Density functional studies of the electronic structure and transport properties of thermoelectric materials

Espen Flage-Larsen

Thesis submitted in partial fulfillment
of the requirements for the degree of
Philosophiae Doctor

Department of Physics
University of Oslo

August 2009

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 906

ISSN 1501-7710

Cover: Inger Sandved Anfinsen.
Printed in Norway: AiT e-dit AS, Oslo, 2009.

Produced in co-operation with Unipub AS.
The thesis is produced by Unipub AS merely in connection with the
thesis defence. Kindly direct all inquiries regarding the thesis to the copyright
holder or the unit which grants the doctorate.

Unipub AS is owned by
The University Foundation for Student Life (SiO)

Summary

The present work was set to study materials, and in particular thermoelectric materials based on first–principle calculations. Density functional theory was used to calculate the electronic structure of skutterudites, Zintl compounds and black phosphorus. Transport properties can be obtained experimentally, but routines for calculating these from first–principles have not been fully developed. Thus, much effort was put into the development, implementation and testing of routines that calculate transport properties based on the electronic structure. Electron transfer analyses were also performed for selected skutterudites and Zintl compounds to explicitly investigate their bonds. The combination of experimental and theoretical studies is mutually beneficial and wherever possible experimental data have been included to verify or further investigate the electronic structure and transport properties.

Acknowledgements

I would like to express gratitude to my supervisors, Johan Taftø and Ole Martin Løvvik for their encouragement, scientific knowledge, enthusiasm and kindness during my three years of work on this thesis. I also appreciate countless hours of discussions with Øystein Prytz, Ole Bjørn Karlsen and Simone Casolo, both on a scientific and personal level. In addition, Spyros Diplas, Terje Finstad, Kjetil Valset, Michael Böttger, Lars–Olav Vestland and Protima Rauwel deserve thanks for their scientific interest.

Georg Kresse, Judith Harl and the rest of the people in the VASP group in Vienna should be thanked for their knowledge, support and kindness during my visits.

During my stay at Caltech, Jeff Snyder, Eric Toberer, Andrew May, Yanzhong Pei and Ali Saramat should be acknowledged for excellent hospitality and great discussions.

I would also like to express my deepest gratitude and love to my family for their everlasting support and understanding. They early inspired the need of curiosity and dedication in everyday life, factors I strongly feel have been major contributors to this thesis. And aside from the scientific aspects, their personal support has always been unsurpassed.

Finally, I would like to thank the Norwegian Research Council for financial support and the NOTUR project for computational resources.

Espen Flage–Larsen, 2009

Preface

This thesis was developed for the partial fulfilment of the requirements for the degree of Philosophiae Doctor at the University of Oslo. All work was funded by the Norwegian Research council through the project "Studies of the electronic structure of materials at the nanoscale". The work started autumn 2006 and ended three years later.

The title of the thesis is "Density functional studies of the electronic structure and transport properties of thermoelectric materials", a title that mainly describes its contents. However, even though this work is biased towards theory, focus incorporating experiments has been important for the published work. The author is a strong believer of the importance of an increased collaboration between experiment and theory.

Several visits to the VASP group in Vienna were done, where we worked out and investigated different ways of calculating the transport properties based on the VASP density functional engine.

Spring 2009 was spent at Jeff Snyder's thermoelectrics group at California Institute of Technology, where the author assisted Jeff Snyder, Eric Toberer and Andrew May with density functional and first–principle transport calculations.

Table of Contents

Chapter 1

Introduction

Conflicts related both directly and indirectly to energy demands are escalating all over the world. A primary concern for developing nations is to obtain long–term sustainable energy resources with as little investment as possible. On the other hand, OECD nations are constantly seeking continuity and increased harvest of established energy resources. Naturally, due to the shared resources, this leads, and will lead to local and global conflicts. Environmental concerns are also becoming more important due to increased emissions from fossil fuels, nuclear waste and local havoc of natural habitats of animals and plants.

Today's global energy consumption is estimated to be close to 500 quadrillion BTU (British Thermal Unit) (see Fig. 1.1). Extrapolation surveys from Energy Information Administration (EIA) [1] show that the world's total energy demands are expected to soar by 44% by 2030 (see Fig. 1.1), much due to economic growth in the developing nations. The OECD nations account for the largest share of today's energy consumption. However, as Fig. 1.1 indicates, this is about to change due to the rapid growth of energy demand in the developing nations, in particular China.

The dominating energy resource is still expected to be liquid (fossil fuels and bio fuels) and it is increasing steadily each year. Coal and renewable energy resources increase the most. These surveys are based on a relatively high fuel cost and thus any unforeseen change would likely yield a different distribution among the energy resources. Even though the transport sector is highly dependent on liquid fuels, we see from Fig. 1.2 that the future increase of energy consumption is mostly due to demand for electricity, where the largest contributors are the Asian nations. It is also interesting to note the projected distribution among the renewable energy resources in Fig. 1.2, where the hydroelectricity now dominates, surpassed in 2030 by the rapid increase of wind energy. The share of the other renewable energy resources are almost an order of magnitude

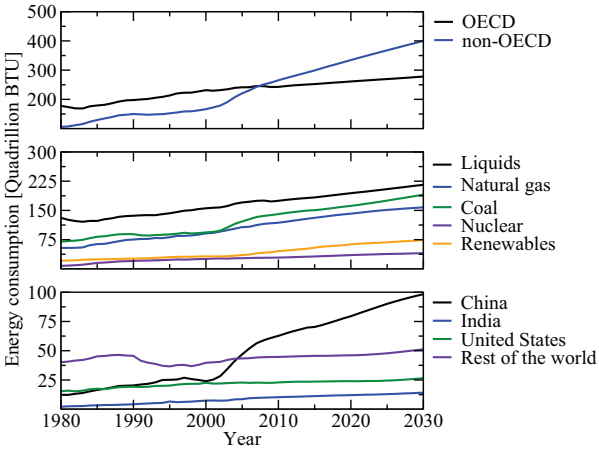


Figure 1.1: The world's energy consumption from 1980 to 2030. Top figure shows OECD and non–OECD energy consumption. The total energy consumption is split into different energy resources in the middle figure, while the lower figure shows a more detailed comparison of the nations' energy consumption. Data from International Energy Outlook 2009, EIA [1].

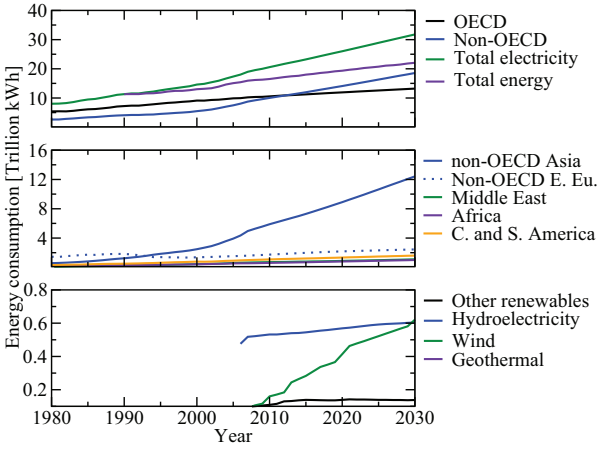


Figure 1.2: The world's net electricity generation from 1980 to 2030. Top figure compares OECD, non–OECD electricity generation and the total energy consumption to the total electricity generation. Middle figure splits the electricity generation into different continents, while the lower figure gives estimates of the OECD Europe nations' renewable electricity generation. Data from International Energy Outlook 2009, EIA [1]

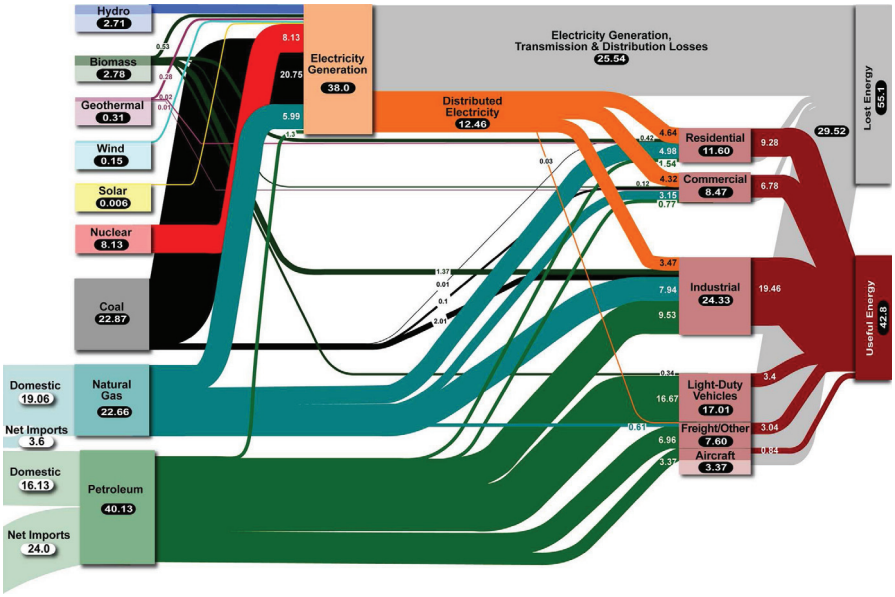


Figure 1.3: United States electricity generation, transmission and energy loss. Figure from Lawrence Livermore National Laboratory [2].

smaller.

Extensive research efforts have been put into traditional energy fields like nuclear, thermal and water energy. In parallel there has been inclining interest in alternative resources, such as solar, wind and ocean wave energy. Most of these energy fields are based on first stage energy exploitation, where the base energy is available in nature. However, second stage energy exploitation is also possible by collecting waste heat from a primary process and then convert this to a given form of distributable energy. The energy loss of different processes is illustrated in Fig. 1.3, where it is clear that over 50% of the total energy harvest potential is lost to waste heat. The potential to exploit heat from factories, power plants, cars, tumble driers, electronic circuits and so forth is thus astonishingly large.

Thermoelectricity is a particular interesting solid–state alternative for converting heat to electricity. Although it can be used on a primary energy source, it is thought to be even more useful in a secondary stage to collect waste heat. Without the need of moving parts, the longevity and adaptability are expected to be excellent. An additional benefit is flexibility, due to the dual nature of the thermoelectric device, which by reversing the voltage can operate both as a heater and a cooler. Simultaneously, waste heat conversion can be exploited in the same device. Unfor-

tunately, the thermoelectric materials used in an assembled thermoelectric device are currently too inefficient to compete with traditional heat conversion cycles. However, comprehensive improvements are expected once a more complete understanding of the physics and chemistry of these materials is achieved.

Motivated by this, we in this work, try to develop a more solid footing for first–principle calculations of materials and in particular, thermoelectric materials. We have tried to keep this work consistent and in close collaboration with experiments and believe this is crucial to obtain breakthroughs in this field. Readers are encouraged to consult the bibliography for more extensive background information.

Bibliography

Thermoelectrics Handbook, edited by D. M. Rowe, CRC, Boca Raton, 2006
Solid State Physics, N. W. Ashcroft and N. D. Mermin, Brooks Cole, 1976
Fundamentals of Semiconductors, P. Y. Yu and M. Cardona, Third edition, Springer-Verlag, Berlin, 2001
Introduction to Solid State Physics, C. Kittel, Seventh edition, John Wiley & Sons, 1996
Density-Functional Theory of Atoms and Molecules, R. G. Parr and W. Yang, Oxford University Press, New York, 1989
Density Functional Theory, R. Dreizler and E. Gross, Plenum Press, New York, 1995
Density Functional Theory, A Practical Introduction, D. Sholl and J. A. Steckel, John Wiley & Sons, New Jersey, 2009

References

[1] International Energy Outlook 2009, `http://eia.doe.gov`.

[2] Electricity Generation, Transmission, and Distribution Losses, `https://eed.llnl.gov/flow/`.

Chapter 2

Thermoelectricity

Simply put, thermoelectricity relates thermal gradients and electrical currents in materials. Through this relation it is possible to convert heat to electricity or use electricity for heating or cooling.

The basic principles of thermoelectricity was discovered two centuries ago, however, due to the lack of understanding of the physical properties of materials it was not until later that significant progress was done. During the 60's and 70's considerable theoretical and experimental work was performed, but it was not before the energy debate was brought to the public during the 90's that the field experienced a rapid expansion of research activities.

Even though basic thermoelectric devices have been operating for decades, serving remote lighthouses, unmanned spacecraft and local cooling, their current efficiency is too mediocre to be a viable alternative for large scale heat conversion applications. For example, current state of the art bulk materials have a figure–of–merit, the factor that determines their efficiency, that is three to four times too low to compete with traditional refrigerator solutions [1]. However it is believed that there is a large potential for optimization.

The main reason for thermoelectric research is due to its unique range of applicability, but in addition, there is a wealth of exciting general physical phenomena governing these devices and in particular the materials that are not yet fully understood. Hence, they also serve as excellent platforms for solid–state research.

In this chapter we briefly cover the history, the basics and the future of thermoelectricity.

2.1 Basic principles

2.1.1 Seebeck effect

What we today call the Seebeck effect was discovered in 1823 by Thomas Johann Seebeck. He placed a compass needle close to two dissimilar metals in contact. While heating one of the conductors the compass needle deflected and Seebeck correctly thought this was caused by magnetic fields generated by the temperature difference. However, he quickly discovered through Ampere's law that it was a current that generated the magnetic field. He concluded that this current was generated by the temperature difference between the two conductors. The Seebeck coefficient α was then defined as the ratio between the induced voltage V and the temperature difference ΔT;

$$\alpha = \frac{V}{\Delta T}. \tag{2.1}$$

2.1.2 Peltier effect

Based on the discoveries of Seebeck, the watchmaker Jean Charles Athanase Peltier detected a complementary effect 12 years later. During an experiment he observed local heating when a current was passed between two different conductors. He failed to explain the nature of this, but used Seebeck's discoveries to try to explain the phenomenon. It was not until 1838 that Lenz discovered that heating and cooling in the junction between the two dissimilar metals was dependent on the current direction. Similar to the Seebeck coefficient, the Peltier coefficient Π was defined as the ratio between power generated at the metal junction Q and the passing current I sent through the circuit;

$$\Pi = \frac{Q}{I}. \tag{2.2}$$

2.1.3 Thomson effect

Additional insight was given when William Thomson (Lord Kelvin) discovered a relation between the Peltier and the Seebeck effect. He deduced that

$$\Pi = \alpha T, \tag{2.3}$$

where T is the temperature. In light of this he predicted a third thermoelectric effect, the Thomson effect, which relates the heating and cooling of a conductor in the presence of a temperature

gradient.

2.1.4 Figure of merit

The figure–of–merit is considered the "goodness" factor of a thermoelectric material and its dimensionless definition is

$$zT = \frac{\alpha^2 \sigma}{\kappa} T, \tag{2.4}$$

where σ and κ are the thermal and electrical conductivity, respectively. From this relation it is apparent that to obtain a large zT value, we would need a large Seebeck coefficient, large electrical conductivity and at the same time a low thermal conductivity. Beside mechanical properties and other manufacturing aspects, the most important property of a thermoelectric material is this zT value. It is common nomenclature to use a large Z for the device and a small z for the material. Due to the importance of zT we will in Sec. 3.4 further discuss its components in light of different electronic structures.

2.2 Thermoelectric device

To put thermoelectricity to practice a thermoelectric device is needed. Such device consists of small thermocouples made of thermoelectric materials, preferably materials with the highest possible zT in the operational temperature range. In addition, there are other factors limiting the choice of materials, like durability and mechanical issues. For historical reasons these devices are often called "Peltier elements". The device is usually constructed of alternating n– and p–type materials, electrically connected in series, while thermally in parallel. In Fig. 2.1 a sketch of a simple one layer thermoelectric device is shown. These devices work on an external load and generate electricity if a thermal gradient normal to the surface is present. The reverse effect is possible if a current is passed through the device, thus generating a thermal gradient. This gradient can then be utilized for heating or cooling.

2.3 Current and future status

2.3.1 Typical materials

To fabricate an efficient thermoelectric device, efficient thermoelectric materials is a prerequisite. Such materials are not strictly metallic, semiconducting or insulating. As a rule of thumb,

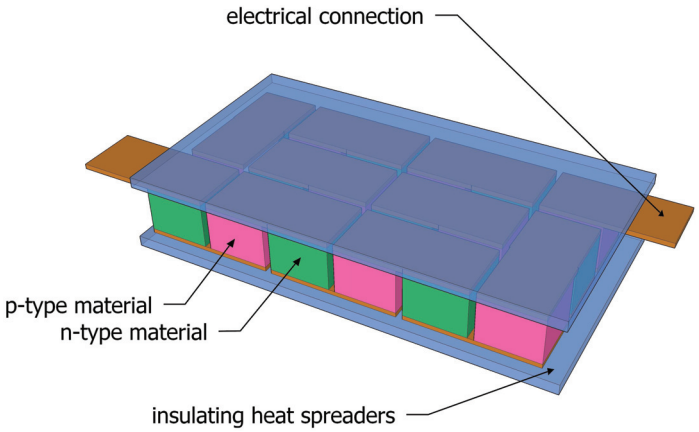


Figure 2.1: A typical (and simple) construction of alternating n– and p–type thermoelectric materials to make a thermoelectric device. The individual parts are connected electrically in series and thermally in parallel. Electrical connectors are soldered at the ends. Picture is made by Michael Böttger, Basic and Applied Thermoelectric Initiative (BATE) [2], 2009.

highly doped semiconductors are preferred based on an acceptable balance between the Seebeck coefficient and the electrical conductivity [3]. In Fig. 2.2 a simplified picture of the competing Seebeck coefficient and the electrical and thermal conductivity are shown. However, it needs to be stressed that this picture is based on simplified models [3]. We will later in this thesis, in Sec. 3.4 discuss band structure properties that give a high zT value (or more specifically its numerator). A comparison of the best bulk materials available today is given in Fig. 2.3, but we would like to emphasize that other promising bulk materials exist as well, such as the $A\mathrm{Zn_2Sb_2}$ compounds (A is rare earth, see Paper IV), the β–$\mathrm{Zn_4Sb_3}$ [4] compound and the $\mathrm{La_3Te_4}$ [5] compound to mention a few.

2.3.2 Current fields of application

The efficiency of existing materials is unfortunately too low for widespread applications. Compression units (refrigerators, steam engines etc.) are more efficient and it is necessary to significantly improve the zT value of the materials to make them a viable alternative for large scale applications. In general it is believed that a zT greater than three (currently, the best bulk materials have $zT \sim 1$, see Fig. 2.3) would be enough to compete with current refrigerator applications [1]. However, even though today's efficiency is low, there are areas of applications where thermoelectric devices are (and have been) favourable. This is due to the longevity, ruggedness

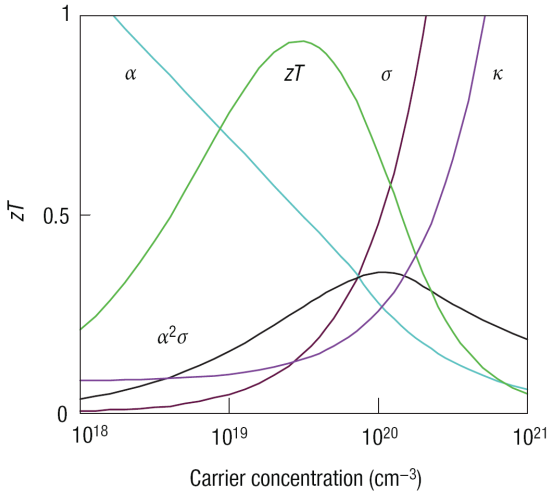


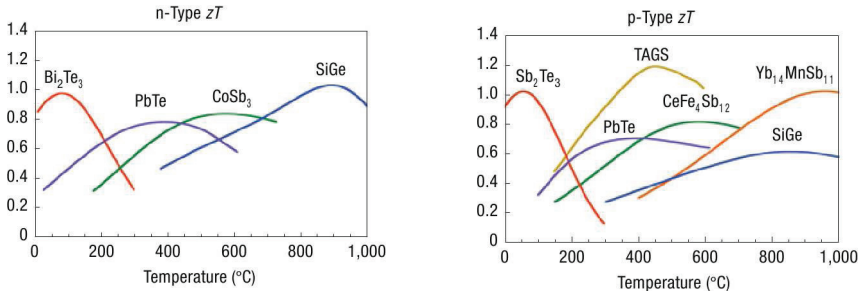
Figure 2.2: Simple overview of the competing processes that need to be overcome to obtain a large figure–of–merit. Figure from Ref. [3]. Here, zT is the figure–of–merit, α is the Seebeck coefficient, σ the electrical conductivity and κ the total thermal conductivity.

and low noise of the thermoelectric devices that are considered to be unsurpassed compared to the traditional solutions. Below we give three examples of areas where thermoelectric devices are currently used.

Radioisotope Thermoelectric Generator

The radioisotope thermoelectric generator (RTG) has been the best option for power generation in outer space devices, such as the Apollo missions, Viking Mars landers, Pioneer 10 and 11, and the Cassini spacecrafts to mention a few. The operational principle is simple, consisting of a radioactive core (usually made of Pu^{238}) encapsulated in a grid of thermoelectric devices. Due to the low temperature in outer space, a temperature gradient is easily maintained between the radioactive core and space. The simple construction quickly paid back in long–term reliability in hostile deep space environments.

The use of radioactive components demands extremely robust and high temperature capable thermoelectric materials, which unfortunately limits the choice of efficient materials. Lead telluride (PbTe), telluride antimonide germanium and silver (TAGS) and silicon germanium (SiGe) have been used on space missions [6]. A radioactive heater is illustrated in Fig. 2.4 which would



(a) n–type (b) p–type

Figure 2.3: Material figure–of–merit, zT as a function of temperature for the best n– and p–type bulk materials available today. Figures from Ref. [3].

be covered by thermoelectric modules and placed inside a space craft. Latest generations of SiGe based RTGs show 8% efficiency with a power–to–weight ratio of 5.1 W/kg [6].

In addition to outer space applications, such devices are also used to power remote light beacons and weather stations.

Temperature controlled seat elements

Increased demand for comfort in vehicles quickly introduced resistive heated seating into modern cars. This design is simple, but lacks the possibility of cooling in hot environments. To provide comfort to the occupants, the automotive industry implemented heat ventilation and air conditioning (HVAC) units. Such units control the air temperature of the whole volume of the vehicle and is thus rather inefficient, and occupants seat and back area are still not cooled to comfort temperatures. In addition, the cool–down or warm–up time for the HVAC systems is often tens of minutes, failing to give immediate comfort temperatures. Some automotive companies invented seating that circulated air from the HVAC system through the seating and back area, helping the comfort problem, but not the efficiency problem. To help remedy the efficiency problem, automotive companies started to investigate whether seat temperatures could be controlled by a thermoelectric device. Amerigon was the first company to introduce such seats back in 1999. Today, these seats (see Fig. 2.5) are standard equipment in many high–end vehicles. The operational principle is still based on air circulating through a perforated seating and back area, but the air is forced through an air–to–air thermoelectric cooler/heater, giving the occupants better individual control of temperature and comfort. In addition, by controlling the seat temperature, the HVAC can be downsized and a potential improvement in fuel economy is possible.

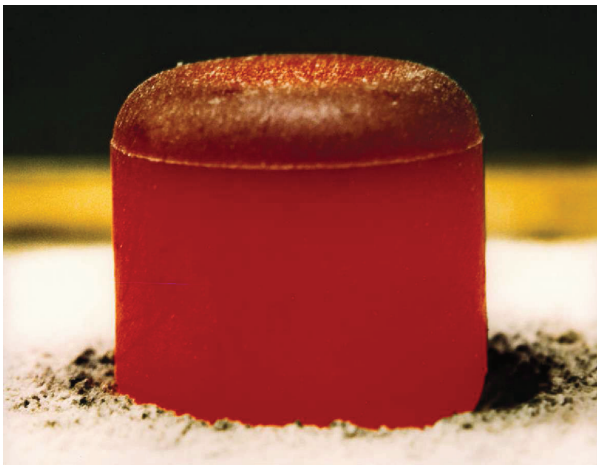


Figure 2.4: Glowing radioisotope thermoelectric generator core. The core is then encapsulated with thermoelectric devices and placed in the space craft. Picture from Jet Propulsion Labs (JPL), NASA, USA.

Other applications of thermoelectric devices

In addition to the previous examples, niche applications of thermoelectric devices are widespread, from simple portable coolers and localized electrical circuit cooling to sophisticated sensor temperature control. However, due to the low efficiency these solutions are not implemented on a large scale.

2.3.3 Future applications

If we manage to improve the efficiency and in addition reduce material cost and toxicity the potential field of application for thermoelectric devices is extremely large. We will now briefly list a few of the potential applications.

Combustion engine exhaust

The idea of retrieving some of the waste heat in the transport sector is relatively new. However, the potential is enormous, with a few tens of kilowatts of expelled heat from small vehicles [6]. The main idea is to use the thermogenerated electricity to replace the alternator and thus decrease

Figure 2.5: Climate controlled seating (CCS) from Amerigon. These seats use thermoelectric elements to heat or cool the air that is circulated through perforated leather areas in the cushion and seat back. Picture from Amerigon, USA.

the motor load to improve fuel economy. General Motors, BMW and Fiat have official programs running. Prototypes from General Motors have shown that it is possible to decrease the total fuel consumption by five percent [7]. Even though this number seems relatively small, there is room for improvements both on the device engineering and the type of materials used. One of the first prototypes made for a Chevrolet Suburban is shown in Fig. 2.6. Today they claim 750W of output power [7] during highway driving.

Other potential applications

Due to the current lack of efficiency, application studies of thermoelectric devices have been limited and their potential applications are only briefly mentioned in survey studies. Here we list some applications that might benefit from thermoelectric devices once their efficiency is improved;

- Three stage solar units. This idea is based on a three stage sandwich design made of solar cells followed by thermoelectric elements and then a cooling stage with circulating water. In this way, electricity is generated from both the solar cells and the thermoelectric elements, while the facility is heated by the heated circulated water.

- Refrigerator and air conditioner. Thermoelectric elements would increase longevity and reduce noise. Additional benefits are more compact design and better possibilities to tailor designs to specific needs.

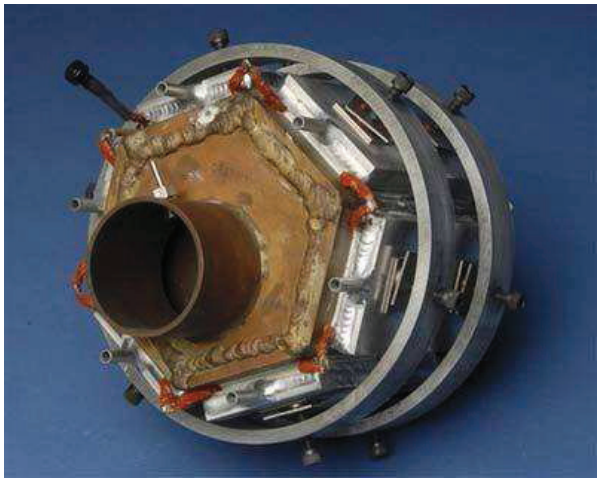


Figure 2.6: A ten–year old exhaust thermoelectric prototype from General Motors. Picture from General Motors, USA.

- Embedded heating and cooling. Localized cooling and heating are becoming increasingly important for the semi–conductor industry. Embedded designs would increase stability and operational temperature range. Electricity could potentially be harvested from local temperature gradients.
- Large scale generation of electricity from industrial waste heat.
- Small scale generation of heat and heat conversion in housings.
- Blanket shields for fission (and maybe in the future fusion) plants using thermoelectric devices to generate electricity from the expelled heat.
- Direct conversion of heat to electricity from furnaces operating a diverse selection of fuels.

References

[1] T. C. Harman, P. J. Taylor, M. P. Walsh, and B. E. LaForge. Quantum dot superlattice thermoelectric materials and devices. *Science*, 297:2229, 2002.

[2] Basic and Applied Thermoelectric Initiative (BATE), `http://www.fys.uio.no/bate`.

[3] G. J. Snyder and E. S. Toberer. Complex thermoelectric materials. *Nature Materials*, 7:105–114, 2008.

[4] G. J. Snyder, M. Christensen, E. Nishibori, T. Caillat, and B. Brummerstedt Iversen. Disordered zinc in Zn_4Sb_3 with phonon-glass and electron-crystal thermoelectric properties. *Nature Materials*, 3:458, 2004.

[5] A. F. May, D. J. Singh, and G. J. Snyder. Influence of band structure on the large thermoelectric performance of lanthanum telluride. *Phys. Rev. B*, 79:153101, 2009.

[6] J. Yang and T. Caillat. Thermoelectric materials for space and automotive power generation. *MRS Bulletin*, 31:224, 2006.

[7] Jihui Yang's talk at MS&T2008 conference in Pittsburgh, US.

Chapter 3

Density functional theory studies of thermoelectric materials

The core of this thesis is based on density functional theory studies of materials, in particular the thermoelectric branch. Such studies are beneficial since perfect structures can be investigated. This is of great importance due to the complexity of the thermoelectric materials. In addition, once the computational resources permit, large scale simulations can be performed to assist in the search for new and promising materials. However, at the same time, density functional studies are burden by numerous accuracy, performance and general implementation problems.

We will thus devote the first two sections of this chapter to explain the idea and basic principles behind density functional theory. At the same time we will touch important aspects that lead to common density functional failures. Then we will move on to a section that covers electron transfer analysis and its use to investigate bonding. Finally, in the last two sections we first develop a way to calculate the transport properties based on the semi–classical Boltzmann equation, and then investigate different band structures and their thermoelectric properties to highlight band structure requirements for good thermoelectric performance.

3.1 Ab–initio solutions of the Schrödinger equation

Basically all non–relativistic quantum mechanical problems boil down to the solution of the Schrödinger equation

$$\mathbf{H}|\Psi\rangle = E|\Psi\rangle, \tag{3.1}$$

where $\mathbf{H}$, $|\Psi\rangle$ and E are the Hamiltonian operator, the eigenvectors (or wavefunctions) and the eigenvalues (or energy levels), respectively. See Ref. [1] (Sec. 6.4) for Dirac notation. The complexity of this equation is well hidden. It can be solved in numerous ways depending on the wanted accuracy and possible ways of formulating the problem.

The exact (to numerical precision) diagonalization is the most direct (read brute force) method to solve Eq. (3.1). However, as we shall see, such a formalism is not feasible for materials (or extended systems in general). These limitations are connected to the accuracy and available computational resources. Nevertheless, it is important to understand its basics. We will now briefly go through the brute force way of solving the Schrödinger equation. From Eq. (3.1) we expand $|\Psi\rangle$ in the space set up by the basis vectors $|\beta\rangle$, such that

$$|\Psi\rangle = \sum_{\beta} \langle\beta|\Psi\rangle|\beta\rangle. \tag{3.2}$$

The eigenvectors $|\beta\rangle$, with quantum numbers β form a basis in the Hilbert space, see for example Ref. [1] (Sec. 6.4) for details. In the Hilbert space they are orthonormal, thus

$$\langle\beta'|\beta\rangle = \delta_{\beta'\beta}. \tag{3.3}$$

Inserting Eq. (3.2) into Eq. (3.1) gives

$$\sum_{\beta} \mathbf{H}|\beta\rangle\langle\beta|\Psi\rangle = \sum_{\beta} E|\beta\rangle\langle\beta|\Psi\rangle, \tag{3.4}$$

and multiplying from the left with $\langle\beta'|$ yields

$$\sum_{\beta} \langle\beta'|\mathbf{H}|\beta\rangle\langle\beta|\Psi\rangle = \sum_{\beta} E\delta_{\beta'\beta}\langle\beta|\Psi\rangle. \tag{3.5}$$

Defining the coefficient $\Psi_\beta = \langle\beta|\Psi\rangle$ gives

$$\sum_{\beta} \langle\beta'|\mathbf{H}|\beta\rangle\Psi_\beta = E\Psi_{\beta'}, \tag{3.6}$$

which can be written as

$$\tilde{\mathbf{H}}\vec{\Psi} = E\vec{\Psi}, \tag{3.7}$$

where $\tilde{\mathbf{H}} = \langle\beta'|\mathbf{H}|\beta\rangle$. To find the eigenvalues, we need to diagonalize the matrix $\tilde{\mathbf{H}}$. In principle, the Hilbert space is infinite, so we should expand all vectors to infinity. Fortunately one can

introduce a cutoff of the expansion. Such cutoffs are a crucial convergence parameter for exact diagonalization methods. The diagonalization is usually done by the use of orthogonal transformations. Common procedures for this are the QR algorithm with implicit shifts, Householder reduction and Lanczos algorithm. They are among the most effective methods for solving eigenproblems. The derivation of these methods is outlined in Ref. [2] (Sec. 11.3) and [3] (Sec. 1.1 for a short introduction). To solve the Schrödinger equation, a proper basis and Hamiltonian operator are needed. For many–particle interactions these are not trivial to formulate due to efficiency and accuracy difficulties. Even though most of these difficulties can be controlled, there is an additional fundamental problem with the exact diagonalization approach; the eigenvectors from Eq. (3.1) contain all degrees of freedom. Hence, for a n particle system

$$\langle x_1 x_2 \dots x_n | \Psi \rangle = \Psi(x_1, x_2, \dots, x_n), \tag{3.8}$$

where $x_i = \{\vec{r}_i, \vec{s}_i\}$, the spatial coordinate vector $\vec{r}_i$ and spin vector $\vec{s}_i$ for particle i. Such eigenstates would require a huge parameter space. For the spatial part, the parameter space scales as $3n$. It is not uncommon in (even not very complex) materials to handle several hundred particles. Hence, such a parameter space is unsolvable (regardless of method). We have to introduce approximations.

3.2 Density functional theory

The idea behind density functional theory (DFT) is to reduce the parameter space. This is done by posing the problem as a function of the particle density, not the spatial and spin coordinates. Instead of handling a spatial $3n$ parameter space, we now have a three dimensional parameter space (or six with spin degrees of freedom), which in principle is computationally constant in n. In this section we will briefly cover the basics of DFT and touch a few topics that are important for its practical applications.

3.2.1 Basics

The era of DFT started in 1964 when Hohenberg and Kohn published a paper about a possible way of solving the interacting electron gas system [4]. We will in this section partly follow this and the additional paper by Kohn and Sham [5]. Hohenberg and Kohn considered a set of enclosed particles moving in an average background potential $V_{\text{ext}}(\vec{r})$ and posed the following two theorems

Theorem 1. *The external potential $V_{\text{ext}}(\vec{r})$ is (to within a constant) a unique functional of the electron density $n(\vec{r})$; since, in turn, $V_{\text{ext}}(\vec{r})$ fixes the Hamiltonian* $\mathbf{H}$ *the many–particle ground state Ψ is a functional of $n(\vec{r})$.*

Proof. If we introduce a potential $V'_{\text{ext}}(\vec{r})$ with ground state Ψ' and assume this also yields the electron density $n(\vec{r})$ an inconsistency appears. This is based on the fact that Ψ and Ψ' should solve the same Schrödinger equation as long as the system is non–degenerate. Define $\mathbf{H}'$, E' to be associated with Ψ' such that $\mathbf{H}' - \mathbf{H} = \mathbf{V}' - \mathbf{V}$. We then have

$$\begin{aligned} E' &= \langle \Psi' | \mathbf{H}' | \Psi' \rangle < \langle \Psi | \mathbf{H}' | \Psi \rangle, \\ &< E + \langle \Psi | \mathbf{V}' - \mathbf{V} | \Psi \rangle, \end{aligned} \tag{3.9}$$

where

$$\mathbf{V}' = \int V'_{ext}(\vec{r}) \Psi^{\dagger}(\vec{r}) \Psi(\vec{r}) d\vec{r}. \tag{3.10}$$

Similarly

$$\begin{aligned} E &= \langle \Psi | \mathbf{H} | \Psi \rangle < \langle \Psi' | \mathbf{H} | \Psi' \rangle, \\ &< E' + \langle \Psi' | \mathbf{V} - \mathbf{V}' | \Psi' \rangle. \end{aligned} \tag{3.11}$$

If we now add Eq. (3.9) and Eq. (3.11) we get

$$E' + E < E' + E, \tag{3.12}$$

an obvious inconsistency. □

Theorem 2. *Given an electron density $n'(\vec{r})$ for a closed system of electrons where there is a well defined minimum*

$$E_0 \leq E[n'(\vec{r})]. \tag{3.13}$$

Only if $n'(\vec{r})$ equals the ground state electron density $n(\vec{r})$ is $E_0 = E$ valid.

Proof. See Ref. [4] and [5] for proof. □

Now, since the ground state depends on $n(\vec{r})$ it is possible to express an energy functional as

$$E[n(\vec{r})] = F[n(\vec{r})] + V_{\text{ext}}[n(\vec{r})], \tag{3.14}$$

where F and V_{ext} are the internal energy and external potential energy functionals (of the electron

density), respectively. The external potential energy depends on the system, while the functional F should in principle be independent of the system and handle all internal interactions. The determination of F and V_{ext} is not trivial due to their inherent many–particle properties.

One year later, in 1965, Kohn and Sham continued the work of Hohenberg and Kohn and came up with a solution to determine F and V_{ext} that is still commonly used today. Their approach is the Kohn–Sham equations which we will briefly discuss below.

To bring the many–particle formulation to a one particle form we first start by writing the F functional as

$$F[n(\vec{r})] = T[n(\vec{r})] + E_{\text{H}}[n(\vec{r})] + E_{\text{xc}}[n(\vec{r})], \tag{3.15}$$

where T and E_{H} are the kinetic and Hartree energy functionals, respectively. The exchange–correlation energy functional E_{xc} contains Fock terms (exchange) excluded from the Hartree energy functional and in addition correlation effects. To determine $n(\vec{r})$ we need to minimize Eq. (3.14). This can be done by the use of a Lagrange multiplier λ under the restriction

$$\frac{\delta}{\delta n}\left(\int n(\vec{r})d\vec{r}\right) = 0. \tag{3.16}$$

This restriction is valid since our system is in a stationary state. Using the Hartree term

$$E_{\text{H}}[n(\vec{r})] = \int\int \frac{n(\vec{r})n'(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}d\vec{r}d\vec{r}\,', \tag{3.17}$$

the external potential

$$V_{\text{ext}}[n(\vec{r})] = \int v_{\text{ext}}(\vec{r})n(\vec{r})d\vec{r}, \tag{3.18}$$

and the expression for F in Eq. (3.15), we can now minimize Eq. (3.14) by using

$$\frac{\delta}{\delta n}\left(E[n(\vec{r})] - \lambda\int n(\vec{r})d\vec{r}\right) = 0, \tag{3.19}$$

which gives by substitution

$$\frac{\delta T[n(\vec{r})]}{\delta n(\vec{r})} + \frac{\delta E_{\text{xc}}[n(\vec{r})]}{\delta n(\vec{r})} + v_{\text{ext}}(\vec{r}) + \int \frac{n'(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}d\vec{r}\,' = 0. \tag{3.20}$$

Let us now look at the similar functionals for a system of non–interacting single particles moving in a potential $v(\vec{r})$

$$E_{\text{s}}[n(\vec{r})] = T_{\text{s}}[n(\vec{r})] + V_{\text{s}}[n(\vec{r})], \tag{3.21}$$

where

$$V_{\rm s}[n(\vec{r})] = \int v(\vec{r})n(\vec{r})d\vec{r}, \tag{3.22}$$

and $T_{\rm s}$ is the kinetic energy functional. Minimizing Eq. (3.21) yields

$$\frac{\delta T_{\rm s}[n(\vec{r})]}{\delta n(\vec{r})} + v(\vec{r}) = 0, \tag{3.23}$$

for which the Schrödinger equation $\mathbf{H}_s\phi_i = \epsilon_i\phi_i$, with Hamiltonian $\mathbf{H}_s$ and eigenvalues ϵ_i is well known. The similarities between Eq. (3.20) and Eq. (3.23) are obvious if we take $T = T_s$ such that

$$T[n(\vec{r})] = -\int \phi_i^{\dagger}(\vec{r})\nabla^2\phi_i(\vec{r})d\vec{r} \tag{3.24}$$

and

$$v(\vec{r}) = v_{\rm ext}(\vec{r}) + \int \frac{n(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|}d\vec{r}\,' + \frac{\delta E_{\rm xc}[n(\vec{r})]}{\delta n(\vec{r})}, \tag{3.25}$$

and

$$n(\vec{r}) = \sum_i^N f_i|\phi_i(\vec{r})|^2, \tag{3.26}$$

for N particles, where f_i is the occupancy of the single particle orbital ϕ_i. Hence, to solve Eq. (3.20) (that is to find the electron density $n_0(\vec{r})$ that minimizes it), we only need to solve the single particle Schrödinger equation

$$\left(-\frac{1}{2}\nabla^2 + v(\vec{r})\right)\phi_i(\vec{r}) = \epsilon_i\phi_i(\vec{r}) \tag{3.27}$$

The total energy E_0 for the minimized electron density $n_0(\vec{r})$ can now be found by expressing $V_{\rm ext}(\vec{r})$ in terms of $v(\vec{r})$ through Eq. (3.18) and Eq. (3.25), such that

$$V_{\rm ext}[\vec{r}] = V_{\rm s}[n_0(\vec{r})] - 2V_{\rm H}[n_0(\vec{r})] - \frac{\delta E_{\rm xc}[n(\vec{r})]}{\delta n(\vec{r})}. \tag{3.28}$$

By substituting this back into Eq. (3.14), using Eq. (3.15) we get

$$E_0 = T_{\rm s}[n_0(\vec{r})] + V_{\rm s}[n_0(\vec{r})] + V_{\rm H}[n_0(\vec{r})] - \frac{\delta E_{\rm xc}[n_0(\vec{r})]}{\delta n_0(\vec{r})} - E_{\rm xc}[n_0(\vec{r})]. \tag{3.29}$$

The first two terms are the energy functionals for the single particle Schrödinger equation, such

that

$$E_0 = \sum_i \epsilon_i + V_{\rm H}[n(\vec{r})] - \frac{\delta E_{\rm xc}[n_0(\vec{r})]}{\delta n_0(\vec{r})} - E_{\rm xc}[n_0(\vec{r})], \tag{3.30}$$

Due to the fact that the change in energy with respect to the electron density is precisely the chemical potential μ, it is customary to define

$$\frac{\delta E_{\rm xc}[n_0(\vec{r})]}{\delta n_0(\vec{r})} \equiv \mu_{\rm xc} \tag{3.31}$$

The electron density $n(\vec{r})$ needs to be solved self–consistently as illustrated in Fig. 3.1. There are several ways to stop the self–consistency cycle. A common way is to check the total energy calculation after each iteration and stop the cycle when the difference is below a supplied cutoff. Another possibility that is also widely used is to compare the newly generated electron density with the previous and enforce a similar supplied cutoff comparison.

From Eq. (3.30) it is clear that the sum of single particle energies ϵ_i is not equal to the total energy E_0. The single particle energy was introduced as a variable and they represent the eigenvalues of an equation that gives the electron density that minimizes Eq. (3.14). As a consequence it is difficult to associate the eigenvalues calculated with this approach to an experimental energy spectrum. The term pseudoparticle was introduced to avoid confusing the readers. However, during the last 50 years, this nomenclature has been lost due to the fact that the calculated energy spectra yield acceptable experimental agreements in an astonishing wide area of systems.

One obvious limitation of the Kohn-Sham equations is that they require closed expressions or other ways of calculating $v_{\rm ext}(\vec{r})$ and $E_{\rm xc}[n(\vec{r})]$. Unfortunately, this is difficult and the search for new and optimized functionals has been an important cornerstone for the modern success of density functional theory. Some like to think of functional generation as an art, and one should not be mistaken believing that all available functionals are equally good for a diverse span of extended systems. In most cases only a few functionals are able to produce acceptable agreement with experiments for benchmark systems. It is tempting to claim that it is the choice of the functional that is the art, and not its generation. We will briefly discuss the properties of exchange–correlation functionals in general. But, before doing so, we will address an issue which extends far beyond density functional theory. In fact it is a general requirement for any matrix equation solver. The choice of the basis set.

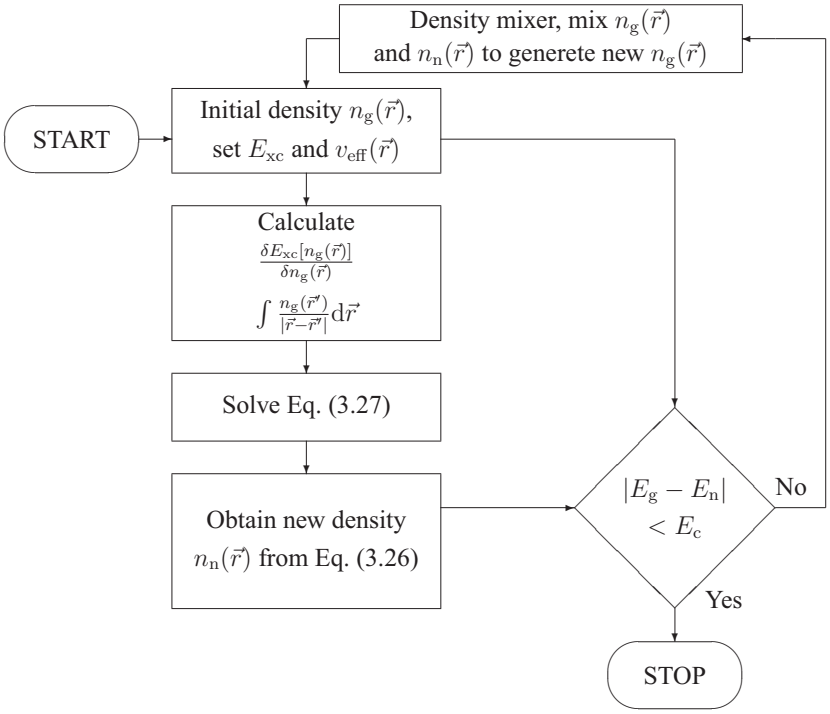


Figure 3.1: Flow chart of a typical Kohn–Sham based density functional calculation where a tolerance in the total energy E_{c} is used as a cutoff to terminate the density functional calculations when the absolute difference between the old E_{g} and the new E_{n} total energy is less than E_c.

3.2.2 Choice of basis set

To expand the eigenvectors $|\Psi\rangle$ it is necessary to choose a basis set that is representative for the system. There exists a wealth of different basis sets, some more general than others. In this section we will briefly cover the basics of the most common basis sets.

Slater determinants

The simplest form of an N electron eigenvector is the product of one electron eigenvectors ϕ_i

$$\Psi(\vec{x}_1, \vec{x}_2 \dots \vec{x}_N) = \phi_1(\vec{x}_1)\phi_2(\vec{x}_2)\cdots\phi_N(\vec{x}_N), \tag{3.32}$$

usually referred to as a Hartree product. To obey the antisymmetry of the Pauli exclusion principle [6], the eigenvectors need to be expanded as Slater determinants of the one electron eigenvectors. The simplest two particle Slater determinant is expanded as

$$\Psi(\vec{x}_1, \vec{x}_2) = \frac{1}{\sqrt{2}} \left(\phi_1(\vec{x}_1)\phi_2(\vec{x}_2) - \phi_1(\vec{x}_2)\phi_2(\vec{x}_1) \right). \tag{3.33}$$

Now, for the many–particle case, this gets far more complicated

$$\Psi(\vec{x}_1, \vec{x}_2 \dots \vec{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{x}_1) & \phi_2(\vec{x}_1) & \dots & \phi_N(\vec{x}_1) \\ \phi_1(\vec{x}_2) & \phi_2(\vec{x}_2) & \dots & \phi_N(\vec{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\vec{x}_N) & \phi_2(\vec{x}_N) & \dots & \phi_N(\vec{x}_N) \end{vmatrix}. \tag{3.34}$$

Such many–particle Slater determinants are very demanding to calculate and only a single Slater determinant is usually implemented in Kohn-Sham DFT (and Hartree–Fock) procedures. But it is important to emphasize that there exist generalized Kohn-Sham schemes that include a combination of several Slater determinants to calculate exchange kernels and correlation effects on a perturbative level. More accurate many–particle Configuration Interaction (CI) or Multi–Configurational Self–Consistent Field (MCSCF) calculations also require a multi–term Slater determinant. For simple molecules it is feasible to use the full Slater approach, but for extended systems this is not an option due to the unfavourable scaling of the calculation of the determinant. To use the Slater approach or not, knowledge about the single particle eigenvectors ϕ_i is necessary. The most basic type of eigenvectors are hydrogen like orbitals, e.g. Slater–Type Orbitals (STO) [7].

Linear combination of atomic orbitals

Linear combination of atomic orbitals (LCAO) is based on single atomic orbitals (e.g. STOs). To set up the basis, the atomic orbitals are expanded up to a given cutoff. To increase flexibility and speed Gaussian–Type Orbitals (GTO) [8] are often used instead of STOs. A single GTO does not reproduce a hydrogen like orbital, but the operator integrals are faster to calculate (they have an explicit form). The atomic orbitals can be efficiently calculated and are still extensively used in molecular calculations. Such basis sets are however prune to failure if the states are delocalized with respect to the atom sites and are thus not optimal for many materials.

Pseudo–potentials

For calculations beyond simple molecules additional approximations of the basis set are necessary to improve performance. One of the first and simplest approaches was the use of pseudo–potentials. The main idea is to separate the core and valence electrons and exclude the core states from the self–consistent calculation. The core states are predetermined for all atoms, based on atomic calculations. The contributions from the core states are included in the Kohn-Sham equations through the external potential. The pseudo–potential approach simplifies the choice of basis set due to the exclusion of a great number of electrons. However, a proper basis set is still needed for the valence electrons and at the same time this basis set should be compatible with the pseudo–potentials such that there is acceptable overlap.

Plane wave based basis functions

Plane wave basis sets exploit the periodicity of structures [9] (Chap. 2 and App. F). Basically plane waves are waves expanded based on the Brillouin zone reciprocal vectors. The plane waves need a large amount of coefficients to describe a localized state, but combined with the pseudo–potential approach they are very efficient and have some advantages over the orbital basis sets (mainly non–local expansions). Plane wave basis sets yield a simple form of the Kohn-Sham equations and are thus easy to implement. Due to the generality of the plane waves the same basis set can be utilized for all atom species.

Linearized Augmented plane waves (LAPW) [10] (Sec. 4.2) incorporate atomic–like states in the atomic region, while envelope functions are used to expand the interstitial (or bonding) states. The envelope functions are determined by plane wave expansions. Similarly, Linear Muffin–Tin–Orbitals (LMTO) [11] are obtained if the envelope functions are expanded by Hankel functions (solutions of the Laplace equation). The main difference between LAPW and LMTO lies in the interstitial region, where the energy dependence is excluded in the LMTO approach. This makes LMTO more feasible for closed packed structures.

A generalization of the LAPW and pseudo–potential method is called the Projected Augmented Wave (PAW) [12] method. It is more efficient while retaining the accuracy of the LAPW approach. As a consequence it is praised as one of the best basis sets for first–principle calculations [13, 14] of materials.

3.2.3 Exchange–correlation functionals

Before addressing the exchange–correlation functionals we will briefly consider an important aspect known as the derivative discontinuity of the exchange–correlation functionals. Unfortunately we will see that this is an inherent feature of the Kohn-Sham equations. This section closely follows the work published by Perdew et al. [15, 16].

Let us start with a system containing $N = N' + \omega$ electrons, where N' is a positive integer and $0 \leq \omega \leq 1$. The introduction of fractional electron numbers is unphysical, but a time averaged mixing of N' and $N' + 1$ is possible [16]. The parameter ω controls the mixing of the $\Psi_{N'}$ and $\Psi_{N'+1}$ states such that

$$N = N' + \omega, \tag{3.35}$$

$$= (1 - \omega)N' + \omega(N' + 1), \tag{3.36}$$

since

$$N = \int n_N(\vec{r})d\vec{r}, \tag{3.37}$$

$$n_N(\vec{r}) = (1 - \omega)n_{N'}(\vec{r}) + \omega n_{N'+1}(\vec{r}). \tag{3.38}$$

From Eq. (3.30) we know that E_0 is a functional of $n(\vec{r})$. Hence E_0 can be written as

$$E_0[n_N(\vec{r})] = (1 - \omega)E_0[n_{N'}(\vec{r})] + \omega E_0[n_{N'+1}(\vec{r})], \tag{3.39}$$

which are continuous lines as a function of N' [16, 17]. Thus, $\partial E_0/\partial N$ should be ill–defined for each integer number of particles in the system.

The ionization potential I for an atom X with N electrons is defined as the process

$$X \rightarrow X^+ + e^-, \tag{3.40}$$

with

$$I(N) = E_0(N - 1) - E_0(N), \tag{3.41}$$

where the outermost and weakest bonded electron is removed to infinity, such that there is no interaction between the ion and the electron. Similarly the electron affinity A is defined as

$$X^- \rightarrow X + e^-, \tag{3.42}$$

with

$$A(N) = E_0(N) - E_0(N+1). \tag{3.43}$$

We can relate I and A to a change in $n(\vec{r})$ such that

$$I(N) = -\left.\frac{\delta E_0}{\delta n_N(\vec{r})}\right|_{N'-1}, \tag{3.44}$$

and

$$A(N) = -\left.\frac{\delta E_0}{\delta n_N(\vec{r})}\right|_{N'+1}. \tag{3.45}$$

Looking back at Eq. (3.29) we see that the density functional derivative of the external potential $V_{\rm s}$ and electrostatic potential $V_{\rm H}$ is inherently continuous, which leaves the kinetic energy $T_{\rm s}$ and exchange–correlation energy $E_{\rm xc}$ to obey the fundamental discontinuity property. The kinetic energy derivative and its discontinuity is trivial due to the direct evaluation of the pseudo particle orbitals. Hence, it is the exchange–correlation functional that ought to reproduce the discontinuities. The difference between the ionization potential and electron affinity is usually defined as $I - A$, which, if the system is charge neutral, determines the fundamental band gap $E_{\rm g}$,

$$\begin{aligned} E_{\rm g}(N) &= I(N) - A(N), && (3.46)\\ &= \left.\frac{\delta T}{\delta n(\vec{r})}\right|_{N+1} - \left.\frac{\delta T}{\delta n(\vec{r})}\right|_{N-1} + \left.\frac{\delta E_{xc}}{\delta n(\vec{r})}\right|_{N+1} - \left.\frac{\delta E_{xc}}{\delta n(\vec{r})}\right|_{N-1}, && (3.47)\\ &= \Delta_{\rm KS} + \Delta_{\rm xc}, && (3.48) \end{aligned}$$

where $\Delta_{\rm KS}$ and $\Delta_{\rm xc}$ are the kinetic and exchange–correlation derivative differences, respectively. For all systems, except the simplest non–interacting ones, $\Delta_{\rm xc} \neq 0$ [15]. This is a paradox, since, by construction, $E_{\rm xc}$ should be continuously differentiable and independent of the energy. Unfortunately there is no easy way to escape this problem, except going past the Kohn-Sham scheme to a more generalized Kohn-Sham scheme (gKS) that incorporates energy dependent functionals based on e.g. Fock terms (through the orbital dependence). We will now present two common ways of expanding the exchange–correlation functionals and introduce a relatively new hybrid approach including exact exchange (Fock) terms in the gKS formalism. The subsequent functionals all have spin–polarized versions, but to simplify notation they are not discussed. These and other failures of modern implementations of density functional theory are frequently discussed in the literature [18, 19, 20] and readers interested in the subject should consult the referred papers.

Local density approximation

One of the first approximations of the exchange–correlation energy was based on the local density approximation (LDA) [4, 5]

$$E_{\mathrm{xc}}^{\mathrm{LDA}} = \int \epsilon_{\mathrm{xc}}^{\mathrm{hom}}(n(\vec{r}))d\vec{r}, \tag{3.49}$$

where $\epsilon_{\mathrm{xc}}^{\mathrm{hom}}(n(\vec{r})) = \epsilon_{\mathrm{x}}^{\mathrm{hom}}(n(\vec{r})) + \epsilon_{\mathrm{c}}^{\mathrm{hom}}(n(\vec{r}))$ is the energy from the homogeneous free electron gas. The exchange contribution $\epsilon_{\mathrm{x}}^{\mathrm{hom}}(n(\vec{r}))$ is known exactly as

$$\epsilon_{\mathrm{x}}^{\mathrm{hom}}(n(\vec{r})) = -\frac{3e^2}{4}\left(\frac{3}{\pi}\right)^{1/3} n(\vec{r})^{4/3}, \tag{3.50}$$

where e is the electron charge. See Ref. [21] (contribution by V. Sahni, page 217). The correlation contribution $\epsilon_{\mathrm{c}}^{\mathrm{hom}}(n(\vec{r}))$ however, has to be parametrized. Most implementations today use a parameterization based on Quantum Monte Carlo (QMC) simulations done by Ceperley and Alder [22]. At first it came as a surprise that LDA gave reasonable agreement with experiments, even for systems well beyond the free electron gas. It was later discovered that this was due to error cancellation in the evaluation of the over–compensated exchange and the under–compensated correlation [23, 24] functionals.

Generalized gradient approximation

There is no doubt that the local form of the local density approximation will produce inaccurate results for systems deviating from the free electron gas regime. Therefore a more general approach was introduced, called the generalized gradient approximation (GGA) [25],

$$E_{\mathrm{xc}}^{\mathrm{GGA}}(n(\vec{r})) = \int f(n(\vec{r}), \nabla n(\vec{r}))d\vec{r}, \tag{3.51}$$

where $f(n(\vec{r}), \nabla n(\vec{r}))$ is a general functional of the density and its gradient. The difference between semi–local GGA functionals depends on the choice of $f(n(\vec{r}), \nabla n(\vec{r}))$ and they can thus differ substantially from each other. A discussion of the wide range of GGA functionals is not the intention of this thesis, so readers are referred to the literature and Ref. [26] for a quick overview (primarily Sec. 1.6). However, it should be emphasized that the GGA approach is still local, but corrected by the gradient $\nabla n(\vec{r})$ such that it is ultimately semi–local.

Hybrid LDA/GGA

One major concern with the general LDA and GGA functionals used today is the lack of delocalization correction [19], which tends to spread the electron charge. This correction is inherent to the exact exchange (Fock terms) [27] and is excluded for the aforementioned approaches due to the implicit energy (through the orbitals) dependence, and thus the performance penalties that follow.

Most hybrid functionals include a quarter portion of exact exchange and evaluate the remainder exchange from the base functional. The concept of the quarter mixing ratio is based on work done by Ernzerhof and coworkers [28, 29, 30]. One of the most common hybrid functionals use the PBE functional as a base and is therefore called PBE0 [31]. It is defined as

$$E_{\mathrm{PBE0}} = \frac{1}{4}E_x^{\mathrm{HF}} + \frac{3}{4}E_x^{\mathrm{PBE}} + E_c^{\mathrm{PBE}}, \tag{3.52}$$

where the subscripts (x) and (c) term the exchange and correlation parts, respectively. The superscript signalizes how to evaluate the given energy, where HF is the Hartree energy and the exact exchange (this is thus the well known Hartree–Fock [27] contribution). The PBE superscript tells us that the exchange and correlation should be evaluated from the PBE functionals. Similarly one can define range separated hybrid functionals, like the HSEx [32], defined as

$$E_{\mathrm{HSEx}} = \frac{1}{4}E_x^{\mathrm{HF,sr}} + \frac{3}{4}E_x^{\mathrm{PBE,sr}} + E_x^{\mathrm{PBE,lr}} + E_c^{\mathrm{PBE}}, \tag{3.53}$$

where (sr) and (lr) describe the short and long range parts of the exchange energy. This range separation is defined by separating the Coulomb kernel such that

$$\frac{1}{r} = \frac{\mathrm{erfc}(\mu r)}{r} + \frac{\mathrm{erf}(\mu r)}{r}, \tag{3.54}$$

where μ determines the range separation. For the HSE03 [32] functional this is defined as $\mu = 0.3$ Å^{-1}. The range separation is motivated for two reasons: first, it allows the calculation of the exact exchange on a coarser grid, thus improving convergence [33]; second, one has the possibility to investigate the range extension of the exact exchange energy. Recent reviews of the hybrid functionals reveal a very good compromise between accuracy and performance for a wide range of benchmark systems [33, 34, 35]. Although, depending on implementation, an order of magnitude increase in computational time has to be expected over the traditional LDA and GGA functionals. Memory requirements also increase due to the calculation and storage of the orbitals. There are still some questions as to whether the quarter portion of exact exchange

can be used for all systems (metallic systems in particular do not show much improvement [33]) or if there should be a headroom for tuning this portion. Future studies will reveal the necessary answers to these questions.

3.3 Electronic structure and bonding

To analyze the electronic structure and bonding of materials it is common to consider the density of states and band structure. They reveal hybridization, chemical shifts and bonding properties indirectly. We like to term such analysis state–resolved. Quantitative experimental verifications can be done by X–ray photoemission spectroscopy (XPS), ultra–violet photoemission spectroscopy (UPS), a diverse selection of band gap measurements and electron energy loss spectroscopy (EELS). However, these techniques usually contain a wealth of effects not present in the calculation of the ground state density of states (e.g. excitations, resonances and secondary effects). Thus, a direct verification of the validity of the calculation can sometimes be difficult.

However, and especially so for density functional calculations, all properties are calculated from the electron density. This density is thus also available for direct analysis and contrary to the state–resolved analysis one can perform spatial–resolved analysis of electron transfer in the system. These transfers are associated with the electron rearrangement to set up the bonds between the ions in the system. Experimental verifications can be done directly by obtaining the structure factors from X–ray and electron diffraction. These structure factors can be transformed to the real space by a Fourier transformation [36] (Sec. 6.5) and are then directly comparable to the electron density. Although such approach is uncommon today, we believe such analysis will be more important in the future.

3.3.1 Electron density analysis

Electron transfer

The electron density contains all necessary information to compute observables within DFT. Since the bonds are inherently determined by electron transfer it also serves as a perfect base for bond analysis and an alternative or complement to the state–resolved density of states analysis that is common practice these days [37] (Chap. 5). The electron transfer ρ_{b} can be determined by

$$\rho_{\mathrm{b}}(\vec{r}) = \rho_{\mathrm{e}}(\vec{r}) - \rho_{\mathrm{r}}(\vec{r}), \tag{3.55}$$

where ρ_c and ρ_r are the crystal and reference electron density, respectively. The crystal electron density is obtained directly from calculations, but ρ_r still needs to be determined.

From a theoretical point of view we can use the procrystal [38] electron density ρ_p as a reference ρ_r. Generated from a superposition of free atomic electron densities embedded in the crystal unit cell, this reference should be free of bonding features.

Different experimental methods exist to determine the electron density and/or electron transfer. One approach is to compare spectroscopic data with an elemental reference ρ_m [39, 40, 41, 42], and the relative intensity difference can be converted to occupancy numbers [43]. However, this approach is burdened by differences in the reference bonds and could yield unreliable results if two elements with different bonding characteristics are used.

A more direct approach is to perform diffraction experiments and refine the structure factors by starting from overlapping atomic orbitals [38, 44, 45]. This reference is in principle equivalent to the previously discussed procrystal. The structure–factor refinement is stopped when sufficient agreement with experimental diffraction intensities is reached.

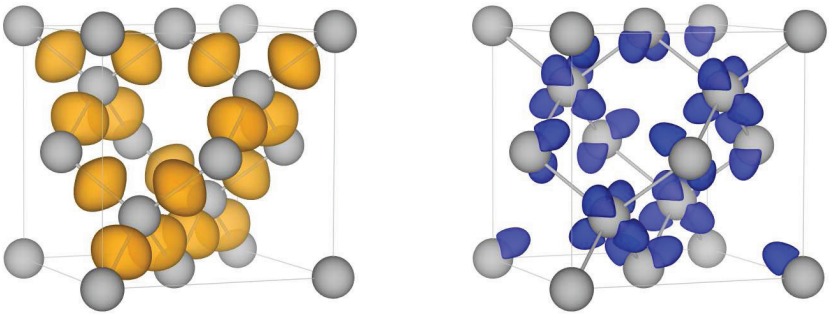
The procrystal represents a consistent reference in electron transfer analysis, but may today be difficult to use as a reference experimentally due to the lack of established and efficient experimental procedures to obtain a large number of structure factors.

3.3.2 A few examples

To demonstrate the spatially resolved electron transfer bond analysis and its strengths, we here give three standard examples; the covalently bonded silicon, the ionic sodium chloride and the metallic aluminium. In addition we give a more complicated example of the filled phosphorus skutterudite $LaFe_4P_{12}$. All calculations were done on a dense augmented grid where the all–electron density was regenerated [13] to get rid of the problematic compensator charges inherent to most of the density functional implementations using plane waves. In all examples the respective procrystal has been used as the electron density reference.

Covalent silicon

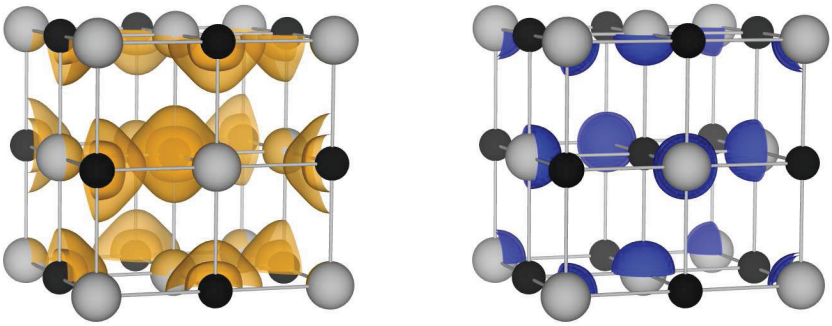
Elemental silicon, Si, was chosen as it is well known from the literature to contain exclusively covalent bonds. It is also relatively simple due to the high symmetry and small primitive cell. In Fig. 3.2 we show the electron accumulation and depletion arising from the positive and negative ρ_b, respectively. The electron accumulation in Fig. 3.2(a) clearly illustrates the covalent character of silicon. The covalent bonds are set up by rearranging the electron density around Si, moving



(a) Electron accumulation (b) Electron depletion

Figure 3.2: (a) Electron accumulation (positive ρ_b) and (b) depletion using the respective pro-crystal as a reference for Si. Isosurfaces are drawn at 0.008/bohr^3.

electron density from non–directional to the directional Si–Si parts of the crystal (see Fig. 3.2(b)).



(a) Electron accumulation (b) Electron depletion

Figure 3.3: (a) Electron accumulation (positive ρ_b) and (b) depletion using the respective pro-crystal as a reference for NaCl. Na in gray and Cl in black. Isosurfaces are drawn at 0.005/bohr^3.

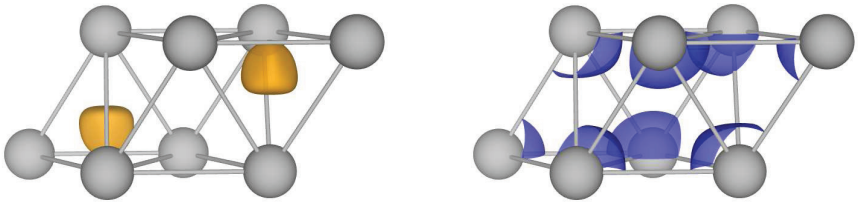
Ionic sodium chloride

Elemental sodium–chloride, NaCl, serves as a prototype example of ionic bonding and is usually used as a textbook example, making the structure and bonding properties well known. Similar to silicon we illustrate the electron accumulation and depletion in Fig. 3.3. From this figure, the ionic character, with local depletion (Na^+) on the Na and local accumulation (Cl^-) on the Cl

sites are reproduced.

Metallic aluminium

To represent metallic bonding, elemental aluminium, Al, was chosen. Similarly to Si and NaCl its structure and bonding properties are well known from the literature and from Fig. 3.4 we see that the metallic character is reproduced. Electron density is depleted close to the Al site and redistributed between the Al sites in a non–directional manner.



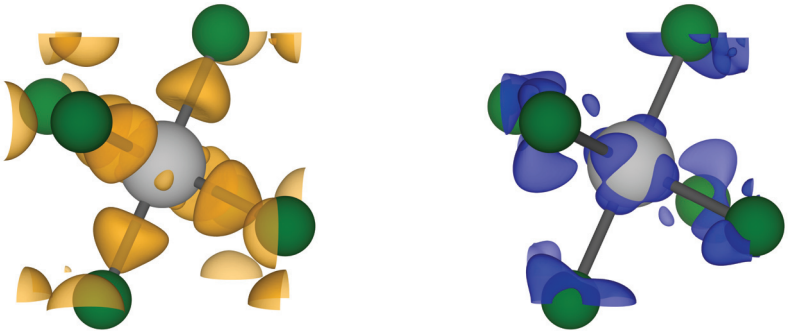
(a) Electron accumulation (b) Electron depletion

Figure 3.4: (a) Electron accumulation (positive ρ_b) and (b) depletion using the respective procrystal as a reference for Al. Isosurfaces are drawn at 0.004/bohr^3.

Phosphorus filled skutterudite $LaFe_4P_{12}$

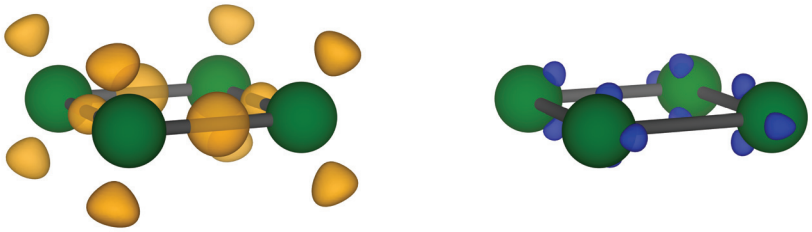
As a final example we include the lanthanum filled iron phosphorus skutterudite $LaFe_4P_{12}$ (see Paper I for structure information). The reason for this is that electron transfer analysis of this compound reveals what seems to be a unique electron arrangement around the lanthanum ions. This could explain why the filled skutterudite structure so successfully reduces the thermal conductivity. In addition, the skutterudite structure contains covalent bonds. In Fig. 3.5 the electron accumulation and depletion are given for the FeP_6 octahedral arrangement. The biased (towards P) covalent bond between Fe and P can be explained from the difference in electronegativity. Similarly in Fig. 3.6, the electron transfers between the P ions in the P_4 ring are covalent and due to the rectangular shape, one bond is stronger, which is nicely reproduced by this analysis.

Finally in Fig. 3.7 the electron transfer around the La ion is shown. The electron accumulations (only partially shown due to the cutout of the unit cell) responsible for the La–P bonds can be seen on the border in the La plane. Close to the La ion, elongated electron accumulation is aligned towards the P ions. These electron accumulations could set up restrictions of the La motion during vibrational perturbations and thus contribute to the significant reduction in the lattice thermal conductivity [46, 47].



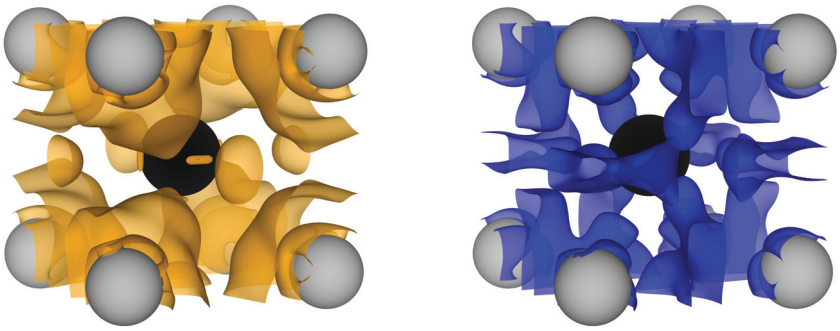
(a) Electron accumulation (b) Electron depletion

Figure 3.5: (a) Electron accumulation (positive ρ_b) and (b) depletion between the grey Fe and green P in the octahedra using the respective procrystal as a reference. Isosurfaces are drawn at 0.024/bohr^3.



(a) Electron accumulation (b) Electron depletion

Figure 3.6: (a) Electron accumulation (positive ρ_b) and (b) depletion between the green P ions using the respective procrystal as a reference. Isosurfaces are drawn at 0.017/bohr^3.



(a) Electron accumulation (b) Electron depletion

Figure 3.7: (a) Electron accumulation (positive ρ_b) and (b) depletion around the black La using the respective procrystal as a reference. The Fe ions are shown in grey. Isosurfaces are drawn at 0.0025/bohr^3.

3.4 Transport properties

The performance of thermoelectric materials is determined by the figure–of–merit zT as previously discussed. To determine zT extensive knowledge about the system is required. More specifically we need to determine the electrical and thermal conductivity, and the Seebeck coefficient, all preferably at different temperatures. Calculations of these properties are not trivial, due to several intricate problems with the first–principle calculations that serve as a baseline for the transport calculations. It should also be mentioned that the experimental determination of these properties is difficult. In this chapter we will introduce a semi–classical procedure to calculate the electrical conductivity σ and the Seebeck coefficient α, which together determine the power factor $\alpha^2\sigma$. In addition we calculate the electrical thermal conductivity such that the figure–of–merit can be determined for a given lattice thermal conductivity.

We now develop a general approach by discussing the strengths, pitfalls and implementation of the semi–classical Boltzmann transport equations. Most of the derivation of the distribution function and the Boltzmann transport equation is based on the textbook of Ashcroft and Mermin [9] (Chap. 12, 13 and 16). Before the explicit Boltzmann equation is derived, we briefly discuss electron transport in solids in general. We then dedicate a section of prediction based on simple band engineering and propose band structure characteristics that give the largest power factor and figure–of–merit. Readers of this section should be aware that where applicable the position and wave vector are dependent on time. The time dependence is shown explicitly where needed. However, for some variables with an implicit time dependence through the position and wave vector (e.g. equilibrium distribution function), explicit time dependence is still shown. This notation is not strictly valid, but has been chosen for simplicity.

3.4.1 Electronic transport in solids

According to Bloch theory [9] (Chap. 12), electrons do not scatter if the lattice is periodic and perfect. Such lattices do not exist in reality and scattering events are usually classified in three main categories;

- Scattering from thermal vibrations in the lattice. Due to vibrations in the lattice each ion will deviate from the equilibrium position and generate imperfections in the crystal. Electrons scatter locally from these imperfections. One further separates these vibrations into acoustical and optical vibrations, based on the vibrational character. Lattice vibrations primarily arise due to thermal energy and are thus strongly temperature dependent.

- Scattering from deviations from the atom positions in the perfect crystal (e.g. impurities and vacancies) and/or periodic lattice (e.g. boundary effects). Usually the two are related and imperfections can in general be local or extended, depending on type.

- Scattering based on strong electron–electron interactions. These types of scattering events arise due to exact exchange (Hartree–Fock) and screening interactions. Typical relaxation times (the time between each collision) for these scattering events are about 10^2-10^6 [9] (Chap. 17) times larger than vibrational scattering and scattering from imperfections and can in most cases be safely discarded. We would have to go to very low temperature to remove any thermal vibration for electron–electron scattering to be an observable effect in samples free of imperfections.

The electronic transport depends on the distribution of electrons. This distribution generally differs from the equilibrium Fermi distribution and we will term its function g in the following. The distribution function will depend on both the position vector $\vec{r}$, the wave vector $\vec{k}$ and time t. We will however suppress the $\vec{r}$ dependence until we discuss the Boltzmann equation in the next section. Before exploring this distribution function, let us first discuss the scattering process. We will disregard spin effects in the derivation, but emphasize that most of these equations are equally valid for bosons and fermions. Let us define a scattering probability $\Gamma_{\vec{k},\vec{k}',n,n'}$, where $\vec{k}$ and $\vec{k}'$ are the initial and final wave vectors, and similar for the band index n. The scattering rate can now be defined as

$$\frac{1}{\tau_n^{\text{out}}(\vec{k},t)} = \int \frac{d\vec{k}'dn'}{\Omega} \left(1 - g_{n'}(\vec{k}',t)\right) \Gamma_{\vec{k},\vec{k}',n,n'}, \tag{3.56}$$

where $\Gamma_{\vec{k},\vec{k}',n,n'}d\vec{k}'/\Omega$ is the probability per unit time that a scattering event from $\vec{k}$ and n to $d\vec{k}'dn'$ within $\vec{k}'$ and n' takes place. The unoccupied distribution function $1-g_{n'}(\vec{k}',t)$ is needed since the occupied (and thus no scattering) states are determined by $g_{n'}(\vec{k}',t)$. By integrating all possible $\vec{k}'$, n' and normalizing using the $\vec{k}$-space volume Ω (taken to be $(2\pi)^3$ or $4\pi^3$, depending on whether to account for spin degeneracy) the total rate is obtained. Notice that the relaxation time $\tau_n(\vec{k})$ is not explicitly a function of $\vec{k}$ (although we still persist on this notation for simplicity), but depends on $\vec{k}$ through $\Gamma_{\vec{k},\vec{k}',n,n'}$ and $g_{n'}(\vec{k}',t)$.

The change in $g_n(\vec{k},t)$ for an infinitesimal time interval dt is the probability for scattering out of $\vec{k}$ times the distribution function $g_n(\vec{k},t)$, that is

$$\left(\frac{dg_n(\vec{k},t)}{dt}\right)^{\text{out}} = -\frac{g_n(\vec{k},t)}{\tau_n^{\text{out}}(\vec{k})}. \tag{3.57}$$

The out scattering reduces the states in $\vec{k}$, hence the minus.

In Eq. (3.56) and Eq. (3.57) scattering out of $\vec{k}$ was defined. However, we simultaneously need to account for the inbound scattering processes. To obtain this we use Eq. (3.56) and substitute $1 - g_{n'}(\vec{k'}, t)$ by $g_{n'}(\vec{k'}, t)$ and $\Gamma_{\vec{k},\vec{k'},n,n'}$ by $\Gamma_{\vec{k'},\vec{k},n',n}$ (now termed $1/\tau_n^{\text{in}}(\vec{k}, t)$). To get the inbound scattering change in $g_n(\vec{k}, t)$ we multiply this by the unoccupied distribution function $1 - g_n(\vec{k}, t)$, such that

$$\left(\frac{dg_n(\vec{k}, t)}{dt}\right)^{\text{in}} = \frac{1 - g_n(\vec{k}, t)}{\tau_n^{\text{in}}(\vec{k})}. \tag{3.58}$$

Adding the outbound and inbound terms yields the total change in $g_n(\vec{k})$ as

$$\left(\frac{dg_n(\vec{k}, t)}{dt}\right)^{\text{total}} = \left(\frac{dg_n(\vec{k}, t)}{dt}\right)^{\text{out}} + \left(\frac{dg_n(\vec{k}, t)}{dt}\right)^{\text{in}}, \tag{3.59}$$

$$= -g_n(\vec{k}, t) \int \frac{d\vec{k'}dn'}{\Omega} \left(1 - g_{n'}(\vec{k'}, t)\right) \Gamma_{\vec{k},\vec{k'},n,n'}$$

$$+ \left(1 - g_n(\vec{k}, t)\right) \int \frac{d\vec{k'}dn'}{\Omega} \left(g_{n'}(\vec{k'}, t)\right) \Gamma_{\vec{k'},\vec{k},n',n}. \tag{3.60}$$

The scattering probability Γ can in general be very complex. For simple impurity scattering one can for example develop an expression for Γ based on the "Golden rule" of Fermi. See for example Ref. [48] (Sec. 3.2) and [1] (Sec. 8.3). For vibrational scattering an entirely different process needs to be developed to also include the electron–phonon interaction. It should be clear to the reader that the general scattering problem is comprehensive. Even many–body problems free of scattering are problematic to solve and it should not come as a surprise that we need to introduce approximations.

We will now assume that a scattering event is not determined by the initial distribution function $g_n(\vec{k})$, and that the scattering events maintain the thermodynamic equilibrium. Furthermore, we assume that the relaxation time $\tau_n(\vec{k}, t)$ is only dependent on the band index n and the wave vector $\vec{k}$. Hence, no implicit dependence through $g_n(\vec{k}, t)$ or Γ is allowed. Now, the in scattering is $g_n^0(\vec{k}, t)$ times the probability $dt/\tau_n(\vec{k})$ of a scattering event in the time interval dt. More specifically

$$\left(\frac{dg_n(\vec{k}, t)}{dt}\right)_{\text{RTA}}^{\text{total}} = \frac{g_n^0(\vec{k}, t) - g_n(\vec{k}, t)}{\tau_n(\vec{k}, t)}, \tag{3.61}$$

where $g_n^0(\vec{k})$ is the equilibrium distribution function. All approximations that led to Eq. (3.61) and hence the equation itself are usually called the relaxation time approximation (RTA). This

approximation is very common and used in almost all transport calculations today. The in and out scattering terms can be identified as

$$\left(\frac{dg_n(\vec{k},t)}{dt}\right)^{\text{in}}_{\text{RTA}} = \frac{g^0_n(\vec{k},t)}{\tau_n(\vec{k},t)}, \tag{3.62}$$

$$\left(\frac{dg_n(\vec{k},t)}{dt}\right)^{\text{out}}_{\text{RTA}} = -\frac{g_n(\vec{k},t)}{\tau_n(\vec{k},t)}. \tag{3.63}$$

We will see later, that further approximations need to be introduced to make the transport calculations tractable for extended systems.

3.4.2 The Boltzmann equation

In this section the Boltzmann equation is derived from the semi–classical equations of motion [9] (Chap. 12) by considering a change in $g_n(\vec{r},\vec{k},t)$. The change of a position vector is known as

$$\frac{d\vec{r}}{dt} = \vec{v}(\vec{r},\vec{k},t), \tag{3.64}$$

where $\vec{v}(\vec{r},\vec{k},t)$ is the quantum mechanical velocity average of the velocity operator. Similarly, the change in wave vector from external forces is

$$\frac{d\vec{k}}{dt} = \frac{1}{\hbar}\vec{F}(\vec{r},\vec{k},t) \tag{3.65}$$

$$= -\frac{e}{\hbar}\left(\vec{E}(\vec{r},\vec{k},t) - \frac{1}{c}\vec{v}(\vec{r},\vec{k},t)\times\vec{H}(\vec{r},\vec{k},t)\right), \tag{3.66}$$

where $\vec{F}(\vec{r},\vec{k},t)$ is the external field that contains both the electric $\vec{E}(\vec{r},\vec{k},t)$ and magnetic $\vec{H}(\vec{r},\vec{k},t)$ fields, while c is the speed of light. A change of $g_n(\vec{r},\vec{k},t)$ in the time interval dt can now be determined by the modified equations of motion, such that

$$g_n(\vec{r},\vec{k},t)d\vec{r}d\vec{k} = g_n\left(\vec{r}-\vec{v}(\vec{r},\vec{k},t)dt, \vec{k} - \frac{\vec{F}(\vec{r},\vec{k},t)}{\hbar}dt, t-dt\right) \times d\vec{r}(t-dt)d\vec{k}(t-dt). \tag{3.67}$$

Through Liouville's theorem [9] (Chap. 12) the $\vec{r}\vec{k}$ phase space is conserved during a dt change. Thus, the infinitesimal terms on both sides can be cancelled. It should be clear that Eq. (3.67)

does not account for scattering events. These terms need to be added and we get

$$\frac{g_n(\vec{r},\vec{k},t)}{dt} = \frac{g_n\left(\vec{r}-\vec{v}(\vec{r},\vec{k},t)dt,\vec{k}-\frac{\vec{F}(\vec{r},\vec{k},t)}{\hbar}dt,t-dt\right)}{dt} + \left(\frac{dg_n(\vec{r},\vec{k},t)}{dt}\right)^{\text{total}}. \tag{3.68}$$

Expanding to linear order in dt and taking the limit $dt \to 0$ reveals the famous Boltzmann equation

$$\frac{dg_n(\vec{r},\vec{k},t)}{dt} + \vec{v}(\vec{r},\vec{k},t)\cdot\frac{\partial g_n(\vec{r},\vec{k},t)}{\partial\vec{r}} + \frac{1}{\hbar}\frac{\partial g_n(\vec{r},\vec{k})}{\partial\vec{k}}\cdot\vec{F}(\vec{r},\vec{k},t) = \left(\frac{dg_n(\vec{r},\vec{k},t)}{dt}\right)^{\text{total}}. \tag{3.69}$$

The expressions on the left hand side control the total change in $g_n(\vec{r},\vec{k},t)$, the change from diffusion and the change from applied fields, respectively. On the right hand side we have the scattering contributions.

3.4.3 Non–equilibrium distribution function

We need knowledge about the non–equilibrium distribution function to be able to utilize the Boltzmann equation and will now go through derivations that lead it to a tractable form. The derivations follow closely the textbook of Ashcroft and Mermin [9] (Chap. 12, 13 and 16).

Assume that the infinitesimal phase space $d\vec{r}d\vec{k}$ contains dN electrons such that

$$dN = g_n(\vec{r},\vec{k},t)\frac{d\vec{r}d\vec{k}}{4\pi^3}. \tag{3.70}$$

The distribution function $g_n(\vec{r},\vec{k},t)$ can now be determined from the time relaxation approximation if we assume that all previous scattering in the volume $d\vec{r}\,'d\vec{k}'$ from t' are scattered into the volume $d\vec{r}d\vec{k}$ at t. This is a reasonable assumption as the particles will follow the semi–classical equations of motion from t' to t. From Liouville's theorem [9] (Chap. 12), $d\vec{r}\,'d\vec{k}' = d\vec{r}d\vec{k}$. The in scattered part of the relaxation time approximation (Eq. (3.62)) can be integrated over t' in infinitesimal dt' parts from $-\infty$ to t, such that the distribution function now becomes

$$\int_{-\infty}^{t}\frac{g_n^0(\vec{r},\vec{k},t')}{\tau_n(\vec{r},\vec{k},t')}dt', \tag{3.71}$$

which need to be modified, since there can be scattering events between t' and t. We need the probability $P_n(\vec{r},\vec{k},t,t')$ of no further scattering after t', but before t, to be multiplied into the integrand. Hence,

$$g_n(\vec{r},\vec{k},t) = \int_{-\infty}^{t} \frac{g_n^0(\vec{r},\vec{k},t')P_n(\vec{r},\vec{k},t,t')}{\tau_n(\vec{r},\vec{k},t')}dt'. \tag{3.72}$$

Substituting this back into Eq. (3.70) yields

$$dN = \frac{d\vec{r}d\vec{k}}{4\pi^3}\int_{-\infty}^{t} \frac{g_n^0(\vec{r},\vec{k},t')P_n(\vec{r},\vec{k},t,t')}{\tau_n(\vec{r},\vec{k},t')}dt'. \tag{3.73}$$

We now need to determine the probability $P_n(\vec{r},\vec{k},t,t')$. This can be done by noting that

$$P_n(\vec{r},\vec{k},t,t') = P_n(\vec{r},\vec{k},t,t'+dt')\left(1-\frac{dt'}{\tau_n(\vec{r},\vec{k},t')}\right), \tag{3.74}$$

since the number of particles that undergo no scattering between t' and t is less than between $t'+dt'$ and t by the ratio $1-dt'/\tau_n(\vec{r},\vec{k},t')$. Taking the limit $dt' \to 0$ we get a differential equation

$$\frac{\partial P_n(\vec{r},\vec{k},t,t')}{\partial t'} = \frac{P_n(\vec{r},\vec{k},t,t')}{\tau_n(\vec{r},\vec{k},t')}, \tag{3.75}$$

with solution

$$P_n(\vec{r},\vec{k},t,t') = \exp\left(-\int_{t'}^{t}\frac{dt''}{\tau_n(\vec{r},\vec{k},t'')}\right), \tag{3.76}$$

subject to the boundary condition $P_n(\vec{r},\vec{k},t,t) = 1$. We can now substitute Eq. (3.75) into Eq. (3.72) and obtain

$$g_n(\vec{r},\vec{k},t) = \int_{-\infty}^{t} g_n^0(\vec{r},\vec{k},t')\left(\frac{\partial}{\partial t'}P_n(\vec{r},\vec{k},t,t')\right)dt'. \tag{3.77}$$

Integrating by parts yields

$$g_n(\vec{r},\vec{k},t) = g_n^0(\vec{r},\vec{k},t) - \int_{-\infty}^{t} P_n(\vec{r},\vec{k},t,t')\left(\frac{d}{dt'}g_n^0(\vec{r},\vec{k},t')\right)dt', \tag{3.78}$$

if we also set the condition $P_n(\vec{r},\vec{k},0,-\infty) = 0$, that is, that no particles can reach t without experiencing scattering events. Next, is the time derivative of the equilibrium distribution function.

Expanding the derivative as

$$\begin{aligned}\frac{dg_n^0(\vec{r},\vec{k},t')}{dt'} &= \frac{\partial g_n^0(\vec{r},\vec{k},t')}{\partial \epsilon_n(\vec{k},t')}\frac{\partial \epsilon_n(\vec{k},t')}{\partial \vec{k}}\cdot\frac{d\vec{k}}{dt'} + \frac{\partial g_n^0(\vec{r},\vec{k},t')}{\partial T(\vec{r},t')}\frac{\partial T(\vec{r},t')}{\partial \vec{r}}\cdot\frac{d\vec{r}}{dt'} \\ &+ \frac{\partial g_n^0(\vec{r},\vec{k},t')}{\partial \mu(\vec{r},t')}\frac{\partial \mu(\vec{r},t')}{\partial \vec{r}}\cdot\frac{d\vec{r}}{dt'},\end{aligned} \tag{3.79}$$

and applying the semi–classical equations of motion from Eq. (3.64) and Eq. (3.66) yields,

$$\begin{aligned}g_n(\vec{r},\vec{k},t) &= g_n^0(\vec{r},\vec{k},t) \\ &- \int_{-\infty}^{t} P_n(\vec{r},\vec{k},t,t')\left[\frac{\partial g_n^0(\vec{r},\vec{k},t')}{\partial \epsilon_n(\vec{k},t')}\vec{v}_n(\vec{k},t')\right. \\ &\times \left.\left(-e\vec{E}(\vec{r},\vec{k},t) - \nabla\mu(\vec{r},t') - \left(\frac{\epsilon_n(\vec{k},t')-\mu(\vec{r},t')}{T(\vec{r},t')}\right)\nabla T(\vec{r},t')\right)\right]dt'.\end{aligned} \tag{3.80}$$

This is a very general function. At first sight it might seem that Eq. (3.80) is independent of a magnetic field. This is not the case as it enters implicitly by modifying $\vec{r}$ and $\vec{k}$. We can induce additional approximations to simplify Eq. (3.80) further. If we assume a spatial uniform electric field and temperature gradient, the $\vec{r}$ dependence drops out of $\vec{E}$ and ∇T. The same assumption can also be made for the relaxation time and the chemical potential (depends on space through T) such that the $\vec{r}$ dependence drops out of the integrand. The equilibrium distribution function $g_n^0(\vec{r},\vec{k},t)$ is dependent on $\vec{r}$ through T and on $\vec{k}$ through $\epsilon_n(\vec{k},t)$ and is also under the previous assumptions independent on $\vec{r}$. In this thesis, we will only work with electrons, so we now abbreviate

$$g_n^0(\vec{r},\vec{k},t) = f(\vec{k},t), \tag{3.81}$$

where $f(\vec{k},t)$ is the equilibrium Fermi distribution function.

Even though we developed an expression for $g_n(\vec{r},\vec{k},t)$, we have yet to give a more specific form of $P_n(\vec{r},\vec{k},t,t')$, except for the one in Eq. (3.76). If we now assume a relaxation time that depends on energy only, the probability becomes trivial from Eq. (3.76), such that

$$P_n(\vec{k},t,t') = e^{-(t-t')/\tau_n(\vec{k})}, \tag{3.82}$$

where the $\vec{k}$ dependence in $\tau_n(\vec{k})$ enters implicitly through $\epsilon_n(\vec{k},t)$. By applying all these ap-

proximation to Eq. (3.80), we obtain

$$\begin{aligned} g_n(\vec{k},t) &= g_n^0(\vec{k},t) \\ &- \int_{-\infty}^{t} e^{-(t-t')/\tau_n(\vec{k})} \left[\frac{\partial f(\vec{k},t')}{\partial \epsilon_n(\vec{k},t')} \vec{v}_n(\vec{k},t') \right. \\ &\times \left. \left(-e\vec{E}(t') - \nabla\mu(t') - \left(\frac{\epsilon_n(\vec{k},t') - \mu(t')}{T(t')} \right) \nabla T(t') \right) \right] dt'. \end{aligned} \tag{3.83}$$

Still, the time integral of Eq. (3.83) is non–trivial. Based on an expansion of Eq. (3.66) we can approximate a change in wave vector $\vec{k}(t)$ from modifications due to $\vec{E}(t)$ as

$$\vec{k}(t) \approx \vec{k}(0) - \frac{e\vec{E}(t)t}{\hbar}. \tag{3.84}$$

The main contribution to the integrand in Eq. (3.83) appears at $t \approx \tau$. Based on the fact that $\tau \propto 10$ fs [9] (Chap. 1) the modifications to $\vec{k}(0)$ is thus rather small. Similarly one can argue that perturbation of the temperature within the same time slit is small. Hence, the time integral can be taken at a constant electric field and temperature gradient. Unfortunately, it is not possible to generate a linear approximation for the perturbation of magnetic fields. However, in the absence of magnetic field, the time dependence of $\vec{k}$ disappears and the time integral is trivial, such that

$$\int_{-\infty}^{t} e^{-(t-t')/\tau_n(\vec{k})} dt' = \tau_n(\vec{k}), \tag{3.85}$$

since t always is positive. The combination of Eq. (3.83) and Eq. (3.85) under the given approximations serve as a base to obtain tractable equations for the transport coefficients. It is important to emphasize that Eq. (3.83) satisfy the Boltzmann equation in Eq. (3.69) and is this a valid solution of the electron transport.

3.4.4 The Boltzmann transport equations

Based on the above derivation of the Boltzmann equation and distribution function we will now develop the specific Boltzmann transport equations. The electrical current density can be defined as

$$\vec{J} = nq\vec{v}, \tag{3.86}$$

where n is the number of charge carriers per volume unit, q is the elementary charge and $\vec{v}$ is the carrier drift velocity. Now, n is determined by the distribution function integrated, while for

electrons, $q = -e$. Since multiple bands can contribute, we need to sum each band index n. This gives

$$\vec{J} = -\frac{e}{\Omega}\sum_n \int g_n(\vec{k})\vec{v}_n(\vec{k})d\vec{k}, \tag{3.87}$$

where Ω is the volume of the $\vec{k}$–space. At this time it is important to notify the reader; one should not be easy minded and believe that Eq. (3.87) is of a general form (e.g. only $\vec{k}$ dependence in the integrand). A significant number of approximations was introduced to reduce the dimension of the distribution function. Generally, it will also depend on $\vec{r}$ and t. The electrical current can also be expressed as

$$\vec{J} = \sigma\vec{E} - \chi\nabla T, \tag{3.88}$$

and similarly for the thermal current

$$\vec{J_Q} = T\chi\vec{E} - \kappa^0\nabla T, \tag{3.89}$$

where σ is known as the electrical conductivity, χ and κ^0 are thermal constants in front of the thermal gradients for the electrical and thermal current, respectively. As usual $\vec{E}$ is the applied electric field and T is the temperature. In general σ, χ and κ_0 are tensors. The Seebeck coefficient α is defined as the ratio between the applied electrical field and the temperature gradient at zero electrical current. More specifically

$$\alpha = \frac{\vec{E}}{\nabla T}, \tag{3.90}$$

which for zero current gives

$$\alpha = \sigma^{-1}\chi, \tag{3.91}$$

from Eq. (3.88). We will now derive expressions for σ and χ. Let us continue to use the simplified distribution function $g_n(\vec{k}, t)$ developed in the last section. By substituting Eq. (3.83) into Eq. (3.87) using Eq. (3.85) and assuming constant (in time) electrical field and temperature gradient we get

$$\vec{J} = \frac{e}{\Omega}\sum_n \int \tau_n(\vec{k})\left(-\frac{\partial f_n(\vec{k})}{\partial\epsilon(\vec{k})}\right)\vec{v}_n(\vec{k})\left(\vec{v}_n(\vec{k})\cdot\left[e\vec{E} + \nabla\mu + \frac{\epsilon(\vec{k})-\mu}{T}\nabla T\right]\right)d\vec{k}. \tag{3.92}$$

Comparing Eq. (3.88) and Eq. (3.92) reveal that the electrical conductivity can be written as

$$\sigma = \frac{e^2}{\Omega}\sum_n \int \tau_n(\vec{k})\left(-\frac{\partial f_n(\vec{k})}{\partial\epsilon(\vec{k})}\right)\vec{v}_n(\vec{k})\vec{v}_n(\vec{k})d\vec{k}. \tag{3.93}$$

We determine χ in the same way, looking at the component in front of the temperature gradient in Eq. (3.92). This gives

$$\chi = \frac{e}{T\Omega}\sum_n \int \tau_n(\vec{k})(\varepsilon(\vec{k})-\mu)\left(-\frac{\partial f_n(\vec{k})}{\partial\epsilon(\vec{k})}\right)\vec{v}_n(\vec{k})\vec{v}_n(\vec{k})d\vec{k}. \tag{3.94}$$

The electronic part of the thermal conductivity, κ^e, relates the thermal current and the thermal gradient when no electrical current is flowing, thus

$$\vec{J}_Q = -\kappa^e \nabla T. \tag{3.95}$$

Under these conditions we see from Eq. (3.88) that

$$\vec{E} = \sigma^{-1}\chi\nabla T, \tag{3.96}$$

substituting this into Eq. (3.89) yields

$$\kappa^e = \kappa^0 - T\chi\sigma^{-1}\chi. \tag{3.97}$$

In general σ, χ and κ_0 are tensors. This complicates the transport coefficients because the product of the velocities cannot be calculated as a direct vector product. If we now index σ and χ we get

$$\sigma_{ij} = \frac{e^2}{\Omega}\sum_n \int \tau_n(\vec{k})\left(-\frac{\partial f_n(\vec{k})}{\partial\epsilon(\vec{k})}\right)v_i^n v_j^n d\vec{k}, \tag{3.98}$$

and

$$\chi_{ij} = \frac{e}{T\Omega}\sum_n \int \tau_n(\vec{k})(\epsilon(\vec{k})-\mu(T))\left(-\frac{\partial f_n(\vec{k})}{\partial\epsilon(\vec{k})}\right)v_i^n v_j^n d\vec{k}, \tag{3.99}$$

where $v_i^n = v_n(k_i)$. We can now use Eq. (3.91), which with indexing becomes

$$\alpha_{kl} = (\sigma^{-1})_{ki}\chi_{il}. \tag{3.100}$$

We can also obtain a similar expression for κ^0, see [9] (Chap. 13), such that

$$\kappa^0_{ij} = \frac{1}{e^2T\Omega}\sum_n \int \tau_n(\vec{k})(\epsilon(\vec{k})-\mu(T))^2\left(-\frac{\partial f_n(\vec{k})}{\partial\epsilon(\vec{k})}\right)v_i^n v_j^n d\vec{k}. \tag{3.101}$$

The electron part of the thermal conductivity can also be indexed as

$$\kappa^e_{pl} = \kappa^0_{pl} - T\chi_{pk}\alpha_{kl}. \tag{3.102}$$

By including a magnetic field (the complete Lorentz force, instead of just $\vec{E}$) we obtain the Hall coefficient. The derivation is not worked out here as it is vastly similar to the derivation for the electrical conductivity, except for additional complications due to the cross product between the velocity and the magnetic field. For a short derivation, see Ref. [49] (Sec. 52.2). On a tensor form it can be written as

$$\sigma_{ijp} = \frac{e^3}{\Omega}\sum_n \int \tau_n(\vec{k})^2 \xi_{pqr}\left(-\frac{\partial f_n(\vec{k})}{\partial \epsilon(\vec{k})}\right) v_i^n v_r^n M^{-1}_{jq} d\vec{k}, \tag{3.103}$$

where ξ represents the Levi–Civita symbol and M^{-1} the inverse effective mass tensor defined as

$$(M^n)^{-1}_{ij} = v_i^n v_j^n. \tag{3.104}$$

Using this, the Hall coefficient can be calculated using

$$R_{klm} = (\sigma^{-1})_{il}\sigma_{ijm}(\sigma^{-1})_{kj.} \tag{3.105}$$

3.4.5 Implementation

To calculate Eq. (3.98), (3.99), (3.101) and (3.103) we need to determine the band structure $\epsilon(n, \vec{k})$ and the velocities $\vec{v}(n, \vec{k})$. The band structure can be obtained through any reliable band structure code. For this work we choose the density functional theory code Vienna *Ab–initio* Simulation Package (VASP) [50, 51, 52, 53, 13] due to its excellent efficiency and frequent updates published by the developers. Additional calculations have to be made to determine the velocities. We will now develop a closed expression for the electron velocity.

Electron velocity

The change of the Hamiltonian operator of an electron in a periodic crystal upon a change in crystal momentum $\vec{k}$, for Bloch states $\psi_n = e^{i\vec{k}\cdot\vec{r}}u_{n,\vec{k}}(\vec{r})$, can be written as

$$H_{\vec{k}+\vec{q}} = H_{\vec{k}} + \frac{\hbar^2}{m}\vec{q}\cdot\left(\frac{1}{i}\nabla + \vec{k}\right) + \frac{\hbar^2}{2m}q^2, \tag{3.106}$$

where $H_{\vec{k}} = \frac{\hbar^2}{2m}\nabla^2$ and $\vec{q}$ is the momentum change. If we now do second order perturbation theory and take $H_{\vec{k}}$ as the base and the rest as $H_{\vec{q}}$ we get

$$\epsilon_n(\vec{k}+\vec{q}) = \epsilon_n^0(\vec{k}) + \langle\psi_n|H_{\vec{q}}|\psi_n\rangle + \sum_{n'\neq n}\frac{|\langle\psi_{n'}|H_{\vec{q}}|\psi_n\rangle|^2}{\epsilon_n^0(\vec{k})-\epsilon_{n'}^0(\vec{k})} + \dots \tag{3.107}$$

A small change $\vec{q}$ can be Taylor expanded as

$$\epsilon_n(\vec{k}+\vec{q}) = \epsilon_n(\vec{k}) + \sum_i \frac{\partial\epsilon_n(\vec{k})}{\partial k_i}q_i + \frac{1}{2}\sum_{ij}\frac{\partial^2\epsilon_n(\vec{k})}{\partial k_i k_j}q_i q_j + O(q^3) \tag{3.108}$$

By substituting Eq. (3.106) into Eq. (3.107), keeping only linear order $\vec{q}$ terms we get

$$\epsilon_n(\vec{k}) = \epsilon_n^0(\vec{k}) + \frac{\hbar^2}{m}\sum_i \langle u_{n,\vec{k}}(\vec{r})|\left(\frac{1}{i}\nabla+\vec{k}\right)|u_{n,\vec{k}}(\vec{r})\rangle q_i + O(q^2) \tag{3.109}$$

If we now compare Eq. (3.108) and Eq. (3.109) we see that

$$\sum_i \frac{\partial\epsilon_n(\vec{k})}{\partial k_i}q_i = \frac{\hbar^2}{m}\sum_i \langle u_{n,\vec{k}}(\vec{r})|\left(\frac{1}{i}\nabla+\vec{k}\right)|u_{n,\vec{k}}(\vec{r})\rangle q_i, \tag{3.110}$$

going back to $|\psi_n\rangle$ this gives

$$\frac{\partial\epsilon_n(\vec{k})}{\partial\vec{k}} = \frac{\hbar^2}{im}\langle\psi_n|\nabla|\psi_n\rangle. \tag{3.111}$$

Now, since $\vec{v} = d\vec{r}/dt$ we know that

$$\frac{d\vec{r}}{dt} = \frac{1}{i\hbar}[\vec{r},H], \tag{3.112}$$

$$= \frac{\vec{p}}{m}, \tag{3.113}$$

$$= \frac{\hbar}{im}\nabla, \tag{3.114}$$

hence

$$\vec{v}_n(\vec{k}) = \frac{1}{\hbar}\frac{\partial\epsilon_n(\vec{k})}{\partial\vec{k}}, \tag{3.115}$$

thus taking the derivative with respect to the $\vec{k}$–points of the band structure. To obtain this derivative, there are three approaches; the first is based on a brute force numerical derivative, the second utilizes a Fourier transform. These approaches do however need a very dense and well

converged band structure. In addition they are semi–classical in nature. The third alternative is to calculate the velocity directly from Eq. (3.111). This however requires knowledge of the single particle wave functions.

The latter procedure include quantum effects and even though they are calculated on a perturbative level, the advantage over the semi–classical approach is evident. Unfortunately such implementation also yield a strong $\vec{k}$–point dependence similar to the one seen in the implementation of the calculation of the optical properties. In fact, Eq. (3.111) is usually implemented to calculate optical properties, and thus, it should have the same convergence character. To calculate Eq. (3.111), $\nabla|\psi_n\rangle$ must be evaluated. The implementation of this is quite cumbersome for plane wave codes, but is explained in great detail in the literature [54]. We chose to use this method as it has the potential of being more accurate and as a bonus it was already implemented in VASP5.X. The only additional implementation needed in VASP, was the extraction and printout of these velocities.

Two approaches to the Boltzmann transport integrals and sums

Unfortunately, implementing Eq. (3.98), (3.99), (3.101) and (3.103) is still not straightforward. We will now dedicate a few paragraphs to discuss this and highlight the major problems.

Imagine a single discretely sampled band. To calculate the electrical conductivity we sum over each sampled $\vec{k}$–point leading to a divergence (if the band is completely flat) in the conductivity when the $\vec{k}$–points approach the continuous limit. Now, imagine a band formed as a delta function in $\vec{k}$–space. This band would at most attribute one $\vec{k}$–point for a given energy slit. A combination of such bands at a given constant sampling, will yield a non–conserved $\vec{k}$–point counting, and will thus without normalization yield a band behaviour that is unphysical. Obviously, neither of these examples are physical but they illustrate the need of a normalization, or weighted mean of the sampled $\vec{k}$–points. Even though Eq. (3.87) is expressed as a integral over $\vec{k}$–space and a sum over n, it is thus more practical in a band structure approach to define it as an energy integral with a weighted mean of the counted energy points in each energy slit.

$$\vec{J} = -\frac{e}{\Omega}\int \vec{J}(\epsilon)d\epsilon, \tag{3.116}$$

where

$$\vec{J}(\epsilon) = \frac{1}{n_k}\sum_{n,\vec{k}} g(\vec{k})\vec{v}_n(\vec{k})\frac{\delta(\epsilon-\epsilon_n(\vec{k}))}{d\epsilon}, \tag{3.117}$$

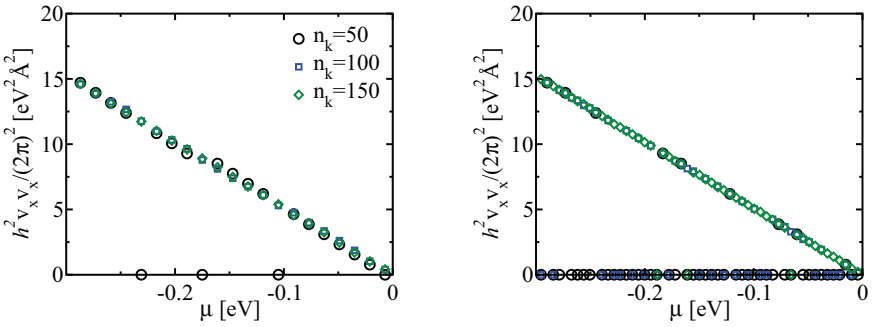
where n_k is the number of sampled $\vec{k}$–points. This approach obtain the average $\vec{J}(\epsilon)$ for each

energy slit and have the advantage of being intuitive. In addition, some closed expressions of electron scattering are energy dependent [48] (Sec. 3.1, specifically Tab. 3.2), making them relatively easy to implement in such an approach due to the explicit energy dependence.

At first, the implementation of Eq. (3.116) seems trivial. But counting the sampled points is not easy to accomplish due to the smeared out Fermi function, especially important at elevated temperatures. A suitable cutoff would be able to partly solve this, but the process is still relatively slow. This is of particular importance when we need to include many $\vec{k}$–points and bands in the calculation.

However, for the purpose of calculating the Seebeck and Hall coefficient, Eq. (3.87) can be used on a complete summation form (sum over both n and $\vec{k}$) since the resolution normalization is inherent. Such implementation is a straight forward calculation of the sums and the ratio of σ and χ (on tensor form) and we will not discuss the practical aspects of it here.

In Fig. 3.8 we show the velocity tensor in one spatial direction as a function of band depth (chemical potential) for two different slit sizes. The velocity tensor have been generated from a single parabolic band with $m^* = m_e$ (free electron) with band maximum at $\mu = 0$. For the



(a) 22 slits (b) 110 slits

Figure 3.8: (a) $\hbar^2 v_{xx} = \partial^2 \epsilon_n(\vec{k})/\partial k_x^2$ for different $\vec{k}$–point samplings n_k for 22 energy slits. (b) Same, but for 110 energy slits (every other point is removed for clarity).

largest slit size (see Fig. 3.8(a)), some points are dropped for $n_k = 50$, while this improves for $n_k = 100$ and $n_k = 150$. In addition, wiggles are observed, originating from slit boundary effects. For the densest $\vec{k}$–point grid the results are acceptable, and approach the linear limit. The same situation is illustrated for the smaller slit size (see Fig. 3.8(b)), but as expected more points are skipped since the $\vec{k}$–point sampling is the same. Again, the dropped points decrease with increasing $\vec{k}$–point grid size. If we now were to include different bands with different curvatures (and hence velocities) an inconsistency appears. Hence, it is important to choose the

right compromise between slit size and $\vec{k}$–point sampling, or to introduce additional routines to fill in the missing points. For the band engineering section (Sec. 3.5) later we manually matched the band curvatures and the slit and $\vec{k}$–point resolution to avoid any significant problems of wiggles or lack of point pickup. For the articles where the Seebeck coefficient have been calculated we used the direct sum approach due to its simpleness and lack of these problems.

Based on random point interpolations routines such as Kriging (see for example the book by A. G. Journel and Ch. J. Huijbregts[55] for a discussion of this technique) or modified Shepard [56] it is possible to adapt this code to utilize on the go interpolations in each slit, thus significantly increasing the flexibility and redundancy by removing the aforementioned problems.

Calculation of the chemical potential

We also want to calculate the temperature dependence of the transport coefficients, but this is not straight forward due to the temperature dependence of the chemical potential. However, we can circumvent the problem by assuming that we have a closed system with a constant number of electrons N and utilize the Fermi function such that

$$\begin{aligned} N &= \frac{1}{\Omega}\sum_n \int 2 f_n(\vec{k}) d\vec{k}, \\ &= \frac{1}{\Omega}\sum_n \int \frac{2}{e^{(\epsilon(\vec{k})-\mu(T))/k_b T}+1} d\vec{k}. \end{aligned} \tag{3.118}$$

At a given temperature this equation can be solved for the chemical potential by a least square approach, or similarly for a given temperature and particle number. We choose the Newton–Raphson method [2] (Sec. 9.4) as it is simple and sufficiently fast and accurate for our purpose.

The important $\vec{k}$–point convergence

The energy slit method needs interpolation routines to be stable and tractable for a wide range of systems. We shall see that the calculation of the transport properties by utilizing the Boltzmann equations is burdened by additional severe $\vec{k}$–point convergence problems related to the calculation of the band structure and the velocities. To illustrate these problems we will use the band structure and velocities for the $SrZn_2Sb_2$ compound. The band structure and velocities are extracted from density functional calculations in VASP based on a PBE+U [57, 58] approach (see paper IV for calculation details). Even though this fails to produce the correct overall band structure (see paper V), the point of this section is to show problems related to

the $\vec{k}$–point convergence. We only present data for the Seebeck coefficient, but the convergence problems are similar for the separate σ and χ tensors. The tensor trace of the Seebeck coefficient $(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3$ is used for the convergence tests. All post–interpolation of the band structure and velocities have been performed by the modified Sheppard interpolation [56]. This interpolation technique has the ability of extracting random points and was chosen with the future possibilities of such randomness in mind (e.g. interpolation along paths and scattered domains).

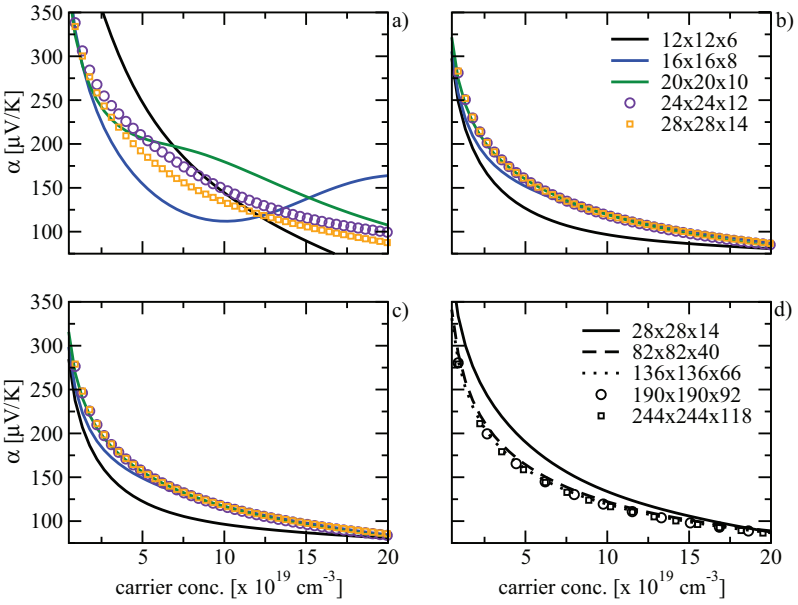


Figure 3.9: Convergence tests of the Seebeck coefficient as a function of carrier concentration for $SrZn_2Sb_2$ at 300 K. In (a), (b) and (c) the convergence as a function of input $\vec{k}$–point grid (see legend) is given for three different interpolation schemes. (a) fills none, (b) four and (c) eight points between the original points in the input grid. The $28\times28\times14$ input $\vec{k}$–point grid is interpolated (see legend for final grid density) and compared in (c).

The severeness of the $\vec{k}$–point convergence is evident from Fig. 3.9, where the Seebeck coefficient versus carrier concentration is investigated for different $\vec{k}$–point grid densities. All transport calculations were done at a fixed temperature of 300 K. In Fig. 3.9(a) no interpolation is performed and, as the input grid becomes increasingly dense, there is trace of a slow convergence. Post-interpolating the output band structure grids by four intermediate points between the originals (Fig. 3.9(b)) immediately improves the convergence. Doubling the point filling in Fig. 3.9(c) yield very similar results and thus indicate a clear convergence. The same behaviour

Table 3.1: $SrZn_2Sb_2$ $\vec{k}$–points used for the Seebeck coefficient convergence tests. Input grid points in the Brillouin zone (BZ) and the irreducible Brillouin zone (IBZ) are also listed.

Original			Interpolated (intermediate points)			
Grid	IBZ	BZ	2	4	6	8
12×12×6	100	864	34×34×16	56×56×26	78×78×36	100×100×46
16×16×8	213	2048	46×46×22	76×76×36	106×106×50	136×136×64
20×20×10	396	4000	58×58×28	96×96×46	134×134×64	172×172×82
24×24×12	667	6912	70×70×34	116×116×56	162×162×78	208×208×100
28×28×14	1030	10976	82×82×40	136×136×66	190×190×92	244×244×118

is seen in Fig. 3.9(d), where a 28×28×14 band structure grid is interpolated with no, two, four, six and then eight intermediate points. The difference between the first two is substantial, but then the results start to converge, suggesting four intermediate points as an acceptable compromise. It is worth pointing out from Table 3.1 that, to obtain such convergence a fairly aggressive scheme is needed. More precisely, even though we already fill the Brillouin zone with 10976 $\vec{k}$–points, we need to post–interpolate this to 1.2 million $\vec{k}$–points, or roughly an order of magnitude increase of points in one direction to observe convergence. For the most aggressive interpolation scheme a convergence within 1–2 μV/K is obtained.

Reasonable convergence was possible for the Seebeck coefficient as a function of the carrier concentration. But, one should not instantly think that this convergence is universal for e.g. the temperature dependence of the Seebeck coefficient for the same band structure grids and interpolation schemes. In Fig. 3.10 a similar convergence test is given for the Seebeck coefficient as a function of the temperature. For the non–interpolated band structure grids in Fig. 3.10(a), no overall convergence can be observed, although, it is signs of converging Seebeck coefficient at higher temperatures. Interpolating the band structure grids in Fig. 3.10 with four intermediate points yields clear signs of convergence. This convergence further improves for the most aggressive eight intermediate point interpolation scheme (see Fig. 3.10(c) and 3.10(d)). Acceptable convergence within 5–10 μV/K is obtained for the most aggressive interpolation scheme, although notice that this is almost an order of magnitude larger than what was possible with the carrier concentration tests under the same conditions.

For readers unfamiliar with density functional calculations and the computational resources needed to do such dense calculations should know that for the 28×28×14 calculations it was necessary to use 32 processors with roughly 15GB of memory on each processor. And this is part of the reason why we choose a simple system like the $SrZn_2Sb_2$ compound. Obtaining such dense grids for larger systems, like the 14–1–11 compounds or the β–Zn_4Sb_3 compound is

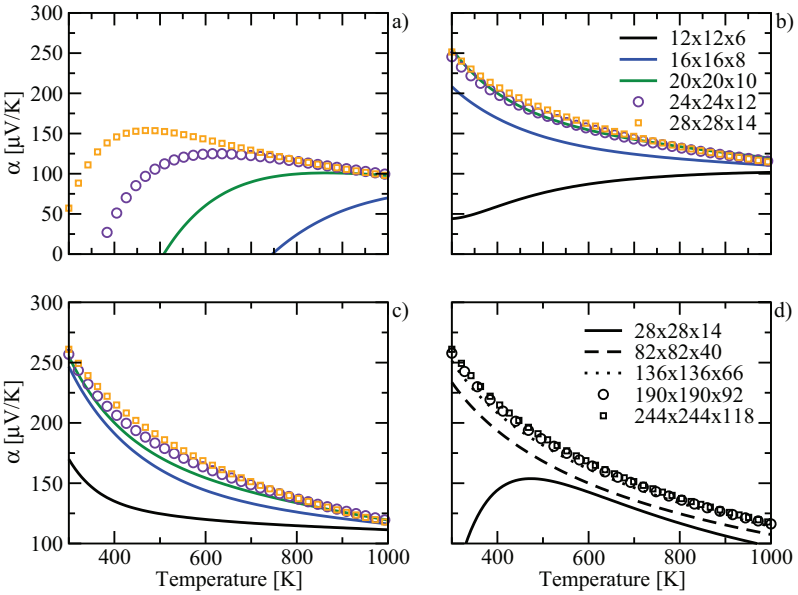


Figure 3.10: Convergence tests of the Seebeck coefficient as a function of temperature for $SrZn_2Sb_2$. Fermi level is set at the top of the valence band. See Fig. 3.9 for additional details.

simply not possible. However, even though a dense grid is needed for the $SrZn_2Sb_2$ compound, case–by–case studies of the convergence is crucial for the system of interest and a warning to the reader is hereby given to carefully assess published results not discussing this.

It would be very beneficial to solve the aforementioned convergence problems, such that this kind of transport calculations could be done for larger systems and maybe in the future also for systems incorporating defects or other imperfections in the crystal.

A possible solution is a grid shift approach [59] currently under development in VASP5.X. The main idea is based on the irreducible Brillouin zone (IBZ) and shifts of this. In principle we would like to utilize the IBZ wherever possible due to the performance benefits. However, figuring out an interpolation (or extrapolation) scheme that operates on the IBZ and implement this into running density functional codes is not trivial. The method under development works by shifting the whole IBZ as shown in Fig. 3.11 and utilize an interpolation and extrapolation of the plane wave coefficients based on the original IBZ. The new shifted IBZ is then diagonalized and the velocities (or other observables) are calculated. Contrary to the current approach of generating the densest possible input grid and applying post–interpolation, this approach has the potential of being more accurate (interpolation/extrapolation directly from the plane wave

coefficients) and it is expected to give a significant improvement of speed and memory usage.

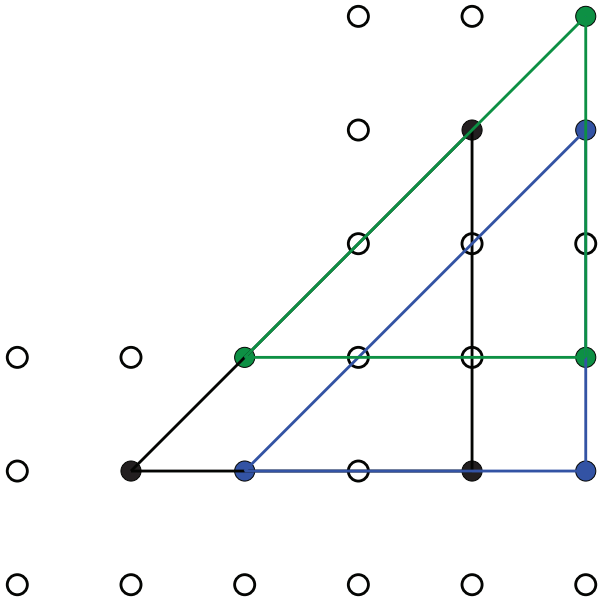


Figure 3.11: Outline of an interpolation approach that utilize irreducible Brillouin zone (IBZ) grid–shifts. A cubic $3\times3\times1$ (one to keep the two dimensionality of the figure) IBZ is shown for simplicity. The black points and lines represent the original IBZ, while the blue and green represent two different shifts. When all shifts are done, the IBZ is up sampled by two additional intermediate points. Increasing the number of shifts, while decreasing the shift length would increase the up sampling.

Even though this approach was initially intended for transport calculations it turns out it could also be used to generate dense density of states, band structures and improve the slow convergence of the optical properties (similar to the transport, due to the calculation of the momentum operator operating on the plane wave functions).

To illustrate some preliminary results of this approach, the density of states of the elemental Al is given in Fig. 3.12. Four base $\vec{k}$–point grids are calculated; $7\times7\times7$, $11\times11\times11$, $33\times33\times33$ and $55\times55\times55$, respectively. Then, the base $11\times11\times11$ grid is interpolated to a $33\times33\times33$ and a $55\times55\times55$ grid. The agreement between the interpolated $33\times33\times33$ and the base $33\times33\times33$ grid is very good with the additional benefit of a significantly decreased computational time as can be seen from Table 3.2. The calculation of the velocities currently suffer from a slight increase in computational time (but significant memory savings) due to dynamic symmetry lockups

Table 3.2: Al density of states calculation times over a set of different $\vec{k}$–point grids. The base 11×11×11 grid is interpolated to 33×33×33 and 55×55×55 $\vec{k}$–points using shifts of the irreducible Brillouin zone and interpolation/extrapolation of the plane wave coefficients.

Grid	Time [s]
11×11×11	28
33×33×33	470
55×55×55	1974
11×11×11→33×33×33	100
11×11×11→55×55×55	189

needed for the explicit use of the IBZ. However, a linear to constant lockup with respect to the number of $\vec{k}$–points is expected when these routines are rewritten to utilize hashed $\vec{k}$–points. This would yield a constant lockup time with some extra overhead (collision checks) compared to the currently unhashed routines. Consequently an order of magnitude speedup is expected. The idea of using hash lockups is not new [60] (Chap. 6) and we will not discuss this further in this thesis.

3.5 Band engineering and the impact on the transport properties

Now that a rigorous calculation of the transport properties can be done, let us investigate how different band structure properties affect the electrical conductivity and the Seebeck coefficient. To do this in a consistent manner we generate fictitious bands based on the modified single parabolic band approximation

$$\epsilon(\vec{k}) = \frac{\hbar^2 k^2}{2m^*} + \Delta\epsilon, \tag{3.119}$$

where m^* is the effective mass and $\Delta\epsilon$ is an energy shift constant. The effective mass will set the curvature of the band. The effective mass is replaced by $m^* = am_e$ for simplicity. The constant a then determines the curvature of the band. A negative a generates a valence band, while a positive a yields a conduction band.

Before generating the bands it is worthwhile to investigate the different parameters that determine the electrical conductivity and the Seebeck coefficient. The electrical conductivity is fully determined by Eq. (3.98), while the Seebeck coefficient uses a combination of Eq. (3.98) and Eq. (3.99). While the relaxation time is of importance we will disregard it until further notice. The two remaining properties in the conductivity tensor are then the velocity tensor and the

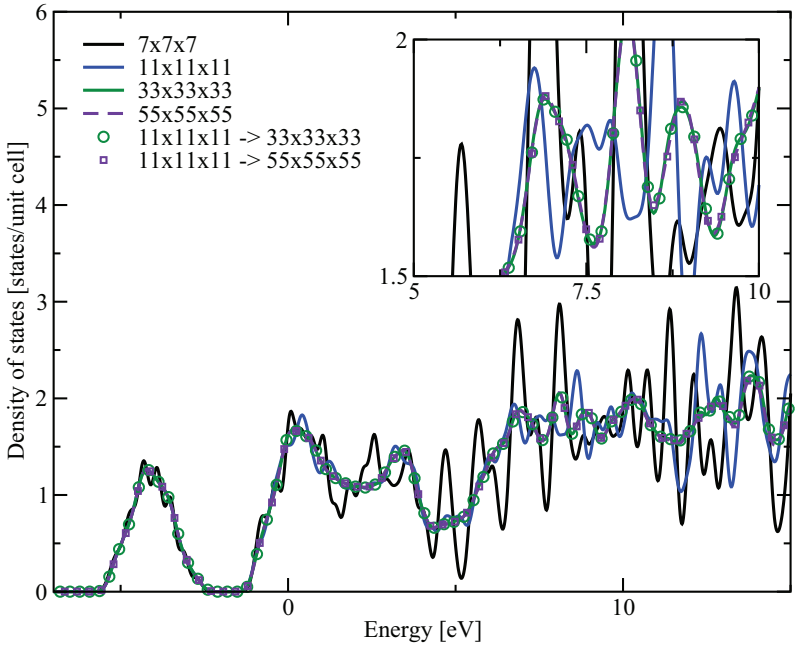


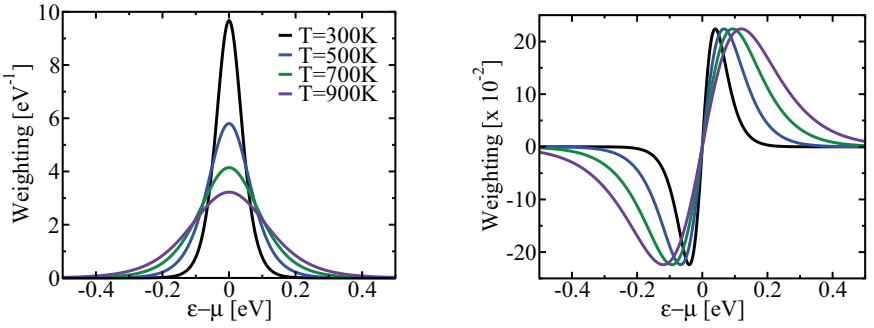
Figure 3.12: Density of states for elemental Al for different sets of $\vec{k}$–point grids and interpolation schemes. The $11\times11\times11$ grid has been interpolated to a $33\times33\times33$ (green circles) and a $55\times55\times55$ (indigo squares) grids to illustrate the potential of the shift approach. A Gaussian smearing of 0.2 eV was added to the calculated results.

derivative of the equilibrium Fermi function. The velocities are dependent on the band structure (and quantum effects beyond what can be observed in the band structure), while the band pickup from the Fermi function derivative is only dependent on the temperature and the difference between the energy and the chemical potential.

To illustrate the band pickup at different temperatures we plot $\partial f/\partial\epsilon$ as a function of $\epsilon-\mu$ in Fig. 3.13(a), where the Fermi energy is set at zero. The derivative distribution is centered at μ, thus any shift of μ would shift the band pickup accordingly. The width of the distribution increases with temperature as is well known from the Fermi function. For 300 K there is hardly any band pickup outside $\epsilon-\mu\pm0.2$ eV, while there for 900K is significant pickup outside this region, dropping to zero at approximately $\epsilon-\mu\pm0.5$ eV. This tells us that the semi–classical conductivity is determined (at all interesting temperatures) by states with energies in a region of ±0.5 eV around the chemical potential. Due to the temperature dependent width of the band pickup, the band velocities close to $\epsilon=\mu$ are dominating at low temperatures, while at elevated

temperatures, velocities away from $\epsilon = \mu$ becomes increasingly more important.

For the χ tensor the behavior is a bit more complicated. The main difference between the pickup factor for χ and σ is an additional linear $\epsilon - \mu$ weighting factor. In Fig. 3.13(b), where $(\epsilon - \mu)\partial f/\partial\epsilon$ is plotted as a function of $\epsilon - \mu$, this is clearly seen. The pickup as a function of temperature is similar to that of the electrical conductivity, but is now multiplied by a linear term. The weighting amplitude is constant for all temperatures and the range pickup is slightly wider, especially at elevated temperatures. But the region of ± 0.5 eV around μ is still valid. Because of the integral over the band pickup and velocity, the asymmetry of the χ band pickup leads to the requirement of band asymmetry around μ to obtain a large χ and thus a large Seebeck coefficient. If the bands are mirrored around μ within a large portion of the pickup region, χ would approach zero and thus yield a small Seebeck coefficient. If we assume that the velocity is constant (e.g.



(a) Electrical conductivity band pickup

(b) χ band pickup

Figure 3.13: (a) Band pickup for the σ (conductivity) tensor at four different temperatures. (b) Similar, but for the χ tensor. See (a) for legend.

isotropic structure with linear bands), the Seebeck coefficient is independent of the velocity. For the single band approximation we see that

$$v = \frac{\hbar k}{m^*}, \tag{3.120}$$

$$= \sqrt{\frac{2\epsilon}{m^*}}. \tag{3.121}$$

Thus, for the single band approximation, the band pickup is multiplied by an additional $\sqrt{\epsilon}$ factor.

Due to the nature of the single band approximation it can immediately be recognized that the electrical conductivity is easily calculated if the effective mass is known. It also turns out that the Seebeck coefficient is independent of the effective mass. Substituting the single parabolic

velocity Eq. (3.121) into Eq. (3.117) (and assuming a single band) the electrical conductivity σ and the thermal coefficient χ are obtained for the single parabolic band approximation as follows

$$\sigma_s = \frac{e^2}{\Omega k_b T}\sqrt{\frac{1}{2\hbar^2 m^*}}\int_0^\infty \tau(\epsilon)\frac{\epsilon^{1/2}}{\cosh((\epsilon-\mu)/k_bT)+1}d\epsilon, \tag{3.122}$$

$$\chi_s = \frac{e}{\Omega k_b T^2}\sqrt{\frac{1}{2\hbar^2 m^*}}\int_0^\infty \tau(\epsilon)\frac{\epsilon^{1/2}(\epsilon-\mu)}{\cosh((\epsilon-\mu)/k_bT)+1}d\epsilon, \tag{3.123}$$

which from Eq. (3.91) gives the Seebeck coefficient

$$\alpha_s = \frac{1}{eT}\left(\frac{\int_0^\infty \tau(\epsilon)\frac{\epsilon^{3/2}}{\cosh((\epsilon-\mu)/k_bT)+1}d\epsilon}{\int_0^\infty \tau(\epsilon)\frac{\epsilon^{1/2}}{\cosh((\epsilon-\mu)/k_bT)+1}d\epsilon}-\mu\right), \tag{3.124}$$

where $\tau(\epsilon)$ is an energy dependent scattering time. The way of maximizing the Seebeck coefficient from the single band approximation is if $\mu \ll \epsilon$ or $\mu \gg \epsilon$. Another option would be to make the denominator (conductivity) approach zero, which is satisfied for the same conditions. Hence, the longer away from the band μ is placed, the larger the Seebeck coefficient. This is somewhat fictitious since the conductivity would be zero and no current would be able to flow, but it illustrates the demand for asymmetry around μ to obtain the highest possible Seebeck coefficient. The conductivity is increased by increasing the velocity (similar to a decrease of the effective mass, or an increase of the band curvature).

To go beyond the single band approximation (but still in the parabolic approximation) we have to sum over the additional bands in Eq. (3.122), (3.123) and (3.125). We will now leave the single band approximation and perform band engineering to illustrate how the electrical conductivity and Seebeck coefficient varies over a wide range of band combinations. All transport calculations in the rest of this thesis are based on the solution of the Boltzmann equation under the approximations that lead to the Boltzmann transport equations in Sec. 3.4.4. These equations are solved using the energy slit implementations presented in section Sec. 3.4.5. The band curvature is determined by the constant a, while the energy shift is determined by $\Delta\epsilon$. Unless otherwise stated, the relaxation time have been taken to be independent of both energy and $\vec{k}$–points and fixed at 1 fs [9] (Chap. 1). For convenience the Fermi level ϵ_f is fixed at $\epsilon = 0$, where $\mu = \epsilon_\mathrm{f}$ at equilibrium.

3.5.1 Chemical potential or carrier concentration

Studies of the transport properties are usually given as a function of carrier concentration or chemical potential. The carrier concentration is readily available from Hall measurements and it is thus more common to use carrier concentration representation when presenting experimental data. However, in this section we give most of the data versus the chemical potential due to its simpleness. But, for the constant scattering and energy filtered power factor we also include the carrier concentration representation. The relation between chemical potential and carrier concentration n_c for a single parabolic band [9] (Chap. 28) is given by

$$n_c(\mu, m^*, T) = \frac{\sqrt{2}}{\pi^2}\left(\frac{m^* k_b T}{\hbar^2}\right)^{3/2} \int_0^\infty \epsilon^{1/2} f(\epsilon, \mu, T) d\epsilon. \tag{3.125}$$

where f is the equilibrium Fermi distribution. This is calculated for illustrative purposes for a set of band curvatures in Fig. 3.14. It is important to emphasize that a sum over Eq. (3.126) is needed if multiple bands are used.

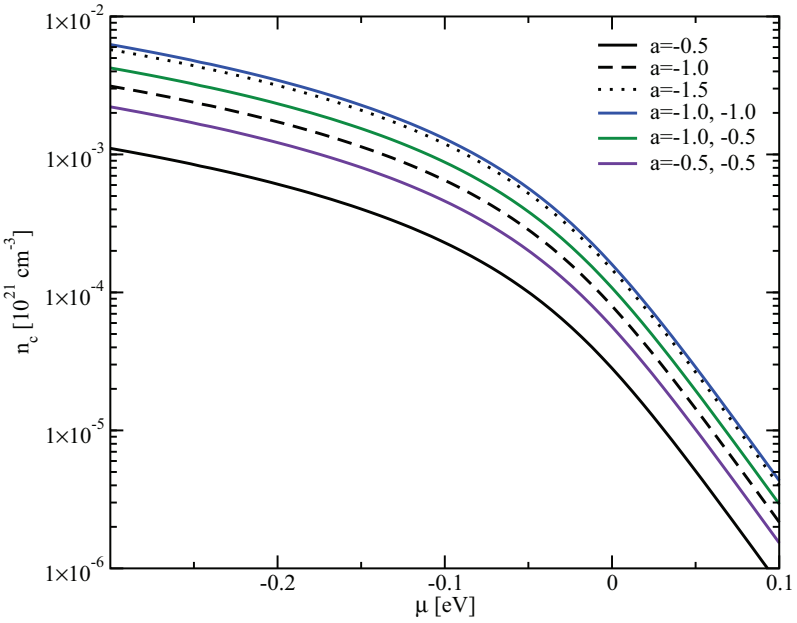
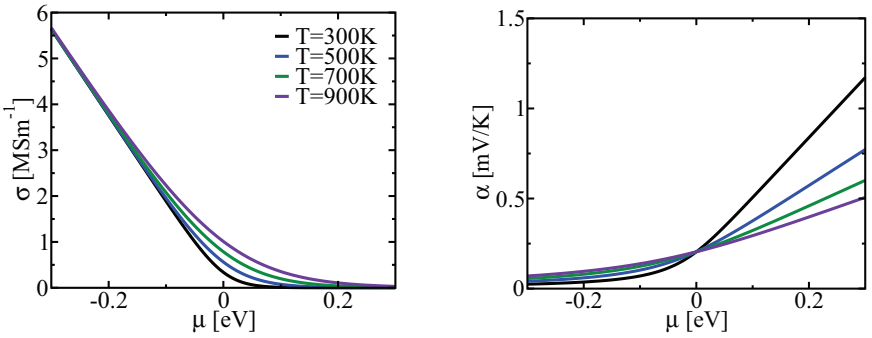


Figure 3.14: Carrier concentration as a function of chemical potential for a set of different effective masses (band combinations). The absolute value of the effective mass parameter a is used in the calculation of Eq. (3.126). The temperature is fixed at 300 K.

3.5.2 Temperature

To explicitly illustrate the temperature dependence of the band pickup we calculate the electrical conductivity and the Seebeck coefficient as a function of μ for different temperatures. The valence band is fixed at $a = -1, \Delta\epsilon = 0$ and thus resemble the free electron mass for simplicity.



(a) Electrical conductivity (b) Seebeck coefficient

Figure 3.15: (a) Electrical conductivity and (b) Seebeck coefficient as a function of chemical potential at four different temperatures. See (a) for legend.

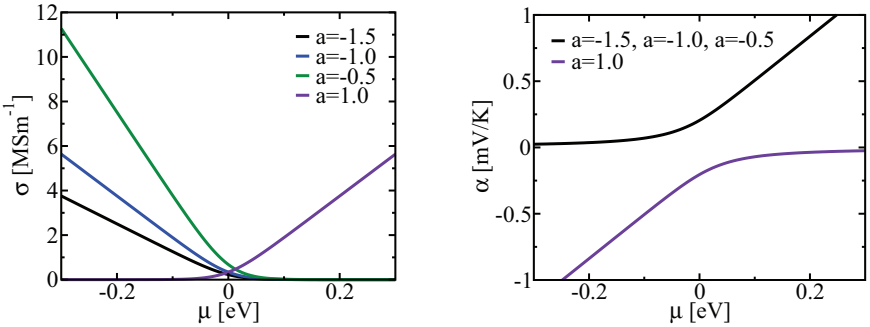
From Fig. 3.15 the increased band pickup is clearly illustrated at elevated temperatures. The fact that the Seebeck coefficient in Fig. 3.15(b) is largest when the band pickup decreases is visible and due to the width of the pickup (and thus smaller conductivity), the Seebeck coefficient at room temperature at $\mu = 0.3$ eV is largest of all four. At higher temperature a decreased slope of the Seebeck coefficient as a function of μ is observed. This is caused by different band pickup between σ and χ as illustrated in Fig. 3.13(a) and Fig. 3.13(b). This behaviour is cancelled between χ and σ at $\mu = 0$, where the asymmetry is equal for all temperatures. However, for $\mu \neq 0$ the symmetry changes due to the band pickup and the curves start to separate. To summarize;

- $\mu > 0$; lower temperatures give higher Seebeck coefficients with the largest slope due to a small denominator (conductivity). The electrical conductivity is small due to the lack of band pickup (except close to $\mu = 0$).
- $\mu < 0$; higher temperatures give higher Seebeck coefficients. The electrical conductivity at different temperatures becomes similar when μ enter the deeper regions of the band due to the band pickup around μ.

3.5.3 Curvature

In the previous section we underlined temperature dependent features of the electrical conductivity and Seebeck coefficient. We now fix the temperature (in this case at 300 K) and investigate the curvature of single parabolic bands (both valence and conduction), but postpone the discussion of the combination until a later section.

The electrical conductivity is proportional to the band curvature. From Fig. 3.16(a), we see that bands with larger velocity results in an increased electrical conductivity. The difference of a valence band and conduction band is also shown for the band curvatures $a = -1.0$ and $a = 1.0$, respectively. The difference is just what we would expect, symmetric around $\mu = 0$.



(a) Electrical conductivity (b) Seebeck coefficient

Figure 3.16: (a) Electrical conductivity and (b) Seebeck coefficient as a function of chemical potential for four different band curvatures, determined by the constant a. The temperature is fixed at 300 K.

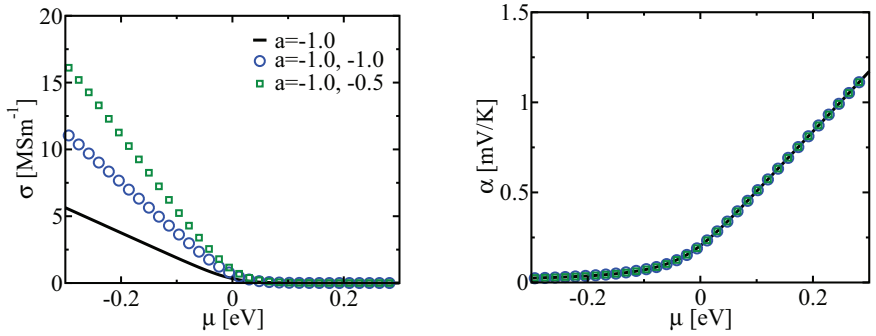
In Fig. 3.16(b) we again (also Eq. (3.125)) see that the Seebeck coefficient is independent of a change in the band curvature (effective mass), except for the obvious valence to conduction band changes. Negative Seebeck coefficients are a signature of positive band curvature, such that n–doped materials will exhibit negative Seebeck values, while the opposite is true for p–doped materials. Hence;

- Increasing the band curvature increases the conductivity, while the sign of the curvature determines whether the conductivity increases for positive or negative μ.
- The Seebeck coefficient does not depend on the band curvature within the single parabolic band approximation. A sign change of the curvature changes the sign and the increasing/decreasing behaviour versus μ of the Seebeck coefficient.

3.5.4 Degeneracy and reciprocal shift

It is now time to determine how the electrical conductivity and the Seebeck coefficient respond to band degeneracy and shifts in reciprocal space. From the implementation of the energy slits it should be obvious that any reciprocal shift does not matter as long as the band edges are retained within the pickup region. The changes of the electrical conductivity due to degeneracy are shown in Fig. 3.17(a), while they for the Seebeck coefficient are shown in Fig. 3.17(b). The difference between a single band and two bands is evident for the conductivity. By adding a second band with a smaller effective mass, the electrical conductivity increases as is expected from the additive behaviour of the bands and the direct dependence of the effective mass.

Any additional band (as long as it is not energy shifted with respect to the others) will however not change the Seebeck coefficient. This is due to the cancellation effect between χ and σ for a constant change in band curvature. Similar to the electrical conductivity, there is also no change if we apply a reciprocal shift to one of the bands. The cancellation effect is a feature of the parabolic band approximation and can also be seen from Eq. (3.125). To summarize this section;



(a) Electrical conductivity (b) Seebeck coefficient

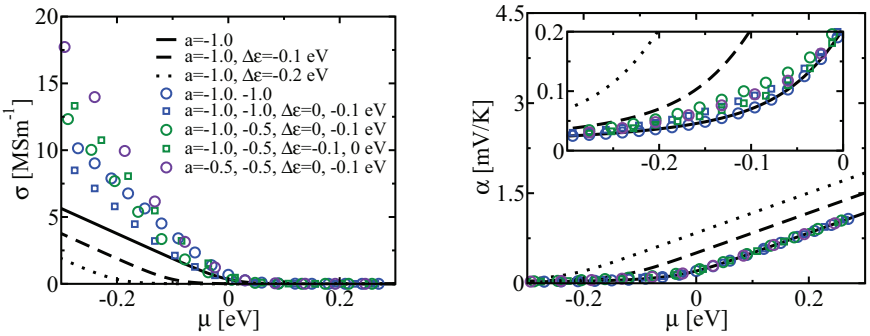
Figure 3.17: (a) Electrical conductivity and (b) Seebeck coefficient as a function of chemical potential, compared for one band, two bands similar curvatures and a two band case with different curvatures. The temperature is fixed at 300 K. See (a) for legend.

- The electrical conductivity is dependent on the band degeneracy. A doubling of the degeneracy accordingly doubles the electrical conductivity if the bands are kept equal.
- The Seebeck coefficient is independent of the degeneracy in the parabolic approximation (as long as the bands are not energy shifted). Deviations in real systems are expected due

to the failure of the parabolic approximation.

3.5.5 Energy shift

We have so far not investigated the effect of energy shifts of the generated bands. To investigate this we use sets of single parabolic bands with different effective masses and energy shifts. In Fig. 3.18(a) we see that a negative energy shift lowers the onset for band pickup and thus lowers the onset of electrical conductivity. By introducing an additional band with the same curvature, but energy shifted from the original, the electrical conductivity increases accordingly as we saw in the previous discussion, although with a modification due to the later band pickup of the shifted band. Again, changes in the effective mass follows the previous discussions. However, for the Seebeck coefficient in Fig. 3.18(b) we see that the largest difference is introduced by the single band energy shifts. As previously discussed the Seebeck coefficient is largest where the conductivity drops to zero. It is also important to notice that it is the first band that dominate the Seebeck coefficient. The introduction of additional bands deeper in energy only modify the Seebeck coefficient in the band pickup region of the additional band. An interesting fact to note is the subtle difference in the Seebeck coefficient between the $a = -1.0, -0.5, \Delta\epsilon = 0, -0.1$ eV and $a = -1.0, -0.5, \Delta\epsilon = -0.1, 0$ eV calculations (see inset in Fig. 3.18(b)). The latter will have a slightly higher electrical conductivity in the region from -0.1 to 0 eV, which yields a lower Seebeck coefficient close to and in this region. The impact of energy shifts can be summarized



(a) Electrical conductivity (b) Seebeck coefficient

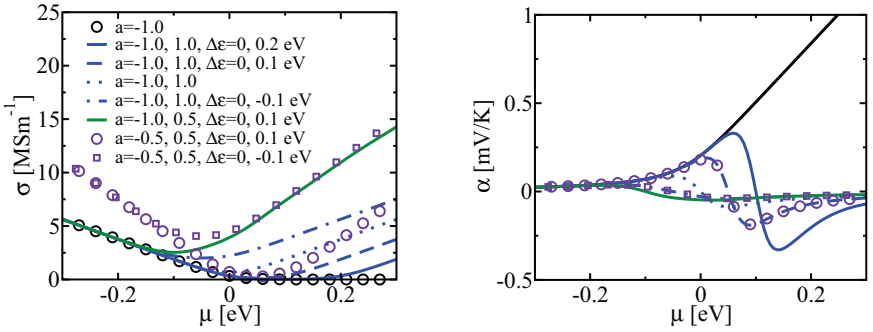
Figure 3.18: (a) Electrical conductivity and (b) Seebeck coefficient as a function of chemical potential, compared for one band with and without energy shifts and two bands with different curvatures and energy shifts. The temperature is fixed at 300 K and the energy shift is given by $\Delta\epsilon$. See (a) for legend.

as follows;

- The electrical conductivity as a function of chemical potential changes drastically between the different energy shifts. The largest conductivity is found for the band combination with the lowest effective mass (largest band curvature).
- The Seebeck coefficient is highly dependent on the first band. Additional bands introduce small (but still important) modifications in the respective pickup regions.

3.5.6 Conduction band

Even though the previous discussions are important, they are somewhat artificial. In most realistic situations, particularly at high temperate and in narrow band gap systems, conduction bands will contribute. In Fig. 3.19(a) the electrical conductivity is given as a function of chemical potential for different combinations of valence and conduction bands. The previous discussion of the change in the conductivity as the effective mass changes still holds, although with additional contributions on the conduction side. When the conduction band is pulled down into the valence region to generate a metal–like system there is a non–zero region of the conductivity in the overlap region between the valence and conduction bands. This even occurs for bands separated by a small gap (because of the distribution at 300 K). The Seebeck coefficient is similarly



(a) Electrical conductivity (b) Seebeck coefficient

Figure 3.19: (a) Electrical conductivity and (b) Seebeck coefficient as function of chemical potential, compared for different combinations of a (negative is valence, positive is conduction) and energy shifts. The temperature is fixed at 300 K and the energy shift is given by $\Delta\epsilon$.

illustrated in Fig. 3.19(b). Here, we start to see many of the commonly known behaviours of the Seebeck coefficient for semi–conductors. For example, introducing a conduction band with the

same effective mass as the valence band, separated by a small band gap gives a transition region in the middle of the gap. In this region, the Seebeck coefficient changes sign from positive to negative when the conduction band starts to dominate (or opposite for the valence band). A negative Seebeck coefficient is associated with n–type behaviour, while p–type is positive. When the gap decreases, the maximum and minimum Seebeck coefficient decreases due to the lack of asymmetry around μ. Again, we see that an effective mass change is ignored if there is an equal change between the conduction and valence band. A decrease of the effective mass in the conduction band with respect to the valence band (or vice versa) results in a decrease of the Seebeck coefficient due to the increased conductivity. To summarize;

- Introducing conduction bands does not modify the original discussion about the electrical conductivity. However, due to the change in sign of the effective mass, we also get conductivity contribution for positive μ. At a sufficiently wide gap (depending on the temperature), there is a region of zero conductivity in the gap region. This disappears when the thermal band pickup enclose the gap (e.g. high temperatures and/or small gap). A clear metallic behaviour is then observed.

- The effect of conduction bands on the Seebeck coefficient is significant. A negative Seebeck coefficient is to be associated with n–doped systems, while a p–type is to be associated with the positive counterpart. The transition between a positive and negative Seebeck coefficient takes place in the middle of the gap region if the valence and conduction bands are symmetric around the gap with a sign change. A small band gap accordingly give a small Seebeck coefficient due to less antisymmetry around μ.

3.5.7 Modifications due to electron scattering

Due to the non–cancelling relaxation time τ from σ, it is somewhat artificial to fix the scattering time without considering basic electron scattering. If we simply set τ to a constant, we get a conductivity that is increasing with increasing number of bands. Hence, the conductivity would keep rising monotonically as the temperature is increased. For real systems this is highly unphysical and we need to introduce a parametrization for the relaxation time to account for at least the majority of effects.

The two dominating scattering mechanisms in doped semi–conductors are scattering by acoustical phonons and impurities. Simple models for both are given by Fistul [48] (Tab. 3.2). In this thesis we have implemented a factor similar to the one given in Eq. (3.2.22) of Fistul [48].

To summarize, Fistul defines

$$\tau^{a,i} = \tau_0^a \frac{|\epsilon|^{3/2}}{|\epsilon|^2 + \xi^2}, \tag{3.126}$$

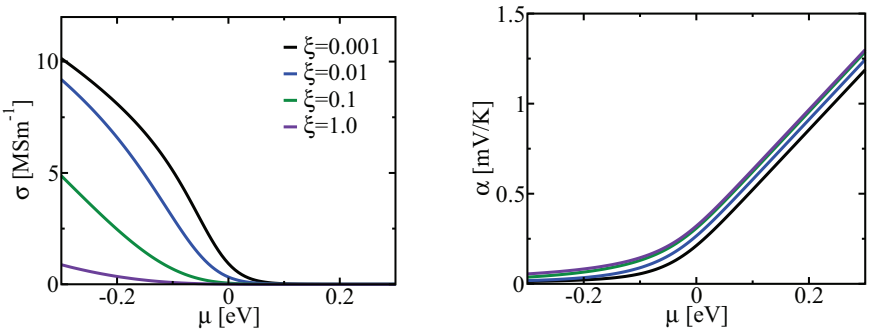
where

$$\xi^2 = \frac{\tau_0^a}{\tau_0^i}, \tag{3.127}$$

as a relaxation time that contains both scattering by acoustical vibrations and impurities through τ_0^a and τ_0^i, respectively. Here, ξ^2 gives the ratio between the acoustical and impurity contributions. The prefactor τ_0^a is given as

$$\tau_0^a = \frac{9}{4\sqrt{2}} \frac{\hbar^4 \omega^2 M}{C^2 a^2 (m^* k_b T)^{3/2}}, \tag{3.128}$$

where C is a Bloch constant, ξ the lattice parameter, ω the sound velocity, a the lattice parameter and M the atomic mass. We take τ_0^a to be constant in energy. Hence, $\tau^{a,i}$ is a function of energy and ξ^2 only. Substituting Eq. (3.127) into the transport equations, we see that τ_0^a cancels for the Seebeck and Hall coefficients, while it is retained for the electrical and thermal conductivity. The cancellation is convenient due to the complex problem of calculating τ_0^a from *ab–initio* principles. Common relaxation times are in the order of tenths of femtoseconds [9] (Chap. 1). To simplify the comparisons we will fix τ_0^a at 1 fs.



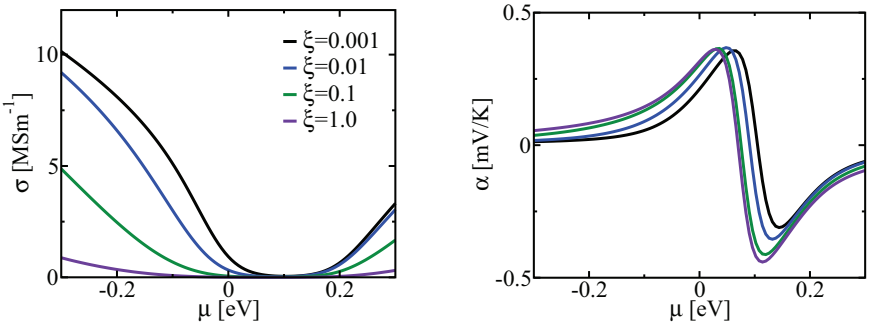
(a) Electrical conductivity (b) Seebeck coefficient

Figure 3.20: (a) Electrical conductivity and (b) Seebeck coefficient as function of chemical potential, compared for different combinations of the scattering parameter ξ. Only one valence band with $a = -1$ is included in the calculation. The temperature is fixed at 300 K. See (a) for legend.

A large ξ means that $\tau_0^a >> \tau_0^i$, thus primarily ionic scattering. For a small ξ the opposite is true. The combined approach of Eq. (3.129) is difficult to compare to a constant scattering time used until now. This is due to the energy dependence (that modify the band pickup) and

quantitative comparisons with the previous results will not be done.

In Fig. 3.20 we compare results given different scattering parameters ξ for the electrical conductivity and Seebeck coefficient. One valence band with $a = -1$ has been included in the calculations. A scattering parameter of $\xi = 1$ clearly yields the largest dampening of the conductivity and hence the largest Seebeck coefficient. Also, the curvature of the electrical conductivity changes quite drastically from $\xi = 1$ to $\xi = 0.001$, the transition from combined ionic and acoustic to predominantly acoustic scattering. It is interesting to note the rather small change of the Seebeck coefficient (compared to the electrical conductivity), which is a consequence of the cancellation between the χ tensor and the electrical conductivity. There is however a constant shift of the Seebeck coefficient for the different scattering parameters.



(a) Electrical conductivity (b) Seebeck coefficient

Figure 3.21: (a) Electrical conductivity and (b) Seebeck coefficient as function of chemical potential, compared for different combinations of the scattering parameter ξ. One valence band and one conduction band with $a = -1$, $\Delta\epsilon = 0$ eV and $a = 1$, $\Delta\epsilon = 0.1$ eV are included in the calculation. The temperature is fixed at 300 K. See (a) for legend.

Introducing a shifted conduction band in Fig. 3.21 yields similar results, although with additional features. Due to the conduction band, a transition from n–type to p–type and the sign change in the Seebeck coefficient are observed. As before, the smallest conductivity gives the largest Seebeck coefficient.

We can briefly summarize the scattering section in a few points;

- Scattering due to ionic impurities and acoustic vibrations scattering alters the electrical conductivity significantly.

- In this simple model, ionic scattering is the strongest electron scatter process.

- Changes in the Seebeck coefficient due to scattering are less significant than the changes observed for the electrical conductivity. The behaviour as a function of chemical potential (and carrier concentration) is fairly similar, but a constant shift is applied.

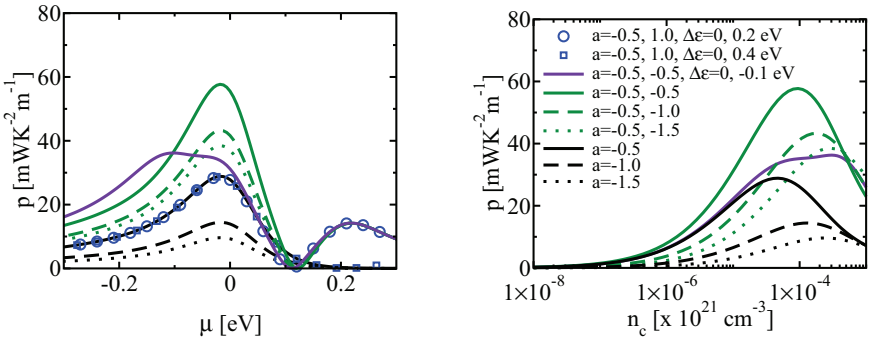
- The difference between the one and two band model is similar to the previous discussion disregarding scattering mechanisms.

3.5.8 Power factor and optimization of electronic properties

It should from the proceeding discussions be apparent that we can not simultaneously obtain a large electrical conductivity and a Seebeck coefficient. To obtain a large zT value, this is precisely what is needed. The combination enters in the power factor $p = \alpha^2\sigma$ in Eq. (2.4). Hence, for thermoelectric optimization, knowledge about this factor is crucial. We will in this section discuss the compromise between the electrical conductivity and the Seebeck coefficient and how to obtain the "optimal" power factor.

Power factor with constant scattering

Let us first consider the power factor in absence of energy dependent scattering with a constant scattering time of 1 fs [9] (Chap. 1). In Fig. 3.22(a) we give the power factor as a function of the chemical potential for different combinations of bands. The maximum power factor is clearly obtained for bands with the smallest effective mass (largest conductivity). The introduction of an additional valence band with $a = -0.5$ increases the power factor, but when a is increased the power factor approach the limit of the original band (compare green solid, dashed and dotted lines to the black solid line in Fig. 3.22). Similar behaviour is seen for the power factor as a function of carrier concentration (only shown for valence bands) in Fig. 3.22(b). This is a very important observation that confirms that a small effective mass is essential for large power factors. The power factor is thus strongly dependent on the electrical conductivity and optimization of this should be a primary concern as long as a decent Seebeck coefficient is met in a given system. The addition of a conduction band introduces a second peak and has a relatively small effect on the peak associated with the valence band. Another interesting observation is the shift of the valence peak if one valence band is downshifted in energy (compare solid purple and green line in Fig. 3.22). This is accompanied by an increase of the peak width at a cost of the height and indicates that the range of operational temperature might be increased by filling the valence (or conduction if n–type is wanted) region with several similar, but energy shifted bands.



(a) As a function of chemical potential (b) As a function of carrier concentration

Figure 3.22: (a) The power factor and as a function of the chemical potential and (b) carrier concentration, compared for different combinations of bands given by the curvature a and the energy shift $\Delta\epsilon$. The temperature is fixed at 300 K. See (b) for legend.

Power factor with energy dependent scattering

In the last section we discovered that the power factor was strongly dependent on the conductivity (or more directly, the effective mass). Due the sensitivity of the electrical conductivity to scattering it is expected that the power factor will change significantly when electron scattering is involved. In Fig. 3.23(a), 3.23(b) and 3.23(c) the power factor for three different scattering parameters $\xi = 0.01, 0.1$ and 1.0 is given.

The result of ionic and acoustical scattering ($\xi = 1.0$) can be understood from Fig. 3.21(a). The slope and magnitude of the electrical conductivity are significantly changed, while the Seebeck coefficient remains virtually unchanged. Thus, as a consequence, the peak of the power factor is downshifted and decreased due to the slope and magnitude changes of the electrical conductivity. A more subtle change is the similarity between the energy shifted calculation (solid indigo line) and the non–shifted calculation (solid green line) for larger ξ. One might expect these changes to be similar to the constant scattering calculations in Fig. 3.22. However, due to the modification in the band pickup (additional energy terms from $\tau^{a,i}$) there is less change when the scattering model is applied.

When acoustical scattering becomes more important ($\xi = 0.1$) we see that the shift and decrease of the valence peak is smaller. And, when the acoustic scattering is predominant ($\xi = 0.01$) we retain similar behaviour to the constant scattering time approach, except for the modification of the band pickup, causing a different behaviour of the shifted band. An increase of the power factor is seen due to the increase of the total scattering time. All these changes can

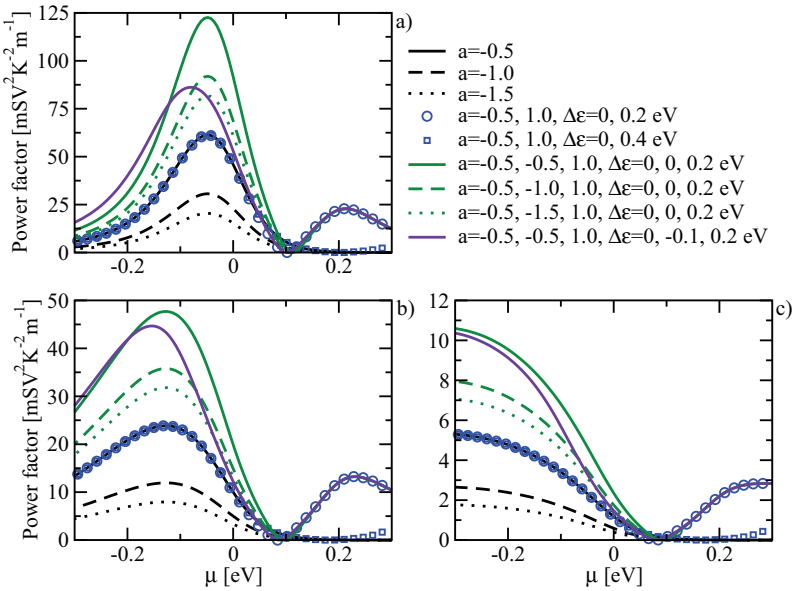


Figure 3.23: The power factor as a function of the chemical potential, compared for different combinations of bands given by the curvature a and the energy shift $\Delta\epsilon$. The temperature is fixed at 300 K. Scattering parameters are $\xi = 0.01, 0.1, 1.0$ for (a), (b) and (c), respectively.

be explained from the electrical conductivity in Fig. 3.21(a).

3.5.9 Figure–of–merit

The previous section discussed band dependencies of the power factor. A strong correlation between the velocity (electrical conductivity) and power factor was shown. However, the figure–of–merit, zT also depends on the electrical and lattice thermal conductivities. From Eq. (2.4) we can separate these thermal conductivities such that

$$zT = \frac{\alpha^2\sigma}{\kappa_e + \kappa_l}T, \tag{3.129}$$

where κ_e and κ_l are the electrical and lattice thermal conductivities, respectively. By rewriting Eq. (3.130) we get

$$zT = \frac{\alpha^2\sigma}{\left(1 + \frac{\kappa_l}{\kappa_e}\right)\kappa_e}T. \tag{3.130}$$

Through Wiedemann-Franz law [61]

$$\kappa_e = L_0 T \sigma, \tag{3.131}$$

where L_0 is the Sommerfield Lorentz constant [9] (Chap. 1), we obtain

$$zT = \frac{\alpha^2}{L_0 \left(1 + \frac{\kappa_L}{\kappa_e}\right)}. \tag{3.132}$$

Generally, L_0 should be replaced by a Lorentz function L, which can is calculated from the ratio between the electrical conductivity and the electrical thermal conductivity tensors. This is clearly shown in Fig. 3.24 for different band combinations. Unless the valence bands are shifted with respect to each other, they do not alter the one band behaviour of the Lorentz function. A small difference between the energy shifted (purple diamonds) and the energy non-shifted (green circles) Lorentz functions can be seen close to the shifted band onset. Adding a conduction band pulls the Lorentz values down towards the L_0 constant. In the metallic (or highly doped) limit, the Lorentz function approaches the commonly used L_0 constant.

To illustrate the band curvature dependence of the figure–of–merit, we calculate the Lorentz tensor and use its trace in Eq. (3.131) and give the figure–of–merit as a function of chemical potential and lattice thermal conductivity in Fig. 3.25 and 3.26. From this it is clear that a large band curvature and low lattice thermal conductivity are needed to obtain a large zT value. Additional bands increase the zT value further and if they are shifted with respect to each other, a slight increase of the width of the zT peak at the cost of height is observed (similar to the power factor). A conduction band yields what is expected, another zT peak associated with it, where the separation between the center zT valence and conduction peak is determined by the the band gap. For a sufficiently large gap the single band zT valence peak is retained. However, by pulling the conduction band into the valence region the zT peak decreases drastically even though the band curvature is large. Notice that the zT value is largest right above the valence band edge onset. Unfortunately this is a region which is difficult to reach, but the closer we are to the termination of the band from below (valence bands), the larger the zT value is, demonstrating also the need for asymmetry.

We can thus summarize;

- By lowering the lattice thermal conductivity, the zT value increases.
- Increasing the band curvature and/or degeneracy increases the zT value.
- Asymmetry around μ is necessary for a large zT value and the chemical potential should

be set close to the band onset in the given temperature range.

- The operational temperature might be improved by adding several bands with sufficient spacing so that continues asymmetries exist for decreasing (increasing for n-type) chemical potential.

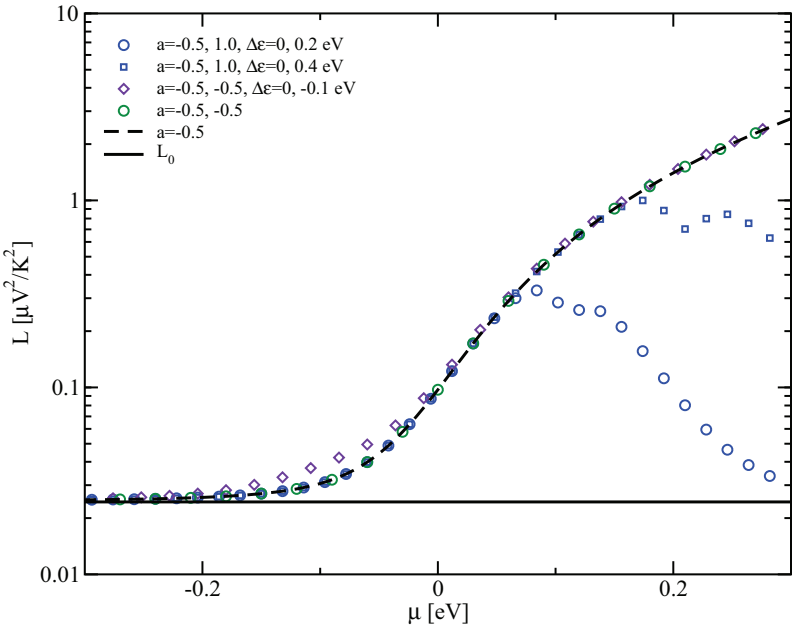


Figure 3.24: Lorentz function for different combinations of bands given by the curvature a and the energy shift $\Delta\epsilon$. Temperature is fixed at 300K. The Sommerfield Lorentz number L_0 is also shown for comparison.

3.5.10 Energy filtering

Albeit this might seem artificial, there is a possible way to increase the power factor beyond the classical approach. What if we introduce an energy barrier that modifies the band pickup? This would increase or decrease the asymmetry around μ and thus modify the Seebeck coefficient significantly. At the same time, the electrical conductivity should also change (within the filtering region), but opposite to the Seebeck coefficient, thus leading to a possible large change in the power factor. Such a barrier could arise from e.g. scattering. For modelling purposes we chose

here a scattering time relaxation to be represented by a Heavyside step function $H(-\epsilon-\epsilon_b)$ times a constant scattering time. By using this function an energy filtered band pickup is generated such that only states with larger energy than the barrier is allowed to be involved in the transport. The reason for choosing such simple function is purely due to implementation reasons and for more realistic calculations it would be beneficial to introduce a more complicated filter function (e.g. at least a Gaussian smeared Heavyside function). We now define the energy dependent scattering time as

$$\tau(\epsilon) = \tau_0 H(|\epsilon| - \epsilon_b), \tag{3.133}$$

where τ_0, ϵ and ϵ_b are the constant relaxation time, the energy and the energy barrier, respectively. The Heavyside step function is defined as

$$H(|\epsilon| - \epsilon_b) = \begin{cases} 0 & |\epsilon| < \epsilon_b \\ \frac{1}{2} & |\epsilon| = \epsilon_b \\ 1 & |\epsilon| > \epsilon_b \end{cases} \tag{3.134}$$

For simplicity we again fix τ_0 at 1 fs [9] (Chap. 1). In Fig. 3.27 we compare the electrical conductivity and the Seebeck coefficient at different energy barriers and different combinations of valence bands. The Seebeck coefficient is significantly enhanced by the introduction of an energy barrier. Increasing the energy barrier further yields an even larger Seebeck coefficient. The electrical conductivity rapidly drops to zero in the region where the Heavyside function is zero, but approaches the unfiltered calculation away from the barrier onset.

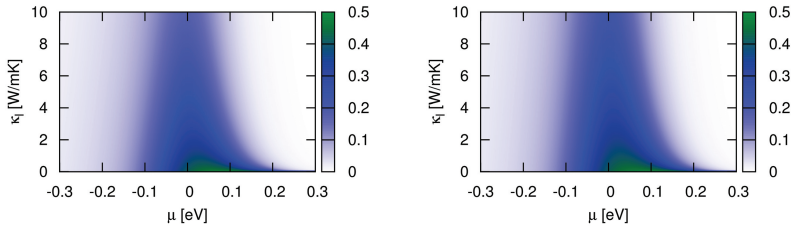
Results of this are apparent when we look at the power factor as a function of chemical potential in Fig. 3.28(a). Close to a doubling of the power factor is observed when we introduce an energy barrier in 0.1 eV steps. This change is dramatic. In Fig. 3.28(b) the power factor as a function of carrier concentration is plotted (only for the valence bands) and it is similarly behaved as the results presented in Fig. 3.28(a).

To illustrate the change in the figure–of–merit due to energy filtering we give the zT value as a function of lattice thermal conductivity and chemical potential in Fig. 3.29 for three different energy filtering schemes, given by the energy barrier ϵ_b. Similarities between the changes due to the energy filtering in the power factor and the zT value are observed. Notice that the energy filtering shifts the zT peak to lower chemical potential, or higher carrier concentration. This is very convenient, since the right size of filtering can tune the zT to the chemical potential that yields the highest band curvature or to chemical potentials that can be reached in real samples.

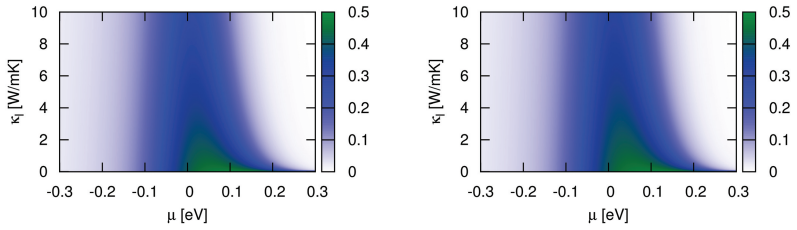
3.5.11 Band structure requirements

In this section we will briefly summarize the band engineering sections by pointing out some key requirements needed of the band structure to be able to obtain a large power factor and figure–of–merit. Based on the previous discussions the reader should consider the following;

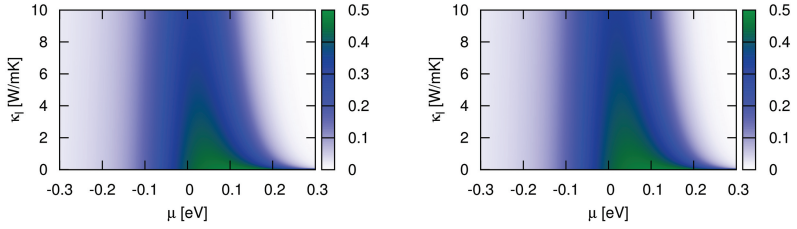
- Low effective masses (high velocity) are important (if not crucial)
- A high level of degeneracy is important (degeneracy in $\vec{k}$–space is equally valid)
- Decent Seebeck coefficient is important (asymmetry around μ)
- Depending on the operational range in temperature or chemical potential there is a compromise between the electrical conductivity and the Seebeck coefficient. If an acceptable Seebeck coefficient is obtained, focus should be shifted towards maximizing the electronic conductivity.
- A wider operational temperature can be obtained (with a compromise of the peak power factor and figure–of–merit value) by shifting (or adding) similar bands with low effective mass downwards (upwards if n–type) in the valence (conduction if n–type) band.
- When the "optimal" power factor is reached and there is little potential for improvement, energy filtering should be introduced to further improve the power factor. If an almost perfect step function can be made, it should be possible to increase the power factor by a factor of 2–4 (and accordingly the zT value).
- Energy filtering can potentially be used to tune the optimal figure–of-merit to a feasible chemical potential and carrier concentration region.



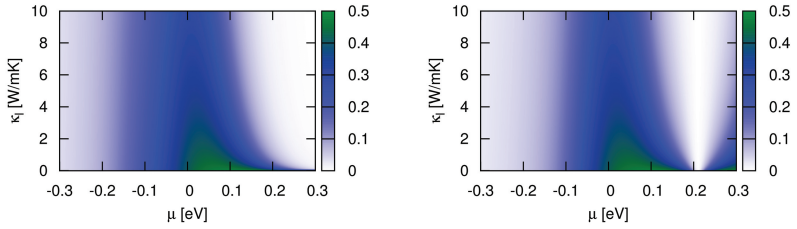
(a) a=-1.5 (b) a=-1.0



(c) a=-0.5 (d) a=-0.5, -1.5

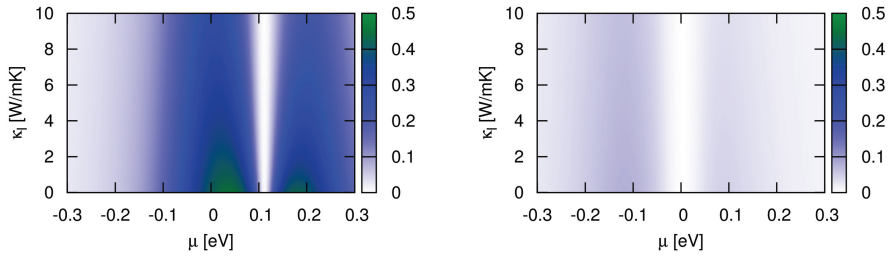


(e) a=-0.5, -1.0 (f) a=-0.5, -0.5



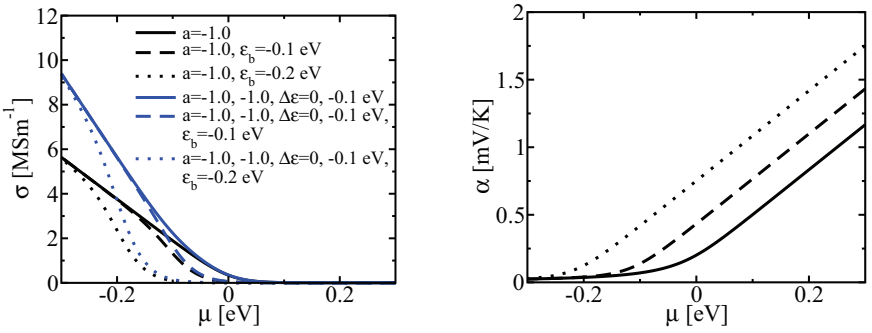
(g) a=-0.5,-0.5, $\Delta\epsilon$=0, -0.1 eV (h) a=-0.5, 1.0, $\Delta\epsilon$=0, 0.4 eV

Figure 3.25: Figure–of–merit, zT, as a function of the chemical potential and lattice thermal conductivity κ_l at different combinations of bands given by the curvatures a and the energy shift $\Delta\epsilon$. The temperature is fixed at 300 K.



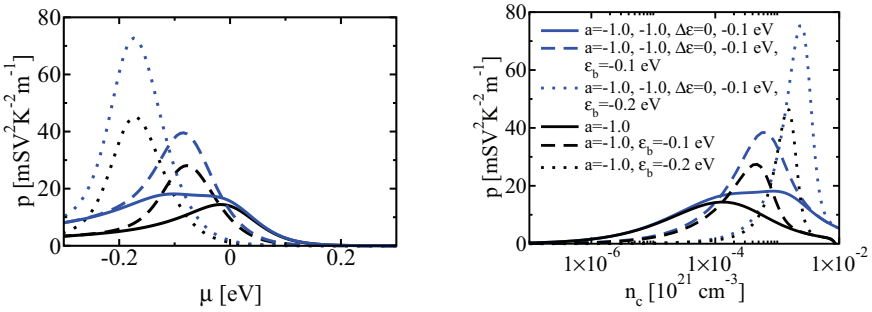
(a) a=-0.5, 1.0, $\Delta\epsilon$=0, 0.2 eV (b) a=-0.5, 1.0, $\Delta\epsilon$=0, -0.1 eV

Figure 3.26: Figure–of–merit, zT, as a function of the chemical potential and lattice thermal conductivity κ_l at different combinations of bands given by the curvatures a and the energy shift $\Delta\epsilon$. The temperature is fixed at 300 K.



(a) Electrical conductivity (b) Seebeck coefficient

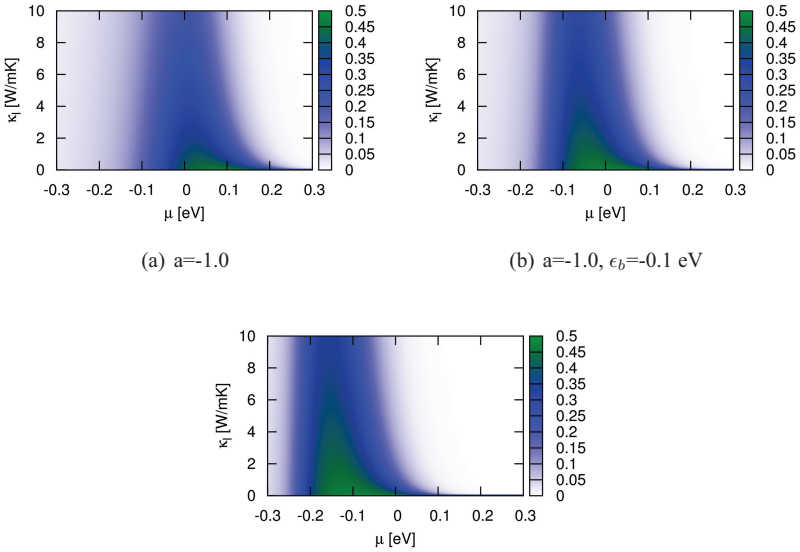
Figure 3.27: (a) Electrical conductivity and (b) Seebeck coefficient as function of chemical potential, compared for different combinations of the energy barrier ϵ_b. One valence band, $a = -1$, is included in the calculation. Data for the two band case is not shown due to the similarities to the one band case (see previous discussions). The temperature is fixed at 300 K and the energy shift is $\Delta\epsilon$. See (a) for legend.



(a) As a function of chemical potential

(b) As a function of carrier concentration

Figure 3.28: (a) Power factor as function of the chemical potential and (b) carrier concentration, compared for different combinations of the energy barrier ϵ_b and bands. The temperature is fixed at 300 K and the energy shift is $\Delta\epsilon$. See (b) for legend.



(a) a=-1.0

(b) a=-1.0, ϵ_b=-0.1 eV

(c) a=-1.0, ϵ_b=-0.2 eV

Figure 3.29: Figure–of–merit, zT, as a function of the chemical potential and lattice thermal conductivity κ_l at $a = -1.0$ and different energy barriers ϵ_b. The temperature is fixed at 300 K.

References

[1] A. I. M. Rae. *Quantum Mechanics*. Taylor & Francis, 2002.

[2] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.

[3] G. Meurant. *The Lanczos and Conjugate Gradient Algorithms: From Theory to Finite Precision Computations (Software, Environments and Tools)*. SIAM, 2006.

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

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

[6] W. Pauli. Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren. *Z. Physik*, 31:765, 1925.

[7] J. C. Slater. Atomic shielding constants. *Phys. Rev.*, 36:57, 1930.

[8] S. F. Boys. Electronic wave functions. A general method of calculation for the stationary states of any molecular system. *Proc. R. Soc. London Ser. A*, 200:542, 1950.

[9] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Thomson Learning, 1976.

[10] D. J. Singh and L. Nordström. *Planewaves, pseudopotentials, and the LAPW method*. 2nd ed., Springer, New York, 2005.

[11] O. K. Andersen. Linear methods in band theory. *Phys. Rev. B*, 12:3060, 1975.

[12] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953, 1994.

[13] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758, 1999.

[14] M. Torrent, F. Jollet, F. Bottin, G. Zérah, and X. Gonze. Implementation of the projector augmented-wave method in the abinit code: Application to the study of iron under pressure. *Comput. Mater. Sci.*, 42:337, 2008.

[15] J. P. Perdew and M. Levy. Physical content of the exact kohn-sham orbital energies: Band gaps and derivative discontinuities. *Phys. Rev. Lett.*, 51:1884, 1983.

[16] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz Jr. Density-functional theory for fractional particle number: Derivative discontinuities of the energy. *Phys. Rev. Lett.*, 49:1691, 1982.

[17] W. Yang, Y. Zhang, and P. W. Ayers. Degenerate ground states and a fractional number of electrons in density and reduced density matrix functional theory. *Phys. Rev. Lett.*, 84:5172, 2000.

[18] A. J. Cohen, P. Mori-Sánchez, and W. Yang. Fractional charge perspective on the band gap in density-functional theory. *Phys. Rev. B*, 77:115123, 2008.

[19] P. Mori-Sánchez, A. J. Cohen, and W. Yang. Buried Ni/Cu(001) interface at the atomic scale. *Phys. Rev. Lett.*, 100:146401, 2008.

[20] A. J. Cohen, P. Mori-Sánchez, and W. Yang. Fractional spins and static correlation error in density functional theory. *J. Chem. Phys.*, 129:121104, 2008.

[21] R. M. Dreizler and E. K. U. Gross. *Density Functional Theory*. Springer, Berlin, 1990.

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

[23] R. Q. Hood, M. Y. Chou, A. J. Williamson, G. Rajagopal, and R. J. Needs. Quantum monte carlo investigation of exchange and correlation in silicon. *Phys. Rev. Lett.*, 78:3350, 1997.

[24] R. Q. Hood, M. Y. Chou, A. J. Williamson, G. Rajagopal, and R. J. Needs. Exchange and correlation in silicon. *Phys. Rev. B*, 57:8972, 1998.

[25] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46:6671, 1992.

[26] C. Fiolhais, F. Nogueira, and M. A.L. Marques. *A primer in Density Functional Theory*. Springer, Berlin, 2003.

[27] V. Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Z. Physik*, 61:126, 1930.

[28] M. Ernzerhof. Construction of the adiabatic connection. *Chem. Phys. Lett.*, 263:499, 1996.

[29] M. Ernzerhof, J. P. Perdew, and K. Burke. Coupling-constant dependence of atomization energies. *Int. J. Quantum Chem.*, 64:285, 1997.

[30] J. P. Perdew, M. Ernzerhof, and K. Burke. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.*, 105:9982, 1996.

[31] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110:6158, 1999.

[32] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Hybrid functionals based on a screened coulomb potential. *J. Chem. Phys.*, 118:8207, 2003.

[33] M. Marsman, J. Paier, and G. Kresse. Hybrid functionals applied to extended systems. *J. Phys.: Condens. Matter*, 20:064201, 2008.

[34] J. Heyd and G. E. Scuseria. Efficient hybrid density functional calculations in solids: Assessment of the heyd-scuseria-ernzerhof screened coulomb hybrid functional. *J. Chem. Phys.*, 121:1187, 2004.

[35] E. N. Brothers, A. F. Izmaylov, J. O. Normand, V. Barone, and G. E. Scuseria. Accurate solid-state band gaps via screened hybrid electronic structure calculations. *J. Chem. Phys.*, 129:011102, 2008.

[36] J. P. Glusker, M. Lewis, and M. Rossi. *Crystal Structure Analysis for Chemists and Biologists*. Wiley-VCH, 1994.

[37] P. Wilkes. *Solid State Theory in Metallurgy*. Cambridge University Press, London, 1973.

[38] L. Wu, Y. Zhu, T. Vogt, H. Su, J. W. Davenport, and J. Taftø. Valence-electron distribution in MgB_2 by accurate diffraction measurements and first-principles calculations. *Phys. Rev. B*, 69:064501, 2004.

[39] A. P. Grosvenor, R. G. Cavell, and A. Mar. X-ray photoelectron spectroscopy study of the skutterudites $LaFe_4Sb_{12}$, $CeFe_4Sb_{12}$, $CoSb_3$, and CoP_3. *Phys. Rev. B.*, 74:125102, 2006.

[40] H. Anno, K. Matsubara, T. Caillat, and J.-P. Fleurial. Valence-band structure of the skutterudite compounds $CoAs_3$, $CoSb_3$, and $RhSb_3$ studied by x-ray photoelectron spectroscopy. *Phys. Rev. B.*, 62:10737, 2000.

[41] J. Graetz, C. C. Ahn, H. Ouyang, P. Rez, and B. Fultz. White lines and d-band occupancy for the 3d transition-metal oxides and lithium transition-metal oxides. *Phys. Rev. B.*, 69:235103, 2004.

[42] Ø. Prytz, J. Taftø, C. C. Ahn, and B. Fultz. Transition metal d-band occupancy in skutterudites studied by electron energy-loss spectroscopy. *Phys. Rev. B.*, 75:125109, 2007.

[43] D. H. Pearson, C. C. Ahn, and B. Fultz. White lines and d-electron occupancies for the 3d and 4d transition metals. *Phys. Rev. B.*, 47:8471, 1993.

[44] J. M. Zuo. Measurements of electron densities in solids: a real-space view of electronic structure and bonding in inorganic crystals. *Rep. Prog. Phys.*, 67:2053, 2004.

[45] J. M. Zuo, M. Kim, M. O'Keeffe, and J. C. H. Spence. Direct observation of d-orbital holes and Cu-Cu bonding in Cu_2O. *Nature*, 401:49, 1999.

[46] G. A. Slack and V. G. Tsoukala. Some properties of semiconducting $IrSb_3$. *J. App. Phys.*, 76:1665, 1994.

[47] C. Uher. Skutterudites: Prospective novel thermoelectrics. *Semiconductors and semimetals*, 69:139, 2001.

[48] V. I. Fistul. *Heavily Doped Semiconductors*. Translated from Russian by Albin Tybulewicz, published in English by Plenum Press, New York, 1969.

[49] R. C. Dorf. *The Electrical Engineering Handbook, Second Edition*. CRC, Boca Raton, 1997.

[50] G. Kresse and J. Hafner. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B*, 48:13115, 1993.

[51] G. Kresse and J. Hafner. Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B*, 49:14251, 1994.

[52] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comp. Mater. Sci.*, 6:15, 1996.

[53] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169, 1996.

[54] M. Gajdos, K. Hummer, G. Kresse, J. Furthmüller, and F. Bechstedt. Linear optical properties in the projector-augmented wave methodology. *Phys. Rev. B*, 73:045112, 2006.

[55] A. G. Journel and C. J. Huijbregts. *Mining geostatistics*. Academic press, London, 1978.

[56] R. J. Renka. Algorithm 661. qshep3d: Quadratic shepard method for trivariate interpolation of scattered data. *ACM Trans. Math. Software*, 14:151, 1988.

[57] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.

[58] A. I. Liechtenstein, V. I. Anisimov, and J. Zaanen. Density-functional theory and strong interactions: Orbital ordering in mott-hubbard insulators. *Phys. Rev. B*, 52:R5467, 1995.

[59] Discussions with Georg Kresse, 2008.

[60] D. Knuth. *The Art of Computer Programming*. Addison-Wesley, Massachusetts, 1998.

[61] G. Wiedemann and R. Franz. Ueber die wärme-leitungsfähigkeit der metalle. *Annalen der Physik*, 165:497, 1853.

Papers

We will here present the papers included in this thesis. All articles are based on density functional calculations and all, except the first paper also include experimental data to confirm or investigate motivating questions of the papers further.

Paper I – Bond analysis of phosphorus skutterudites: elongated lanthanum electron buildup in $LaFe_4P_{12}$

Status: Submitted to Computational Material Science, 29.06.2009

Authors: Espen Flage–Larsen, Ole Martin Løvvik, Øystein Prytz and Johan Taftø

Recently there have been much experimental and theoretical work regarding the correlated motion of the filler ions in the skutterudite structure. We thus wanted to investigate the charge surroundings of the filler ions and intended to perform a strict electron charge analysis to get a better spatial view of the electronic structure close to the filler ions. However, such charge analysis is complicated by the lack of a well defined charge reference.

From a modelling point of view, we have the option of generating a procrystal containing a superposition of free atomic electron densities. Such a procystal contains overlapping densities, but is by definition without bonds. The bonds can then be analysed and visualized from the difference between the crystal and procrystal electron density.

We performed such an analysis for the phosphorus skutterudites CoP_3 and $LaFe_4P_{12}$.

All calculation, visualisations, most of the interpretation and the write up of the article was done by Espen Flage–Larsen, while Ole Martin Løvvik assisted.

Paper II – The influence of exact exchange corrections in van der Waals layered narrow band gap black phosphorus

Status: Submitted to Journal of Physics: Condensed Matter, 11.08.2009

Authors: Øystein Prytz and Espen Flage–Larsen

Based on the development of new hybrid functionals for use in the density functional framework we wanted to do a comparison of these new functionals and how they performed in a combined covalent and van der Waals bond dominated structure. The van der Waals interaction is purely determined by long–range correlation forces and is not included in any of the traditional functionals. However, it is well known that many of the DFT functionals include residue exchange couplings at long range, which effectively introduces an artificial bond in the van der Waals region. By the inclusion of Fock terms (exact exchange) this can be corrected such that the functional is free of van der Waals bonds.

The black phosphorus was chosen as a test case based on previous work on phosphorus in the electron microscope. Also, the black phosphorus contains important covalent bonds that need to be handled correctly before a satisfactory agreement can be met with the experiments. In this work we obtained electron energy loss spectra (EELS) and compared them to real space multiple scattering (RSMS) and density functional theory (DFT) calculations. Even in the RSMS calculations most features except the band gap were accounted for. However, a dispersion in the conduction band was lacking. Similar results were obtained using the standard functionals (PZ–LDA) and (PBE–GGA). When the hybrid functionals (HSE03 and PBE0) were applied a significant dispersion was added together with an opening of a gap. Relaxation of the structure was also done using the hybrid functionals, which gave indications that PBE0 was free of the exchange residue due to a non–converged and drifting inter–layer distance.

This paper is not following the red line of first–principle studies of thermoelectric materials that is the focus of this thesis. However, it is still included to demonstrate the importance of investigating new functionals. In particular, the hybrid functionals will be of importance to many thermoelectric materials that contains d and f states close to the Fermi level (see for example paper V).

Øystein Prytz did the experimental work and the RSMS calculations, while Espen Flage–Larsen did the DFT part. The interpretation of the results and writeup of the article were a joint effort.

Paper III – Electron energy loss spectroscopy of the $L_{2,3}$ edge of phosphorus skutterudites and electronic structure calculations

Status: Published in Physical Review B, 11.08.2009

Authors: Ragnhild Sæterli, Espen Flage–Larsen, Øystein Prytz, Johan Taftø, Knut Marthinsen and Randi Holmestad

Ref: Phys. Rev. B 80, 075109 (2009)

Motivated for the same reasons as in the first paper we wanted to do a combined analysis of electron energy loss spectroscopy (EELS) and density functional calculations (DFT) to further investigate the phosphorus bonds. Experimental investigations of the phosphorus energy loss edges were lacking in the literature and data was acquired and compared to density functional calculations of the conduction density of states. A good agreement between experiments and calculations was found. Real space multiple scattering (RSMS) calculations were also performed to calculate the EELS spectra directly. The agreement of the EELS spectra and RSMS calculation was satisfactory.

The experimental work was done by Ragnhild Sæterli, DFT calculations were performed by Espen Flage–Larsen, and the RSMS calculations by Øystein Prytz. The writing of the article was a collaboration between all three, although Ragnhild did much of the writing of the introduction and the interpretation of the experimental results.

Paper IV – Electronic structure and transport in thermoelectric compounds $A\mathbf{Zn_2Sb_2}$ (A = Sr, Ca, Yb, Eu)

Status: Submitted to Dalton Transactions, 14.07.2009

Authors: Eric S. Toberer, Andrew F. May, Brent Melot, Espen Flage–Larsen and Jeff Snyder

The aim of this invited paper was to summarize previous work on the 122 Zintl compounds and put them into a larger thermoelectric context. In addition the thermoelectric properties of the (up until then undocumented) $SrZn_2Sb_2$ were included and analysed. The results was found to fit well within the 122 context.

Thermoelectric measurements and synthesis were done by Eric Toberer, Andrew May and Jeff Snyder. Density functional and transport calculations were done by Espen Flage–Larsen. Initial results of Brent Melot were reworked and charge analyses were added. The interpretation of the thermoelectric properties was initially done by Eric Toberer and Andrew May with additional interpretations and comments by Espen Flage–Larsen.

Paper V – Valence band study of thermoelectric Zintl $SrZn_2Sb_2$ and $YbZn_2Sb_2$

Status: Submitted to Physical Review B, 13.08.2009

Authors: Espen Flage–Larsen, Spyros Diplas, Eric S. Toberer and Andrew F. May

This paper sought to address several issues regarding the calculations in Paper IV. Due to the need for a band gap of the $SrZn_2Sb_2$, a semi–empirical correction was added to be able to obtain reasonable transport properties. Based on previous discoveries and experience of Paper II, calculations using hybrid functionals were compared to the previous calculations in Paper IV and X–ray photoemission spectra. Excellent corrections of the Zn d peak, Yb f peak and band gap were shown for the hybrid functionals. In addition, there have been some discussions in the literature regarding additional inherent Yb^{3+} in the $YbZn_2Sb_2$. Due to non-existing density of states for the Yb_2O_3 (should nominally contain Yb^{3+}) in the literature, this was calculated. However, these calculations did not reproduce the claimed Yb^{3+} region and it is thus an XPS effect due to Yb $4d$–$4f$ resonances as previously indicated in literature. The $YbZn_2Sb_2$ sample was investigated as a function of *in–situ* Ar^+ scattering before doing an XPS sweep. Significant reduction in the Yb^{3+} peaks was found. However, even after one hour of sputtering (unrealistically long) there were still signs of Yb^{3+} in the sample, either due to Yb_2O_3 in the sample (not only surface) or inherent Yb^{3+}. Structure relaxation was then done for a $1\times1\times3$ cell with three Yb layers and no forced symmetry. No significant deviation from the reported structure was found. In addition, relaxation of the same cell, but with one Yb removed did not yield significant occupancy change to conclude that inherent Yb^{3+} was likely. However, these models were simple and larger superstructure and defect models need to be considered before drawing conclusions.

Density functional calculations were performed by Espen Flage–Larsen, while Spyros Diplas obtained experimental X–ray photoemission spectra. Eric Toberer and Andrew May did the sample synthesis. The writeup and interpretations was mostly done by Espen Flage–Larsen with assistance of Spyros Diplas on the oxygen discussion.