of the problem.

3.1.1 Theoretical basis

The Hamiltonian of N interacting electrons immerse in an external potential V_{ext} (such as the one generated by the ionic background) can be generally written as:

$$\hat{H} = -\sum_{i}^{N} \frac{\nabla^2}{2} + \frac{1}{2} \sum_{i \neq j}^{N} \frac{1}{\mathbf{r}_i - \mathbf{r}_j} + V_{ext} \tag{3.3}$$

where the first term represent the kinetic energy, the second term is the repulsive Coulomb electron-electron energy with $\mathbf{r}_{ij}$ electrons coordinates, and the third term represents the effect of the external potential. At the base of Density Functional Theory (DFT) there is the assumption that the entire system of interacting electrons can be described just by the knowledge of the non-interacting system of electrons. This is a crucial first distinction between DFT and other previous theories. Indeed, considering for example the HF theory, the solution of the problem of N interacting electrons goes through the minimization w.r.t. a wave function of the form:

$$\Psi_{el}(\mathbf{r}_1, .., \mathbf{r}_N) \tag{3.4}$$

This depends on $3N$ variables: three spatial directions for each of the N electrons. Furthermore, this wave function is usually constructed considering the Slater determinant of the single electron wave function and this does not allow us to treat the electronic interactions in a correct way.

In DFT a crucial role is played by the electron density of the system

$$\rho(\mathbf{r}) = \int |\Psi_{el}(\mathbf{r}_1, .., \mathbf{r}_N)|^2 \, d\mathbf{r}_2 .. d\mathbf{r}_N \tag{3.5}$$

The crucial point here is that we need to solve an equation that depends only on the 3 spatial variables, whatever is the number of electrons N. This significantly reduces the size of the problem we have to solve.

DFT is based on the two fundamental theorems from P. Hohenberg and W. Kohn [46]:

- For each system of interacting particles in an external potential V_{ext}, this potential is uniquely determined from the ground-state particle density $\rho_0(\mathbf{r})$
- We can define a universal energy functional $E_{HK}[\rho]$ such that for each external potential V_{ext}

  $$E_{HK} = F_{HK}\left[\rho(\mathbf{r})\right] + \int V_{ext}(\mathbf{r})\rho(\mathbf{r})\, d\mathbf{r} \tag{3.6}$$

  where F_{HK} is a functional of the density and does not depend on the external potential. For each V_{ext}, the exact ground-state energy follows from the minimization of E_{HK} and the density that minimize this quantity is the exact ground-state density $\rho_0(\mathbf{r})$

In few words the first theorem states that the ground-state density $\rho_0(\mathbf{r})$ determines all the ground-state properties of the system, and the second one gives us a variational principle to compute the E_{HK} functional.

The mathematical development of DFT is due to W. Kohn and L. J. Sham [47]. Starting from a system of non-interacting particles under the influence of an effective potential, and assuming that the ground-state density of the non-interacting system of particles is the same as the ground-state density of the real system. Such a density can be written in the form

$$\rho(\mathbf{r}) = \sum_{i}^{occ} \left|\varphi_i(\mathbf{r})\right|^2 \tag{3.7}$$

where the sums runs over the occupied states and $\varphi_i(\mathbf{r})$ are the *Kohn-Sham* (KS) orbitals. Within this formalism, the kinetic energy term writes

$$T_0[\rho] = -\sum_i^{occ} \langle\varphi_i| \frac{\nabla^2}{2} |\varphi_i\rangle \tag{3.8}$$

and the Hartree energy term, which represents the classical Coulombian interaction between electrons, writes as

$$E_H[\rho] = \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\,\mathbf{r} d\,\mathbf{r}' \tag{3.9}$$

Finally, the KS approach consists in expressing the energy functional as:

$$E_{KS}\,[\rho] = T_0\,[\rho] + E_H\,[\rho] + E_{xc}\,[\rho] + \int V_{ext}(\mathbf{r})\rho(\mathbf{r})\,d\mathbf{r} \tag{3.10}$$

where E_{xc} is the *exchange-correlation* functional. This is a purely quantum quantity since it derives from Pauli principle and contains all the correlations effects between electrons that are not included in the kinetic and coulombian terms of the non-interacting system. By applying the variational principle described in the second HK theorem to the functional in Eq. (3.7) we can write

$$\frac{\delta E_{KS}}{\delta\rho(\mathbf{r})} = \frac{\delta T_0}{\delta\rho(\mathbf{r})} + \frac{\delta E_H[\rho]}{\delta\rho(\mathbf{r})} + \frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})} + \frac{\delta E_{ext}[\rho]}{\delta\rho(\mathbf{r})} \tag{3.11}$$

where we are considering an orthonormal set of KS orbitals $\langle\varphi_i|\varphi_j\rangle = \delta_{ij}$. From Eqs. (3.5) and (3.6) one gets

$$\frac{\delta\rho(\mathbf{r})}{\delta\varphi_i^*(\mathbf{r})} = \varphi_i(\mathbf{r}), \qquad \frac{\delta T_0}{\delta\varphi_i^*(\mathbf{r})} = -\frac{\hbar^2}{2m_0}\nabla^2\varphi_i(\mathbf{r}) \tag{3.12}$$

Putting this information into Eq. (3.10) we can now define the so-called *Kohn-Sham equations*

$$\left(-\frac{1}{2}\nabla^2 + V_{KS}(\mathbf{r})\right)\varphi_i(\mathbf{r}) = \epsilon_i\varphi_i(\mathbf{r}) \tag{3.13}$$

The term between parenthesis is the KS Hamiltonian. Eq. 3.13 can be seen as the single particle Schrödinger equation under the action of an effective potential that includes the Hartree potential, the exchange-correlation potential, and the external potential, that can be thus written as

$$\begin{aligned} V_{KS}(\mathbf{r}) &= V_H(\mathbf{r}) + V_{xc}(\mathbf{r}) + V_{ext}(\mathbf{r}) \\ &= \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}' + \frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})} + V_{ext}(\mathbf{r}) \end{aligned} \tag{3.14}$$

We now have a set of equations that have an exact analytical derivation and will allow us to solve the problem of the interacting electrons in an exact way. This is the main difference between DFT and the wave functions methods, such as HF: DFT is in principle an exact theory. Unfortunately one should know the exact form of the exchange-correlation potential to correctly address this problem, and so far nobody was able to give a form to such potential. Anyway, one can use approximations of this functional, that bring to a self-consistent resolution of the KS equations.

There exist two well-known approximations for the exchange-correlation functional:

- LDA (Local Density Approximation);
- GGA (Generalized Gradient Approximation).

In the LDA approximation [47, 52] all the non local effects are neglected, and only the value in a given point of the space defines the contribution to that point to the E_{xc}. This is generally expressed as

$$E_{xc}^{LDA}[\rho] = \int \epsilon_{xc}^{hom}\left[\rho(\mathbf{r})\right]\rho(\mathbf{r})\,d\mathbf{r} \tag{3.15}$$

where ϵ_{xc}^{hom} is the exchange-correlation energy per electron of the electron homogeneous gas with density ρ. The GGA approximation [52, 53] adds a term that considers the gradient of the electron density in a given point in the space, such that the distribution is not anymore homogeneous, taking somehow into account a non-local contribution. This is typically expressed as

$$E_{xc}^{GGA}[\rho] = \int \epsilon_{xc}\left[\rho(\mathbf{r}), \nabla\rho(\mathbf{r})\right]\rho(\mathbf{r})\; d\mathbf{r} \tag{3.16}$$

3.1.2 Practical use of DFT: Basis sets and pseudopotentials

To solve Eq. (3.12) one needs to introduce a basis set. We can indeed expand the KS wave function in a basis set $\{\phi_i\}$ such that

$$\varphi(\mathbf{r}) = \sum_i c_i\phi_i(\mathbf{r}), \tag{3.17}$$

where c_i are the coefficients of the expansion. In the case of crystals, a good choice for the basis set lies in plane waves because in this way we can easily describe the periodicity of solids. Considering an orthonormal basis set, we have

$$\sum_i \left(\langle\phi_j|\,\hat{H}_{KS}\,|\phi_i\rangle - \epsilon\delta_{ij}\right)c_i = 0. \tag{3.18}$$

In the case of one electron moving in an effective potential $V_{KS}(\mathbf{r})$ that has the same periodicity of the crystal ($V_{KS}(\mathbf{r}+\mathbf{R}) = V_{KS}(\mathbf{r})$, where $\mathbf{R}$ is the lattice vector of the primitive cell), for the Bloch theorem [54] we can write the wave function as

$$\varphi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}) \tag{3.19}$$

where n is the band index, $\mathbf{k}$ is a point in the Brillouin Zone (BZ), and $u_{n,\mathbf{k}}(\mathbf{r})$ is a function with the same periodicity of the lattice such that $(u_{n,\mathbf{k}}(\mathbf{r}+\mathbf{R}){=}u_{n,\mathbf{k}}(\mathbf{r}))$. The inclusion of the plane waves allows us to write the function $u_{n,\mathbf{k}}(\mathbf{r})$ as:

$$u_{n,\mathbf{k}} = \sum_{\mathbf{G}} c_{n,\mathbf{k}+\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{3.20}$$

where $c_{n,\mathbf{k}+\mathbf{G}}$ are the expansion coefficients and $\mathbf{G}$ is the reciprocal lattice vector. Combining Eq. (3.18) and Eq. (3.19) we finally have

$$\varphi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{n,\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}. \tag{3.21}$$

However, there are two practical problems: the first one is related to the fact that, in principle, this basis set involves an infinite sum over $\mathbf{G}$ vectors. Once again, the use of plane waves helps us to solve this problem in a straightforward way. Indeed these are solutions of the non-interacting problem with kinetic energy of the form

$$T_0 = \frac{1}{2} \left|\mathbf{k}+\mathbf{G}\right|^2 \tag{3.22}$$

The sum over the $\mathbf{G}$ vector is typically truncated in such a way that we can take into account only solutions with energy smaller than a *cut-off* energy [52], defined as

$$E_{cut} = \frac{1}{2} G^2_{cut}. \tag{3.23}$$

The infinite sum reduces then to a finite sum of the form

$$\varphi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{|\mathbf{k}+\mathbf{G}| \leq G_{cut}} c_{n,\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}. \tag{3.24}$$

The second problem is that the electronic wave function becomes really oscillating when approaching the nuclei. To reproduce this behaviour, we need a large number of plane waves. This problem is solved by the use of *pseudopotentials* (see Fig. 3.1). The concept of pseudopotentials is mainly based on the assumption that the properties of a material depend mostly on the valence electrons (frozen-core approximation). This is typically done in DFT calculations, where only the valence electrons are taken into account, and the core electrons are included in the ionic potential. This allows us to reduce the number of plane waves that have to be used.

A recent work in which the quality of pseudopotentials, considering also a comparison with the so-called *all-electron* codes, can be found in Ref. [55].

3.1.3 DFT problems

So far the general aspects of DFT have been described. We have also mentioned that by knowing the exact exchange-correlation functional many ground-state properties would have been computed exactly.

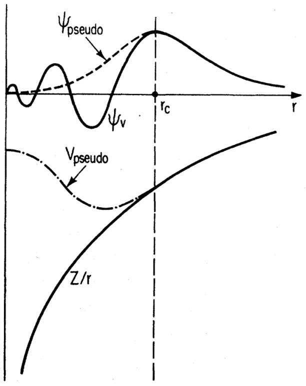


Figure 3.1: Schematic illustration of all-electron (solid lines) and pseudo-electron (dashed lines) potentials and their corresponding wave functions. The radius at which all-electron and pseudo-electron values match is designated r_c (From ref. [3]).

Nowadays, DFT is widely used even without this knowledge, and even wrongly sometimes (i.e., to compute non ground-state properties). This is mainly due to two reasons: the first one is that it is computationally fast and easy to implement. The second reason, that somehow bring us to the incorrect use of DFT, is that this is a well-established theory. This means, that not only the theoretical aspects are known, but also the limitations are well known [56]. Indeed, just to cite some common examples, it is known that DFT tends to underestimate the band gap, as well as reaction barriers, it is unable to describe localized electronic states etc. Since the band gap problem and description of localized electrons is an issue that somehow we experienced in this work, it is worth spending a few words on that.

3.1.3.1 Band gap problem

Considering a system with N electrons, the band gap can be defined as the difference between the largest addition energy and smallest removal energy [57]. If we define the ionization energy as $IE = E_N - E_{N-1}$, where E_N is the energy of the system with N electrons, and the electron affinity $EA = E_{N+1} - E_N$, then for the band gap we have

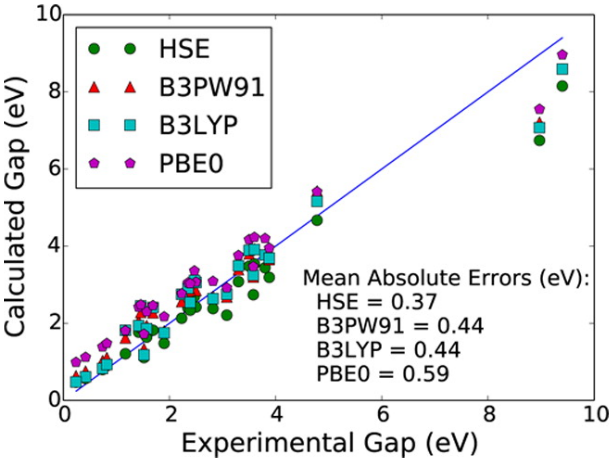
$$E_g = EA - IE = E_{N+1} - E_{N-1} - 2E_N. \tag{3.25}$$

However, KS band gaps are strongly underestimated w.r.t. experiments up to 50%. At this stage, it is worth to remember that KS eigenvalues and eigenvectors are purely mathematical quantities without any physical meaning, but this is not an explanation to this problem. The problem here is more deep, because it lies in the dependence of the exact energy functional upon the number of electrons and the inability of approximate functionals to reproduce it.

For the sake of completeness, we also have to mention that DFT being an independent particle approximation to the many-body interacting problem, quantities as Coulomb screening are of course wrong. To take into account this many-body interactions one should extend the treatment to the many-body perturbation theory going through the use of the *GW* methodology [58]. However, in this thesis, we were not concerned about this aspect.

Another way to overcome this problem is the use of hybrid functionals. In Fig. 3.2 we report the comparison of the experimental band gap and the band gap computed with different hybrid functionals [4]. The starting point of this approach is that the exchange-correlation functional can be split in two contributions: the exchange contribution E_x, that takes into account the exact exchange as considered in the HF theory, and the correlation functional E_c. Hybrid functionals try to improve semi-locals functionals, such as GGA and LDA, by adding explicit non-local properties of the electron density through a mixing of E_x^{DFT} and E_x^{HF}. The exact HF

Figure 3.2: Calculated vs experimental band gaps for the functionals studied in Ref. [4].



exchange in terms of KS orbitals, can be defined as:

$$E_x^{HF} = -\frac{1}{2}\sum_{\mathbf{k},n}\sum_{\mathbf{k}',n'}\int\int\frac{\varphi^*_{\mathbf{k},n}(\mathbf{r})\varphi_{\mathbf{k}',n'}(\mathbf{r})\varphi^*_{\mathbf{k}',n'}(\mathbf{r})\varphi_{\mathbf{k},n}(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} \tag{3.26}$$

Unfortunately, for periodic systems E_x^{HF} converges very slowly with distance. For this reason Heyd *et al* [59] proposed the separation of this term into a long-range part and a short-range part, in which only the short-range part is mixed:

$$E_{xc}(\mu,\alpha) = \alpha E_x^{HF,sr,\mu} + (1-\alpha)E_x^{PBE,sr,\mu} + E_x^{PBE,lr,\mu} + E_c^{PBE}. \tag{3.27}$$

Here α and μ are screening parameters. This is the idea behind the HSE hybrid functional. However it is worth mentioning that, even if these functionals give a better prediction of certain physical properties than semi-local functionals, as the band gap, they are computationally more expensive due to the bi-electron integrals (see Eq. (3.26)). Indeed, hybrid functionals are currently only implemented at ground-state (DFT) level. As a result, it is not possible to compute optical properties within Density Functional Perturbation Theory using hybrid functionals. Furthermore, these functionals are all parametrized and the value of the parameter it is usually fitted to get the correct experimental value.

3.1.3.2 GGA+U

An other problem that is worth discussing is the difficulty of semi-locals functionals in dealing with localized electrons such as d electrons in transition metal oxides. Indeed, there is a tendency to over delocalize electrons. This is mainly due to the non physical self-interaction (see Eq. (3.9)) of the electrons, that cancels out for many properties, but represents a huge issue when dealing with localized states. Naturally, the combination of these two errors can bring to totally wrong predictions. For instance, a small gap semiconductor ($E_g < 1.5\ eV$) can be seen as a metal in DFT.

One popular and computationally cheap method to correct for this self interaction error is to add a Hubbard U term to the KS Hamiltonian. This leads to the so-called LDA+U and GGA+U approaches. Following the work proposed by Dudarev *et al* [60], the GGA+U energy is given by

$$E^{\mathrm{GGA+U}} = E^{\mathrm{GGA}} + \frac{U-J}{2}\sum_{\sigma}\left[\sum_{m,m'} n^{\sigma}_{m,m'} n^{\sigma}_{m',m}\right] \tag{3.28}$$

The matrix $n^{\sigma}_{n,m}$ is the occupation matrix for a d orbital. The U and J parameters are respectively the on-site Coulomb and exchange parameters. These two parameters are typically combined in one effective U parameter ($U_{\text{eff}} = U - J$). In practice, the U value is seen as a parameter that has to be calibrated to get an agreement with respect to experimental results, in analogy with the parameters present in the hybrid functionals.

3.2 Density Functional Perturbation Theory

DFT formalism gives access to the ground-state properties of the system. However in this thesis we would like to investigate the optical properties of the crystals. To do that, we have to perturb the system from equilibrium applying a macroscopic electric field. This perturbation is typically small, and hence it can be accounted for via a perturbative approach. The interest in this methodology lies essentially in the fact that many interesting quantities are directly related to successive derivatives of the total energy with respect to different perturbations. Indeed, in a more general perspective, it can be shown that by focusing on quantities associated with first, and second order derivatives of the total energy with respect to atomic position $\mathbf{R}$, homogeneous strain ξ and homogeneous electric field $\mathcal{E}$, all the functional properties of interest can be defined, namely forces $\mathbf{F}$, stress tensor σ, polarization $\mathbf{P}$, interatomic force constants C, internal strain coupling parameter g, dynamical Born effective charges Z^*, clamped-ion elastic constants c^0 indirect clamped-ion piezoelectric tensor e^0, and optical dielectric tensor ε^∞ [11].

Two approaches have been developed for casting DFT within perturbation theory: a Green function approach, proposed by Baroni [61], and a variational approach, proposed by Gonze [62]. We

E	1^{st}-order	2^{nd}-order		
		$\frac{\partial}{\partial \mathbf{R}}$	$\frac{\partial}{\partial \xi}$	$\frac{\partial}{\partial \mathcal{E}}$
$\frac{\partial}{\partial \mathbf{R}}$	$\mathbf{F}$	C	g	Z^*
$\frac{\partial}{\partial \xi}$	σ	g	c^0	e^0
$\frac{\partial}{\partial \mathcal{E}}$	$\mathbf{P}$	Z^*	e^0	ε^∞

Table 3.1: Physical quantities related to the first and second order derivatives of the total energy E (see Ref. [11]).

will briefly see both approaches and how DFPT can be used to describe optical responses. At this stage, it is worth mentioning that DFPT being derived from a static theory, we can only access the response in that limit ($\omega = 0$ eV). Finally an extension of DFPT based on the $2n+1$ theorem is proposed for the calculation of the third order derivative related to the SHG susceptibility [63, 64].

3.2.1 Greens function method

The basic ansatz of perturbation theory is that every quantity (wave functions, energies, ...) can be written as a perturbation series. This basically means that we can write any quantity as ($X = H, E_i, \psi_i$):

$$X(\lambda) = X^{(0)} + \lambda X^{(1)} + \lambda^2 X^{(2)} + ... \tag{3.29}$$

where the expansion coefficients can be written as:

$$X^{(n)} = \frac{1}{n!} \frac{d^n X}{d\lambda^n}\bigg|_{\lambda=0} \tag{3.30}$$

.

Starting from the KS Hamiltonian (see Eqs. (3.13) (3.14)) we can use the perturbative approach to write:

$$H_{KS}(\lambda)\left|\varphi_n(\lambda)\right\rangle = \epsilon_n(\lambda)\left|\varphi_n(\lambda)\right\rangle . \tag{3.31}$$

where all the quantities are already considered in the KS basis set. Considering now the expansion of the Hamiltonian term, the wave functions and the energies, we can write

$$\begin{aligned}
&H_{KS}^{(0)}\left|\varphi^0\right\rangle + \\
&\lambda\left(H_{KS}^{(0)}\left|\varphi^1\right\rangle + H_{KS}^{(1)})\left|\varphi^0\right\rangle\right) + \\
&\lambda^2\left(H_{KS}^{(0)}\left|\varphi^2\right\rangle + H_{KS}^{(1)}\left|\varphi^1\right\rangle + H_{KS}^{(2)}\left|\varphi^0\right\rangle\right) + ... = \\
&\epsilon^{(0)}\left|\varphi^0\right\rangle + \\
&\lambda\left(\epsilon^{(0)}\left|\varphi^1\right\rangle + \epsilon^{(1)})\left|\varphi^0\right\rangle\right) + \\
&\lambda^2\left(\epsilon^{(0)}\left|\varphi^2\right\rangle + \epsilon^{(1)}\left|\varphi^1\right\rangle + \epsilon^{(2)}\left|\varphi^0\right\rangle\right) + ...
\end{aligned} \tag{3.32}$$

where the different perturbation orders are clear. It is worth mentioning that the normalization condition also has to be fulfilled ($\langle\varphi_i(\lambda)|\varphi_i(\lambda)\rangle = 1$), leading to

$$\begin{aligned}
&\left\langle\varphi_i^0\middle|\varphi_i^{(0)}\right\rangle + \\
&\lambda\left(\left\langle\varphi_i^0\middle|\varphi_i^{(1)}\right\rangle + \left\langle\varphi_i^1\middle|\varphi_i^{(0)}\right\rangle\right) + \\
&\lambda^2\left(\left\langle\varphi_i^0\middle|\varphi_i^{(2)}\right\rangle + \left\langle\varphi_i^1\middle|\varphi_i^{(1)}\right\rangle + \left\langle\varphi_i^2\middle|\varphi_i^{(0)}\right\rangle\right) + \\
&... = 1
\end{aligned} \tag{3.33}$$

In order for this to hold for any value if λ, we must have:

$$\begin{aligned}
&\left\langle \varphi_i^0 \middle| \varphi_i^{(0)} \right\rangle = 1 \\
&\left\langle \varphi_i^0 \middle| \varphi_i^{(1)} \right\rangle + \left\langle \varphi_i^1 \middle| \varphi_i^{(0)} \right\rangle = 0 \\
&\left\langle \varphi_i^0 \middle| \varphi_i^{(2)} \right\rangle + \left\langle \varphi_i^1 \middle| \varphi_i^{(1)} \right\rangle + \left\langle \varphi_i^2 \middle| \varphi_i^{(0)} \right\rangle = 0 \\
&\ldots
\end{aligned} \tag{3.34}$$

Looking at the first-order corrections to the energies and the wave functions we have

$$H_{KS}^{(0)} \left| \varphi_i^{(1)} \right\rangle + H_{KS}^{(1)} \left| \varphi_i^0 \right\rangle = \epsilon_i^{(0)} \left| \varphi_i^{(1)} \right\rangle + \epsilon_i^{(1)} \left| \varphi_i^{(1)} \right\rangle \tag{3.35}$$

For the first-order corrections to the energies, we can recover the *Hellman-Feynman theorem*

$$\epsilon_i^{(1)} = \left\langle \varphi_i^{(0)} \middle| H_{KS}^{(1)} \middle| \varphi_i^{(0)} \right\rangle . \tag{3.36}$$

And, looking at the first-order corrections to the wave functions, we recover the so called *Sternheimer equation.*

$$\left(H_{KS}^{(0)} - \epsilon_i^{(0)} \right) \left| \varphi_i^{(1)} \right\rangle = - \left(H_{KS}^{(1)} - \epsilon_i^{(1)} \right) \left| \varphi_i^{(0)} \right\rangle \tag{3.37}$$

If we now express the 1^{st} order wave function in terms of a combination of the 0^{th} order ones in KS basis

$$\left| \varphi_i^{(1)} \right\rangle = \sum_j c_{ij}^{(1)} \left| \varphi_j^{(0)} \right\rangle \tag{3.38}$$

we have

$$c_{ij}^{(1)} = \frac{\left\langle \varphi_j^{(0)} \middle| H_{KS}^{(1)} \middle| \varphi_i^{(0)} \right\rangle}{\epsilon_j^{(0)} - \epsilon_i^{(0)}} \tag{3.39}$$

These equations form a self-consistent set that must be solved in order to determine the behaviour of the perturbed system. Looking at the main DFT quantity, namely the electron density, and neglecting the spin, at the first-order we have:

$$\rho^{(1)}(\mathbf{r}) = 2\sum_{i}^{N}\sum_{j\neq i}\varphi_n^{(0)*}(\mathbf{r})\varphi_m^{(0)*}(\mathbf{r})\frac{\left\langle \varphi_m^{(0)} \middle| H_{KS}^{(1)} \middle| \varphi_n^{(0)} \right\rangle}{\epsilon_n^{(0)} - \epsilon_m^{(0)}} \tag{3.40}$$

The electron density only responds to perturbations that couple the valence and conduction many-folds. To calculate the projection of the first-order wave functions onto the conduction bands many-fold, we can introduce the projector on the conduction states P_c defined as

$$P_c = \sum_c \left|\varphi_c^{(0)}\right\rangle \left\langle\varphi_c^{(0)}\right| = 1 - \sum_v \left|\varphi_v^{(0)}\right\rangle \left\langle\varphi_v^{(0)}\right| \tag{3.41}$$

where the sum over v runs over the occupied (valence) states. In this framework, the Sternheimer equation can be written as

$$P_c(H_{KS} - \epsilon_v^0)P_c\left|\varphi_v^{(1)}\right\rangle = -P_c H_{KS}^{(1)}\left|\varphi_v^0\right\rangle \tag{3.42}$$

and we then have

$$\left|\varphi_v^{(1)}\right\rangle = G_v H_{KS}\left|\varphi_v^0\right\rangle \tag{3.43}$$

where G_v is the Green function operator projected onto the conduction bands

$$G_v = \sum_c \frac{\left|\varphi_c^{(0)}\right\rangle \left\langle\varphi_c^{(0)}\right|}{(\epsilon_v^{(0)} - \epsilon_c^{(0)})} \tag{3.44}$$

This method is also called *Green's function technique* for dealing with the Sternheimer equation. Solution of this linear problem requires only the knowledge of the occupied (valence) states, due to the projector, thus, even if the conductions states appear in Eq. (3.44) they are never explicitly required.

3.2.2 Variational approach

As we have already mentioned, the solution of the Schrödinger equation can be achieved by minimizing the total energy with respect to the wave function. A similar approach can be used in perturbation theory. This represents an alternative to the solution of the Sternheimer equation, even though both bring to the same physical results.

Minimization of the energy is possible through this variational expression

$$E^{(0)} = \min_{\varphi_\alpha^{(0)}} \left\{ \sum_{\alpha}^{occ} \left\langle \varphi_\alpha^{(0)} \right| H_{KS}^{(0)} \left| \varphi_\alpha^{(0)} \right\rangle \right\} \tag{3.45}$$

under the condition

$$\left\langle \varphi_\alpha^{(0)} \middle| \varphi_\beta^{(0)} \right\rangle \quad \forall\ \alpha, \beta \in \{occ\} \tag{3.46}$$

Expansion of the previous equation gives access to the perturbative approach. The second order derivative of the energy and the first-order wave function can be determined simultaneously through the minimization of the following equation

$$\begin{aligned} E^{(2)} = \min_{\varphi_\alpha^{(1)}} \Bigg\{ \sum_{\alpha} \Big[& \left\langle \varphi_\alpha^{(1)} \right| (H_{KS} - \epsilon_\alpha)^{(0)} \left| \varphi_\alpha^{(1)} \right\rangle + \\ & \left\langle \varphi_\alpha^{(0)} \right| H_{KS}^{(2)} \left| \varphi_\alpha^{(0)} \right\rangle + \left\langle \varphi_\alpha^{(0)} \right| H_{KS}^{(1)} \left| \varphi_\alpha^{(1)} \right\rangle + \\ & + \left\langle \varphi_\alpha^{(1)} \right| H_{KS}^{(1)} \left| \varphi_\alpha^{(0)} \right\rangle \Big] \Bigg\} \end{aligned} \tag{3.47}$$

under the constraints:

$$\left\langle \varphi_\alpha^{(1)} \middle| \varphi_\beta^{(0)} \right\rangle + \left\langle \varphi_\alpha^{(0)} \middle| \varphi_\beta^{(1)} \right\rangle = 0 \ \ \forall\ \alpha, \beta \in \{occ\} \tag{3.48}$$

The variational approach basically translates into DFPT by considering the solution of the following equation:

$$\begin{aligned} E^{(2)}\left[\varphi^{(0)}, \varphi^{(1)}\right] = \sum_{\alpha}^{occ} \Big[& \left\langle \varphi_{\alpha}^{(1)} \right| H_{KS}^{(0)} - \epsilon_{\alpha}^{(0)} \left| \varphi_{\alpha}^{(1)} \right\rangle + \\ & \left\langle \varphi_{\alpha}^{(0)} \right| V_{\text{ext}}^{(2)} \left| \varphi_{\alpha}^{(0)} \right\rangle + \left\langle \varphi_{\alpha}^{(0)} \right| V_{\text{ext}}^{(1)} \left| \varphi_{\alpha}^{(1)} \right\rangle + \\ & \left\langle \varphi_{\alpha}^{(0)} \right| V_{hxc}^{(1)} \left| \varphi_{\alpha}^{(1)} \right\rangle + \left\langle \varphi_{\alpha}^{(1)} \right| V_{hxc}^{(1)} \left| \varphi_{\alpha}^{(0)} \right\rangle \Big] + \\ & \frac{1}{2} \int \int K_{xc}(\mathbf{r}, \mathbf{r}') \rho^{(1)}(\mathbf{r}) \rho^{(1)}(\mathbf{r}') \, d\mathbf{r} \, d\mathbf{r}' + \\ & \frac{1}{2} \int \int \frac{\rho^{(1)}(\mathbf{r}) \rho^{(1)}(\mathbf{r}') \, d\mathbf{r} \, d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} + \frac{1}{2} \frac{d^2 E_{hxc}}{d\lambda^2} \end{aligned} \qquad (3.49)$$

where the Kernel $K_{xc}(\mathbf{r}, \mathbf{r}')$ is defined as

$$K_{xc}(\mathbf{r}, \mathbf{r}') = \frac{\delta^2 E_{xc}[\rho^{(0)}]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')} \qquad (3.50)$$

and the first-order change in the wave function, $\varphi_{\alpha}^{(1)}$ are varied under the constraint

$$\left\langle \varphi_{\alpha}^{(0)} \middle| \varphi_{\beta}^{(1)} \right\rangle = 0 \quad \forall \; \alpha, \beta \in \{occ\} \qquad (3.51)$$

Once again, both approaches are equivalent, leading to the same physical results.

More interestingly, the perturbative approach gives access to further energy derivatives. We can indeed derive two major variational perturbation theorems [49]:

- A variational principle can be exhibited for the even order of the perturbations

$$E^{(2n)} = \min_{\varphi^{(n)}} \left\{ \tilde{E}_{(\lambda)} \left[\sum_{i=0}^{n-1} \lambda^i \varphi_0^{(i)} + \lambda^n \varphi^{(n)} \right] \right\}^{2n} \tag{3.52}$$

- The energy up to the order $(2n+1)$ can be expressed through an expression that requires only the knowledge of the wave function up to the order n.

$$E^{(2n+1)} = \min_{\varphi^{(n)}} \left\{ \tilde{E}_{(\lambda)} \left[\sum_{i=0}^{n} \lambda^i \varphi_0^{(i)} \right] \right\}^{(2n+1)} \tag{3.53}$$

the last of these theorems is known as the $(2n+1)$ theorem and we will see in the next section how this has played a crucial role in our results.

3.3 Electric field perturbation

The response of a system to an external perturbation such as an electric field is generally given by the susceptibility tensor, that is the proportionality coefficient between the macroscopic polarization and the electric field. The most general expression writes

$$\mathbf{P} = \chi \mathcal{E} \tag{3.54}$$

As we have seen in Chap. 2, in a semiconductor the polarization can be expressed as a Taylor expansion of the macroscopic electric field

$$P_i = P_i^0 + \sum_j \chi_{ij}^{(1)} \mathcal{E}_j + \sum_j \chi_{ijk}^{(2)} \mathcal{E}_j \mathcal{E}_k + ... \tag{3.55}$$

Within DFT and its perturbative extension we have seen how to compute the density ρ, the wave functions φ, and energies E at different orders. Since nowadays these quantities are easy accessed by many DFT codes (see for example [65–67]) it is interesting to see how we can make a connection with the response of the materials, and thus the susceptibility at the different orders.

3.3.1 Dielectric tensor

At the linear order of the perturbation of Eq.(3.55), the susceptibility is related to static dielectric function through the relationship

$$\varepsilon_{\infty}^{\alpha\beta} = \delta_{\alpha\beta} + 4\pi\chi_{\alpha\beta} \tag{3.56}$$

This can be also defined as [68]:

$$\varepsilon_{\infty}^{\alpha\beta} = \delta_{\alpha\beta} + 4\pi\frac{\partial \mathbf{P}_{\alpha}}{\partial E_{\beta}} \tag{3.57}$$

Here we are considering only the electronic contribution to the dielectric tensor at clamped ions. However, there is a problem due to the fact that the electric field is generally described by a potential of the form

$$V_{\text{ext}} = \mathcal{E} \cdot \mathbf{r} \tag{3.58}$$

that is neither lattice periodic nor bounded from below. However, as can be seen in Sec. 3.2.1, we only need off-diagonal matrix elements, justifying the use of the following alternative expression

$$\langle\varphi_m|\,\mathbf{r}\,|\varphi_n\rangle = \frac{\langle\varphi_m|\,[H_{KS},\mathbf{r}]\,|\psi_n\rangle}{\epsilon_m - \epsilon_n} \quad \forall\ m \neq n \tag{3.59}$$

Now the quantity $|\bar{\varphi}_n^\alpha\rangle = P_c r_\alpha |\varphi_n\rangle$ can be rewritten considering the solution of a linear system

$$(H_{KS} - \epsilon_n) |\bar{\varphi}_n^\alpha\rangle = P_c [H_{KS}, r_\alpha] |\varphi_n\rangle . \tag{3.60}$$

Finally, the static dielectric tensor at clamped ions can be rewritten as

$$\varepsilon_{\alpha\beta}^\infty = \delta_{\alpha\beta} - \frac{16\pi}{N\Omega} \sum_{\mathbf{k}} \sum_v \left\langle \bar{\varphi}_n \middle| \frac{\partial \varphi_{m,\mathbf{k}}}{\partial \mathcal{E}_\beta} \right\rangle \tag{3.61}$$

3.3.2 Second harmonic generation tensor

One of the most important achievements of the perturbative approach, at least for the sake of this manuscript, is that the derivatives of the total energy can give us access to the properties of the perturbed system we are interested in. Within this spirit, Dal Corso *et al.* [63, 64] showed that extending this approach to higher order derivatives is possible to compute the response of a system beyond the linear regime. Indeed, we have that the second order susceptibility $\chi^{(2)}$ is related to the third order derivative of the total energy w.r.t. the electric field perturbation. However, from the $2n+1$ theorem (see Eq. (3.53)) we have that the latter can be computed just by the knowledge of the response of the system at the first-order. These are the same quantities necessary to compute the dielectric tensor.

Following the notation in Ref. [69], the SHG susceptibility tensor writes:

$$\chi_{ijk}^{(2)} = -\frac{3}{\Omega_0} E^{\mathcal{E}_i \mathcal{E}_j \mathcal{E}_k} \tag{3.62}$$

where

$$E^{\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k}=\frac{1}{6}\left.\frac{\partial^3 E}{\partial\mathcal{E}_i\partial\mathcal{E}_j\partial\mathcal{E}_k}\right|_{\mathcal{E}_i=0,\mathcal{E}_j=0,\mathcal{E}_k=0} \tag{3.63}$$

The general expression of third order energy derivative w.r.t. the electric field perturbation as implemented in the ABINIT [66] is the following

$$\begin{aligned}\tilde{E}^{\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k} =& \sum_\alpha \Big[\left\langle\psi_\alpha^{\mathcal{E}_i}\right|(T+V_{\text{ext}})^{\mathcal{E}_j\mathcal{E}_k}\left|\psi_\alpha^{(0)}\right\rangle \\ &+\left\langle\psi_\alpha^{\mathcal{E}_i}\right|(T+V_{\text{ext}}+V_{hxc})^{\mathcal{E}_j}\left|\psi_\alpha^{\mathcal{E}_i}\right\rangle \\ &+\left\langle\psi_\alpha^{(0)}\right|(T+V_{\text{ext}})^{\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k}\left|\psi_\alpha^{(0)}\right\rangle \\ &+\left\langle\psi_\alpha^{(0)}\right|(T+V_{\text{ext}})^{\mathcal{E}_i\mathcal{E}_j}\left|\psi_\alpha^{\mathcal{E}_k}\right\rangle\Big] \\ &-\sum_{\alpha,\beta}\Lambda_{\beta\alpha}^{\mathcal{E}_j}\left\langle\psi_\alpha^{\mathcal{E}_i}\middle|\psi_\beta^{\mathcal{E}_k}\right\rangle \\ &+\frac{1}{6}\int d\mathbf{r}\,d\mathbf{r}'\,d\mathbf{r}''\frac{\delta^3 E_{hxc}[\rho^{(0)}]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')\delta\rho(\mathbf{r}'')}\rho^{\mathcal{E}_i}(\mathbf{r})\rho^{\mathcal{E}_j}(\mathbf{r}')\rho^{\mathcal{E}_k}(\mathbf{r}'') \\ &+\frac{1}{2}\int d\mathbf{r}\,d\mathbf{r}'\frac{d}{d\mathcal{E}_j}\left.\frac{\delta^2 E_{hxc}[\rho^{(0)}]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\right|_{\mathcal{E}=0}\rho^{\mathcal{E}_i}(\mathbf{r})\rho^{\mathcal{E}_k}(\mathbf{r}') \\ &+\frac{1}{2}\int d\mathbf{r}\frac{d^2}{d\mathcal{E}_i d\mathcal{E}_k}\left.\frac{\delta^2 E_{hxc}[\rho^{(0)}]}{\delta\rho(\mathbf{r})}\right|_{\mathcal{E}=0}\rho^{\mathcal{E}_j}(\mathbf{r}) \\ &+\frac{1}{6}\left.\frac{d^3 E_{hxc}[\rho^{(0)}]}{d\mathcal{E}_i d\mathcal{E}_j d\mathcal{E}_k}\right|_{\mathcal{E}=0}\end{aligned} \tag{3.64}$$

with

$$V_{hxc}^{\mathcal{E}_j}=\int\frac{\delta^2 E_{hxc}[\rho^{(0)}]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\rho^{\mathcal{E}_j}(\mathbf{r}')d\mathbf{r}'+\frac{d}{d\mathcal{E}_j}\left.\frac{\delta E_{hxc}[\rho^{(0)}]}{\delta\rho(\mathbf{r})}\right|_{\mathcal{E}=0} \tag{3.65}$$

and

$$\Lambda_{\beta\alpha}^{\mathcal{E}_j}=\left\langle\psi_\beta^{(0)}\right|(T+V_{\text{ext}}+V_{hxc})^{\mathcal{E}_j}\left|\psi_\alpha^{(0)}\right\rangle \tag{3.66}$$

Defining $E^{\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k}$ as the sum over permutations of the 3 perturbations $\tilde{E}^{\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k}$ we finally have

$$\begin{aligned} E^{\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k} =& \frac{1}{6}\Big(\tilde{E}^{\mathcal{E}_i\mathcal{E}_j\mathcal{E}_k} + \tilde{E}^{\mathcal{E}_i\mathcal{E}_k\mathcal{E}_j} + \tilde{E}^{\mathcal{E}_j\mathcal{E}_i\mathcal{E}_k} \\ & + \tilde{E}^{\mathcal{E}_j\mathcal{E}_k\mathcal{E}_i} + \tilde{E}^{\mathcal{E}_k\mathcal{E}_j\mathcal{E}_i} + \tilde{E}^{\mathcal{E}_k\mathcal{E}_i\mathcal{E}_j}\Big) \end{aligned} \tag{3.67}$$

Further details about the implementation and the analytic derivation of this quantities can be found in Ref. [69] and references therein. Once again, it is worth to remember that, the general expression of the SHG susceptibility tensor depends on the frequencies of the optical electric field. Here we are computing the tensor at $\omega = 0\ eV$ and as a consequence the computed $\chi^{(2)}$ satisfies the Kleinman's symmetry condition (see Sec. 2.2.4), which means that it is symmetric under the permutation of the $i,\ j,\ k$ index. In order to obtain its frequency dependence, one would need either to apply the formalism of the Time Dependent Functional Theory (TDDFT) or to use expressions that involves a sum over excited states (see for example Refs. [70, 71].

3.4 Automatization of the calculations: the high-throughput approach

The development of efficient ab-initio software packages and the constant increase of the computational power of modern supercomputer at our disposal opened a new approach in Material Science. Many physical quantities are made nowadays accessible from first principles. From this point of view, the success of DFT and DFPT stems from their reliability and low computational cost. As a result, first-principles calculations have recently been combined with a high-throughput (HT) approach. The idea behind HT ab-initio computing is to evaluate these properties for several thousands of known or predicted compounds [17, 50]. The combination of these two methods enables the creation of large databases with material properties that would be prohibitive (in time and cost) for experimental measurements. By screening those databases, new materials, targeting specific applications, can be identified. Successful discoveries (i.e., prediction confirmed in the laboratory) include materials for batteries, hydrogen production and storage, thermoelectrics and photovoltaics (see, e.g., Refs. [17] and [18]). However, HT techniques have not only allowed to search for materials candidates for a specific application. Indeed, Databases can be also analyzed using data mining techniques, aiming at identifying trends and limitations that can give a further insight into the comprehension of the materials properties, or even make predictions for unknown compounds through Machine Learning (see, for example, Refs. [19] and [20]). Before moving to large scale calculation it is mandatory to implement a robust procedure to handle the whole process and provide a reliable set of input parameters that suits most of the possible cases considered. In this regard, it has been shown how the validation of the results is an element of great relevance when approaching the HT regime. Due to the amount of data generated during high-throughput computational calculations, a careful storage of the data in a database is necessary. This database stores information for each computed compound: initial and relaxed crystal structures, chemical formula, total energy computed DFT and DFPT parameters. Following this trend, several open materials repository have been created containing a number of properties obtained from DFT calculations thanks to ad-hoc frameworks developed to help the automation of the whole process (see, e.g., Refs. [17] and [18]).

The response related to the first-order change in the wave function, thus the second order derivative of the energy (see Table 3.1) can be performed very efficiently using DFPT. Works in which DFPT has been combined with HT has been recently presented (see for example Refs. [22, 23, 72, 73]). In this perspective, the idea is to extend the current quantities stored in the materials repository beyond ground-state theory. As an example, the Materials Project recently included in its repository quantities such as the phonon bandstructure, static dielectric function, Born effective charges and other quantities obtainable via DFPT.

In the next two sections the guidelines for the creation of two databases based on DFPT calculations, and containing the optical properties of materials are given. In the first one, the materials optical properties in the linear regime, electronic and structural properties are computed via DFPT and DFT using the VASP [65] code. In the second one, the nonlinear SHG tensor has been computed via DFPT and the $(2n + 1)$ theorem as implemented in ABINIT [69]. For reasons that will be clear soon, it is worth to mention that the current implementation only supports LDA functionals.

3.4.1 Linear database

For the construction of the linear DB, we start from the relaxed structures available in the Materials Project (MP) repository [74]. Their thermodynamical stability can be assessed by the energy above hull E_{hull} [75, 76]: for a stable compound, $E_{\text{hull}} = 0\,\text{meV/atom}$, and the stability decreases as E_{hull} increases. Here, we extract the materials with $E_{\mathrm{hull}} \leq 25$ meV/atom [77]. We also include a few exceptions (with $E_{\mathrm{hull}} > 25$ meV/atom) already investigated previously in the literature for technological applications. Among all the possible candidates, a set of 4040 semiconductors, for which also electronic structural properties are present in the MP, has been extracted. The 4040 selected materials cover a broad range of chemistries (oxides, fluorides, sulfides, ...) with various compositions (binaries, ternaries, ...). However, a significant fraction of those (3375 out of 4040) are oxides, since they show important applications in many sectors (semiconductor industry, catalysts, ...) with an exceptionally broad range of electronic properties (see Ref. [78]).

For all those structures, the static part of the refractive index n_s is computed in the framework of DFPT. All the calculations are performed with the VASP software package [65], adopting the projector augmented wave (PAW) method [79], and using the Generalized Gradient Approximation (GGA) for the exchange-correlation functional as parameterized by Perdew, Burke, and Ernzerhoff (PBE) [53]. The workflow to handle the calculation is implemented in the Fireworks package When dealing with oxides including elements with partially occupied d electrons (such as V, Cr, Mn, Fe, Co, Ni, or Mo) a Hubbard-like Coulomb U term is added to the GGA (GGA+U) [60] to correct the spurious GGA self-interaction energies, adopting the U values advised by the MP [80]. We use the suggested value proposed from MP for the DFT static calculation (density of 1000 points per reciprocal atoms) also for the k-point sampling and the energy cut off. The JDOS of the materials are calculated from the band structures available in the MP [81]. For the band gap, we focus on the direct band gap, E_g^d, since optical processes are related to vertical transitions. Once again, it is worth pointing out that DFT is known to underestimate the band gap up to the 50% with respect to experiments (see for example Ref. [82, 83]), while a tendency to overestimate n_s is to be expected. Further details on the validation of a similar workflow and on the error of the refractive index computed via DFPT can be found in Ref [22].

Finally, the results (dielectric function, refractive index, space groups, etc.) are stored using the MongoDB database engine, a NoSQL database program that favors the use of JSON-like documents [84]. These data are publicly available online in CSV (Comma Separated Value) format (see Ref. [24]).

3.4.2 Nonlinear database

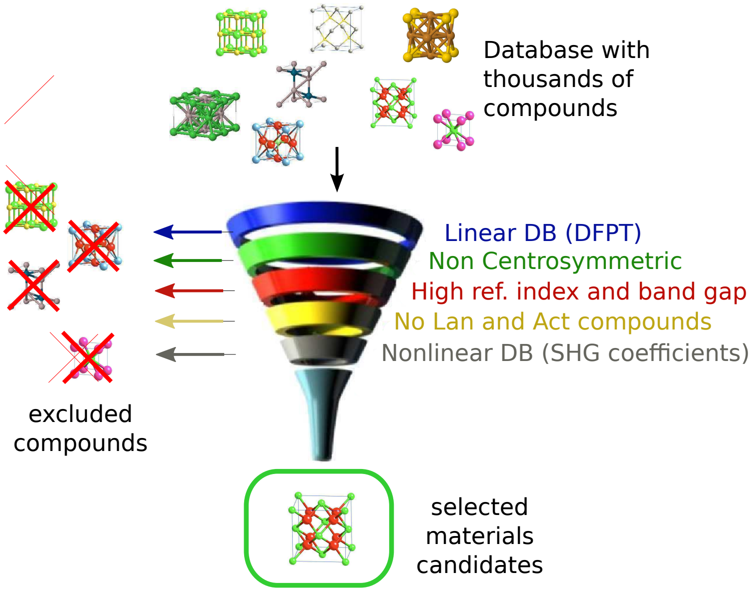
To compute the nonlinear SHG tensor d_{ij}, we used the open source code ABINIT. The workflow for the HT calculation of the coefficients have been implemented in the Abiflows package. The Pymatgen and Abipy python packages are used to generate inputs and analyze the results.

The theoretical framework that enables us to compute the d_{ij} tensor has already been introduced in Sec. 3.3.2. However, it is worth mentioning that in the ABINIT implementation available at the beginning of our work (see Ref. [69]), this quantity was only accessible by activating as well the calculation of the response of the material with respect to atomic displacements, namely phonons. However, perturbations that involve the atomic displacements can be much more complex and heavy from a computational point of view. We thus modified the implementation by including a new tag for the

ABINIT variable *nlflag* (*nlflag*=3, see ABINIT documentation for further details [85]). In this way we only computed the SHG tensor, that is independent from the phonon perturbations and that is the quantity in which we are mainly interested at this stage.

The screening of the 457 materials started from the initial list of candidates contained in the linear DB (4040). The full screening procedure is depicted in Fig. 3.3

Figure 3.3: Sketch of the screening procedure followed to build up the database containing the nonlinear optical coefficients of the semiconductors analyzed in this study.

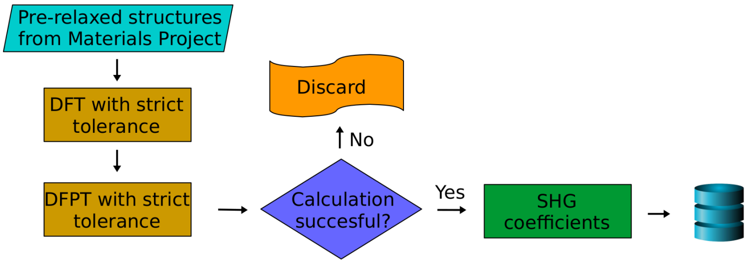


At this point is worth to remind that only materials with lack of inversion symmetry show non zero SHG susceptibility. We then extract a list of 824 non centrosymmetric candidates. A further screening is based on discarding Lan-Act materials, reducing the number to 695. The main reason is due to the lack of xc functionals for these elements. At this stage, the materials that reached the required accuracy in all the calculation steps are then 457. In fact, the calculation of the second order susceptibility tensor requires a high accuracy for the calculation of the total energy (up to 10^{-22} Ha), wave functions (up to 10^{-22}), and their derivatives with respect to the external perturbation (up to 10^{-22} atomic unit). Such an accuracy is difficult to achieve for some systems, such as magnetic systems, that have been thus discarded from our analysis. Finally, to optimize the materials search from an applied point of view, the renormalization of the band gap has been computed so far to 369 out of the 457 within the HSE06 hybrid functional, as implemented in VASP.

For the calculation of the d_{ij} tensor we select the 457 pre-relaxed structures from the MP. The workflow implemented in Abipy and Abiflows is organized as follows: We first use self-consistent and non self-consistent DFT calculations to compute the wave function and the density. The following step consists in running DFPT simulations to obtain the second order derivatives of the energy w.r.t. the electric field. This is basically achieved in two steps: the calculation of the derivative of the wave-function w.r.t. their wave vector, and the calculation of the derivative of the first order wave function w.r.t. the applied external electric field. If the calculations are completed correctly and high accuracy is reached, the set of derivatives is then used in the last step that involves the calculation of the third order derivative w.r.t. the electric field to obtain the d_{ij} tensor. This workflow is outlined in Fig. 3.4 and is part of a more general DFPT workflow that has been implemented in the python packages already mentioned, which allows to access all the quantities that can be computed in the framework of DFPT.

For this work, we used the LDA xc functionals [86]. We then

Figure 3.4: Schematic overview of the work flow implemented to compute the SHG tensor for the 457 semiconductors.



have chosen the Norm-Conserving pseudopotentials available in the PseudoDojo [87]. The cutoff was chosen independently for each material according to the values suggested in the PseudoDojo. These pseudopotentials and cutoff have been carefully tested w.r.t. all electron codes. For the sampling of the Brillouin Zone, convergence is achieved for a density of 3000 points per reciprocal atom (using a grid that respects the symmetry of the system) .For the few materials for which experimental data are available in the literature the converged results are in pretty good agreement with the experimental measurements (in a range between 10% and 15 %).

Finally the results are stored in the Mongo DB database engine. Even though the only new quantity, w.r.t. the linear DB, is the SHG tensor, we stored many of the physical quantities that were previously computed to check the consistency between the two implementations.

4

Results: Linear and nonlinear material databases

In the second chapter of this thesis, we have introduced the theoretical framework and the optical properties we want to deal with. In the third chapter, DFT and DFPT have been introduced as a way to compute the optical materials properties, and we also introduced how these approaches can be merged with the HT technique in order to generate a large set of data.

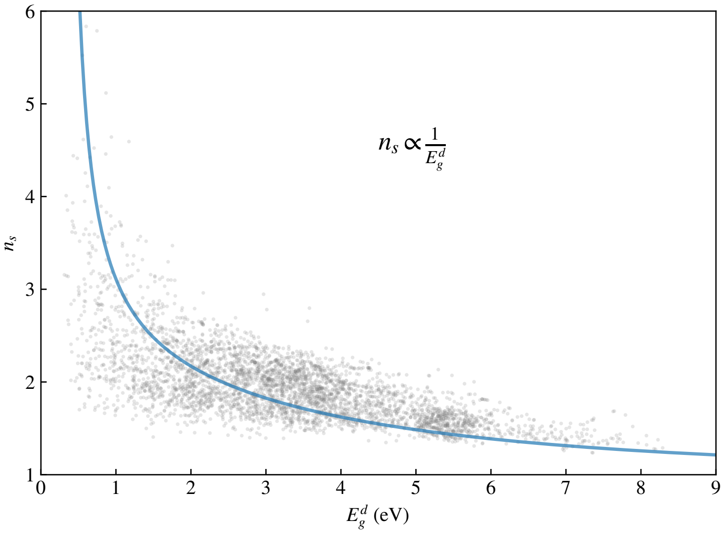
In what follows our results are presented. In the first part, we show the analysis of the linear optical properties for the 4040 compounds. The main focus for this part will be on materials with both high refractive index and band gap. However, these two properties are typically described by an inverse correlation with high refractive index appearing in small gap materials and vice-versa. Our data confirm the general inverse trend between refractive index and band gap but interesting outliers are also identified. The data are then analyzed via data mining to obtain a descriptive model that can help in the comprehension of the data distribution. This work has been published in Physical Review Materials [24].

In the second part we show a database containing more than 400 semiconductors, screened out from the first DB, for which we compute the SHG susceptibility tensor. The analysis of the latter can open different perspectives: from the data mining comprehension of the data distribution, to the analysis including machine learning model, passing through an extension of the computed quantities that would be helpful from the point of view of materials discovery.

4.1 Linear optical materials database

We investigate the relationship between the refractive index and the band gap using a first-principles HT approach relying on DFT and DFPT. Our aim is to provide a statistical, “data driven”, analysis based on a large set of 4040 semiconductors. Calculated data confirm the global inverse trend between those two properties, as recently discussed in similar works [22, 23, 88]. However, there is also a wide spread of the data around this general tendency, pointing out some outliers with both relatively high refractive index and wide band gap among which well-known materials (TiO_2, $LiNbO_3$, ...), already widely used for optical applications, and other materials, not yet considered for such applications (Ti_3PbO_7, $LiSi_2N_3$, BeS, ...). By mapping all the compounds onto a two-state system, a simple model is derived some descriptors of which can be accessed from the electronic structure. The density of states (DOS) at the valence and conduction band edges as well as the effective masses of those bands are found to play a critical role for achieving a high refractive index and a wide band gap simultaneously. Indeed, the availability of a large number of weakly dispersive states for optical transitions can partly counterbalance the inverse relationship between the refractive index and the band gap. Based on these considerations, we focus on the 3375 oxides present in the data set. We examine these materials in terms of their chemistry and pinpoint the most interesting ones. Before proceeding with the analysis of the results, it is useful to take a look at our raw data in Fig. 4.1. Starting from Eq. (2.10), the inverse relationship between the refractive index and the band gap it is already evident. Indeed, the imaginary part of the dielectric function can be written as [89]

Figure 4.1: Calculated data points (refractive index n_s vs. band gap E_g^d) for the 4040 materials considered here represented by grey circles. The blue line represents a guide line to indicate the inverse relationship between these two quantities.



$$\varepsilon_2(\omega) = \frac{4\pi^2}{\omega^2} \sum_{v,c} \int_{\text{BZ}} \frac{2d\mathbf{k}}{(2\pi)^3} \mid \hat{e} \cdot M_{cv}(\mathbf{k}) \mid^2 \times \delta(\epsilon_c(\mathbf{k}) - \epsilon_v(\mathbf{k}) - \omega) \quad (4.1)$$

where $\hat{e}$ is the polarization vector in the direction of the electric field and $M_{cv}(\mathbf{k})$ is the dipole matrix element for a transition from a valence state $\epsilon_v(\mathbf{k})$ to a conduction state $\epsilon_c(\mathbf{k})$. The imaginary part written in this form, that is anyway obtainable considering the quantum approach discussed in Sec. 2.2.2, will be useful to have direct connection with the joint density of states (JDOS) in what follows. The real part of the dielectric function and the imaginary part are related through the Kramers-Kronig relation

$$\varepsilon_1(\omega) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{\omega' \varepsilon_2(\omega')}{\omega'^2 - \omega^2} \, d\omega' \tag{4.2}$$

where P indicates the principal part of the integral. We have then for the real part of the dielectric function

$$\begin{aligned} \varepsilon_1(\omega) = 1 + 8\pi \sum_{v,c} \int_{\mathrm{BZ}} \frac{2d\mathbf{k}}{(2\pi)^3} \\ \frac{\mid \hat{e} \cdot M_{cv}(\mathbf{k}) \mid^2}{\epsilon_c(\mathbf{k}) - \epsilon_v(\mathbf{k})} \frac{1}{(\epsilon_c(\mathbf{k}) - \epsilon_v(\mathbf{k}))^2 - \omega^2} \end{aligned} \tag{4.3}$$

From these considerations, the inverse relationship between refractive index (see Eq. 2.10) and band gap is clear, but it is as well clear that there is a large spread of the points around the blue curve in Fig. 4.1, used as a guideline and it is also evident that the band gap is not a sufficient quantity to properly describe the data trend.

4.1.1 Global trend

Various models have been proposed in the literature to describe the inverse relationship between the refractive index and the band gap.

In Fig. 4.2, the models proposed by Ravindra *et al.* [5] (green line):

$$n_s = 4.084 - 0.62E_g^d,$$

Moss [6] (red line):

$$n_s = \left(\frac{95}{E_g^d}\right)^{1/4},$$

Hervé and Vandamme [7] (cyan line):

$$n_s = \sqrt{1 + \left(\frac{13.6}{E_g^d + 3.47}\right)^2},$$

Reddy and Anjaneyulu [8] (magenta line):

$$n_s = \left(\frac{154}{E_g^d - 0.365}\right)^{1/4},$$

and Kumar and Singh [9] (yellow line):

$$n_s = 3.3668\left(E_g^d\right)^{-0.32234}$$

are superimposed on our calculated data. A detailed discussion of these models can be found in Ref. [90]. It is clear that none of them follow closely the trend of the data (their mean absolute errors (MAE) range from 0.42 to 0.91) nor do they account for the wide spread of the points. It should however be mentioned that all these models were built up using a small set of experimental data points (≤ 100).

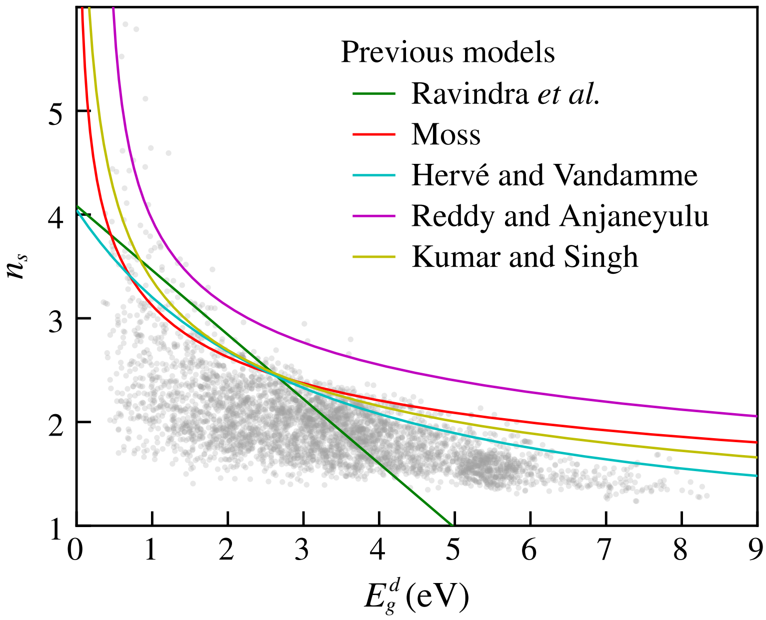


Figure 4.2: Comparison of the calculated data points (refractive index n_s vs. band gap E_g^d) with various well-known empirical and semi-empirical models [5–9]. The data points for the 4040 materials considered here are represented by grey circles, while the models are indicated by solid lines.

For these reasons, we build a new model to describe the data, starting from the basic relationship between refractive index and dielectric function, that in the static case writes

$$n_s = \sqrt{\varepsilon_{1s}}. \tag{4.4}$$

Here the subscript s represents the static limit ($\omega = 0$ eV). In the static limit, Eq. (4.2) becomes:

$$\varepsilon_{1s} = 1 + \frac{2}{\pi} \int_0^\infty \frac{\varepsilon_2(\omega)}{\omega} \, d\omega. \tag{4.5}$$

In principle, the above integral has to be taken from zero to infinity. In practice, a typical $\varepsilon_2(\omega)$ spectrum usually reveals well-separated peak regions, with little overlap, due to different absorption processes. Therefore, one can set an upper frequency limit ω_{max} which is high enough compared to the optical absorption processes of interest here, but small compared to other ones. In this work, ω_{max} is defined in such a way that:

$$\left(\int_0^{\omega_{\mathrm{max}}} \frac{\varepsilon_2(\omega)}{\omega} \, d\omega \Big/ \int_0^{\infty} \frac{\varepsilon_2(\omega)}{\omega} \, d\omega \right) \geq 99\%. \tag{4.6}$$

As already stated in the previous chapter, the value of ε_{1s} and hence n_s can be directly calculated using DFPT at low computational cost [48, 61, 68, 91, 92]. Indeed, conduction states do not need to be taken into account in contrast with the sum over states formulation within the random-phase approximation [93, 94]. The drawback of the DFPT approach is that only the static limit of the dielectric function is computed and hence the frequency dependence is not available. This can be partly circumvented as follows. We first introduce the JDOS:

$$j(\omega) = \sum_{v,c} \int_{\mathrm{BZ}} \frac{2d\mathbf{k}}{(2\pi)^3} \delta(\epsilon_c(\mathbf{k}) - \epsilon_v(\mathbf{k}) - \omega), \tag{4.7}$$

which can easily be obtained from DFT calculations of the electronic band structure. We note its similarity with Eq. (4.1). As a result, we define a frequency-dependent transition probability $k(\omega)$ such that:

$$\varepsilon_2(\omega) = \frac{4\pi^2}{\omega^2} k(\omega) j(\omega). \tag{4.8}$$

We note that, if the matrix elements $\mid \hat{e} \cdot M_{cv}(\mathbf{k}) \mid^2$ were all equal to a constant K, we would simply have $k(\omega) = K$. In Fig. 4.3, we show, as an example, a comparison between the frequency-dependent transition probability $k(\omega)$ and the constant value K for a real material. Consequently, a simple approximation for the

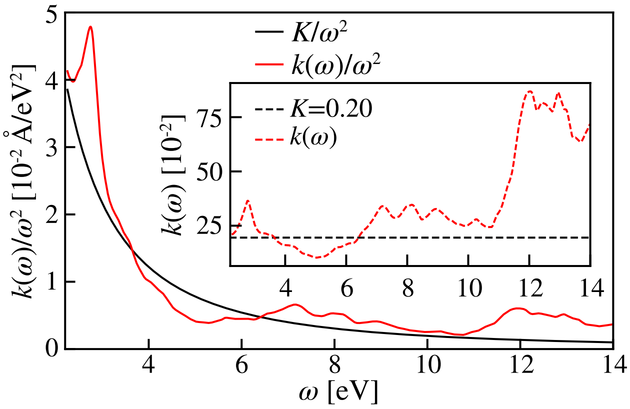


Figure 4.3: Comparison of the frequency-dependent transition probability $k(\omega)/\omega^2$ and its constant value K/ω^2 obtained from Eq. (4.10) for a real material (TiO$_2$, mp-2657). In the inset the comparison is given for $k(\omega)$ and K. All functions are shown in a frequency range $[2, \omega_{max}]$, with ω_{max} equal to 14 eV for this material.

imaginary part of the dielectric function can be obtained as [89]:

$$\tilde{\varepsilon}_2(\omega) = 4\pi^2 K \frac{j(\omega)}{\omega^2}. \tag{4.9}$$

The value of K is determined such that $\tilde{\varepsilon}_2(\omega)$ also satisfies the Kramers-Kronig relation given by Eq. (4.5). This is strictly equivalent to defining K as a weighted average of $k(\omega)$ as follows:

$$K = \int_0^{\omega_{\max}} k(\omega)\frac{j(\omega)}{\omega^3}d\omega \bigg/ \int_0^{\omega_{\max}} \frac{j(\omega)}{\omega^3}d\omega \tag{4.10}$$

A comparison of $\varepsilon_2(\omega)$ with $\tilde{\varepsilon}_2(\omega)$ is given in Fig. 4.4. Using this approximation for the imaginary part of the dielectric function,

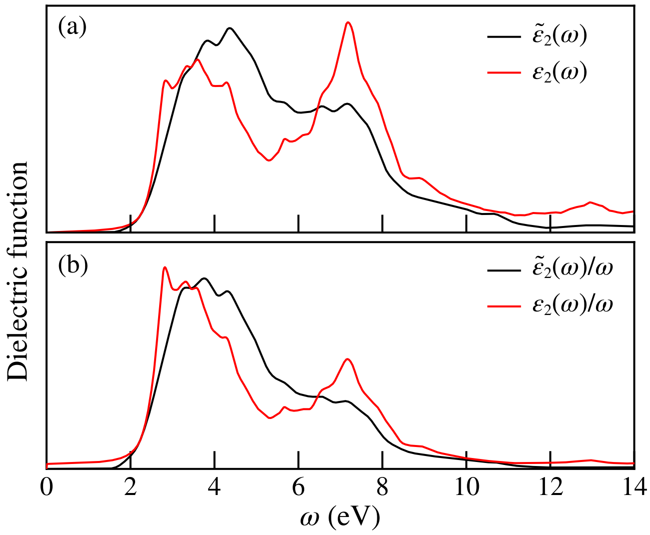


Figure 4.4: Comparison of the imaginary part of (a) the dielectric function $\varepsilon_2(\omega)$ and (b) $\varepsilon_2(\omega)/\omega$ considering the two methodologies of calculation for a real material (TiO$_2$, mp-2657). The red curves are obtained averaging the diagonal components of the DFT imaginary part of the dielectric function (Eq. (4.1)). The black curves are obtained via a renormalization of the $j(\omega)$ (Eq. (4.9)).

Eq. (4.5) can be rewritten as follows:

$$\begin{aligned}\varepsilon_{1s} &= 1 + \frac{2}{\pi}\int_0^{\omega_{\max}} \frac{\tilde{\varepsilon}_2(\omega)}{\omega}\,d\omega \\ &= 1 + 8\pi K \int_0^{\omega_{\max}} \frac{j(\omega)}{\omega^3}\,d\omega. \end{aligned} \tag{4.11}$$

Introducing the integral of the JDOS J, we further define the effective frequency ω_{eff}:

$$\begin{aligned}\omega_{\text{eff}} &= \left(\frac{2}{\pi}\int_0^{\omega_{\max}} \omega^2 \tilde{\varepsilon}_2(\omega)\,d\omega\right)^{\frac{1}{3}} \\ &= \left(8\pi K \int_0^{\omega_{\max}} j(\omega)\,d\omega\right)^{\frac{1}{3}} = (8\pi K J)^{\frac{1}{3}} \end{aligned} \tag{4.12}$$

and the average optical gap ω_g:

$$\begin{aligned}\omega_g &= \left(\int_0^{\omega_{\max}} \omega^2 \tilde{\varepsilon}_2(\omega)\, d\omega \Big/ \int_0^{\omega_{\max}} \frac{\tilde{\varepsilon}_2(\omega)}{\omega}\, d\omega \right)^{\frac{1}{3}} \\ &= \left(\int_0^{\omega_{\max}} j(\omega)\, d\omega \Big/ \int_0^{\omega_{\max}} \frac{j(\omega)}{\omega^3}\, d\omega \right)^{\frac{1}{3}} . \end{aligned} \tag{4.13}$$

Finally, we can thus write:

$$n_s^2 = \varepsilon_{1s} = 1 + \left(\frac{\omega_{\text{eff}}}{\omega_g} \right)^3 . \tag{4.14}$$

The model resulting from the present study is reported in Fig. 4.5. It captures the trend better than all previous models (MAE=0.33 considering $\omega_{\text{eff}} = 12.10$ eV, calculated by fitting Eq.(4.14) in the last square sense) and the spread in the data can be accounted for through the parameter ω_{eff}.

In our model we thus map each material onto the simplest system that one can think of for describing optical transitions: a two-state (E_1, E_2) system with a transition characterized by (i) an energy $\omega_g = E_2 - E_1$, (ii) a probability K, and (iii) a degeneracy factor $J = n_1 n_2$, where n_1 (resp. n_2) is the degeneracy of the state E_1 (resp. E_2). In the mapping procedure, which is schematically illustrated in the left panel of Fig. 4.6, ω_g is obtained as the weighted average of the transitions contributing to the optical properties (it will hence be referred to as the average optical gap). J is simply the integral of $j(\omega)$, the corresponding joint density of states (JDOS), and K is the average probability of those transitions.

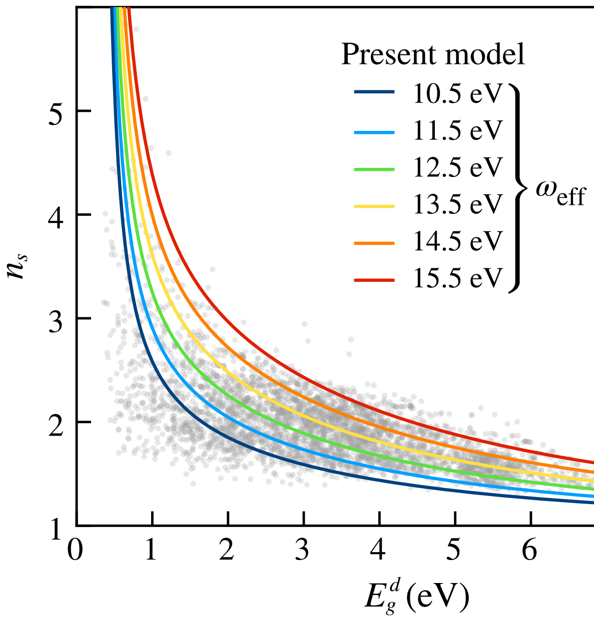


Figure 4.5: Comparison of the calculated first-principles data points (refractive index n_s vs. band gap E_g^d) with the model described by Eq. (4.18). The data points for the 4040 materials considered here are represented by grey circles, while the models are indicated by solid lines. Different values of the parameter ω_{eff} have been considered, accounting for the spread in the data points.

As can be seen from Fig. 4.6, the average optical band gap ω_g is related to the direct band gap E_g^d by:

$$\omega_g = E_g^d + \Delta. \tag{4.15}$$

The value of Δ is material dependent since it is influenced by the dispersion of the valence and conduction bands involved in the transition and their distribution in energy and, indirectly, by the direct band gap E_g^d as shown in Fig. 4.7.

The calculated values of ω_g (see Eq. 4.13) and E_g^d, achieved via ab-initio computations, are shown in Fig. 4.8 for all the materials

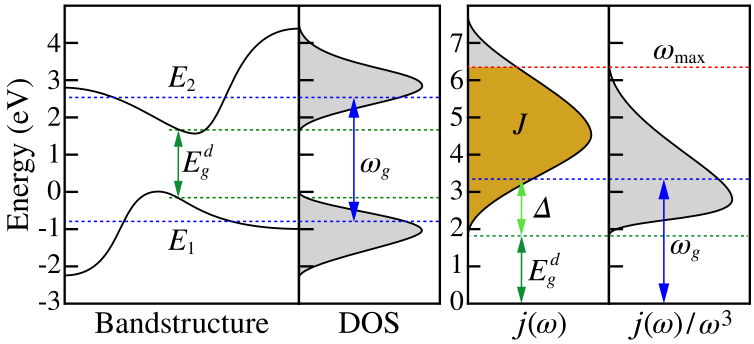


Figure 4.6: (Left panel) Schematic illustration of the mapping procedure from the electronic structure [band structure and DOS in solid black lines] is replaced by a two-state system (E_1, E_2 in dashed blue lines). (Right panel) Optical functions $j(\omega)$, the JDOS, and $j(\omega)/\omega^3$. The direct band gap E_g^d and the average optical gap ω_g are indicated by green and blue dotted lines, respectively. The difference Δ between ω_g and E_g^d is also reported in light green. The optical function $j(\omega)/\omega^3$ is used to determine the upper frequency limit $\omega_{\max}$ for the optical absorption processes, as indicated by the red dotted line. The integral of $j(\omega)$ up to $\omega_{\max}$ leads to the value of J, the degeneracy factor of the transitions between the two states.

considered here.

We can describe the relationship between the quantities in Eq. (4.15) by the following equation (see Appendix A for a more detailed description):

$$\omega_g = E_g^d + \alpha + \frac{\beta}{E_g^d} \tag{4.16}$$

where α=6.74 eV and β=-1.19 eV^2. However, we note that there is a wide spread of the data around the interpolated value (which translates into a quite large MAE of 1.20 eV for the fit).

This can be traced back to the dependence of ω_g on the width of the JDOS (see Fig. A.2) or, in other words, distribution in energy of the transitions. The simplest physical quantity that can account

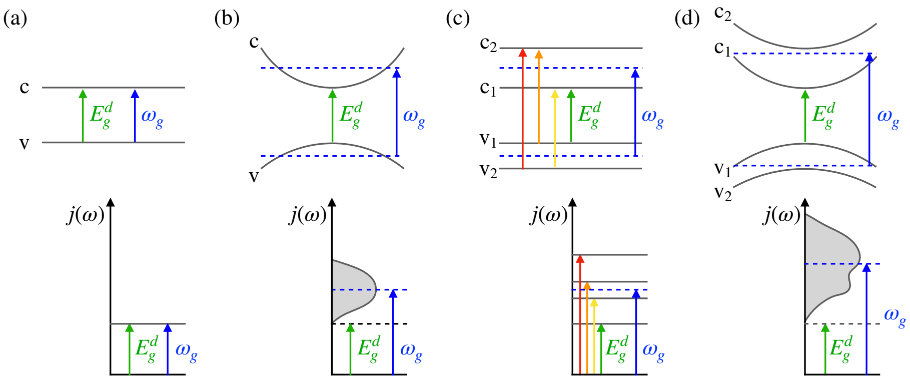


Figure 4.7: Schematic illustration of the dependence of the average optical gap ω_g on the width of the JDOS $j(\omega)$. Starting from a two-state system (a) with flat bands for which ω_g coincides with the direct band gap E_g^d and $j(\omega)$ is a Dirac peak, the graphs shows how ω_g is affected by (b) the dispersion of the bands which increases the width of the JDOS, (c) the band distribution in energy when new flat bands are added leading to new Dirac peaks in $j(\omega)$, and (d) the combination of both.

for this is the inverse effective mass of the transition [95] defined by

$$\frac{1}{\mu} = \frac{1}{m_v^*} + \frac{1}{m_c^*} \tag{4.17}$$

where m_v^* and m_c^* are respectively the effective mass of the valence and conduction states averaged over the three possible directions. The details of the calculation of m_v^* and m_c^* are given in Refs. [81, 96]. By coloring the data points according to $1/\mu$ in Fig. 4.8, we note that the larger μ (the smaller the dispersion of the bands), the smaller ω_g. To improve the visualization, the full set of data has been split according to the values of $1/\mu$. For each panel, the dashed line represents Eq. (4.16) with the the coefficients α and β reported above, and the the colored lines represent the same equation by fitting those coefficient considering each subset of data. The remaining spread in the data (other than the one coming from μ) is difficult to quantify by a simple physical quantity. Part of it can probably be attributed to the distribution of the bands in energy.

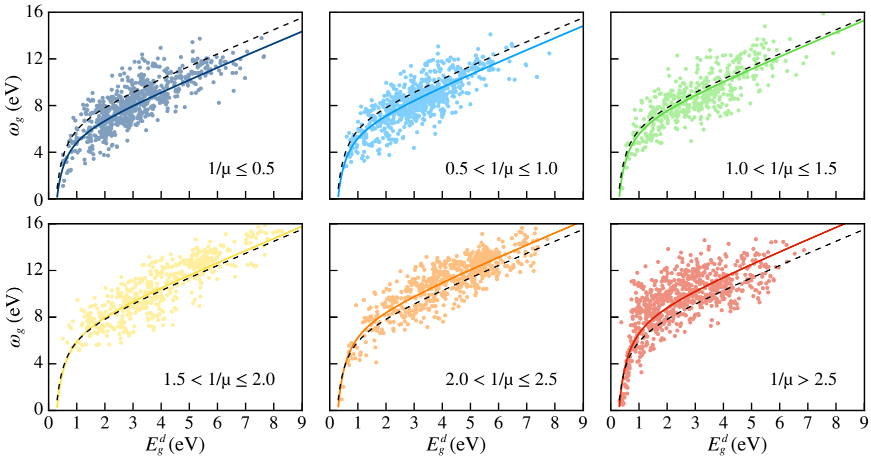


Figure 4.8: Calculated values of the average optical gap ω_g as a function of the direct band gap E_g^d (both in eV) for the 4040 materials considered in this study, split considering the different values of the effective mass $1/\mu$. In each panel, the dashed black line corresponds to $\omega_g = E_g^d + 6.74 - 1.19/E_g^d$ which was obtained by fitting all the data, while the colored line is obtained considering only the data in the subset represented in the panel.

In principle, α and β in Eq. (4.16) depend on the width of the JDOS. In practice, in the rest of the paper we assume α and β as constants, i.e. considering the fit calculated on the overall set of data (α=6.74 eV and β=-1.19 eV^2) to ease the discussion and the analysis.

Combining Eqs. (4.14) and (4.16), we obtain a direct relationship between the average static refractive index n_s and the direct band gap E_g^d:

$$n_s = \sqrt{1 + \frac{\omega_{\text{eff}}^3}{\left(E_g^d + \alpha + \frac{\beta}{E_g^d}\right)^3}} \tag{4.18}$$

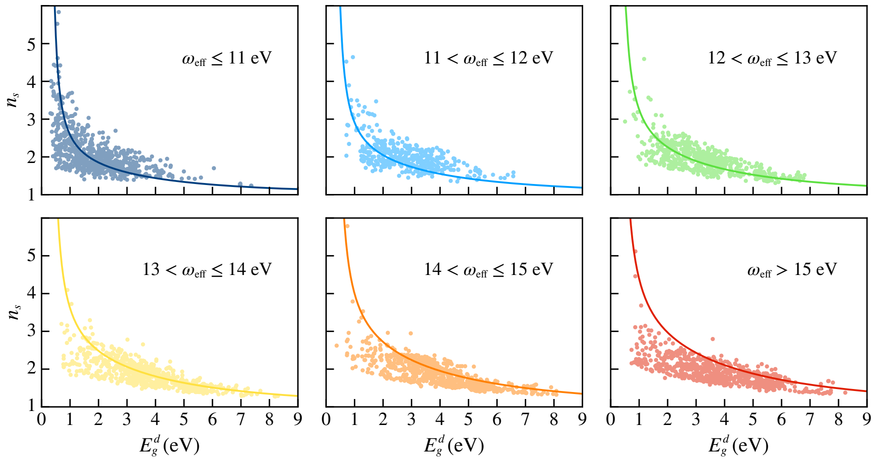


Figure 4.9: Calculated values of the static refractive index n_s as a function of the direct band gap E_g^d for the 4040 materials considered in this study, split considering the different values of the effective frequency ω_{eff}. The solid lines correspond to Eq. (4.18) using the the same values of ω_{eff} reported in Fig. 4.5.

which can be compared to all the calculated data, as shown in Fig. 4.9. Here, the full set of data has been split according to the values of ω_{eff} for a better visualization. This quantity has been computed by reverting Eq. (4.14). In each panel, the colored lines were obtained using the values of ω_{eff} indicated in Fig. 4.5. Globally, the data follow the trend of Eq. (4.18) as represented by these lines, confirming the inverse relationship between refractive index and band gap. The agreement is quite good given the approximations that are being made for the fit of ω_g as a function of E_g^d. In particular, the points with large (respectively, small) effective masses can fall significantly above (respectively, below) the corresponding curve (given that the latter is obtained for an average value of the effective mass).

From our analysis, it is clear that the effective frequency ω_{eff} (combining the integral of the JDOS J and the average transition probability K) and the effective mass μ (as well as the distribution in energy of the bands) play a key role in counterbalancing the effect of the band gap on the refractive index. The former is the numerator of the fraction appearing in Eq. (4.18), so the larger ω_{eff} the higher n_s. The latter acts on the denominator by limiting the difference between the direct band gap and the average optical gap: the larger μ, the smaller ω_g and hence the higher n_s.

At this stage, we would like to emphasize that the model that we propose in Eq. (4.18) is not predictive. Indeed, while ω_g can be determined directly from the electronic structure of the compounds, ω_{eff} (and more precisely K) cannot. We leave it for another study to analyze whether machine learning might help to overcome this limitation.

4.1.2 Outliers

As far as the combination of high refractive index and high band gap is concerned, the most interesting materials are those lying above the curve corresponding to the value ω_{eff}=12.10 eV, calculated by fitting in the last square sense Eq. (4.18) to the full set of our data. Such materials have either a large value of ω_{eff} (i.e. following the general trend of the curves) or of μ (i.e. due to the spread of the data). Among those, we found various compounds commonly used for optical devices, a few examples of which are reported in Table 4.1. In contrast, to the best of our knowledge, some of these outliers have not yet been considered as optical materials (for instance, Ti_3PbO_7, $LiSi_2N_3$, BeS, ...). In Appendix E we provide various tables with the 10 materials with the highest refractive index for a given direct band gap range.

Having a high value of the refractive index, for the compounds listed in Table 4.1 we also expect a high response in the nonlinear regime. This aspect, based on the Miller's empirical rule [10], will be further analyzed in the next section. In particular, $LiTaO_3$, $LiNbO_3$, LiB_3O_5, and BaB_2O_4 are known to have high nonlinear second order coefficients. They are thus commonly used for Second Harmonic Generation (SHG), to convert the incoming light from UV, or even deep UV, to the visible spectral range (see for example Refs. [32, 33]). In contrast, both TiO_2 phases (anatase and rutile) are centro-symmetric and they do not show any response at the second order. But, because of their refractive index, they have been recently investigated as optical switching devices and waveguides (see for example Refs. [97–99]).

4.1.3 Trend in oxides

As we have already stated, most of the compounds in the DB are oxides (3375 out of 4040). These are important compounds because they show important applications in many sectors with an exceptionally broad range of electronic properties (see for instance Ref. [78]) We focus on the chemical composition and the electronic structure of the materials making the connection with the optical properties of the 3375 oxides.

To properly describe the data distribution, we introduced the effective frequency ω_{eff} that is related to both J and K (see Fig. D.1). Although both these quantities are important to obtain the correct ω_{eff} for each material, only J can be deduced from the electronic

Formula	MP-id	n_s	E_g^d	ω_{eff}	ω_g	m_v^*	m_c^*	μ
$LiTaO_3$	mp-3666	2.25	3.71	14.44	9.05	3.48	1.44	1.02
$LiNbO_3$	mp-3731	2.33	3.41	13.00	7.91	3.53	1.60	1.10
LiB_3O_5	mp-3660	1.62	6.35	16.13	13.70	6.77	1.20	1.02
rutile-TiO_2	mp-2657	2.84	1.78	10.88	5.66	2.53	1.00	0.72
anatase-TiO_2	mp-390	2.60	2.35	10.73	5.96	1.95	1.85	0.95
BaB_2O_4	mp-540659	1.63	4.60	13.46	11.29	15.01	0.61	0.58

Table 4.1: List of known outliers (i.e., lying above the curve corresponding to ω_{eff}=12.10 eV). The chemical formula, MP identification (MP-id), average refractive index, direct band gap E_g^d (in eV), the effective frequency ω_{eff} (in eV), the average optical gap ω_g (in eV) and the average effective mass of the transitions μ are shown for each material.

structure of the compounds. This is the main reason why our model cannot be predictive. To the best of our knowledge, there is no way to predict the probability transition K just considering the band structure. Systematic correlations between K and materials properties are still under investigations. This an interesting aspect for which it would be worth performing also some machine learning analysis.

In order to analyze the trend in terms of their chemistry, the compounds are organized in four different classes: two groups of transition metal oxides (TMOs), lanthanide oxides, and main-group oxides. Materials with actinide elements are not taken in account in our analysis. These classes are created as follows: the groups of the TMOs include compounds in which there is at least one TM element with the d shell not completely filled and no lanthanide elements. The lanthanide oxides class contains compounds in which there is at least one lanthanide element, but no TM elements (747). Finally, the remaining oxides that do not contain any of the above mentioned elements are included in the main-group. The TMOs have been further split in two groups considering not only the TM element but also its oxidation state [100]. In the first group, the d shell of the TM element is empty (e.g. V^{5+}) and therefore the electronic transitions from the top of the valence to the bottom of the conduction states are expected to be from the O 2p to the TM d-orbitals (671). In the second group, the TM d shell is partially filled (e.g. V^{4+}, V^{3+}) and thus the transitions are expected to be from the TM filled d states to the empty ones (303). Finally, all the compounds that contain TM elements in which the d shell is completely filled (e.g. Zn^{2+}, Cu^{+}) are included in the main group (1520).

For each class, the probability density function is computed for the distribution of the refractive index as function of the band gap via a Kernel-Density Estimation (KDE) using a Gaussian kernel (see Ref. [101] for further details). In Fig. 4.10, we only represent each class by an ellipse that contains the main data distributions and is obtained as follows. Its center is located at the average value of the direct gap and refractive index for the corresponding distribution. The orientation and lengths of its axes are determined using principal component analysis for the materials which belong to the region with a density larger than 75%. The curve reported in the figure is obtained from Eq. (4.18) using $\omega_{\text{eff}} = 12.10\ eV$. As we already stated, materials falling above this curve are the most interesting ones.

The importance of the flatness of the bands at the edges of the valence and conduction bands has already been underlined in Ref. [23]. This means that the presence of d and f-orbitals can be helpful. Indeed, the most interesting compounds (i.e. those located mostly above this curve) come from the first group of TMOs and the lanthanide oxides, in which those orbitals are present close to the VBM and CBM. These are the most suitable for applications that require both a wide band gap and a high refractive index. This is especially true for applications for which the absorption edge is at the limit of the visible region (experimental $E_g^d{\sim}3$ eV). For applications in the UV (experimental $E_g^d{\sim}6$ eV), the compounds in the main group of elements reveal to be the most promising.

In the following subsections, we describe the peculiarities of the four different classes, focusing on a typical example for each of them. For the four representative materials, we focus on the electronic structure and on a brief description of the optical functions $j(\omega)$, and $j(\omega)/\omega^3$. We also highlight the different relevant quantities (E_g^d, ω_g, and ω_{eff}).

Before proceeding with the discussion of the different classes, it is worth stressing again that the electronic structures used in this study are taken directly from the Materials Project repository. They have been obtained in the framework of DFT using the PBE exchange-correlation functional, that is known to underestimate the band gap with respect to experiments.

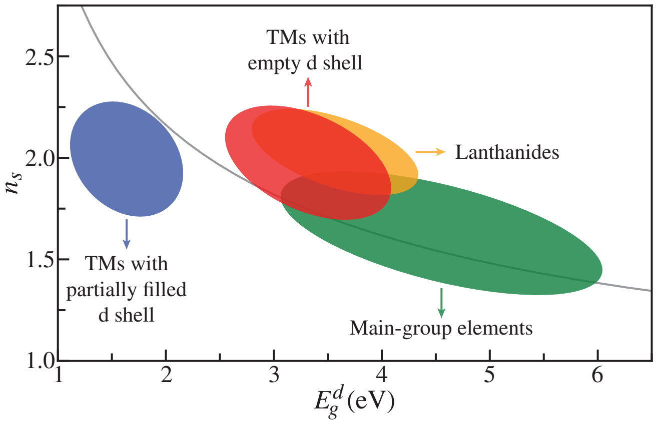


Figure 4.10: Static refractive index n_s as a function of the direct band gap E_g^d for the four classes of materials (first and second groups of TMOs in red, and blue respectively, main-group elements in green, lanthanides in orange) considered in this study. Each class is represented by an ellipse (see text) indicating the main distribution of the materials that belong to this class. The solid line correspond to Eq. (4.18) with ω_{eff}=12.10 eV.

4.1.3.1 TMOs with empty d shell (1^{st} group)

Many materials from this class are of high technological interest as dielectrics and as lenses for optical devices both in the linear and nonlinear regime [78].

As can be seen in Fig. 4.11, the materials from this class show a relatively high value for both the refractive index and the band gap.

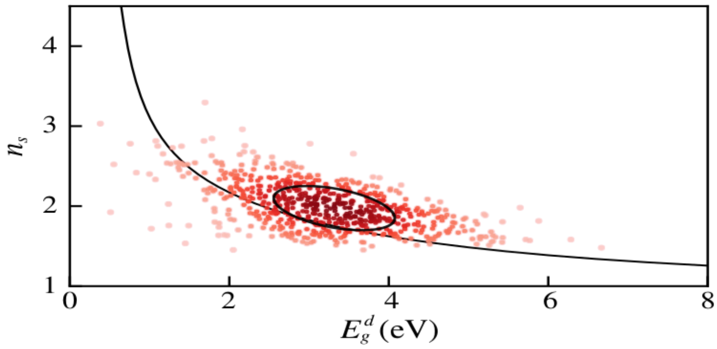


Figure 4.11: Static refractive index n_s as a function of the direct band gap E_g^d for the TMOs with empty d shell (1^{st} group) (671 materials). The solid line corresponds to Eq. (4.18) with ω_{eff}=12.10 eV. The probability density function is computed in the distribution for the refractive index as a function of the band gap via a Kernel-Density Estimation (KDE) using a Gaussian kernel. The darker region represents the higher value for the distribution density.

TiO_2 is a typical material of this group. For this compound, the Ti oxidation state is +4 (empty d shell). This binary oxide compound still generates great interest for the construction of optical devices (see for example Refs. [102, 103]). It is indeed one of the materials with the highest value of the refractive index, while retaining a high transparency throughout the visible region. The electronic structure and the optical functions for the rutile phase (mp-2657) are shown in Fig. 4.15(a). These are representative of those for other known TiO_2 phases (anatase, brookite, and monoclinic) and for other compounds in this class (e.g. ZrO_2, V_2O_5, $LiNbO_3$, ...).

In all these materials, the bands at the valence and conduction edges are quite flat. The main contribution to the top valence states originates from the O 2p-orbitals, while that to the bottom conduction bands comes from the d-orbitals of the transition metal (Ti 3d in the case of TiO_2). The flat nature of the bands at the edge of the band structure in this material can also be appreciated looking at the different values of the effective masses (m_v^*=2.53, m_c^*=1.00, μ=0.72). As a consequence the DOS is actually quite high at the band edges. This leads to an important $j(\omega)$ originating from the transitions from the O 2p-orbitals to the transition metal d orbitals, and hence to a large value of J. Further, these materials have a wide band gap that arises mainly from electron repulsion effects [104, 105]. This translates into a high refractive index (n_s=2.85).

In summary, for transition metals oxides from the first group, the wide band gap (which pushes the refractive index downwards) is compensated by a large number of available transitions from the top of the valence band to bottom of the conduction bands due to both the flatness of the band structure and to the high density of states at the band edges. These materials are thus very interesting candidates for further investigations.

4.1.3.2 TMOs with partially filled d shell (2^{nd} group)

In general, TMOs from the second group have a smaller gap than those from the first group. This obviously pushes their refractive index upwards. However, contrary to the first group of TMOs, the majority of the data fall below the curve (see Fig. 4.12).

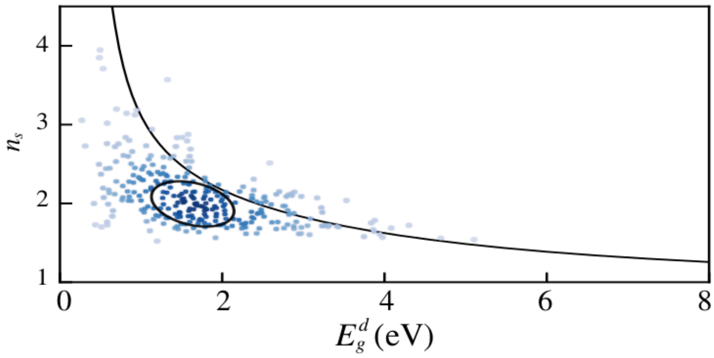


Figure 4.12: Static refractive index n_s as a function of the direct band gap E_g^d for the TMOs with partially filled d shell (2^{nd} group) (303 materials). The solid line corresponds to Eq. (4.18) with ω_{eff}=12.10 eV. The probability density function is computed in the distribution for the refractive index as a function of the band gap via a Kernel-Density Estimation (KDE) using a Gaussian kernel. The darker region represents the higher value for the distribution density.

These materials could be considered good candidates for optical applications that require a moderate transparency (e.g. in the visible range) and a high refractive index. We focus on Cr_2O_3 (mp-19399) as an illustrative example with features common to other compounds of this class (e.g. PtO_2, NiO...). In this case the Cr oxidation state is +3 (partially filled d shell). Chromium oxides are widely used in many sectors such as, for example, catalysis, solar energy applications, and others (further information can be found in Ref. [106]). Since this material shows a magnetic ordering, in Fig. 4.15(b) is reported its electronic structure for both spin components separately, and the optical functions resulting from the combination of both of them. The bands at the edge of the band structure (spin up component) show a flat nature. This is further emphasized looking at the values of the effective masses (m_v^*=4.29, m_c^*=2.53, μ=1.59). At the bottom of conduction states (spin up component), as in the previous case, the main contribution comes from the d-orbitals of the TM (Cr 3d in the case of Cr_2O_3). One of the main difference lies in the contribution of the d-orbitals in the valence states, leading to a more hybridized character. The presence of an important amount of d states both at the top of the valence and at the bottom of the conduction leads to a decrease of the band gap with respect to the TMOs with empty d shell [104, 105]. Due to the flatness of the bands, the DOS at the edge of the band structure is quite high giving an important $j(\omega)$. However, in this case, the JDOS gives a more broad spectrum with respect to the TiO_2 case, leading to larger values of both ω_g and ω_{eff}, and a slightly smaller value of the refractive index for Cr_2O_3 (n_s=2.51).

In summary, the main difference with respect to the first group of TMOs lies in the valence bands in which there is a strong contribution from the d-orbitals. So, despite their lower gap, the TMOs from the second group show a similar refractive index.

As a final remark, it is worth to mentioning that, for this class of materials, DFT is known to predict wrong band gap and dispersion due to the presence of partially filled d-orbitals. To this purpose, as mentioned in Sec. 3.4.1, a Hubbard-like Coulomb U term was added (GGA+U) [60, 80].

4.1.3.3 Main-group oxides

The main-group oxides show a higher diversity than the other classes. Indeed in this class we can find oxides that contain elements such as Zn, Cd and Hg in their oxidation state +2 such that they have a full d shell in valence, as well as Si, Ge, etc. Although the compounds in this class show a common behavior in terms of their electronic structure, due to the diversity of the materials their band gaps display a more important spread, as can be seen in Fig. 4.13.

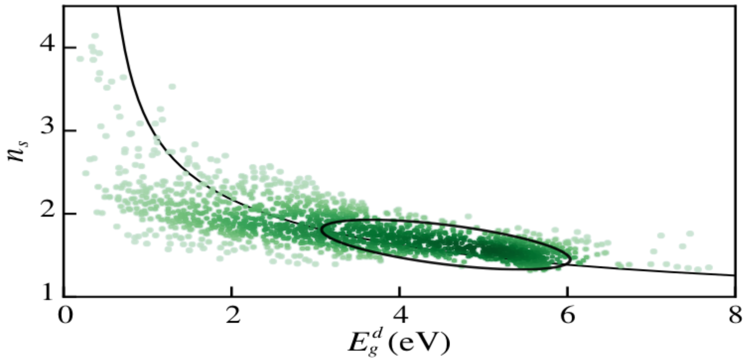


Figure 4.13: Static refractive index n_s as a function of the direct band gap E_g^d for the main-group oxides (1520 materials). The solid line corresponds to Eq. (4.18) with ω_{eff}=12.10 eV. The probability density function is computed in the distribution for the refractive index as a function of the band gap via a Kernel-Density Estimation (KDE) using a Gaussian kernel. The darker region represents the higher value for the distribution density.

Most of the materials belonging to this class are commonly used as insulators. A prototypical example of this class SiO_2. This material is used for many devices and one of the most known applications is found in the amorphous silica phase, used for optical fibers [107]. The electronic structure and the optical functions for the β-cristobalite $I\bar{4}2d$ tetragonal form (mp-546794) are shown in Fig 4.15(d). Here, as in the case of the first group of transition metals oxides, O 2p states lead to quite flat valence bands and increase the DOS at the valence edge. In contrast, the conduction bands are very dispersive, showing almost a free-electron like parabolic character which directly translates into the values of the effective masses (m_v^*=4.15, m_c^*=0.56, μ=0.49). This results in a small contribution to the DOS at the bottom of the conduction states. Consequently, the JDOS $j(\omega)$ [Fig. 4.15(c)] does not show any clear peak close to the absorption edge and the refractive index is quite low (n_s=1.48). Indeed the value of J for this material is smaller than the one of both the representative candidates previously described. Furthermore, compared to the other cases, the value of ω_g is much higher than that of E_g^d ($\omega_g > 2E_g^d$), and it is much closer to ω_{eff}.

4.1.3.4 Lanthanide oxides

Lanthanide oxides tend to have a wide band gap, while still showing a high refractive index (see Fig. 4.14). The oxides contained in this class have common features with the TMOs with empty d shell. Indeed, the two respective ellipses are almost superimposed on each other in Fig. 4.10. As an illustrative example, we have

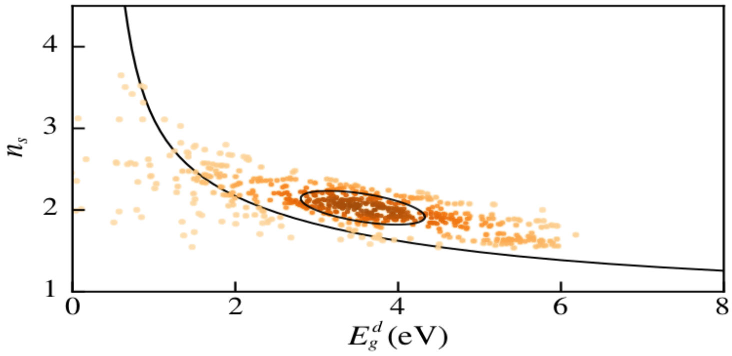


Figure 4.14: Static refractive index n_s as a function of the direct band gap E_g^d for the lanthanide oxides (747 materials). The solid line corresponds to Eq. (4.18) with ω_{eff}=12.10 eV. The probability density function is computed in the distribution for the refractive index as a function of the band gap via a Kernel-Density Estimation (KDE) using a Gaussian kernel. The darker region represents the higher value for the distribution density.

chosen Nd_2O_3 (mp-1045). These compounds are typically known because they show good luminescence properties and they can be used as fluorescent materials in lighting applications (more information can be found in Ref. [108]). Looking at the electronic structure, the flat nature of the bands can be appreciated at the top of the valence states coming mainly from the O 2p-orbitals. At the bottom of the conduction states, there is a slightly more dispersive behavior that comes mainly from the d-orbitals (Nd 4d in the case of Nd_2O_3). This is also evident looking at the values of the effective mass (m_v^*=6.81, m_c^*=0.53, μ=0.50). As a result, the DOS at the top of the valence is quite important (as in the case of the 1^{st} group of TMOs), but the slightly more dispersive nature of the bands at the bottom of the conduction leads to a less pronounced DOS. Anyway, $j(\omega)$ shows a pretty well-defined peak centered at around ω_g. The large availability of states for an electronic transition from valence to conduction turns into a large value of ω_{eff}, leading to reasonably high value of the refractive index (n_s=2.06). All these considerations show the similarity between this class of oxides and the 1^{st} group of TMOs, and explain why the refractive index remains high in despite the wide band gap.

Finally, we would like to emphasize that the results for the lanthanide oxides need to be taken with great caution. Indeed, it is not a simple task to accurately compute materials with f electrons from first principles relying on pseudopotentials. In many cases, these electrons are frozen in the core, which may lead to a lack of accuracy.

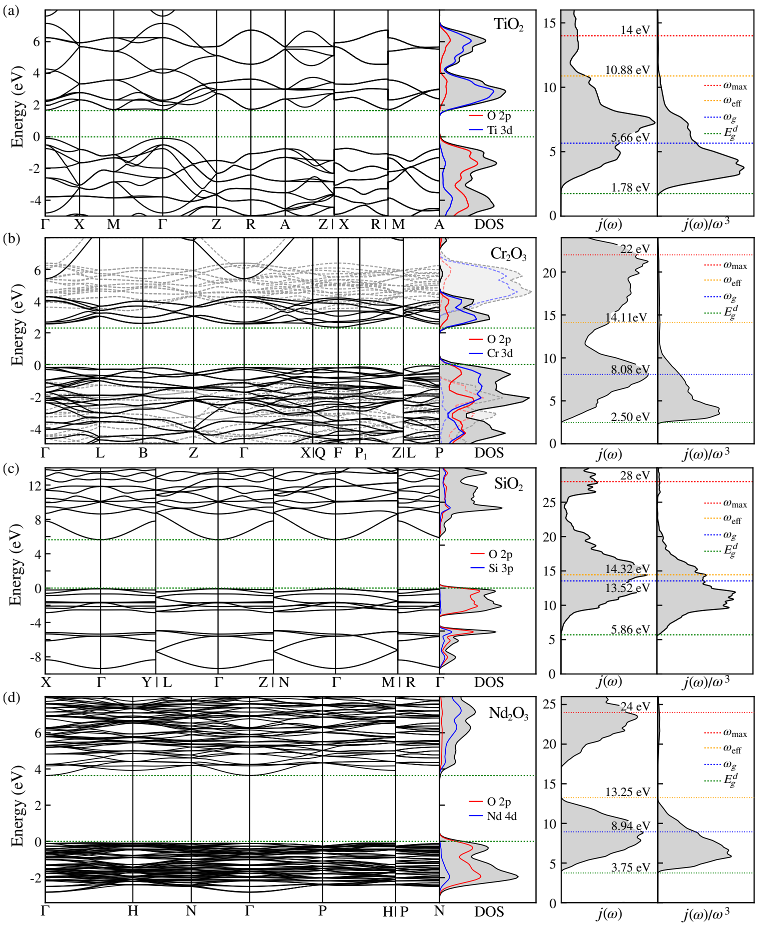


Figure 4.15: Electronic structure [band structure and density of states (DOS)] and optical functions [$j(\omega)$ and $j(\omega)/\omega^3$ in arbitrary units] for (a) TiO$_2$, (b) Cr$_2$O$_3$, (c) SiO$_2$, and (d) Nd$_2$O$_3$. The direct band gap E_g^d, average optical gap ω_g, effective frequency ω_{eff}, and upper limit of integration $\omega_{\max}$ are indicated by green, blue, orange, and red dotted lines. The four materials have been selected as representatives of the first and second groups of TMOs, the main-group oxides, and lanthanide oxides, respectively. For Cr$_2$O$_3$ (which shows a magnetic ordering), the electronic structure of both spin components are reported separately (the spin down component is indicated by the use of lighter colors and dashed lines) while the optical functions are the sum of both of them.

4.2 Nonlinear optical materials database

By selecting the non-centrosymmetric materials present in the linear optical materials DB (see Sec. 3.4.2), we built up a new set of 824 compounds aiming at calculating their second order susceptibility. At this stage, we have computed the nonlinear properties for 457 materials, which have been inserted in a new database. This number is intended to increase in the future. The analysis of this nonlinear optical materials DB is mainly focused on two aspects: (ii) a descriptive model that can characterize the trend of the materials distribution, in analogy with what we have done for the analysis of the linear DB; (ii) a method that can improve the screening procedure to accelerate materials discovery. In both cases, the analysis is carried out considering one material as an example to start our investigation. Our plan is to expand this analysis to the whole dataset. Two different distributions of the 457 materials are shown in Fig. 4.16. In the panel (a), the nonlinear effective coefficient d_{eff}, computed as described in Eq. (2.54), is plotted versus the direct band gap E_g^d, with the refractive index n_s indicated by the color of the data points. In the panel (b), the nonlinear effective coefficient d_{eff} is plotted as a function of the refractive index n_s, coloring the data points as a function of the direct band gap E_g^d. In the panel (a), the dashed line indicates the proportionality relation between the nonlinear and the linear optical coefficient, as described by the Miller's rule written in the form

$$d_{\mathrm{eff}} \propto \delta n_s^6 \tag{4.19}$$

where δ is the Miller's coefficient which is taken to be constant and equal to 0.05 pm/V. Assuming that this empirical rule is valid, the large spread in the data distribution around the dashed line shows that the Miller's coefficient cannot be materials independent, as already stated in Sec. 2.2.3. Further, it is also evident that a large linear optical coefficients is not a sufficient condition to have a large response also in the nonlinear regime. To this purpose, further analyses on the Miller's coefficient have to be carried out to identify a physical descriptor that can help us to understand the data distribution, hence leading to a further comprehension of the phenomena of the SHG. This aspect will be reported in Sec. 4.2.1, taking as example $\mathrm{LiNbO_3}$, a well-known material in the field of nonlinear optics.

In panel (b) of Fig. 4.16, the dependency of the nonlinear coefficient as a function of the direct band gap is shown. The dashed line is used as a guide to highlight the inverse relationship between these two quantities. Once again, assuming the validity of the Miller's rule, and given the descriptive model of Eq. (4.18), it is straightforward to assume that there is a proportionality given by the following relationship

$$d_{\text{eff}} \propto \frac{1}{(E_g^d)^9} \tag{4.20}$$

between the nonlinear coefficient d_{eff} and the band gap E_g^d. However, to the best of our knowledge, this relationship has been addressed in the literature only by few authors (see, for instance, Ref. [109]). As it is evident, also in this case there is a large spread in the data distribution. However, in the framework of this thesis, Fig. 4.16(b) will be only used for screening purpose, thus we will focus on the search for good nonlinear optical materials.

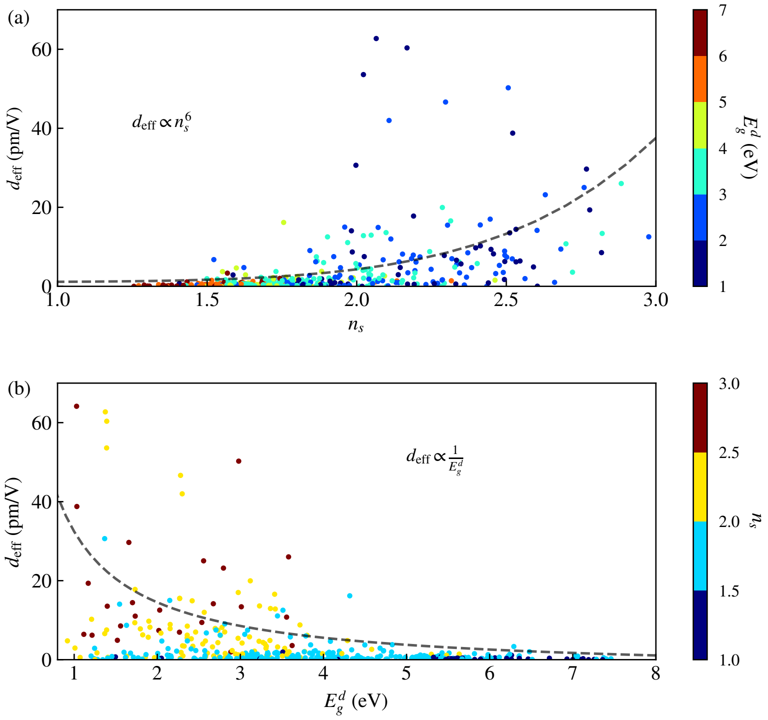


Figure 4.16: (a) Calculated values of the nonlinear effective coefficient d_{eff} as a function of the refractive index n_s for the 457 materials considered in this study. The color of the data points represents the direct band gap E_g^d. The dashed line indicates the proportionality between nonlinear and linear optical properties as predicted by Miller [10]. (b) Calculated values of the nonlinear effective coefficient d_{eff} as a function of the direct band gap E_g^d for the 457 materials considered in this study. The color of the data points indicates the refractive index n_s. The dashed line represents the inverse proportionality relation between nonlinear effective d_{eff} coefficient and the direct band gap E_g^d.

Regarding the spread, it is however worth mentioning that, given the form of the Miller's rule as written in Eq. (2.40), one can expect that a quantity such as the band gap can be inserted in the analysis through the linear susceptibility. Based on the model that describes the relationship between the nonlinear and the linear optical properties, a model that relates the nonlinear properties to the band gap can be deduced as a natural consequence.

4.2.1 Global trend

The comprehension of Miller's rule, and hence of the contribution of the Miller's coefficient δ, has attracted particular interest since its deduction. As we have already shown, this plays an important role for a proper description of this relationship. This would make it possible to determine the nonlinear optical properties of a material just from the linear ones, that are more easily accessible both from a first-principles and an experimental point of view.

To this purpose, here we mention the work of Levine [110] and Bell [111]. Although they had used slightly different models, they arrived at the same conclusion for which the nonlinear susceptibility can be related to the anharmonic motion of the bond-charge of the system. Further details about bond-charge models can be found, for instance, in Ref. [112]. However, to the best of our knowledge, these models have been applied only to the relatively simple case of binary compounds.

The Miller's coefficient can be naively seen as the coefficient of proportionality between the linear and the nonlinear optical properties. By rewriting Eq. (4.19), and considering the relationship between the refractive index and the dielectric function in the static limit (see Eq. (4.4)) we can compute the Miller's coefficient for each material as

$$\delta = \frac{d_{\text{eff}}}{\varepsilon_s^3} \tag{4.21}$$

For the different materials in our data set, the distribution of the nonlinear coefficient d_{eff} as a function of the Miller's coefficient δ is shown in Fig. 4.17. Based on these data points, δ takes values in the range $[0, 2]$ pm/V. This indicates, once again, that this quantity is material dependent. It is interesting to look at the compounds with $\delta > 1$ pm/V. Some of them are reported in Table 4.2. By looking at the structure of these compounds, some similarity can be found, as for instance, the fact that these compounds are layered materials (see Fig. 4.18). This is a further confirmation that a particular atomic distribution can favor the nonlinear process. This finding will be further analyzed in future works.

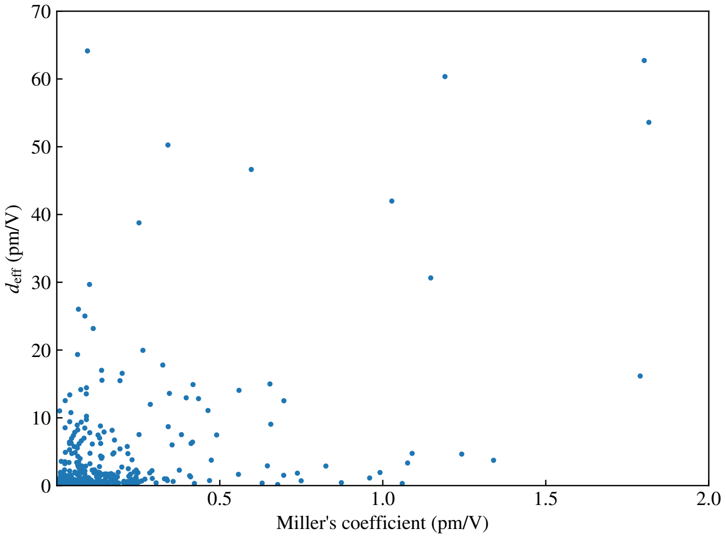


Figure 4.17: Calculated values of the nonlinear effective coefficient d_{eff} as a function of the Miller's coefficient computed as reported in Eq. (4.21).

There is another crucial aspect that has to be taken into account, namely the symmetry of the system. Indeed, the first condition needed to have a nonlinear optical response is that the system has to be non-centrosymmetric. In other words, the Miller's coefficient is null in the case of centrosymmetric materials.

Formula	MP-id	n_s	d_{eff}	δ
K_2S_3	mp-7667	2.07	62.72	2.07
PtF_4	mp-8943	2.17	60.35	1.19
Rb_2S_3	mp-7446	2.02	53.60	1.03
SBr	mp-28099	2.11	41.98	1.82
ClO_3	mp-22869	2.01	30.64	1.15

Table 4.2: Materials with a Miller's coefficient δ larger than 1 pm/V. The chemical formula, MP identification (MP-id), average refractive index n_s, nonlinear effective coefficient d_{eff} (in pm/V), and the Miller's coefficient δ (in pm/V) are shown for each material.

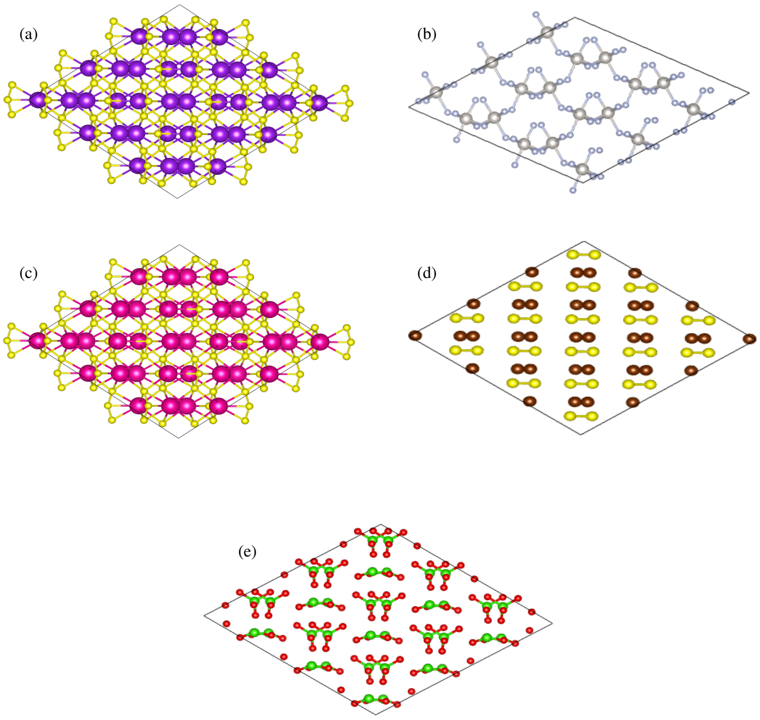


Figure 4.18: Structures of the compounds in Table 4.2. (a) K_2S_3; (b) PtF_4; (c) Rb_2S_3; (d) SBr; (e) ClO_3

Having all these considerations in mind, we start our analysis considering, as an example, the case of Lithium Niobate (LNB) in its

	d_{31}	d_{32}
This work	29.45	7.76
Experiment	25.2	4.6
Previous calculation	30.21	8.01

Table 4.3: Independent elements of the SHG d_{ij} tensor, compared with experimental results [12], and a previous theoretical work [13]. The data are presented in absolute value. The experimental values are measured at a wavelength of 1064 μm.

ferro and para-electric phases. In both phases, LNB has a trigonal crystal system, and it is thus a uniaxial optical crystal. The ferroelectric phase is widely known in the field of nonlinear optics because of its broad range of applications and its high response in the linear and nonlinear regime ($n_s = 2.35$, $d_{\text{eff}} = 16.50$ pm/V), while having a quite large band gap ($E_g^d = 3.41$ eV using PBE [53] xc functional, $E_g^d = 4.77$ eV using HSE [59] xc functional). Indeed, it has been used, for instance, as a modulator for optical fibers [113] and holographic applications [114]. At a temperature of 1480 K, a phase change occurs in this material: below this temperature LNB is ferroelectric, while it is paraelectric above this critical temperature [115]. Being non-centrosymmetric, the ferroelectric phase is the only one that is SHG active. Since it is widely used, many experimental data are available for comparison with our calculated results. This is summarized in Table 4.3, showing a pretty good agreement.

The paraelectric phase is instead centrosymmetric, and does not show any nonlinear optical effect at the second order. Yet, it is interesting to compare these two phases, since they have similar structural parameters. In both cases, we have 10 atoms (2 niobium atoms, 2 lithium atoms, 6 oxygen atoms) in the unit cell and the values of the lattice parameters are also pretty similar. The only difference lies, of course, in the space group, that is R3c for the ferroelectric phase, and R-3c for the paraelectric phase.

We then started our analysis considering the two structures stored in the MP repository. As already explained in Sec. 3.4.2, the calculation of the d_{ij} is possible only considering LDA xc functionals [86, 87], at the current level of the implementation in ABINIT [69]. We then execute a structural relaxation for these two phases. Some information as the lattice parameters, the lattice angles, and band gap, computed considering such a functional can be found in Table 4.4

The atomic structure it is shown in Fig. 4.19 for both phases. The niobium atoms (green) are at the center of an octahedron formed by six oxygen atoms (red). In the paraelectric phase (panel (a)), where indeed we have the inversion symmetry, the Nb atoms are exactly at the center of these octahedra, while in the ferroelectric phase (panel (b)) the octahedra are distorted, and the Nb atoms are not located at their centers, breaking then the inversion symmetry. Looking at the lithium atoms (blue), they are at the center of a planar triangle formed by 3 O atoms for the paraelectric phase. After the phase change, the Li atoms are immersed in octahedra formed by six O atoms. As in the Nb, the octahedra are distorted and the Li atoms are not located at their centers.

The atomic positions of the two phases are also reported in Table 4.5, in which the 3^{rd} column represents the distance between the position of each atom from its centrosymmetric configuration to the non-centrosymmetric one (and vice-versa). There is a small displacement of the Nb atoms, while the largest displacement is given by the two Li atoms and there is a moderate displacement for the O atoms. In fact, as already discussed, there is a change of the

	Paraelectric phase	Ferroelectric phase
Material-id	mp-3731	mp-552588
Crystal system	Trigonal	Trigonal
Space group	R-3c	R3c
Point group	-3m	3m
a($\AA$)	5.1221	5.0649
c($\AA$)	5.3989	5.4173
α(degree)	61.6819	62.1293
γ(degree)	60.00	60.00
Band gap (eV)	2.67	3.50

Table 4.4: Structural parameters of $LiNbO_3$ considering its ferroelectic and paraelectric phase.

atomic environment in which the Li atoms are immersed going from the centrosymmetric to the non-centrosymmetric configuration.

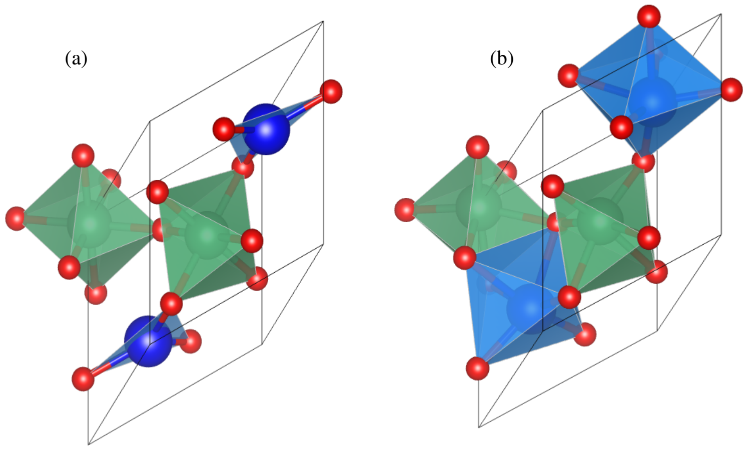


Figure 4.19: Atomic structure of $LiNbO_3$. (a) Paraelectric (centrosymmetric) phase; (b) Ferroelectrics (non-centrosymmetric) phase. In both panels, the Nb atoms are colored in green, the O atoms in red, and the Li atoms in blue.

By interpolating between these two structures, we generated 4 intermediate structures depicted in Fig. 4.20. It is interesting to look at the 1st step (moving the atoms from the centrosymmetric to the non-centrosymmetric configuration) in panel (a). This represents

	Ferroelectric phase	Paraelectric phase	d
Li_1	0.2500, 0.2500, 0.2500	0.2177, 0.2177, 0.3470	0.5868
Li_2	0.7500, 0.7500, 0.7500	0.7177, 0.7177, 0.8470	0.6133
Nb_1	0.5000, 0.5000, 0.5000	0.5013, 0.5013, 0.4962	0.0332
Nb_2	1.0000, 1.0000, -0.0001	1.0013, 1.0013, -0.0038	0.0743
O_1	0.1253, 0.3747, 0.7500	0.1328, 0.3931, 0.6925	0.3078
O_2	0.2500, 0.8747, 0.2500	0.2815, 0.8931, 0.1925	0.3423
O_3	0.3747, 0.7500, 0.7500	0.3931, 0.7815, 0.6925	0.3277
O_4	0.6253, 0.2500, 0.2500	0.6328, 0.2815, 0.1925	0.3395
O_5	0.7500, 0.1253, 0.7500	0.7815, 0.1328, 0.6925	0.3212
O_6	0.8747, 0.6253, 0.2500	0.8931, 0.6328, 0.1925	0.3099

Table 4.5: Atomic position of $LiNbO_3$ in fractional coordinates considering its paraelectric and ferroelectic phase. The third column represents the distance (in $\AA$) between each atom in the two configurations.

the first structure for which the inversion symmetry is broken, and the system is then SHG active. In this structure, the inversion symmetry is broken from the displacement of the NB and Li atoms, that are not anymore at the center of the environment made of the O atoms. However, the Li atoms are still surrounded by 3 O atoms. Looking at the second step (panel (b)) the Li atoms make additional bonds with the O atoms, and they are immersed in an octahedron formed by the latter. As it is evident, at this step the Li atoms are quite far from the center of the octahedra and approaches their center only in the further steps (panels (c) and (d)), reaching the equilibrium position when the system is in its ferroelectric phase. However, we have already discussed that even in this configuration the Li atoms are not located at the center of the octahedra.

We have then in total 6 structures for which we are interested in the variation of the optical properties in the linear and nonlinear regime. These results are shown in Fig. 4.21. Of course, the centrosymmetric structure corresponds to the paraelectric phase and

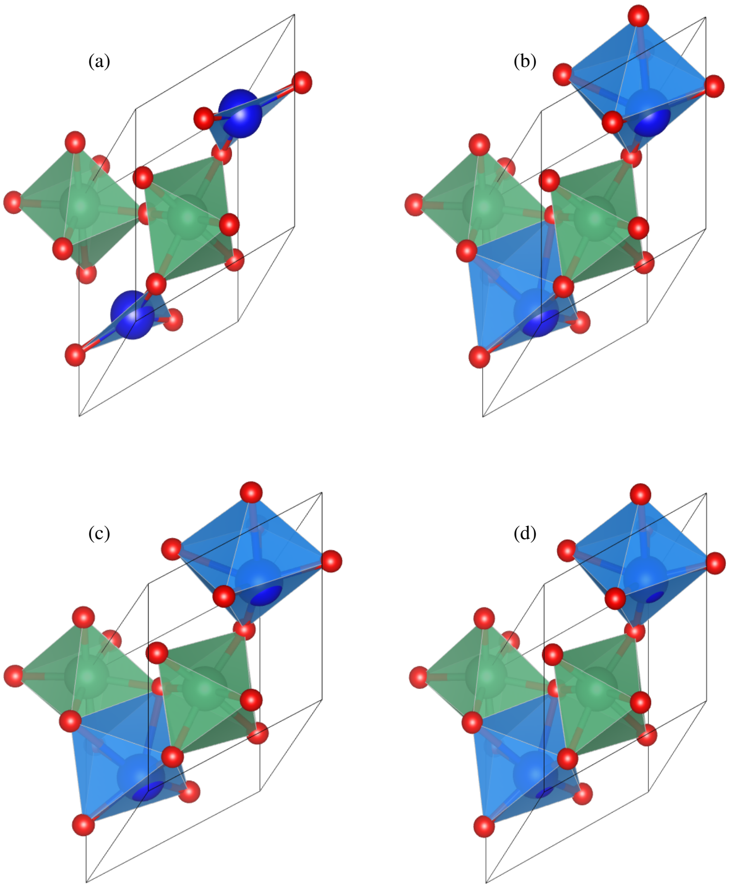


Figure 4.20: Intermediate LNB structures obtained by interpolating the atomic positions from its paraelectric configuration to the ferroelectric one. In each panel, the Nb atoms are colored in green, the O atoms in red, and the Li atoms in blue. Panel (a) represents the first atomic displacement from the centrosymmetric configuration for which the inversion symmetry is broken. Here Nb and Li atoms are not anymore at the center of the environment formed by the O atoms. In panel (b), the atomic displacement produces a change in the environment in which the Li atoms are immersed, and they make additional bonds with the O atoms. They are immersed now in octahedra formed by 6 O atoms, in which the Li atoms are not at the center. In panel (c) and (d), the atomic environment does not change anymore, but yet there is an atomic displacement that leads to further deformation of the octahedra.

the non-centrosymmetric one to the ferroelectric phase. In panel

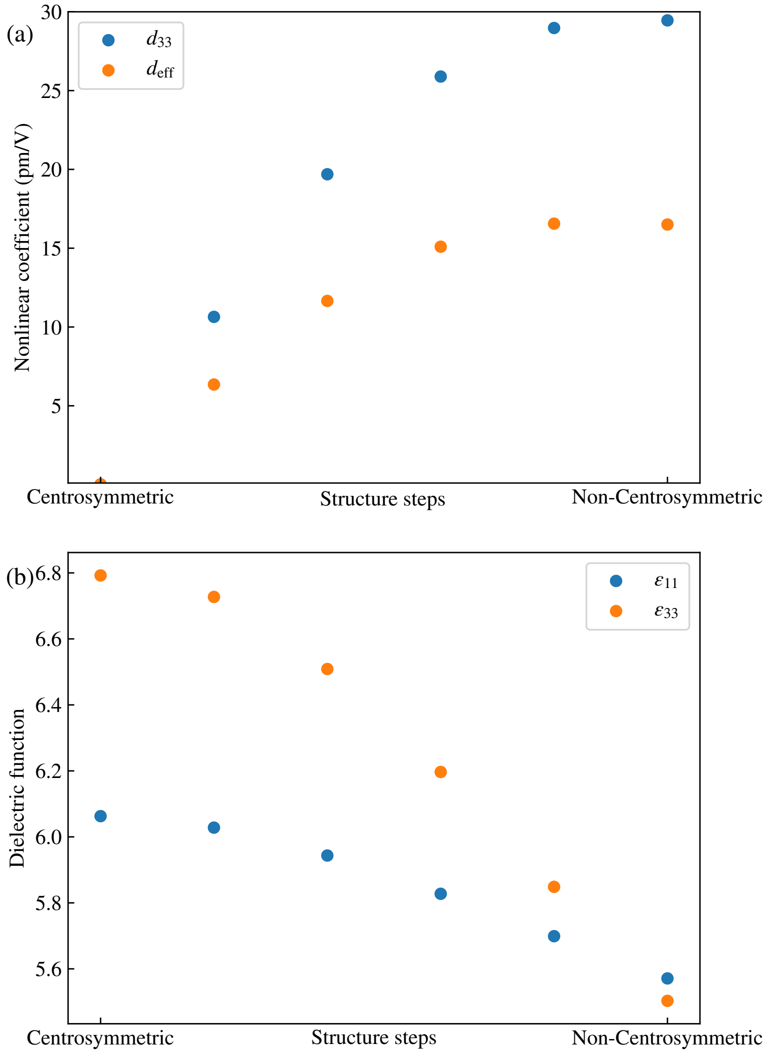


Figure 4.21: (a)Nonlinear optical properties for the different structural steps in $LiNbO_3$, going from a paraelectric (centrosymmetric) to a ferroelectric (non-centrosymmetric) atomic configuration. The blue dots represent the d_{33} element of the SHG tensor while the oranges correspond to the nonlinear effective coefficient d_{eff}. (b)Linear optical properties for the different structural steps in $LiNbO_3$, going from a paraelectric (centrosymmetric) to a ferroelectric (non-centrosymmetric) atomic configuration. The blue dots represent the ε_{11} element of the dielectric tensor while the oranges correspond to the ε_{33} element.

(a), the variation of the nonlinear optical coefficient d_{eff} (orange points) and the d_{33} (blue points) tensor element is shown as a function of the structural steps. As expected, going from the centrosymmetric to the non-centrosymmetric configuration, the nonlinear coefficients increase. It is even more interesting to take look at the behavior of the linear optical coefficients as a function of the structural steps. In Fig 4.21 the variation of the tensor elements of the dielectric function ε_{11} (in blue) and ε_{33} are shown. In both cases, the linear optical properties decrease going from the centrosymmetric to the non-centrosymmetric configuration. This is, in principle, in contrast with the Miller's rule. These observations, are summarized in Fig. 4.22, where the non linear coefficient as a function of the linear one is plotted for each structural step. However, given

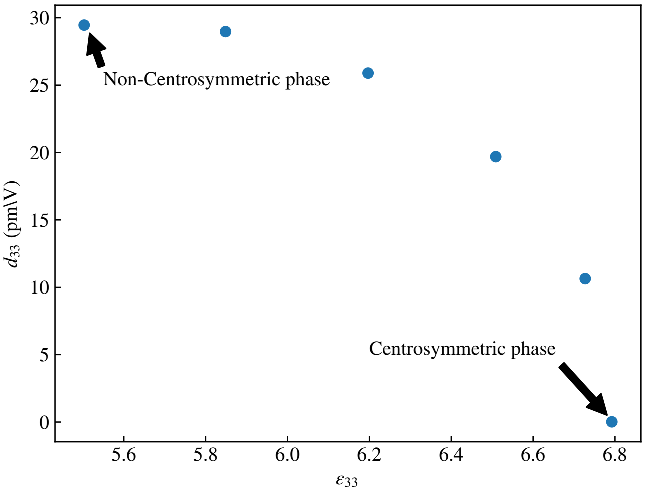


Figure 4.22: d_{33} component of the SHG tensor as a function of the ε_{33} component for the different structural steps in LNB.

the inverse dependency of the refractive index w.r.t. the band gap, a decrease of the linear optical properties can be expected (See Fig. 4.23).

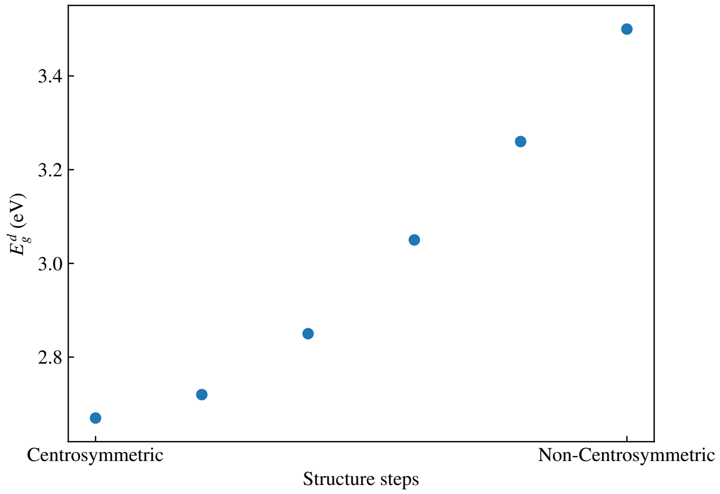


Figure 4.23: Direct band gap (E_g^d in eV) variation as a function of the different structural steps in LNB.

Indeed, as shown in Fig. 4.24, going from the paraelectric phase (in red) to the ferroelectric one (in blue), there is an increase of the band gap. This is due to a change in the electronic band structure, in which the bands at the edge of the conduction states become flatter, having then an opening of the band gap. The electronic structural properties for all the 6 structures derived are given in Appendx G. Anyway, as already mentioned, this is an aspect that explains the behavior of the linear optical coefficients, but not the one of the nonlinear optical coefficients.

By investigating the variation of the Miller's coefficient, we can see that there is an increase of this quantity going from the centrosymmetric to the non-centrosymmetric configuration (Fig. 4.25). In the case of the d_{33} coefficient (blue points), this increase seems to be linear w.r.t. the structural steps, while in the case of d_{eff} (orange points) there is still an increase, but it is now less pronounced.

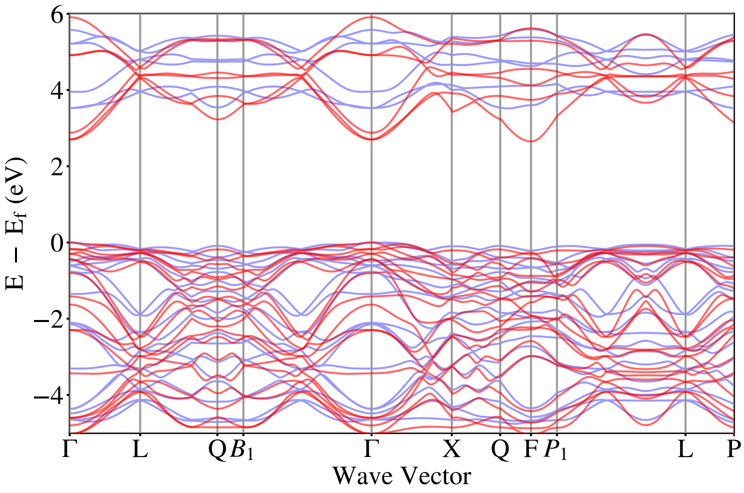


Figure 4.24: Electronic band structure of LNB. In red it is indicated the paraelectric phase, while the blue lines indicate the ferroelectric configuration.

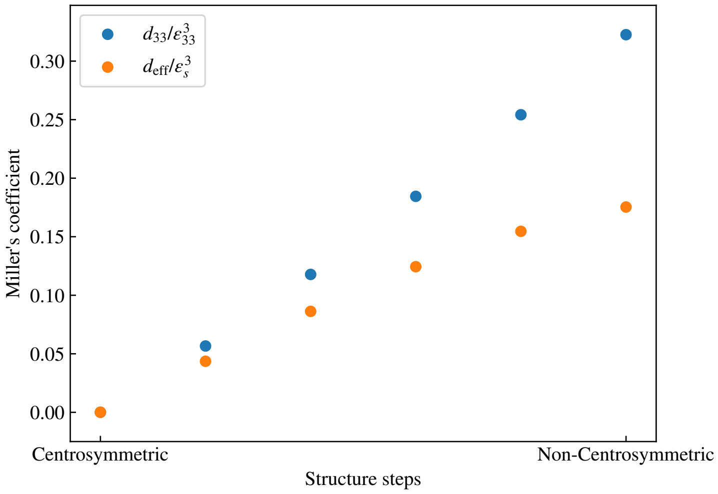


Figure 4.25: Variation of the Miller's coefficient, computed as reported in Eq. (4.21), as a function of the structural steps for the d_{33} tensor element (blue dots), and for the effective nonlinear coefficient d_{eff} (orange dots).

In summary, starting from a centrosymmetric configuration and by moving the atoms from their position, in such a way that we can reach the non-centrosymmetric configuration, there is a significant change in the value of the Miller's coefficient δ, that generates an increase of the nonlinear optical coefficients. Looking at the different structural steps, we have highlighted the change of the atomic environment in which the Li atoms are immersed.

In our opinion, this is another clear evidence of the importance of the Miller's coefficient, that cannot be considered material independent. Further, all these considerations suggest that a physical descriptor that can explain the data distribution and thus the variation of the Miller's coefficient has to be searched in the structural features of the compounds. In our future plans, we also have in mind to validate the relationship proposed by Scandolo and Bassani (see Eq. (2.46)) [39] to compute the Miller's coefficient in the form

$$\chi^{(2)}_{ijk}(0,0) = -\frac{\left\langle \frac{\partial^3 V}{\partial x_i x_j x_k} \right\rangle_0}{2N^2} \chi^{(1)}_{ii}(0)\chi^{(1)}_{jj}(0)\chi^{(1)}_{kk}(0) \tag{4.22}$$

as already showed in Chap. 2. We remind that N is the density of electrons and $V(\mathbf{x})$ is the external potential experienced by the electrons, where the average, indicated by the angle bracket, is performed in the ground state of the system. From our conclusions, we expect that by computing the Miller's coefficient by the means of the derivative of the external potential, we will have the same trend as in Fig. 4.25. If this is, this approach can be easily extended to all the materials contained in the DB, since this quantity is in principle obtainable from the ground-state of the system.

4.2.2 Outliers

One of the main purposes of building DBs containing materials properties is that by screening those DBs, we can find new optimal candidates for a particular application. This aspect, at least from an introductory point of view, will be the main focus of this last section.

So far, by screening the list of non-centrosymmetric materials, we have computed the linear and nonlinear optical coefficients for more than 400 materials. At this point, it is interesting to look at the outliers, meaning materials with large nonlinear coefficient d_{eff} and large band gap E_g^d, as we have done in Sec. 4.1.2. As a matter of fact, we have retrieved the known materials already discussed in that section, such as $LiTaO_3$, and $LiNbO_3$. An update of Table 4.1, in which the nonlinear effective coefficient d_{eff} for these materials and other materials mentioned in this thesis such as KH_2PO_4 and $TiPbO_3$ is shown in Table 4.6. As already stated in the previous section, $LiNbO_3$ is confirmed as a material with large linear and nonlinear coefficients. In contrast, to the best of our knowledge, some of these outliers have not yet been considered as NLO (as for instance PON, B_2O_3, PbO, Sb_2WO_6, ..).

However, by looking at the requirements that a material should satisfy to be a good nonlinear optical material (see Sec. 2.1), we do not have enough information to investigate the quality of a NLO material at this stage.

Indeed, as a first criterion, depending on the particular application for which our material is made, the appropriate absorption edge should be achieved. To this purpose, it is worth reminding, once again, that DFT tends to underestimate the band gap up to 50%.

Formula	MP-id	n_s	E_g^d	d_{eff}
$PbTiO_3$	mp-20459	2.63	2.79	23.19
$KNbO_3$	mp-7375	2.28	3.12	19.95
$LiNbO_3$	mp-3731	2.33	3.41	16.56
$LiTaO_3$	mp-3666	2.25	3.71	8.79
KH_2PO_4	mp-6268	1.87	3.05	6.41
BaB_2O_4	mp-5730	1.69	4.85	2.07
KH_2PO_4	mp-696752	1.52	5.64	0.29

Table 4.6: List of known NLO outliers . The chemical formula, MP identification (MP-id), average refractive index, direct band gap E_g^d (in eV), and the nonlinear effective coefficient d_{eff} (in pm/V) are shown for each material.

This means that the band gap should be corrected to get a more reliable set of data. This has been performed for a set of 369 materials, by means of the HSE hybrid functional [59] (see Sec. 3.1.3.1). The result of this correction is shown in Fig. 4.26.

To improve the comparison, both GGA (red points) and HSE (blue points) data are shown. It is evident, as expected, that there is blue-shift, that opens the materials band gap. Being materials dependent, this shift is not constant.

Furthermore, in our DFPT calculation, the impact of the phase-matching angle is ignored. For a given material, if the requirement of phase-matching is not satisfied, the effective value of the nonlinear optical coefficients can be drastically reduced, affecting the quality of the NLO crystal.

A further improvement would be given thus by exploring if there is a set of angles for which the materials in our DB satisfy the phase-matching condition (see Sec. 2.3). For the purpose of including phase-matching parameters, we selected a material from our DB, and we show the approach we are using to perform on this analysis.

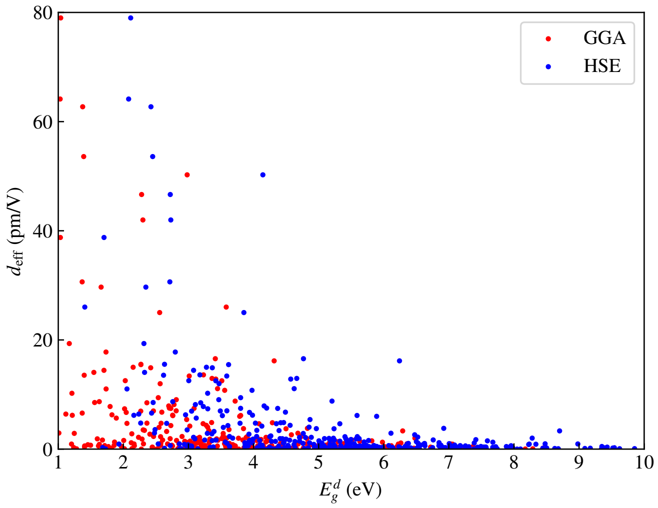


Figure 4.26: Calculated values of the nonlinear effective coefficient d_{eff} as a function of the direct band gap E_g^d, computed in the GGA approximation (red points), and HSE approximation (blue points).

As a material candidate we consider here Phosphate OxiNitride (PON, mp-753671). This compound has an orthorhombic unit cell. It is a biaxial crystal which belongs to the space group $I2_12_12_1$ (point group 222). For this material, we have computed a value $n_s = 1.79$ of the refractive index with a moderate birefringence ($\Delta_s = 0.10$), and $d_{\text{eff}} = 1.68$ pm/V for the nonlinear optical coefficient. For the band gap, we have $E_g^d = 5.51$ eV using the PBE xc functional[53], that becomes $E_g^d = 6.76$ eV when HSE [59] is used instead. This means that, if this material satisfied all the possible criteria, it would be a good candidate for the application that need deep UV NLO materials.

As reported in Ref. [32], these materials are of particular interest from both academic and technological standpoints, finding applications in the semiconductor manufacturing, laser systems, and attosecond pulse generation, for instance. A good NLO material should possess the following requirements:

- Wide UV transparency range ($E_g > 6.2$ eV);
- Large SHG coefficients (compared to the $d_{36} = 0.39$ pm/V of KH_2PO_4).
- Moderate birefringence (Δn_s $0.07 - 0.10$) to achieve phase-matching in the deep UV

In addition to these requirements, chemical stability and resistance to laser damage are required. Anyway, these further requirements, combined with the fact that large high optical quality crystals have to be grown, are requirements that have to be tested in the experiments as a last confirmation about the quality of the NLO material. PON seems then to satisfy the first requirements that an NLO material should possess.

PON has recently attracted some new interest also from an experimental point of view, because of its similarity with silica (SiO_2), to which it is isosteric. This similarity, was already highlighted by Léger at the end of the 90's [116]. Silica is one of the most studied materials because of its unique properties and the vast range of applications, that cover nonlinear optical phenomena as well, and because of the wide variety of crystal structures known for this compound. Baumann *et al.* [117, 118], recently found new possible crystalline structures for PON, and, as reported in their work, the number of possible structure types for PON could even exceed that for SiO_2, because of the ability of nitride ions to connect with more than two neighboring $P(O,N)_4$ tetrahedra.

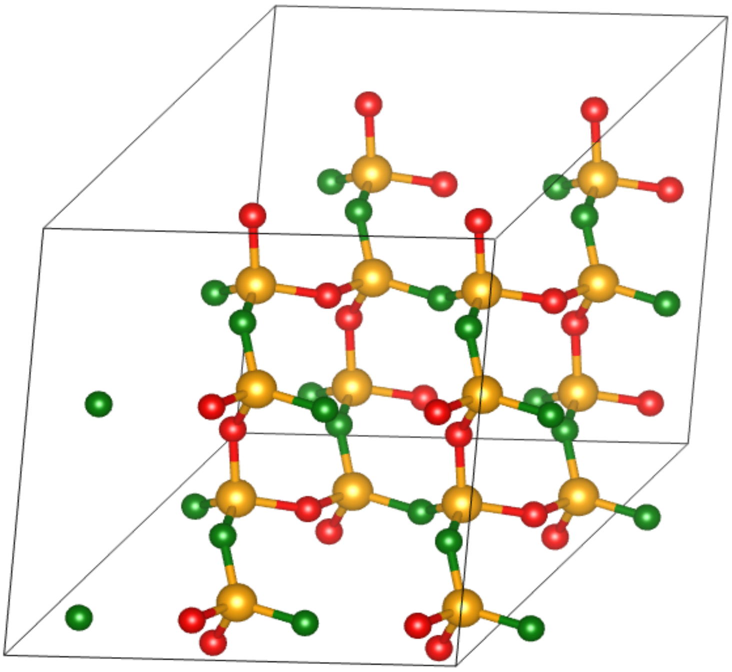


Figure 4.27: Atomic structure of PON, considering a (3×3×3) supercell. The P atoms are colored in yellow, the O atoms are in red, and the N atoms are colored in green.

Looking indeed at the structure of PON, shown in Fig. 4.27, we have six atoms in the unit cell: 2 phosphorous atoms (yellow), 2 nitrogen atoms (green), and 2 oxygen atoms (red). As in the structure of silica, where the silicon atom is surrounded by 4 Oxygen atoms forming a tetrahedron, in PON, the P atoms are immersed in a tetrahedron formed by the 2 O atoms and the 2 N atoms. Furthermore, it is interesting to take a look at Table 4.7, where the refractive index and the nonlinear effective coefficient computed in our framework for PON and two well-known silica phases are reported. From our data, PON has indeed a higher linear and nonlinear response than silica.

Before moving to the calculation of the phase match angle, there

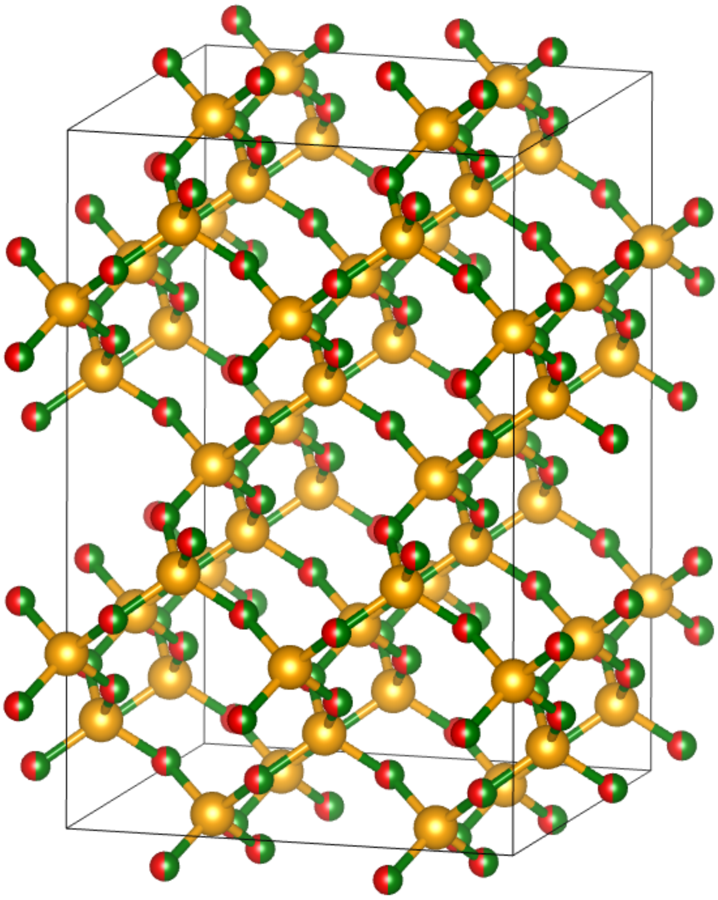


Figure 4.28: Atomic structure of PON, considering a (2×2×2) supercell. The P atoms are colored in yellow, the O atoms are in red, and the N atoms are colored in green.

is another point that is worth clarifing. Indeed, as also reported in Baumann's works, this material shows a disorder in its structure at the level of the arrangement of the oxygen and nitrogen atoms. This is somehow a limitation, because while from our calculated data this material falls into a non-centrosymmetric space group, the disorder can bring it to another space group that is centrosymmetric. The partial occupation of the sites (1/2 O, 1/2 N) is reported in Fig. 4.28. Starting from this cristobalite phase (space

Formula	MP-id	n_s	d_{eff}
PON	mp-753671	1.79	1.79
SiO_2 (α-quartz)	mp-6930	1.55	0.31
SiO_2 (cristobalite)	mp-7000	1.55	0.29

Table 4.7: Effective nonlinear coefficient d_{eff} (in pm/V) and average refractive index n_s for phosphorous oxynitride, α-quartz silica, and cristobalite silica. The data are taken from the calculation stored in our database.

group I$\bar{4}$2d), present in the Inorganic Crystal Structure Database (ICSD) [119], we derived 9 possible structures that include all the possible O-N ordering. These are shown in Fig. 4.29.

In each structure, the P atoms are immersed in the tetrahedron formed by the O and N atoms. Further information on these structures is given in Table 4.8. In column 4, the DFT total energies are shown. The difference in the total energy between all the derived structures is small, confirming the disorder in the ordering of N and O atoms. In columns 4 and 5, the linear and nonlinear optical properties are shown, respectively. We find then that all the 9 structures are non-centrosymmetric, showing a response in the nonlinear regime. Further, the average value of the nonlinear effective coefficient for the 9 structures is $d_{\text{eff}} = 1.65$ pm/V, that is pretty close to the data in our DB ($d_{\text{eff}} = 1.68$ eV).

Finally, we move to the analysis of the phase match (PM) parameters, for PON. As already discussed, the phase-matching condition $n(2\omega) = n(\omega)$ is impossible to achieve. However, playing with the birefringence properties of the materials, it is still possible to achieve angular phase-matching. To compute PM angles, we follow the approach of Ref. [45] briefly described in Sec. 2.3.1. In order to get the refractive index in an extended frequency range we need to compute the dielectric function, as described in Eq. (2.10).

Crystal system	Space group	Point group	Unit cell	Energy	n_s	d_{eff}
Orthorhombic (I)	$I2_12_12_1$	222	6	-307.3632	1.78	1.39
Monoclinic (I)	$P2_1$	2	12	-307.3080	1.79	1.79
Monoclinic (II)	$P2$	2	12	307.3575	1.77	1.51
Monoclinic (III)	Cc	m	6	-307.3606	1.77	2.25
Triclinic (I)	$P1$	1	12	-307.3330	1.78	1.54
Triclinic (II)	$P1$	1	12	-307.3338	1.78	1.58
Monoclinic (IV)	$P2$	2	12	-307.3046	1.79	1.61
Tetragonal	P-4	-4	12	-307.2580	1.81	1.83
Orthorhombic (II)	$P222_1$	222	12	-307.3546	1.77	1.37

Table 4.8: List of possible O-N ordering for PON structure. The crystal system, space group, point group, number of atoms in the unit cell, total DFT energy per formula unit (in eV), average refractive index n_s, and effective nonlinear coefficient d_{eff} (in pm/V) are shown for each possible ordering of the O-N sites.

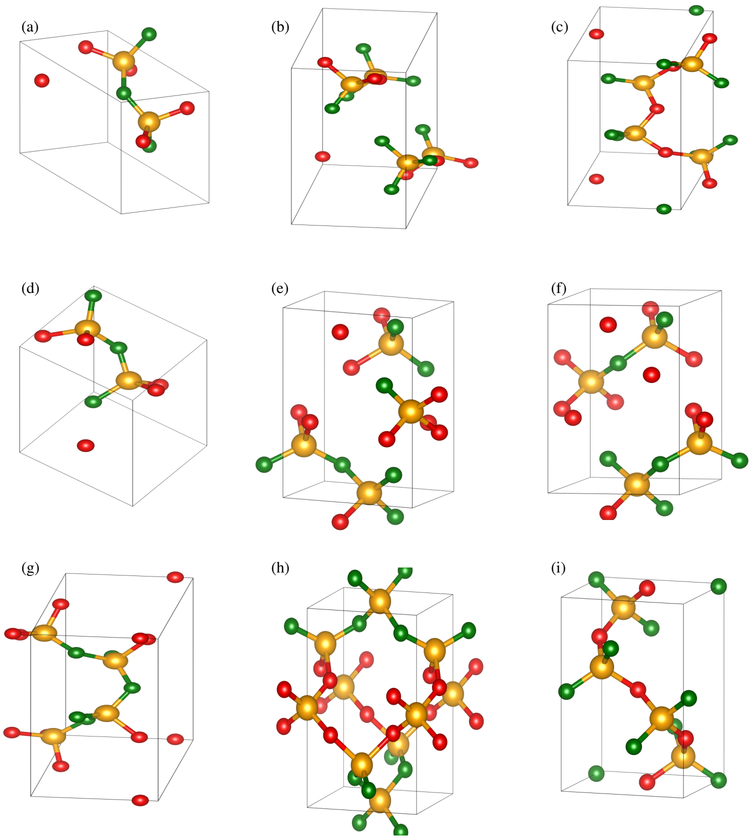


Figure 4.29: Structure of phosphorous oxynitride considering all the possible O-N arrangements. These structures are derived starting from the orthorhombic structure of PON (Fig. 4.27) stored in our DB. For each structure, the P atoms are colored in yellow, the O atoms are in red, and the N atoms are colored in green. (a) Orthorhombic (I); (b) Monoclinic (I); (c) Monoclinic (II); (d) Monoclinic (III); (e) Triclinic (I); (f) Triclinic (II); (g) Monoclinic (IV); (h) Tetragonal; (i) Orthorhombic (II);

At this stage, for the calculation of the dielectric function we used the so-called Independent Particle Approximation (IPA), or *Sum-Over-States* approach (see, for instance, Ref. [120]). This basically means that the evaluation of the imaginary part of the dielectric function is given by a form similar to Eq. (4.1). Within this approach, the energy transitions and matrix elements are taken from

DFT calculations of the electronic structure. To compensate the well-known systematic underestimation of the band gap, a scissor operator, calculated in order to match the HSE band gap, is considered. Furthermore, local-field and many-body effects are not taken into account in this case [121]. Convergence studies have been conducted for the important parameters such as the k-point sampling and the number of unoccupied bands. In Fig. 4.30, the different components of the real and imaginary part of the dielectric function are shown.

The convergence of the different parameters should not to be underestimated, because a small change in the value of the refractive index could produce a large change in the calculated PM angles. This is in line with good practice in HT calculations: when screening good materials, thus decreasing their number, the computational requirements on the parameters have to be quite stringent, to improve the prediction power. This is basically why, at this stage, we limit our calculation to the IPA case. As a matter of fact, even with the computational power available today, techniques such as the Random Phase Approximation and Bethe-Salpeter are still prohibitive for a large set of materials. However, once a good NLO candidate has been individuated in the DB, the calculation of the dielectric function within these approaches can be useful as a further confirmation of our results.

The refractive index of PON is shown in Fig. 4.31. To compute the PM properties, we have to check the fundamental and second harmonic value of the refractive index at that point. In this case, we consider for the fundamental light an energy of 1.17 eV, that corresponds to a wavelengths of 1064 nm and is indicated with a green dashed line, and the second harmonic value of 2.33 eV, that

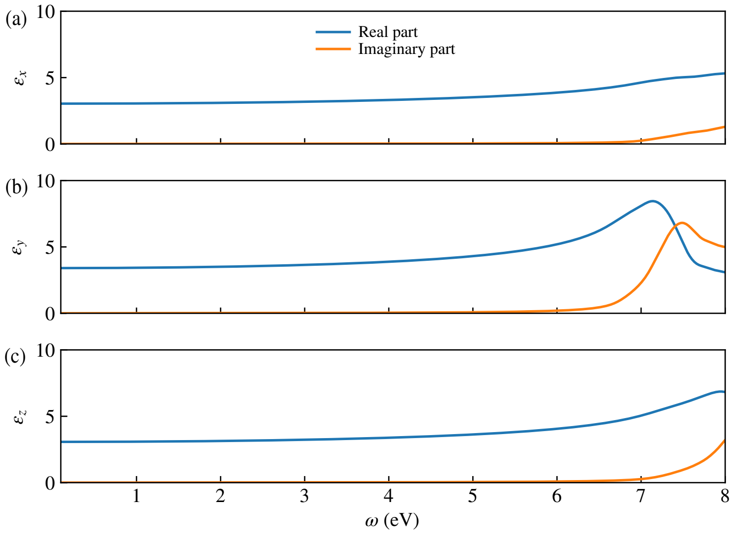


Figure 4.30: Real and imaginary part of the dielectric function of PON as a function of the frequency ((a) x component, (b) y component, (c) z component).

corresponds to a wavelength of 532 nm and is indicated with a red dashed line. These are typical wavelength of interest for optical applications. Of course, the same approach can be used as well with other values of the frequencies.

The results for type I and type II are shown in Fig. 4.32. Type I phase-matching means that, for instance, in SHG the two fundamental beams have the same polarization, perpendicular to that of the generated wave. Conversely, in Type II phase-matching, the two fundamental beams have different polarization directions [33].

For particular combinations of the angles (θ_m, ϕ_m), the so-called Phase match angles, phase-matching is achieved in the PON crystal.

For us, all these considerations, are a confirmation of PON as a good NLO optical material. However, once again, the final confirmation

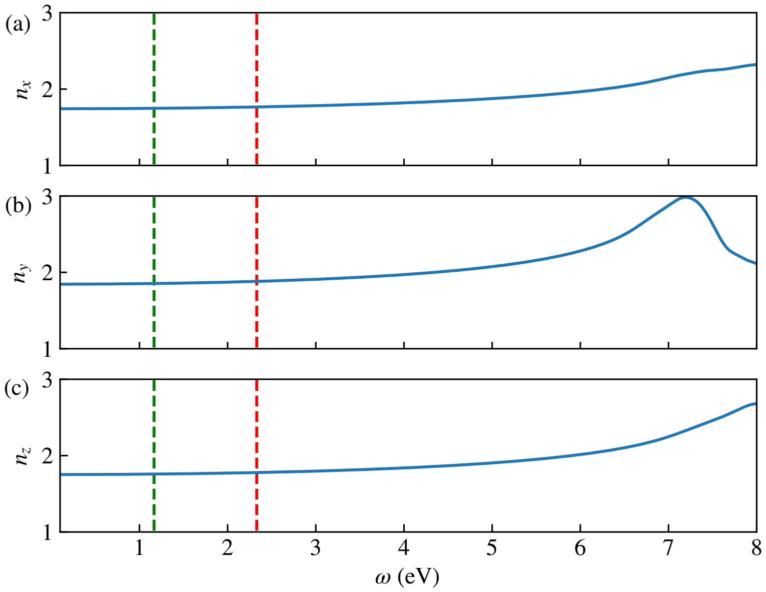


Figure 4.31: Refractive index of PON as a function of the frequency ((a) x component, (b) y component, (c) z component). The vertical dashed green and red lines correspond respectively to a frequency value of $\omega = 1.17$ eV (512 nm) and $\omega = 2.33$ eV (1024 nm).

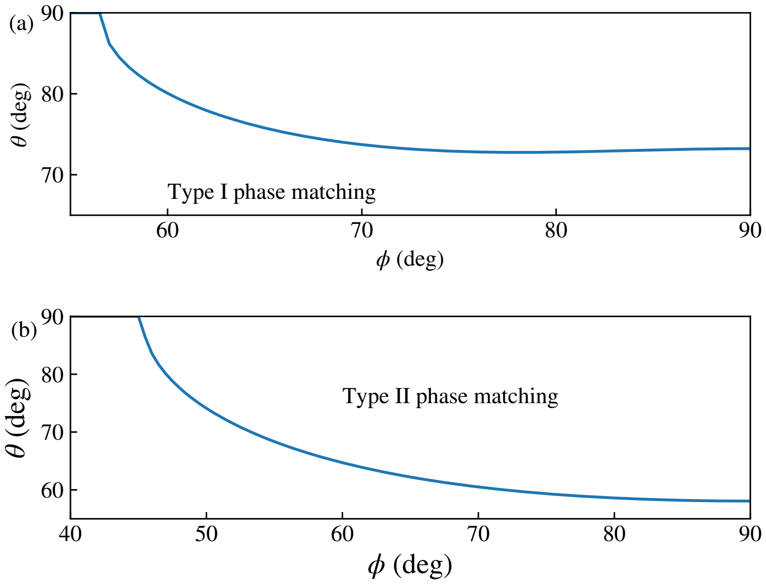


Figure 4.32: Type I and Type II phase match angles for PON.

is up to the experimental analysis. As a final consideration, we want to remark that the inclusion of the phase match angles for each material in our DB, will help us to continue the screening of materials for the search of a good NLO material.

5

Conclusions and future works

In this thesis, we have investigated the optical properties of semiconductors in the linear and in the nonlinear regime. We have combined Density Functional Perturbation Theory with the high-throughput computational approach to construct large databases containing the optical material properties such as the refractive index in the linear regime, and the nonlinear Second Harmonic Generation coefficient in the nonlinear regime.

After an overview on optical materials and the requirements that such materials need to meet in order to be high-performing optical materials, Density Functional Theory and Density Functional Perturbation Theory have been reviewed as methods to compute structural, electronic, and optical properties of materials. The combination of these two first-principles methods with the high-throughput approach has enabled us to compute the material properties for a large set of semiconductors. We then created two databases: one containing more than 4000 semiconductors in which the linear optical properties have been stored, and a second one containing at this stage more than 400 semiconductors, for which the nonlinear optical properties have been computed. The discussion of the results has been carried out separately for these two sets of data.

The initial list of candidates has been selected among the compounds present in the Materials Project repository, for which we collected materials properties such as structural properties (atomic positions, space group, ...) and electronic structural properties (band structures, density of states, ...). In the first part, we have computed the electronic and optical properties of 4040 semiconductors, calculating their band gap E_g^d and static refractive index n_s in the framework of density functional theory and density functional perturbation theory. Our data confirm the inverse relationship between n_s and E_g^d, but outliers are identified that combine a wide band gap with a high refractive index. Some of these are well-known optical materials (e.g. TiO_2, $LiNbO_3$, ...) while others have never been considered in this framework to the best of our knowledge (e.g., Ti_3PbO_7, $LiSi_2N_3$, BeS, ...).

Beyond the simple screening of a computational materials database to search for compounds with certain properties, the large amount of data provided through high-throughput computations can also be used to better understand the factors governing important material properties.

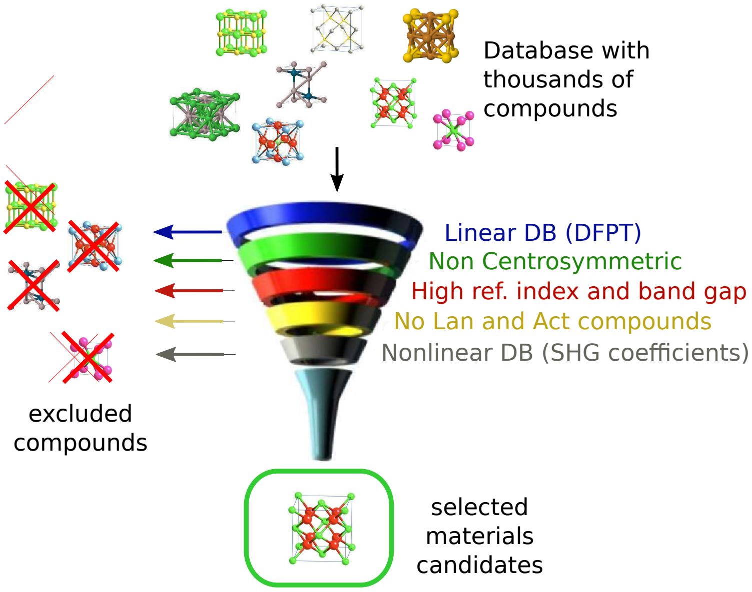
Thus, by mapping all the compounds onto a two-state system, two main descriptors are identified: the average optical gap and the effective frequency. While the former can be deduced directly from the electronic structure, the latter cannot. This limits the predictive power of our model and calls for further analysis (e.g., using a machine-learning approach). However, the model highlights that the decrease of n_s with E_g^d can be partly counterbalanced by a high number and density of available transitions from the top of the valence band to the bottom of the conduction band. This is directly related to the density of states at the edges of those bands and to the effective mass of such states.

By considering the compounds based on their chemical composition, we have then extracted some common features that can be useful in achieving a wide band gap dielectric. We have found that materials belonging to the first class of transition metal oxides and lanthanide oxides are the most promising ones for optical applications that require a wide band gap and a high refractive index.

The material database containing the linear optical materials properties served as a starting point to collect the optical properties in the nonlinear regime. We extracted 824 non-centrosymmetric materials out of the list of the 4040. The Second Harmonic Generation coefficient d_{eff} has been then computed so far for 457 semiconductors via DFPT and the $(2n+1)$ theorem. For the sake of convenience, the screening procedure indicated at the end of Chap. 3, is reported again here in Fig. 5.1.

Inspired by the Miller's rule, that relates the nonlinear optical susceptibility with the linear one, we were interested in the relationship between the d_{eff} and n_s. The explanation of such an empirical rule is of fundamental importance. Indeed, this would make it possible to determine the nonlinear optical properties of material just from the linear ones, that are more easily accessible both from a first-principles and an experimental point of view. From our data, we can confirm that materials with a high nonlinear coefficient also have a high refractive index. However, having a high response in the linear regime is not a sufficient condition to have a high nonlinear response as well. If we thus assume the validity of Miller's rule,

Figure 5.1: Sketch of the screening procedure followed to build up the database containing the nonlinear optical coefficients of the semiconductors analyzed in this study.



further investigation of the Miller's coefficient assumes a crucial part in the comprehension of the data distribution.

In order to find descriptors that explain the variation of the Miller's coefficient we considered Lithium Niobate, a well-known material in nonlinear optics. This material, as confirmed from our results, is one of the materials with the largest nonlinear coefficients in our data collection, namely, an outlier, having also a rather large band gap. Furthermore, this material shows two phases: a ferroelectric one, that is non-centrosymmetric, and thus Second Harmonic Generation (SHG) active, and a paraelectric phase that, being centrosymmetric, does not have any response in the nonlinear regime. While there are many similarities in the structures of these two phases, like the same number of atoms in the unit cell and the same crystal geometry, the atomic environments change going from one structure to the other.

Interpolating between the atomic positions of the centrosymmetric phase of Lithium Niobate and the non-centrosymmetric one we generated 6 intermediate structures for which we studied the variation of the optical coefficients as a function of the structural steps. This result is, in principle, in contrast with the Miller's rule. Indeed, by moving the atoms from their centrosymmetric configuration to the non-centrosymmetric one, as expected, we obtain an increase of the nonlinear optical coefficient d_{eff}, while the refractive index is decreasing. Although the decrease of the linear properties can be explained by looking at the band structure of the two phases, also making a connection with the descriptive model discussed for the comprehension of the data distribution in the linear database, the increase of the nonlinear coefficient cannot be explained. By looking then at the Miller's coefficient, we have an increase in this quantity, that explains the behavior of d_{eff}. We can conclude then, from our data, that the Miller's coefficient is material dependent, and has a high impact in the explanation of the data distribution, following the empirical Miller's rule.

As a future perspective, we think that it would be interesting to compute the Miller's coefficient in the form proposed by Scandolo and Bassani which only involves the calculation of the third derivative of the external potential and the calculation of the linear susceptibilities (see Eq. (2.46)). From our conclusions, we expect that by computing the Millers coefficient by the means of the derivative of the external potential, we will have the same trend as in Fig. 4.24. If this is the case, this approach can be easily extended to all the materials contained in the DB, since this quantity is in principle obtainable from the ground-state of the system, giving us further insight in the comprehension of the phenomena.

The last part of this thesis is focused on the description of a method for the search of new optimal candidates for nonlinear optical applications. By looking then at the distributions of materials considering the nonlinear effective coefficient d_{eff} vs. the band gap E_g^d, the expected inverse relationship is found. Having in mind the well-known band gap underestimation coming from Density Functional Theory calculations, we corrected the band gap by means of the HSE hybrid functional. Among the outliers that combine a wide band gap with high nonlinear optical coefficients, some of these are well-known compounds (e.g. $LiNbO_3$, $LiTaO_3$, $KTiPO_5$). In contrast, to the best of our knowledge, some of these outliers have not yet been considered as nonlinear optical materials (as for instance PON, B_2O_3, PbO, Sb_2WO_6, ..).

We considered then Phosphorous OxyNitride (PON) as an interesting example. From our calculation, this material looks indeed an interesting nonlinear optical material candidate for applications that fall in the deep-UV region of the electromagnetic spectrum. To further improve the quality of our material search, we computed the phase match angles for this material. Indeed, as highlighted many times, the nonlinear efficiency can be drastically reduced if the materials do not satisfy the phase-matching condition. Since phase-matching can be achieved by playing with the birefringence properties of the materials, we computed the refractive index of PON in the framework of the Independent Particle Approximations, aiming at calculating the phase match angles. We have found that for particular combinations of the angles (θ_m , ϕ_m), the so-called phase match angles, phase-matching is achieved in a PON crystal.

Knowing that the final confirmation is only given after the experimental analysis, we think that phosphorous oxynitride could be a good candidate, as an optical material, to be further investigated in the future.

It is worth stressing that our conclusions are inferred merely from statistical considerations. Such approaches can help in the understanding and construction of optical devices in a wide range of applications. Of course, for the final goal of speeding up and improving the material optimization process, the interaction through the exchange of data between the computational predictions and experimental results, which has the final word about the confirmation of the performances of the material, plays a key role.

For the future, our intentions are to increase the number of materials contained in our databases and to extend the range of properties stored for each material. Moreover, given the large amount of data generated, it would be interesting to apply machine learning techniques to generate predictive models of the optical properties of materials, and further improve the comprehension of optical phenomena. Finally it would be interesting to continue the investigation of the optical properties of the interesting candidates through methods that go beyond Density Functional Theory, such as many-body perturbation theory for instance, for which nowadays the calculation of Second Harmonic Generation spectra is also available. This would, of course, improve the material comprehension and would also be helpful to further support the experimental analysis because most of the experimental measurements are typically done at common communication wavelengths (i.e. around 800 nm or 1500nm).

We believe that the design process of optical materials will greatly benefit from the work proposed in this thesis. The most impactful subsequent steps to be undertaken should be:

- Miller's rule investigation
- Increase the information for each material contained in the databases in order to improve the material screening procedure
- Systematic approach to investigate angular phase matching
- Test machine learning techniques to further investigate the dataset at our disposal with the materials response in the linear and nonlinear optical regime
- Extend the analysis considering the frequency dependence of the susceptibility, at least for the most interesting materials

A

Hints about the relationship between ω_g and E_g^d

The relationship between the average optical gap ω_g and the direct band gap E_g^d can formally be deduced from Eq. (4.13), which only involves the JDOS $j(\omega)$. However, the form of $j(\omega)$ is often complex, which makes it impossible to solve the equation analytically.

Here, the solution is first derived for a simple model JDOS:

$$j(\omega) = \begin{cases} -\frac{6}{(2\sigma)^3}(\omega - E_g^d)\left(\omega - (E_g^d + 2\sigma)\right) & \text{for } E_g^d \le \omega \le E_g^d + 2\sigma \\ \\ 0 & \text{elsewhere} \end{cases} \tag{A.1}$$

This corresponds to a JDOS showing a single symmetric parabolic peak with a width σ (see Model 1 in Fig. A.1(a)). The analytic solution of Eq. (4.13) for this model JDOS is given by:

$$\omega_g^3 = -\frac{4}{3}\sigma^3 \left[\ln\left(\frac{E_g^d + 2\sigma}{E_g^d}\right) - \frac{2\sigma(E_g^d + \sigma)}{E_g^d(E_g^d + 2\sigma)} \right]^{-1} \tag{A.2}$$

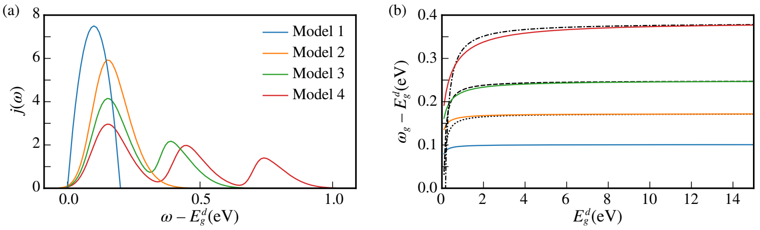


Figure A.1: (a) JDOS models (see text) and (b) corresponding relation between the average optical gap ω_g and ω_g-E_g^d (in eV) from Eq. (4.13). All models show the same trend which can be fitted using Eq. (A.3). The results obtained by truncating the development to n=1 (as in Eq. 4.16) are represented for Models 2, 3, and 4 using dotted, dashed, dot-dashed lines, respectively.

Using the properties of the logarithm and its Taylor expansion, we can finally write:

$$\omega_g = E_g^d + \sum_{n=0}^{\infty} \frac{c_n(\sigma)}{(E_g^d)^n} \tag{A.3}$$

where the coefficients $c_n(\sigma)$ are functions of the width σ. By truncating the development to n=1, we find Eq. (4.16).

Next, more complex JDOS models are considered. Model 2 consists of a skew normal distribution:

$$j(\omega) = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(\omega-\mu)^2}{2\sigma^2}} \left\{ 1 + \mathrm{erf}\left[\frac{\gamma(\omega-\mu)}{\sigma\sqrt{2}} \right] \right\} \tag{A.4}$$

where A, μ, σ, and γ are the amplitude, the position, the width, and the skewness of the peak, respectively. Models 3 and 4 are obtained by summing two and three such skew normal distributions.

For all models, an analytic solution of Eq. (4.13) is out of reach but ω_g can be computed numerically for different values of the peak position in order to determine its dependence with respect to E_g^d. In all cases, it is found that the solution has the form of Eq. (A.3) and that its truncation to n=1 provides a good approximation (see Fig. A.1(b)).

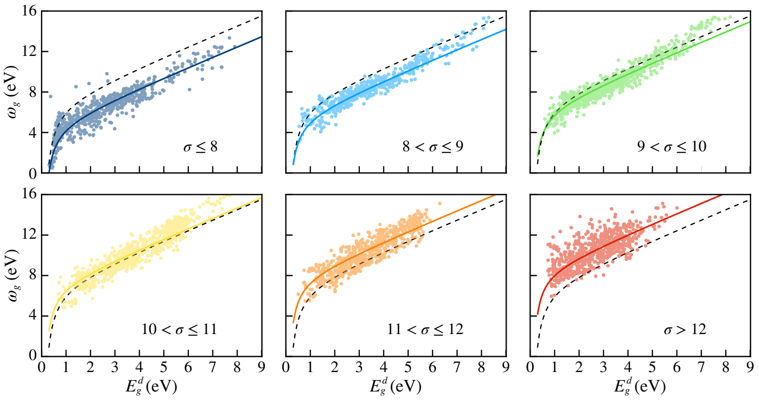


Figure A.2: Splitting of the ω_g-E_g^d data points considering the different ranges of the width σ. For each panel, the dashed black line corresponds to $\omega_g = E_g^d + 6.85 - 1.68/E_g^d$ which was obtained by fitting all the data, while the colored lines are obtained considering only the data in the subset represented in the panel.

Furthermore, playing with the parameters, we confirm that, just like for Model 1, the coefficients $c_n(\sigma)$ in Eq. (A.3) clearly depend on the width σ. These findings suggest that Eq. (A.3) truncated to n=1 could be used for any JDOS. This is illustrated in Fig. A.2 for all our calculated data. The dependence of the coefficients $c_n(\sigma)$ on the width σ has been highlighted by splitting the data into 6 groups according to the width of the real JDOS σ (computed as the difference between the mean value of the JDOS and the direct gap). This width depends on the dispersion of the bands, hence on the effective mass, and their distribution in energy, as illustrated in Fig. 4.7. This justifies the use of Eq. (4.16).

B

Comparison of the n_s-E_g^d models

As already mentioned, different empirical or semi-empirical models have been proposed for the expected inverse relationship between the refractive index n_s and the direct band gap E_g^d. A review of such models was recently proposed by Tripathy [90]. In Fig. B.1, we report the distributions of the absolute errors on the refractive index for various explicit functions of the direct band gap E_g^d compared to the DFPT computed data for the 4040 materials.

We consider various descriptors of the distribution of the absolute errors: the mean absolute error (MAE), the 25th, 50th, 75th percentiles (P25, P50, and P75, respectively). All of them indicate that our new model describes the calculated DFPT data better than the previously proposed ones.

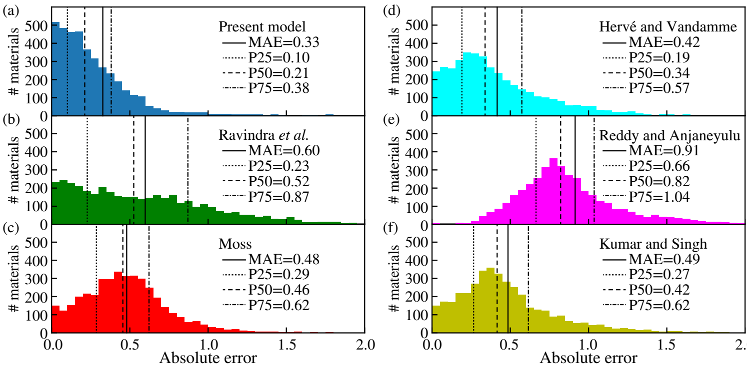


Figure B.1: Distribution of the absolute errors on the refractive index (for various explicit function of the direct band gap E_g^d vs. DFPT) for the 4040 materials. Our model (a) is compared to some well known empirical and semiempirical relations: (b) Ravindra *et al.* [5], (c) Moss [6], (d) Hervé and Vandamme [7], (d) Reddy and Anjaneyulu [8], and (e) Kumar and Singh [9]. For all the models, the mean absolute error (MAE), the 25th, 50th, 75th percentiles (P25, P50, and P75) are indicated.

C

Comparison with the Penn Model

The model presented here can directly be connected to the so-called Penn model in which the static refractive index is given by the following equation:

$$n_s^2 = \varepsilon_{1s} = 1 + \left(\frac{\omega_p}{\omega_g}\right)^2, \tag{C.1}$$

where ω_p is the Drude plasma frequency which is defined through the *f-sum rule*:

$$\int_0^\infty \omega\varepsilon_2(\omega)d\omega = \frac{\pi}{2}\omega_p^2. \tag{C.2}$$

Introducing the imaginary part of the dielectric function given by Eq. (4.9) into Eq. (C.2) we then obtain:

$$\omega_p^2 = 8\pi K \int_0^\infty \frac{j(\omega)}{\omega}d\omega. \tag{C.3}$$

If we now consider the two-state system previously introduced, whose JDOS is given by $j(\omega) = J\delta(\omega - \omega_g)$, the resulting plasma frequency is simply given by:

$$\omega_p^2 = \frac{8\pi K J}{\omega_g} = \frac{\omega_{\text{eff}}^3}{\omega_g}. \tag{C.4}$$

The two expressions for the static refractive index given by Eq. (4.14) and Eq. (C.1) are thus strictly equivalent. Consequently, the model presented in this study is closely related to Penn model. Our model has, however, an important advantage over Penn model for analyzing the data in that the effective frequency ω_{eff} is clearly independent of the average optical gap ω_g (it only depends on the integral of the JDOS J and the average transition probability K), while the plasma frequency ω_p is not.

D

Average transition probability vs. integral of the JDOS

From Eq. (4.12) we have that ω_{eff} is related to the product of the average transition probability K and the integral of the JDOS J. In Fig. D.1, we show the distribution of all our data points as a function of these two quantities (shown in a logarithmic scale for sake of clarity). The data points have been colored according to ω_{eff}. Though K is smaller than 1 for the vast majority of materials, both K and J impact the value of ω_{eff}.

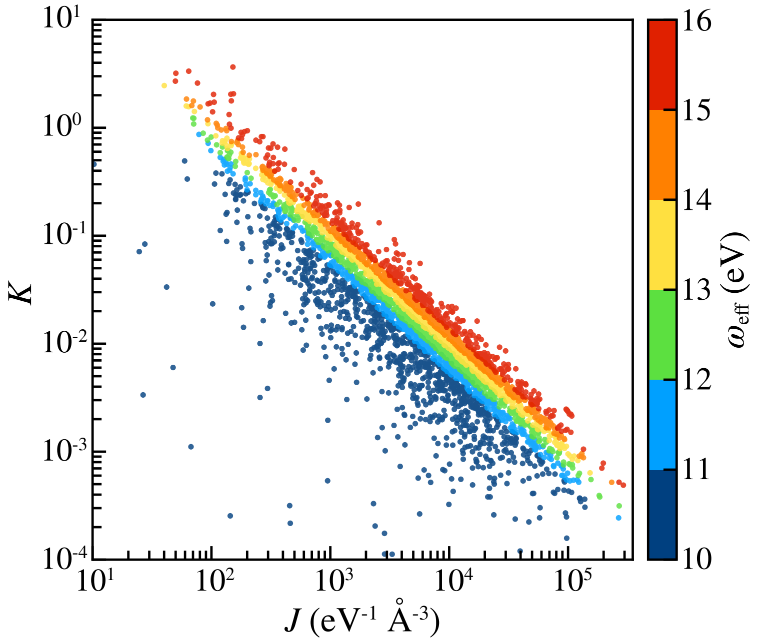


Figure D.1: Calculated values of the average transition probability K and integral of the JDOS J computed up to a frequency $\omega_{\max}$ for the complete dataset of materials. The data are reported as solid circles the color of which refers to the value of ω_{eff} as indicated in the color bar.

E

Tables of compounds

In this section, we provide various tables with the 10 materials with the highest refractive index for a given direct band gap range. In each table the chemical formula, MP identification (mp-id), average refractive index (n_s), diagonal components of the refractive index tensor (n_1, n_2, n_3), direct band gap in eV (E_g^d), the effective frequency in eV (ω_{eff}), the average optical gap in eV (ω_g) and the average effective mass of the transitions μ are shown. The compounds are sorted by the value of the average refractive index.

Table E.1: List of compounds with $E_g^d \leq 2.0\ eV$.

Formula	mp-id	n_s	n_1	n_2	n_3	E_g^d	ω_{eff}	ω_g	μ
TePb	mp-19717	5.12	5.12	5.12	5.12	0.87	17.47	5.96	0.08
FeS_2	mp-1522	4.59	4.41	4.77	4.60	1.17	11.88	4.37	0.83
Tl_2S	mp-667	4.46	4.48	4.58	4.32	0.86	19.12	7.18	0.25
SnSe	mp-691	4.25	4.60	4.25	3.91	0.59	7.62	2.95	0.24
RuS_2	mp-2030	4.09	4.09	4.09	4.09	0.91	13.35	5.32	0.34
Sb_2Se_3	mp-2160	3.90	4.30	4.28	3.10	0.80	10.76	4.45	0.47
GeSe	mp-700	3.79	3.97	3.78	3.62	0.93	13.97	5.89	0.33
Te_2Mo	mp-602	3.69	4.28	4.28	2.53	1.07	10.09	4.32	0.48
SnS	mp-2231	3.68	3.85	3.58	3.60	1.07	15.37	6.61	0.24
Sm_2Te_3	mp-684911	3.68	3.68	3.68	3.67	0.99	9.84	4.23	0.32

Table E.2: List of compounds with $2.0 \leq E_g^d \leq 3.0\ eV$.

Formula	mp-id	n_s	n_1	n_2	n_3	E_g^d	ω_{eff}	ω_g	μ
TiO_2	mp-34688	2.61	2.65	2.65	2.52	2.35	10.77	5.99	0.94
$YbTe_2$	mp-1779	2.60	2.60	2.60	2.60	2.46	12.98	7.23	0.19
Bi_2O_3	mp-23262	2.57	2.61	2.63	2.48	2.46	12.79	7.19	0.58
Ti_3PbO_7	mp-504427	2.57	2.60	2.59	2.51	2.32	14.01	7.90	114.11
CaTe	mp-1519	2.57	2.57	2.57	2.57	2.62	11.96	6.74	0.23
$LiBi_3I_2O_4$	mp-775430	2.56	2.62	2.62	2.44	2.25	12.40	7.00	0.55
B_8O	mp-530033	2.56	2.55	2.57	2.56	2.46	18.29	10.33	1.11
GeI_2	mp-27922	2.56	2.80	2.80	2.07	2.42	11.67	6.59	1.58
$Bi_4I_2O_5$	mp-30130	2.55	2.59	2.52	2.55	2.37	11.74	6.65	0.97
$TiPbO_3$	mp-20459	2.54	2.57	2.57	2.47	2.80	14.44	8.21	0.68

Table E.3: List of compounds with $3.0 \leq E_g^d \leq 4.0\ eV$.

Formula	mp-id	n_s	n_1	n_2	n_3	E_g^d	ω_{eff}	ω_g	μ
ZrSO	mp-3519	2.78	2.78	2.78	2.78	3.01	12.71	6.73	0.70
HfSO	mp-7787	2.65	2.65	2.65	2.65	3.56	13.22	7.25	0.70
$LiNb_3O_8$	mp-3368	2.39	2.45	2.38	2.34	3.00	12.15	7.25	3.62
$Ta_2Pb_2O_7$	mp-755663	2.39	2.40	2.37	2.39	3.11	15.36	9.17	5.62
Nb_2ZnO_6	mp-17177	2.38	2.48	2.32	2.34	3.30	13.40	8.02	3.25
$LaTa_7O_{19}$	mp-14485	2.38	2.35	2.35	2.43	3.13	13.36	8.00	2.68
BiClO	mp-22939	2.37	2.50	2.50	2.10	3.00	13.45	8.07	0.48
$NdTa_7O_{19}$	mp-14676	2.37	2.34	2.34	2.42	3.17	13.91	8.35	3.26
YTa_7O_{19}	mp-772036	2.36	2.34	2.34	2.42	3.16	13.81	8.31	3.45
$BaTiO_3$	mp-5020	2.35	2.35	2.35	2.35	2.97	12.74	7.69	1.11

Table E.4: List of compounds with $4.0 \leq E_g^d \leq 5.0\ eV$.

Formula	mp-id	n_s	n_1	n_2	n_3	E_g^d	ω_{eff}	ω_g	μ
ZrO_2	mp-755089	2.26	2.28	2.20	2.30	4.15	13.78	8.59	1.35
$ScTaO_4$	mp-558781	2.25	2.27	2.30	2.20	4.03	13.28	8.31	1.61
HfO_2	mp-1858	2.24	2.22	2.24	2.27	4.41	15.02	9.44	1.22
$PrScO_3$	mp-559756	2.22	2.24	2.23	2.19	4.26	14.05	8.90	1.15
$Sr_2Zr_7O_{16}$	mp-770419	2.22	2.20	2.20	2.25	4.02	13.42	8.51	3.67
$NdScO_3$	mp-31117	2.21	2.24	2.22	2.18	4.29	13.97	8.88	1.24
ThO_2	mp-643	2.20	2.20	2.20	2.20	4.51	13.40	8.56	0.83
HfO_2	mp-775757	2.20	2.23	2.21	2.16	4.02	14.84	9.48	1.44
$SmScO_3$	mp-31118	2.20	2.22	2.21	2.16	4.36	13.78	8.81	1.41
HfO_2	mp-352	2.19	2.22	2.21	2.14	4.10	14.85	9.51	1.76

Table E.5: List of compounds with $5.0 \leq E_g^d \leq 6.0\ eV$.

Formula	mp-id	n_s	n_1	n_2	n_3	E_g^d	ω_{eff}	ω_g	μ
BeS	mp-422	2.29	2.29	2.29	2.29	5.63	18.11	11.17	0.24
$BeSiN_2$	mp-7913	2.13	2.16	2.14	2.11	5.30	18.77	12.30	0.39
$LiSi_2N_3$	mp-5853	2.07	2.09	2.05	2.08	5.48	17.50	11.77	0.42
DyClO	mp-755323	2.07	2.09	2.09	2.02	5.17	14.68	9.88	0.45
HoClO	mp-29731	2.07	2.09	2.09	2.02	5.18	14.10	9.50	0.61
$HfSiO_4$	mp-4609	1.98	1.97	1.99	1.98	5.65	15.77	11.03	2.10
$ErBO_3$	mp-10791	1.95	1.94	1.96	1.94	5.15	15.93	11.32	2.68
$Pr_3Si_2ClO_8$	mp-554826	1.93	1.93	1.93	1.93	5.15	14.79	10.56	1.25
$Nd_2Be_2SiO_7$	mp-9077	1.92	1.89	1.93	1.93	5.03	15.86	11.41	0.64
$Y_2Be_2SiO_7$	mp-6655	1.90	1.87	1.91	1.91	5.19	15.57	11.31	0.63

Table E.6: List of compounds with $6.0 \leq E_g^d \leq 7.0\ eV$.

Formula	mp-id	n_s	n_1	n_2	n_3	E_g^d	ω_{eff}	ω_g	μ
$BeAl_2O_4$	mp-3081	1.78	1.78	1.78	1.78	6.12	18.71	14.45	0.37
$MgAlBO_4$	mp-8376	1.72	1.72	1.72	1.72	6.23	17.76	14.20	0.42
LiCl	mp-22905	1.72	1.72	1.72	1.72	6.25	15.92	12.75	0.38
LaF_3	mp-905	1.70	1.70	1.70	1.70	6.04	10.20	8.24	9.19
$NaPr_2S_2O_8F_3$	mp-560673	1.70	1.70	1.71	1.68	6.18	14.97	12.13	6.05
CaB_2O_4	mp-8056	1.69	1.73	1.71	1.64	6.09	15.18	12.31	0.60
$Al_6B_5O_15F_3$	mp-6738	1.68	1.68	1.68	1.67	6.17	17.08	14.03	0.53
$BaBePO_4F$	mp-754604	1.67	1.67	1.67	1.67	6.24	15.22	12.52	0.47
SiO_2	mp-549166	1.65	1.82	1.57	1.57	6.00	16.26	13.53	0.47
LiB_3O_5	mp-3660	1.62	1.63	1.64	1.60	6.35	16.13	13.70	1.02

F

Importance of the exchange-correlation functional

It is clear that the exchange-correlation (XC) functional will affect the electronic structure (the band gap and, possibly, the band dispersion) and the optical properties. Here, we analyze how using HSE instead of PBE changes the calculated value of ω_g for the top materials in the Tables E.1-E.6. To this end, in Fig. F.1, we compare the JDOS $j(\omega)$ and $j(\omega)/\omega^3$ computed with computed with both XC functionals. Obviously, the PBE gaps are smaller than the HSE ones so a scissor Δ was applied to the PBE electronic structure in order to match the HSE gap. A comparison of the average gap ω_g computed with PBE+Δ_{HSE} and HSE for the selected materials is also shown in Table F.1. For each material the value of the Δ_{HSE} scissor operator is reported. Looking at the different ω_g values one can see that there is a small difference considering the two XC functionals, with an absolute error ranging from 0.15 to 1.09 eV.

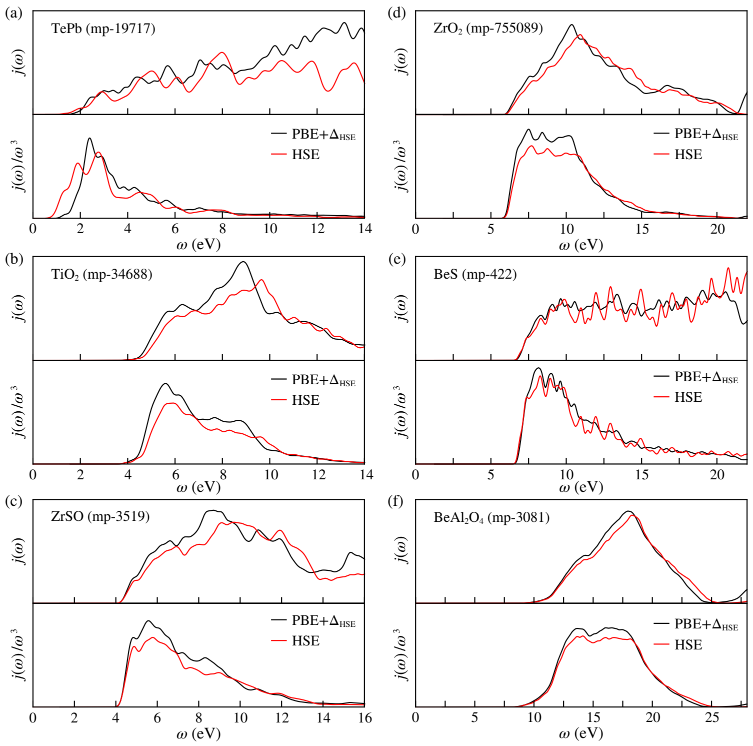


Figure F.1: Optical functions $j(\omega)$ and $j(\omega)/\omega^3$ computed with PBE+Δ_{HSE} and HSE for the top materials in the Tables E.1-E.6.

Table F.1: Comparison of the ω_g values (in eV) computed considering PBE+Δ_{HSE} and HSE for the top materials in the Tables E.1-E.6. For each material also the value of the scissor operator Δ_{HSE} (in eV) is reported.

Formula	MP-id	Δ_{HSE}	ω_g (PBE+Δ_{HSE})	ω_g (HSE)
TePb	mp-19717	0.54	7.09	6.00
TiO_2	mp-34688	1.57	7.73	8.00
ZrSO	mp-3519	1.30	8.24	8.39
ZrO_2	mp-755089	1.84	10.52	10.75
BeS	mp-422	1.26	12.41	12.64
$BeAl_2O_4$	mp-3081	2.25	16.59	16.79

G

$LiNbO_3$: Electronic structures

In this section, we provide the electronic structure for the 6 different stuctures of $LiNbO_3$ (see Sec. 4.2.1). For each structure, the electronic band structure and the projected density of states (DOS) is given.

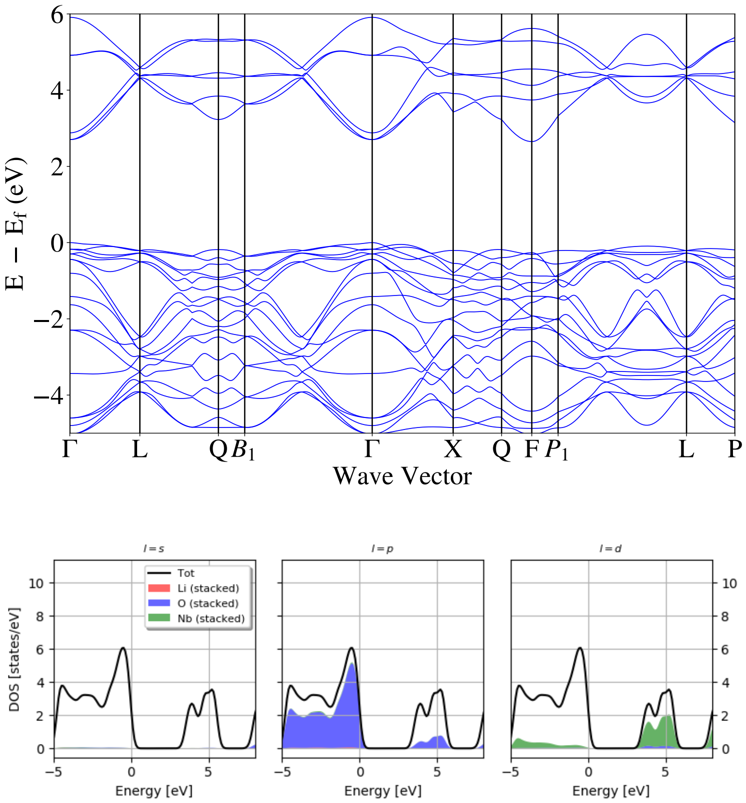


Figure G.1: Electronic structural information of $LiNbO_3$ in its centrosymmetric phase (Fig. 4.19(a)).

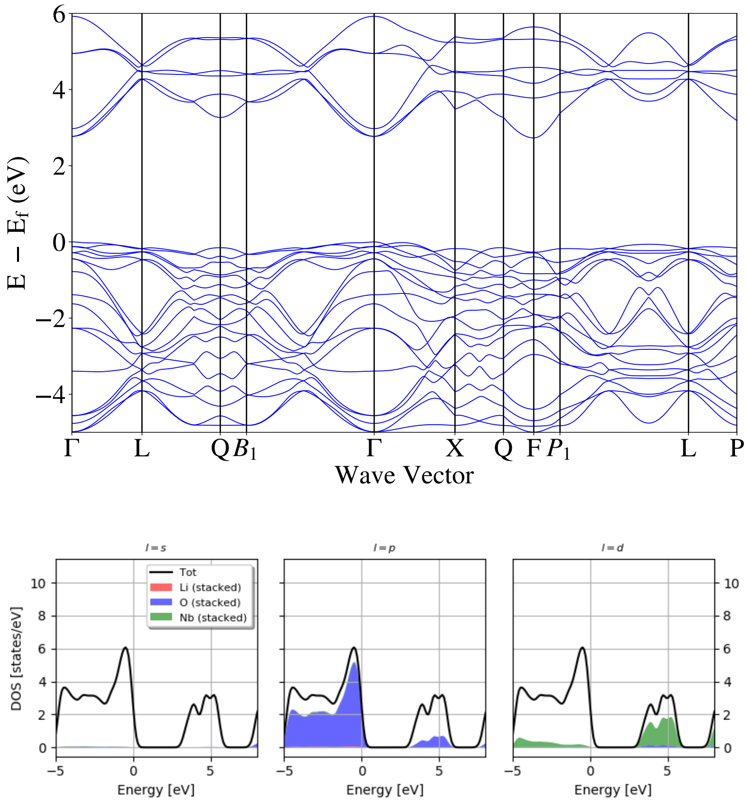


Figure G.2: Electronic structural information of $LiNbO_3$ after a first displacement of the atoms from its centrosymmetric positions (Fig. 4.20(a)).

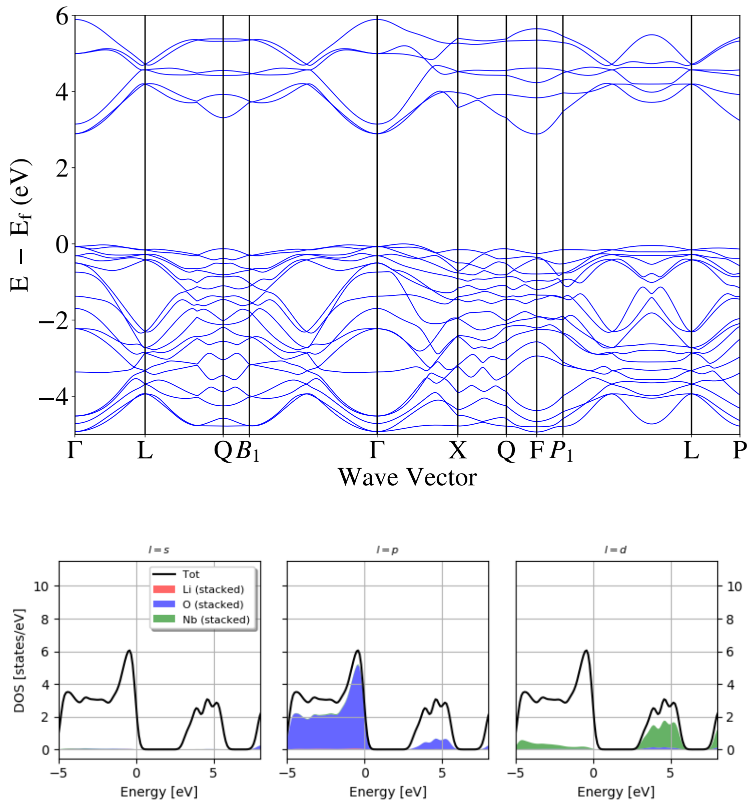


Figure G.3: Electronic structural information of $LiNbO_3$ after a second displacement of the atoms from its centrosymmetric positions (Fig. 4.20(b)).

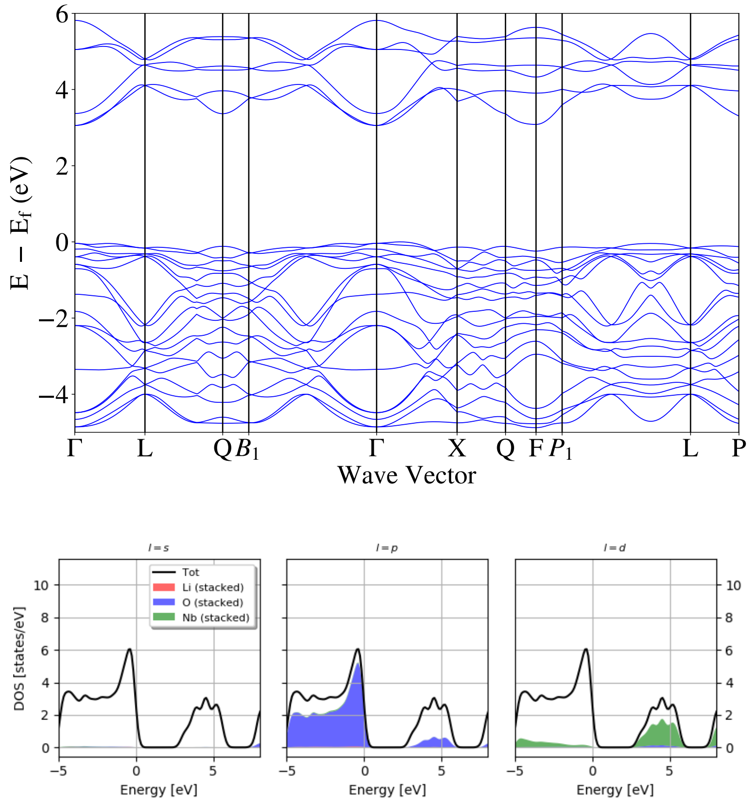


Figure G.4: Electronic structural information of LiNbO$_3$ after a third displacement of the atoms from its centrosymmetric positions (Fig. 4.20(c)).

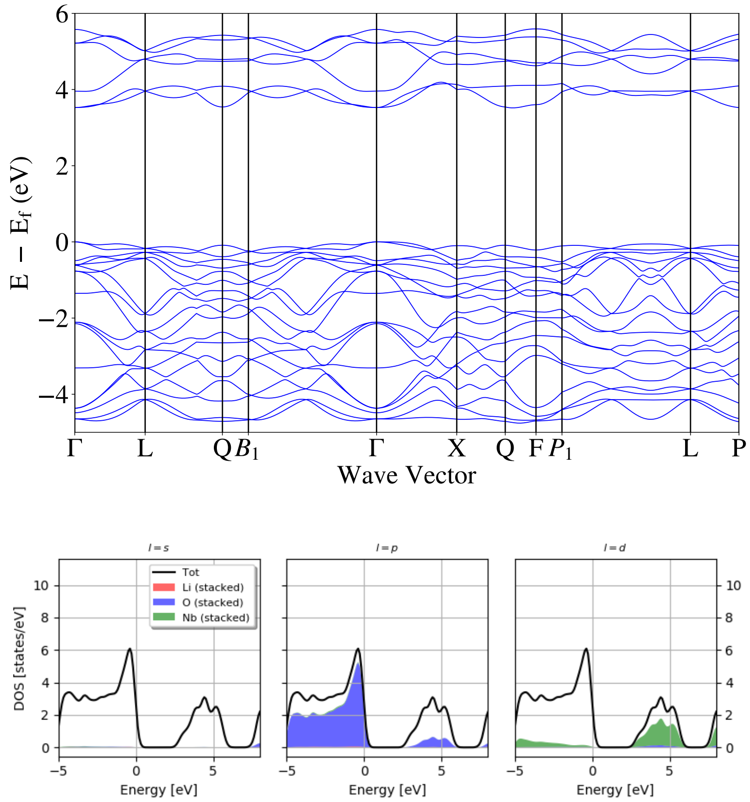


Figure G.5: Electronic structural information of $LiNbO_3$ after a fourth displacement of the atoms from its centrosymmetric positions (Fig. 4.20(d)).

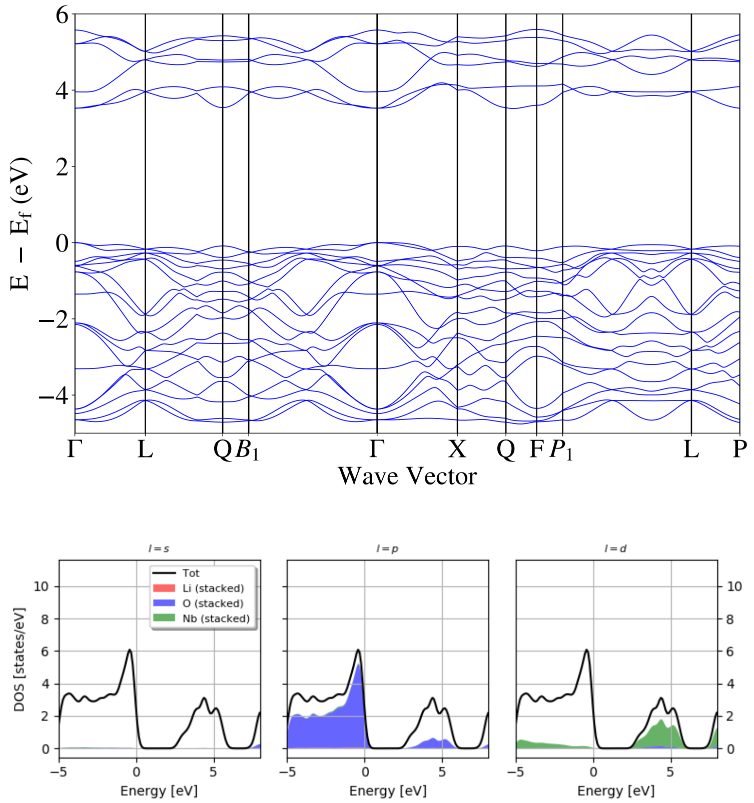


Figure G.6: Electronic structural information of $LiNbO_3$ in its non-centrosymmetric phase (Fig. 4.19(b)).

Bibliography

[1] J. M. Rondinelli and E. Kioupakis. Predicting and designing optical properties of inorganic materials. *Annu. Rev. Mater. Res.*, 45(1):491–518, 2015. doi: 10.1146/annurev-matsci-070214-021150. URL `http://dx.doi.org/10.1146/annurev-matsci-070214-021150`.

[2] Xiaobo Yin, Ziliang Ye, Daniel A. Chenet, Yu Ye, Kevin O'Brien, James C. Hone, and Xiang Zhang. Edge nonlinear optics on a MoS_2 atomic monolayer. *Science*, 344 (6183):488–490, 2014. ISSN 0036-8075. doi: 10.1126/science.1250564. URL `https://science.sciencemag.org/content/344/6183/488`.

[3] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, 64:1045–1097, Oct 1992. doi: 10.1103/RevModPhys.64.1045. URL `https://link.aps.org/doi/10.1103/RevModPhys.64.1045`.

[4] Alejandro J. Garza and Gustavo E. Scuseria. Predicting band gaps with hybrid density functionals. *The Journal of Physical Chemistry Letters*, 7(20):4165–4170, 2016. doi: 10.1021/acs.jpclett.6b01807. URL `https://doi.org/10.1021/acs.jpclett.6b01807`.

[5] N. M. Ravindra, Sushil Auluck, and V. K. Srivastava. On the Penn gap in semiconductors. *Phys. Status Solidi B*, 93(2):K155–K160, 1979. doi: 10.1002/pssb.

2220930257. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pssb.2220930257.

[6] T. S. Moss. Relations between the refractive index and energy gap of semiconductors. *Phys. Status Solidi B*, 131(2):415–427, 1985. doi: 10.1002/pssb.2221310202. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pssb.2221310202.

[7] P. Hervé and L.K.J. Vandamme. General relation between refractive index and energy gap in semiconductors. *Infrared Phys. Technol.*, 35(4):609 – 615, 1994.

[8] R. R. Reddy and S. Anjaneyulu. Analysis of the Moss and Ravindra relations. *Phys. Status Solidi B*, 174(2): K91–K93, 1992. doi: 10.1002/pssb.2221740238. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pssb.2221740238.

[9] V Kumar and JK Singh. Model for calculating the refractive index of different materials. *Indian J. Pure Appl. Phys.*, 48: 571–574, 2010.

[10] Robert C. Miller. Optical second harmonic generation in piezoelectric crystal. *Appl. Phys. Lett.*, 5(1):17–19, 1964. doi: 10.1063/1.1754022. URL https://doi.org/10.1063/1.1754022.

[11] Philippe Ghosez and J Junquera. First-principles modeling of ferroelectric oxide nanostructures. *Handbook of theoretical and computational nanotechnology*, pages 623–728, 2006.

[12] Ichiro Shoji, Takashi Kondo, Ayako Kitamoto, Masayuki Shirane, and Ryoichi Ito. Absolute scale of second-order

nonlinear-optical coefficients. *J. Opt. Soc. Am. B*, 14(9):2268–2294, Sep 1997. doi: 10.1364/JOSAB.14.002268. URL http://josab.osa.org/abstract.cfm?URI=josab-14-9-2268.

[13] P Hermet, M Veithen, and Ph Ghosez. First-principles calculations of the nonlinear optical susceptibilities and raman scattering spectra of lithium niobate. *Journal of Physics: Condensed Matter*, 19(45):456202, oct 2007. doi: 10.1088/0953-8984/19/45/456202. URL https://doi.org/10.10882F0953-89842F192F452F456202.

[14] Anubhav Jain, Kristin A. Persson, and Gerbrand Ceder. Research update: The materials genome initiative: Data sharing and the impact of collaborative ab initio databases. *APL Materials*, 4(5):053102, 2016. doi: 10.1063/1.4944683. URL https://aip.scitation.org/doi/abs/10.1063/1.4944683.

[15] M. L. Green, C. L. Choi, J. R. Hattrick-Simpers, A. M. Joshi, I. Takeuchi, S. C. Barron, E. Campo, T. Chiang, S. Empedocles, J. M. Gregoire, A. G. Kusne, J. Martin, A. Mehta, K. Persson, Z. Trautt, J. Van Duren, and A. Zakutayev. Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies. *Applied Physics Reviews*, 4(1):011105, 2017. doi: 10.1063/1.4977487. URL https://doi.org/10.1063/1.4977487.

[16] Juan J. de Pablo, Barbara Jones, Cora Lind Kovacs, Vidvuds Ozolins, and Arthur P. Ramirez. The materials genome initiative, the interplay of experiment, theory and computation. *Current Opinion in Solid State and Materials Science*, 18(2):99 – 117, 2014. ISSN 1359-0286. doi: https://doi.org/10.1016/j.cossms.2014.02.003.

URL http://www.sciencedirect.com/science/article/pii/S1359028614000060.

[17] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy. The high-throughput highway to computational materials design. *Nat. Mater.*, 12:191–201, 2013. ISSN 1476-1122. doi: http://dx.doi.org/10.1038/nmat3568.

[18] Anubhav Jain, Yongwoo Shin, and Kristin A. Persson. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.*, 1:15004, 2016. URL http://dx.doi.org/10.1038/natrevmats.2015.4.

[19] Dane Morgan, Gerbrand Ceder, and Stefano Curtarolo. High-throughput and data mining with ab initio methods. *Meas. Sci. Technol.*, 16(1):296–301, December 2004. ISSN 0957-0233, 1361-6501. doi: 10.1088/0957-0233/16/1/039. URL http://dx.doi.org/10.1088/0957-0233/16/1/039.

[20] Katja Hansen, Grégoire Montavon, Franziska Biegler, Siamac Fazli, Matthias Rupp, Matthias Scheffler, O. Anatole von Lilienfeld, Alexandre Tkatchenko, and Klaus-Robert Müller. Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theory Comput.*, 9(8):3404–3419, July 2013. ISSN 1549-9618, 1549-9626. doi: 10.1021/ct400195d. URL http://dx.doi.org/10.1021/ct400195d.

[21] de Pablo, J. Juan, Nicholas E. Jackson, Michael A. Webb, Long-Qing Chen, Joel E. Moore, Dane Morgan, Ryan Jacobs, Tresa Pollock, Darrell G. Schlom, Eric S. Toberer, James Analytis, Ismaila Dabo, Dean M. DeLongchamp, Gregory A. Fiete, Gregory M. Grason, Geoffroy Hautier, Yifei Mo, Krishna

Rajan, Evan J. Reed, Efrain Rodriguez, Vladan Stevanovic, Jin Suntivich, Katsuyo Thornton, and Ji-Cheng Zhao. New frontiers for the materials genome initiative. *npj Computational Materials*, 5(1):41, April 2019. ISSN 2057-3960. URL https://doi.org/10.1038/s41524-019-0173-4.

[22] Ioannis Petousis, Wei Chen, Geoffroy Hautier, Tanja Graf, Thomas D. Schladt, Kristin A. Persson, and Fritz B. Prinz. Benchmarking density functional perturbation theory to enable high-throughput screening of materials for dielectric constant and refractive index. *Phys. Rev. B*, 93:115151, 2016. doi: 10.1103/PhysRevB.93.115151. URL https://link.aps.org/doi/10.1103/PhysRevB.93.115151.

[23] Ioannis Petousis, David Mrdjenovich, Eric Ballouz, Miao Liu, Donald Winston, Wei Chen, Tanja Graf, Thomas D. Schladt, Kristin A. Persson, and Fritz B. Prinz. High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials. *Sci. Data*, 4:160134, 2017. URL http://dx.doi.org/10.1038/sdata.2016.134.

[24] Francesco Naccarato, Francesco Ricci, Jin Suntivich, Geoffroy Hautier, Ludger Wirtz, and Gian-Marco Rignanese. Searching for materials with high refractive index and wide band gap: A first-principles high-throughput study. *Phys. Rev. Materials*, 3:044602, Apr 2019. doi: 10.1103/PhysRevMaterials.3.044602. URL https://link.aps.org/doi/10.1103/PhysRevMaterials.3.044602.

[25] T. W. Odom, R. M. Dickson, M. A. Duncan, and W. Tan. Shining a light on the molecular and nanoscopic worlds. *ACS Photonics*, 2:787–789, 2015. doi: 10.1021/acsphotonics.

5b00337. URL https://pubs.acs.org/doi/abs/10.1021/acsphotonics.5b00337.

[26] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich. Generation of optical harmonics. *Phys. Rev. Lett.*, 7:118–119, Aug 1961. doi: 10.1103/PhysRevLett.7.118. URL https://link.aps.org/doi/10.1103/PhysRevLett.7.118.

[27] T. H. Maiman. Optical and microwave-optical experiments in ruby. *Phys. Rev. Lett.*, 4:564–566, Jun 1960. doi: 10.1103/PhysRevLett.4.564. URL https://link.aps.org/doi/10.1103/PhysRevLett.4.564.

[28] Jiesu Wang, Kuijuan Jin, Hongbao Yao, Junxing Gu, Xiulai Xu, Chen Ge, Can Wang, Meng He, and Guozhen Yang. Temperature-dependent phase transition in barium titanate crystals probed by second harmonic generation. *Applied Physics Letters*, 112(10):102904, 2018. doi: 10.1063/1.5023293. URL https://doi.org/10.1063/1.5023293.

[29] Sava A. Denev, Tom T. A. Lummen, Eftihia Barnes, Amit Kumar, and Venkatraman Gopalan. Probing ferroelectrics using optical second harmonic generation. *Journal of the American Ceramic Society*, 94(9):2699–2727, 2011. doi: 10.1111/j.1551-2916.2011.04740.x. URL https://ceramics.onlinelibrary.wiley.com/doi/abs/10.1111/j.1551-2916.2011.04740.x.

[30] Moritz Förderer, Tihomir Georgiev, Matias Mosqueira, Rainer H. A. Fink, and Martin Vogel. Functional second harmonic generation microscopy probes molecular dynamics with high temporal resolution. *Biomed. Opt. Express*, 7(2):525–541, Feb 2016. doi: 10.1364/

BOE.7.000525. URL `http://www.osapublishing.org/boe/abstract.cfm?URI=boe-7-2-525`.

[31] Jung-Hwan Song, Arthur J. Freeman, Tarun K. Bera, In Chung, and Mercouri G. Kanatzidis. First-principles prediction of an enhanced optical second-harmonic susceptibility of low-dimensional alkali-metal chalcogenides. *Phys. Rev. B*, 79:245203, Jun 2009. doi: 10.1103/PhysRevB.79.245203. URL `https://link.aps.org/doi/10.1103/PhysRevB.79.245203`.

[32] T. T. Tran, H. Yu, J. M. Rondinelli, K. R. Poeppelmeier, and P. S. Halasyamani. Deep ultraviolet nonlinear optical materials. *Chem. Mater.*, 28(15):5238–5258, 2016. doi: 10.1021/acs.chemmater.6b02366. URL `http://dx.doi.org/10.1021/acs.chemmater.6b02366`.

[33] Robert W. Boyd. *Nonlinear Optics, Third Edition*. Academic Press, Inc., Orlando, FL, USA, 3rd edition, 2008. ISBN 0123694701, 9780123694706.

[34] F. Zernike and J.E. Midwinter. *Applied Nonlinear Optics*. Dover books on physics. Dover Publications, 2006. ISBN 9780486453606. URL `https://books.google.be/books?id=Y3Wsyo7TgdkC`.

[35] Geoffrey New. *Introduction to Nonlinear Optics*. Cambridge University Press, 2011.

[36] Y. R. Shen. *The principles of nonlinear optics*. Hoboken, 2003. ISBN 0471430803.

[37] J. Ducuing. *Nonlinear Optics*. Academic press, 1975.

[38] David J. Griffiths and Darrell F. Schroeter. *Introduction to Quantum Mechanics*. Cambridge University Press, 3^{rd} edition, 2018.

[39] S. Scandolo and F. Bassani. Miller's rule and the static limit for second-harmonic generation. *Phys. Rev. B*, 51:6928–6931, Mar 1995. doi: 10.1103/PhysRevB.51.6928. URL `https://link.aps.org/doi/10.1103/PhysRevB.51.6928`.

[40] S. Scandolo and F. Bassani. Kramers-Kronig relations and sum rules for the second-harmonic susceptibility. *Phys. Rev. B*, 51:6925–6927, Mar 1995. doi: 10.1103/PhysRevB.51.6925. URL `https://link.aps.org/doi/10.1103/PhysRevB.51.6925`.

[41] J. F. Nye. *Electronic structure and the properties of solids: the physics of the chemical bond*. Oxford University Press, U.S.A., 2002.

[42] D. A. Kleinman. Nonlinear dielectric polarization in optical media. *Phys. Rev.*, 126:1977–1979, Jun 1962. doi: 10.1103/PhysRev.126.1977. URL `https://link.aps.org/doi/10.1103/PhysRev.126.1977`.

[43] J E Midwinter and J Warner. The effects of phase matching method and of uniaxial crystal symmetry on the polar distribution of second-order non-linear optical polarization. *British Journal of Applied Physics*, 16(8):1135–1142, aug 1965. doi: 10.1088/0508-3443/16/8/312. URL `https://doi.org/10.10882F0508-34432F162F82F312`.

[44] S. K. Kurtz and T. T. Perry. A powder technique for the evaluation of nonlinear optical materials. *Journal of Applied Physics*, 39(8):3798–3813, 1968. doi: 10.1063/1.1656857. URL https://doi.org/10.1063/1.1656857.

[45] J. Q. Yao and Theodore S. Fahlen. Calculations of optimum phase match parameters for the biaxial crystal $KTiOPO_4$. *Journal of Applied Physics*, 55(1):65–68, 1984. doi: 10.1063/1.332850. URL https://doi.org/10.1063/1.332850.

[46] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864–B871, 1964. doi: 10.1103/PhysRev.136.B864. URL http://link.aps.org/doi/10.1103/PhysRev.136.B864.

[47] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965. doi: 10.1103/PhysRev.140.A1133. URL http://link.aps.org/doi/10.1103/PhysRev.140.A1133.

[48] S. Baroni, P. Giannozzi, and A. Testa. Green's-function approach to linear response in solids. *Phys. Rev. Lett.*, 58:1861–1864, 1987. doi: 10.1103/PhysRevLett.58.1861. URL http://link.aps.org/doi/10.1103/PhysRevLett.58.1861.

[49] X. Gonze. Perturbation expansion of variational principles at arbitrary order. *Phys. Rev. A*, 52:1086–1095, 1995. doi: 10.1103/PhysRevA.52.1086. URL http://link.aps.org/doi/10.1103/PhysRevA.52.1086.

[50] Geoffroy Hautier, Anubhav Jain, and Shyue Ping Ong. From the computer to the laboratory: materials discovery and design using first-principles calculations. *J. Mater. Sci.*,

47(21):7317–7340, 2012. ISSN 1573-4803. doi: 10.1007/s10853-012-6424-0. URL `http://dx.doi.org/10.1007/s10853-012-6424-0`.

[51] B. H. Bransden and C. J. Joachain. *Physics of Atoms and Molecules*. Pearson Education, 2003.

[52] Richard M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511805769.

[53] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996. doi: 10.1103/PhysRevLett.77.3865. URL `http://link.aps.org/doi/10.1103/PhysRevLett.77.3865`.

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

[55] Kurt Lejaeghere, Gustav Bihlmayer, Torbjörn Björkman, Peter Blaha, Stefan Blügel, Volker Blum, Damien Caliste, Ivano E. Castelli, Stewart J. Clark, Andrea Dal Corso, Stefano de Gironcoli, Thierry Deutsch, John Kay Dewhurst, Igor Di Marco, Claudia Draxl, Marcin Dułak, Olle Eriksson, José A. Flores-Livas, Kevin F. Garrity, Luigi Genovese, Paolo Giannozzi, Matteo Giantomassi, Stefan Goedecker, Xavier Gonze, Oscar Grånäs, E. K. U. Gross, Andris Gulans, François Gygi, D. R. Hamann, Phil J. Hasnip, N. A. W. Holzwarth, Diana Iuşan, Dominik B. Jochym, François Jollet, Daniel Jones, Georg Kresse, Klaus Koepernik, Emine Küçükbenli, Yaroslav O. Kvashnin, Inka L. M. Locht, Sven Lubeck, Martijn Marsman, Nicola Marzari, Ulrike Nitzsche, Lars Nordström, Taisuke Ozaki, Lorenzo Paulatto,

Chris J. Pickard, Ward Poelmans, Matt I. J. Probert, Keith Refson, Manuel Richter, Gian-Marco Rignanese, Santanu Saha, Matthias Scheffler, Martin Schlipf, Karlheinz Schwarz, Sangeeta Sharma, Francesca Tavazza, Patrik Thunström, Alexandre Tkatchenko, Marc Torrent, David Vanderbilt, Michiel J. van Setten, Veronique Van Speybroeck, John M. Wills, Jonathan R. Yates, Guo-Xu Zhang, and Stefaan Cottenier. Reproducibility in density functional theory calculations of solids. *Science*, 351(6280), 2016. ISSN 0036-8075. doi: 10.1126/science.aad3000. URL https://science.sciencemag.org/content/351/6280/aad3000.

[56] Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang. Insights into current limitations of density functional theory. *Science*, 321(5890):792–794, 2008. ISSN 0036-8075. doi: 10.1126/science.1158722. URL https://science.sciencemag.org/content/321/5890/792.

[57] P. Giannozzi. Numerical methods for electronic structure, chapter 2, February 2010. URL http://www.fisica.uniud.it/~giannozz/Didattica/MetNum/LectureNotes/metnum-cap2.pdf.

[58] Wilfried G. Aulbur, Lars Jönsson, and John W. Wilkins. Quasiparticle calculations in solids. 54:1 – 218, 2000. ISSN 0081-1947. doi: https://doi.org/10.1016/S0081-1947(08)60248-9. URL http://www.sciencedirect.com/science/article/pii/S0081194708602489.

[59] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Erratum: hybrid functionals based on a screened coulomb potential [J. Chem. Phys. 118, 8207 (2003)]. *The Journal*

of Chemical Physics, 124(21):219906, 2006. doi: 10.1063/1.2204597. URL `https://doi.org/10.1063/1.2204597`.

[60] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton. Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+U study. *Phys. Rev. B*, 57:1505–1509, 1998. doi: 10.1103/PhysRevB.57.1505. URL `http://link.aps.org/doi/10.1103/PhysRevB.57.1505`.

[61] S. Baroni and R. Resta. *Ab initio* calculation of the macroscopic dielectric constant in silicon. *Phys. Rev. B*, 33:7017–7021, 1986. doi: 10.1103/PhysRevB.33.7017. URL `http://link.aps.org/doi/10.1103/PhysRevB.33.7017`.

[62] X. Gonze. Adiabatic density-functional perturbation theory. *Phys. Rev. A*, 52:1096–1114, 1995. doi: 10.1103/PhysRevA.52.1096. URL `http://link.aps.org/doi/10.1103/PhysRevA.52.1096`.

[63] Andrea Dal Corso and Francesco Mauri. Wannier and bloch orbital computation of the nonlinear susceptibility. *Phys. Rev. B*, 50:5756–5759, Aug 1994. doi: 10.1103/PhysRevB.50.5756. URL `https://link.aps.org/doi/10.1103/PhysRevB.50.5756`.

[64] Andrea Dal Corso, Francesco Mauri, and Angel Rubio. Density-functional theory of the nonlinear optical susceptibility: Application to cubic semiconductors. *Phys. Rev. B*, 53:15638–15642, Jun 1996. doi: 10.1103/PhysRevB.53.15638. URL `https://link.aps.org/doi/10.1103/PhysRevB.53.15638`.

[65] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.*, 6(1):15–50, 1996. ISSN 0927-0256. doi: http://dx.doi.org/10.1016/0927-0256(96)00008-0. URL `http://www.sciencedirect.com/science/article/pii/0927025696000080`.

[66] X. Gonze, F. Jollet, F. Abreu Araujo, D. Adams, B. Amadon, T. Applencourt, C. Audouze, J.-M. Beuken, J. Bieder, A. Bokhanchuk, E. Bousquet, F. Bruneval, D. Caliste, M. Côté, F. Dahm, F. Da Pieve, M. Delaveau, M. Di Gennaro, B. Dorado, C. Espejo, G. Geneste, L. Genovese, A. Gerossier, M. Giantomassi, Y. Gillet, D.R. Hamann, L. He, G. Jomard, J. Laflamme Janssen, S. Le Roux, A. Levitt, A. Lherbier, F. Liu, I. Lukaevi, A. Martin, C. Martins, M.J.T. Oliveira, S. Ponc, Y. Pouillon, T. Rangel, G.-M. Rignanese, A.H. Romero, B. Rousseau, O. Rubel, A.A. Shukri, M. Stankovski, M. Torrent, M.J. Van Setten, B. Van Troeye, M.J. Verstraete, D. Waroquiers, J. Wiktor, B. Xu, A. Zhou, and J.W. Zwanziger. Recent developments in the ABINIT software package. *Comput. Phys. Commun.*, 205:106–131, August 2016. ISSN 0010-4655. doi: 10.1016/j.cpc.2016.04.003. URL `https://doi.org/10.1016/j.cpc.2016.04.003`.

[67] P Giannozzi, O Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, N Colonna, I Carnimeo, A Dal Corso, S de Gironcoli, P Delugas, R A DiStasio, A Ferretti, A Floris, G Fratesi, G Fugallo, R Gebauer, U Gerstmann, F Giustino, T Gorni, J Jia, M Kawamura, H-Y Ko, A Kokalj, E Kçkbenli, M Lazzeri, M Marsili, N Marzari, F Mauri, N L Nguyen, H-V

Nguyen, A Otero de-la Roza, L Paulatto, S Poncé, D Rocca, R Sabatini, B Santra, M Schlipf, A P Seitsonen, A Smogunov, I Timrov, T Thonhauser, P Umari, N Vast, X Wu, and S Baroni. Advanced capabilities for materials modelling with quantum ESPRESSO. *Journal of Physics: Condensed Matter*, 29(46):465901, oct 2017. doi: 10.1088/1361-648x/aa8f79. URL https://doi.org/10.10882F1361-648x2Faa8f79.

[68] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni. Ab initio calculation of phonon dispersions in semiconductors. *Phys. Rev. B*, 43:7231–7242, 1991. doi: 10.1103/PhysRevB.43.7231. URL http://link.aps.org/doi/10.1103/PhysRevB.43.7231.

[69] M. Veithen, X. Gonze, and Ph. Ghosez. Nonlinear optical susceptibilities, raman efficiencies, and electro-optic tensors from first-principles density functional perturbation theory. *Phys. Rev. B*, 71:125107, Mar 2005. doi: 10.1103/PhysRevB.71.125107. URL https://link.aps.org/doi/10.1103/PhysRevB.71.125107.

[70] James L. P. Hughes and J. E. Sipe. Calculation of second-order optical response in semiconductors. *Phys. Rev. B*, 53:10751–10763, Apr 1996. doi: 10.1103/PhysRevB.53.10751. URL https://link.aps.org/doi/10.1103/PhysRevB.53.10751.

[71] Eleonora Luppi, Hannes Hübener, and Valérie Véniard. Ab initio second-order nonlinear optics in solids: Second-harmonic generation spectroscopy from time-dependent density-functional theory. *Phys. Rev. B*, 82:235201, Dec 2010. doi: 10.1103/PhysRevB.82.235201. URL https://link.aps.org/doi/10.1103/PhysRevB.82.235201.

[72] Guido Petretto, Xavier Gonze, Geoffroy Hautier, and Gian-Marco Rignanese. Convergence and pitfalls of density functional perturbation theory phonons calculations from a high-throughput perspective. *Computational Materials Science*, 144:331 – 337, 2018. ISSN 0927-0256. doi: https://doi.org/10.1016/j.commatsci.2017.12.040. URL `http://www.sciencedirect.com/science/article/pii/S0927025617307243`.

[73] Guido Petretto, Shyam Dwaraknath, Henrique P.C. Miranda, Donald Winston, Matteo Giantomassi, Michiel J. van Setten, Xavier Gonze Kristin A. Persson, Geoffroy Hautier, and Gian-Marco Rignanese. High-throughput density-functional perturbation theory phonons for inorganic materials. *Scientific Data*, 5:180065, 2018. doi: https://doi.org/10.1038/sdata.2018.65.

[74] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson. The materials project: A materials genome approach to accelerating materials innovation. *APL Mater.*, 1(1):011002, 2013. ISSN 2166532X. doi: 10.1063/1.4812323. URL `http://link.aip.org/link/AMPADS/v1/i1/p011002/s1&Agg=doi`.

[75] S. P. Ong, W. Lei, B. Kang, and G. Ceder. $LiFePO_2$ phase diagram from first principles calculations. *Chem. Mater.*, 20(5):1798–1807, 2008. doi: 10.1021/cm702327g. URL `http://dx.doi.org/10.1021/cm702327g`.

[76] H. Chen, G. Hautier, A. Jain, C. Moore, K. Byoungwoo, R. Doe, L. Wu, Y. Zhu, Y. Tang, and G. Ceder. Carbonophosphates: A new family of cathode materials for li-ion batteries identified computationally. *Chem. Mater.*, 24 (11):2009–2016, 2012. doi: 10.1021/cm203243x. URL `http://dx.doi.org/10.1021/cm203243x`.

[77] Geoffroy Hautier, Shyue Ping Ong, Anubhav Jain, Charles J. Moore, and Gerbrand Ceder. Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability. *Phys. Rev. B*, 85(15):155208, April 2012. ISSN 1098-0121, 1550-235X. doi: 10.1103/physrevb.85.155208. URL `http://dx.doi.org/10.1103/physrevb.85.155208`.

[78] P. A. Cox. *Transition Metal Oxides: An Introduction to Their Electronic Structure and Properties.* Oxford, 2010.

[79] M. Gajdoš, K. Hummer, G. Kresse, J. Furthmüller, and F. Bechstedt. Linear optical properties in the projector-augmented wave methodology. *Phys. Rev. B*, 73:045112, 2006. doi: 10.1103/PhysRevB.73.045112.

[80] A. Jain, G. Hautier, S. P. Ong, C. J. Moore, C. C. Fischer, K. A. Persson, and G. Ceder. Accurate formation enthalpies by mixing GGA and GGA+U calculations. *Phys. Rev. B*, 84: 045115, 2011. doi: 10.1103/PhysRevB.84.045115.

[81] Francesco Ricci, Wei Chen, Umut Aydemir, G. Jeffrey Snyder, Gian-Marco Rignanese, Anubhav Jain, and Geoffroy Hautier. An ab initio electronic transport database for inorganic materials. *Sci. Data*, 4:170085, 07 2017. URL `http://dx.doi.org/10.1038/sdata.2017.85`.

[82] Luis A. Agapito, Stefano Curtarolo, and Marco Buongiorno Nardelli. Reformulation of DFT+U as a pseudohybrid hubbard density functional for accelerated materials discovery. *Phys. Rev. X*, 5(1):011006, January 2015. ISSN 2160-3308. doi: 10.1103/physrevx.5.011006. URL http://dx.doi.org/10.1103/physrevx.5.011006.

[83] M. K. Y. Chan and G. Ceder. Efficient band gap prediction for solids. *Phys. Rev. Lett.*, 105(19):196403, November 2010. ISSN 0031-9007, 1079-7114. doi: 10.1103/physrevlett.105.196403. URL http://dx.doi.org/10.1103/physrevlett.105.196403.

[84] Mongodb. https://www.mongodb.com/.

[85] Abinit. https://www.abinit.org/.

[86] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048–5079, May 1981. doi: 10.1103/PhysRevB.23.5048. URL https://link.aps.org/doi/10.1103/PhysRevB.23.5048.

[87] M.J. van Setten, M. Giantomassi, E. Bousquet, M.J. Verstraete, D.R. Hamann, X. Gonze, and G.-M. Rignanese. The pseudodojo: Training and grading a 85 element optimized norm-conserving pseudopotential table. *Computer Physics Communications*, 226:39 – 54, 2018. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2018.01.012. URL http://www.sciencedirect.com/science/article/pii/S0010465518300250.

[88] Kanghoon Yim, Youn Yong, Joohee Lee, Kyuhyun Lee, Ho-Hyun Nahm, Jiho Yoo, Chanhee Lee, Cheol Seong Hwang,

and Seungwu Han. Novel high-κ dielectrics for next-generation electronic devices screened by automated ab initio calculations. *NPG Asia Mater.*, 7:e190, 2015. URL http://dx.doi.org/10.1038/am.2015.57.

[89] G. F. Bassani and G. P. Parravicini. *Electronic states and optical transitions in solids*. Pergamon Press, 1975.

[90] S.K. Tripathy. Refractive indices of semiconductors from energy gaps. *Opt. Mater.*, 46:240, 2015.

[91] X. Gonze. First-principles responses of solids to atomic displacements and homogeneous electric fields: Implementation of a conjugate-gradient algorithm. *Phys. Rev. B*, 55: 10337–10354, 1997. doi: 10.1103/PhysRevB.55.10337. URL http://link.aps.org/doi/10.1103/PhysRevB.55.10337.

[92] X. Gonze and C. Lee. Dynamical matrices, born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B*, 55:10355–10368, 1997. doi: 10.1103/PhysRevB.55.10355. URL http://link.aps.org/doi/10.1103/PhysRevB.55.10355.

[93] Stephen L. Adler. Quantum theory of the dielectric constant in real solids. *Phys. Rev.*, 126:413–420, Apr 1962. doi: 10.1103/PhysRev.126.413. URL https://link.aps.org/doi/10.1103/PhysRev.126.413.

[94] Nathan Wiser. Dielectric constant with local field effects included. *Phys. Rev.*, 129:62–69, Jan 1963. doi: 10.1103/PhysRev.129.62. URL https://link.aps.org/doi/10.1103/PhysRev.129.62.

[95] A.M. Fox. *Optical Properties of Solids*. Oxford master series in condensed matter physics. Oxford University Press, 2001. ISBN 9780198506126. URL `https://books.google.be/books?id=-5bVBbAoaGoC`.

[96] Geoffroy Hautier, Anna Miglio, David Waroquiers, Gian-Marco Rignanese, and Xavier Gonze. How does chemistry influence electron effective mass in oxides? a high-throughput computational analysis. *Chem. Mater.*, 26(19):5447–5458, 2014.

[97] Christopher C. Evans, Jonathan D. B. Bradley, Erwin A. Mart-Panameo, and Eric Mazur. Mixed two- and three-photon absorption in bulk rutile (TiO_2) around 800 nm. *Opt. Express*, 20(3):3118, January 2012. ISSN 1094-4087. doi: 10.1364/oe.20.003118. URL `http://dx.doi.org/10.1364/oe.20.003118`.

[98] Christopher C. Evans, Katia Shtyrkova, Jonathan D. B. Bradley, Orad Reshef, Erich Ippen, and Eric Mazur. Spectral broadening in anatase titanium dioxide waveguides at telecommunication and near-visible wavelengths. *Opt. Express*, 21(15):18582, July 2013. ISSN 1094-4087. doi: 10.1364/oe.21.018582. URL `http://dx.doi.org/10.1364/oe.21.018582`.

[99] Christopher C. Evans, Katia Shtyrkova, Orad Reshef, Michael Moebius, Jonathan D. B. Bradley, Sarah Griesse-Nascimento, Erich Ippen, and Eric Mazur. Multimode phase-matched third-harmonic generation in sub-micrometer-wide anatase TiO_2 waveguides. *Opt. Express*, 23(6):7832, March 2015. ISSN 1094-4087. doi: 10.1364/oe.23.007832. URL `http://dx.doi.org/10.1364/oe.23.007832`.

[100] David Waroquiers, Xavier Gonze, Gian-Marco Rignanese, Cathrin Welker-Nieuwoudt, Frank Rosowski, Michael Göbel, Stephan Schenk, Peter Degelmann, Rute André, Robert Glaum, and Geoffroy Hautier. Statistical analysis of coordination environments in oxides. *Chem. Mater.*, 29(19):8346–8360, September 2017. ISSN 0897-4756, 1520-5002. doi: 10.1021/acs.chemmater.7b02766. URL `http://dx.doi.org/10.1021/acs.chemmater.7b02766`.

[101] David W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3):605–610, 1979. doi: 10.1093/biomet/66.3.605. URL `+http://dx.doi.org/10.1093/biomet/66.3.605`.

[102] Christopher C. Evans, Chengyu Liu, and Jin Suntivich. Low-loss titanium dioxide waveguides and resonators using a dielectric lift-off fabrication process. *Opt. Express*, 23(9):11160–11169, May 2015. doi: 10.1364/OE.23.011160. URL `http://www.opticsexpress.org/abstract.cfm?URI=oe-23-9-11160`.

[103] Christopher C. Evans, Chengyu Liu, and Jin Suntivich. TiO_2 nanophotonic sensors for efficient integrated evanescent raman spectroscopy. *ACS Photonics*, 3(9):1662–1669, 2016. doi: 10.1021/acsphotonics.6b00314. URL `https://doi.org/10.1021/acsphotonics.6b00314`.

[104] J. Zaanen, G. A. Sawatzky, and J. W. Allen. Band gaps and electronic structure of transition-metal compounds. *Phys. Rev. Lett.*, 55:418–421, 1985. doi: 10.1103/PhysRevLett.55.418. URL `https://link.aps.org/doi/10.1103/PhysRevLett.55.418`.

[105] Walter A Harrison. *Physical Properties Of Crystals: Their Representation by Tensors and Matrices*. Courier Corporation, 2012.

[106] M. M. Abdullah, Fahd M. Rajab, and Saleh M. Al-Abbas. Structural and optical characterization of Cr_2O_3 nanostructures: Evaluation of its dielectric properties. *AIP Advances*, 4(2):027121, 2014. doi: 10.1063/1.4867012.

[107] J. M. Senior. *Optical Fiber Communications Principles and Practice*. Pearson Education, 2009.

[108] R. Bazzi, A. Brenier, P. Perriat, and O. Tillement. Optical properties of neodymium oxides at the nanometer scale. *J. Lumin.*, 113(1):161–167, 2005. ISSN 0022-2313. doi: https://doi.org/10.1016/j.jlumin.2004.09.120. URL `http://www.sciencedirect.com/science/article/pii/S0022231304004387`.

[109] A.G. Jackson, M.C. Ohmer, and S.R. LeClair. Relationship of the second order nonlinear optical coefficient to energy gap in inorganic non-centrosymmetric crystals. *Infrared Physics & Technology*, 38:233244, 06 1997. doi: 10.1016/S1350-4495(97)00017-0.

[110] B. F. Levine. Electrodynamical bond-charge calculation of nonlinear optical susceptibilities. *Phys. Rev. Lett.*, 22:787–790, Apr 1969. doi: 10.1103/PhysRevLett.22.787. URL `https://link.aps.org/doi/10.1103/PhysRevLett.22.787`.

[111] M. I. Bell. Frequency dependence of Miller's rule for nonlinear susceptibilities. *Phys. Rev. B*, 6:516–521, Jul 1972. doi: 10.

1103/PhysRevB.6.516. URL https://link.aps.org/doi/10.1103/PhysRevB.6.516.

[112] J. C. Phillips. Covalent bond in crystals. II. partially ionic binding. *Phys. Rev.*, 168:905–911, Apr 1968. doi: 10.1103/PhysRev.168.905. URL https://link.aps.org/doi/10.1103/PhysRev.168.905.

[113] E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P. F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. McBrien, and D. E. Bossi. A review of lithium niobate modulators for fiber-optic communications systems. *IEEE Journal of Selected Topics in Quantum Electronics*, 6 (1):69–82, Jan 2000. ISSN 1077-260X. doi: 10.1109/2944.826874.

[114] Lambertus Hesselink, Sergei S. Orlov, Alice Liu, Annapoorna Akella, David Lande, and Ratnakar R. Neurgaonkar. Photorefractive materials for nonvolatile volume holographic data storage. *Science*, 282(5391):1089–1094, 1998. ISSN 0036-8075. doi: 10.1126/science.282.5391.1089. URL https://science.sciencemag.org/content/282/5391/1089.

[115] M. Veithen and Ph. Ghosez. First-principles study of the dielectric and dynamical properties of lithium niobate. *Phys. Rev. B*, 65:214302, May 2002. doi: 10.1103/PhysRevB.65.214302. URL https://link.aps.org/doi/10.1103/PhysRevB.65.214302.

[116] J. M. Léger, J. Haines, C. Chateau, G. Bocquillon, M. W. Schmidt, S. Hull, F. Gorelli, A. Lesauze, and R. Marchand. Phosphorus oxynitride PON, a silica analogue: structure and

compression of the cristobalite-like phase; P–T phase diagram. *Physics and Chemistry of Minerals*, 28(6):388–398, Jul 2001. ISSN 1432-2021. doi: 10.1007/s002690100161. URL https://doi.org/10.1007/s002690100161.

[117] Dominik Baumann, Robin Niklaus, and Wolfgang Schnick. A high-pressure polymorph of phosphorus oxonitride with the coesite structure. *Angewandte Chemie International Edition*, 54(14):4388–4391, 2015. doi: 10.1002/anie.201410526. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201410526.

[118] Dominik Baumann, Stefan J. Sedlmaier, and Wolfgang Schnick. An unprecedented AB_2 tetrahedra network structure type in a high-pressure phase of phosphorus oxonitride (PON). *Angewandte Chemie International Edition*, 51(19):4707–4709, 2012. doi: 10.1002/anie.201200811. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201200811.

[119] Mariette Hellenbrandt. The Inorganic Crystal Structure Database (ICSD) – Present and future. *Crystallography Reviews*, 10(1):17–22, 2004. doi: 10.1080/08893110410001664882. URL https://doi.org/10.1080/08893110410001664882.

[120] S. Sharma and C. Ambrosch-Draxl. Second-harmonic optical response from first principles. *Physica Scripta*, T109:128, 2004. doi: 10.1238/physica.topical.109a00128. URL https://doi.org/10.12382Fphysica.topical.109a00128.

[121] M. Rohlfing and S. G. Louie. Electron-hole excitations and optical spectra from first principles. *Phys. Rev. B*, 62:4927–4944, 2000. doi: 10.1103/PhysRevB.62.4927. URL `http://link.aps.org/doi/10.1103/PhysRevB.62.4927`.