



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„High-throughput GW Calculations“

verfasst von / submitted by

Florian Ellinger, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 876

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Physik UG2002

Betreut von / Supervisor:

Univ.-Prof. Dr. techn. Cesare Franchini

Contents

1	Introduction	3
1.1	Shortcomings of Density Functional Theory	3
1.2	The GW Approach - Correcting DFT Calculations	3
1.2.1	GW Calculations - Conventional Approach	4
1.2.2	The Basis Set Extrapolation Method	4
1.3	Automated Calculations for Large Material Sets	5
2	Theory	7
2.1	Physical Theory	7
2.1.1	Important Terms	7
2.1.2	Electronic Systems	8
2.1.3	Electronic Band Structure	10
2.1.4	Electronic Band Gaps	10
2.2	Computational Theory	11
2.2.1	Important Terms	11
2.2.2	VASP - Vienna Ab initio Simulation Package	13
2.2.3	Density Functional Theory	14
2.2.4	The GW Approach	16
2.2.5	The GW Approximation as Implemented in VASP	18
2.2.6	Converged GW Band Gap	19
2.2.7	Basis Set Correction of the Correlation Energy	21
3	Methods	23
3.1	High Throughput GW Calculations	23
3.1.1	Implementing the Basis Set Extrapolation	23
3.1.2	Automating GW Calculations	25
3.2	Overview of the Implementation	26
3.2.1	Directory Structure	26
3.2.2	Scripts and Programs	28
3.3	Computational Details	34
3.3.1	G0W0 Calculations	35
3.3.2	DFT Calculations	36
3.3.3	Band Structure Calculations	37
3.4	Calculating the Band Gaps	37
3.4.1	Gamma Gap	37
3.4.2	Smallest Indirect Gap	38
3.4.3	Smallest Direct Gap	38

4	Results	39
4.1	Validating the Method	39
4.1.1	Comparison to the Original Basis Set Extrapolation	39
4.1.2	Wrongly Placed or Occupied Bands	42
4.1.3	Accuracy of the Memory Conserving Variant	43
4.1.4	Influence of Parallelization	44
4.2	Automatic GW series	46
4.2.1	Automatically Calculated Band Gaps	46
4.2.2	Comparison to Experimental Gaps	49
4.3	Possibility of a Norm Correction	52
4.4	Band Structures at DFT Level	53
5	Discussion	55
5.1	Challenges during the Project	55
5.1.1	Encountered Problems and Stability	55
5.1.2	Correct K-point Correction	55
5.1.3	Memory Requirements	56
5.1.4	Code Instability due to Large Band Numbers	56
5.2	Feasibility of Automatic GW Calculations	56
5.2.1	Machine Learning Approaches	57
5.2.2	Bringing GW Calculations to AiiDA	57
	Acknowledgments	59
	Bibliography	61
	Appendix	63

1 Introduction

In materials science, computer simulations have become an integral part for the analysis and discovery of novel materials. With increasing computational power, previously very challenging methods are nowadays available broad use. Just like a few years back with the surge of Density Functional Theory (DFT) and its wide application, the GW method is on the brink of introducing a new standard for *ab initio* simulations.

1.1 Shortcomings of Density Functional Theory

After the development of DFT by Hohenberg and Kohn back in 1964 this method has been refined over the years and was long the gold-standard for *ab initio* calculations. Through continuous enhancements, such as implementing better exchange-correlation potentials, today calculated quantities match experimental values to a very high degree.

Although, through the inherently inaccurate treatment of correlation effects some flaws of the method are not completely compensated yet. One of the biggest problems here is the systematic underestimation of a materials fundamental electronic band gap.

Unfortunately, the fundamental gap is detrimental for modern material applications which require semi-conducting or insulating behavior, such as in information technology, electronics, solar cells, energy conversion, and many more. Additionally, the electronic band structure, to which the band gap is of course directly related, defines many other material properties, such as effective mass, Fermi velocity and optical transitions. An inaccurate description of it is a great obstacle in the practical usability of DFT simulations.

1.2 The GW Approach - Correcting DFT Calculations

The GW approximation (G is the one particle Green function, W the screened Coulomb potential) was introduced as a beyond Hartree-Fock method, substituting the static Coulomb potential with a dynamically screened one. For the accurate treatment of correlation effects in materials it is important to find a good estimation of the electronic self-energy. The first full calculation of this property was done by Hedin, back in 1965.[3] The GW approximation as used today was then introduced as the first term of an expansion of this self-energy in powers of the screened Coulomb potential and has found many applications.[4, 5, 6]

Since this can be seen as a perturbation of an unperturbed ground state Hamiltonian, the GW method is also closely related to quantum-mechanical perturbation theory. To employ such a method, an initial guess of the unperturbed system is necessary. This can be done systematically by performing an *ab initio* DFT calculation and use the resulting Kohn-Sham

1 Introduction

orbitals as a starting point of the GW calculations. In this perspective, the GW method is used to correct correlation effects, which are treated inaccurately in DFT calculations.

For a long time it was nearly impossible to perform GW calculations for many problems, due to very demanding computational costs. Through new computing clusters, equipped with an ever-growing amount of memory, this situation has changed and GW calculations can be performed successfully even for large systems.

1.2.1 GW Calculations - Conventional Approach

Although non self-consistent GW calculations, called single-shot GW (or G0W0), are relatively easy to perform, one such calculation alone does not guarantee converged results. One way to improve this situation is to either self-consistently vary the screened Coulomb potential W (G0W) or both, the potential and the Green function G (fully self-consistent GW). Although this seems to be a good way to get good results for all kinds of properties, such self-consistent calculations can be very difficult to perform.

The standard procedure of producing accurate results at GW level is to manually converge computational parameters, such as the energy cut-off (also for the auxiliary basis set), number of k-points, number of energy bands, and for self-consistent or half self-consistent calculations also the number of iterations (see section (2.2.6)). Practically this means increasing each of the different parameters systematically and performing several calculations in succession, while monitoring the results. Above approach is often referred to as the *conventional method* (or *conventional approach*) of calculating converged results with GW simulations, which will also be the term for it used in this thesis.

Unfortunately, this method has two significant downsides: (1) It is a lot of work setting up all the different G0W0 calculations and converging several parameters manually and (2), probably even worse, is that this method does not stand on a strict mathematical foundation. That quantities, such as the band gap, show a convergence while changing mentioned parameters is a purely heuristic find and can not be expressed formally. A relatively new method, the basis set extrapolation proposed by Klimeš *et al.* in reference [2], makes here a difference by being rigorously derived from the mathematical background of the GW approach.

1.2.2 The Basis Set Extrapolation Method

In their paper from 2014, Klimeš *et al.* introduce a new way for calculating converged band gaps on GW level.[2]

Basis set extrapolation uses the fact, that errors which arise due to the inaccurately treated correlation energy of the system can be reduced by including more (empty) energy bands in the calculation. In fact, the error shows a $1/N$ behavior, where N is the number of electronic bands.¹ This means by extrapolating to an infinite set of bands, the error can be corrected and

¹As shown later in section (2.2.7), in the derivation of the basis set correction, the number of plane waves in the auxiliary basis set N_{pw}^X occurs instead of the number of bands N_B . However, these are equivalent formulations because the number of bands and energy cut-offs (determining the number of plane waves) have to be increased simultaneously, which in turn stand in a fixed relation to each other where the energy cut-offs of the normal basis set (E_{cut}) governs the one of the auxiliary basis set ($E_{cut}^X = \frac{2}{3}E_{cut}$).

the correct band gap retrieved.

This method not only reduces the number of G0W0 necessary to be performed, it also does not require to converge parameters manually or monitor results in the process. This is the key difference important for this project. The fact, that this method is straight forward to implement and does not require a subjective evaluation if convergence is reached makes it possible to automate the whole procedure without the need of complicated algorithms.

One downside, however, is that with this method only single results are corrected, such as the fundamental band gap or quasi particle energies. While the conventional method provides a set of parameters for which most results found in the output should be well converged (including band structure, density of states, dielectric function, and optical spectra), the basis set extrapolation fails at doing that. This means there exists no single output file with accurate data, in contrast to the conventional method once the optimal parameters are found.

1.3 Automated Calculations for Large Material Sets

A crucial, if not the most important part of modern computer science is the access to large data sets. As in many different fields, in computational solid state and materials physics too, the arrival of data driven methods (machine learning) seems like a promising new approach to many already solved or still unsolved problems.[7] The success of such methods depends greatly on the data set available, however.

For GW calculations such data collections are practically non-existent, since simulations meant tedious one-by-one investigations of single materials in the past. With the new basis set extrapolation scheme as a starting point for the automation of GW calculations this situation could change in the near future. Since most of the workload of setting up calculations and parsing results can be translated into scripts, large sets of materials can be analyzed quickly. This is also the main idea of various software suites specialized in high-throughput simulations, such as AiiDA.[8] GW calculations in the form of basis set extrapolation can easily be implemented in such environments.

This larger scope shifts the paradigm away from single, sophisticated calculations towards larger test series. Furthermore, the fine tuning is done more so by changing hyper-parameters (fixed values for all calculations, as called for machine learning codes) and not by adjusting all parameters for each material individually. As a consequence, the results themselves have to be viewed as a collective entity. The result data sets accuracy has to be improved as a whole while retaining a practical throughput rate.

With the goal of databases containing data from materials in the thousands, this trade-off is well worth it in the early stages. Furthermore, machine learning techniques need first and foremost a large quantity of data to be applicable at all. A higher quality of data then improves the methods accuracy. With this in mind, it only makes sense to lay a focus on the quantitative yield of a method as a starting point and then try to improve the average accuracy of the whole result set. To provide such a database and the means to fill it represents the final goal of the project the work done here is part of.

2 Theory

In this section a theoretical overview is given, to create a better understanding of underlying concepts used in this work. The topics are split up in (1) explaining the physical background and (2) the computational theory and methods used, since these are the two main pillars the project is built on.

At the beginning of both parts, relevant terms which might occur more frequently during the thesis are explained briefly.

In the section for physical theory, basics about the description of electronic systems are given before relevant material properties like the band structure and the fundamental band gap are covered.

For the computational part, theoretical details regarding *ab initio* simulations in general, density functional theory (DFT), and GW calculations are discussed. Afterwards, the conventional method of calculating a converged GW level band gap is presented, before the basis set extrapolation is introduced in a short summary of what can be found in the original proposal of the method.[2]

2.1 Physical Theory

Many important properties of materials arise due to their electronic configuration. Especially for technical applications crucial quantities like the band gap, conductivity, optical properties, and many more can either directly or indirectly be related to their electronic structure.

2.1.1 Important Terms

To understand the topics discussed below, some important terms will be explained here briefly.

reciprocal space: Materials are usually described in two different spaces: real and reciprocal space. Real space corresponds to our actual world, where we describe the materials structure by defining a lattice through basis vectors. We call this the systems Bravais lattice, which represents together with the atomic basis the crystal symmetry. Atom positions in the crystal can then be expressed by linear combinations of these basis vectors.

For reciprocal space, the topic gets a bit less intuitive. The reciprocal lattice corresponding to the real lattice of a material is given by all wave vectors \mathbf{k} , that define plane waves with the periodicity of the given Bravais lattice.[9] This can be explained in a mathematical way by defining the systems Bravais lattice as $\mathbf{R} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3$. The reciprocal lattice is then given by all vectors $\mathbf{G} = h \mathbf{b}_1 + k \mathbf{b}_2 + l \mathbf{b}_3$ which fulfill

$$e^{i\mathbf{G}\cdot\mathbf{R}} = 1. \quad (2.1)$$

2 Theory

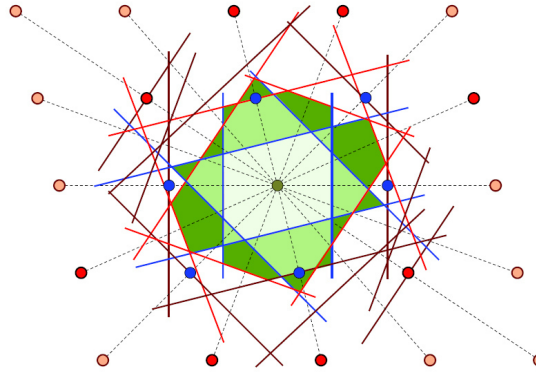


Figure 2.1: Graphical representation of the Brillouin zones. The most relevant first Brillouin zone is marked light green at the center. In increasing darkness the second and third Brillouin zone is also displayed. Mark that all zones have the same area, so the first zone is not included in the second, and so on. Figure taken from reference [9]

The reciprocal space is especially important for describing optical properties and is used in many different experiments, e.g. scattering experiments for structure analysis or the measurement of electron dispersion to determine the band structure.

Brillouin zone: This is the smallest possible unit cell in reciprocal space. It is constructed by taking half the connection line of one reciprocal lattice point to each of its nearest neighbors. A graphical representation can be found in figure (2.1). Due to the crystals periodicity, it is often sufficient to discuss properties like the band structure or phonon spectrum only in the first Brillouin zone. In real space, the smallest unit cell constructed the same way as the Brillouin zone is referred to as Wigner-Seitz cell.

correlation: Phenomena originating from electron-electron interactions are often generalized as correlation effects. These include mechanisms involving the electron spins or dipole interactions, and other many-body effects. Since it is impossible to account for all these interactions analytically or treat them accurately in simulations, these contributions have to be approximated in most cases. For this reason, exchange-correlation potentials are implemented to account for such contributions. Finding an appropriate method to do this influences the calculated properties of materials significantly.[9]

exchange: Exchange effects occur when two electrons swap their places in a material. Such mechanics contribute to the systems energy, but are also involved in physical phenomena such as covalent bonding or magnetism.[9]

2.1.2 Electronic Systems

As fermions, electrons have to obey the Pauli principle and thus must be described in the frame of quantum mechanics. Even at first glance simple properties like the heat capacity as well as

much more complex phenomena like the formation of polarons can not be explained classically.

In the center of quantum-mechanical models stands the Schrödinger equation and its solutions, the wave functions. By finding and solving the right equation, all properties of the investigated system can theoretically be calculated based on the wave function describing it. Because of this, many problems in quantum mechanics have to do with finding an adequate Schrödinger equation and its solutions.

Hamilton Operator for Electronic Systems

Generally, all electronic systems can be described as a many-body problem involving the electrons and atom cores. The corresponding Hamilton Operator contains parts which describe the individual energies of electrons and cores, as well as their interaction.

$$\hat{H} = T_e + T_n + V_{ee} + V_{nn} + V_{en} \quad (2.2)$$

- kinetic energies of the electrons

$$T_e = -\frac{\hbar^2}{2m_e} \sum_{i=1}^N \nabla_i^2 \quad (2.3)$$

- kinetic energies of the cores

$$T_n = -\frac{\hbar^2}{2M_n} \sum_{n=1}^M \nabla_n^2 \quad (2.4)$$

- Coulomb interaction of the electrons

$$V_{ee} = \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{i,j=1, i \neq j}^N \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (2.5)$$

- Coulomb interaction of the cores

$$V_{nn} = \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_{n,m=1, n \neq m}^M \frac{Z_n Z_m e^2}{|\mathbf{R}_n - \mathbf{R}_m|} \quad (2.6)$$

- Coulomb interaction between electrons and cores

$$V_{en} = -\frac{1}{4\pi\epsilon_0} \sum_i^N \sum_n^M \frac{Z_n e^2}{|\mathbf{r}_i - \mathbf{R}_n|} \quad (2.7)$$

The kinetic energies, as well as the interaction between electrons and cores lower the total energy of the system. Where as the repulsive potential between the electrons or the cores contributes an increase in energy. If the wave function is an eigenfunction of this Hamiltonian, the corresponding eigenvalue is the total energy of the system.

2 Theory

The Schrödinger equation for such systems is then given by

$$\hat{H}\psi = E\psi, \quad (2.8)$$

where ψ is the wave function and E the systems total energy. Since the Schrödinger equation is effectively solved in reciprocal space in most simulations, not the energy eigenvalues in real space (E_n), but rather the eigenvalues in reciprocal space ($E_n(\mathbf{k})$) are calculated most of the time.

2.1.3 Electronic Band Structure

The band structure represents the energy eigenvalues $E_n(\mathbf{k})$ of the electron wave functions, depending on the wave vector \mathbf{k} . Similarly to other quantum-mechanical systems, only discrete energy levels are allowed for the electrons to occupy. These energy levels are the bands, denoted by the band index n and each of them having a finite energy bandwidth. Between these bands are gaps, the so called band gaps, representing regions of energy the electrons cannot occupy. If there is a gap located between the fully occupied band with the highest energy (valence band) and the unoccupied band at the lowest energy (conduction band) the material is an insulator or semi-conductor and this gap is called the *fundamental band gap*. The topology of said band structure can be very complicated, due to complex interactions between atoms in the material.

In crystals, only wave vectors inside the first Brillouin zone have to be considered due to the system periodicity. Parts of the band structure which lie outside of this zone can be shifted inside through translations by multiples of the reciprocal lattice vector \mathbf{G} . [9] Because of this, it is sufficient to consider paths between high-symmetry points of this cell in experiments and calculations to investigate the systems band structure. An example for this can be found in figure (2.2). Suggested paths, which should lead to detailed band structures for different unit cells can be found in reference [10].

2.1.4 Electronic Band Gaps

The band gap refers to a range of energy between two bands that cannot be occupied by electrons. As visible in figure (2.2), there is no connection between the valence (highest band below 0) and the conduction band (lowest band above 0). This means it is not possible for electrons to change from the valence to the conduction band without the excitation by external energy. Such band gaps are called *fundamental energy gaps* and give rise to semi-conducting and insulating behavior of materials.

The fundamental band gap size influences the electrical behavior of the material to a very high degree. While materials with very large band gaps are insulators, a smaller band gap (as for silicon in figure (2.2)) allows for semi-conducting properties. A material is a conductor, if there is no fundamental gap at all and electrons can move from valence to conduction band freely.

There are also different types of band gaps, depending on the type of fundamental excitation one can differentiate between direct and indirect gaps. Direct gaps can be accessed more easily

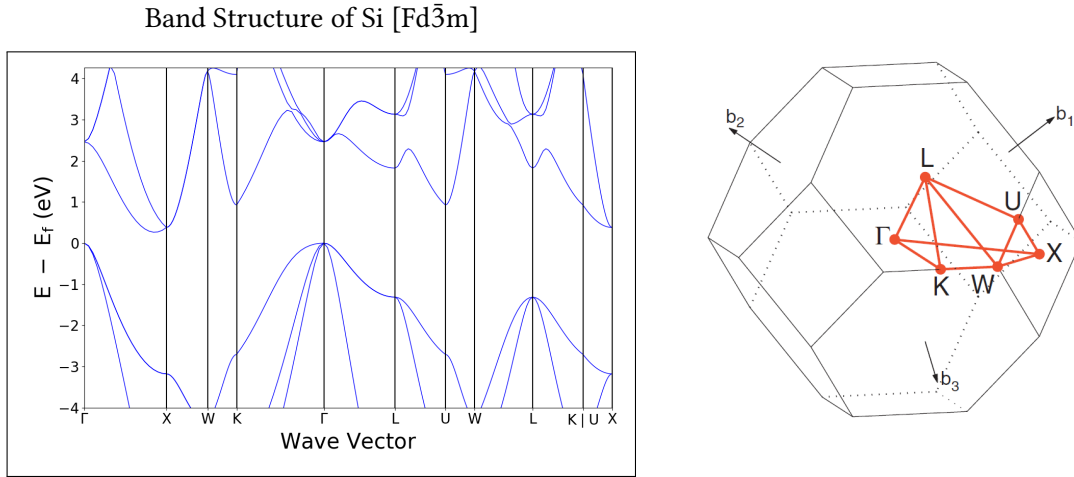


Figure 2.2: Left: Band structure of silicon in face-centered cubic configuration calculated for this project. The considered path between high symmetry points in the first Brillouin zone is denoted at the x-axis. Right: Visualization of the first Brillouin zone and the path between high symmetry points. Figure taken from reference [10]

in experiments and applications, due to the fact that electrons can be excited into the conduction bands with the use of photons only. This is because photons carry no impulse they could transfer to the electron, which means that the excitation can only occur at a fixed value of the wave vector \mathbf{k} in reciprocal space. To address indirect gaps, phonons¹ have to be involved in the process to make a change in momentum possible. Because of this, the process is much more complicated to investigate or use practically.

The fundamental band gap size and its type is of great importance for different kinds of modern day applications. Finding materials with a band gap in the right energy range, or even constructing one (band gap engineering), is a key topic of modern day solid state physics.

2.2 Computational Theory

After a quick introduction to the physical side of theory, computational backgrounds will be explained below.

2.2.1 Important Terms

Some terms important for the presented project will be elaborated below. This should give a good overview over relevant concepts occurring in this thesis.

¹While photons are used for the quantization of electromagnetic fields, phonons are correspondingly used to quantize lattice vibrations. They can be described as quasi-particles with energy $E = \hbar\omega$ and carry a quasi-impulse $\mathbf{p} = \hbar\mathbf{k}$.

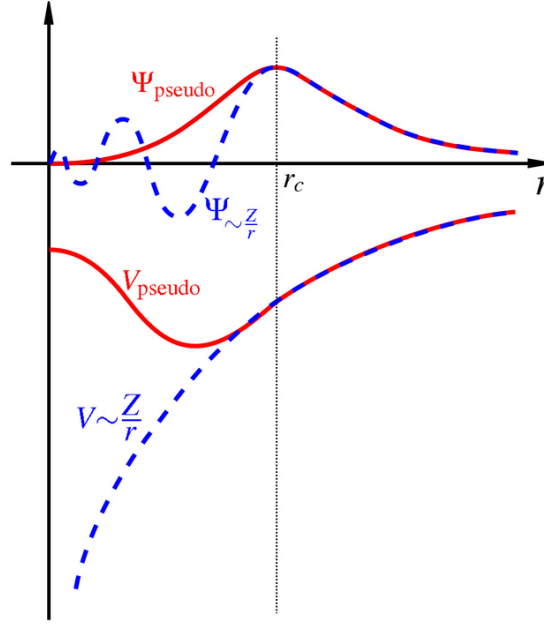


Figure 2.3: Graphical representation of the concept behind pseudo-potentials. Below the radius r_c the diverging potential is replaced by a finite one to avoid rapid oscillations in the resulting wave function.

ab initio simulation: In these kind of simulations no empirical data is used during the calculation. Simulated systems are often constructed out of *pseudo-potentials* (such as in DFT and G0W0 calculations done for this project) and the dynamics are described through mathematical concepts.

pseudo-potential: Pseudo-potentials approximate the atomic wave function for the use in calculations. Real systems show a complicated behavior such as rapid oscillations of the wave function or divergences of the potential near the atom core region, which makes the description difficult with basis functions used in simulation codes.

Therefore this problematic region is often approximated by a different function, which can be better described computationally, and fulfills certain criteria. Examples for the latter would be that the norm of the wave function is accurately represented (*norm-conserving* pseudo-potentials) or that the pseudo wave function in the core region is especially smooth (*ultrasoft* pseudo-potential). A graphical example of such a potential can be seen in figure (2.3).

These potentials are usually not created by the user, since generating a realistic pseudo-potential is very complicated and involves difficult simulations in the form of all-electron codes. However, once such a pseudo-potential is constructed for an element it can be reused in all future calculations as long as the code supports its standard.

For this project so called projector-augmented wave (PAW) *PBE* pseudo-potentials were used.[11][12] For these potentials special GW-versions are available, constructed specifically for these simulations. They include a larger number of plane waves leading to larger energy cut-offs, such that scattering properties are well reproduced even in a relatively wide region

above the Fermi energy. By constructing the potentials this way, a larger number of empty bands can be treated accurately, which the convergence of GW quasi-particle energies crucially depends on.

Perdew, Burke and Ernzerhof (PBE) exchange-correlation potential: Because of their complexity, exchange and correlation effects in materials cannot be described analytically during an *ab initio* simulation. As an approximation different potentials can be used to describe these effects. Commonly used examples are the local density approximation (*LDA*) or the general gradient approximation (*GGA*). The *PBE* potential is a variation of the general gradient approximation and used in the calculations for this project.[13]

Hartree-Fock approximation: In this approximation, the all-electron problem arising in electronic systems is reduced to a one-particle problem. This is done by introducing the Fock-operator, which replaces the many-body Hamiltonian in the systems Schrödinger equation. The problem is transformed into a set of single-particle equations, where all electrons move in a effective field, often referred to as *mean field*. The GW approximation can be seen as a generalization of this, where the bare Coulomb interaction is replaced by a dynamically screened one.[14]

quasi-particle shift: Additionally to electronic band gaps, the quasi-particle shifts (QP-shift, QS) are given as results for each material. This property corresponds to the shift experienced by the energy eigenvalues due to the correction applied through the G0W0 calculation. It is calculated as the difference between corrected valence band energies (after G0W0 has been applied) and the initial bands (Kohn-Sham eigenvalues as calculated in DFT) at the corresponding k-point. In most cases this property is negative, since the GW correction leads to a further opening of the band gap, i. e. lowering of the valence band energies.

2.2.2 VASP - Vienna Ab initio Simulation Package

All calculations for this project are done using the Vienna Ab initio Simulation Package, more broadly known as VASP.[15, 16]

This code for *ab initio* simulations uses quantum mechanical calculations with plane waves as a basis set. With periodic boundary conditions and calculations done in reciprocal space it is especially suited for the simulation of periodic systems, such as crystals and solids in general.

In the following, important files used by VASP will be introduced briefly:

INCAR: This file contains flags which determine general properties for the calculation. Here one can switch on or off certain functionalities of VASP, set computational parameters, and also define the simulation type, i.e. if a DFT (see section (2.2.3)) or GW (see section (2.2.4)) calculation should be performed. Different values for the energy cut-off (*ENCUT*) or the number of energy bands (*NBANDS*) are for this project relevant parameters that can be defined here.

2 Theory

POSCAR: The structural information of the system is stored in this input file. Its lattice vectors, atom types and their positions are defined here. This file is used in the automatic GW procedure as main input, since all information necessary for setting up the calculation can be extracted here.

POTCAR: Atom specific pseudo-potentials used to simulate the system atoms are provided in the *POTCAR* file. It is a very important decision to choose a suitable pseudo-potential for the problem at hand, since the calculation performance and its results are directly tied to it. For this project non-norm conserving *PAW-PBE* potentials were used for most calculations. Because of the high memory requirement of the basis set extrapolation, these potentials were chosen to be the most efficient. Employing norm conserving potentials would lead to very large basis sets, although the results would be more accurate for some systems due to the reduced norm violation.[17] Norm conserving potentials were only used on a subset of materials to validate the method and compare the results to the original proposal of the basis set extrapolation.[2]

KPOINTS: The last important input is the *KPOINT* file. Here, the density of the computational grid in reciprocal space is defined. The number of k-points can severely influence the accuracy of the calculation, so an adequate choice is crucial.

Since the extrapolation to an infinite basis set shows low sensitivity to the number of k-points, most of the G0W0 calculations can be done using a coarse grid.[17] Only one calculation has to be done for correcting to a higher density grid afterwards, as shown figures (3.1) and (3.2).

The *KPOINTS* file also allows to define a path between high symmetry points in the reciprocal unit cell, along which the band structure is calculated. These paths were taken from reference [10].

2.2.3 Density Functional Theory

Although this project mainly is about the application of GW calculations, the importance of density functional theory as a preliminary step cannot be neglected. Every G0W0 calculation needs an initial guess for the wave function and its derivative. Since these are usually provided through DFT calculations the theory behind them will be quickly summarized in the following section.

In quantum mechanical simulations, one of the big goals is solving the systems Schrödinger equation, i.e. finding the corresponding wave function. Once this is achieved one can calculate rather easily expectation values of operators, which then represent properties of the system at hand. Although this seems easy theoretically speaking, in practice it is nothing that can be done analytically for most real systems.

Finding an easy, efficient, and still relatively accurate way of determining the systems wave function is an important topic in computational science. Density functional theory is one approach of solving this problem. The core idea around this method revolves is the Hohenberg-Kohn theorem.

The Hohenberg-Kohn Theorem

In their famous paper 1964 Hohenberg and Kohn set the foundation of modern DFT. Their theorem contains two main statements, often referred to as separate theorems abbreviated as *HK1* and *HK2*. Firstly, it was stated that an universal functional for the energy $E[n(\mathbf{r})]$ can be defined which is valid for any external potential V_{ext} . This functional depends only on the charge density $n(\mathbf{r})$ (*HK1*). Secondly, for any given external potential the ground state energy is the global minimum of this functional and no energy below this true ground state energy can be found for the system. The density $n(\mathbf{r})$ (equation (2.9)) that minimizes $E[n(\mathbf{r})]$, n_0 , is the ground state density and can be found through variational principles (*HK2*).

$$n(\mathbf{r}) = N \int d^3r_2 \int d^3r_3 \dots \int d^3r_N \psi^*(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.9)$$

Effectively this means that instead of dealing with a functional of N vectorial variables one can work with a function only depending on one vectorial variable instead. This is possible, because the ground state wave function not only has to reproduce the ground state density, but also has to minimize the systems energy. Knowing the ground state charge density implicitly contains all needed information for recreating the ground state wave function in the given potential.[18]

Starting from that, multiple methods can be developed using this systematically as an advantage over calculations only using the wave function.

The Kohn-Sham Equations

The Kohn-Sham equations (equation(2.10) to (2.12)) are the core of most DFT-calculations. They map the real, interacting system to an auxiliary construct with no interaction between its components but the same charge density. Through the iteration of self consistent loops this density can be determined.

$$v_s(\mathbf{r}) = v(\mathbf{r}) + v_H(\mathbf{r}) + v_{xc}(\mathbf{r}) \quad (2.10)$$

$$\left[-\frac{\hbar^2 \nabla^2}{2m} + v_s(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}) \quad (2.11)$$

$$n(\mathbf{r}) \equiv n_s(\mathbf{r}) = \sum_i^N f_i |\phi_i(\mathbf{r})|^2 \quad (2.12)$$

The main idea here is to make an initial guess of the systems ground state charge density $n(\mathbf{r})$, calculating the corresponding potential $v_s(\mathbf{r})$, solving the differential equation (2.11), and with its solution finding a new, and better guess of the initial density by using equation (2.12).

In many cases, the wanted wave function can be found rather fast and reliably this way, making such an iterative method of solving the systems Schrödinger equation a powerful tool for computational quantum mechanics.[18]

2.2.4 The GW Approach

The GW approach can be classified as post-DFT method, i.e. is used to further refine or correct data acquired with density functional theory. This is achieved by replacing the approximate v_{xc} above by calculating the explicit self-energy, using the Green function of the system (G) and the screened Coulomb potential (W). In the following some basic principles of this method will be explained. All concepts presented are taken from a very detailed derivation of the method that can be found in reference [14].

Green Function of Electronic Systems

The single-particle Green function for solid state systems, or systems where its properties are mainly dictated by the electronic configuration in general, can be defined as shown below.

$$iG(x, x') = \langle N | T[\hat{\psi}(x)\hat{\psi}^\dagger(x')] | N \rangle \quad (2.13)$$

Here $|N\rangle$ is the N-electron ground state, $\hat{\psi}(x)$ a field operator in Heisenberg representation, and T the time-ordering operator. Above Green function can be physically interpreted as the probability amplitude, that for

- $t' > t$ a hole created at x will propagate to x' and for
- $t > t'$ an electron added at x' will propagate to x .

The photo-emission spectrum in the limit of large kinetic electron energies (*Sudden approximation*) can now be directly related to above single particle Green function. Therefore, we can obtain - besides the expectation values of any single-particle operator and the ground state energy of the system - also its one-electron excitation spectrum.

The Hedin Equations

In the following the Hedin equations as a core concept of the GW approximation will be introduced briefly. Again, all information was taken from reference [14] and a detailed derivation containing all details can be found there.

Starting now from the Heisenberg equation of motion for the Green functions field operator $\hat{\psi}(x)$

$$i \frac{\partial \hat{\psi}(x)}{\partial t} = [\hat{\psi}(x), \hat{H}] \quad (2.14)$$

one can find the equation of motion for the Green function as

$$[i \frac{\partial}{\partial t} - h_0(x)]G(x, x') - \int dx'' M(x, x'')G(x'', x') = \delta(x - x'). \quad (2.15)$$

In above equation (2.15) M is the mass operator, containing the Hartree potential v_H and the self-energy Σ .

With the use of a functional derivative we can now separate this mass operator and extract the self-energy. With this the equation of motion takes the following form:

$$[i \frac{\partial}{\partial t} - H_0(x)]G(x, x') - \int dx'' \Sigma(x, x'')G(x'', x') = \delta(x - x'). \quad (2.16)$$

The Hartree potential is now absorbed in H_0 , as well as the external field $\phi(\mathbf{x}, t)$ which was introduced to perform the functional derivative and can later be set to zero.

$$H_0 = h_0 + v_H + \phi \quad (2.17)$$

Through further investigations, which will not be presented in detail here but can be found in reference [14], one can now derive a set of coupled integral equations, the so called Hedin equations² (2.18) - (2.21).

$$\Sigma(1, 2) = i \int d(34) G(1, 3^+)W(1, 4)\Lambda(3, 2, 4) \quad (2.18)$$

$$G(1, 2) = G_0(1, 2) + \int d(34) G_0(1, 3)\Sigma(3, 4)G(4, 2) \quad (2.19)$$

$$\Lambda(1, 2, 3) = \delta(1 - 2)\delta(2 - 3) + \int d(4567) \frac{\delta\Sigma(1, 2)}{\delta G(4, 5)} G(4, 6)G(7, 5)\Lambda(6, 7, 3) \quad (2.20)$$

$$W(1, 2) = v(1, 2) + \int d(34) v(1, 3)P(3, 4)W(4, 2) \quad (2.21)$$

In above equations, several quantities occur:

Σ is the self-energy

W is the screened Coulomb potential, given by

$$W(1, 2) = \int d3 \epsilon^{-1}(1, 3)v(3 - 2) \quad (2.22)$$

itself containing the inverse dielectric function $\epsilon^{-1} = \frac{\delta V(1)}{\delta \phi}$, where $V = V^H + \phi$ is the total and v the standard Coulomb potential

Λ is the vertex function

G_0 is the Green function corresponding to $\Sigma = 0$

²A shorter notation is used in the Hedin equations to allow a more compact form. In all equations 1 corresponds to the variable-tuples $x_1 = (\mathbf{r}_1, t_1)$, $2 \equiv x_2 = (\mathbf{r}_2, t_2)$, and so on.

2 Theory

P is the polarization function which describes the change in the charge density ρ upon a change in the total field V , and is given by

$$P(1, 2) = \frac{\delta\rho(1)}{\delta V(2)} = -i \int d3d4 G(1, 3)\Lambda(3, 4, 2)G(4, 1^+). \quad (2.23)$$

Another, related quantity describing the change in the charge density upon a change in the external field ϕ is the response function R , given by

$$R(1, 2) = \frac{\delta\rho(1)}{\delta\phi(2)}. \quad (2.24)$$

These two quantities themselves are related to the inverse dielectric function ϵ^{-1} via

$$\epsilon^{-1} = \frac{\delta V}{\delta\phi} = 1 + v \frac{\delta\rho}{\delta\phi} = 1 + v \frac{\delta\rho}{\delta V} \frac{\delta V}{\delta\phi}. \quad (2.25)$$

Generally speaking, the GW method is a generalization of the Hartree-Fock approximation, where the bare Coulomb interaction v is substituted by a screened interaction W . The GW approximation used in most implementations can formally be obtained by neglecting the second term in the vertex function (2.20), thus setting $\Lambda(1, 2, 3) = \delta(1 - 2)\delta(2 - 3)$. [14]

2.2.5 The GW Approximation as Implemented in VASP

Like mentioned above, also in VASP the approximation is implemented by simplifying the vertex function (2.20). This leads to simplified expressions for both, the polarization function (2.23) and the self-energy (2.18):

$$P(1, 2) = -iG(1, 2)G(2, 1) \quad (2.26)$$

$$\Sigma(1, 2) = iG(1, 2)W(1, 2) \quad (2.27)$$

Furthermore, by transforming the problem into the frequency domain (reciprocal space), a spectral representation for the Green function can be introduced as

$$G(\mathbf{r}, \mathbf{r}', \omega) = \sum_i \frac{\psi_i(\mathbf{r}, \omega)\psi_i^\dagger(\mathbf{r}', \omega)}{\omega - E_i(\omega)}. \quad (2.28)$$

In above equation, ψ_i are the solutions to the following quasi-particle equation:

$$H_0(\mathbf{r})\psi_i(\mathbf{r}, \omega) + \int d^3r' \Sigma(\mathbf{r}, \mathbf{r}', \omega)\psi_i(\mathbf{r}', \omega) = E_i(\omega)\psi_i(\mathbf{r}, \omega) \quad (2.29)$$

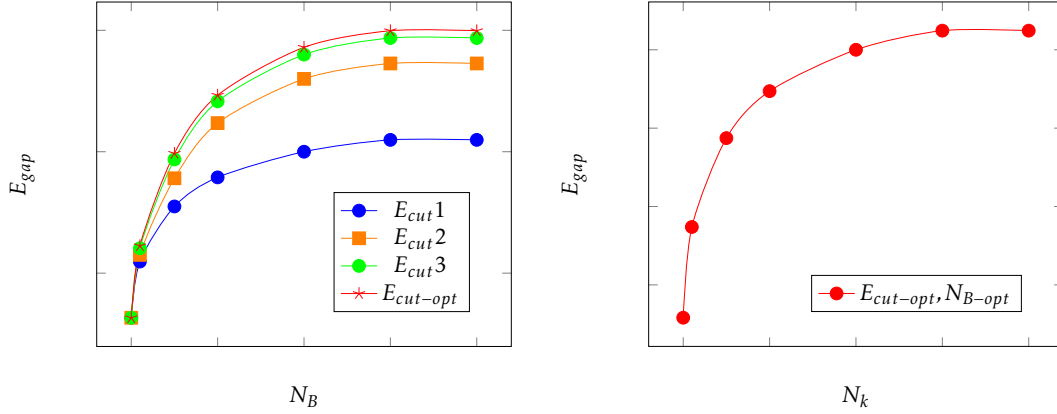


Figure 2.4: Schematic representation of the conventional method for calculating converged GW band gaps. First the energy cutoff and the number of bands is changed until the gap energy converges (left), then the number of k-points is determined using the optimal values $E_{cut-opt}$ and N_{B-opt} for these two quantities (right).

Solving this eigenvalue problem and therefore determining the quasi-particle energies E_i corresponds to the here employed G0W0 calculation. This means the DFT Kohn-Sham energies form the starting wave function are corrected once, by calculating the polarization function P , the screened Coulomb potential W , and the self-energy Σ .

2.2.6 Converged GW Band Gap

In this section, two different procedures will be explained for calculating a converged electronic band gap using G0W0 simulations: (1) The so-called *conventional method*, as established standard procedure, and (2) the newly proposed *basis set extrapolation*. The latter will be used for this project in an altered, more memory conserving but in turn slightly less accurate version.

Conventional Method

The concept of what is called the "conventional method" of retrieving a converged GW-level band gap is purely based on heuristics. It was observed, that in many cases one has to converge several parameters before the accurate results are retrieved. These parameters are for G0W0 calculations: (1) the energy cut-offs of the plane wave basis set E_{cut} and the response function E_{cut}^χ , as well as (2) their corresponding number of included energy bands N_B and N_B^χ , (3) the number of k-points used for the calculation grid N_k , and (4) the number of calculation points on the frequency grid N_ω .

By increasing these values progressively, the quasi-particle shift - and thus the correction of DFT-level band gaps - should converge to the best achievable value.

A standard procedure of this approach, where the fundamental band gap is monitored as convergence indicator, looks as follows:

2 Theory

1. The band gap is calculated in a single-shot GW calculation (G0W0) using fixed values for all parameters E_{cut} , E_{cut}^χ , N_B , N_{pw}^χ , N_k , and N_ω .
2. An example for this can be found in figure (2.4). By subsequently increasing the energy cutoff E_{cut} and corresponding number of energy bands N_B the gap should converge at a certain point, revealing the optimal values $E_{cut-opt}$ and N_{B-opt} . If this is achieved, $E_{cut-opt}$ and N_{B-opt} are fixed and the number of k-points N_k has to be converged the same way.
3. This procedure must be performed carefully with each parameter, while always keeping in mind that some parameters might influence each other.
4. A single-shot GW calculation using all optimal parameters should now give a converged band gap.

Clearly, arriving at a converged band gap with this procedure is quite a tedious task, since numerous single-shot GW calculations must be done to achieve convergence. However, many converged properties can be extracted from the output files of a single G0W0 calculation, once the final set of parameters is found.

Another, even more severe downside is the fact, that there is presently no rigorous mathematical proof that this method indeed yields the correct and converged band gap for the simulated system.[17] Extensive testing and comparing the results with experimental findings suggests that this approach is valid in most cases, although a solid mathematical foundation would be much appreciated, rather than a purely heuristic one.

Basis Set Extrapolation

Due to a very slow convergence of the quasi-particle energies it was found that an enormous amount of empty bands need to be included in the single-shot GW calculations to find accurate band gaps for some systems.[19] Because of this, and other observations, it was suggested that the quasi-particle energies in the GW approach show a convergence behavior of $1/N_B$, where N_B is the number of energy bands included in the basis set of the calculation.[2]

For the practical application this means the accurate gap should be found through extrapolation to an infinitely large basis set. This is done by performing several G0W0 calculations while increasing the energy cutoffs E_{cut} and E_{cut}^χ , as well as their corresponding number of energy bands N and N^χ (which is equivalent to increasing the number of energy bands N_B in the basis set and N_{pw}^χ in the auxiliary basis set of the calculation).

It is not recommended to change the two energy cutoffs E_{cut} and E_{cut}^χ independently. Because of this, the VASP default of $E_{cut}^\chi = \frac{2}{3}E_{cut}$ has been kept for all calculations done here.

In contrast to the conventional method this approach stands on a formal basis. The important points of the mathematical reasoning will be summarized below. A detailed derivation and all here mentioned information, as well as additional relevant references, can be found in reference [2], where this basis set extrapolation method was proposed.

The most crucial approximation for this derivation is that the high-energy unoccupied states occurring in the calculations can be represented by plane waves. This can be justified due to the

fact, that for high energies the kinetic term $-\Delta/2$ will be the dominant part of the Hamiltonian and it can be well described by plane waves.

2.2.7 Basis Set Correction of the Correlation Energy

In the following a short summary of the derivation used for the basis set correction will be given, investigating the convergence behavior of the direct second-order Møller-Plesset energy (dMP2). The full version of the derivation and all information used below can be found in reference [2].

The main errors that have to be corrected arise from the inaccurately treated correlation energy of the system, specifically due to a plane-wave cut-off E_{cut}^χ imposed on the response function of the GW calculation. It restricts the auxiliary basis set constructed from \mathbf{G} (in the following also \mathbf{G}') to energies below the cut-off, such that $\frac{|\mathbf{G}|^2}{2} < E_{cut}^\chi$.

Starting from a more general form of the direct correlation energy E^{RPA} for simulations involving the random phase approximation (RPA) we can see that for large wave vectors \mathbf{G} the polarizability χ_0 becomes small and the correlation reduces to the leading-order term.

$$E^{RPA} \approx Tr[\frac{1}{2}(\chi_0 v)^2] \quad (2.30)$$

In above equation v is the Coulomb operator, the second order term E^{dMP2} can be written as

$$E^{dMP2}(\mathbf{G}', \mathbf{G}) = -2 \sum_{ij}^{occ} \sum_{ab}^{unocc} \sum_{\mathbf{G}\mathbf{G}'} \langle i|\mathbf{G}|a\rangle \langle a|-\mathbf{G}'|i\rangle \frac{4\pi}{\Omega \mathbf{G}^2} \times \\ \langle j|\mathbf{G}'|b\rangle \langle b|-\mathbf{G}|j\rangle \frac{4\pi}{\Omega \mathbf{G}^2} \frac{1}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j} \quad (2.31)$$

where $\{a, b\}$ and $\{i, j\}$ are the indices for the unoccupied and occupied states respectively, and epsilon the corresponding energies. We are now interested in the energy error, which arises because of the cutoff. For this we investigate the energy corresponding to auxiliary states \mathbf{G} and \mathbf{G}' which have no overlap with occupied orbitals, specifically where

$$\langle i|-\mathbf{G}|j\rangle \approx 0. \quad (2.32)$$

Through different assumptions³ and using the Fourier transformed charge density (here given

3

- Allow indices a and b in equation (2.31) to run over all states, instead of only unoccupied ones. This can be justified by equation (2.32).
- Assumption that we can describe high-energy occupied states as plane waves of the form $\psi_a(\mathbf{r}) = \exp(i\mathbf{G}_a \cdot \mathbf{r})/\sqrt{\Omega}$ with eigenenergies $\epsilon_a = \mathbf{G}_a^2/2$.
- Approximate $\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j \approx (\mathbf{G}^2 + \mathbf{G}'^2)/2$ due to the assumption that for occupied states components with small wave vectors dominate.
- Use $\sum_a |a\rangle \langle a| = 1$.

2 Theory

for orbital i) equation (2.33)

$$\rho_i(\mathbf{G}) = \frac{1}{\Omega} \int_{\Omega} d\mathbf{r} \psi_i^*(\mathbf{r}) \psi_i(\mathbf{r}) \exp(i\mathbf{G} \cdot \mathbf{r}) \quad (2.33)$$

we can rewrite equation (2.31) in the more compact form (2.34).

$$E^{dMP2}(\mathbf{G}, \mathbf{G}') = -\frac{1}{2} |\rho(\mathbf{G} - \mathbf{G}')|^2 \frac{4\pi}{\mathbf{G}'^2} \frac{4\pi}{\mathbf{G}^2} \frac{2}{\mathbf{G}^2 + \mathbf{G}'^2}. \quad (2.34)$$

With the cutoff condition and introducing $G_{cut} = \sqrt{2E_{cut}^\chi}$ we arrive at the energy error ΔE as final solution.

$$\Delta E = \frac{4}{3} \frac{1}{G_{cut}^3} \sum_{\mathbf{g}} |\rho(\mathbf{g})|^2 \approx \frac{2}{9\pi^2} \frac{\Omega^2}{N_{pw}^\chi} \sum_{\mathbf{g}} |\rho(\mathbf{g})|^2 \quad (2.35)$$

Equation (2.35) tells us that the error, which arises due to the inaccurate treatment of the direct correlation energy, has to shrink with increasing number of plane waves N_{pw}^χ included in the auxiliary basis set of the calculation. For $N_{pw}^\chi \rightarrow \infty$ the error should therefore vanish.

3 Methods

In this chapter it will be discussed, how the theoretical concepts were implemented and applied practically.

First, a general overview how to calculate a converged band gap on GW level through the means of basis set extrapolation will be given. Then, the actual implementation used for this project will be broken down and explained in detail. Finally, it will be discussed how the results were calculated from the data produced.

Through all states of implementing the presented automatic GW scheme, the focus was on keeping everything as general as possible. The procedure had to be universally applicable for many different materials, with optimally no user interference and still good accuracy.

3.1 High Throughput GW Calculations

The automatic procedure for performing the basis set extrapolation scheme consists mainly of individual BASH scripts, connecting the different calculation steps. That such an easy way of automating something as complex as fundamental band gap calculations on GW level is possible, highlights the simple structure of the basis set extrapolation, in comparison to the conventional scheme.

There are only two kinds of input files necessary for each material: (1) the structure file containing atom sorts and positions (*POSCAR*) and (2) the respective pseudo-potential files for each atom sort (*POTCAR*). All material specific information can be extracted from these files, including the energy cut-off (*ENCUT*) from the potential file, which then determines the number of bands included in the calculation (*NBANDS*), as well as from the structure file the material symmetry, necessary for the band structure calculation.

3.1.1 Implementing the Basis Set Extrapolation

While implementing the basis set extrapolation scheme, many suggestions from the original proposal [2] were taken into account. The procedure was slightly edited by substituting the most memory demanding (high *ENCUT*) calculation by a less accurate one.

First, the original procedure will be introduced briefly. This method consists of three low-density k-point grid calculations for extrapolation and one high-density grid calculation for the k-point correction. This setup seems to provide good accuracy in terms of the predicted fundamental band gap for most systems.

A graphical representation of the scheme can be seen in figure (3.1), and explained the following way:

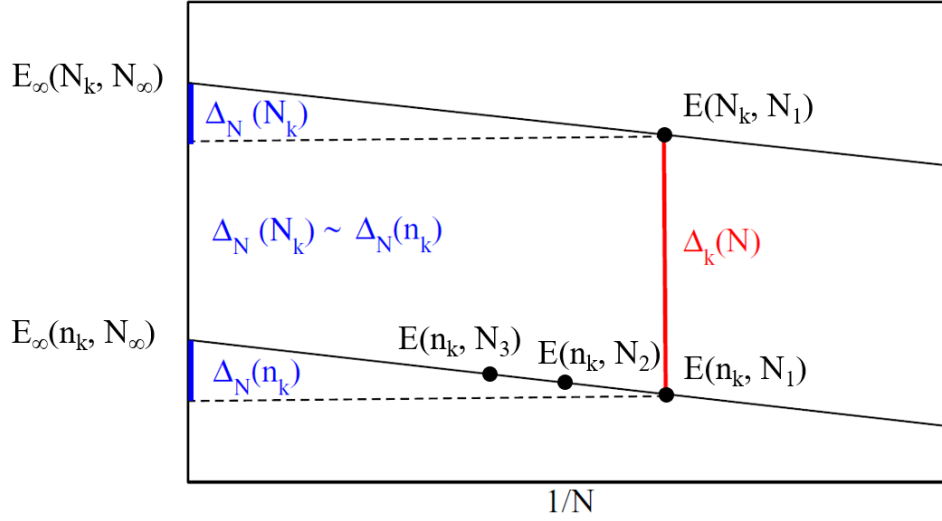


Figure 3.1: Implementation of the basis set extrapolation as proposed by Klimeš *et al.* in reference [2]. The graphic was taken from reference [17] and differently labeled for better understanding.

1. Three single shot GW calculations including a different number of empty bands (N_1 , N_2 , N_3) and using a coarse k-point grid (n_k) are done for extrapolating to an infinite basis set. This is possible, since the extrapolation shows to be very insensitive to the number of k-points used. The band number corresponds to three different energy cutoff values. For E_1 the highest $ENMAX$ value of the used *POTCAR* is chosen as cut-off, then $E_2 = 1.25 \times E_1$ and $E_3 = 1.57 \times E_1$.
2. A forth single shot GW calculation is then performed with a dense grid (N_k), but with a low number of empty bands (N_1, E_1). The result is used to correct the extrapolated gap to the dense k-point solution.

This corresponds to the following equation for the final, corrected band gap:

$$E_\infty(N_k, N_\infty) \approx E_\infty(n_k, N_\infty) - E(n_k, N_1) + E(N_k, N_1) \quad (3.1)$$

Because of the rather restrictive memory requirements, the original procedure was altered to be more feasible. For this the extrapolation was done by using lower energy cut-offs: When E_{pot} is the highest $ENMAX$ value of the potentials, the extrapolation is done using $E_1 = 0.75 \times E_{pot}$, $E_2 = E_{pot}$, and $E_3 = 1.25 \times E_{pot}$. This means the highest cut-off is substituted by a low-energy one. The k-point correction was still done at E_{pot} to ensure a more accurate gap.

$$E_\infty(N_k, N_\infty) \approx E_\infty(n_k, N_\infty) - E(n_k, N_2) + E(N_k, N_2) \quad (3.2)$$

Although this seems to be quite a big change, the overall trend of the extrapolation is not much influenced and the resulting band gaps differ by less than 7 %, as tests show (see section

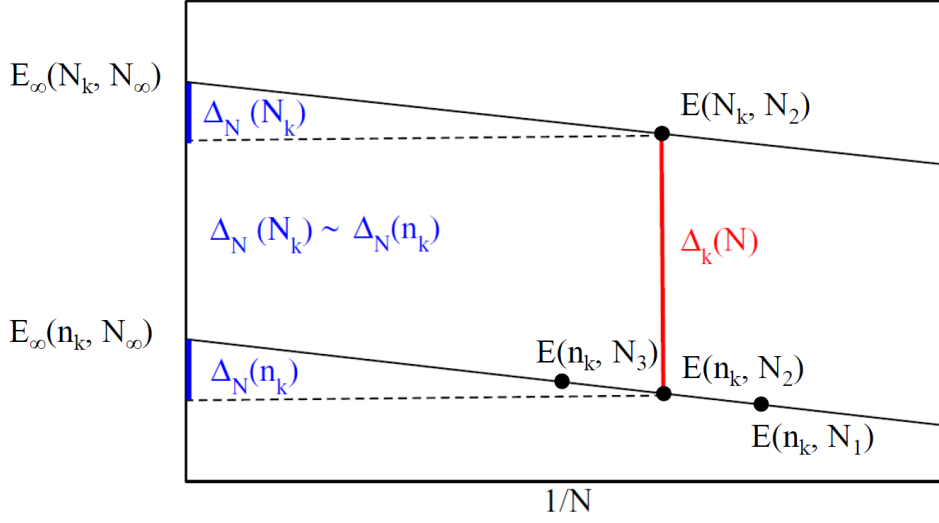


Figure 3.2: Implementation of the basis set extrapolation actually used for the project. Again, this graphic was taken from reference [17] and modified accordingly.

(4.1.3)). Additionally, this was a necessary measure to ensure reliable calculations during and preventing out-of-memory events. This procedure is illustrated in figure (3.2).

3.1.2 Automating GW Calculations

The main topic of this project is how this previously discussed basis set extrapolation now can be used for the automation of GW calculations. While doing simulations for large sets of materials it is crucial to reduce the time spent on repetitive tasks, such as setting up and scheduling each calculation manually. Ideally, the user would only spend time on analyzing and interpreting the results of simulations, omitting these repetitive tasks completely.

With this goal in mind, some requirements on the procedure were made:

- Trustful results and performing as accurate calculations as possible is still one of the top priorities.
- The second most important requirement is reducing the required amount of user interference while doing the simulations.
- The procedure should be robust and universally applicable for different materials.
- A set of parameters should be found which is fitting for most materials, minimizing the cases where a special treatment is necessary for achieving meaningful results.
- Lastly, there should be a balance between keeping the computational cost low and still producing accurate data.

3.2 Overview of the Implementation

The requirements mentioned above were taken as a guideline for implementing the GW procedure. After some different approaches, creating an infrastructure consisting out of BASH scripts was chosen to be the most intuitive way to go. An overview of the final procedure can be found in figure (3.3).

Several different steps are necessary for performing a basis set extrapolation a single material: (1) At least four G0W0 calculations, each of them requiring their own preliminary DFT calculation, as a first step. (2) From all of them the band gap has to be extracted, these are then used for the (3) basis set extrapolation afterwards. (4) Additionally, one DFT calculation is used to extract the band structure of the system on DFT level.

3.2.1 Directory Structure

There are several important directories used by the scripts introduced. They are mentioned below to get a better understanding of the implementation.

repositories: Several repositories are used for holding the input files (*POSCAR*, *POTCAR*, k-point paths for the band structure) and the scripts used for the calculations.

feed: This is a temporary repository for the input structures to be analyzed. The *setup-calculation script* iterates through all *POSCARs* provided in this folder, until it reaches an empty file called *DEADEND*. This signals the code that it should not move on to the next structure and the series is stopped.

bs_extrapolation: In this directory the working folders for the calculations are created by the *setup-calculation script*. It also contains two important auxiliary files: *matcount* and *fincount*. The first contains information on which structure file is the next to use for calculation, the second how many parts of the basis set extrapolation have already finished. A history on which materials have been analyzed is also saved in a log file in this folder.

data: Several important output files are stored in this directory. Besides the *VASP* standard output files *OUTCAR* and *vasprun.xml*, intermediate result files containing the raw gap data are also kept here.

With the files stored here it is possible to reproduce results presented here. Especially for high throughput calculations it is crucial to save only relevant data in a well organized repository, to keep the required amount of memory low and accessibility high.

results: This directory contains all final results. Besides an individual file for each calculation with all the corrected gaps, the band structure images and a file collecting all gaps as an overview is also stored here.

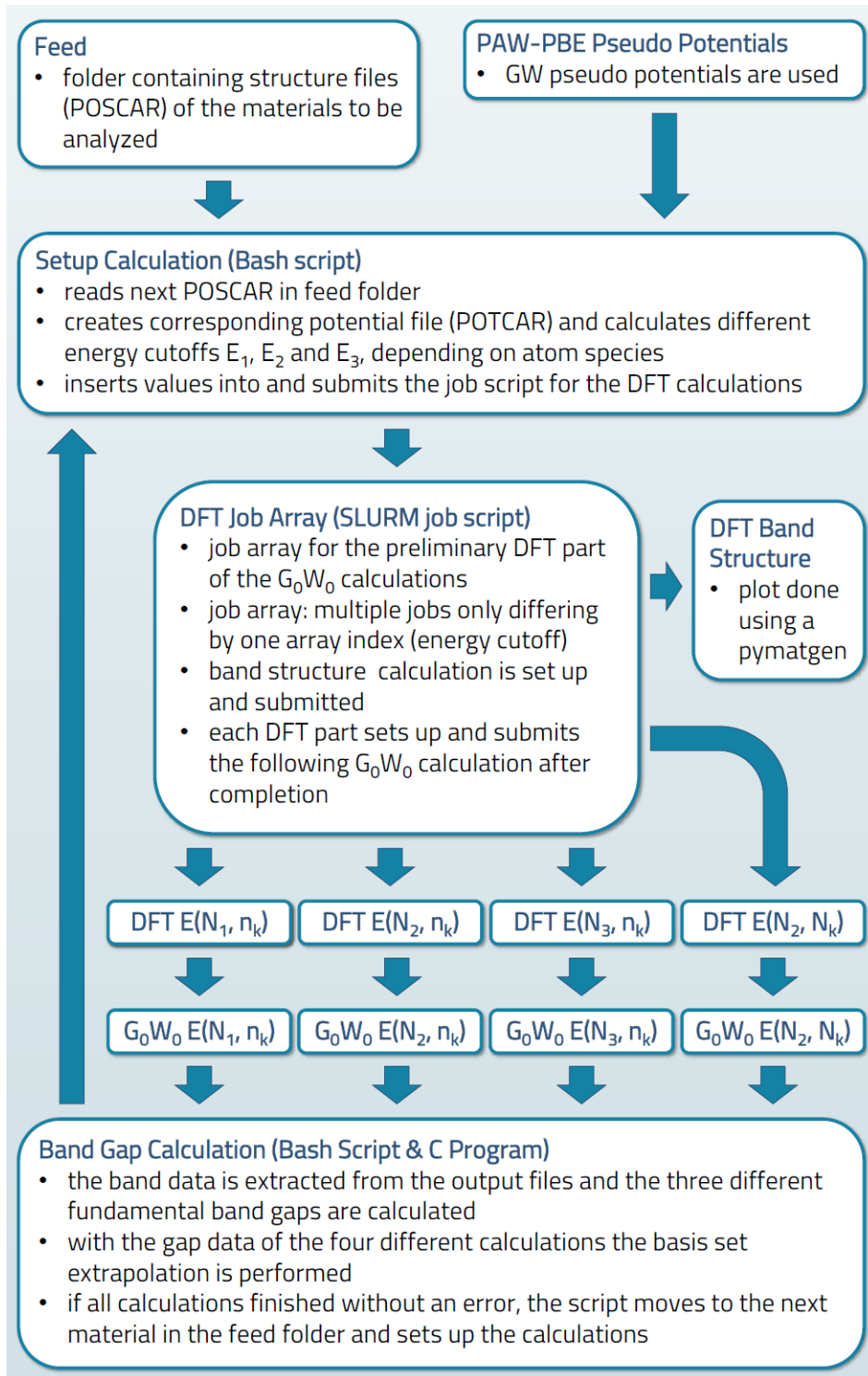


Figure 3.3: Structure of the automatic GW scheme: N_i are the different number of energy bands used for the extrapolation, E_i are the corresponding energy cut-off values ($i = 1, 2, 3$). n_k denotes a calculation done using a coarse k-point grid, where N_k stands for a dense grid respectively.

3.2.2 Scripts and Programs

In this section the different scripts and programs used in the automatic procedure will be briefly broken down in their input, output and what their respective function is.

These include not only Bash scripts for calculating band gaps, or programs for doing the actual extrapolation to an infinite basis set, but also the job scripts which are submitted to the computing cluster.

[Bash] Setup-Calculation Script

Input:

- *POSCAR* of the material to simulate
- *POTCARs* to construct the material
- material counter (*matcount*) which *POSCAR* is next to be read in

Output:

- calculation ID
- *POTCAR* file corresponding to the investigated material
- first job script to submit (*GW1*)

Function:

- increase counter and read in next *POSCAR*, if the file *DEADEND* is reached abort the calculation series
- define the calculation ID (*material_timestamp*)
- create the folder environment for the calculation to be performed
- copy the *POSCAR* file to the calculation directory
- create the *POTCAR* corresponding to the material at hand
- calculate the different energy cut-offs (*ENCUT*) for the basis set extrapolation based on information contained in the *POTCAR*
- setup the first job script (*GW1*) and submit the calculation

This script is the starting point of every basis set extrapolation procedure. Its functionality can be split up in several different parts:

1. First, it checks if there are structure files (*POSCAR*) in the feeding folder. If not, or all materials provided have already been analyzed (script reached *DEADEND*), it stops the test series. If there are still materials left it continues as follows.

2. It fetches the next *POSCAR* provided in the feeding folder. Then the script creates necessary working- and output- folders for the simulation, using the material name (sum formula and international symbol for its configuration) and a time stamp as naming convention.
3. As a next step, the to the material corresponding *POTCAR* is created with the atom information from the *POSCAR* file.
4. According to the potentials used, the different energy cut-off values (*ENCUT*) for the basis set extrapolation are calculated and the *GW1 script* is copied to the working folder, edited and submitted.

[SLURM array job] GW1 Script

Input:

- material definition (sum formula and international configuration symbol)
- calculation ID (material_timestamp)
- energy cut-off values for the basis set extrapolation

Output:

- *INCAR* and *KPOINTS* files for the different DFT calculations involved in the basis set extrapolation
- *INCAR*, *KPOINTS* and job files (*GW2*) for the G0W0 calculations
- *INCAR* and job file (*Band Structure Script*) for the band structure calculation

Function:

- submit an array job with three jobs, each corresponding to one energy cut-off value for the basis set extrapolation
- for the lowest energy cut-off:
 - * setup the preliminary DFT calculation (*INCAR*, low-density *KPOINTS*)
 - * execute *VASP*
 - * setup G0W0 calculation (folders, *INCAR*)
 - * call the *GW2* script to create the G0W0 job file
 - * submit GW job
- for standard energy cut-off (highest *ENMAX* in *POTCAR*):
 - * setup the preliminary DFT calculation (*INCAR*, low-density *KPOINTS*)
 - * execute *VASP*

- * setup G0W0 calculation (folders, *INCAR*)
- * call the *GW2* script to create the G0W0 job file
- * submit GW job
- * setup DFT calculation for k-point correction (*INCAR*, high-density *KPOINTS*)
- * execute *VASP*
- * setup G0W0 calculation (folders, *INCAR*)
- * call the *GW2* script to create the G0W0 job file
- * submit GW job for k-point correction
- for highest energy cut-off:
 - * setup the preliminary DFT calculation (*INCAR*, low-density *KPOINTS*)
 - * execute *VASP*
 - * setup band structure calculation (folders, *INCAR*)
 - * call k-point generation script for band structure calculation (*generate_kpoints*)
 - * call script to setup the band structure job file (*bandstructure* script)
 - * submit band structure job
 - * setup G0W0 calculation (folders, *INCAR*)
 - * call the *GW2* script to create the G0W0 job file
 - * submit GW job

This is the first job script submitted to the computing cluster. It creates an array job with three parts, each one performing a preliminary DFT calculations as input to the G0W0 simulations used in the basis set extrapolation.

The script also automatically sets up the remaining input files (*INCAR*, *KPOINTS*), before starting the calculation. After every finished DFT calculation¹ the *GW2* script is called to continue with the G0W0 part. Additionally, after the DFT calculation involving the highest number of bands the *band structure script* is called.

¹What is here referred to as one DFT calculation consists in reality out of two separate calculations: (1) First, a self consistent DFT calculation is performed to find the electronic ground state of the system at hand and (2) then the flag *LOPTICS* = *.TRUE.* and the final number of bands *N* is added in the *INCAR*, before only one electronic iteration (*NELM* = 1) is performed to write all files required as an input for the G0W0 step afterwards.

[Bash] GW2 Script

Input:

- calculation ID
- energy cut-off
- number of nodes and computing cores

Output:

- final results and output files of the G0W0 calculations used for the basis set extrapolation

Function:

- setup job file for the G0W0 calculation with the given parameters (*ENCUT*, number of nodes and cores)
- function of the job file:
 - * execute *VASP*
 - * create output folders in the data repository and store important result files
 - * call *Calculate-Gap* script
 - * remove large files like *WAVECAR* or *WAVEDER* to clear up space
 - * call the *Check-Finish* script

This script creates the SLURM job script for the G0W0 calculation, performed after each of the preliminary DFT simulations. Inside the job file, after a successful GW calculation the *calculate-gap* script is called to analyze the output data, before the *check-finish* script is invoked as a last step.

[Bash] Generate k-points Script

Input:

- *OUTCAR* of the preliminary DFT calculation
- *KPOINT* file from the repository containing the right path through reciprocal space for the unit cell

Output:

- *KPOINT* file used for the band structure calculation

Function:

- read in the *OUTCAR* file produced by the preliminary DFT calculation at highest energy cut-off and find unit cell symmetry
- fetch the right *KPOINTS* file from the repository fitting the materials crystal structure

This script looks at which crystal symmetry has been found by *VASP* during the DFT calculation and fetches the right *KPOINTS* file for the band structure calculation from the repository. Some systems require the calculation of material specific parameters, for example when distortions or elongations are involved. For these more complicated crystal structures there was no automatic procedure implemented and the band structure calculation omitted.

[SLURM job] Band Structure Script

Input:

- calculation ID
- SLURM job array ID

Output:

- output files of the band structure calculation
- number of bands

Function:

- run *VASP* to perform the band structure calculation
- store relevant files (*OUTCAR*, *INCAR*, *POSCAR*, *KPOINTS*, *vasprun.xml*)
- call the *calculate bands* script to calculate and the *plot bands* script to plot the band structure
- store the band structure and its plot as final results

This script performs a simple band structure calculation at DFT level. Afterwards, the Python Materials Genomics (pymatgen) package [20] is used to extract and plot the band structure from output produced by this calculation.

[Bash] Calculate Gap Script

Input:

- *OUTCAR* of the G0W0 calculation

Output:

- raw data three different types of band gaps and quasi particle shifts (direct at Gamma, smallest direct, smallest indirect) at the different energy cut-offs used for the basis set extrapolation
- corresponding number of bands to each gap

Function:

- extract number of bands from the provided *OUTCAR*
- calculate band gap and QP-shift at Gamma, store values together with the number of bands in an output file
- find conduction band minimum and valence band maximum
- calculate smallest indirect gap between these two points and corresponding QP-shift and store values together with the number of bands in an output file
- find smallest direct gap at any k-point and the corresponding QP-shift and store values together with the number of bands in an output file

This script is invoked after every G0W0 calculation. It calculates the different band gaps (direct gap at Gamma, and smallest indirect gap, smallest direct gap), as well as the corresponding quasi-particle shifts from data contained in the *OUTCAR* file.

The data is stored several output files (one for each kind of gap and QP-shift) to be used as an input for the *basis set extrapolation* program.

As later mentioned in section (5.1.2) one has to be cautious when calculating the k-point correction because of the different density of the k-point grids.

[Bash] Check Finish Script

Input:

- counter indicating how many G0W0 calculations already finished (stored in the *fincount* file)

Output:

- updated counter

Function:

- read in the counter of already finished G0W0 calculations
- if all are finished, reset the counter and call the *calculate gap* script to finish the basis set extrapolation and the *setup calculation* script to start analyzing the next material
- if not all have finished yet update the counter

This script is called at the end of each single-shot GW calculation. It checks how many of the four G0W0 parts of the basis set extrapolation have already finished. If all of them are done, the *basis set extrapolation* program is called over auxiliary scripts. Afterwards, if everything finished successfully, the *setup calculation* script is started again with the next material to investigate.

[C] Basis Set Extrapolation Program

Input:

- calculation ID
- output files from the *calculate gaps* script containing band gaps and QP-shifts at different values of *NBANDS* (number of bands)

Output:

- final, extrapolated band gaps and QP-shifts

Function:

- for each band gap type read in the values at different number of bands
- perform a linear extrapolation to an infinitely large basis set ($1/NBANDS \rightarrow 0$)
- gather and write all results into the final result file and store it

This program takes the stored data on band gaps and QP shifts as an input. It performs the basis set extrapolation and k-point correction for all these quantities. The corrected gaps and quasi-particle shifts are stored in the final result file.

3.3 Computational Details

All computational details will be discussed in this section. A goal of the project was to find a computational setup which delivers accurate results, while being resource friendly and univer-

sally applicable to as many systems as possible.

All calculations were performed on the Vienna Scientific Cluster 3 (VSC3), using its standard nodes. Each node is equipped with 16 computing cores distributed over two processors, backed by 64 GB of memory, and are interconnected via a high-speed InfiniBand fabric.[21]

3.3.1 G0W0 Calculations

An example *INCAR* file, as used for the G0W0 employed in the basis set extrapolation can be found below.

```
!G0W0 INCAR                                !smearing
ALGO = GW0                                ISMEAR = 0
LSPECTRAL = .TRUE.                        SIGMA = 0.05
NOMEGA = 96                               EDIFF = 1E-6
ENCUT = [E1, E2, E3]
NBANDS = [N1, N2, N3]                !parallelization
MAXMEM = 5000                             KPAR = 2
```

Since the high-throughput should be as simple as possible, many parameters were kept at the VASP default. For example, the energy cut-off of the auxiliary basis set E_χ , defined by the flag *ENCUTGW* was kept at the default value of $2/3 \times$ the cut-off *ENCUT*.

Switching on the spectral method via *LSPECTRAL* helps to reduce the computational workload of G0W0 calculations at the cost of higher memory requirements, the flag *MAXMEM* sets the maximally available memory per computing core (in MB), while also influencing the allocation of the response function. The rank of said response function is determined by *NOMEGA*, which should be dividable by the number of computing cores for optimal parallelization. This value also influences the results accuracy [17] and a value around 100 was deemed sufficient.

The energy cutoff *ENCUT* and corresponding number of band *NBANDS* was, of course, individually chosen and varied for the sake of performing the basis set extrapolation ($[E_1, E_2, E_3]$, $[N_1, N_2, N_3]$).

For these calculations, 8 nodes with 12 computing cores each were used at the VSC3, resulting in a total of 96 cores working in parallel. Although each node would be equipped with a maximum of 16 cores, more memory was required for the large response functions. By only using 12 cores on each node, the memory per core was increased from 4 GB to 5 GB and thus sufficiently large for the simulations.

For the calculations with norm conserving pseudo-potentials to compare the automatic procedure in section (4.1.1) to results obtained in reference [2], the flag *NMAXFOCKAE* = 2 was added to the G0W0 *INCAR*. This is suggested and used in said reference to ensure a better augmentation of the pseudo wave functions on the plane wave grid.

3 Methods

The calculations using norm conserving pseudo-potentials were done on the VSC4 due to the excessive amount of computational resources required. Also here, the amount of tasks per node (number of cores used per node) had to be reduced to ensure sufficient memory. The calculations were done on 8 nodes, each using only 24 of the available 48 cores to increase the memory per core from 2 GB to 4 GB.

Parallelization flags, such as *NCORE* or *NPAR* are not supported for the G0W0 calculations and thus have to be kept at default, even for the preliminary DFT calculations.

3.3.2 DFT Calculations

The DFT calculations are the starting point for each G0W0 calculation. While these simulations are rather simple to perform and require only little computational resources in comparison to GW simulations, the resulting wave function has to be discretized in a fitting way. In practice, this means that the same number of computing cores has to be employed as are used later in the G0W0 simulations. Only then the wave function will be parallelized in the right way.

Although less resources would be necessary, because of the reasons above already here 96 cores were used to perform the DFT calculations.

In practice, the DFT calculation part itself corresponds to two separate simulations. For the first step (*INCAR 1*), a standard DFT calculation is performed to find the electronic ground state. This calculation is done at the corresponding energy cut-off needed for the basis set extrapolation, and an *exact diagonalization* algorithm (*ALGO = Exact*) is chosen to prevent errors.

```
!DFT INCAR 1
ALGO = Exact
ENCUT = [E1, E2, E3]

!smearing
ISMEAR = 0
SIGMA = 0.05
EDIFF = 1E-6

!parallelization
KPAR = 2
```

Once this is done, only one electronic update step (*NELM = 1*) is done in the second step (*INCAR 2*). Taking the ground state wave function from the first part as input, empty bands are added to reach the maximal number of bands compatible with the energy cut-off. After only one step, the wave function derivative (*WAVEDER*) and the frequency dependent dielectric matrix are calculated (*LOPTICS = .TRUE.*). These are needed for the GW calculation afterwards to serve as input. Since the electronic ground state was already found in the first DFT part, this step is only needed to write the output with the right number of bands.

```

!DFT INCAR 2
ALGO = Exact
NELM = 1
ENCUT = [E1, E2, E3]
NBANDS = [N1, N2, N3]
LOPTICS = .TRUE.

!smearing
ISMEAR = 0
SIGMA = 0.05
EDIFF = 1E-6

!parallelization
KPAR = 2

```

3.3.3 Band Structure Calculations

The band structure on DFT level was also calculated for each system as part of the automatic scheme. For this step, the largest energy cutoff (E_3) was used for maximal accuracy. The k-point paths between the high symmetry points in reciprocal space were taken from reference [10], for the Python Materials Genomics package (*pymatgen*) was used.[20]

```

!band structure INCAR
ICHARG = 11
LORBIT = 11
ENCUT = E3

!smearing
ISMEAR = 0
SIGMA = 0.1

```

3.4 Calculating the Band Gaps

When performing experiments to investigate the electronic properties of materials, different kinds of band gaps can be found depending on the electronic excitation. Because of this, three different types of band gaps were calculated for this project.

To calculate the band gaps, the *Calculate-Gap script* retrieves the conduction- and valence-band energies at the corresponding k-points from the *OUTCAR* file.

3.4.1 Gamma Gap

The direct band gap at the Gamma point is the first one which was calculated. This band gap can often be found as a reference in computational physics publications, what makes it an important quantity to investigate. Because of the k-point grid, band gaps can not be calculated at arbitrary points in reciprocal space. For many calculations this discrete lattice has its center at Gamma, making it often a good choice to use.

This gap is calculated by simply taking the energy difference of the valence and conduction band at the k-point located at Gamma ($[0, 0, 0]$).

3.4.2 Smallest Indirect Gap

Indirect band gaps are more complicated to investigate than direct ones, at least in experiments. This is because to address an indirect gap, phonons have to be involved in the excitation of electrons. This makes it a much more complicated process to describe or even use practically. Nevertheless, if the indirect gap is much smaller than the direct gap, or the crystal symmetry promotes these kinds of excitation, it is also an important quantity.

The script finds the valence band maximum (VBM) and conduction band minimum (CBM) for calculating the gap between those two points.

Interpreting the indirect band gap at GW level has to be done carefully, however. Since the band energies are only known at very few points in reciprocal space (k-points), one has to make sure that the gap between these points is physically relevant. Thus, this value is not as significant as the Gamma gap in some cases for comparing results between different simulations or experiments.

3.4.3 Smallest Direct Gap

In most experimental setups, this is the gap which is investigated. Many experiments involve photonic excitation, which only responds to direct gaps and usually the energy dissipates using the smallest one. This gap is calculated by finding the smallest band gap at any one k-point.

Unfortunately, the same problem as already mentioned for the smallest indirect gap occurs here for GW calculations. The gap is only calculated on select few points in the reciprocal unit cell, if the systems smallest direct gap lies somewhere in between it cannot be retrieved. This makes it especially problematic when trying to compare computational results to experiments.

4 Results

For this project, a broad range of materials was analyzed automatically, without any user interference during the whole procedure. The material set contains 69 insulators and semi-conductors. Mostly mono- and diatomic materials were investigated, but also a small amount of compounds with more atoms in the unit cell. All computational parameters were set at the beginning and are the same for every material, with the exception of automatically calculated material specific parameters such as the exact energy cut-off or number of bands for the basis set extrapolation. This might cause some inaccuracies in the results, but optimizing all parameters for each system individually would go against the paradigm that is tried to be enforced here.

For all systems the fundamental band gap at GW level was calculated, using a less memory demanding variant of basis set extrapolation as proposed in reference [2]. During the procedure, three different gap types (direct at the Gamma point, smallest indirect gap, smallest direct gap at any one k-point) were calculated, as well as the band structure at DFT level.

4.1 Validating the Method

The ultimate goal of performing simulations is producing data which predicts reality as close and reliable as possible.

Since the here introduced automatic procedure is an early attempt at adopting a newly developed method for high-throughput calculations this goal is not the first criteria to look at, however. In reference [2] it was shown that the basis set extrapolation is able to produce very high quality gap data when performed carefully. It only makes sense to try reaching similarly accurate results through an automatic procedure as a first test of feasibility.

4.1.1 Comparison to the Original Basis Set Extrapolation

As a first validation, some results obtained with the procedure used here should be compared to data from reference [2], where the basis set extrapolation was introduced and investigated very carefully (see tables (4.1) and (4.2)). The calculations performed by Klimeš *et al.* were focused on very safe and accurate results, optimizing parameters for each material separately, while also using the best choice of pseudo-potentials in the form of norm-conserving PAW potentials.

As already mentioned, the procedure here does not allow to perform such an individual treatment of every single system and a general set of parameters should be found which works universally for most materials investigated. Because of this, it has to be expected that the results deviate from the accurate data, depending on the sensitivity to sub-optimal parameter choices.

Material	Gap@ Γ [eV] ref. [2]	Gap@ Γ [eV] NC PP	Gap@ Γ [eV] mem. cons.
AlAs	2.99	3.11 (+ 0.12)	3.31 (+ 0.32)
AlP	4.23	4.32 (+ 0.09)	4.47 (+ 0.24)
AlSb	2.40	2.57 (+ 0.17)	2.65 (+ 0.25)
BN	11.33	11.30 (- 0.03)	11.61 (+ 0.28)
C	7.43	7.44 (+ 0.01)	7.49 (+ 0.06)
CdO	<i>1.50</i>	<i>1.72 (+ 0.22)</i>	<i>1.69 (+ 0.19)</i>
CdS	2.15	<i>1.06 (- 1.09)</i>	2.18 (+ 0.03)
CdSe	1.52	1.83 (+ 0.31)	1.61 (+ 0.09)
CdTe	1.57	1.74 (+ 0.17)	1.80 (+ 0.23)
GaAs	1.23	1.32 (+ 0.09)	2.08 (+ 0.85)
GaN	2.85	3.02 (+ 0.17)	3.94 (+ 1.09)
GaP	2.62	2.69 (+ 0.07)	3.36 (+ 0.73)
GaSb	<i>0.68</i>	<i>0.61 (- 0.07)</i>	1.54 (+ 0.78)
InAs	<i>0.23</i>	<i>0.18 (- 0.05)</i>	<i>0.30 (+ 0.07)</i>
InP	1.23	1.39 (+ 0.16)	1.43 (+ 0.20)
InSb	<i>0.25</i>	<i>0.16 (- 0.09)</i>	<i>0.33 (+ 0.08)</i>
MgO	7.55	7.56 (+ 0.01)	7.86 (+ 0.31)
Si	3.25	3.24 (- 0.01)	3.28 (+ 0.03)
SiC	7.35	7.39 (+ 0.04)	7.52 (+ 0.17)
ZnS	3.46	3.48 (+ 0.02)	3.54 (+ 0.08)
ZnSe	2.55	2.74 (+ 0.19)	2.50 (- 0.05)
ZnTe	2.27	2.39 (+ 0.12)	2.35 (+ 0.08)

Table 4.1: Comparison of calculated fundamental band gaps at Gamma to the results presented in reference [2]. For the middle column a very similar setup as proposed in said reference was used, employing the original basis set extrapolation and norm conserving pseudo-potentials. Although norm conserving, the potentials used here are not exactly the same ones as in the reference and thus also the energy cut-offs and band numbers for the extrapolation are slightly different. This can lead to significant differences in the results, as shown later in section (4.1.4). In the last column results obtained with the memory conserving variant are presented, which was used to determine the final result set of fundamental gaps. Results in italic show wrong band occupations in DFT and after one G0W0 calculation, as discussed later in section (4.1.2)

For this validation, a setup was created which reproduces the one used in the reference paper as close as possible, while still using the automatic calculation framework.

Norm conserving pseudo-potentials were used, as well as the original energy cut-off progression of $(E_1, 1.25 \times E_1, 1.57 \times E_1)$. The first cut-off value E_1 was chosen as the highest suggested value in the used potentials ($ENMAX$ in the corresponding potential file). This is already a difference between the calculations done here, since in reference [2] all potentials were optimized and consequently different values for the energy cut-off were chosen. Together with probably different parallelization settings this leads to deviations in the numbers of energy bands N_B

Material	ΔIP_{VB} [eV] ref. [2]	ΔIP_{VB} [eV] NC PP	ΔIP_{VB} [eV] mem. cons.
AlAs	-1.01	-0.85 (+ 0.16)	-0.81 (+ 0.20)
AlP	-0.96	-0.87 (+ 0.09)	-0.90 (+ 0.06)
AlSb	-0.84	-0.69 (+ 0.15)	-0.78 (+ 0.06)
BN	-1.53	-1.48 (+ 0.05)	-1.45 (+ 0.08)
C	-1.22	-1.18 (+ 0.04)	-1.11 (+ 0.11)
CdO	-0.95	-0.97 (- 0.02)	-0.96 (- 0.01)
CdS	-1.11	-0.66 (+ 0.45)	-1.15 (- 0.04)
CdSe	-1.15	-1.13 (+ 0.02)	-1.03 (+ 0.12)
CdTe	-0.99	-0.88 (+ 0.11)	-0.90 (+ 0.07)
GaAs	-0.91	-0.72 (+ 0.19)	-0.76 (+ 0.15)
GaN	-1.20	-1.25 (- 0.05)	-1.26 (- 0.06)
GaP	-0.86	-0.74 (+ 0.12)	-0.83 (+ 0.03)
GaSb	-0.77	-0.54 (+ 0.23)	-0.77 (+ 0.00)
InAs	-0.83	-0.68 (+ 0.15)	-0.66 (+ 0.17)
InP	-0.78	-0.73 (+ 0.05)	-0.78 (+ 0.00)
InSb	-0.75	-0.55 (+ 0.20)	-0.69 (+ 0.06)
MgO	-2.16	-2.01 (+ 0.15)	-2.02 (+ 0.14)
Si	-0.72	-0.67 (+ 0.05)	-0.72 (+ 0.00)
SiC	-1.05	-1.00 (+ 0.05)	-0.99 (+ 0.06)
ZnS	-1.33	-1.18 (+ 0.15)	-1.16 (+ 0.17)
ZnSe	-1.34	-1.24 (+ 0.10)	-1.05 (+ 0.29)
ZnTe	-1.07	-0.93 (+ 0.14)	-0.89 (+ 0.18)

Table 4.2: Comparison of quasi-particle shifts ΔIP_{VB} proposed in reference [2] and calculated in this project. These shifts are calculated with respect to the valence band maximum, i. e. a negative shift means a lowering in the corresponding energy after the G0W0 step.

used for the extrapolation.

Later in section (4.1.4) it is shown that even slight changes in these numbers can lead to significant differences in the gap value. This means although using norm conserving pseudo-potentials and good settings it has to be expected that the results presented in the paper are not completely accurately reproduced.

Another difference is the reduced number of frequency points for the response function, since all calculations done for this project were done using $NOMEGA = 96$, while the results shown in reference [2] are produced using $NOMEGA = 200$. As claimed in the paper, this should not change the results significantly (for most materials well below 50 meV), however it still might add to some errors and was a necessary step to ensure that there is enough memory available.

Nevertheless, for many materials quite good accuracy was achieved, as shown in the tables (4.1) and (4.2). Many fundamental band gaps and quasi-particle shifts calculated here show good agreement with the optimized values in the reference.

4 Results

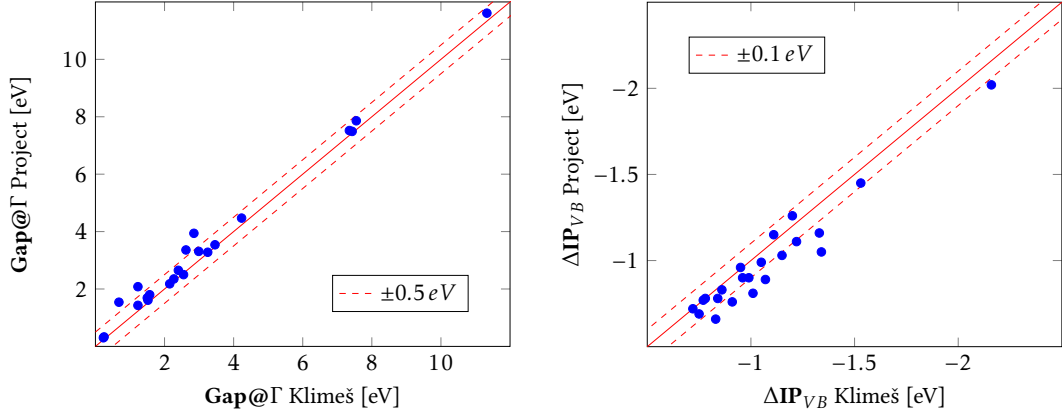


Figure 4.1: Comparison between band gaps and quasi-particle shifts acquired here in the project and data as proposed by Klimeš *et al.* in the original basis set extrapolation paper.[2]

Results achieved with the here adopted standard procedure, using the memory conserving variant of $(0.75 \times E_1, E_1, 1.25 \times E_1)$ and no norm conserving pseudo-potentials, can also be found in these tables. They match the other results relatively well, for some materials the fundamental gap compares even better to the reference than the ones computed using norm conserving potentials. For others, like all the gallium compounds, it can be seen that the norm violation leads to significant errors.

Materials which are marked in *italic* show wrong band ordering and occupation in DFT, leading to metallic properties. This will be further discussed in section (4.1.2)

A graphical representation can be found in figure (4.1), where the accurate gaps from reference [2] are compared to the results obtained with the here adopted, less accurate procedure. Due to the almost consistently higher values of the Gamma gaps it is clearly visible that the band gap correction is reduced because of the lower accuracy.

4.1.2 Wrongly Placed or Occupied Bands

Some materials show metallic band properties in DFT, with wrongly placed or occupied bands. To be precise, CdO, CdS, GaSb, InAs, and InSb show all such behavior. This was also observed in reference [2], where it was claimed that the energies of the degenerate cation p-orbitals incorrectly lie above the anion s-orbitals at Gamma. Additionally, these s-orbitals are fully occupied and the three degenerate ones only partially.

While the band ordering gets fixed after the G0W0 step, the wrong occupation still cannot be fixed with only one GW iteration, such that the conduction band minimum is still occupied and below the Fermi level. A possibility would be to do more iterations of the GW step, which leads to an opening of the right fundamental gap and correction of the occupation.

However, the extrapolation should still be based on G0W0 calculations only. Because of this, the problem was treated as suggested in reference [2], where the threefold degenerate bands were identified as the true valence orbitals after a single G0W0 step. Then the gap was

Material	original [eV]	mem. cons. [eV]	Error (abs) [eV]	Error [%]
AlAs	3.29	3.36	0.07	+2.13
AlP	4.41	4.51	0.10	+2.27
AlSb	2.61	2.73	0.12	+4.60
BN	11.64	11.80	0.17	+1.46
C	7.53	7.58	0.05	+0.66
CdO	15.33	15.38	0.05	+0.33
CdS	2.15	2.23	0.08	+3.72
CdSe	1.63	1.62	0.01	-0.61
CdTe	1.82	1.82	0.00	+0.00
GaAs	2.15	2.00	0.15	-6.98
GaN	3.91	4.04	0.13	+3.32
GaP	3.41	3.31	0.10	-2.93
InP	1.49	1.41	0.08	-5.37
MgO	7.67	8.12	0.45	+5.87
Si	3.29	3.28	0.01	-0.30
SiC	7.70	7.57	0.13	-1.69
ZnS	3.58	3.55	0.03	-0.84
ZnTe	2.38	2.32	0.06	-2.52

Table 4.3: Comparison of Gamma band gaps calculated using the original procedure with higher energy cut-offs and the memory conserving variant described in section (3.1.1).

calculated between these bands and the next one higher in energy, although this band is still occupied. This is not the best approach, since the band energies change further during the correction of the band occupation through more GW steps. Although moving away from single G0W0 calculations and employing an more or less arbitrary amount of iterations until the occupation is fixed would again lead to an semi-empirical approach and it is clear why the authors refrained from it in the reference.

For FeO, MnO, and NiO metallic solutions are found in DFT, which again cannot be corrected by only one G0W0 iteration. Additionally, all calculations performed here were done without taking into account magnetic moments, probably leading to an even larger error for these magnetic materials.

4.1.3 Accuracy of the Memory Conserving Variant

In this project a slight change was made to the original basis set extrapolation, as it was proposed in reference [2]. Here, the calculation at the highest energy cut-off was replaced by a less accurate one, as described in section (3.1.1).

In table (4.3) the fundamental band gaps at Gamma can be found for a test set of some materials, once calculated with the original and once with the memory conserving basis set extrapolation scheme. It was found that this less accurate variant introduces errors well below 10%, with a maximum error of 6.98% for GaAs.

4 Results

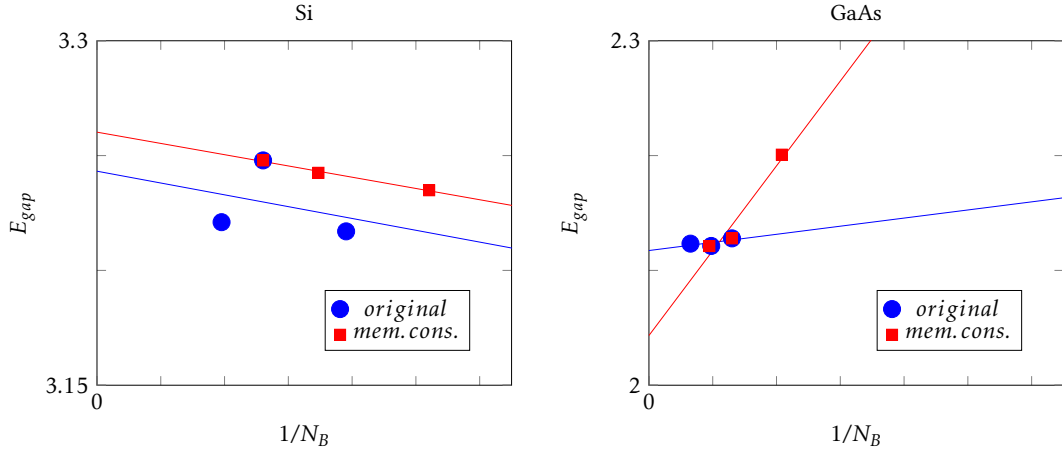


Figure 4.2: Comparison between extrapolations done within the original basis set extrapolation scheme (E_1 , $1.25 \times E_1$, $1.57 \times E_1$) and the memory conserving variant ($0.75 \times E_1$, E_1 , $1.25 \times E_1$). Left: For Si the memory conserving scheme reproduces the trend very well. Only a small error in the gap energy (0.01 eV) is introduced as shift of the extrapolation function. Right: For GaAs the trend is not reproduced correctly. It can be seen that the calculation at the lowest energy cut-off introduces a large error (0.15 eV), resulting in a higher difference between the fundamental gaps found by the two variants.¹

In figure (4.2) a possible explanation can be found, when this variant yields a good gap and when it introduces a large error. For Si the error obtained was very small and can be interpreted as the shift experienced by the extrapolation. For GaAs, however, the overall trend is changed by replacing the one calculation. This leads to a significant change in the predicted gap and for such cases the memory conserving variant is a sub-optimal choice.

Although some errors are quite significant for high-accuracy calculations, in the frame of this automated scheme and the fact, that it was a necessary step to guarantee sufficient memory for all materials investigated, the additional error is in an acceptable range for most materials.

Furthermore, in section (4.1.4) it is shown that even for slight changes in the number of bands used for the extrapolation, large deviations of the trends have to be expected. This seems to be mostly due to the small number of only three data points used for the procedure and the fact, that the fundamental gap not always shows a perfect linear trend.

4.1.4 Influence of Parallelization

The number of energy bands included in calculations cannot be chosen arbitrary, for a given energy cut-off there is a maximum number of bands which are supported. This number is also

¹The energy cut-offs are chosen such that for both, the original and the memory conserving variant, $E_1^{original} = E_1^{mem.cons.}$. However, it is visible that the fundamental gaps are not calculated at exactly the same number of bands in figure (4.2). This is because the number of bands must be compatible to the number of computing cores used in the calculation and calculations using the original variant were done at different parallelization settings. As discussed later, this can significantly influence the result, but is not relevant for the explanation when the memory conserving variant is a good choice and when it falls short.

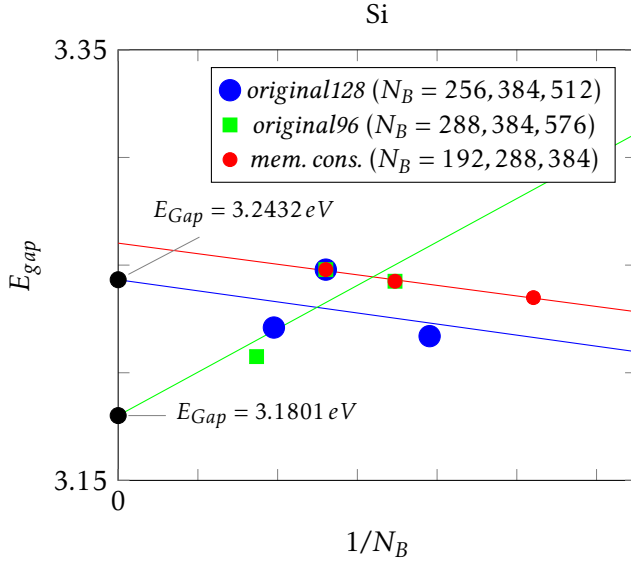


Figure 4.3: Three different basis set extrapolations for silicon. For the first two (*original128* and *original96*) the original scheme was used ($E_1, 1.25 \times E_1, 1.57 \times E_1$), although once using 128 cores (*original128*) and once using 96 respectively (*original96*). The third extrapolation was done with the memory conserving variant ($0.75 \times E_1, E_1, 1.25 \times E_1$) using 96 cores. The trend shown in the *original96* extrapolation is completely different than for the other two, although using a very similar set of parameters as the memory conserving one. This leads to an error of 0.06 eV (or approximately 2%) of the estimated gap.

influenced by the parallelization settings, i. e. over how many computing cores the calculation is distributed. The number of bands has to be chosen as a multiple of the number of cores, constraining which values can be chosen for the extrapolation.

The basis set extrapolation is done with only three data points, what makes the result quite sensitive to outliers. This can be seen already in figure (4.2) for GaAs, where one data point changes the predicted fundamental gap significantly. In this example, this might be because of the low energy cut-off for said point and thus a lower accuracy of the corresponding calculation. However, similar behavior is also found for very comparable calculations such as shown in figure (4.3).

Here, two very similar basis set extrapolations (*original128* and *original96*) are presented, both using similar numbers of bands ($N_B^1 = 256, 384, 512$ and $N_B^2 = 288, 384, 576$) but different numbers of cores, with 128 cores employed for *original128* and 96 cores for *original96*. It is clearly visible that these two setups produce vastly different trends in the extrapolation and consequently different fundamental band gaps.

Furthermore, even with the memory conserving extrapolation shown in figure (4.3) having a very similar setup compared to the *original96* extrapolation, the trend of *original128* matches much better. This hints strongly, that the procedure is very sensitive to even slight changes in the basis sets used for the extrapolation and even similar setups can produce significantly deviating results.

To increase the accuracy and reliability of results it might be necessary to use more than three data points for the extrapolation to guarantee a reproduction of the right trend when using this procedure. However, this would mean even larger memory requirements since the number of bands can only be increased in multiples of the computing cores used for the calculations.

The error arising due to this uncertainty in the extrapolation is given for each quantity calculated in this project and presented in tables (4.4a) and (4.4b). The derivation of this error can be found in the appendix.

4.2 Automatic GW series

The automatic basis set extrapolation scheme, visualized in figure (3.3), was applied on a set of 69 different materials, consisting of different insulators and semi-conductors. Most of them are mono- or diatomic, but also some materials containing more than two atoms per unit cell were included.

For all materials three different types of fundamental band gaps were calculated: (1) the direct gap at the Gamma point, (2) the smallest direct gap at any k-point, as well as (3) the smallest indirect gap between any two k-points.

Furthermore, for each gap the quasi-particle shift (ΔIP) with respect to the corresponding valence band maximum (VBM) was calculated. This property is the shift which occurs in the band energies due to corrections caused by the G0W0 calculation. For most materials this results in a further opening of the fundamental gap compared to DFT results, which corresponds to a lowering of the VBM and thus a negative sign of the shifts.

As discussed in the previous sections, several problematic aspects can contribute to an error in the determined values of these gaps. Not only is the extrapolation to an infinite basis set very sensitive to the numbers of bands used for the three calculations it is based on, but also norm violations due to the lack of norm conserving pseudo-potentials, as well as the less accurate memory conserving scheme may negatively influence the results.

For a better quantification of the error the gaps might have, for each value the uncertainty arising due to the extrapolation is given. This value is an indicator how good the data reproduces a linear trend, where a lower value means less deviation from the linear fit. An example for this would be the red data set (small points) in figure (4.3). The other two lines (green squares and large blue dots) in the same figure have a larger uncertainty. A description how this error is calculated can be found in the appendix.

4.2.1 Automatically Calculated Band Gaps

In this section all band gaps and quasi-particle shifts obtained with the automatic basis set extrapolation scheme are presented. The collected data can be found in tables (4.4a) and (4.4b). For many systems the smallest direct gap is located at Gamma, in such cases only the Gamma gap is presented. The smallest indirect band gap in these tables corresponds to the minimal energy difference between the valence band maximum (VBM) and conduction band minimum (CBM). In some cases this coincides either with the Gamma gap, the smallest direct gap, or both. If this is the case, these gaps are also not explicitly mentioned. To summarize, if only the direct gap at the Gamma point is mentioned this means there is no smaller gap between any other k-points.

The total computation time, once the scheme was successfully implemented, was approximately one week. Since everything was scheduled automatically and in serial, the only unproductive time was when some of the calculations were pending due to a full cluster. No user interference was necessary and all results were automatically parsed and collected. Depending on the system, each part of the basis set extrapolation (one DFT calculation with G0W0 afterwards) took between 15 minutes and several hours, where the calculation with a dense k-point grid took always the longest to finish.

Material [configuration]	Gamma		Indirect		Direct	
	Gap [eV]	Δ IP [eV]	Gap [eV]	Δ IP [eV]	Gap [eV]	Δ IP [eV]
AgCl [Fm $\bar{3}$ m]	4.90 \pm 0.04	-1.24 \pm 0.05	2.32 \pm 0.07	-1.56 \pm 0.04		
AgF [Fm $\bar{3}$ m]	3.93 \pm 0.75	-3.16 \pm 0.78				
AgI [F $\bar{4}$ 3m]	2.53 \pm 0.08	-1.38 \pm 0.01				
AlAs [F $\bar{4}$ 3m]	3.31 \pm 0.00	-0.81 \pm 0.02	2.17 \pm 0.04	-0.81 \pm 0.02		
AlN [P $\bar{6}$ ₃ mc]	6.28 \pm 0.09	-1.46 \pm 0.02				
AlP [F $\bar{4}$ 3m]	4.47 \pm 0.03	-0.90 \pm 0.02	2.42 \pm 0.03	-0.90 \pm 0.02		
AlSb [F $\bar{4}$ 3m]	2.65 \pm 0.05	-0.78 \pm 0.02	1.73 \pm 0.03	-0.78 \pm 0.02		
BaBiO ₃ [Pm $\bar{3}$ m]	6.79 \pm 0.01	-1.17 \pm 0.02	2.08 \pm 0.02	-0.41 \pm 0.01	4.79 \pm 0.03	-1.21 \pm 0.01
BaF ₂ [Fm $\bar{3}$ m]	9.93 \pm 0.02	-2.38 \pm 0.01				
BaO [Fm $\bar{3}$ m]	5.85 \pm 0.01	-1.27 \pm 0.03	3.61 \pm 0.01	-1.30 \pm 0.03		
BaTiO ₃ [P4mm]	3.95 \pm 0.15	-0.93 \pm 0.03	3.51 \pm 0.10	-0.98 \pm 0.01		
BaTiO ₃ [Pm $\bar{3}$ m]	3.75 \pm 0.10	-0.96 \pm 0.03	3.56 \pm 0.06	-0.97 \pm 0.01		
BaTiO ₃ [R3m]	4.71 \pm 0.07	-0.99 \pm 0.01	4.23 \pm 0.02	-1.05 \pm 0.01		
BeO [P $\bar{6}$ ₃ mc]	10.94 \pm 0.03	-2.04 \pm 0.12				
BeTe [F $\bar{4}$ 3m]	4.20 \pm 0.04	-0.87 \pm 0.04	1.80 \pm 0.03	-0.87 \pm 0.04		
BN [F $\bar{4}$ 3m]	11.61 \pm 0.02	-1.45 \pm 0.07	6.55 \pm 0.12	-1.45 \pm 0.07		
C [Fd $\bar{3}$ m]	7.49 \pm 0.06	-1.11 \pm 0.06	6.26 \pm 0.04	-1.11 \pm 0.06		
C [P $\bar{6}$ ₃ /mmc]	8.22 \pm 0.10	-1.19 \pm 0.04	4.30 \pm 0.01	-0.69 \pm 0.01		
C [R $\bar{3}$ m]	7.92 \pm 0.08	-1.05 \pm 0.10	3.87 \pm 0.05	-0.63 \pm 0.06	4.55 \pm 0.07	-0.62 \pm 0.06
CaI ₂ [P $\bar{3}$ m1]	6.32 \pm 0.01	-1.61 \pm 0.01	6.02 \pm 0.00	-1.61 \pm 0.01		
CaO [Fm $\bar{3}$ m]	6.82 \pm 0.06	-1.51 \pm 0.00	6.39 \pm 0.02	-1.51 \pm 0.00	6.80 \pm 0.04	-1.56 \pm 0.01
CaS [Fm $\bar{3}$ m]	6.56 \pm 0.00	-1.26 \pm 0.01	4.32 \pm 0.02	-1.26 \pm 0.01	5.19 \pm 0.03	-1.32 \pm 0.02
CdCl ₂ [R $\bar{3}$ m]	5.68 \pm 0.23	-1.69 \pm 0.17	5.53 \pm 0.06	-1.81 \pm 0.04		
CdO [Fm $\bar{3}$ m]	1.69 \pm 0.06	-0.96 \pm 0.06				
CdS [F $\bar{4}$ 3m]	2.18 \pm 0.01	-1.15 \pm 0.05				
CdS [Fm $\bar{3}$ m]	2.54 \pm 0.01	-0.81 \pm 0.04	0.82 \pm 0.03	-0.18 \pm 0.01		
CdS [P $\bar{6}$ ₃ mc]	2.28 \pm 0.02	-1.16 \pm 0.03				
CdSe [F $\bar{4}$ 3m]	1.61 \pm 0.04	-1.03 \pm 0.08				
CdTe [F $\bar{4}$ 3m]	1.80 \pm 0.00	-0.90 \pm 0.00				
CsAu [Pm $\bar{3}$ m]	7.98 \pm 0.06	-1.71 \pm 0.03	2.36 \pm 0.02	-1.26 \pm 0.03	3.03 \pm 0.02	-1.26 \pm 0.03
CuCl [F $\bar{4}$ 3m]	1.61 \pm 0.09	-1.01 \pm 0.02				
FeAl [Pm $\bar{3}$ m]	2.45 \pm 0.11	-0.09 \pm 0.00	2.26 \pm 0.00	-0.02 \pm 0.01	1.35 \pm 0.04	-0.02 \pm 0.01
FeO [Fm $\bar{3}$ m]	1.27 \pm 0.02	0.61 \pm 0.01	0.63 \pm 0.00	0.38 \pm 0.02		
GaAs [F $\bar{4}$ 3m]	2.08 \pm 0.03	-0.76 \pm 0.00	1.99 \pm 0.00	-0.76 \pm 0.00		
GaN [F $\bar{4}$ 3m]	3.94 \pm 0.01	-1.26 \pm 0.00				

Table 4.4a: Calculated fundamental G0W0 band gaps (**Gap**) (direct gap at **Gamma**, smallest **direct** gap at any k-point, smallest **indirect** gap) and corresponding quasi-particle shifts (Δ IP) of the valence band maximum. The latter two gap types are only given if different from the fundamental gap at Gamma. Gaps reported in italic have show wrong band occupations or metallic solutions, see section (4.1.2).

4 Results

Material [configuration]	Gamma		Indirect		Direct	
	Gap [eV]	Δ IP [eV]	Gap [eV]	Δ IP [eV]	Gap [eV]	Δ IP [eV]
GaP [F $\bar{4}$ 3m]	3.36 \pm 0.02	-0.83 \pm 0.01	2.32 \pm 0.00	-0.83 \pm 0.01		
GaSb [F $\bar{4}$ 3m]	1.54 \pm 0.03	-0.77 \pm 0.00	1.45 \pm 0.01	-0.77 \pm 0.00		
HfO ₂ [Fm $\bar{3}$ m]	6.69 \pm 0.07	-1.38 \pm 0.00			5.62 \pm 0.02	-1.40 \pm 0.03
HgO [Imm2]	13.68 \pm 0.01	-0.98 \pm 0.02	3.30 \pm 0.01	-0.98 \pm 0.02	3.20 \pm 0.01	-1.07 \pm 0.02
InAs [F $\bar{4}$ 3m]	<i>0.30 \pm0.01</i>	<i>-0.66 \pm0.00</i>				
InN [F $\bar{4}$ 3m]	<i>0.32 \pm0.07</i>	<i>-0.81 \pm0.00</i>				
InP [F $\bar{4}$ 3m]	1.43 \pm 0.02	-0.78 \pm 0.01				
InSb [F $\bar{4}$ 3m]	<i>0.33 \pm0.03</i>	<i>-0.69 \pm0.00</i>				
KCl [Fm $\bar{3}$ m]	8.42 \pm 0.01	-2.52 \pm 0.04				
LiF [Fm $\bar{3}$ m]	13.73 \pm 0.26	-3.13 \pm 0.23				
LiH [Fm $\bar{3}$ m]	22.59 \pm 0.01	-1.64 \pm 0.07			5.06 \pm 0.03	-1.55 \pm 0.00
Mg ₂ Si [Fm $\bar{3}$ m]	2.31 \pm 0.05	-0.58 \pm 0.03			0.34 \pm 0.04	-0.14 \pm 0.01
MgO [Fm $\bar{3}$ m]	7.86 \pm 0.02	-2.02 \pm 0.03				
MnO [Fm $\bar{3}$ m]	<i>0.82 \pm0.13</i>	<i>0.34 \pm0.02</i>				
MnS [Fm $\bar{3}$ m]	2.77 \pm 0.08	0.28 \pm 0.03	3.54 \pm 1.08	0.13 \pm 0.09	0.11 \pm 0.14	0.13 \pm 0.09
MoS ₂ [R3m]	3.99 \pm 0.02	-0.68 \pm 0.01	2.51 \pm 0.06	-0.48 \pm 0.03	2.78 \pm 0.00	-0.48 \pm 0.03
MoSe ₂ [R3m]	3.88 \pm 0.03	-0.82 \pm 0.00	1.85 \pm 0.08	-0.51 \pm 0.01	2.86 \pm 0.03	-0.51 \pm 0.01
NaCl [Fm $\bar{3}$ m]	9.44 \pm 0.02	-2.33 \pm 0.01				
NaCl [Pm $\bar{3}$ m]	10.31 \pm 0.06	-2.19 \pm 0.09			7.91 \pm 0.00	-2.19 \pm 0.03
NaI [Fm $\bar{3}$ m]	5.85 \pm 0.01	-1.77 \pm 0.01				
NiO [Fm $\bar{3}$ m]	<i>3.09 \pm0.00</i>	<i>0.19 \pm0.01</i>	2.28 \pm 0.02	0.14 \pm 0.01	2.29 \pm 0.03	-0.25 \pm 0.02
PbF ₂ [Fm $\bar{3}$ m]	11.23 \pm 0.10	-1.85 \pm 0.00	6.71 \pm 0.02	-1.64 \pm 0.01	6.85 \pm 0.02	-1.67 \pm 0.01
PbSe [Fm $\bar{3}$ m]	7.68 \pm 0.01	-0.84 \pm 0.01			1.52 \pm 0.00	-0.15 \pm 0.00
Si [Fd $\bar{3}$ m]	3.28 \pm 0.00	-0.72 \pm 0.01	1.14 \pm 0.03	-0.72 \pm 0.01		
Si [I4 ₁ /amd]	3.20 \pm 0.14	-0.32 \pm 0.03			2.36 \pm 0.01	-0.35 \pm 0.03
Si [P6 ₃ /mmc]	18.80 \pm 0.03	-0.36 \pm 0.02	2.06 \pm 0.05	-0.33 \pm 0.02	4.92 \pm 0.04	-0.33 \pm 0.02
SiC [F $\bar{4}$ 3m]	7.52 \pm 0.07	-0.99 \pm 0.02	2.47 \pm 0.01	-0.99 \pm 0.02	5.81 \pm 0.02	-1.17 \pm 0.03
SrO [Fm $\bar{3}$ m]	5.81 \pm 0.10	-1.48 \pm 0.04	5.37 \pm 0.01	-1.48 \pm 0.04	5.47 \pm 0.01	-1.52 \pm 0.03
TlBr [Pm $\bar{3}$ m]	9.31 \pm 0.26	-1.43 \pm 0.00			3.33 \pm 0.00	-1.33 \pm 0.00
TlCl [Pm $\bar{3}$ m]	11.45 \pm 0.06	-1.47 \pm 0.02			3.67 \pm 0.03	-1.37 \pm 0.02
WS ₂ [R3m]	4.03 \pm 0.00	-0.63 \pm 0.01	2.76 \pm 0.01	-0.59 \pm 0.01	2.85 \pm 0.01	-0.59 \pm 0.01
ZnS [F $\bar{4}$ 3m]	3.54 \pm 0.01	-1.16 \pm 0.03				
ZnSe [F $\bar{4}$ 3m]	2.50 \pm 0.02	-1.05 \pm 0.00				
ZnTe [F $\bar{4}$ 3m]	2.35 \pm 0.00	-0.89 \pm 0.00				

Table 4.4b: Continuation of calculated fundamental G0W0 band gaps (**Gap**) (direct gap at **Gamma**, smallest **direct** gap at any k-point, smallest **indirect** gap) and corresponding quasi-particle shifts (Δ **IP**) of the valence band maximum. The latter two gap types are only given if different from the fundamental gap at Gamma. Gaps reported in italic have show wrong band occupations or metallic solutions, see section (4.1.2).

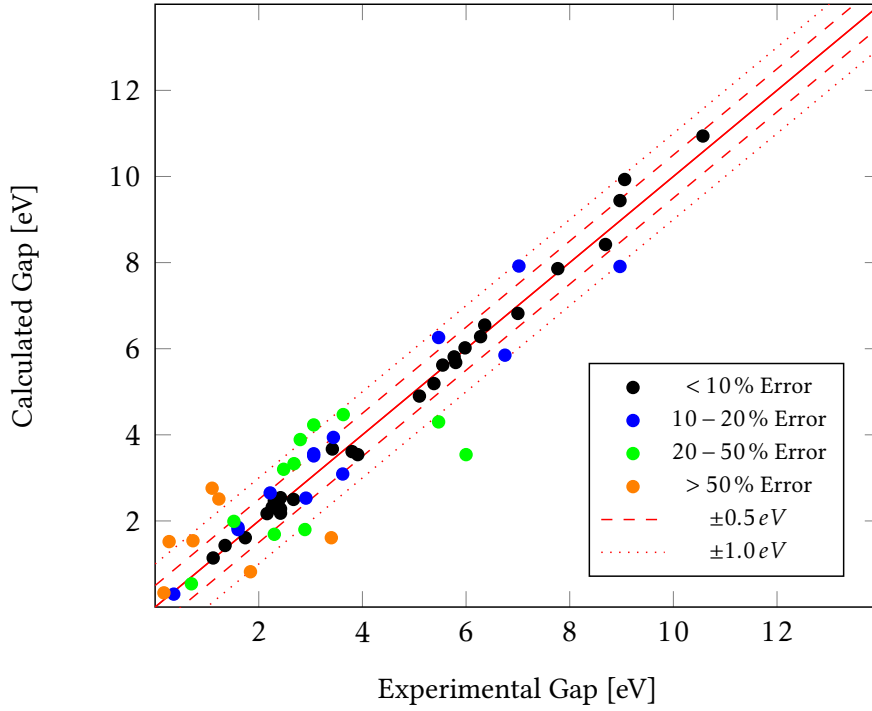


Figure 4.4: Comparison of calculated band gaps to experimental values. For gaps without any transition (direct/indirect) specified the best fitting gap was chosen from the calculated ones. Experimental data from references [22, 23, 24, 25, 26, 27, 28].

4.2.2 Comparison to Experimental Gaps

In this section the calculated band gaps should be compared to gaps determined in experiments. Even if the automatic procedure has yet to reach similar accuracy as performing GW calculations manually, a first look at how the results compare to experiments helps to understand the quality of produced data.

For each material, the corresponding gap to the transition measured in experiments (direct/indirect, (d)/(i) as indicated in tables (4.5a) and (4.5b)) was chosen for comparison and for the graphical representation in figure (4.4). If there was no transition specified in the literature (denoted as (-) in the tables) the best fitting gap was chosen. This does not always mean that this has to be the right choice.

Many gaps compare quite well to values found in experiments, with a majority of gaps being at least 20 % accurate with respect to the experimental gap. For 32 materials the band gap error was even below 10 %. These materials can be found in table (4.5a), all other gaps with errors <10 % are listed in table (4.5b).

As mentioned in reference [1], all gaps in the simulations are calculated at zero temperature. This inevitably leads to errors and deviations between experiments and calculations as performed here. Most experimental gaps used here are measured at room temperature (300 K),

4 Results

Material	Exp. Gap [eV]	Cal. Gap [eV]	Error (abs) [eV]	Error [%]
AgCl [Fm $\bar{3}$ m]	5.1 (-)	4.90	0.20	-3.92
AlAs [F $\bar{4}$ 3m]	2.16 (i)	2.17	0.01	+0.46
AlN [P $\bar{6}$ ₃ mc]	6.28 (d)	6.28	0.00	+0.00
BaF ₂ [Fm $\bar{3}$ m]	9.06 (-)	9.93	0.89	+9.82
BaO [Fm $\bar{3}$ m]	3.8 (-)	3.61	0.19	-5.00
BeO [P $\bar{6}$ ₃ mc]	10.57 (d)	10.94	0.37	+3.50
BN [F $\bar{4}$ 3m]	6.36 (i)	6.55	0.19	+3.00
CaI ₂ [P $\bar{3}$ m1]	5.98 (-)	6.02	0.04	+0.67
CaO [Fm $\bar{3}$ m]	7 (-)	6.82	0.18	-2.57
CaS [Fm $\bar{3}$ m]	5.38 (-)	5.19	0.19	-3.53
CdCl ₂ [R $\bar{3}$ m]	5.8 (-)	5.68	0.12	-2.07
CdS [F $\bar{4}$ 3m]	2.42 (d)	2.18	0.24	-9.92
CdS [Fm $\bar{3}$ m]	2.42 (d)	2.54	0.12	+4.96
CdS [P $\bar{6}$ ₃ mc]	2.42 (d)	2.28	0.14	-5.79
CdSe [F $\bar{4}$ 3m]	1.74 (d)	1.61	0.13	-7.47
CsAu [Pm $\bar{3}$ m]	2.3 (-)	2.36	0.06	+2.53
GaP [F $\bar{4}$ 3m]	2.26 (i)	2.32	0.06	+2.65
HfO ₂ [Fm $\bar{3}$ m]	5.55 (-)	5.62	0.07	+1.26
InP [F $\bar{4}$ 3m]	1.35 (d)	1.43	0.08	+5.93
KCl [Fm $\bar{3}$ m]	8.69 (-)	8.42	0.27	-3.11
LiF [Fm $\bar{3}$ m]	14.20 (-)	13.73	0.47	-3.31
Mg ₂ Si [Fm $\bar{3}$ m]	2.27 (d)	2.31	0.04	-1.76
MgO [Fm $\bar{3}$ m]	7.77 (d)	7.86	0.09	+1.16
NaCl [Fm $\bar{3}$ m]	8.97 (d)	9.44	0.47	+5.24
Si [Fd $\bar{3}$ m]	1.12 (i)	1.14	0.02	+1.79
SiC [F $\bar{4}$ 3m]	2.3 (i)	2.47	0.17	+7.39
SrO [Fm $\bar{3}$ m]	5.77 (-)	5.81	0.04	+0.69
TlCl [Pm $\bar{3}$ m]	3.42 (-)	3.67	0.25	+7.31
ZnS [F $\bar{4}$ 3m]	3.91 (-)	3.54	0.37	-9.46
ZnSe [F $\bar{4}$ 3m]	2.67 (d)	2.50	0.17	-6.37
ZnTe [F $\bar{4}$ 3m]	2.38 (d)	2.35	0.03	-1.26

Table 4.5a: Comparison of calculated fundamental band gaps to experimental values, with an error < 10 % with respect to the experimental band gap. The indicators ((d) direct, (i) indirect, (-) not specified) denote the experimentally measured transition. Experimental data from references [22, 23, 24, 25, 26, 27, 28].

for some values at 0 K were available and preferably taken.

Besides all experimental uncertainties, which might not be clearly stated in every source such as purity or configuration of the material tested, there are also computational uncertainties which could make the results appear worse. As mentioned before, the band gaps are only calculated between select few points in reciprocal space, depending on the density of k-points

Material	Exp. Gap [eV]	Cal. Gap [eV]	Error (abs) [eV]	Error [%]
AgI [F43m]	2.91 (d)	2.53	0.38	-13.06
AlSb [F43m]	2.22 (-)	2.65	0.43	+19.37
BaTiO ₃ [P4mm]	3.06 (-)	3.51	0.45	+14.71
BaTiO ₃ [Pm3m]	3.06 (-)	3.56	0.50	+16.34
C [Fd3m]	5.47 (i)	6.26	0.79	+14.44
C [R3m]	7.02 (-)	7.92	0.90	+12.82
CdTe [F43m]	1.6 (-)	1.80	0.20	+12.50
GaN [F43m]	3.44 (d)	3.94	0.50	+14.53
InAs [F43m]	0.36 (d)	0.30	0.06	-17.58
MoSe ₂ [R3m]	1.6 (-)	1.85	0.25	+15.63
NaCl [Pm3m]	8.97 (d)	7.91	1.06	-11.82
NaI [Fm3m]	6.75 (d)	5.85	0.90	-13.33
NiO [Fm3m]	3.62 (d)	3.09	0.53	-14.64
AgF [Fm3m]	2.8 (-)	3.89	1.09	+38.93
AlP [F43m]	3.63 (d)	4.47	0.84	+23.14
BaTiO ₃ [R3m]	3.06 (-)	4.23	1.17	+38.24
BeTe [F43m]	2.89 (i)	1.80	1.09	-37.72
C [P6 ₃ /mmc]	5.47 (i)	4.30	1.17	-21.39
CdO [Fm3m]	2.30 (-)	1.69	0.61	-26.47
GaAs [F43m]	1.52 (-)	1.99	0.47	+30.92
HgO [Imm2]	2.48 (-)	3.20	0.72	+29.03
InN [F43m]	0.7 (d)	0.54	0.16	-22.86
MnS [Fm3m]	6 (i)	3.54	2.46	-41.00
TlBr [Pm3m]	2.68 (-)	3.33	0.65	+24.25
CuCl [F43m]	3.4 (d)	1.61	1.79	-52.65
GaSb [F43m]	0.73 (d)	1.54	0.81	+110.96
InSb [F43m]	0.17 (d)	0.33	0.16	-95.65
MnO [Fm3m]	1.84 (-)	0.82	1.02	-55.43
MoS ₂ [R3m]	1.23 (i)	2.51	1.28	+105.07
PbSe [Fm3m]	0.27 (d)	1.52	1.25	+462.96
WS ₂ [R3m]	1.1 (-)	2.76	1.66	+150.91

Correlation between calculated and experimental gaps: 0.97

Linear fit: $y = 0.96x + 0.15$

Table 4.5b: Continuation of the comparison of calculated fundamental band gaps to experimental values. Gaps here have from bottom to top an error of 10-20 %, 20-50 %, or >50 % with respect to the experiment. The indicators ((d) direct, (i) indirect, (-) not specified) denote the experimentally measured transition. The correlation coefficient is calculated between all calculated and experimental gaps, this holds also true for the linear fit. Experimental data from references [22, 23, 24, 25, 26, 27, 28].

in the automatically generated mesh. For all low k-point density calculations here, the energy eigenvalues are only accessible at four points in the *OUTCAR* file. Since the band structure at GW level is not available, it is quite possible to miss the right band gap if for example the valence band maximum or the conduction band minimum lies somewhere between those few points.

This is why it is easier to compare such results to other computational results, like done in section (4.1.1). Here, the point in reciprocal space where the fundamental gap is calculated at is well defined, making it clear which transition is investigated. Some of the results listed in table 4.5b with a very large error could be such cases.

Generally speaking, the results are quite satisfactory. Taking into account how fast, easy, and computationally cheap the final procedure is, reaching for approximately half of the test set a gap accuracy with an error below 10 % with respect to the experimental value found in literature undermines the feasibility of such automated procedures. A sample correlation of 0.97 underlines that the calculated results stand in good agreement with experimental values.

Discussing these results one has to keep in mind how small changes in the procedure (see section (4.1.4)) can already introduce errors of 100 meV or more. Additionally, results were achieved without the use of norm conserving pseudo-potentials and a less accurate procedure to reduce the memory requirement. With enough computational resources available to employ better potentials and allow the usage of the original variant, this procedure would probably yield already much better fundamental band gaps for the majority of materials. Increasing the number of data points for the correction would also greatly narrow down the error range, but is leading again to greater computational costs.

4.3 Possibility of a Norm Correction

One of the main sources of errors in the band gap when calculated using basis set extrapolation has its origin in the norm violation of the pseudo-potentials. This can be mostly mitigated by constructing norm-conserving potentials for the simulations. However, the usage of such requires severely larger basis sets and thus even more memory.

The method used for correcting the gaps to a high density k-point grid is possible in a very simple way. This is because the slope of the linear fit resulting from the extrapolation is very insensitive to the number of k-points.[17] Similarly, it was tested if the additional norm violation of not using norm-conserving pseudo-potentials could be corrected in a simple way too.

In section (4.1.1) it can be seen that the quasi-particle shift seems to be much less sensitive to an incorrectness of the norm. Since the band topology does not change significantly by going from DFT to GW, the idea was to correct the norm violation on DFT level.

For this, $\Delta E_{DFT}(n_k, E_2)$ was introduced as the difference between DFT level band gaps at Gamma, using one time norm-conserving potentials and the other time not. This correction was also calculated using the coarse k-point n_k grid and standard energy cut-off E_1 . The whole basis set correction procedure is then given by:

$$E_\infty(N_k, N_\infty) \approx E_\infty(n_k, N_\infty) - E(n_k, N_2) + E(N_k, N_2) + \Delta E_{DFT}(n_k, E_2) \quad (4.1)$$

Material	ref. Gap@ Γ [eV]	calc. Gap@ Γ [eV]	corr. Gap@ Γ [eV]	Error [%]
AlAs	2.99	3.36	3.09	3.38 (-8.98)
AlP	4.23	4.51	4.24	0.30 (-6.38)
AlSb	2.4	2.73	2.52	5.02 (-8.64)
GaAs	1.23	2.00	1.76	42.88 (-20.05)
GaP	2.62	3.31	3.07	17.15 (-9.11)
GaSb	0.68	1.46	1.27	86.37 (-28.74)
InP	1.23	1.41	1.15	6.37 (-8.20)
MgO	7.55	8.12	8.00	5.98 (-1.57)
Si	3.25	3.28	3.28	1.05 (-0.02)
SiC	7.35	7.57	7.58	3.07 (+0.01)

Table 4.7: Comparison of band gaps proposed in reference [2], calculated gaps in this project, and the same gaps after the norm correction. The error in percent with respect to the reference gaps is also given, as well as the error reduction due to the norm correction in brackets.

Quite surprisingly, this correction indeed positively influences band gap value. For all materials the corrected band gap had a better or at least the same agreement to the gaps in reference [2], compared to the uncorrected one. (see table (4.7))

However, for many materials the correction does not compensate the norm violation completely, hinting that such an easy correction is not possible. This can be explained partially by the fact that severe errors, like the band inversion in InAs, of course do not get corrected by this procedure, at least for some systems.

Furthermore, there was only a small number of gaps investigated and more tests have to be made before any general conclusions are possible. Probably a correction in a similar fashion at GW level could lead to even better results and could quite possibly be supported by some machine learning techniques.

4.4 Band Structures at DFT Level

Additionally to the band gaps at GW level, the band structure at DFT level was determined for all systems. This step is also included in the automatic procedure and done during the preliminary DFT calculations before the G0W0 simulations. The plots themselves were created using the Python Materials Genomics package (*pymatgen*), which is a powerful open source tool for the analysis of materials.[20] The k-point paths through the reciprocal unit cell were taken from reference [10].

Two examples band structures are shown here (Cu_2O , FeAl), all the others that were computed can be found in the appendix. All segments between the high symmetry points in reciprocal space are calculated with 20 intersections to guarantee a smooth and accurate curves.

The band structure calculations were just a small effort in comparison to the GW parts of the project. Their execution was generally unproblematic, fast and succeeded for all systems.

4 Results

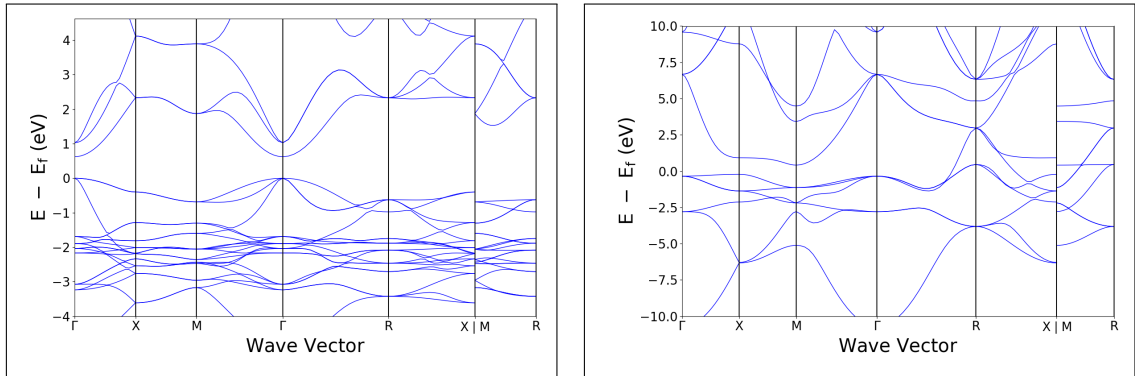


Figure 4.5: Examples for the band structure plots done for all materials. Left: Cu_2O in in cuprite configuration ($\text{Pn}\bar{3}\text{m}$) with an fundamental band gap between valence and conduction band. Right: FeAl in CsCl structure showing metallic properties without a band gap in DFT.

Even though the band structures are only calculated only on DFT level, this information is relevant even for the GW level results. During the quasi-particle correction, achieved by the G_0W_0 calculation, the band structure topology does not change much. Since calculating band structures at GW level is a very difficult task, relevant information such as the position of band extrema or gaps can be retrieved from the DFT level structure.

5 Discussion

In the course of this project the feasibility of high-throughput GW calculations was tested and the results compared to both, other GW and experimental data. In the following the presented results, as well as observations during the work will be discussed.

5.1 Challenges during the Project

Dealing with GW calculations is never an easy task. Even if the basis set extrapolations brings many simplifications and reduces the work necessary to calculate converged gaps, many challenges remain. Some of the most predominant issues which had to be resolved are the extensive memory requirements and code instability.

5.1.1 Encountered Problems and Stability

In the final iteration of the automatic GW procedure, none of the calculations failed during the investigation of 69 materials. To be precise, the procedure did not stop due to an error until all materials were investigated. Although, as later discussed for about half of the materials the calculated band gaps show unsatisfying accuracy, but this is more of a conceptual problem.

The calculations for materials showing wrong band occupation can also be seen as not successful, since here the fundamental gap could not be calculated automatically and further investigation was necessary.

On the technical side, the most challenging part was providing enough memory for the high-bands G0W0 calculations, causing severe difficulties mainly at the early stages of the project. Once this issue was resolved the procedure proofed to be very stable and easy edit, without encountering any further problematic errors.

5.1.2 Correct K-point Correction

Another intricate detail causing some problems and wrong band gaps was during the analysis of band data in the *OUTCAR* file. Here the issue was to ensure that all band gaps and the k-point correction is calculated at the same point in reciprocal space. Since the k-point mesh is automatically generated by the code, number and position of the k-points can vary between different steps. To explain that further, the calculation used for the k-point correction uses a more dens mesh, leading to more points in reciprocal space where the band energies are calculated. In practice this means that, for example the smallest band gap might be found at k-point #3 for low-density grids, but for high-density k-point grids this point is denoted as k-point #12. Here the actual positions in the reciprocal unit cell have to be used to avoid errors and to correct the k-point density with the right fundamental gap.

5.1.3 Memory Requirements

The basis set extrapolation method requires in its simplest form only four G0W0 calculations, which is a significantly lower number of simulations that have to be performed. However, this comes at the cost of an significantly increased memory requirement due to the extensive band set that have to be included.

The main component influenced by these bands is the response function. Its size is dependent on the number of basis functions in the auxiliary basis set, which in turn scales with the energy cut-off which has to be significantly increased during the extrapolation. The response function also depends on the number of frequency points used for the simulations, defined by the flag *NOMEGA*. As proposed in reference [17], this parameter has to be set to a high enough value to retrieve accurate results and a number around 100 (96 because of parallelization) was deemed as sufficient.

This leads to response functions with sizes of several GB, quickly exceeding the accessible memory on the cluster. Even worse, when the spectral method is used for better performance (*LSPECTRAL = .TRUE.*) the response function has to be kept in memory at all times during the calculation.

To finally fulfill these excessive memory requirements it was necessary not to use the full number of cores on each computing node (12 out of 16), resulting in an increased amount of memory per core. Taking these measures it was possible to guarantee that all calculations have enough memory to run stable.

5.1.4 Code Instability due to Large Band Numbers

Although the GW implementation in VASP is of course stable, there are several details which have to be accounted for to successfully perform these simulations.

This begins during the DFT step for producing the starting wave function used as input for the following GW step. It is not recommended to use the maximum number of empty bands already during the electronic self-consistency loop, although required later in the G0W0 calculations. This is because the algorithms used for calculating the wave function and energy may struggle with the large number of empty bands, leading to instability and eventual failure of the procedure.

To avoid this, the ground state is found with a normal set of bands. Afterwards the empty bands are added during a not self-consistent one-step DFT calculation, where also the wave function derivative is calculated.

5.2 Feasibility of Automatic GW Calculations

The results of this project lead to several different conclusions regarding the feasibility of such an automatic procedure to calculate GW level band gaps.

From a conceptual perspective, this project has shown that the basis set extrapolation indeed can be automated through very simple means. The implementation of a high-throughput procedure is straight forward and also very stable and reliable, once the parameters are adjusted accordingly.

The quality of the results, however, draws a mixed picture. For approximately 50 % of the investigated materials the band gap was very accurately predicted with much lower effort than usual. Looking at the quite high error of the other half, this is not a good success rate to rely on.

Although for many materials the gap is very accurate, it would probably be necessary to employ norm-conserving pseudo-potentials to guarantee a good results for the majority of compounds. However, this is still computationally difficult to fulfill right now. Using such potentials per default would lead to a severe increase in memory requirement and CPU time.

5.2.1 Machine Learning Approaches

It may be possible to preemptively make a guess of the expected norm violation when employing standard potentials and then automatically make a choice if norm-conserving potentials would be required. Such an approach could be realized by machine learning algorithms, specifically trained to make such an estimation based on atom types and positions in the material. Including such an algorithm in the automatic procedure could prevent severely wrong gaps.

This approach can be even stretched out further, once enough good quality GW level band gap data is collected. Using such a database could lead to data driven algorithms able predict the QP shift or GW level gap directly from DFT level data. Such a model would need a vast amount of data and has to be set up with great care, but directly correcting the band gap produced by DFT calculations without performing any GW simulations would be several orders of magnitude faster. In reference [7] some approaches to predict the experimental gap from DFT data, which is quite a challenge to reach an accuracy worth the effort.

Trying to predict GW level data might be much simpler, since there are theoretically no hidden uncertainties in contrast to experiments. This is because of the fact that the code used for GW calculations should always lead to the same result when the same input is given, which is also known exactly. Fitting a machine learning model to this different, but also purely mathematical problem could be achieved more easy.

5.2.2 Bringing GW Calculations to AiiDA

This project has shown that the basis set extrapolation method can be automated in a simple, straight-forward way. By eliminating the challenging task of systematically evaluating convergence, sophisticated procedures as implemented in reference [1] are not required here. This opens the possibility to transfer the procedure with rather low effort to already existing software packages, specifically designed for high-throughput calculations and the analysis of large data sets.

A very future-oriented application providing such an environment is AiiDA (Automated Interactive Infrastructure and Database for Computational Science), developed and introduced by Pizzi *et al.* [8]. With a strong focus on automation and producing results in an uniform and easy-to-share format, it is clear to see that the next step would be transferring the automated GW scheme to this software suite.

Already existing plug-ins support rather simple VASP applications, such as electronic or structural relaxation procedures. Implementing GW calculations which provide converged

5 Discussion

results was not possible yet, due to the heuristic nature of the conventional scheme. However, the here introduced method based on basis set extrapolation would not be much different to implement as any other single-calculation plug-in.

Large scale GW calculations would open up new possibilities in computational materials science. Salvaging codes specialized in high-throughput calculations will be essential to achieve this task and the basis set extrapolation method might be a crucial part of such procedures.

Acknowledgments

I wish to express my deepest gratitude to my Supervisor, Univ.-Prof. Dr. techn. Cesare Franchini, who always found the time to help whenever I encountered any problems or needed guidance throughout my project. Thanks to him I was not only able to make new, exciting experiences, but also meet and work together with many different and interesting people. Through his excellent support and scientific expertise I was able to grow as a scientist myself and I have learned a lot during my work at his group here in Vienna.

I also want to thank all of my friends and family for supporting me during my studies, helping me through tough and stressful times during the last few years. Here I want to especially mention Astrid Ellinger and Julia Kodnar, who always supported me wholeheartedly and who I could rely on in times of need.

Furthermore, I want to thank all colleagues from my work group. During my time at the office they were always able to offer helpful insights and advice, and became also good friends of mine.

The computational results presented have been achieved [in part] using the Vienna Scientific Cluster (VSC).

Bibliography

- [1] M. J. van Setten et al. “Automation methodologies and large-scale validation for GW: Towards high-throughput GW calculations”. In: *Phys. Rev. B* 96, 155207 (2017).
- [2] Jiri Klimeš, Merzuk Kaltak, and Georg Kresse. “Predictive GW calculations using plane waves and pseudopotentials”. In: *Phys. Rev. B* 90, 075125 (2014).
- [3] Lars Hedin. “New Method for Calculating the One-Particle Green’s Function with Application to the Electron-Gas Problem”. In: *Phys. Rev.* 139 (3A Aug. 1965), A796–A823. DOI: 10.1103/PhysRev.139.A796. URL: <https://link.aps.org/doi/10.1103/PhysRev.139.A796>.
- [4] G. Strinati, H. J. Mattausch, and W. Hanke. “Dynamical Correlation Effects on the Quasiparticle Bloch States of a Covalent Crystal”. In: *Phys. Rev. Lett.* 45 (4 July 1980), pp. 290–294. DOI: 10.1103/PhysRevLett.45.290. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.45.290>.
- [5] Mark S. Hybertsen and Steven G. Louie. “Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies”. In: *Phys. Rev. B* 34 (8 Oct. 1986), pp. 5390–5413. DOI: 10.1103/PhysRevB.34.5390. URL: <https://link.aps.org/doi/10.1103/PhysRevB.34.5390>.
- [6] M. van Schilfgaarde, Takao Kotani, and S. Faleev. “Quasiparticle Self-Consistent GW Theory”. In: *Phys. Rev. Lett.* 96 (22 June 2006), p. 226402. DOI: 10.1103/PhysRevLett.96.226402. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.96.226402>.
- [7] J. Schmidt et al. “Recent advances and applications of machine learning in solid-state materials science”. In: *npj Computational Materials volume 5, Article number: 83* (2019).
- [8] Giovanni Pizzi et al. “AiiDA: Automated Interactive Infrastructure and Database for Computational Science”. In: *arXiv:1504.01163v2 [physics.comp-ph]* (2015).
- [9] Rudolf Gross and Achim Marx. *Festkörperphysik*. Oldenbourg Wissenschaftsverlag, 2012. ISBN: 9783486712940.
- [10] Wahyu Setyawan and Stefano Curtarolo. “High-throughput electronic band structure calculations: Challenges and tools”. In: *Computational Materials Science* 49, 299–312 (2010).
- [11] P. E. Blöchl. “Projector augmented-wave method”. In: *Phys. Rev. B* 50, 17953 (1994).
- [12] G. Kresse and J. Joubert. “From ultrasoft pseudopotentials to the projector augmented wave method”. In: *Phys. Rev. B* 59, 1758 (1999).

Bibliography

- [13] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized Gradient Approximation Made Simple”. In: *Phys. Rev. Lett.* 77 (18 Oct. 1996), pp. 3865–3868. DOI: 10.1103/PhysRevLett.77.3865. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.77.3865>.
- [14] F. Aryasetiawan and O. Gunnarsson. “The GW method”. In: *Rep. Prog. Phys.* 61, 237-312 (1998).
- [15] G. Kresse and J. Hafner. In: *Phys. Rev. B* 47 558 (1993).
- [16] G. Kresse and J. Hafner. In: *Phys. Rev. B* 54 11169 (1996).
- [17] Zeynep Ergönenc et al. “Converged GW quasiparticle energies for transition metal oxide perovskites”. In: *Phys. Rev. Mat.* 2, 024601 (2018).
- [18] Klaus Capelle. “A bird’s-eye view of density-functional theory”. In: *Brazilian Journal of Physics* arXiv:cond-mat/0211443v5 [cond-mat.mtrl-sci] (2006).
- [19] B.-C. Shih et al. “Quasiparticle band gap of ZnO: high accuracy from the conventional G0W0 approach”. In: *Phys. Rev. Lett.* 105, 146401 (2010).
- [20] Shyue Ping Ong et al. “Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis”. In: *Computational Materials Science* 68 (2013), pp. 314–319. ISSN: 0927-0256. DOI: <https://doi.org/10.1016/j.commatsci.2012.10.028>. URL: <http://www.sciencedirect.com/science/article/pii/S0927025612006295>.
- [21] *VSC3 Technical Data, October 19*. URL: <https://typo3.vsc.ac.at/systems/vsc-3/>.
- [22] Safa O. Kasap and Peter Capper. “Springer handbook of electronic and photonic materials”. In: *Springer*. pp. 54, 327. ISBN 978-0-387-26059-4 (2006).
- [23] D. A. Evans et al. “Determination of the optical band-gap energy of cubic and hexagonal boron nitride using luminescence excitation spectroscopy”. In: *Journal of Physics: Condensed Matter*. 20 (7): 075233 (2008).
- [24] Claus F. Klingshirn. “Semiconductor optics”. In: *Springer*. p. 127. ISBN 978-3-540-61687-0 (1997).
- [25] K. Kobayashi and J. Yamauchi. “Electronic structure and scanning- tunneling- microscopy image of molybdenum dichalcogenide”. In: *Physical Review B*. 51 (23): 17085–17095 (1995).
- [26] S. Chakrabarty and K. Chatterjee. “Synthesis and Characterization of Nano- Dimensional Nickelous Oxide (NiO) Semiconductor”. In: *Journal of Physical Sciences*, Vol. 13, 2009, 245-250 (2009).
- [27] W. H. Strehlow and E. L. Cook. “Compilation of Energy Band Gaps in Elemental and Binary Compound Semiconductors and Insulators”. In: *J. Phys. Chem. Ref. Data*, 2, 163-200 (1973).
- [28] Ya Zhuo, Aria Mansouri Tehrani, and Jakoah Brgoch. “Predicting the Band Gaps of Inorganic Solids by Machine Learning”. In: *J. Phys. Chem. Lett.*, 9 1668-1673 (2018).
- [29] Robert Nau. *Statistical forecasting: notes on regression and time series analysis*. URL: <https://people.duke.edu/~rnau/mathreg.htm>.

Appendix

Abstract

This project aims to efficiently calculate *ab initio* fundamental band gaps beyond density functional theory (DFT) for large material sets. The GW approach (G is the one particle Green function, W the screened Coulomb potential) is used as post DFT method to correct made approximations, specifically electronic correlation contributions. Performing even the simplest type of GW calculations, so called single-shot GW calculations (referred to as G0W0), usually requires a significant amount of computational resources and user interaction. This greatly inhibits an efficient and quick analysis of larger material sets. In practice, the standard procedure (or *conventional method*) for calculating only one converged fundamental band gap typically involves several G0W0 calculations for the purpose of converging different computational parameters separately. These include not only the energy cut-off of the basis set and the corresponding number of energy bands, but also said two parameters for the auxiliary basis set, as well as the density of the k-point grid. All of these quantities are increased gradually, until the fundamental gap does not change anymore and convergence is reached. Because of such involved convergence requirements and large amounts of user interference it is highly impractical to automate this approach. Only through the implementation of sophisticated methods (as done in reference [1]) fundamental band gaps can be computed automatically through this conventional method.

These complications can be circumvent by applying the so called basis set extrapolation, developed by Klimeš *et al.*[2] In its proposal it was shown that the band gap error arising in GW calculations due to truncation of the plane wave expansion behaves like $1/N_B$, where N_B is the number of energy bands included in the simulation.[2] Thus, the gap error can be corrected by extrapolating to the limit of an infinite set of bands. Calculating converged GW level band gaps with this method requires only a total number of four G0W0 calculations. Additionally, by removing the necessity of converging multiple parameters manually, an automatic scheme can be implemented rather easily.

In my work I have developed such an automated scheme for calculating well converged fundamental band gaps on GW level, using basis set extrapolation. With this automatic procedure three different types of fundamental gaps (direct gap at Gamma, smallest indirect gap, smallest direct gap) were calculated fully automatic for a material set of 69 materials, consisting mostly of mono- and diatomic insulators and semi-conductors. Additionally to the fundamental band gaps at GW level, band structures at DFT level were also calculated within the procedure.

The ultimate goal of this automatic approach is to produce reliable GW data for large sets of materials with less effort and making the results accessible to public via an online database.

Zusammenfassung

Das Ziel dieses Projektes ist die effiziente Berechnung von fundamentalen Energie-Bandlücken für große Datensätze von Materialien unter der Verwendung von *ab initio* Methoden. Zur Berechnung dieser Bandlücken kommt die GW-Methode (G steht für die Einteilchen-Greenfunktion, W für das abgeschirmte Coulombpotenzial). Sie ist eine post-DFT Methode und wird dazu verwendet notwendige Annäherungen in der Dichtefunktional-Theorie (DFT) zu korrigieren, konkret werden angenäherte Elektronen-Korrelationseffekte exakt berechnet.

Bereits die einfachste Variante der GW-Methode, die Verwendung von sogenannten *single-shot GW* (G0W0) Berechnungen, benötigt ein hohes Maß an Zeitaufwand und Computer-Ressourcen. Folglich ist die Analyse von großen Datensätzen mit dieser Methode ein sehr aufwändiger Prozess. Zusätzlich benötigt die Berechnung von nur einer konvergierten Energie-Bandlücke nicht lediglich eine G0W0 Berechnung, sondern mehrere davon sind notwendig, um ein korrektes Ergebnis zu erhalten. Bei der konventionellen Methode (*conventional method*), wird eine Vielzahl an G0W0 Berechnungen dafür verwendet, den optimalen Wert für mehrere Eingabe-Parameter manuell zu finden. Diese Parameter sind unter anderem der Energie Cutoff der verwendeten Basisfunktionen und die dazugehörige Anzahl an Energiebändern, sowohl für das Standard-Basisset, als auch für das Hilfs-Basisset. Die Dichte der k-Punkte im reziproken Raum muss ebenfalls optimiert werden. All diese Werte müssen manuell schrittweise erhöht werden, bis sich die Energie-Bandlücke nicht mehr ändert. Diese Prozedur mit ihren komplizierten Konvergenz-Kriterien kann nur schwer automatisiert werden und sehr aufwändige Methoden müssen verwendet werden, um fundamentale Energie-Bandlücken automatisch zu berechnen.[1]

Eine neue Methode, die diese Komplikationen umgeht, ist die Basisset-Extrapolation, entwickelt von Klimeš *et al.*[2] In ihrer Publikation wurde gezeigt, dass der Fehler, kommend von dem Abbruch der Basisset-Entwicklung, ein Verhalten von $1/N_B$ zeigt, wobei N_B die Anzahl der in der Berechnung inkludierten Energiebänder ist.[2] Dies hat zur Folge, dass der Fehler durch eine Extrapolation zur Grenze von unendlich vielen inkludierten Bändern korrigiert werden kann. Diese Methode verwendet lediglich vier G0W0 Berechnungen und keine Parameter müssen manuell optimiert werden, wodurch eine Automatisierung wesentlich einfacher umzusetzen ist.

In meiner Arbeit habe ich unter Verwendung der Basisset-Extrapolation eine automatische Prozedur zur Berechnung von konvergierten, fundamentalen Energie-Bandlücken entwickelt. Mit dieser Prozedur wurden drei verschiedene Arten von Bandlücken (direkte Bandlücke am Gammapunkt, die kleinste indirekte Bandlücke, und die kleinste direkte Bandlücke) vollautomatisch für einen Datensatz von 69 Materialien, bestehend aus hauptsächlich ein- und zweiatomigen Halbleitern und Isolatoren, berechnet. Zusätzlich zur den angeführten fundamentalen Bandlücken auf GW-Level wurden auch die Bandstrukturen der Materialien auf DFT-Level im Zuge der automatischen Prozedur berechnet.

Das finale Ziel dieser automatischen Prozedur ist qualitativ hochwertige Daten auf GW-Level für große Datensätze von Materialien einfacher berechnen zu können und online in einer öffentlichen Datenbank zugänglich zu machen.

Regression Error

Due to the fact that calculated data points do not reproduce an exact linear trend, an error in the estimated gap is introduced when using a linear regression. This error is often referred to as *standard error of the estimate* or *standard error of the constant* of the linear regression.

In this section the formalism, how the error was calculated is briefly presented.

As a starting point, the sample standard deviation is calculated for the errors of the data points. Since the band gaps do not have an error themselves and are seen as exact, one has to calculate this quantity as the standard deviation of the dependent variables (band gaps) times the square root of one minus the correlation squared (equivalent to the r^2 value, which is minimized during the regression).

$$\sigma_{err}^s = \sigma_{gaps}^s \times \sqrt{1 - r^2} \quad (5.1)$$

The r^2 value is referred to as the *coefficient of determination* and an indicator for how good the trend is reproduced by the regression.

Next, the standard error of the regression model s can be calculated using this standard deviation.

$$s = \sigma_{err}^s \times \sqrt{\frac{n-1}{n-2}} \quad (5.2)$$

Where the last term simplifies to $\sqrt{2}$, since only $n = 3$ data points are used for the regression here. Using these two quantities, the standard error of the estimate (error of the extrapolated band gap) SE can now be calculated as follows.

$$SE = \frac{s}{\sqrt{n}} \times \sqrt{1 + \frac{\bar{x}^2}{\sigma_x^2}} \quad (5.3)$$

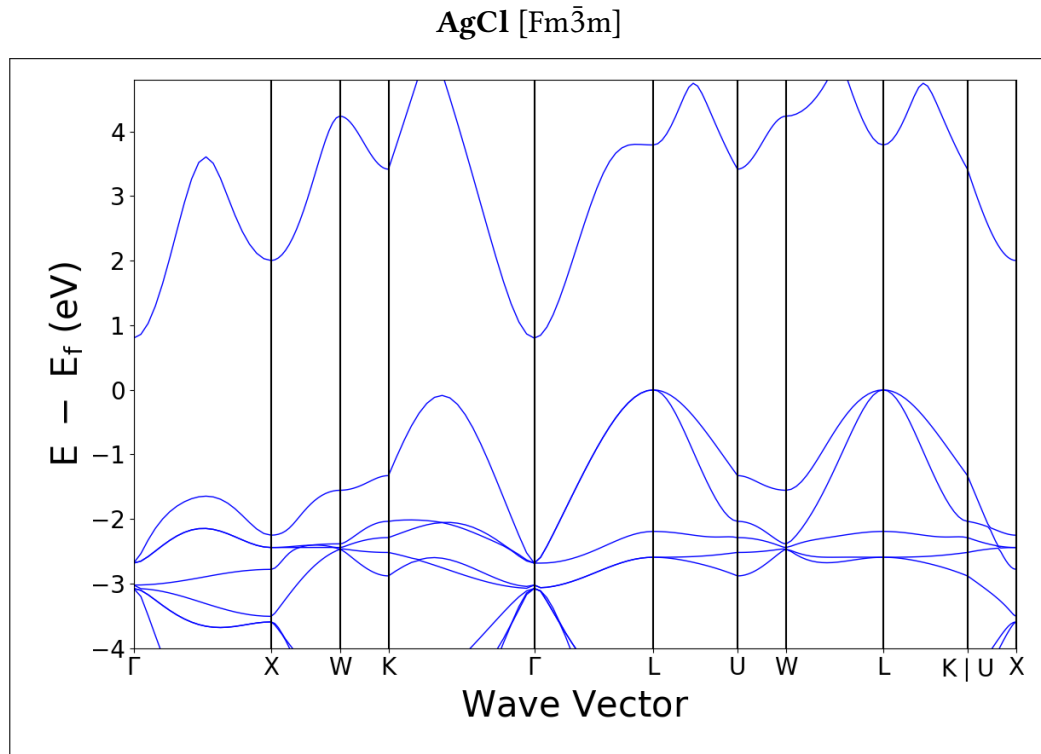
In above equation, $n = 3$ is again the number of data points used for the regression, \bar{x} is the average of the independent variable (here the number of bands N) and σ_x^2 is its variance. Derivations presented here are all taken from reference [29].

Electronic Band Structures

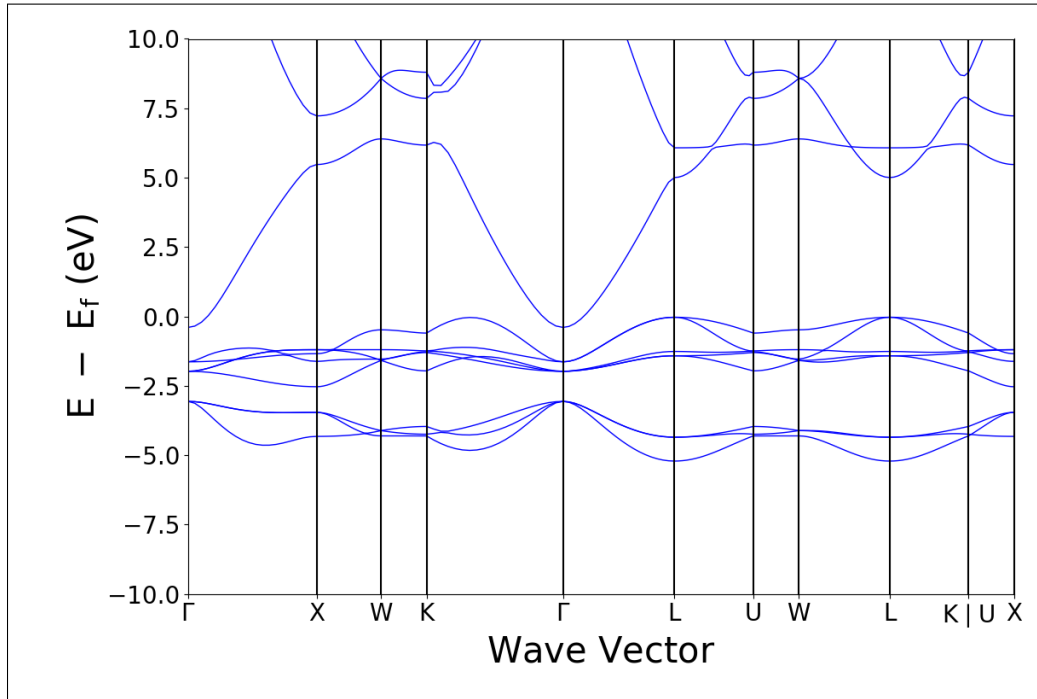
In this section all calculated band structures will be given. The Brillouin zone paths are all taken from ref. [10], energies are given in electron volts.

For all bands, the Fermi energy as found in the *OUTCAR* is subtracted such that the occupied states are below zero, the unoccupied ones above.

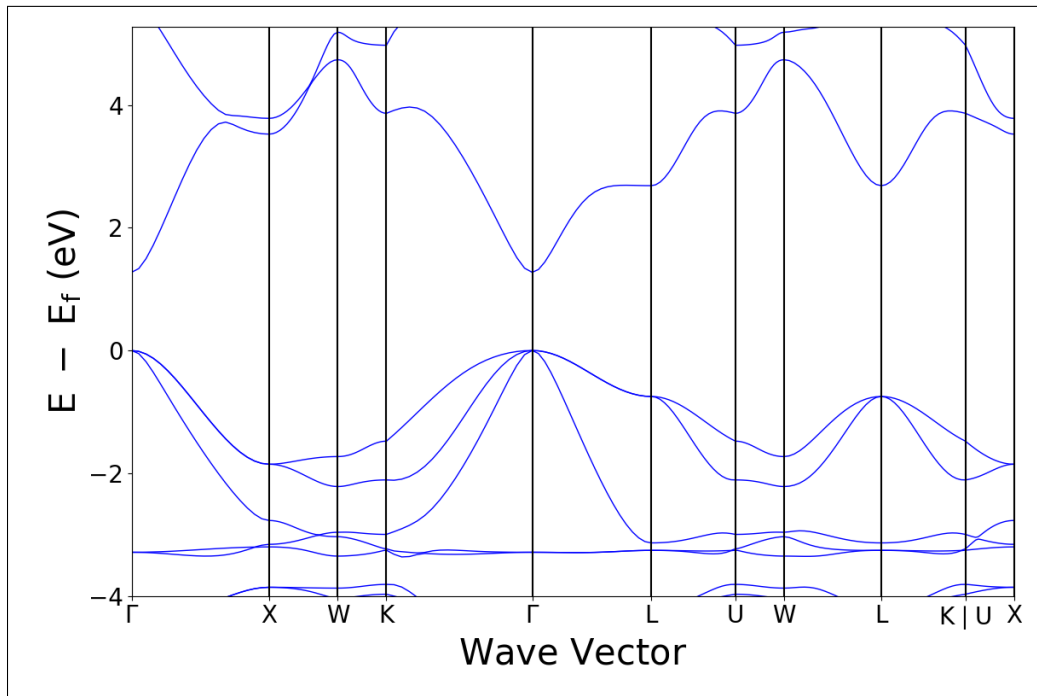
The band structure of systems which require the calculation of material specific parameters due to distorted or more complicated reciprocal unit cells (e.g. body-centered tetragonal in ref. [10]) was not calculated. This is because no routines were implemented, which could have calculated such parameters automatically.



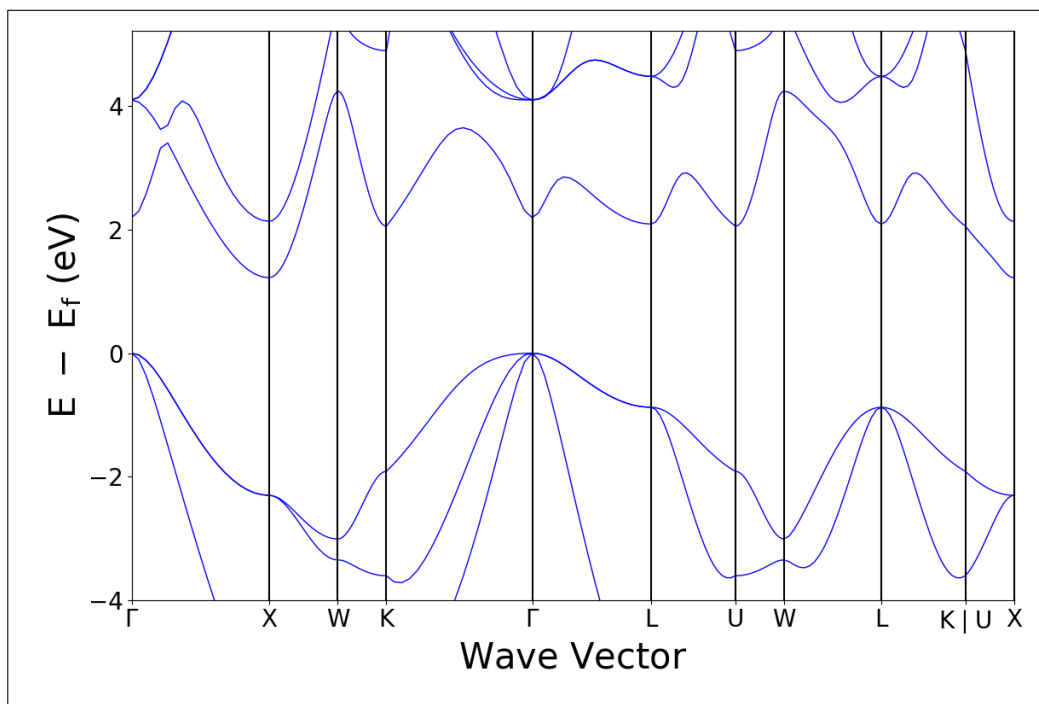
AgF [Fm $\bar{3}$ m]



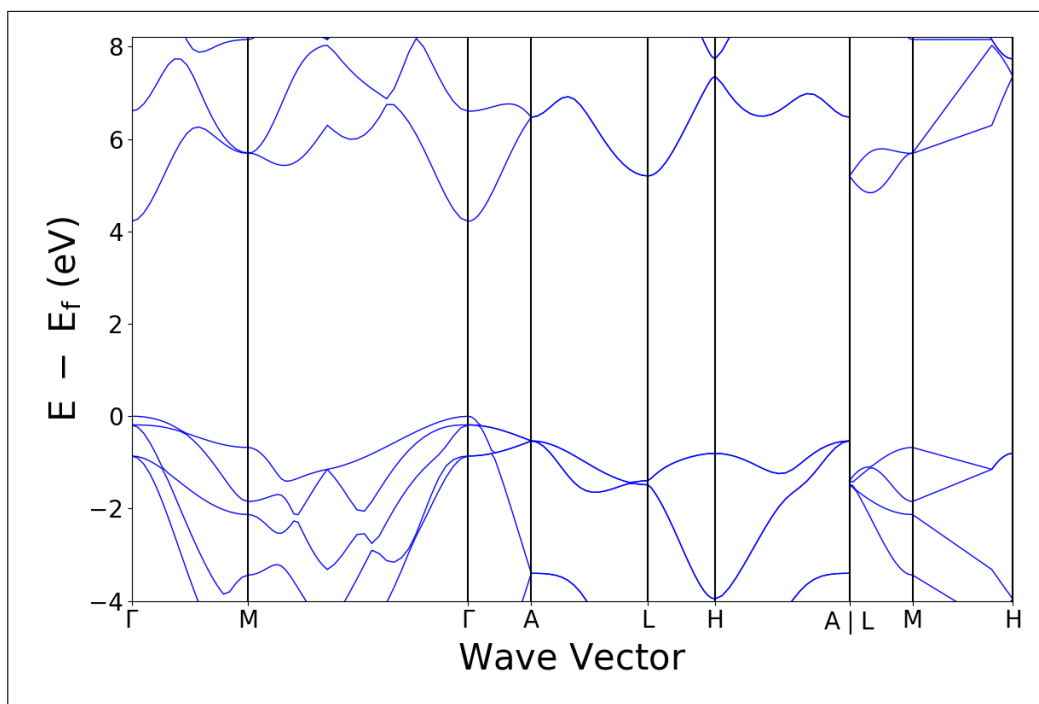
AgI [F4 $\bar{3}$ m]



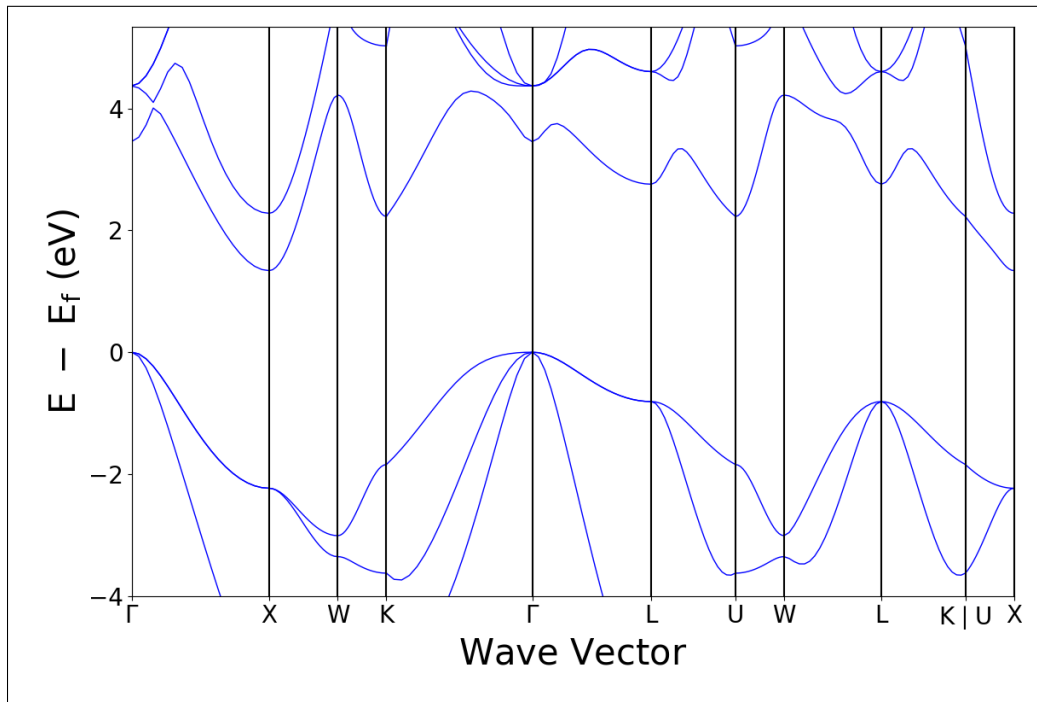
AlAs [$F\bar{4}3m$]



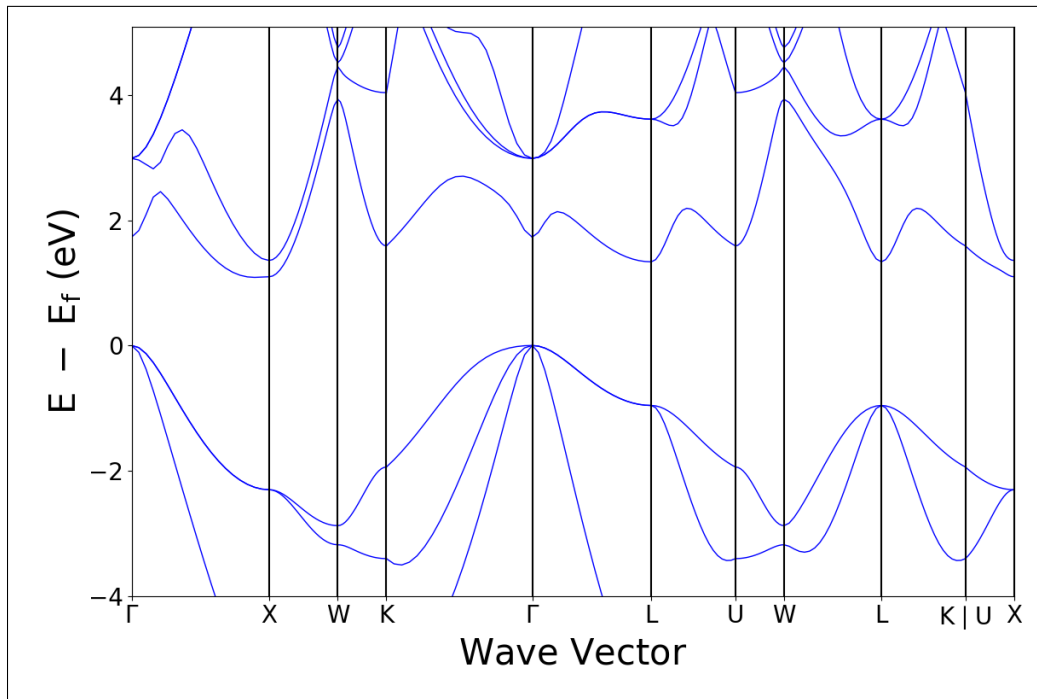
AlN [$P6_3mc$]



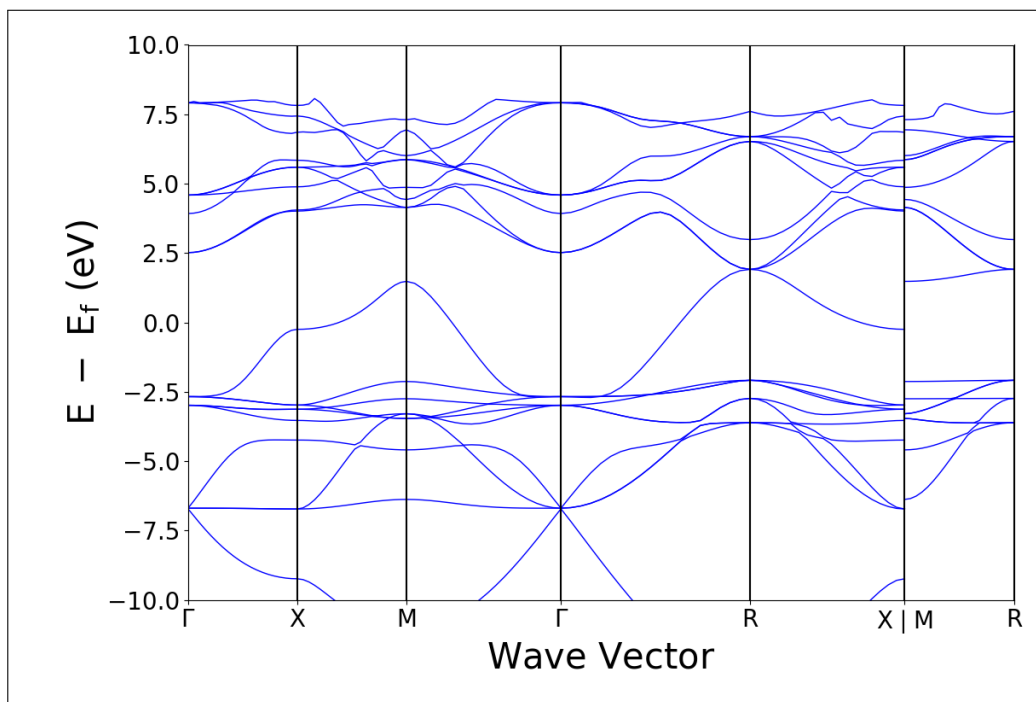
AlP [$F\bar{4}3m$]



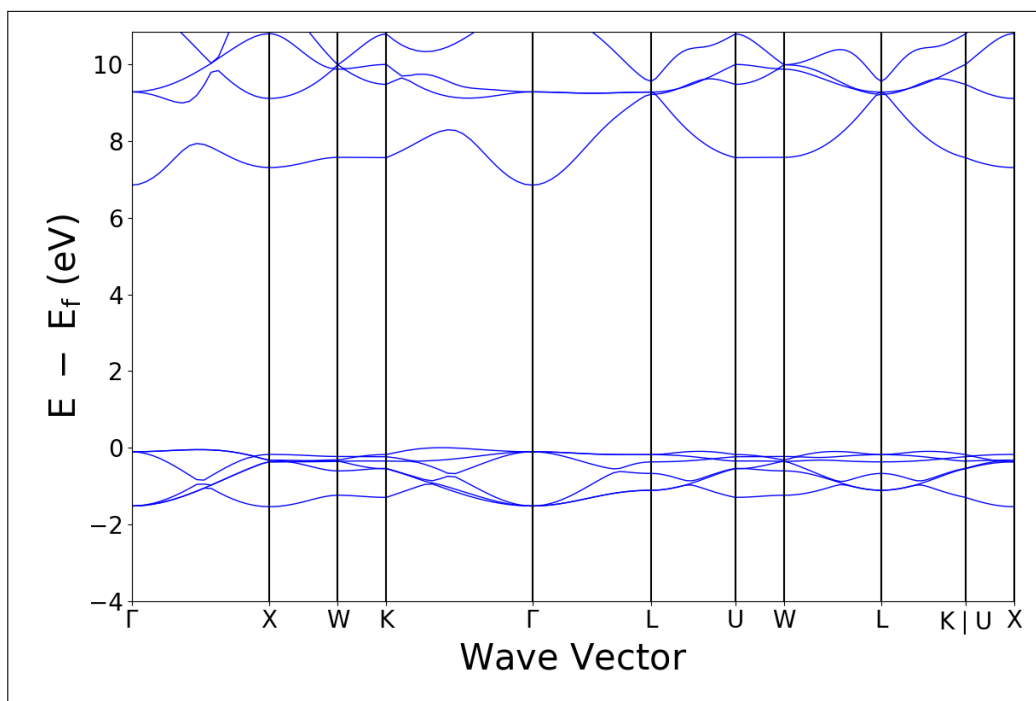
AlSb [$F\bar{4}3m$]



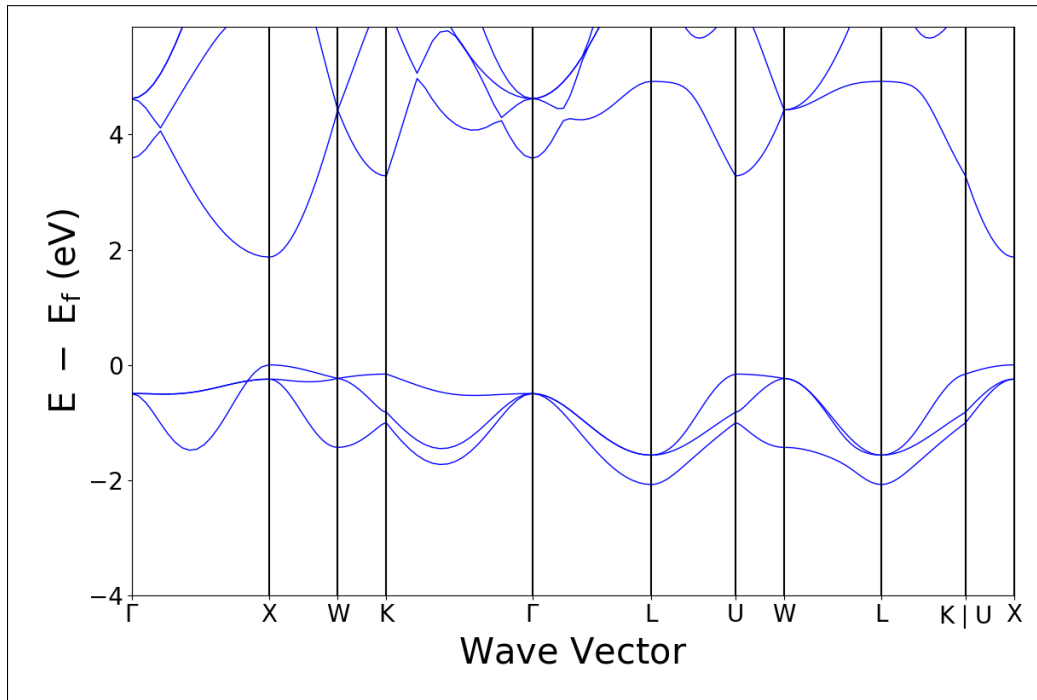
BaBiO₃ [Pm $\bar{3}$ m]



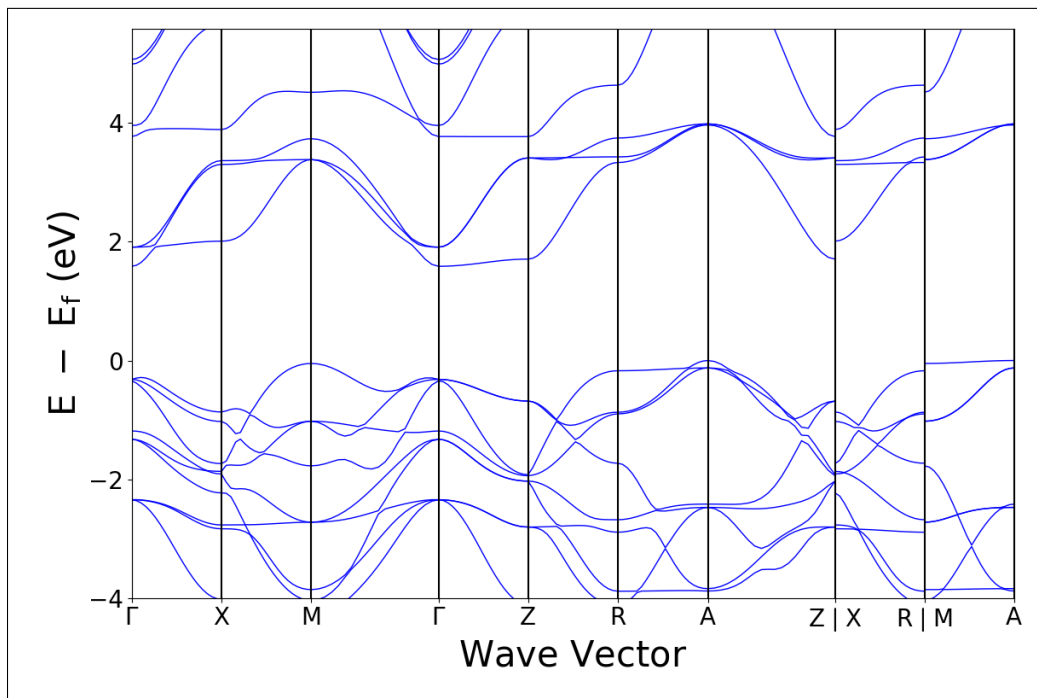
BaF₂ [Fm $\bar{3}$ m]



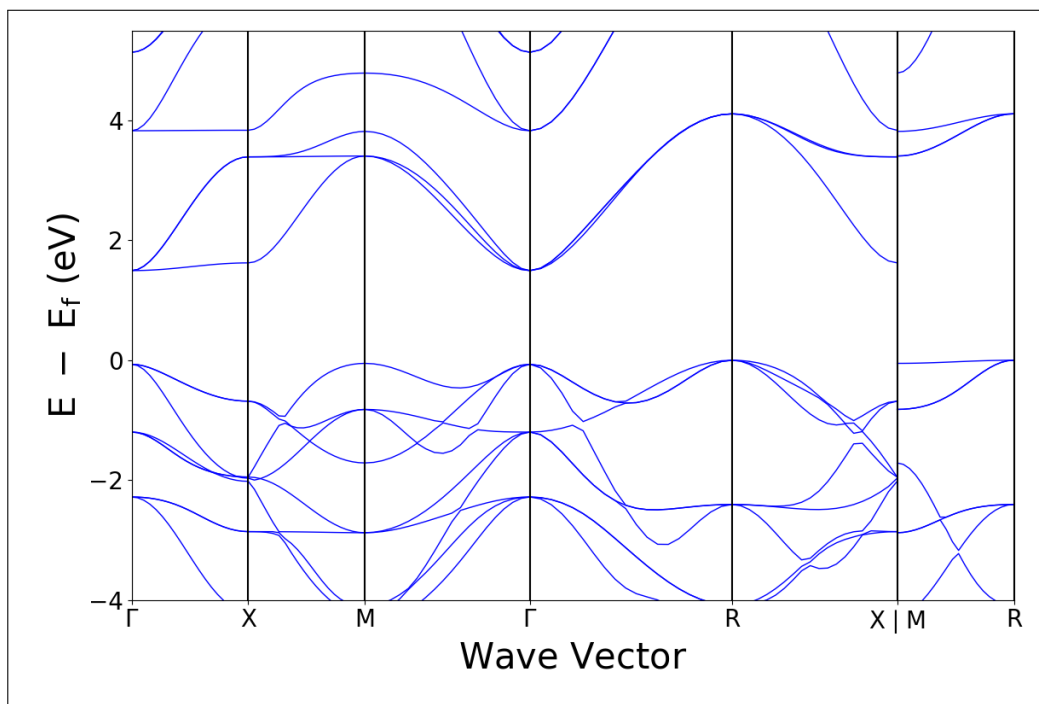
BaO [$Fm\bar{3}m$]



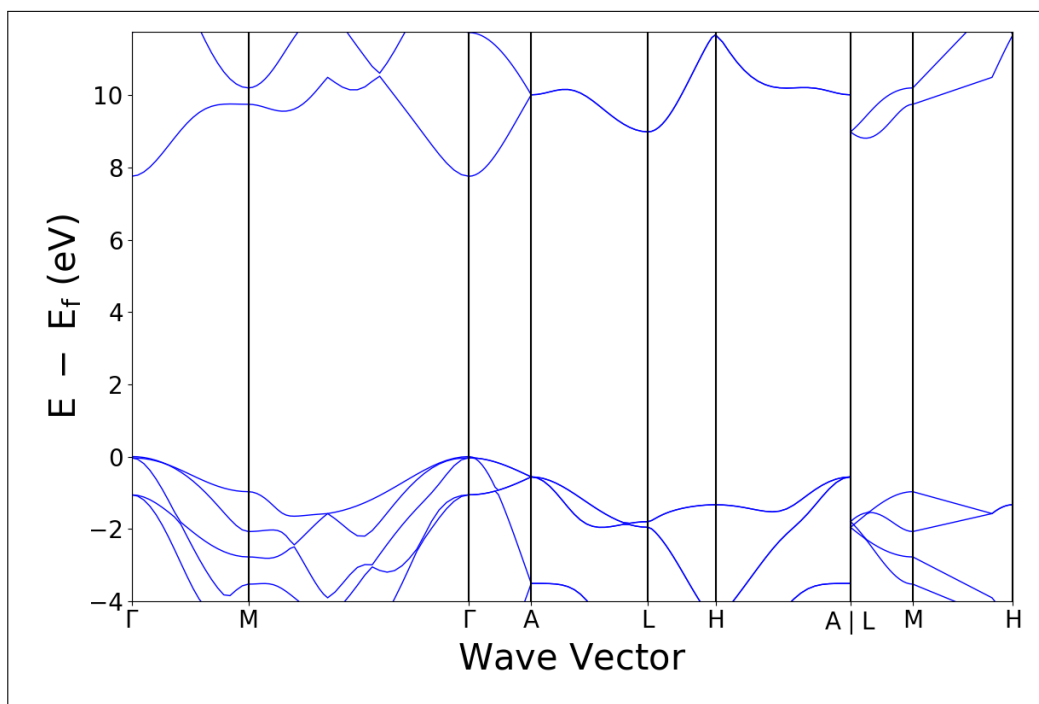
BaTiO₃ [$P4mm$]



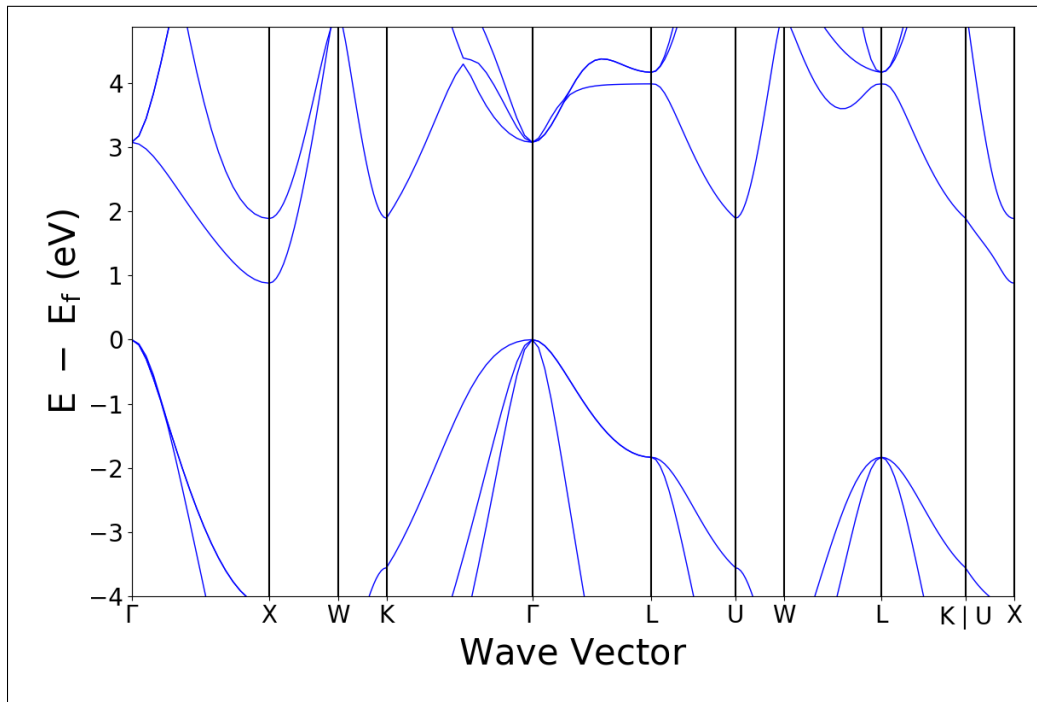
BaTiO₃ [Pm $\bar{3}$ m]



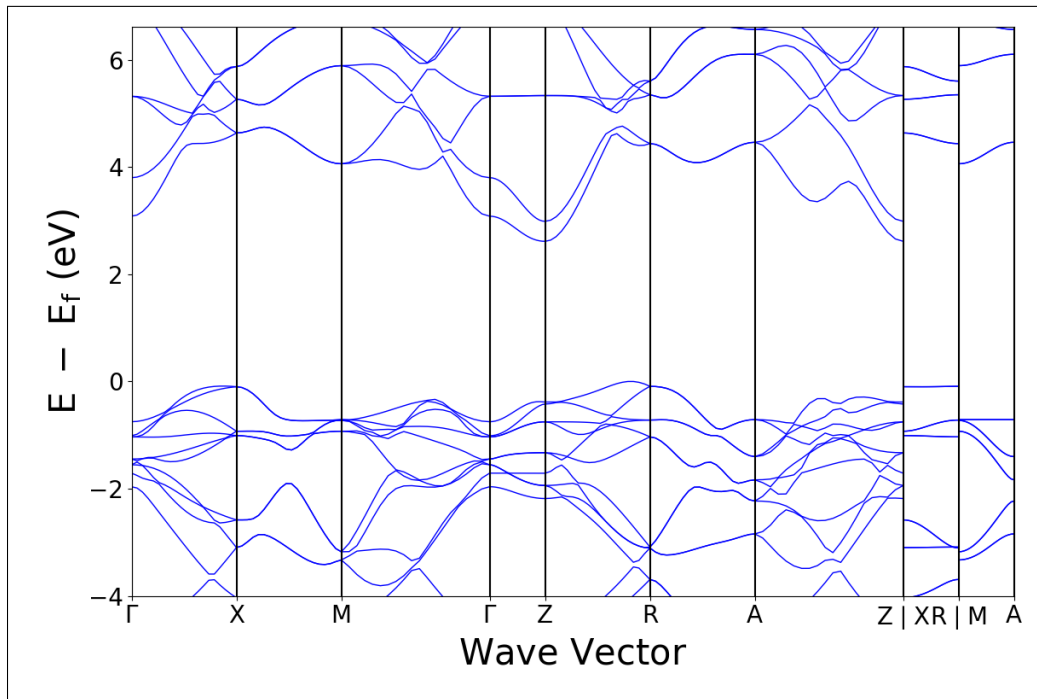
BeO [P6₃mc]



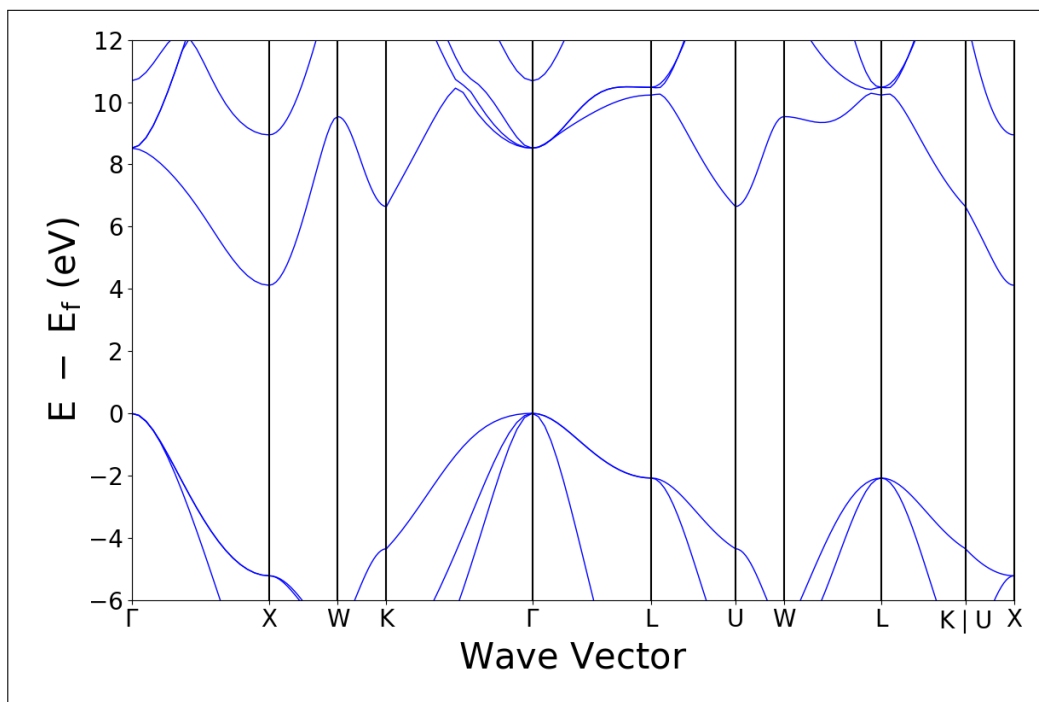
BeTe [$F\bar{4}3m$]



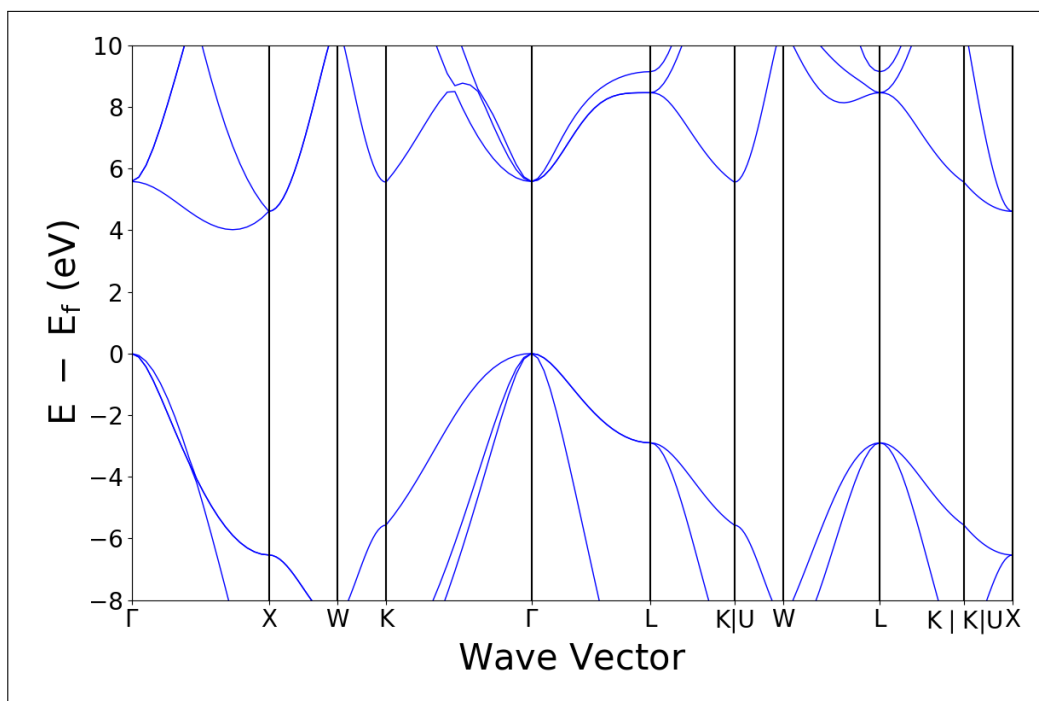
BiOCl [$P4/nmm$]



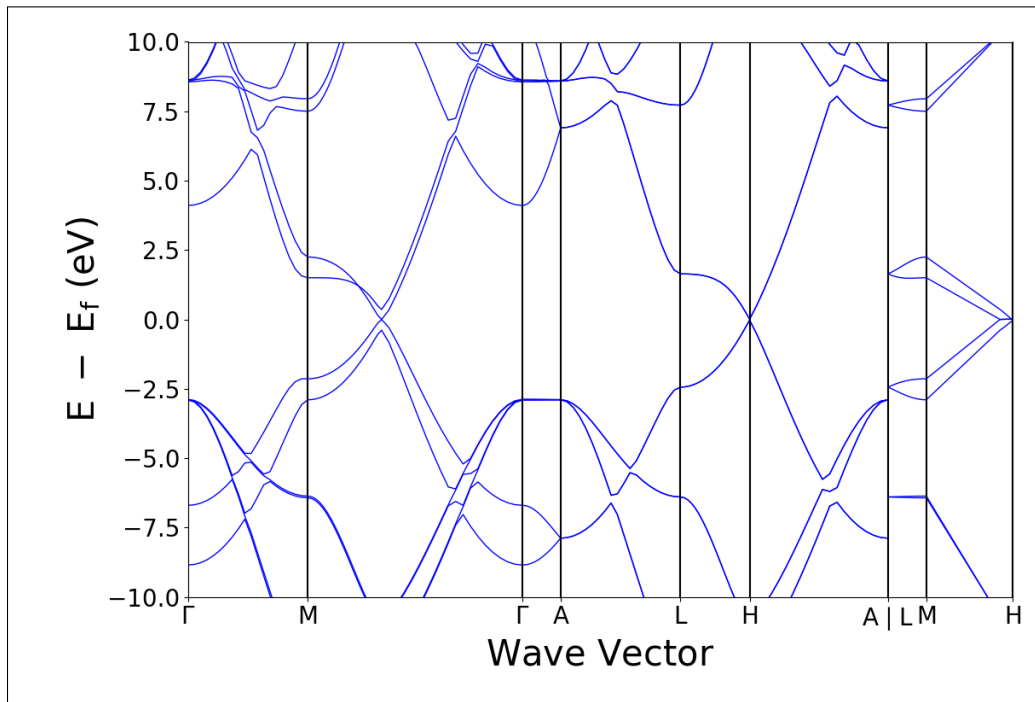
BN [$F\bar{4}3m$]



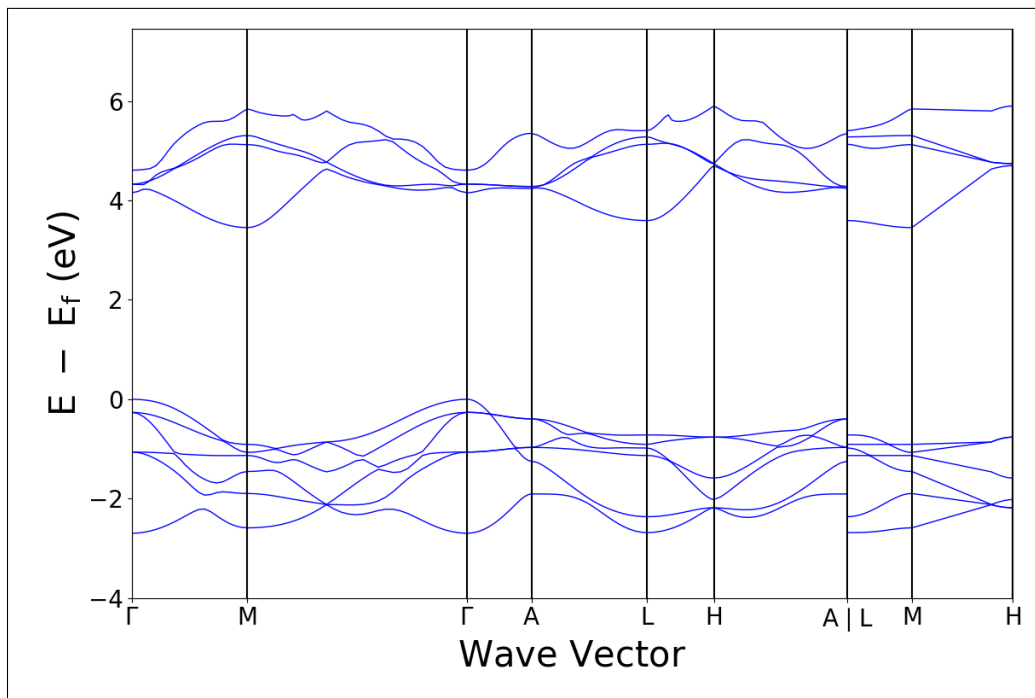
C [$Fd\bar{3}m$]



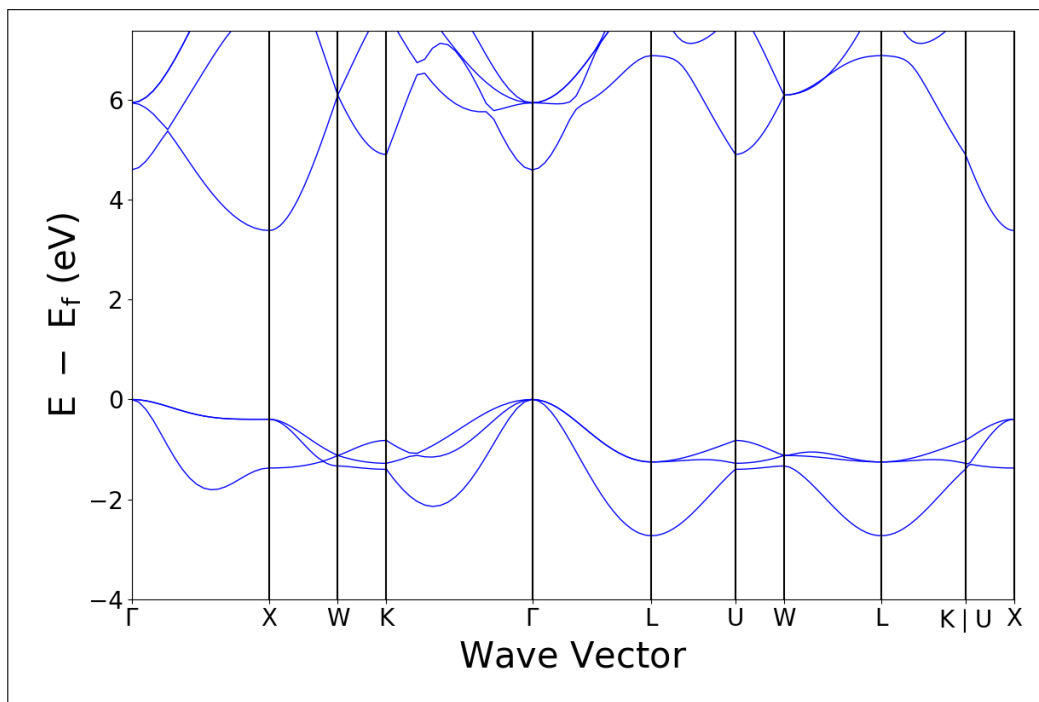
C [$P6_3/mmc$]



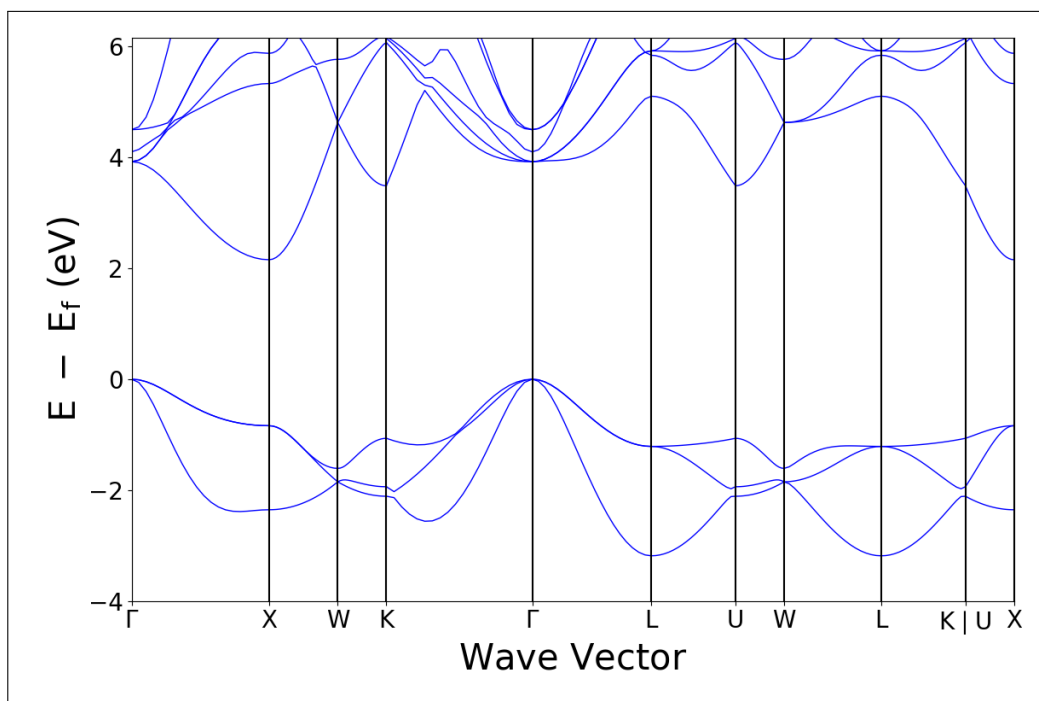
CaI₂ [$P\bar{3}m1$]



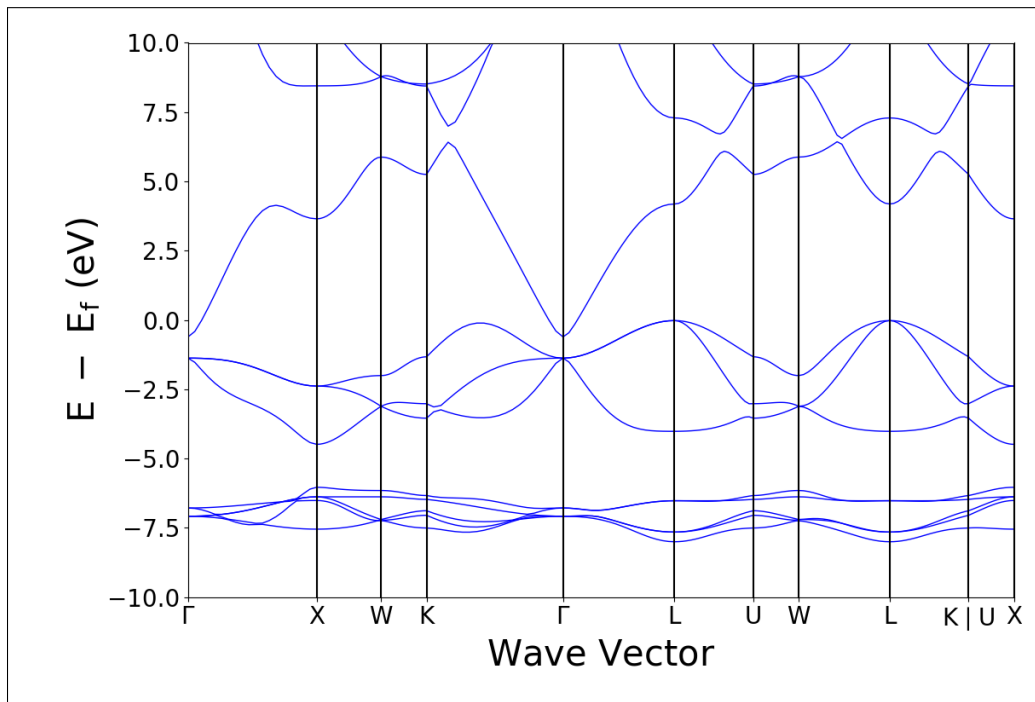
CaO [Fm $\bar{3}$ m]



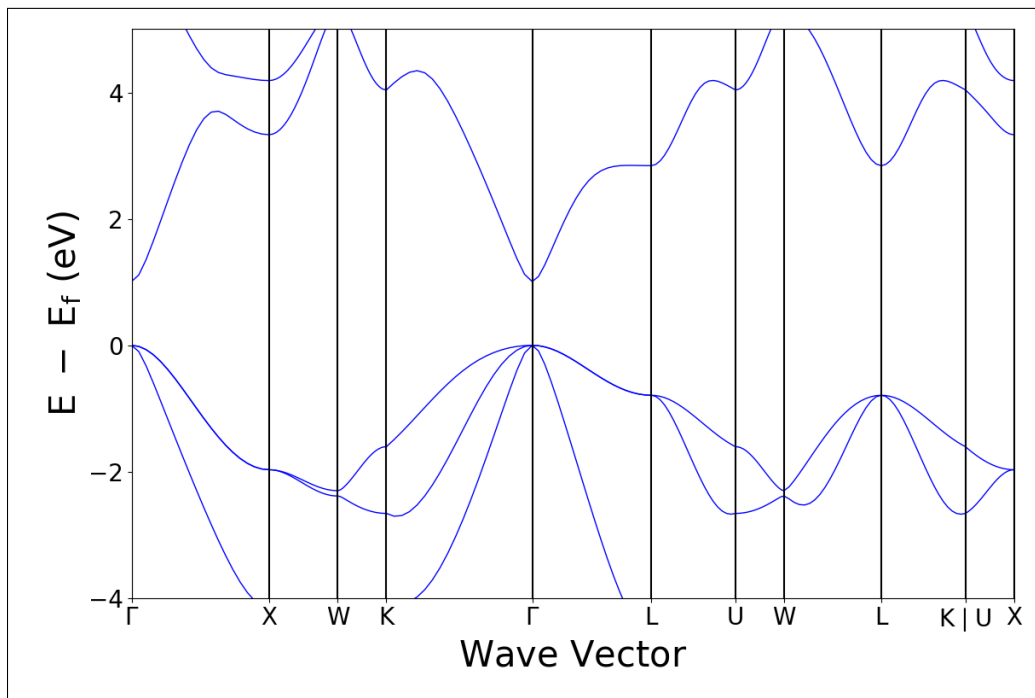
CaS [Fm $\bar{3}$ m]



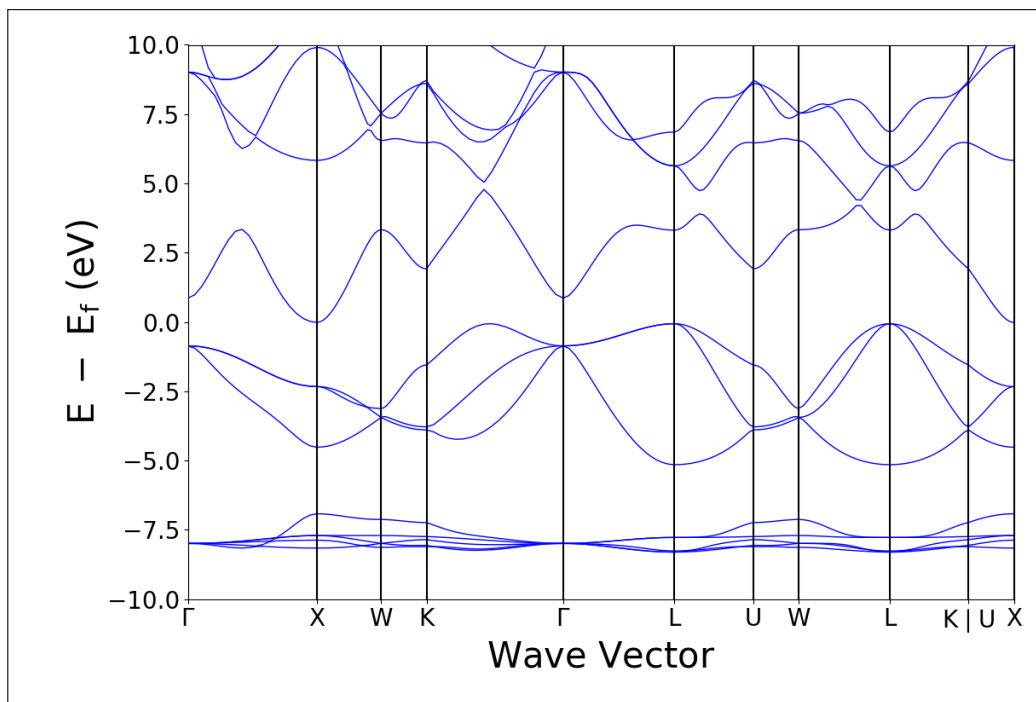
CdO [$Fm\bar{3}m$]



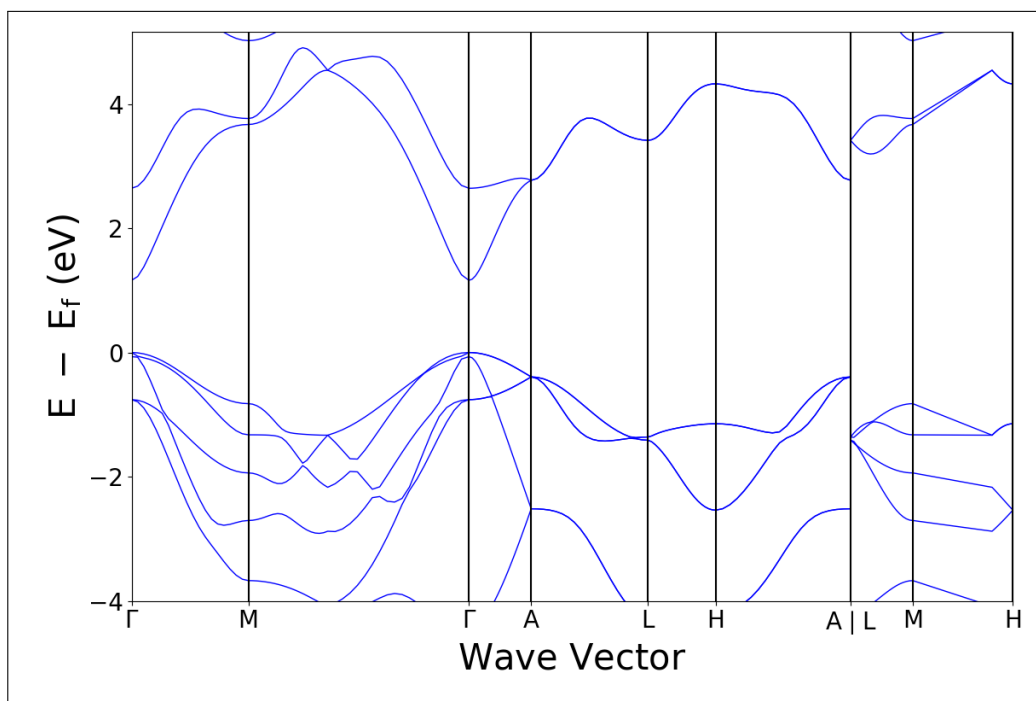
CdS [$F\bar{4}3m$]



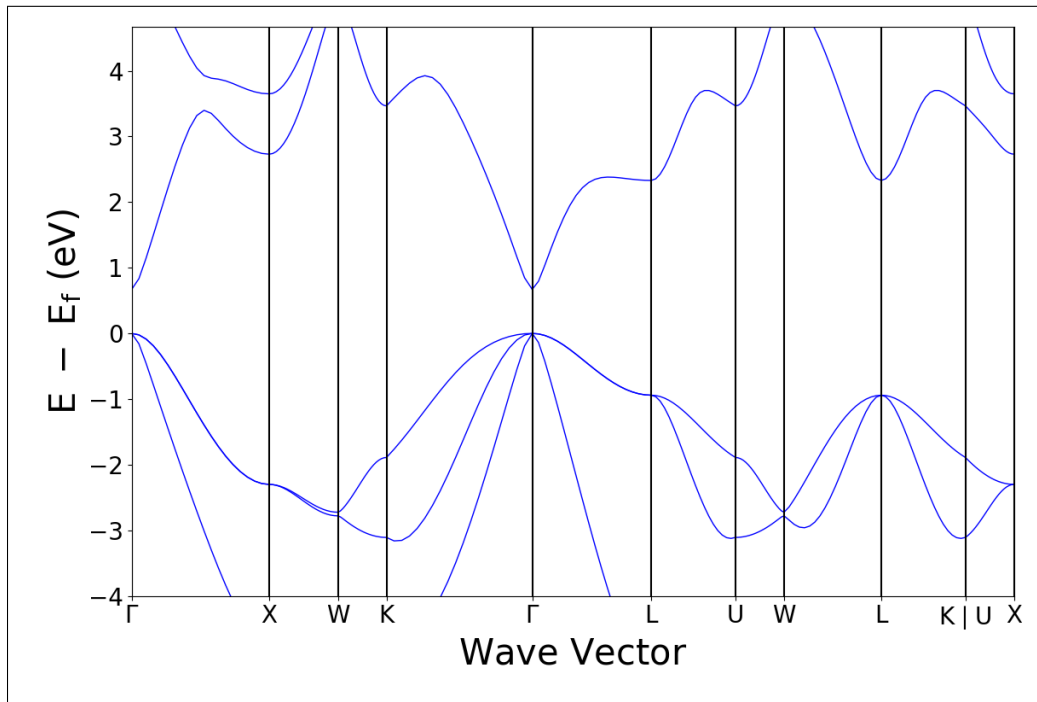
CdS [Fm $\bar{3}$ m]



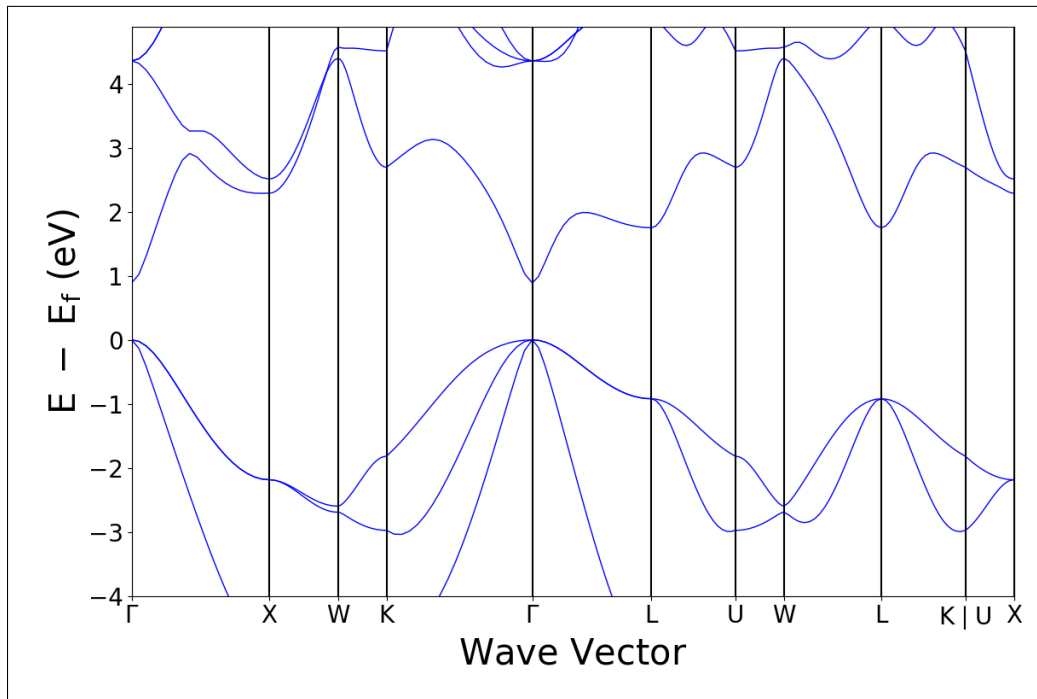
CdS [P6 $_3$ mc]



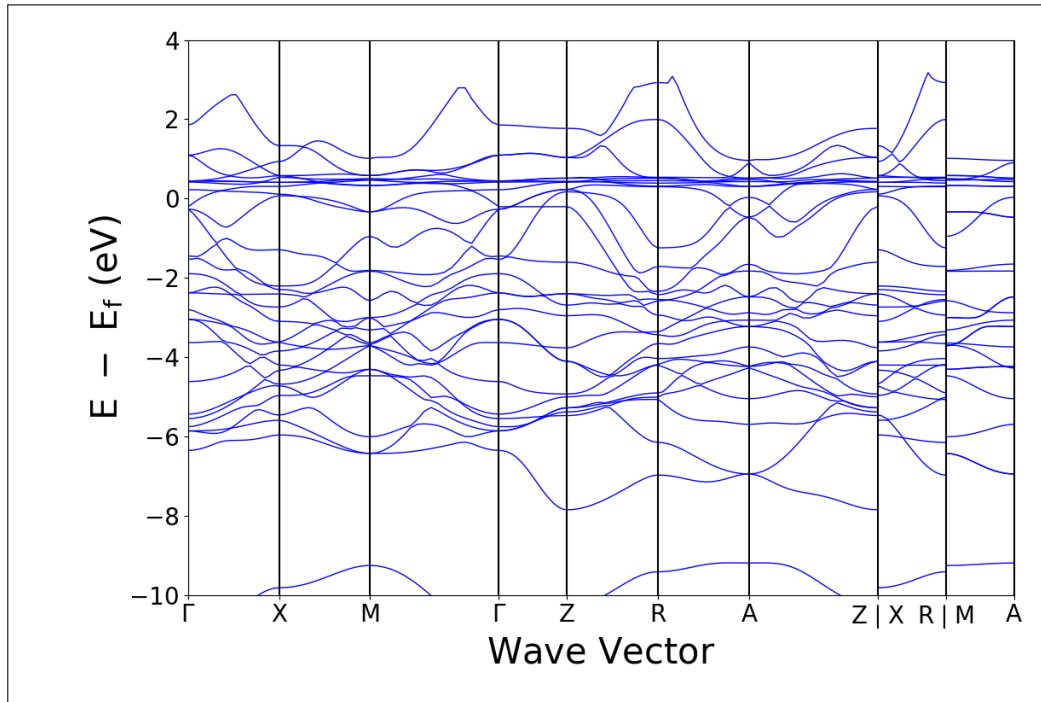
CdSe [$F\bar{4}3m$]



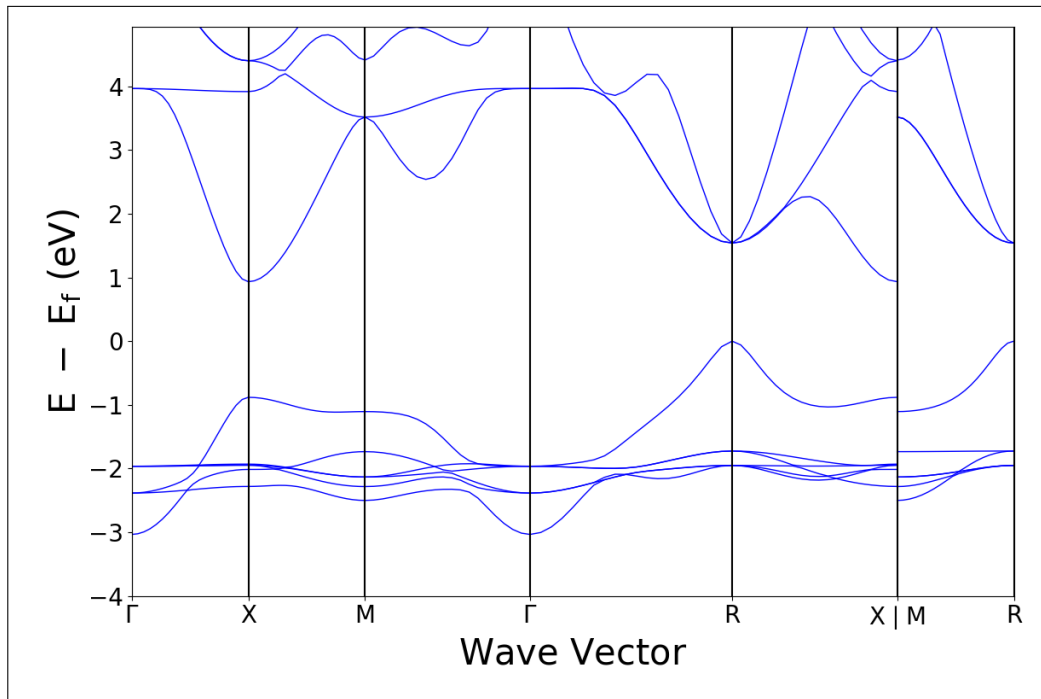
CdTe [$F\bar{4}3m$]



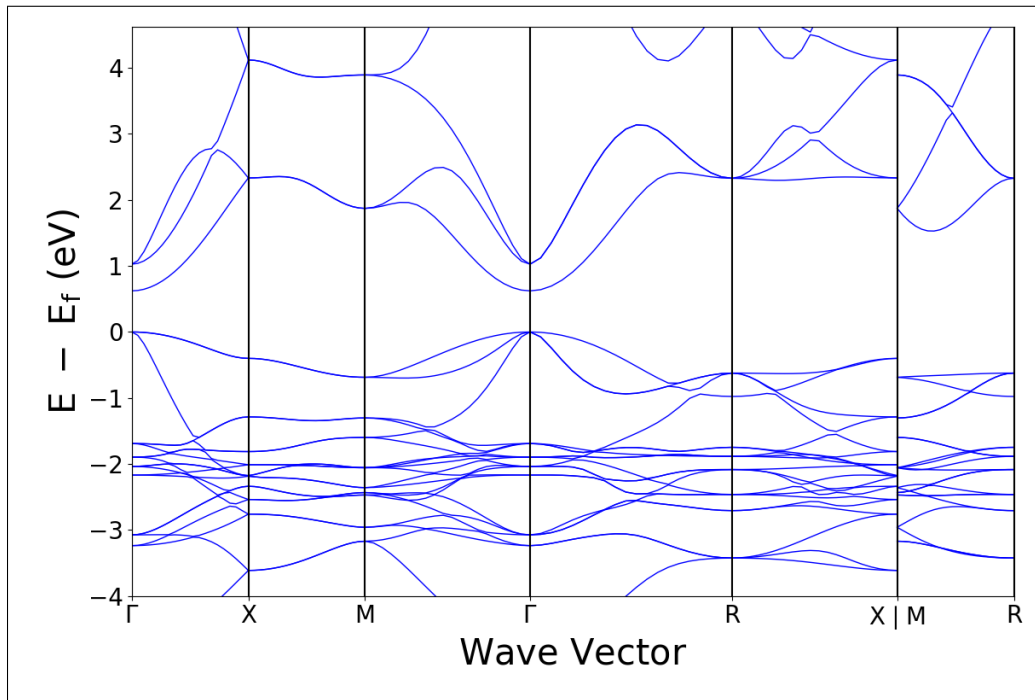
CePt₃Si [P4mm]



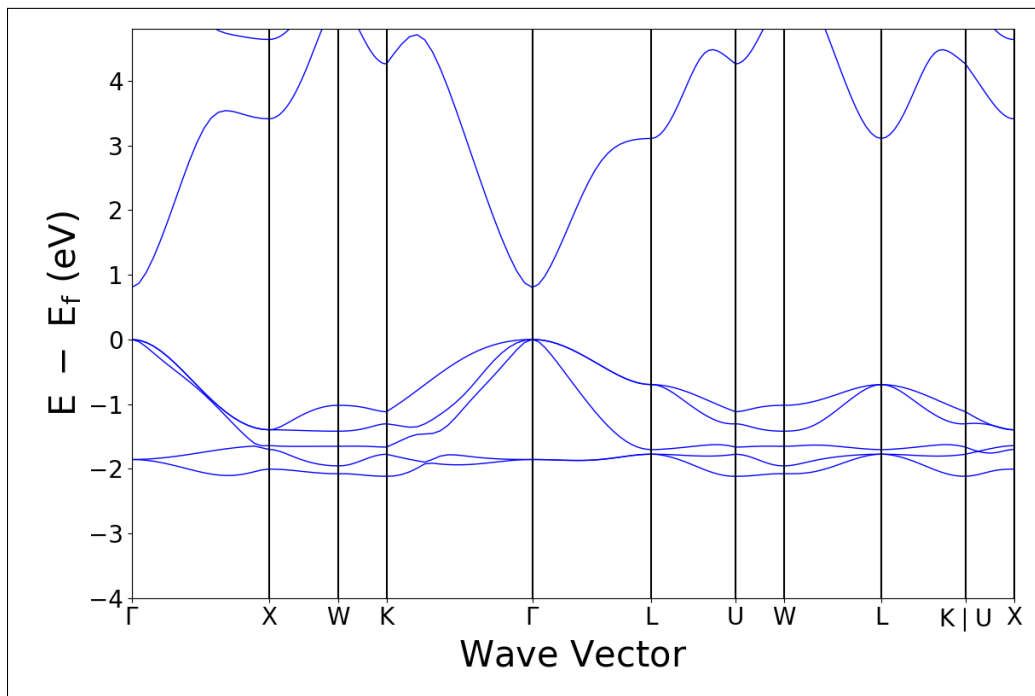
CsAu [Pm $\bar{3}$ m]



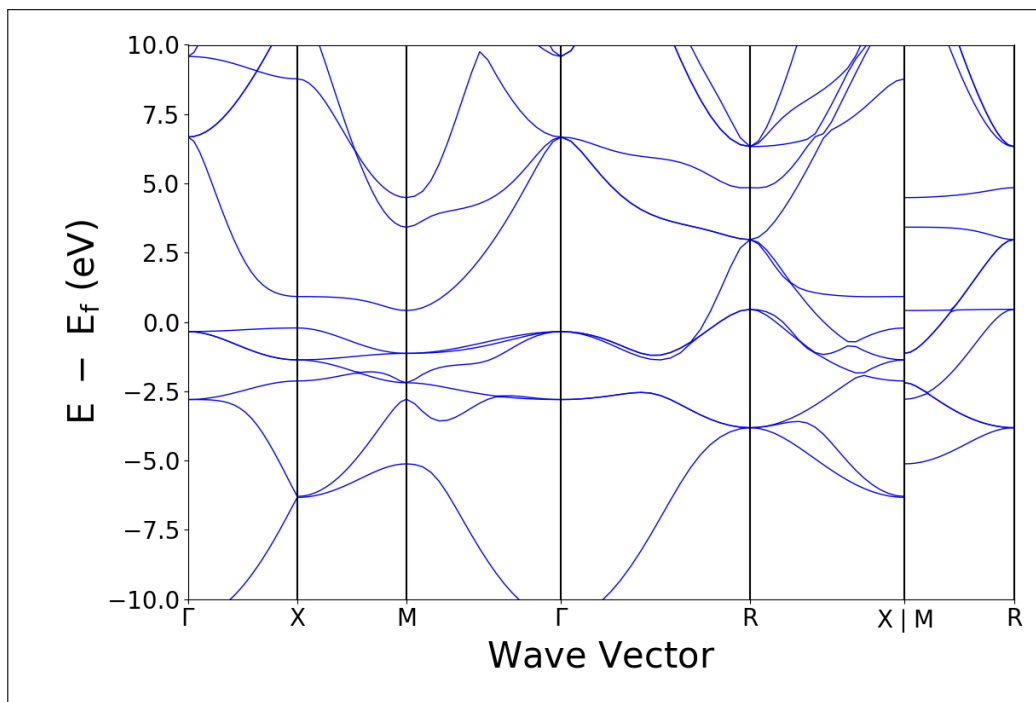
Cu₂O [Pn $\bar{3}$ m]



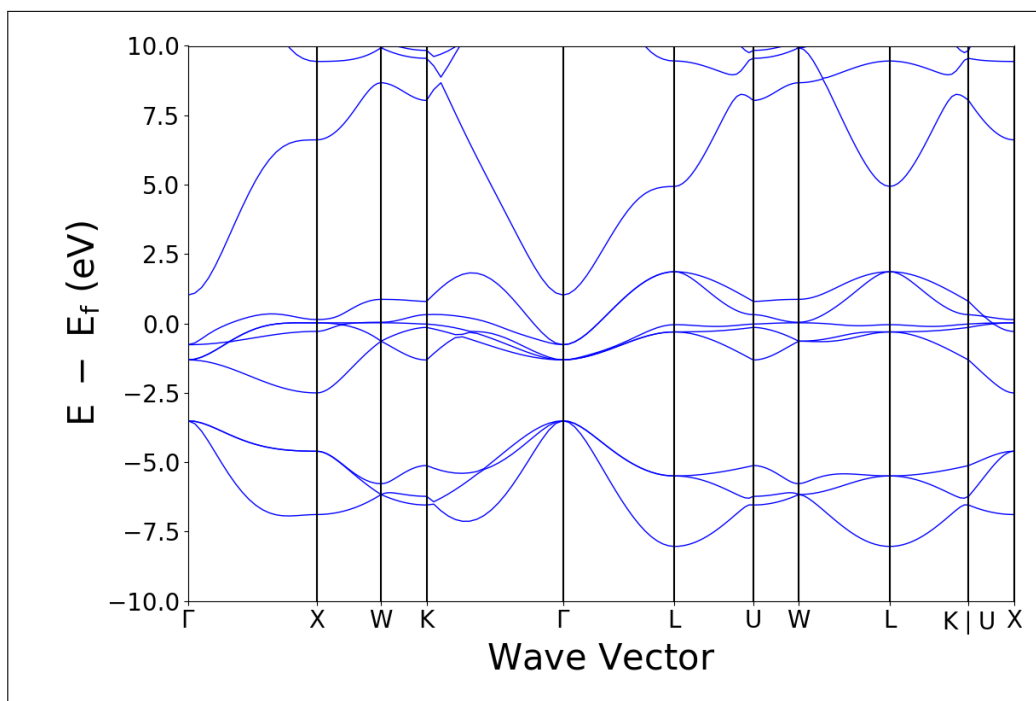
CuCl [F $\bar{4}$ 3m]



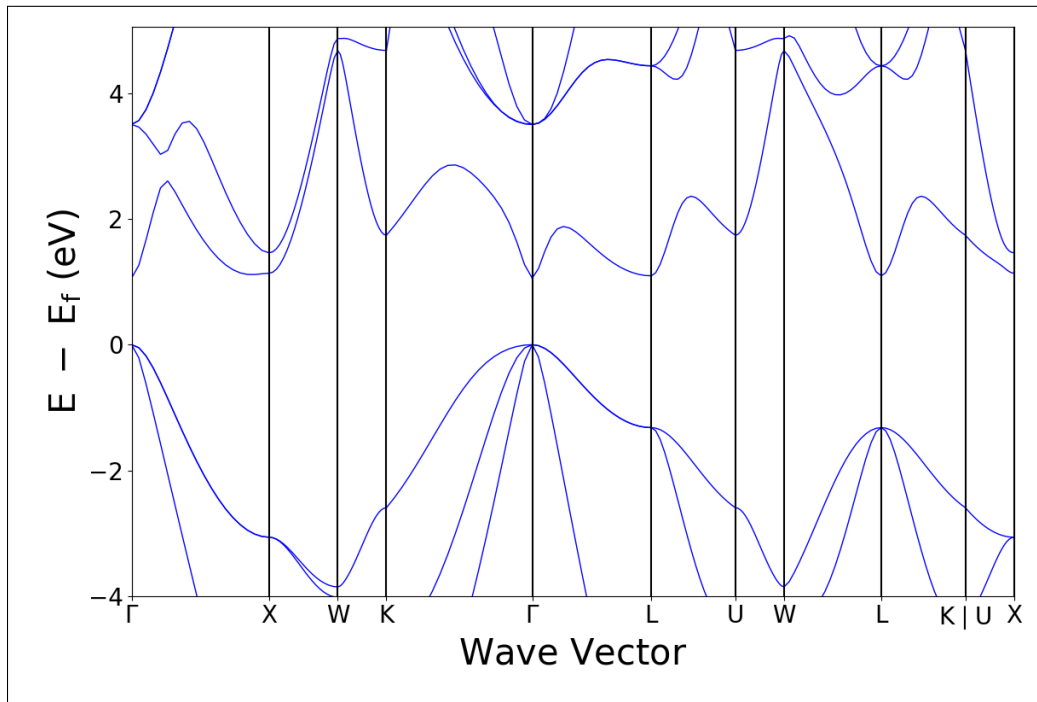
FeAl [$\text{Pm}\bar{3}\text{m}$]



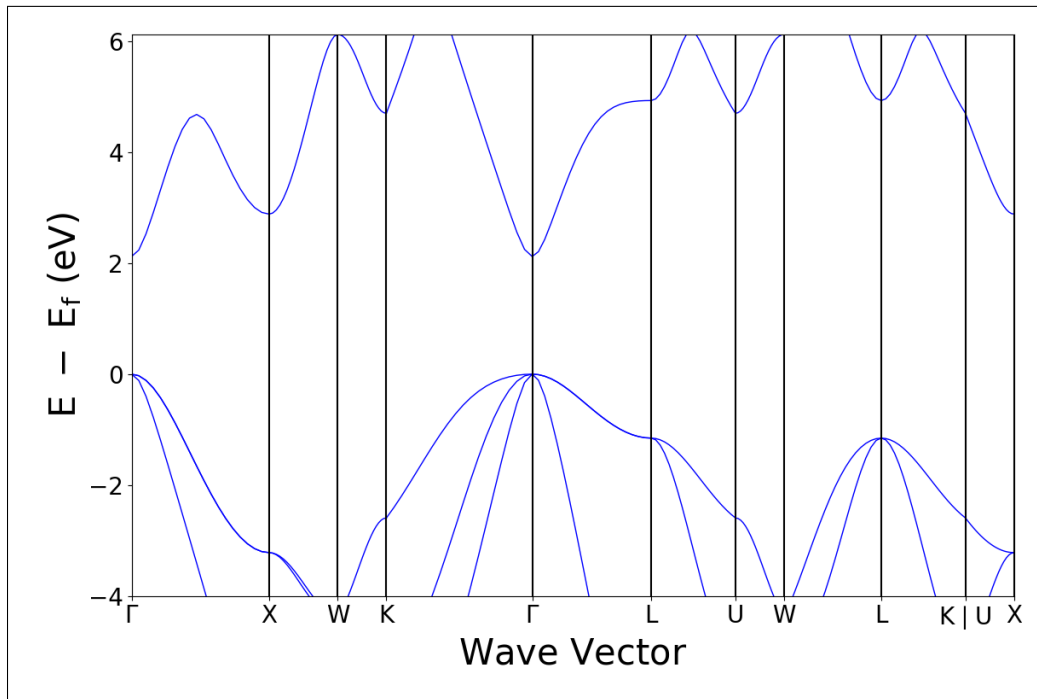
FeO [$\text{Fm}\bar{3}\text{m}$]



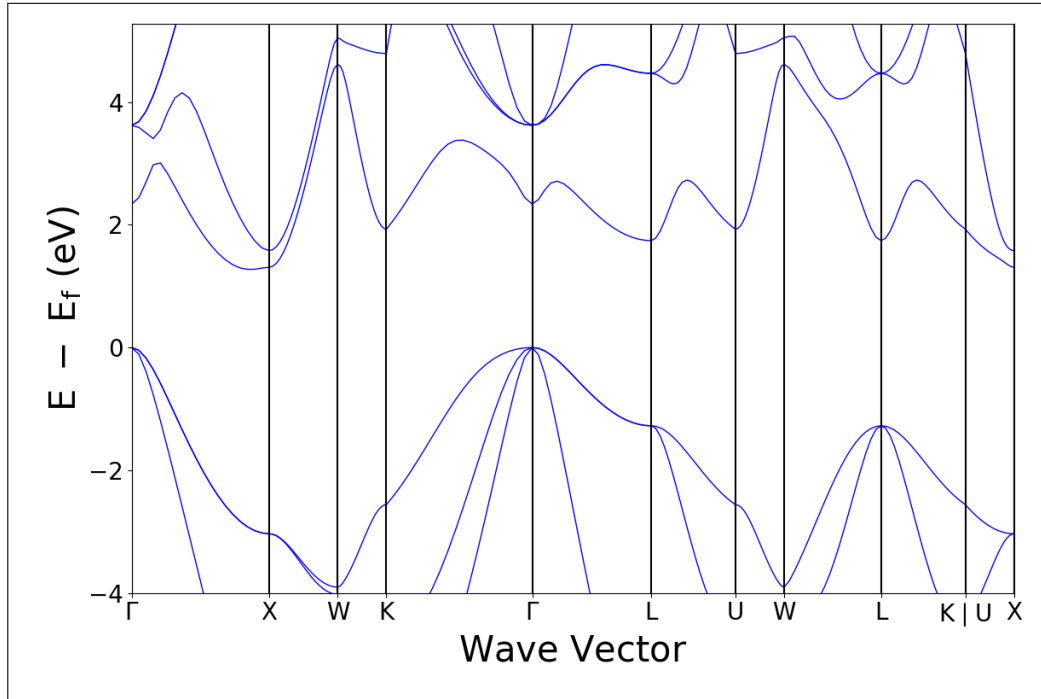
GaAs [$F\bar{4}3m$]



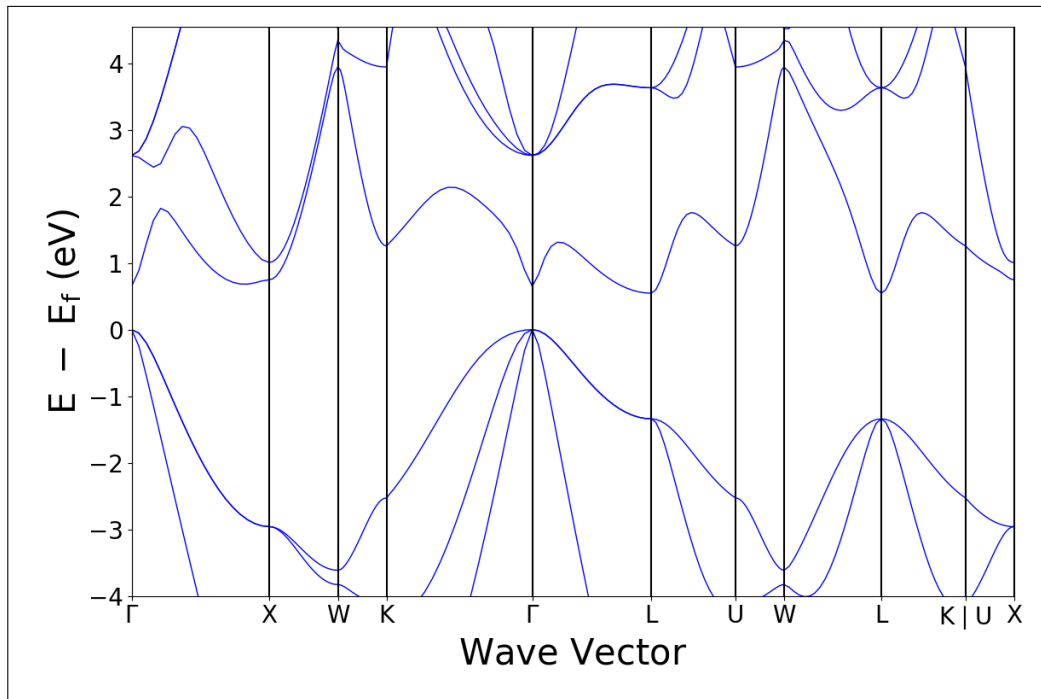
GaN [$F\bar{4}3m$]

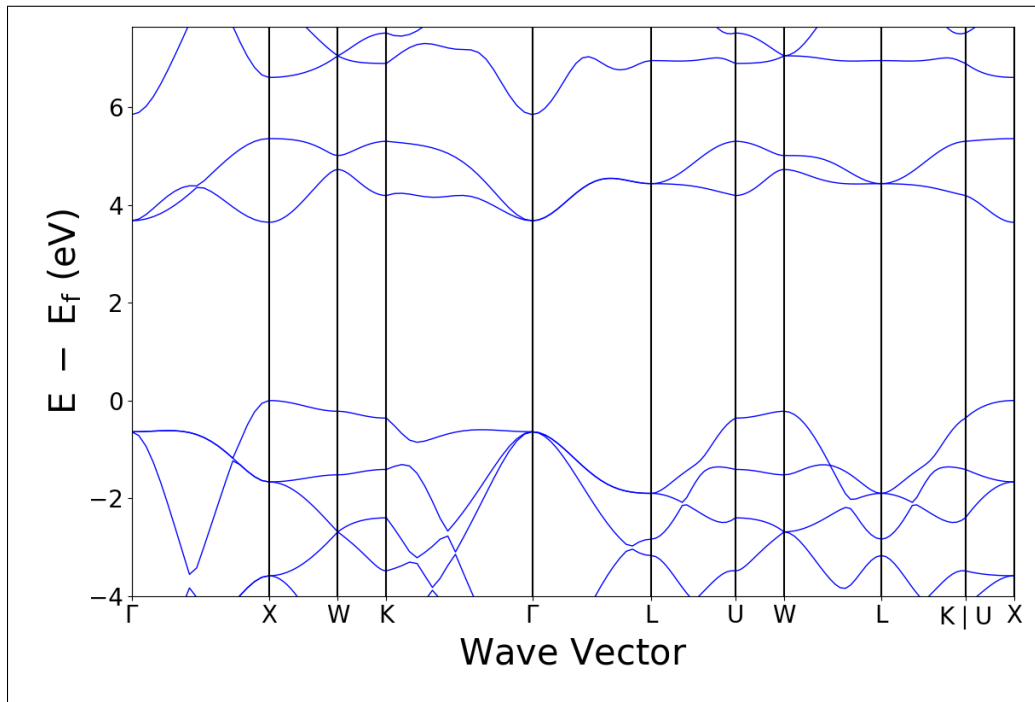
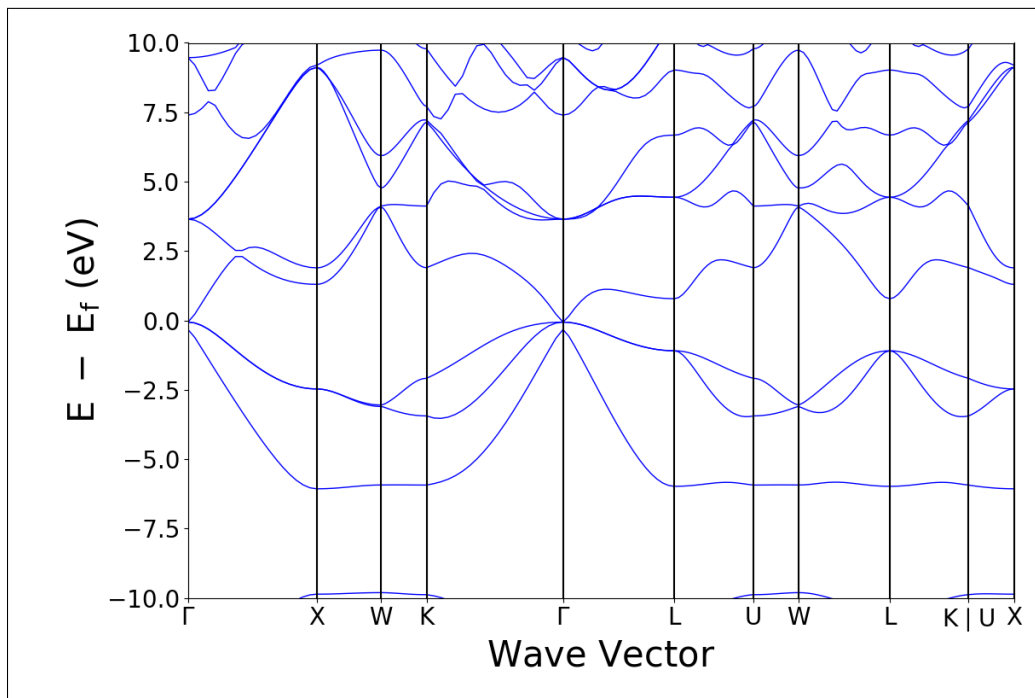


GaP [$F\bar{4}3m$]

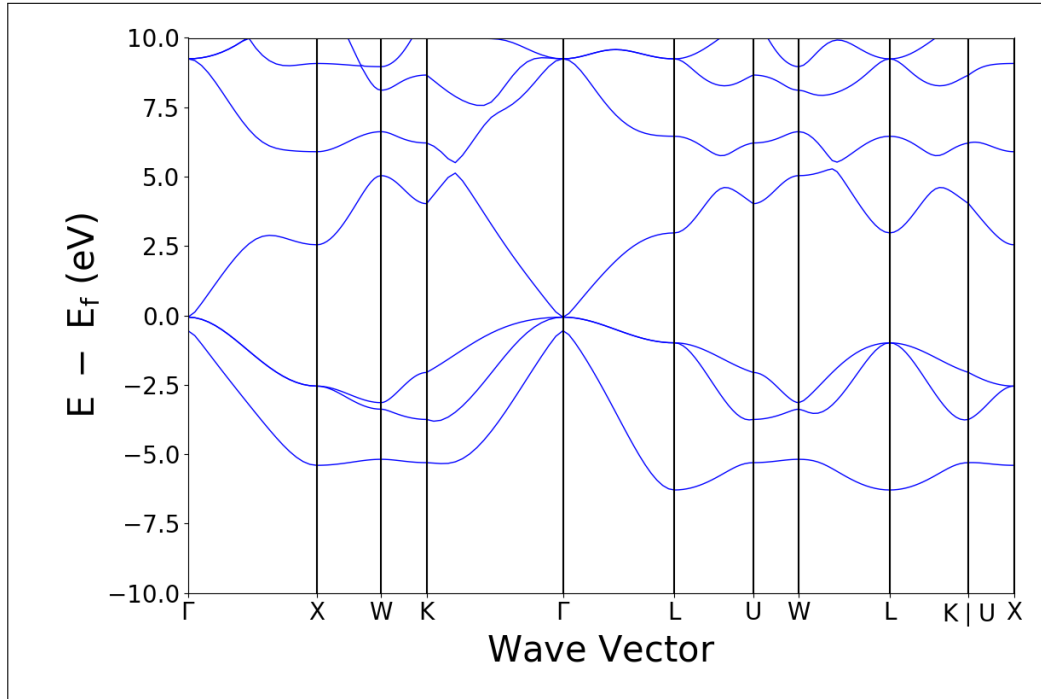


GaSb [$F\bar{4}3m$]

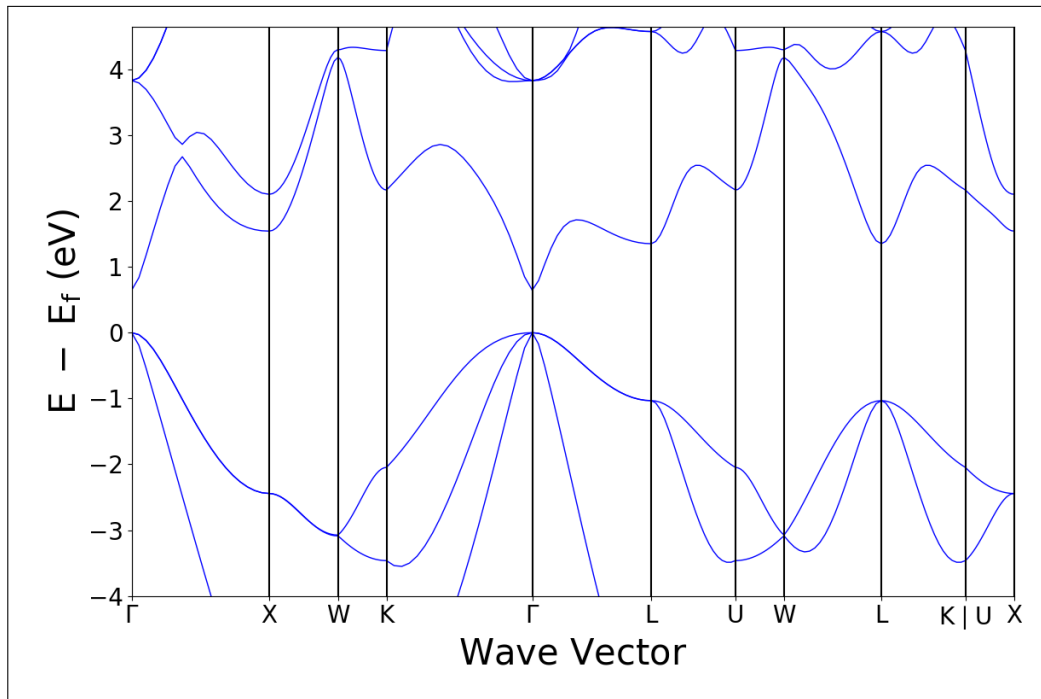


HfO₂ [Fm $\bar{3}$ m]**InAs** [F $\bar{4}$ 3m]

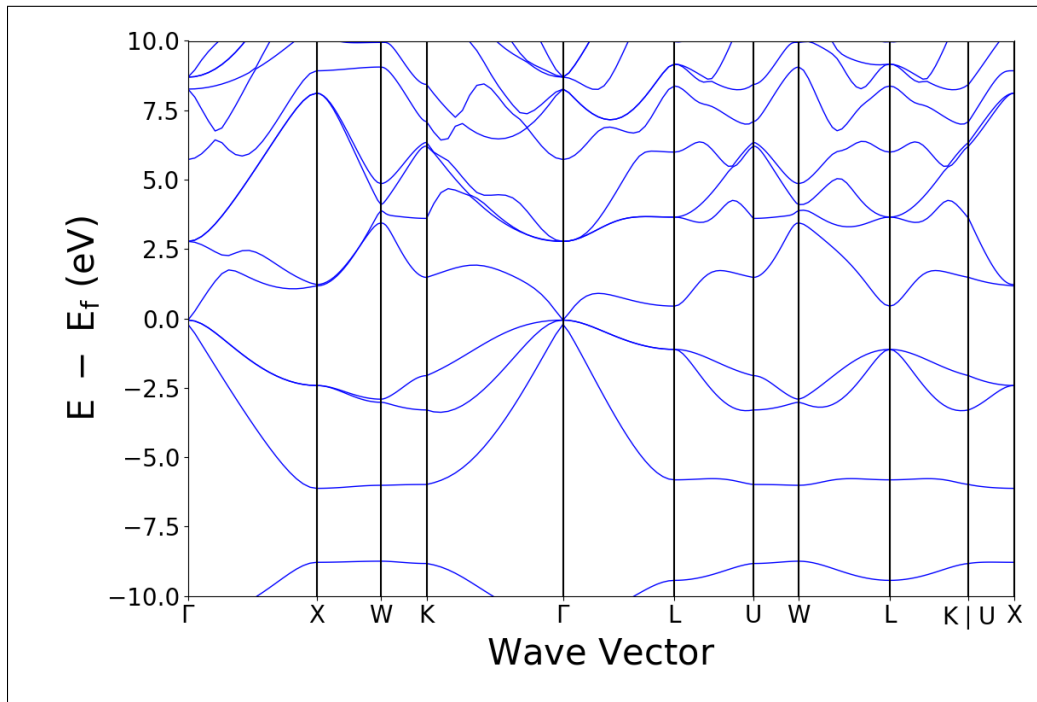
InN [$\bar{F}43m$]



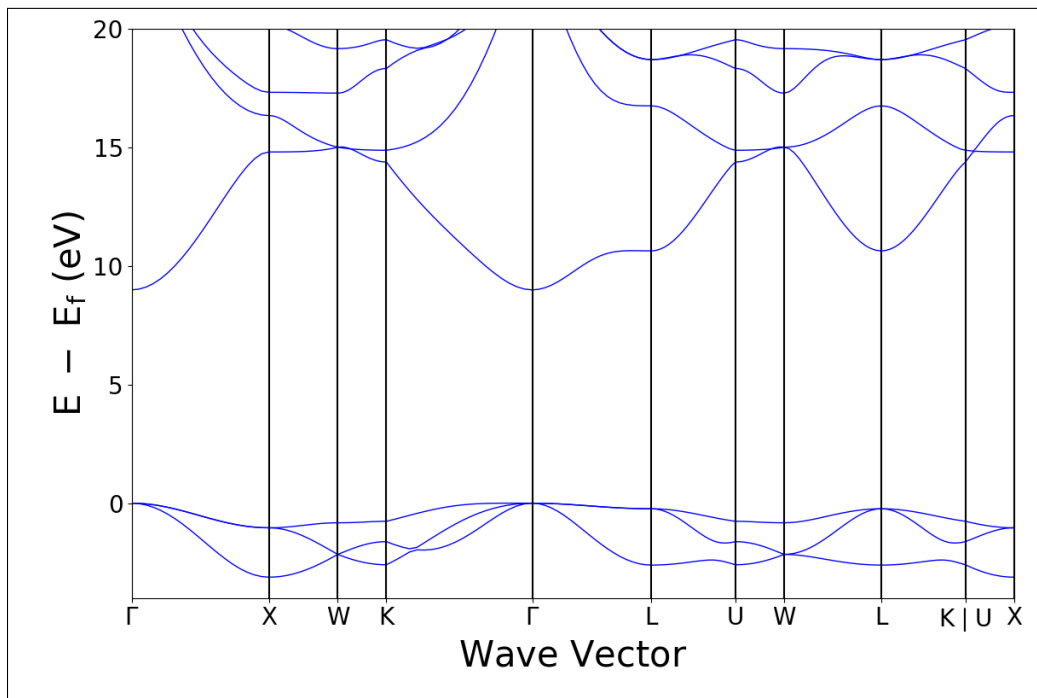
InP [$\bar{F}43m$]



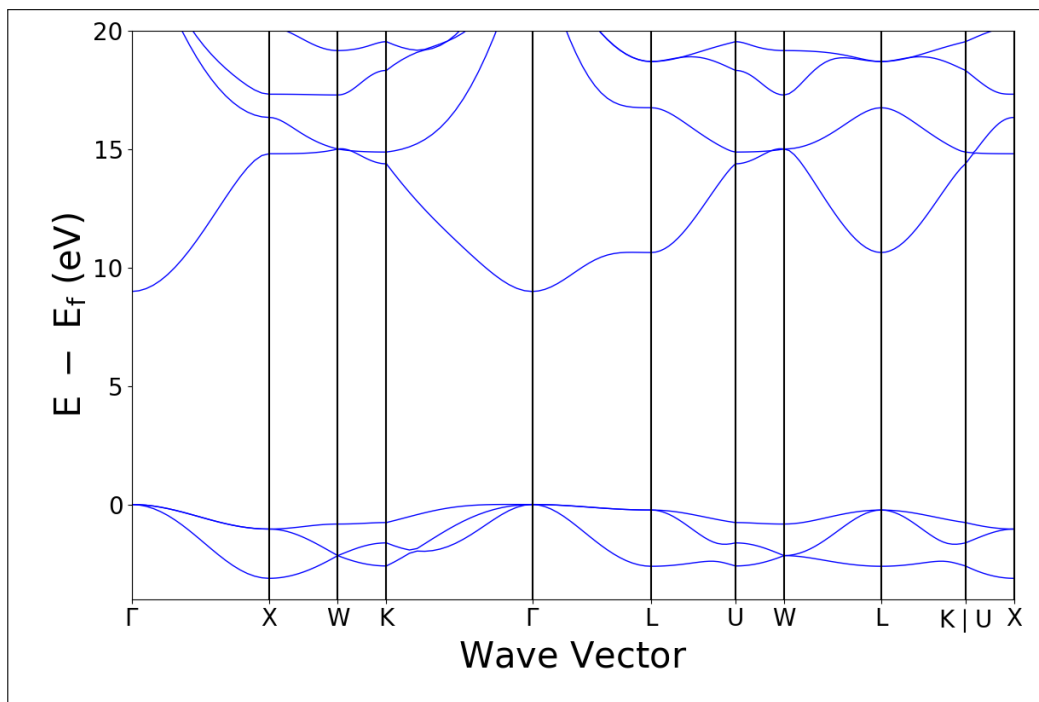
InSb [$F\bar{4}3m$]



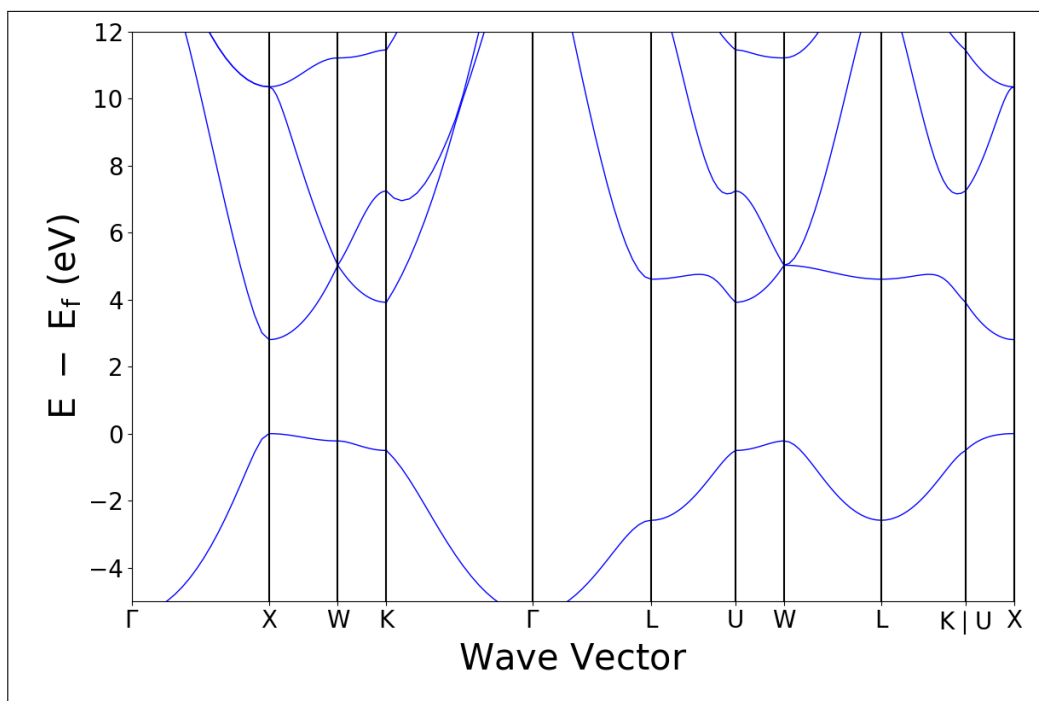
LiF [$Fm\bar{3}m$]

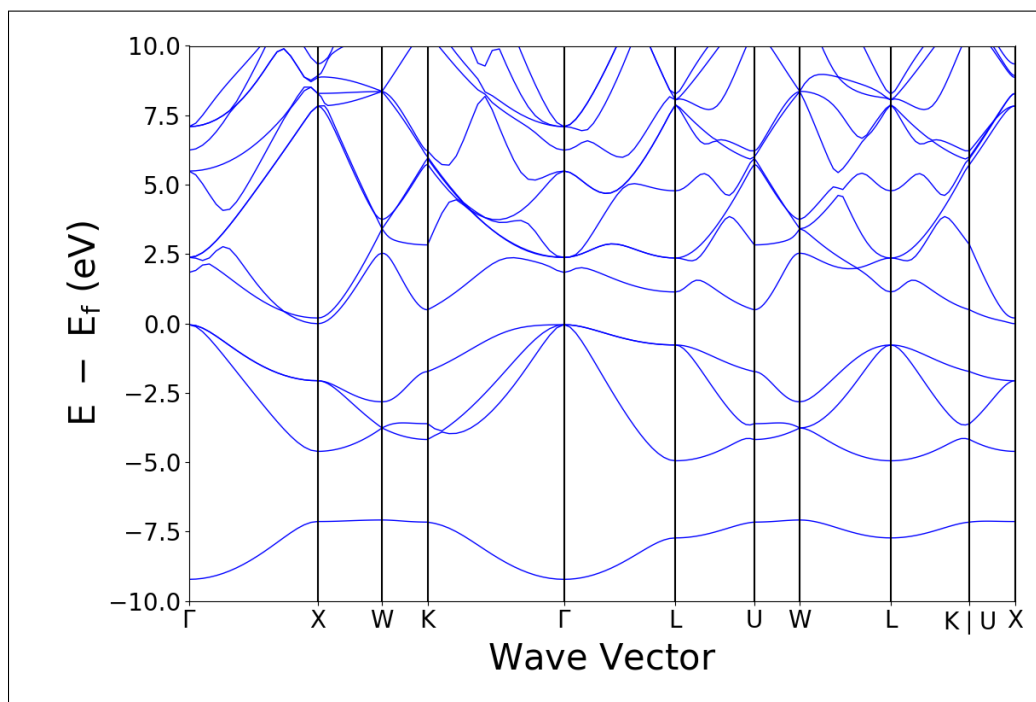
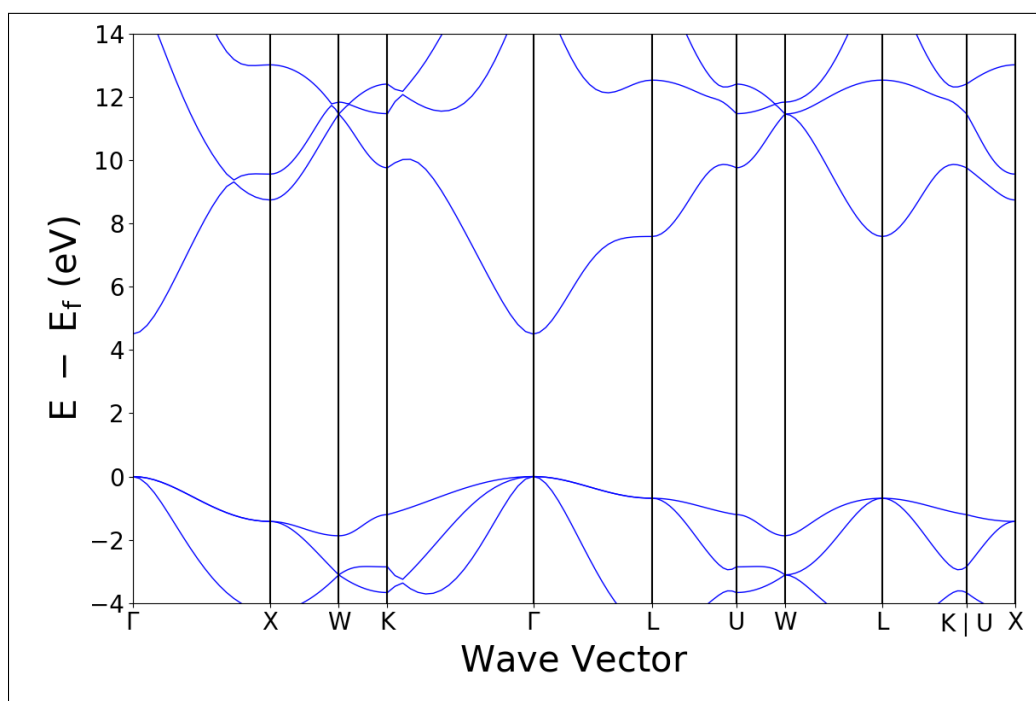


LiF [Fm $\bar{3}$ m]

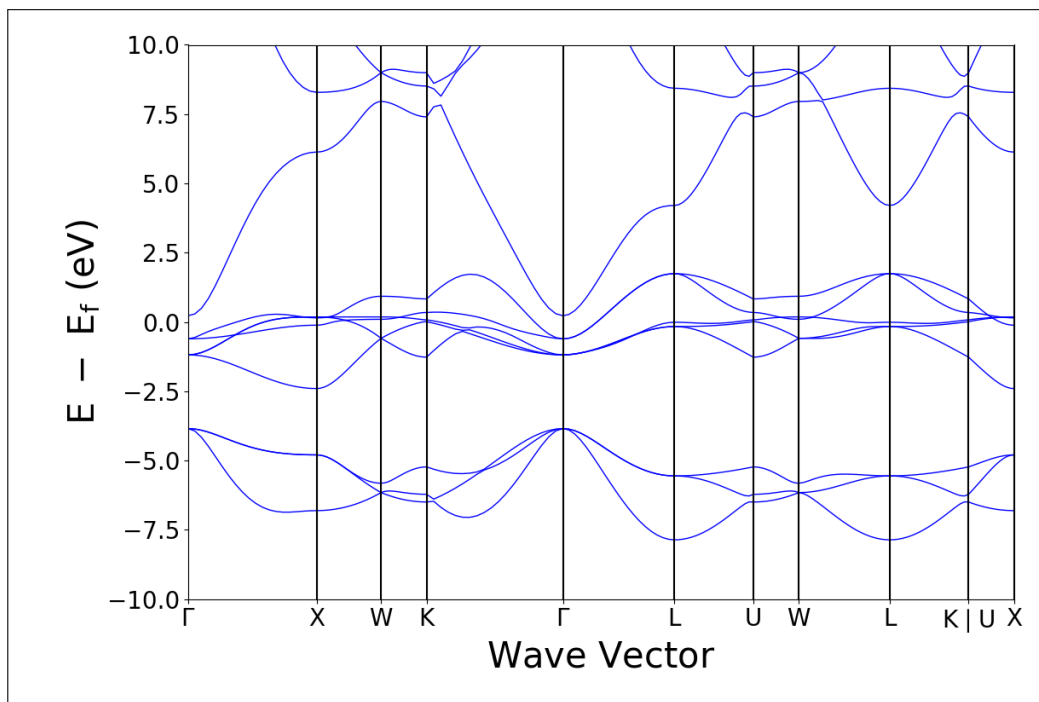


LiH [Fm $\bar{3}$ m]

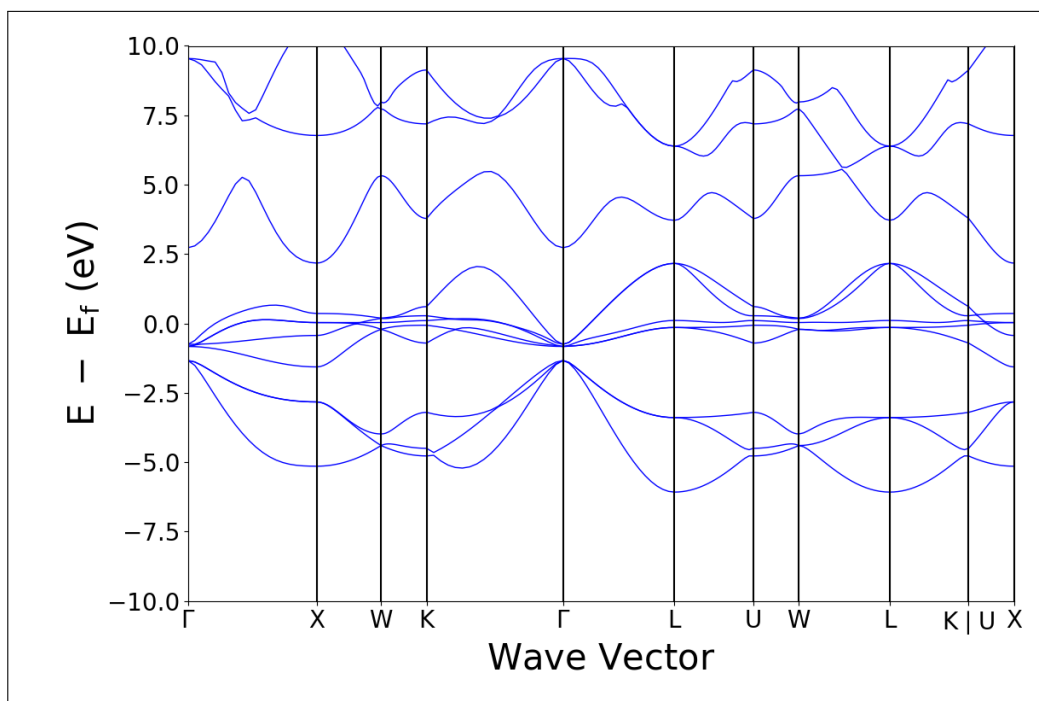


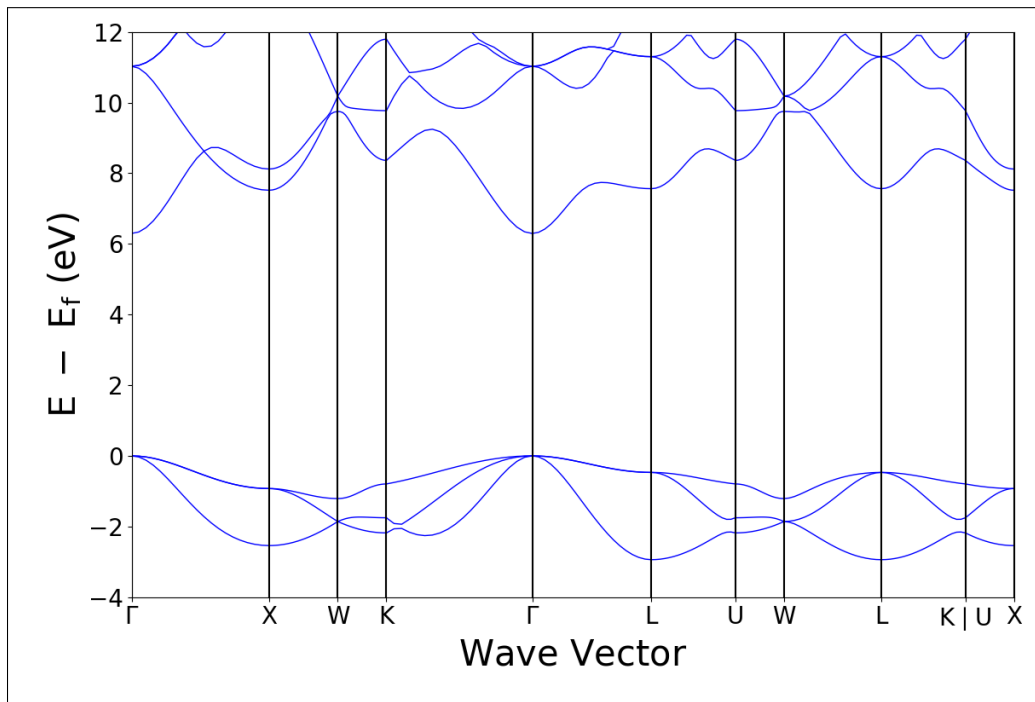
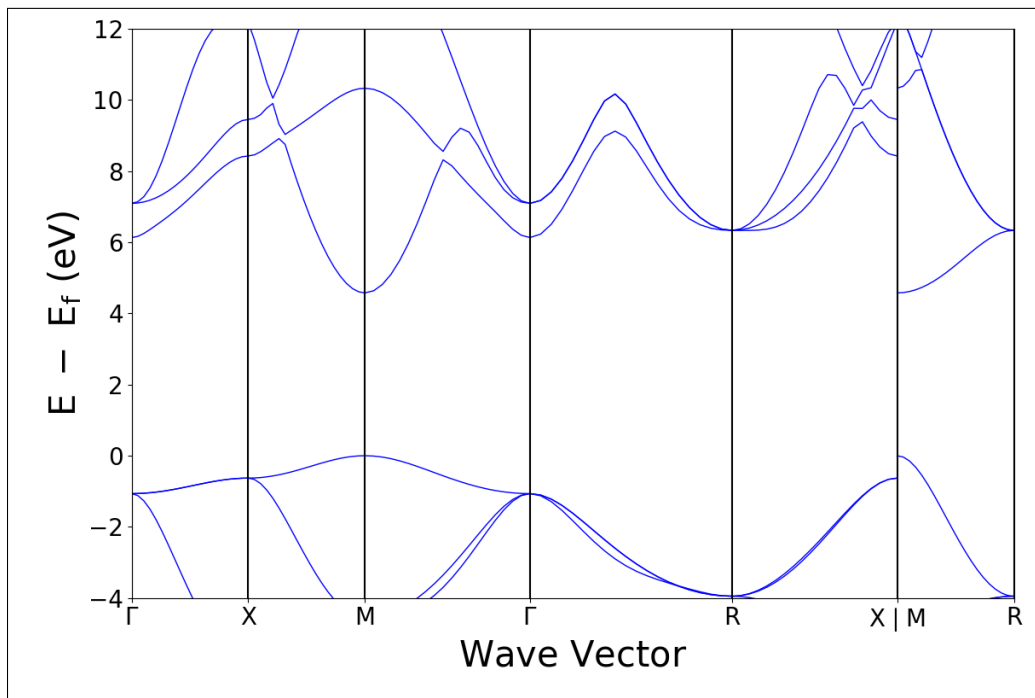
Mg₂Si [Fm $\bar{3}$ m]**MgO** [Fm $\bar{3}$ m]

MnO [$Fm\bar{3}m$]

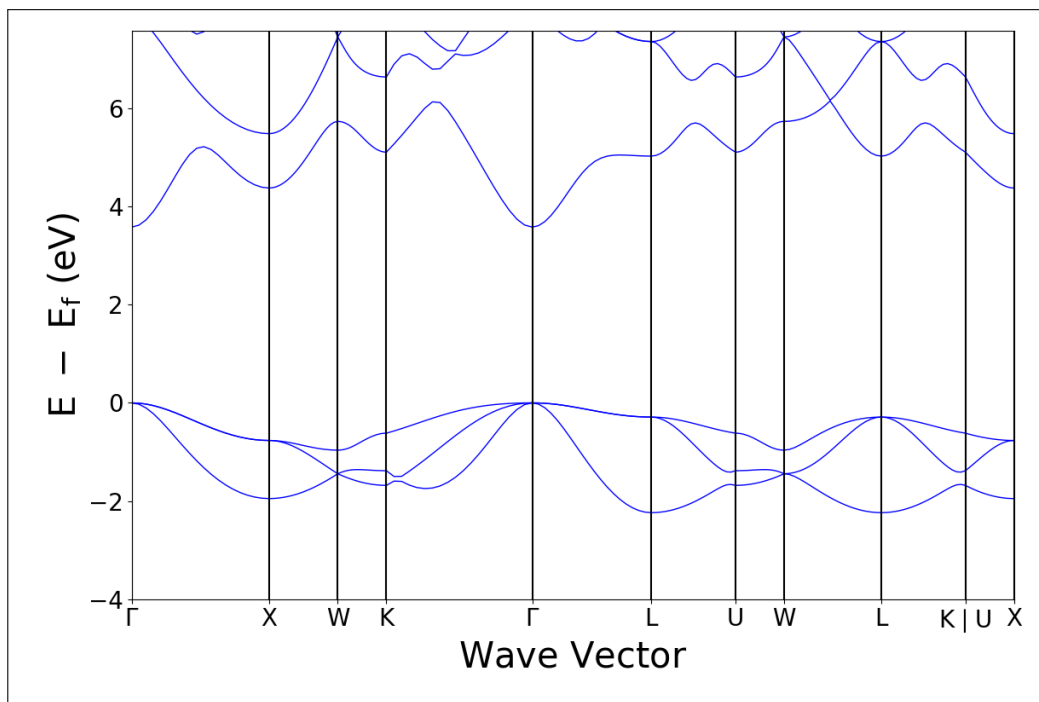


MnS [$Fm\bar{3}m$]

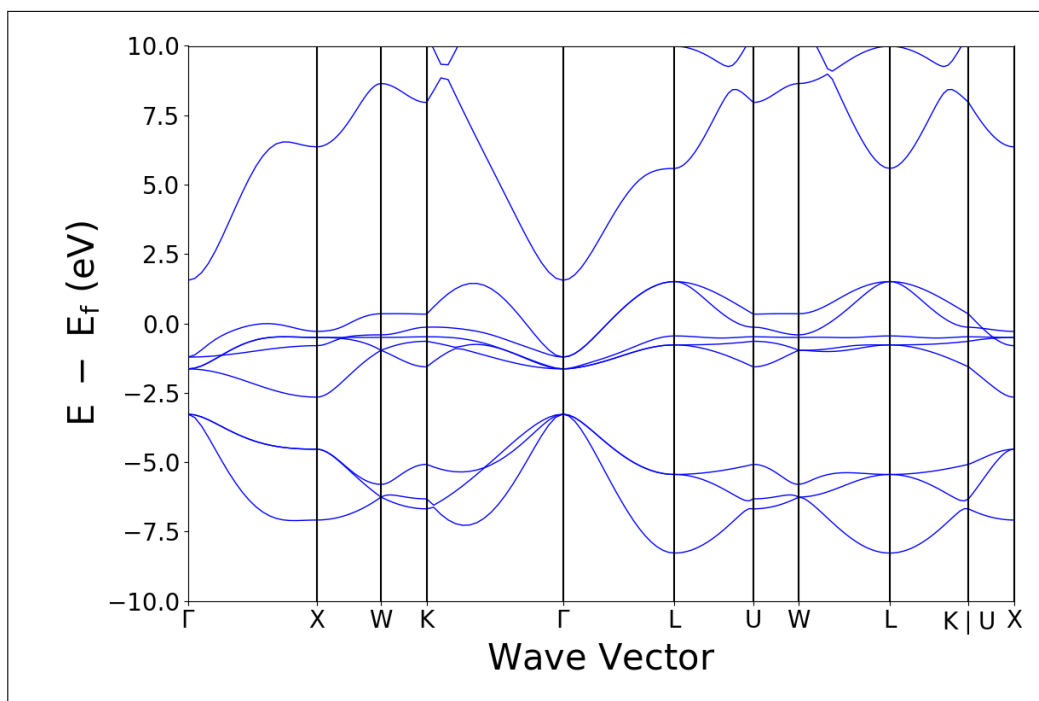


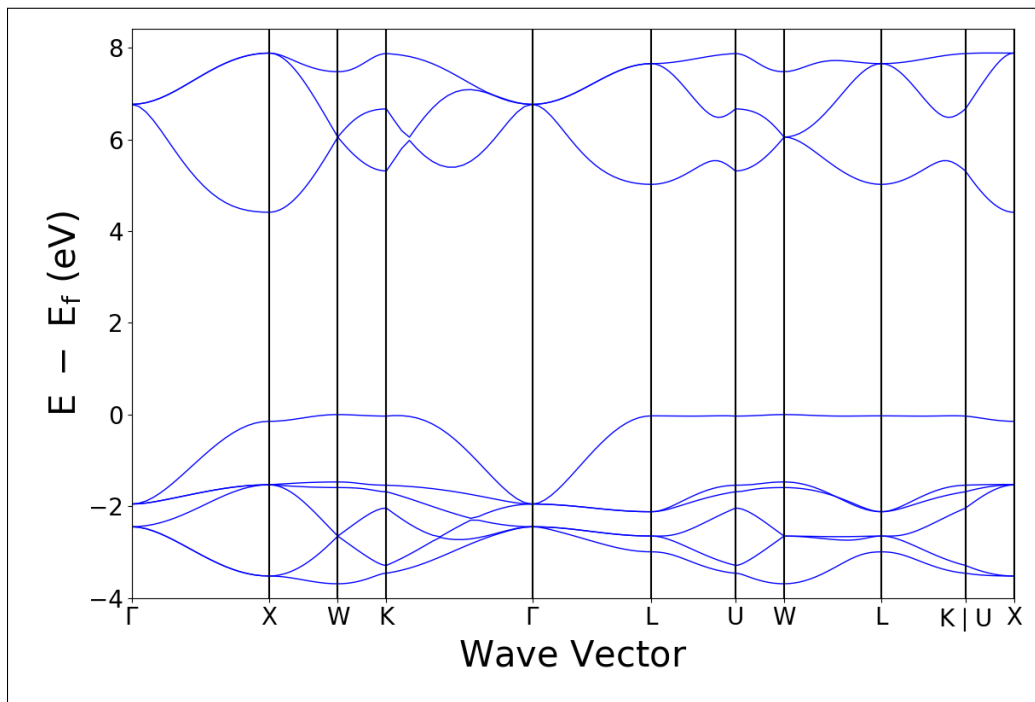
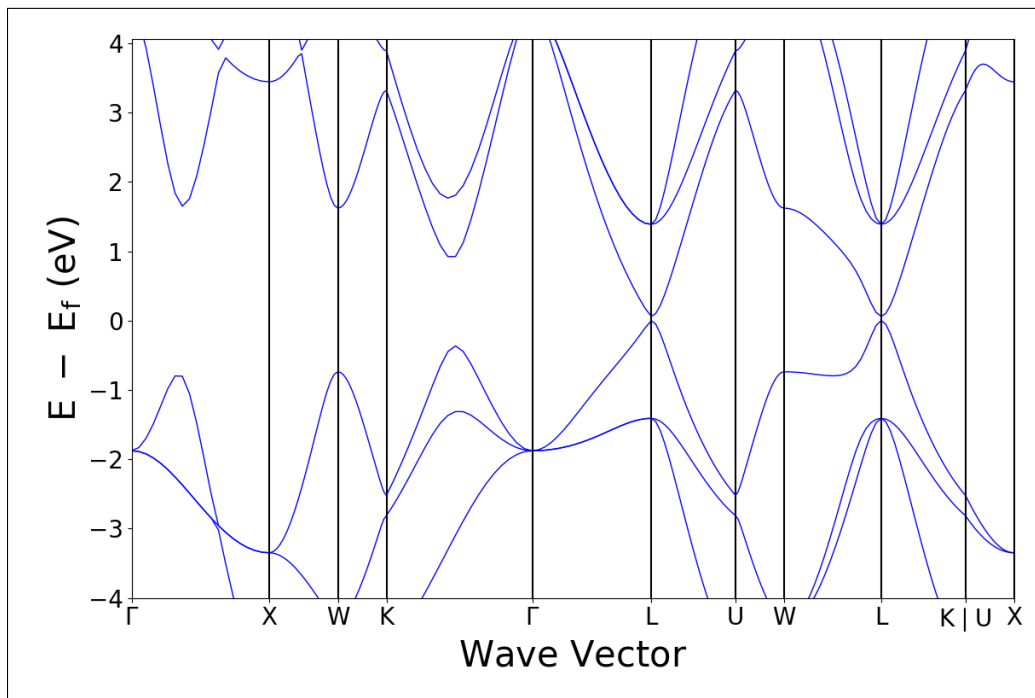
NaCl [$Fm\bar{3}m$]**NaCl** [$Pm\bar{3}m$]

NaI [Fm $\bar{3}$ m]

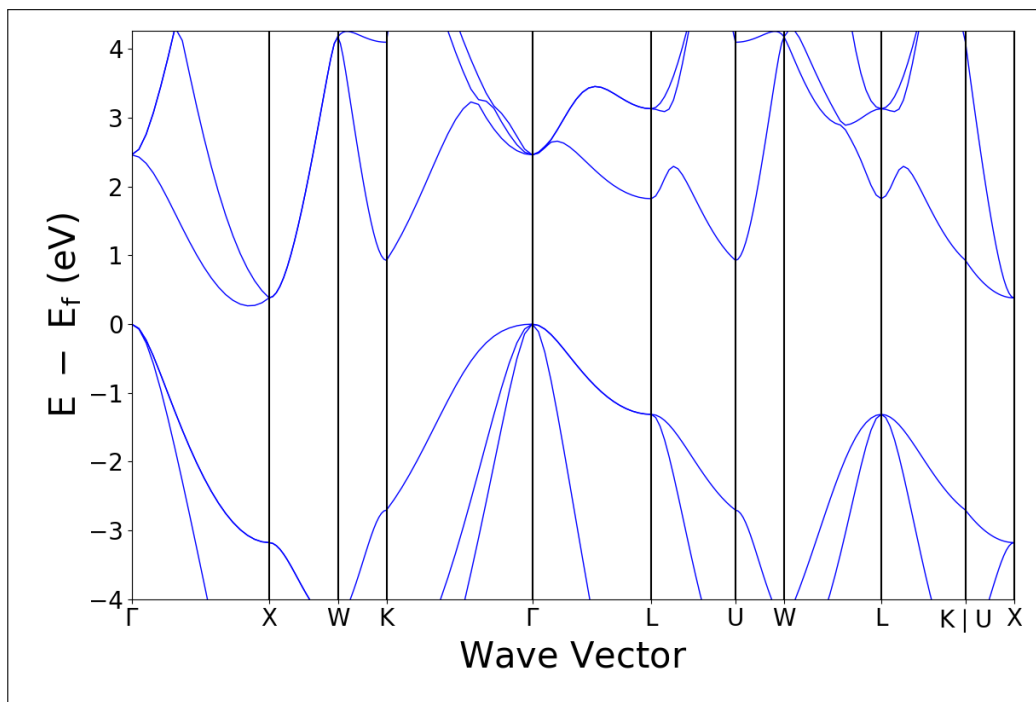


NiO [Fm $\bar{3}$ m]

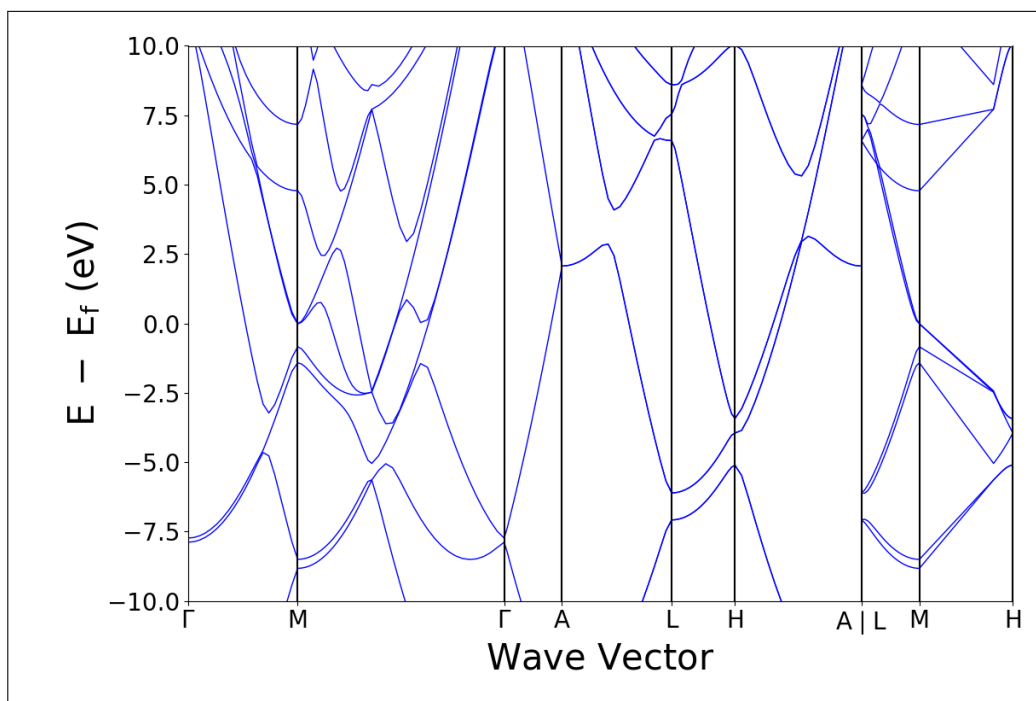


PbF₂ [Fm $\bar{3}$ m]**PbSe** [Fm $\bar{3}$ m]

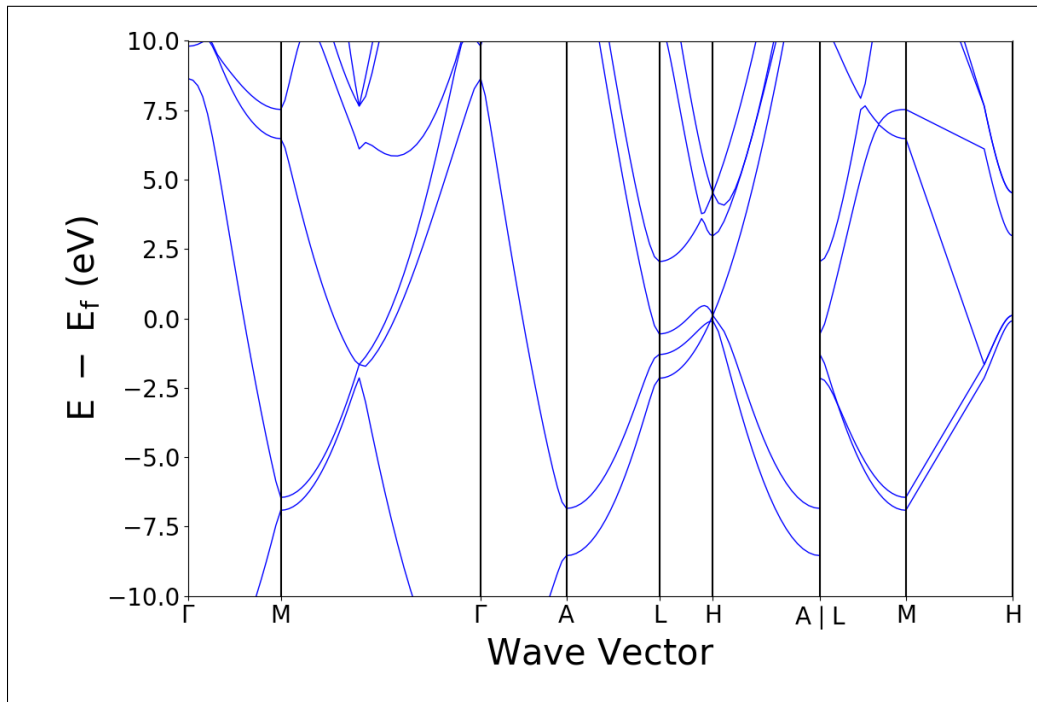
Si [Fd $\bar{3}$ m]



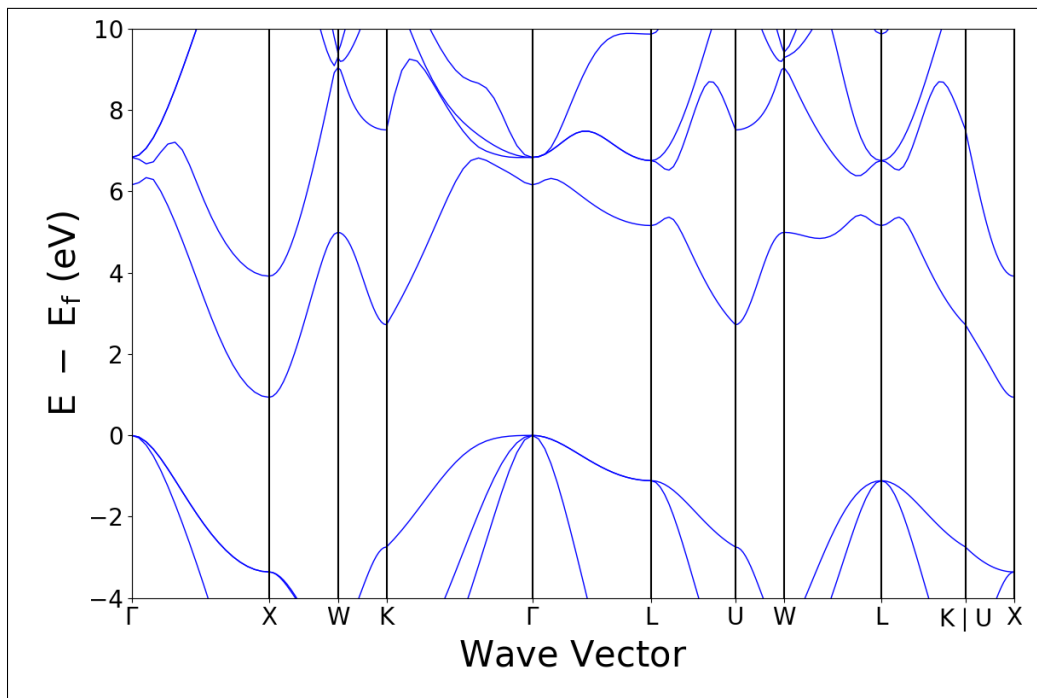
Si [P6 $_3$ /mmc]



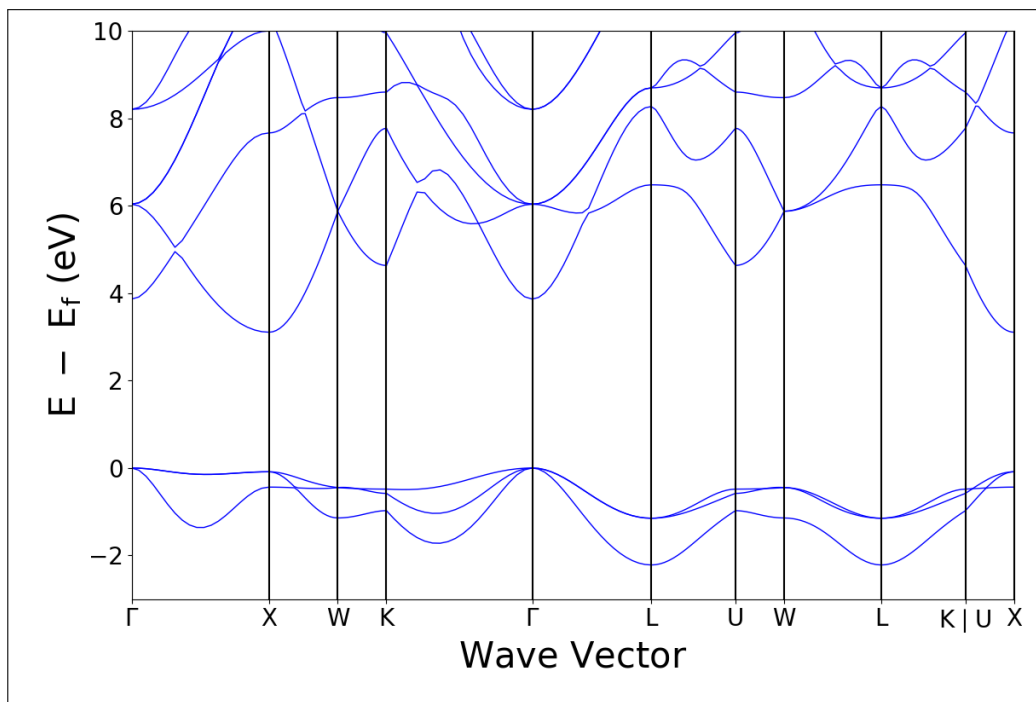
Si [P6/mmm]



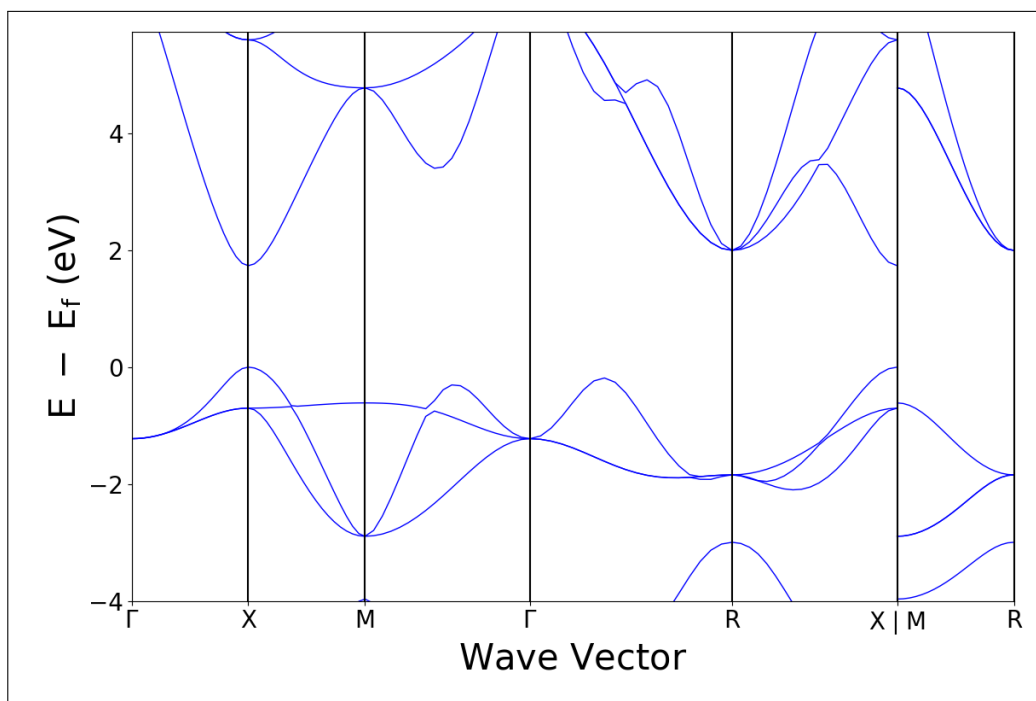
SiC [F43m]



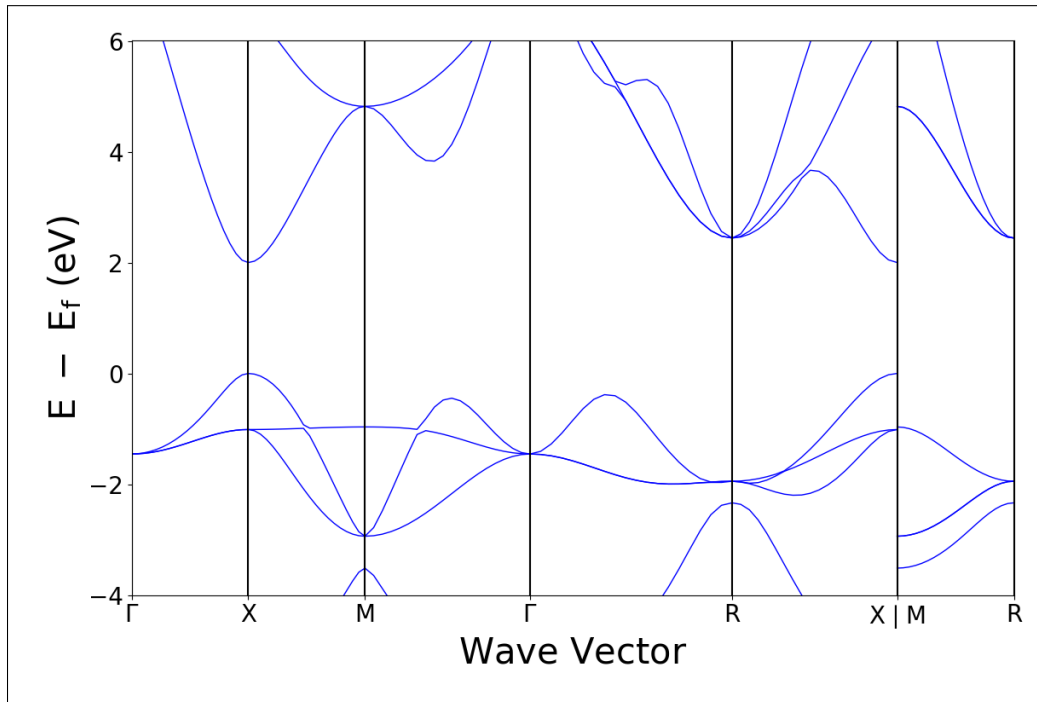
SrO [$Fm\bar{3}m$]



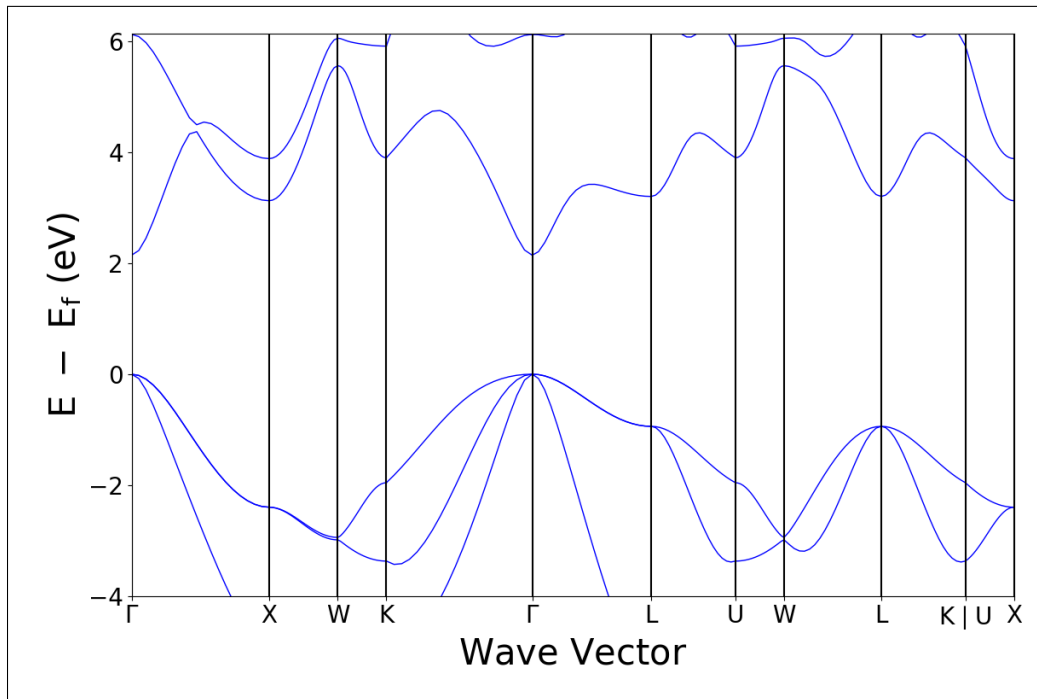
TlBr [$Pm\bar{3}m$]



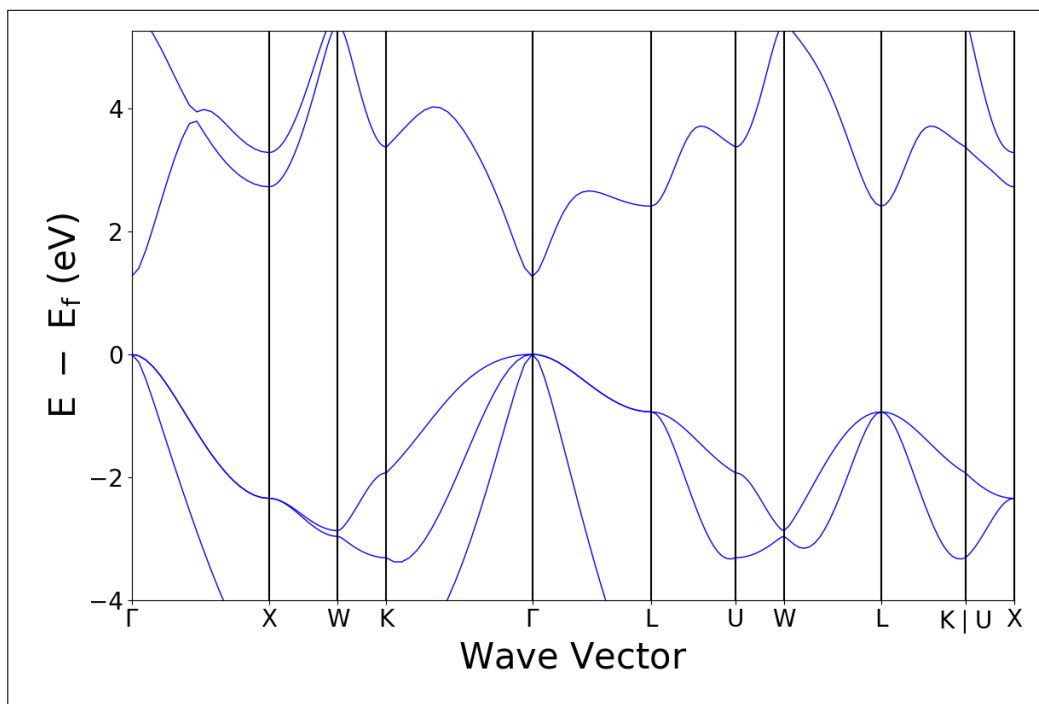
TlCl [$\text{Pm}\bar{3}\text{m}$]



ZnS [$\text{F}\bar{4}3\text{m}$]



ZnSe [F $\bar{4}3m$]



ZnTe [F $\bar{4}3m$]

