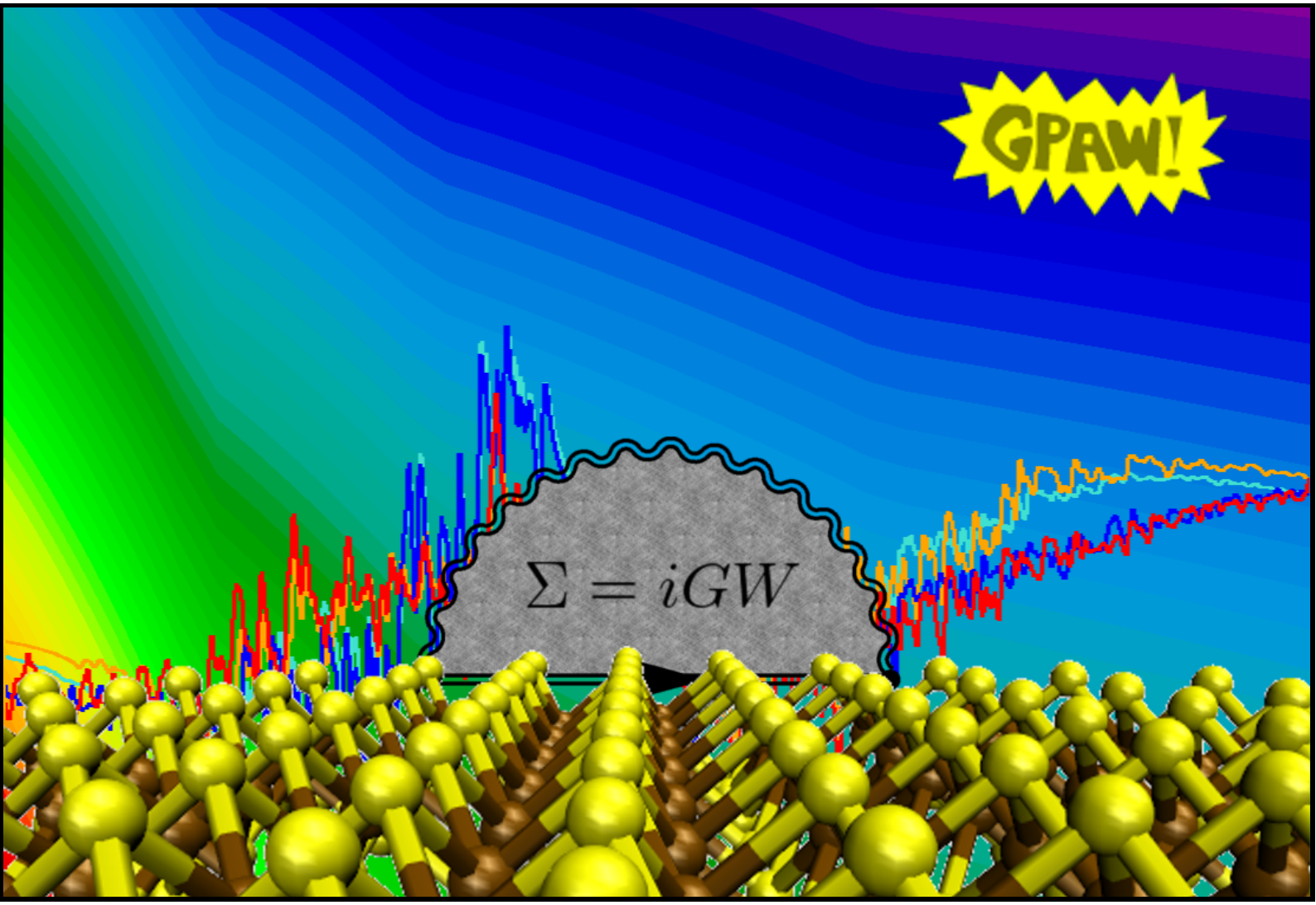
Quasiparticle GW calculations within the GPAW electronic structure code

Hüser, Falco Jonas; Thygesen, Kristian Sommer

Publication date:
2013

Document Version
Publisher's PDF, also known as Version of record

Quasiparticle GW calculations within the GPAW electronic structure code



Falco Jonas Hüser

Center for Atomic-scale Materials Design
Department of Physics
Technical University of Denmark

Quasiparticle GW calculations within the GPAW electronic structure code

Falco Jonas Hüser

Ph.D. Thesis
October 2013

Center for Atomic-scale Materials Design
Department of Physics
Technical University of Denmark

PREFACE

This thesis is submitted in candidacy for the Ph.D. degree from the Technical University of Denmark (DTU). It is based on the work carried out from November 2010 to October 2013 at the Center for Atomic-scale Materials Design (CAMD), Department of Physics, Technical University of Denmark, under the supervision of Professor Kristian S. Thygesen and Associate Professor Jakob Schiøtz.

Many people have helped me along the way and made this project possible.
First of all, I would like to thank:

Kristian Thygesen, for giving me the best support I could have wished for.
Jakob Schiøtz, for taking over as co-supervisor.
Thomas Bligaard, for bringing me here.
Jun Yan, for the good start.
Jens Jørgen Mortensen, for answering all my stupid questions.
Thomas Olsen, for the great collaborations.
Ivano Castelli, for doing the right things at the right times.

I would also like to thank all the great people, who have made these three years so wonderful:

My family, for everything. All the hardworking proofreaders: Marta, Rasmus, Simone, Filip, JJ, Ivano, Jón, Juanma and Thomas. Marianne, Stavroula and Helle, for their kind help. Ole, for taking care of Niflheim. Juanma, for sharing his knowledge. Jens Jørgen, for at tale dansk. Cheryl and Alan at Stanford. The fun office: Vladimir, Jakob, Marco and Nonni. The rest of the six fabulous five: Elisa P., Ana-Sofia, Ivano, Marta and Arnau. Carlo, Elisa B., Rasmus, Jón, Simone, Kirsten, George, Isabela, Tao, Fabio, David McC., Morten, Ask, Irek, Elisabetta, Gabriella and Jakob for all the good times. Marcel, Susanne, Rolf, Barbara, Christina, Teresa, Simon, Andrea, Jonathan and Johanna, for not forgetting me. Gemma Solomon, for what there is to come. Jonas.

ABSTRACT

The GPAW electronic structure code, developed at the physics department at the Technical University of Denmark, is used today by researchers all over the world to model the structural, electronic, optical and chemical properties of materials. They address fundamental questions in material science and use their knowledge to design new materials for a vast range of applications. Todays hottest topics are, amongst many others, better materials for energy conversion (e.g. solar cells), energy storage (batteries) and catalysts for the removal of environmentally dangerous exhausts.

The mentioned properties are to a large extent governed by the physics on the atomic scale, that means pure quantum mechanics. For many decades, Density Functional Theory has been the computational method of choice, since it provides a fairly easy and yet accurate way of determining electronic structures and related properties. However, it has several drawbacks. A conceptual problem is the difficulty of interpreting the calculated results with respect to experimentally measured quantities, resulting in, for example, the "band gap problem" in semiconductors. A practical issue is the necessity of adapting the method with respect to the system one wants to investigate by choosing a certain functional or by tuning parameters.

A succesful alternative is the so-called GW approximation. It is mathematically precise and gives a physically well-founded description of the complicated electron interactions in terms of screening. It provides a direct link to experimental observables through the concept of quasiparticles. Furthermore, it is parameter-free and thereby equally applicable to different kinds of systems. Its downside lies in its immense computational costs that limit its use in practice. Often, only the G_0W_0 approach is considered, which can be regarded as the lowest level of the GW approximation.

This thesis documents the implementation of the G_0W_0 approximation in GPAW. It serves two purposes: First, it can be read as a manual by anyone who is interested in doing GW calculations with GPAW. All features and requirements are explained in detail and many examples are given. This provides a full understanding of how the code works and how the outcome should be interpreted. Secondly, it gives an extensive discussion of calculated results for the electronic structure of 3-dimensional, 2-dimensional and finite systems and comparison with other implementations, methods and experiments. It shows that bandstructures, band gaps and ionization potentials can be obtained accurately with G_0W_0 for many different materials. But also exceptions are pointed out, where higher levels of the GW approximation might be necessary.

RESUMÉ

GPAW er et program, der bruges til at beregne elektroniske strukturer og er blevet udviklet på Institut for Fysik på Danmarks Tekniske Universitet. Det benyttes i dag af forskere i hele verden til at modellere materialers fysiske, elektroniske, optiske og kemiske egenskaber. Herved undersøges grundliggende problemstillinger indenfor materialvidenskab og resultaterne har en lang række anvendelser. Nogle af tidens mest spændende emner er at finde nye materialer til bæredygtig energiproduktion (f. eks. solceller), energiopbevaring (batterier) og katalysatorer, der nedbryder miljøskadelige udstødninger.

De ovennævnte egenskaber afgøres hovedsageligt af fysikken på den atomare skala, dvs. kvantemekanikken. I mange årtier har tæthedsfunktionalteori været den fortrukne beregningsmetode, for den er en forholdsvis enkel, men samtidig nøjagtig metode, til at bestemme den elektroniske struktur og de relaterede egenskaber. Men den viser sig også at have visse ulemper. Et konceptuelt problem er at relatere de beregnede resultater med eksperimentelle målinger. Det fører, f. eks. til det såkaldte båndgab-problem i halvledere. Et praktisk problem er, at man er nødt til at vælge et bestemt funktionale som er velegnet for systemet i undersøgelsen eller at tilpasse en parameter.

Den såkaldte GW approksimation tilbyder et lovende alternativ. Den er matematisk præcis og beskriver fysikken for den komplicerede elektron-vekselvirkning i den meningsfulde form af afskærmningseffekter. Beregninger kan knyttes direkte til eksperimentelle resultater ved hjælp af konceptet kvasi-partikler. Derudover er metoden fri for parametre og kan anvendes til mange forskellige slags systemer. Praktisk sætter den høje kompleksitet dog grænser. Ofte bruges kun G_0W_0, der kan betegnes som det laveste niveau af GW approksimationen.

Denne afhandling dokumenterer implementeringen af G_0W_0 approksimationen i GPAW. Den har to mål: For det første kan den læses som en brugermanual til dem, der selv vil lave GW beregninger med GPAW. Alle funktioner beskrives i detaljer ved hjælp af mange eksempler. På den måde forklares der grundigt om kodens drift og om hvordan resultaterne bør opfattes. For det andet diskuteres og vurderes beregninger af den elektroniske struktur af 3-dimensionale, 2-dimensionale og finite systemer ved at sammenligne med resultater fra andre implementeringer, metoder og eksperimenter. Båndstrukturer, båndgabs og ionisationspotentiale for mange forskellige materialer kan præcist bestemmes med G_0W_0. Men der vises også undtagelser, hvor der kan være brug for et højere niveau af GW approksimationen.

LIST OF INCLUDED PAPERS

Paper I

"Quasiparticle GW calculations for solids, molecules, and two-dimensional materials"
Falco Hüser, Thomas Olsen, and Kristian S. Thygesen
Physical Review B **87**, 235132 (2013).

Paper II

"Stability and bandgaps of layered perovskites for one- and two-photon water splitting"
Ivano E. Castelli, Juan María García-Lastra, Falco Hüser, Kristian S. Thygesen, and Karsten W. Jacobsen
New Journal of Physics **15**, 105026 (2013).

Paper III

"On the convergence of many-body excited state calculations for monolayer MoS_2"
Falco Hüser, Thomas Olsen, and Kristian S. Thygesen
accpeted for Physical Review B.

OTHER PAPERS

"Accurate and Efficient Bandgap Calculations for Screening Studies of New Materials"
Ivano E. Castelli, Falco Hüser, Mohnish Pandey, Anubhav Jain, Kristin Persson, Gebrand Ceder, Kristian S. Thygesen, and Karsten W. Jacobsen
to be submitted

CONTENTS

INTRODUCTION

Computational atomic-scale materials design is about the theoretical prediction of new materials that possess certain physical and chemical characteristics which enable their use for new technologies. Materials, that can possibly be used to build better and faster electronic devices, more powerful and longlasting batteries for electric cars, efficient solar cells or catalysts that reduce toxic gases in industrial exhausts, just to name a few.

The structures are simulated atom by atom on the computer and suitable theoretical models are used to calculate their properties. At the atomic scale, the physics is determined completely by quantum mechanics, where particles are diffuse objects with a limited probability of existing at a certain point in time and space, mathematically represented by wavefunctions. Interacting particles behave collectively and are thus attributed to many-body wavefunctions. The electronic structure of a system is entirely described by its Hamiltonian. It collects the operators for the kinetic energies of the electrons and the ionic cores, the Coulomb attraction and repulsion between all charged particles and interactions with external fields. The wavefunctions and the energy spectrum of the system are given by its eigenfunctions and eigenvalues, respectively. This is written down in compact form in the Schrödinger equation. However, it is in practice impossible to solve the Schrödinger equation directly, other than for the simplest model systems. Leaving out the ionic contributions (which can normally be separated), the complications arise from the electron-electron interactions.

Various approximations have been developed and successfully applied over the years. The simplest idea is to treat the electrons as if they were not interacting with each other. Then, the many-body wavefunction is just a product of single-particle wavefunctions, or, when taking Pauly's exclusion principle into account, a Slater determinant, which is fully antisymmetric under exchange of two particles. This is the idea behind Hartree-Fock theory [1], where the electron-electron interactions are reduced to a Hartree term, which describes the Coulomb repulsion of an electron with the total electron density, and an exchange term, which accounts for the antisymmetric nature of electrons. All that is left out here, is what is usually refered to as "correlation". In Density Functional Theory [2], the system of interacting electrons is mapped onto an auxiliary system of effectively non-interacting electrons under the requirement that the electron density, which is as opposed to the wavefunction an observable, of these two systems are identical. In this scheme, wavefunctions and energies are given as eigenfunctions

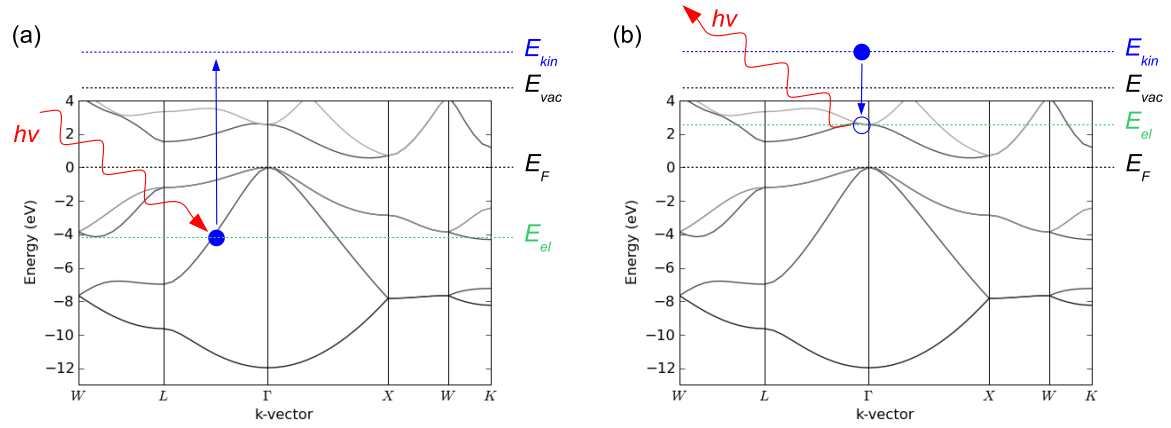


Figure 1: Schematic picture of (a) Photo Electron Spectroscopy (PES) and (b) Inverse Photo Electron Spectroscopy (IPS). In (a), an electron is removed from the system by absorption of a photon with energy $h\nu$. In (b), an electron is added to the system under emission of a photon. The binding energy (relative to the vacuum level, E_{vac}) is then given as $E_{\mathrm{el}} = E_{\mathrm{kin}}$ - $h\nu$, where E_{kin} is the kinetic energy of the free electron. E_{el} corresponds to the energy of a state in a valence and conduction band, respectively. E_F is the Fermi level.

and -values of an effective Kohn-Sham Hamiltonian [3]. However, complications do not vanish, but are only transferred into an exchange-correlation functional. Development of elaborate functionals has been work in progress for many decades and great results have been achieved. Today, Density Functional Theory is one of the most common methods for calculating electronic structures. Still, a number of problems cannot be overcome: A practical issue is that there is a variety of different functionals to choose from and many of them are designed for special purposes, often by tuning parameters to fit experimental data. In this sense, it is not a 100 % *ab-initio* method. More fundamentally, it is a groundstate theory. That means that in principle, total energies can be calculated exactly, whereas the Kohn-Sham wavefunctions and eigenvalues lack a meaningful physical interpretation. Physicists usually think of an electronic structure as a series of bands, which are being filled up by a certain number of electrons, N. The occupied (valence) and unoccupied (conduction) states are separated by the Fermi level, E_F. In experiment, the energies of the valence and conduction bands are typically measured by Photo Electron and Inverse Photo Electron Spectroscopy, respectively, as sketched in Fig. 1. These processes include the removal or addition of one electron and are thus not properties of the N-electron groundstate. The fundamental energy gap is defined as the difference in the lowest electron addition and removal energies, $E_{\mathrm{el}}^{\pm}$, and can be written as:

$$E_{\mathrm{gap}} = E_{\mathrm{el}}^{+} - E_{\mathrm{el}}^{-} = E_0^{N+1} + E_0^{N-1} - 2E_0^{N}, \tag{1}$$

where E_0^N and $E_0^{N\pm1}$ are the groundstate energies of the system with N and $N \pm 1$ electrons, respectively. Eq. (1) allows in principle for determining the gap from three groundstate calculations for the neutral and single positive and negative charged system. However, this cannot be done for periodic systems like semiconductors. In terms of (exact) Kohn-Sham energies, the gap is given by:

$$E_{\mathrm{gap}} = \varepsilon_{N+1}^{\mathrm{KS}}(N) - \varepsilon_{N}^{\mathrm{KS}}(N) + \Delta_{xc} = E_{\mathrm{gap}}^{\mathrm{KS}} + \Delta_{xc}, \tag{2}$$

where $\varepsilon_i^{\mathrm{KS}}(N)$ is the i-th eigenvalue of the N-electron system and Δ_{xc} the derivative discontinuity [4], which in practice can only be estimated.

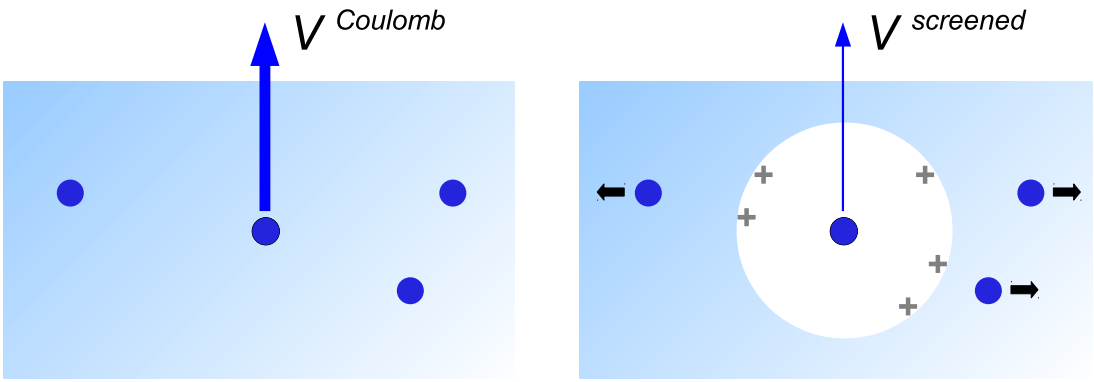


Figure 2: Schematic picture of the concept of screening. On the left side, electrons are non-interacting and the potential seen outside emerging from one electron is the full Coulomb potential. Interacting electrons tend to repel each other. As sketched on the right side, this leads to the formation of an effective positively charged cloud surrounding each electron. This Coulomb hole screens the Coulomb potential.

An alternative approach that avoids these problems is established by many-body perturbation theory. Instead of trying to solve the Schrödinger equation for the wavefunctions, one wants to determine the Green's Function, which describes the propagation of a particle (electron or hole) through the groundstate of the system. This corresponds exactly to the situations depicted in Fig. 1. These and similar excitations are "quasi" single-particle-like and called quasiparticles. The important difference to a real single-particle excitation is that the full response of all particles in the system to that excitation is included in the quasiparticle itself, e.g. all correlation effects. Quasiparticles have in general finite lifetimes and their spectrum, $\{\varepsilon^{\mathrm{QP}}_{i\pm}\}$, is given by the poles of the Green's Function. The fundamental gap is then simply:

$$E_{\mathrm{gap}} = \varepsilon^{\mathrm{QP}}_{0+} - \varepsilon^{\mathrm{QP}}_{0-}. \tag{3}$$

Hence, $\varepsilon^{\mathrm{QP}}_{0\pm}$ are equal to the lowest electron removal and addition energies, $E^{\pm}_{\mathrm{el}}$, introduced above.

The Green's Function can be defined through an equation of motion, which contains the self-energy operator, a non-local and energy-dependent analogous of the exchange-correlation potential. Unfortunately, it is practically impossible to calculate it exactly. However, it is straightforward to expand it systematically using perturbation theory. In the GW approximation [5], the self-energy is taken to first order in the screened potential. This seems rather crude at first sight, but turns out to give an excellent description of weak correlation. The basic idea of screening is illustrated in Fig. 2: When an electron is added to the system, it polarizes its surrounding and thereby induces its own Coulomb hole, which reduces the potential.

This thesis presents GW calculations for solids, molecules and two-dimensional materials. Quasiparticle energies were obtained with first-order perturbation theory from Kohn-Sham wavefunctions and eigenvalues in the so-called G_0W_0 or non-selfconsistent GW approximation. It is organized as follows:

- Chapter 1 introduces the theory of Green's Functions and sets the mathematical framework for the GW approximation.

- Chapter 2 contains all computational details of the implementation developed in this project as well as extensive convergence tests.

- Chapter 3 presents calculations for semiconductors, insulators and metals. Bandstructures and band gaps are compared to results from literature, other methods and experiments, where available.
- Chapter 4 discusses special issues that occur for two-dimensional systems for the examples of single sheets of graphene/hexagonal boron-nitride and MoS_2.
- Chapter 5 focusses on the Ionization Potentials of different molecules and gives insight into the structure of the self-energy.

Each chapter is preceded by a seperate introduction and can be read to a large extent independently from the rest of the thesis.

CHAPTER 1

THEORETICAL BACKGROUND

This chapter gives a brief introduction to the underlying theory of the GW approximation as pointed out by Hedin in 1965 [5]. The theoretical description is given in the framework of many-body perturbation theory (MBPT) in which the central quantities are the Green's Function G and the self-energy Σ. In principal, they contain all information on a given system, similar to the Hamiltonian and wavefunctions as defined by the Schrödinger equation. As opposed to groundstate theories like density functional theory (DFT), MBPT inherently offers ways to calculate excited state properties and the corresponding wavefunctions can be interpreted as quasiparticle (QP) states. This allows for the calculation of the fundamental band gap, for example, which is defined by the difference of electron removal and addition energies. Since all other electrons in the system will respond to that additional electron or hole, the gap is clearly not a groundstate property. These complicated interactions are known as correlation effects or screening. A further major problem of most DFT functionals is the self-interaction error [6] which arises from incomplete cancellation of the interaction of an electron with itself in the Hartree and exchange-correlation terms.

In MBPT, the Green's Function is a solution to the equation of motion in which the self-energy appears as a non-Hermitian, nonlocal and frequency-dependent operator. Determining Σ is therefore the key to finding the electronic structure.

As will be shown in Section 1.3, the self-energy can in principal be evaluated exactly through a set of four coupled integro-differential equations, known as Hedin's equations. However, this turns out to be impossible to do in practice, even for simple systems. In fact, it is as complicated as solving the Schrödinger equation directly (or as finding the one true exchange-correlation functional in DFT) and therefore, approximations need to be made. It is, however, possible to write down systematic expansions of the self-energy and various approaches exist, depending on different aspects of the underlying physics. Feynman diagrams provide an easy and instructive way of interpreting and calculating the different terms, and some of them can be summed up to infinite order. Exxpanding the self-energy to first order in the screened potential W reads simply: $\Sigma = iGW$. This turns out to give an astonishingly good description of the physics of weakly-correlated materials and has become the highly successful GW approxima-

tion.

A mathematical rigorous introduction to quantum-mechanical Green's Functions and Feynman diagrams can be found in Ref. [7], whereas the GW approximation is discussed in detail in the reviews [8–10].

Throughout this chapter, spin indices are suppressed in order to simplify the notation. The extension to spin-dependent quantities is straightforward. Atomic units ($\hbar = m_e = e = 1$) are used.

1.1 Green's Functions

In second quantization, the time-ordered single-particle Green's Function reads:

$$G(\mathbf{r},t;\mathbf{r}',t') = \left\langle \mathcal{T}\left\{\hat{\psi}(\mathbf{r},t)\hat{\psi}^\dagger(\mathbf{r}',t')\right\}\right\rangle, \tag{1.1}$$

where $\hat{\psi}$ and $\hat{\psi}^\dagger$ denote fermionic annihilation and creation operators, respectively, which fulfill the anticommutator relations:

$$\left\{\hat{\psi}(\mathbf{r},t),\hat{\psi}^\dagger(\mathbf{r}',t')\right\}_{t=t'} = \delta(\mathbf{r}-\mathbf{r}'). \tag{1.2}$$

The expectation value $\langle \dots \rangle$ is to be taken with respect to the N-particle groundstate of the system. $\mathcal{T}$ is the time-ordering operator which ensures that the field operators on which it acts are ordered in ascending time argument t from right to left:

$$\mathcal{T}\left\{\hat{\psi}(\mathbf{r},t)\hat{\psi}^\dagger(\mathbf{r}',t')\right\} = \begin{cases} \hat{\psi}(\mathbf{r},t)\hat{\psi}^\dagger(\mathbf{r}',t') & \text{for } t > t' \\ -\hat{\psi}^\dagger(\mathbf{r}',t')\hat{\psi}(\mathbf{r},t) & \text{for } t < t'. \end{cases} \tag{1.3}$$

With this, the physical interpretation of the Green's Function (1.1) becomes clear: It describes the propagation of an electron created at space coordinate $\mathbf{r}'$ and time t' and annihilated at another point in space, $\mathbf{r}$, and a later time t through the groundstate of the system. The opposite holds for a hole. Note that even though the Green's Function (1.1) describes the propagation of a single particle, the full information of the interacting N-electron system is contained through the expectation value.

For a system of electrons which interact via the Coulomb potential $V(\mathbf{r},\mathbf{r}') = \frac{1}{|\mathbf{r}-\mathbf{r}'|}$, e.g. where the Hamiltonian takes the form:

$$\hat{H} = \int d\mathbf{r}\, \hat{\psi}^\dagger(\mathbf{r},t)h_0(\mathbf{r})\hat{\psi}(\mathbf{r},t) + \frac{1}{2}\int d\mathbf{r}\int d\mathbf{r}'\, \hat{\psi}^\dagger(\mathbf{r},t)\hat{\psi}^\dagger(\mathbf{r}',t')V(\mathbf{r},\mathbf{r}')\hat{\psi}(\mathbf{r}',t')\hat{\psi}(\mathbf{r},t), \tag{1.4}$$

the evolution of the Green's Function is governed by an equation of motion: [1]

$$(\omega - h_0(\mathbf{r}) - V_H(\mathbf{r}))\, G(\mathbf{r},\mathbf{r}';\omega) - \int d\mathbf{r}''\Sigma(\mathbf{r},\mathbf{r}'';\omega)G(\mathbf{r}'',\mathbf{r}';\omega) = \delta(\mathbf{r}-\mathbf{r}'), \tag{1.5}$$

where h_0 collects all one-body terms such as the kinetic energy and interaction with an external potential. $V_H(\mathbf{r}) = \int d\mathbf{r}' V(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')$ is the Hartree potential with the electron density given by the diagonal of the Green's Function: $\rho(\mathbf{r}) = G(\mathbf{r}t;\mathbf{r}t^+)$. [2] In frequency domain, the Green's Function reads: $G(\mathbf{r},\mathbf{r}';\omega) = \int d(t-t')\, G(\mathbf{r},t;\mathbf{r}',t')\exp(i\omega(t-t'))$. Σ is called the self-energy and is in general a dynamic, non-local and non-Hermitian operator.

The Green's Function can be expressed in the spectral representation, also known as Lehmann representation:

$$G(\mathbf{r},\mathbf{r}';\omega) = \int d\omega' \, \frac{A(\mathbf{r},\mathbf{r}';\omega)}{\omega - \omega' + i\eta\,\mathrm{sgn}(\omega')}, \tag{1.6}$$

with an infinitesimally small, positive η, which ensures that G is analytic along the real axis. The spectral function is linked to the imaginary part of the Green's Function as:

$$A(\mathbf{r},\mathbf{r}';\omega) = -\frac{1}{\pi}\mathrm{sgn}(\omega)\mathrm{Im}\left\{G(\mathbf{r},\mathbf{r}';\omega)\right\}. \tag{1.7}$$

[1] This can be derived with the use of $\frac{d}{dt}\hat{A}(t) = i[\hat{H},\hat{A}(t)] + \partial_t\hat{A}(t)$ for any quantum-mechanical operator $\hat{A}(t)$ and representing the field operators in the Heisenberg picture: $\hat{\psi}(\mathbf{r},t) = e^{i\hat{H}t}\hat{\psi}(\mathbf{r})e^{-i\hat{H}t}$.

[2] with $t^+ \equiv t + \delta$ for positive $\delta \to 0$

For a system of non-interacting electrons Eq. (1.5) reduces to:

$$\Big(\varepsilon_j - h_0(\mathbf{r})\Big) G_0(\mathbf{r}, \mathbf{r}'; \omega) = \delta(\mathbf{r} - \mathbf{r}'), \tag{1.8}$$

and the non-interacting Green's Function takes the simple form:

$$G_0(\mathbf{r}, \mathbf{r}'; \omega) = \sum_j \frac{\phi_j(\mathbf{r})^* \phi_j(\mathbf{r}')}{\omega - \varepsilon_j + i\eta \mathrm{sgn}(\varepsilon_j - \mu)}, \tag{1.9}$$

where $\{\phi_i\}$ are single-particle wavefunctions. The corresponding spectral function is a sum of δ-functions at the orbital energies $\omega = \varepsilon_i$:

$$A_0(\mathbf{r}, \mathbf{r}'; \omega) = \sum_j \phi_j(\mathbf{r})^* \phi_j(\mathbf{r}') \delta(\omega - \varepsilon_j). \tag{1.10}$$

The interacting Green's Function (also called "full" or "dressed" Green's Function) is connected to G_0 through Dyson's equation:

$$G(\mathbf{r}, \mathbf{r}'; \omega) = G_0(\mathbf{r}, \mathbf{r}'; \omega) + \int d\mathbf{r}_1 \int d\mathbf{r}_2 \, G(\mathbf{r}, \mathbf{r}_1; \omega) \Sigma(\mathbf{r}_1, \mathbf{r}_2; \omega) G_0(\mathbf{r}_2, \mathbf{r}'; \omega). \tag{1.11}$$

In a simplified notation, this reads $G = G_0 + G\Sigma G_0$ and is illustrated graphically in Fig. 1.1 using standard Feynman diagrams. It can also be used as a definition for the self-energy: $\Sigma = G^{-1} - G_0^{-1}$. Similarly, the non-interacting and the full Green's Function can symbolically be written as $G_0(z) = (z - \hat{h}_0 - V_H)^{-1}$ and $G(z) = (z - \hat{H})^{-1}$, respectively, where $\hat{H} = \hat{h}_0 + V_H + \Sigma$ and z is a complex number. In this definition, the self-energy collects all electron-electron interactions that go beyond the Hartree level, that means all exchange and correlation contributions. Therefore $\Sigma = \Sigma_{xc}$ can be regarded as a non-local and energy-dependent analogous of the exchange-correlation potential in DFT.

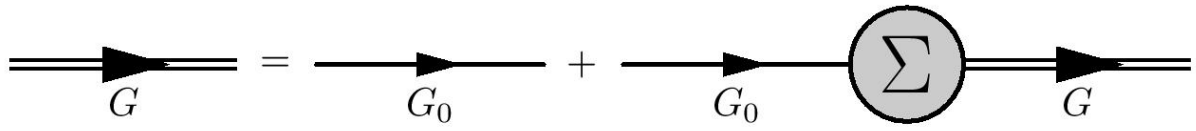


Figure 1.1: Schematic representation of the Dyson equation (1.11). By iteratively inserting the same definition for G on the right-hand side, it becomes an infinite expansion in powers of the self-energy: $G = G_0 + G_0\Sigma G_0 + G_0\Sigma G_0\Sigma G_0 + G_0\Sigma G_0\Sigma G_0\Sigma G_0 + \ldots$.

1.2 Quasiparticles

An alternative way of writing Eq. (1.6) is to expand G in the full complex plane in a set of complete basis functions $\{\psi_i^{\mathrm{QP}}\}$: [3]

$$G(\mathbf{r},\mathbf{r}';z) = \sum_j \frac{\psi_j^{\mathrm{QP}}(\mathbf{r})^*\psi_j^{\mathrm{QP}}(\mathbf{r}')}{z-\varepsilon_j^{\mathrm{QP}}}. \tag{1.12}$$

In the discrete part of the spectrum, $\{\varepsilon_i^{\mathrm{QP}}\}$ are solutions to the quasiparticle equation:

$$\left(\varepsilon_i^{\mathrm{QP}} - h_0(\mathbf{r}) - V_H(\mathbf{r})\right)\psi_i^{\mathrm{QP}}(\mathbf{r}) - \int d\mathbf{r}'\Sigma(\mathbf{r},\mathbf{r}';\varepsilon_i^{\mathrm{QP}})\psi_i^{\mathrm{QP}}(\mathbf{r}') = 0. \tag{1.13}$$

These functions are the quasiparticle states and the energies correspond to excitation energies:

$$\begin{aligned} \psi_{i-}^{\mathrm{QP}}(\mathbf{r}) &= \left\langle N-1, i \middle| \hat{\psi}(\mathbf{r}) \middle| N, 0 \right\rangle \quad \text{and} \quad \varepsilon_{i-}^{\mathrm{QP}} = E_0^N - E_i^{N-1} \quad \text{when} \quad \varepsilon_i^{\mathrm{QP}} < \mu \\ \psi_{i+}^{\mathrm{QP}}(\mathbf{r}) &= \left\langle N, 0 \middle| \hat{\psi}(\mathbf{r}) \middle| N+1, i \right\rangle \quad \text{and} \quad \varepsilon_{i+}^{\mathrm{QP}} = E_i^{N+1} - E_i^N \quad \text{when} \quad \varepsilon_i^{\mathrm{QP}} \geq \mu \quad , \end{aligned} \tag{1.14}$$

where $|N,0\rangle$ stands for the groundstate of the N-particle system and $|N\pm1,i\rangle$ for the i-th excited state of the $N\pm1$-particle system. Accordingly, E_0^N and $E_i^{N\pm1}$ are the total energies. $\mu = E_0^{N+1} - E_0^N$ is the chemical potential. From Eq. (1.14), it becomes clear that the quasiparticle states describe the removal or addition of an electron and the corresponding energies are electron removal and addition energies. For $i = 0$, they are equal to the negative ionization potential (IP) and electron affinity (EA), respectively. Thus, the fundamental band gap is given as:

$$E_{\mathrm{gap}} = IP - EA = \varepsilon_{0+}^{\mathrm{QP}} - \varepsilon_{0-}^{\mathrm{QP}} = E_0^{N+1} + E_0^{N-1} - 2E_0^N. \tag{1.15}$$

In principal, the QP energies and wavefunctions are not equal to the eigenvalues and eigenfunctions defined by Eq. (1.5):

$$(\varepsilon_n(\omega) - h_0(\mathbf{r}) - V_H(\mathbf{r}))\,\psi_n(\mathbf{r},\omega) - \int d\mathbf{r}'\Sigma(\mathbf{r},\mathbf{r}';\omega)\psi_n(\mathbf{r}',\omega) = 0, \tag{1.16}$$

with which the Green's Function can be expressed as:

$$G(\mathbf{r},\mathbf{r}';\omega) = \sum_m \frac{\psi_m(\mathbf{r},\omega)^*\psi_m(\mathbf{r}',\omega)}{\omega-\varepsilon_m(\omega)}. \tag{1.17}$$

These eigenvalues are in general complex and frequency-dependent and the eigenfunctions are non-orthogonal. However, for $\omega_i = \mathrm{Re}\,\{\varepsilon_n(\omega_i)\} = \varepsilon_i^{\mathrm{QP}}$, the eigenvector $\psi_n(\mathbf{r},\omega_i)$ coincides with the QP wavefunction $\psi_i^{\mathrm{QP}}(\mathbf{r})$ (except for normalization) and is denoted $\psi_i(\varepsilon_i^{\mathrm{QP}}) = \psi_i^{\mathrm{QP}}/\|\psi_i^{\mathrm{QP}}\|^2$. If the imaginary part of $\varepsilon_n(\omega_i)$ is small, the spectrum shows a peak at the quasiparticle energy. Its broadening is related to the lifetime of the quasiparticle. In other cases, where $(\omega - \mathrm{Re}\,\{\varepsilon_n(\omega)\})$ and $\mathrm{Im}\,\{\varepsilon_n(\omega)\}$ are small, so-called satellites appear in the spectrum [8]. In the continuous part of the spectrum, G posesses a branch cut and the quasiparticle energies become complex. The real part of $\varepsilon_i^{\mathrm{QP}}$ represents some average energy of a group of excited states and the imaginary part the spread in energy of these states [5].

[3] This follows directly from Eq. (1.1) by inserting the identity $1 = \sum_i |N\pm1,i\rangle\langle N\pm1,i|$, performing Fourier transformation and using analytical continuation.

Only for non-interacting electrons, the eigenvalues are real and the QP wavefunctions can be written as single Slater determinants. The excitations are then true single-particle excitations with energies $\omega = \varepsilon_j$.

The norm of the quasiparticle wavefunction (1.14) is given by:

$$\|\psi_i^{\mathrm{QP}}\|^2 = \left\langle \psi_i(\varepsilon_i^{\mathrm{QP}}) \middle| 1 - \Sigma'(\varepsilon_i^{\mathrm{QP}}) \middle| \psi_i(\varepsilon_i^{\mathrm{QP}}) \right\rangle^{-1} \equiv Z_i, \tag{1.18}$$

where $\Sigma'(\varepsilon_i^{\mathrm{QP}}) = \frac{d}{d\omega}\Sigma(\omega)\Big|_{\omega=\varepsilon_i^{\mathrm{QP}}}$. For non-interacting electrons, the QP norm can be either 1 or 0, corresponding to single- and multiple-particle excitations, respectively. In weakly correlated systems, states with norm ~ 1 are "quasi" single-particle excitations and only those are usually called quasiparticles.

1.3 Hedin's Equations

A formally exact way of calculating the self-energy is given by a set of four coupled equations, known as Hedin's equations:

$$\text{self-energy:} \qquad \Sigma(1,2) = i \int d(34)\, G(1,3)\Gamma(3,2,4)W(4,1^+), \tag{1.19}$$

$$\text{screened potential:} \qquad W(1,2) = V(1,2) + \int d(34)\, V(1,3)P(3,4)W(4,2), \tag{1.20}$$

$$\text{polarization:} \qquad P(1,2) = -i \int d(34)\, G(1,3)G(4,1^+)\Gamma(3,4,2), \tag{1.21}$$

$$\text{vertex function:} \qquad \Gamma(1,2,3) = \delta(1,2)\delta(1,3) + \int d(4567)\, \frac{\partial \Sigma(1,2)}{\partial G(4,5)} G(4,6)G(7,5)\Gamma(6,7,3), \tag{1.22}$$

where (j) denotes $(\mathbf{r}_j, t_j)$ and $V(1,2) = V(\mathbf{r}_1, \mathbf{r}_2)\delta(t_1 - t_2)$ is the Coulomb potential.

The screened potential W can also be expressed through the dielectric function $\epsilon = 1 - VP$ as $W(1,2) = \int d(3)\, \epsilon^{-1}(1,3)V(3,2)$.

These equations could in principal be solved self-consistently, along with the Dyson equation (1.11), starting from a trial Green's Function, e.g. G_0, and then iterating until the self-energy converges. This is however, due to their complicated structure, impossible to do in practice.

1.4 The GW approximation

A simple ansatz can be made by setting the second term in the vertex function to zero, which yields $P = -iGG$ and $\Sigma = iGW$. This choice seems at first somewhat arbitrary. However, it gains a clear physical interpretation when compared to Hartree-Fock theory (HF), in which the self-energy is given as a product of the Green's Function and the bare Coulomb interaction: $\Sigma^{\mathrm{HF}} = iGV$. Here, electron-electron interaction only occurs through the Hartree- and the exchange potential, that means that there is no correlation – the electrons are quasi independent. On the other hand, correlation is to a large extent determined by screening (in fact, for weakly correlated systems, these two terms are often used interchangeably). Thus, by replacing the

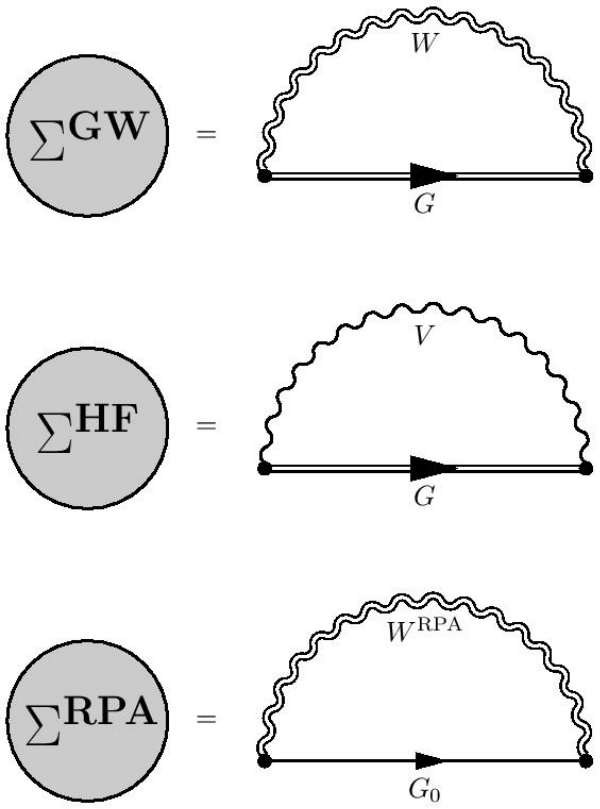


Figure 1.2: Comparison of the GW, HF and RPA self-energies. The screened potential is linked to the full Coulomb potential through a Dyson-like equation $W = V + VGGW$ and $W^{\mathrm{RPA}} = V + VG_0G_0W^{\mathrm{RPA}}$ (see Fig. 1.3).

bare Coulomb interaction V by the screened interaction W in the self-energy, dynamical correlation is introduced. In Fig. 1.2, the Feynman diagram for the GW self-energy is shown, along with corresponding expressions for the HF and the Random Phase Approximation (RPA) for comparison.

In real space and time domain, the GW self-energy is simply given as a product:

$$\Sigma(\mathbf{r}, t; \mathbf{r}', t') = iG(\mathbf{r}, t; \mathbf{r}', t')W(\mathbf{r}, t; \mathbf{r}', t'), \tag{1.23}$$

which becomes a convolution in frequency domain:

$$\Sigma(\mathbf{r}, \mathbf{r}'; \omega) = \frac{i}{2\pi} \int d\omega' \, e^{i\delta\omega'} G(\mathbf{r}, \mathbf{r}'; \omega + \omega')W(\mathbf{r}, \mathbf{r}'; \omega'), \tag{1.24}$$

where the inifinitesimal δ ensures the correct time-ordering in case of a static potential, $W(\omega = 0)$. Using the spectral representation for the Green's Function and an analogous expression for the screened interaction, the real part of Σ breaks into two parts [11]: $\Sigma = \Sigma^{\mathrm{COH}} + \Sigma^{\mathrm{SEX}}$ with the first term arising from the poles in the Green's Function and the second from the poles in W. Σ^{SEX} can be identified as a dynamically screened version of the Fock exchange term and is therefore called "screened exchange". Σ^{COH} describes the dynamic interaction of a particle with the charge that it induces in its surrounding, that means a dynamical "Coulomb hole".

1.4.1 Linearized QP equation

The GW approximation provides a direct guideline on how to find the quasiparticle wavefunctions and their spectrum: First, one has to construct the Green's Function and the screened potential from an initial guess, e.g. from a system of (effectively) non-interacting electrons with wavefunctions $\{\psi_i^s\}$ and energies $\{\varepsilon_i^s\}$. This allows then for the calculation of Σ and determination of new wavefunctions and energies from the quasiparticle equation (1.13). From these, a new Green's Function and potential can be build up and the whole procedure can be iterated until self-consistency is reached. However, Eq. (1.13) requires the self-energy itself to be given at the quasiparticle energy $\varepsilon_i^{\mathrm{QP}}$ which is exactly the quantity one wants to find. This is extremely complicated to solve directly. Instead, the QP equation can be linearized using first-order perturbation theory: Assume that the initial wavefunctions are solutions to $\hat{H}^{\mathrm{eff}} = \hat{h}_0 + V_H + V_{xc}$, where V_{xc} is some effective exchange-correlation potential. Then the full Hamiltonian differs from $\hat{H}^{\mathrm{eff}}$ by $\Sigma_{xc} - V_{xc}$. If the initial wavefunctions are close to the true QP wavefunctions, this difference will be small and one can use perturbation theory in $(\Sigma_{xc} - V_{xc})$. In first order, this yields for the quasiparticle energies:

$$\varepsilon_i^{\mathrm{QP}} = \varepsilon_i^s + Z_i^s \cdot \langle \psi_i^s | \Sigma_{xc}(\varepsilon_i^s) - V_{xc} | \psi_i^s \rangle \,, \tag{1.25}$$

with a renormalization factor:

$$Z_i^s = \langle \psi_i^s | 1 - \Sigma'_{xc}(\varepsilon_i^s) | \psi_i^s \rangle^{-1} \,. \tag{1.26}$$

Z_i^s is an approximation to the QP norm (1.18) and is a measure for how well the quasiparticle wavefunction is represented by ψ_i^s, e.g. if the quasiparticle state can be described by effectively non-interacting electrons.

Within this approach, new wavefunctions can be found by replacing V_{xc} with the effective self-energy operator $\sum_{ij} |\psi_i^s\rangle\langle\psi_i^s|\Sigma_{xc}(\varepsilon_i^s))|\psi_j^s\rangle\langle\psi_j^s|$ in the Hamiltonian and finding the new eigenvectors. This is know as quasiparticle self-consistent GW (QP*sc*GW) [12], as opposed to the full self-consistent GW (*sc*GW) [13], in which the self-energy is calculated for all frequencies and the Green's Function is evaluated through Dyson's equation (1.11).

1.4.2 G_0W_0

In practice, Kohn-Sham orbitals and eigenvalues from a DFT calculation are often used as input for a GW calculation and the quasiparticle spectrum is evaluated non-selfconsistenly from Eq. (1.25) without updating the Green's Function or the screened potential, that means only one iteration is made. This is known as the "one-shot" GW or G_0W_0 approximation and has become a standard tool in electronic structure theory. W_0 is hereby equal to the RPA screened potential, as depicted in Fig. 1.3. Even though this approach is based on several crude simplifications, namely: 1. the GW approximation itself, 2. non-selfconsistency and 3. linearization of the QP equation, it gives a very good balance between accuracy and computational costs. Its great success can be understood by the following:

1. The GW approximation is physically well motivated for weakly correlated systems by the concept of screening, as outlined in the beginning of this section.

2. Self-consistency does not necessarily improve results. This is due to the fact that additional terms are introduced to the self-energy, which would cancel out when the full many-body theory is considered, e.g. by taking the vertex function (1.22) into account [14]. Calculating vertex corrections, however, is enormously costly.

3. The linearized QP equation is a good approximation as long as the initial wave functions already describe the true QP wavefunctions fairly well, even though this may not hold for the eigenvalues. Then, G_0W_0 can give significant corrections to the energies, allowing for accurate bandstructure calculations.

Of course, there are several drawbacks and limitations: A major problem is the starting point issue, since the results depend on the initial wavefunctions and energies and can vary significantly depending on the input. A well-considered choice may be crucial in order to minimize errors. This issue can only be overcome by performing self-consistency. Otherwise, it is a true *ab-initio* method, meaning that no empirical parameters are required, and it is system-independent (which is not the case for most funtionals in DFT). Unfortunately, all this comes at the price of a much higher computational cost, as will be elucidated in detail in the next chapter. And while the lower part of the spectrum can be calculated with good accuracy, G_0W_0 usually goes completely wrong in the high energy range and thus cannot describe satellites, for instance. Finally, for strongly correlated materials, the whole quasiparticle picture does not hold and the GW approximation is expected to fail.

Several approaches exist, that go beyond the one-shot approximation without having to deal with all of the complications and problems of the full self-consistent scheme. In the eigenvalue-*sc* GW, for instance, only the energies are being updated during the iterations while the wavefunctions are being kept on the Kohn-Sham level [15]. Furthermore, energies and/or wavefunctions can be updated in the Green's Function only with a fixed initial screened potential, which corresponds to GW_0 [16, 17].

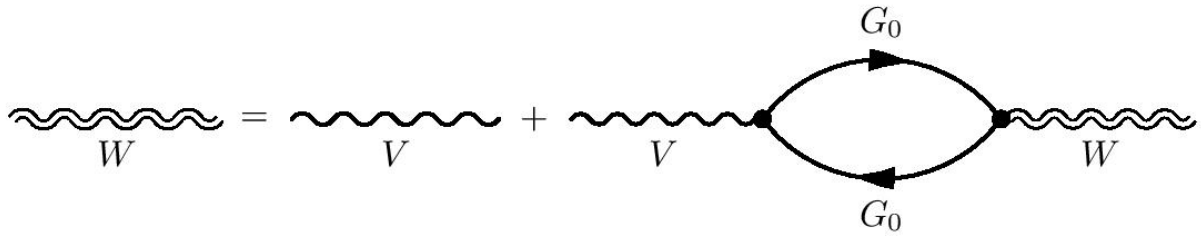


Figure 1.3: Definition of the screened potential W in the G_0W_0 approximation. Similar to Fig. 1.1, it gives an infinite sum over bubble diagrams (G_0G_0) and is equal to the screened potential in the Random Phase Approximation. Each bubble corresponds to the creation and annihilation of one electron-hole pair. Within this approximation, these pairs are non-interacting.

CHAPTER 2

IMPLEMENTATION IN GPAW

GPAW is an electronic structure code based on the projector-augmented wave (PAW) method [18, 19], in which the true wavefunctions are replaced by smooth auxiliary wavefunctions inside atom-centered augmentation spheres. A detailed description of the GPAW code is given in Ref. [20]. Originally, wavefunctions were represented on a real-space grid [21], but later, linear combination of atomic orbitals (LCAO) basis sets [22] and more recently, plane wave representation have been introduced.

In this chaper, all details of the implementation of the G_0W_0 approximation within GPAW are presented. It follows mainly Ref. [23]. The GW self-energy is calculated from the inverse dieelectric function, ϵ^{-1}, in the Random Phase Approximation [24] given as a matrix in a plane wave basis. The frequency-dependence of ϵ can either be evaluated explicitly on a grid ("full frequency dependent method") or modelled in the Plasmon Pole Approximation (PPA) by Godby and Needs [25]. Furthermore, the static limit leads to the so-called static COHSEX approximation.

Special care is required for the divergent terms of the screened potential in the long wavelength limit $q \to 0$. This divergence can be treated both analytically and numerically. Another feature is a truncation scheme for the Coulomb potential. This is immensely important for two-dimensional materials in supercell calculations in order to eliminate spurious interaction effects between periodically repeated layers.

The calculation of the GW self-energy, Σ, includes sums over k points, both occupied and empty bands as well as plane waves and, for the full frequency dependent method, an integration over frequencies – in principal up to infinity. In practice, all summations have to be limited and integrations must be carried out numerically, imposing a number of convergence issues. This also makes GW calculations much more complicated and computationally demanding as compared to groundstate DFT.

2.1 Calculation of the self-energy

The G_0W_0 self-energy can be split into two contributions: $\Sigma^{GW} = V_X + \Sigma_c$, where V_X is the nonlocal exchange potential as in Hartree-Fock theory, and Σ_c is the correlation part. In the following, the latter will simply be denoted as the self-energy $\Sigma = \Sigma_c$. By introducing the difference between the screened and the bare Coulomb potential $\overline{W} = W - V$, it reads:

$$\Sigma(\mathbf{r}, \mathbf{r}'; \omega) = \frac{i}{2\pi} \int d\omega' \, G(\mathbf{r}, \mathbf{r}'; \omega + \omega') \overline{W}(\mathbf{r}, \mathbf{r}'; \omega'). \tag{2.1}$$

In this way, the exact exchange is separated from the actual GW calculation. Since the screened potential approaches the bare Coulomb potential for large ω, $\overline{W}$ goes to zero and the frequency integration becomes numerically stable.

Using Bloch states $|n\mathbf{k}\rangle$, where n and $\mathbf{k}$ denote band index and k-point index, respectively, for the spectral representation of the Green's Function (1.9) and expanding in plane waves, the diagonal terms become:

$$\begin{aligned} \Sigma_{n\mathbf{k}} &\equiv \langle n\mathbf{k}|\Sigma(\omega)|n\mathbf{k}\rangle \\ &= \frac{1}{\Omega} \sum_{\mathbf{G}\mathbf{G}'} \sum_{\mathbf{q}}^{1.\mathrm{BZ}} \sum_{m}^{\mathrm{all}} \frac{i}{2\pi} \int_{-\infty}^{\infty} d\omega' \, \overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega') \frac{\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}) \rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}')^*}{\omega + \omega' - \varepsilon^s_{m\mathbf{k}-\mathbf{q}} + i\eta \, \mathrm{sgn}(\varepsilon^s_{m\mathbf{k}-\mathbf{q}} - \mu)}, \end{aligned} \tag{2.2}$$

with the pair density matrices defined as:

$$\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}) \equiv \left\langle n\mathbf{k} \middle| e^{i(\mathbf{q}+\mathbf{G})\mathbf{r}} \middle| m\,\mathbf{k}-\mathbf{q} \right\rangle. \tag{2.3}$$

$\Omega = \Omega_{\mathrm{cell}} \cdot N_{\mathbf{k}}$ is the total volume, Ω_{cell} the volume of the unit cell and $N_{\mathbf{k}}$ the number of k points. The sums in Eq. (2.2) run over plane waves with wave vectors $\mathbf{G}$ and $\mathbf{G}'$, all differences $\mathbf{q}$ between k points in the first Brillouin zone and all band indices m, respectively. The wavefunctions and corresponding eigenvalues, $\varepsilon^s_{n\mathbf{k}}$, are taken from a Kohn-Sham groundstate calculation. The potential reads:

$$\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = \frac{4\pi}{|\mathbf{q}+\mathbf{G}|} \left(\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) - \delta_{\mathbf{G}\mathbf{G}'} \right) \frac{1}{|\mathbf{q}+\mathbf{G}'|}, \tag{2.4}$$

where $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega)$ is the inverse dielectric matrix, which is obtained in the Random Phase Approximation with a symmetrized Coulomb kernel in $\mathbf{G}$ and $\mathbf{G}'$:

$$\epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = \delta_{\mathbf{G}\mathbf{G}'} - \frac{4\pi}{|\mathbf{q}+\mathbf{G}|} \chi^0_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) \frac{1}{|\mathbf{q}+\mathbf{G}'|}, \tag{2.5}$$

from the non-interacting, time-ordered density response function:

$$\chi^0_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = \frac{2}{\Omega} \sum_{\mathbf{k}}^{1.\mathrm{BZ}} \sum_{n,n'} \left(f^s_{n\mathbf{k}} - f^s_{n'\mathbf{k}+\mathbf{q}} \right) \frac{\rho^{n\mathbf{k}}_{n'\mathbf{k}+\mathbf{q}}(\mathbf{G}) \rho^{n\mathbf{k}}_{n'\mathbf{k}+\mathbf{q}}(\mathbf{G}')^*}{\omega + \varepsilon^s_{n\mathbf{k}} - \varepsilon^s_{n'\mathbf{k}+\mathbf{q}} + i\eta \, \mathrm{sgn}(\varepsilon^s_{n'\mathbf{k}+\mathbf{q}} - \varepsilon^s_{n\mathbf{k}})}, \tag{2.6}$$

with occupation numbers $f^s_{n\mathbf{k}}$. Details on the implementation of the linear density response function and the calculation of the pair density matrices with PAW corrections are given in Ref. [24].

The quasi-particle spectrum is then obtained from Eq. (1.25) as:

$$\varepsilon^{\mathrm{QP}}_{n\mathbf{k}} = \varepsilon^s_{n\mathbf{k}} + Z^s_{n\mathbf{k}} \cdot \mathrm{Re} \left\langle n\mathbf{k} \middle| \Sigma(\varepsilon^s_{n\mathbf{k}}) + V_x - V_{xc} \middle| n\mathbf{k} \right\rangle, \tag{2.7}$$

with a renormalization factor given by:

$$Z^s_{n\mathbf{k}} = \left(1 - \mathrm{Re}\left\langle n\mathbf{k}\middle|\Sigma'(\varepsilon^s_{n\mathbf{k}})\middle|n\mathbf{k}\right\rangle\right)^{-1}. \tag{2.8}$$

The derivative of the self-energy with respect to the frequency is calculated analytically from Eq. (2.2):

$$\Sigma'(\varepsilon^s_{n\mathbf{k}}) = -\frac{1}{\Omega}\sum_{\mathbf{G}\mathbf{G}'}\sum_{\mathbf{q}}^{1.\mathrm{BZ}}\sum_{m}^{\mathrm{all}}\frac{i}{2\pi}\int_{-\infty}^{\infty}d\omega'\,\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega')\frac{\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G})\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}')^*}{\left(\varepsilon^s_{n\mathbf{k}}+\omega'-\varepsilon^s_{m\mathbf{k}-\mathbf{q}}\pm i\eta\right)^2}, \tag{2.9}$$

where $\pm = \mathrm{sgn}(\varepsilon^s_{m\mathbf{k}-\mathbf{q}} - \mu)$.

The calculation of the exact exchange $\langle n\mathbf{k}|V_x|n\mathbf{k}\rangle$ contributions is done seperately in a different part of the GPAW code. This can therefore be done on a different level of accuracy than for the self-energy.

In the current implementation, $\Sigma_{n\mathbf{k}}$ is only evaluated for energies $\omega = \varepsilon^s_{n\mathbf{k}}$ and only its real part is stored. This means, that no further information on the spectrum like quasiparticle lifetimes, line shapes and satellites is available. For semiconductors, however, there exists an energy region around the quasiparticle gap for which the imaginary part of the self-energy is zero and quasiparticle peaks become renormalized δ-functions. The size of this region is determined by the underlying Kohn-Sham bandgap in the G_0W_0 approximation [26]. Since the main focus of this work is the calculation of quasiparticle bandstructures around the Fermi level, this simplification is reasonable. On the other hand, an extension to analysing the complex and frequency-dependent self-energy is trivial and may be done in the future. This will in particular be of interest for metallic systems [27, 28].

Furthermore, only the diagonal terms of the self-energy are evaluated. In principal, determining the off-diagonal elements, $\langle n\mathbf{k}|\Sigma(\omega)|n'\mathbf{k}\rangle$, would allow for calculation of quasiparticle wavefunctions and subsequently lead to the (quasiparticle) self-consistent GW method.

2.1.1 Full frequency-dependent method

The frequency integration in Eq. (2.2) can be carried out for positive values of ω' only due to time-reversal symmetry of the screened potential, $\overline{W}(-\omega) = \overline{W}(\omega)$, by rewriting the integral as:

$$\begin{aligned} I(\omega) &\equiv \int_{-\infty}^{\infty}d\omega'\,\frac{\overline{W}(\omega')}{\omega+\omega'-\varepsilon^s_{m\mathbf{k}-\mathbf{q}}\pm i\eta} \\ &= \int_{0}^{\infty}d\omega'\,\overline{W}(\omega')\left(\frac{1}{\omega+\omega'-\varepsilon^s_{m\mathbf{k}-\mathbf{q}}\pm i\eta}+\frac{1}{\omega-\omega'-\varepsilon^s_{m\mathbf{k}-\mathbf{q}}\pm i\eta}\right) \end{aligned} \tag{2.10}$$

Then, two different ways of calculating $\Sigma_{n\mathbf{k}}$ are available:

In the first method, the double sum over $\mathbf{G}$ and $\mathbf{G}'$ is carried out first as a matrix multiplication of $\rho(\mathbf{G})\rho^*(\mathbf{G}')$ and $\overline{W}_{\mathbf{G}\mathbf{G}'}$. Then, the frequency integration is performed numerically. This is done seperately for each pair of $(n\,\mathbf{k})$ and $(m\,\mathbf{k}-\mathbf{q})$.

The second method reverses this order and is similar to a Hilbert transform: The numerical frequency integration is done first, but for $\varepsilon^s_{m\mathbf{k}-\mathbf{q}} > \mu$ and $\varepsilon^s_{m\mathbf{k}-\mathbf{q}} < \mu$ separately, denoted by I^+

and I^-, respectively. Defining $\overline{\omega} = \omega - \varepsilon^s_{m\mathbf{k}-\mathbf{q}}$, four cases for the integral can be distinguished:

$$\overline{\omega} \geq 0 \quad \text{and} \quad \varepsilon^s_{m\mathbf{k}-\mathbf{q}} > \mu :$$
$$\int_0^\infty d\omega' \, \overline{W}(\omega') \left(\frac{1}{|\overline{\omega}| + \omega' + i\eta} + \frac{1}{|\overline{\omega}| - \omega' + i\eta} \right) = I^+(|\overline{\omega}|),$$
$$\overline{\omega} \geq 0 \quad \text{and} \quad \varepsilon^s_{m\mathbf{k}-\mathbf{q}} < \mu :$$
$$\int_0^\infty d\omega' \, \overline{W}(\omega') \left(\frac{1}{|\overline{\omega}| + \omega' - i\eta} + \frac{1}{|\overline{\omega}| - \omega' - i\eta} \right) = I^-(|\overline{\omega}|),$$
$$\overline{\omega} < 0 \quad \text{and} \quad \varepsilon^s_{m\mathbf{k}-\mathbf{q}} > \mu :$$
$$\int_0^\infty d\omega' \, \overline{W}(\omega') \left(\frac{1}{-|\overline{\omega}| + \omega' + i\eta} + \frac{1}{-|\overline{\omega}| - \omega' + i\eta} \right) = -I^-(|\overline{\omega}|),$$
$$\overline{\omega} < 0 \quad \text{and} \quad \varepsilon^s_{m\mathbf{k}-\mathbf{q}} < \mu :$$
$$\int_0^\infty d\omega' \, \overline{W}(\omega') \left(\frac{1}{-|\overline{\omega}| + \omega' - i\eta} + \frac{1}{-|\overline{\omega}| - \omega' - i\eta} \right) = -I^+(|\overline{\omega}|),$$

which can be summarized as:

$$I(\omega) = \mathrm{sgn}(\overline{\omega}) I^{\mathrm{sgn}(\overline{\omega}) \cdot \mathrm{sgn}(\varepsilon^s_{m\mathbf{k}-\mathbf{q}} - \mu)}(|\overline{\omega}|). \tag{2.11}$$

Summing over $\mathbf{G}$ and $\mathbf{G}'$ then gives the contributions to $\Sigma(\omega)$ for every $(m\,\mathbf{k}–\mathbf{q})$ represented on a finite, positive frequency grid, $\{\overline{\omega}_i\}$. The self-energy at the input eigenvalue, $\Sigma(\omega = \varepsilon^s_{n\mathbf{k}})$, is found by linear interpolation between the two closest points on the grid with $\overline{\omega}_i \leq |\varepsilon^s_{n\mathbf{k}} - \varepsilon^s_{m\mathbf{k}-\mathbf{q}}| < \overline{\omega}_{i+1}$, again for every m and $\mathbf{q}$ seperately.

The same methods apply for the derivative.

2.1.2 Plasmon Pole Approximation

In the Plasmon Pole Approximation (PPA), all the transitions from occupied to unoccupied states $n \to n'$ that sum up to the to the inverse dielectric function (similar to Eq. (2.6)) are averaged to form one single collective excitation, known as plasmon:

$$\begin{aligned} \epsilon^{-1}(\omega) &\propto \sum_{n \to n'} \frac{R_{n\to n'}}{\omega - \omega_{n\to n'} + i\eta} - \frac{R^*_{n\to n'}}{\omega + \omega_{n\to n'} - i\eta} \\ &\approx \frac{R}{\omega - \tilde{\omega} + i\eta} - \frac{R}{\omega + \tilde{\omega} - i\eta}, \end{aligned} \tag{2.12}$$

with some averaged spectral function R. The imaginary part consists only of single peaks at the main plasmon frequencies, $\pm\tilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})$. Thus, $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega)$ can be modeled as:

$$\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = \delta_{\mathbf{G}\mathbf{G}'} + R_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) \left(\frac{1}{\omega - \tilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) + i\eta} - \frac{1}{\omega + \tilde{\omega}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) - i\eta} \right), \tag{2.13}$$

where the spectral function, $R_{\mathbf{G}\mathbf{G}'}(\mathbf{q})$, is assumed to be real. The two terms account for positive and negative frequencies, respectively. Using the Sokhatsky-Weierstrass theorem,

$$\lim_{\eta \to 0^+} \frac{1}{x \pm i\eta} = \mathcal{P}\left\{ \frac{1}{x} \right\} \mp i\pi\delta(x), \tag{2.14}$$

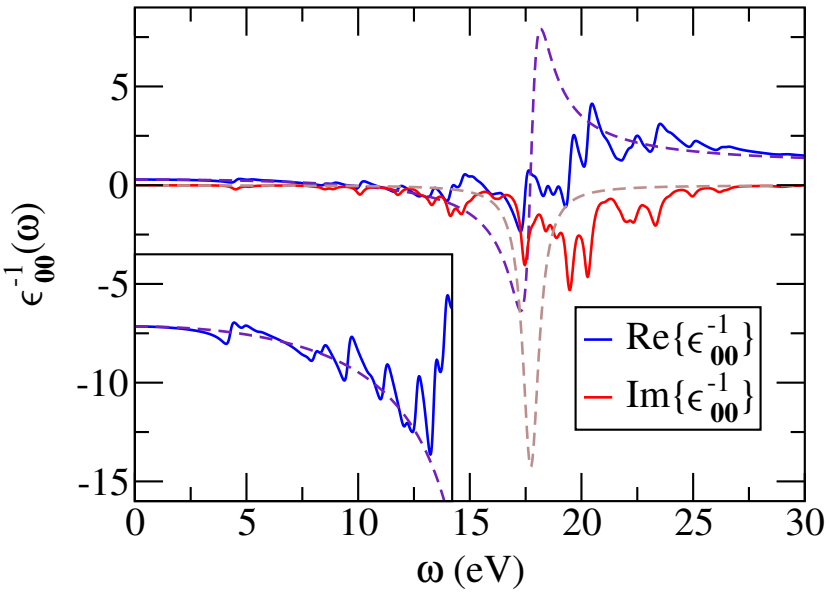


Figure 2.1: Real and imaginary parts of the head of the inverse dielectric function $\epsilon^{-1}_{\mathbf{00}}(\mathbf{q},\omega)$ of a silicon bulk test system for $\mathbf{q} = (1/2, 1/2, 1/2)$. The PPA model (dashed lines) is compared to the fully frequency-dependent *ab-initio* results (full lines). A broadening of $\eta = 0.2\,\text{eV}$ and $E_0 = 1\,\text{Hartree}$ have been used. The imaginary part of the PPA function consists of a single peak at the plasmon frequency. The real part is given by the Kramers-Kronig relation: $\text{Re}\{\epsilon^{-1}(\omega)\} = 1/\pi \int_{-\infty}^{\infty} d\omega'\, \text{Im}\{\epsilon^{-1}(\omega')\}/(\omega' - \omega)$. The inset is a zoom-in on the y-axis. Both in the low and high energy range, the overall shape of $\epsilon^{-1}(\omega)$ is well described by the model function.

where $\mathcal{P}$ denotes the Cauchy principal value, the real and imaginary parts of the potential $\overline{W} = (\epsilon^{-1} - \delta)V$ are given as:

$$\text{Re}\left\{\overline{W}_{\mathbf{GG'}}(\mathbf{q},\omega)\right\} = R_{\mathbf{GG'}}(\mathbf{q})\mathcal{P}\left\{\frac{1}{\omega - \tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})} - \frac{1}{\omega + \tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})}\right\}\frac{4\pi}{|\mathbf{q}+\mathbf{G}||\mathbf{q}+\mathbf{G'}|}, \tag{2.15}$$

and

$$\text{Im}\left\{\overline{W}_{\mathbf{GG'}}(\mathbf{q},\omega)\right\} = -\pi R_{\mathbf{GG'}}(\mathbf{q})\left(\delta(\omega - \tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})) + \delta(\omega + \tilde{\omega}_{\mathbf{GG'}}(\mathbf{q}))\right)\frac{4\pi}{|\mathbf{q}+\mathbf{G}||\mathbf{q}+\mathbf{G'}|}, \tag{2.16}$$

respectively. Similar expressions are found for the non-interacting Green's Function, so that the convolution in Eq. (2.1) can be carried out analytically and the real part of the self-energy (2.2) becomes:

$$\begin{aligned}\text{Re}\{\Sigma_{n\mathbf{k}}\} \;=\; \text{Re}\Bigg\{ & -\frac{1}{\Omega}\sum_{\mathbf{GG'}}\sum_{\mathbf{q}}^{\text{1.BZ}}\sum_{m}^{\text{all}}\frac{1}{2\pi}\frac{4\pi R_{\mathbf{GG'}}(\mathbf{q})}{|\mathbf{q}+\mathbf{G}||\mathbf{q}+\mathbf{G'}|}\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G})\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G'})^{*} \\ & \times\left(\pm\pi\left(\frac{1}{\omega - \varepsilon^{s}_{m\mathbf{k}-\mathbf{q}} + \tilde{\omega}_{\mathbf{GG'}}(\mathbf{q}) - i\eta} - \frac{1}{\omega - \varepsilon^{s}_{m\mathbf{k}-\mathbf{q}} - \tilde{\omega}_{\mathbf{GG'}}(\mathbf{q}) + i\eta}\right)\right. \\ & \left.-\pi\left(\frac{1}{\omega - \varepsilon^{s}_{m\mathbf{k}-\mathbf{q}} + \tilde{\omega}_{\mathbf{GG'}}(\mathbf{q}) \pm i\eta} + \frac{1}{\omega - \varepsilon^{s}_{m\mathbf{k}-\mathbf{q}} - \tilde{\omega}_{\mathbf{GG'}}(\mathbf{q}) \pm i\eta}\right)\right)\Bigg\},\end{aligned} \tag{2.17}$$

where the infinitesimal η is maintained to ensure numerical stability, when the denominator goes to 0.

The model dielectric function (2.13) is required to reproduce the *ab-initio* dielectric function in the static limit $\omega_1 = 0$ and at some frequency $\omega_2 = iE_0$. The latter is chosen to be imaginary, since $\epsilon^{-1}(\omega)$ is smooth along the imaginary axis. This method is known as the Plasmon Pole Approximation of Godby and Needs [25]. From

$$\epsilon^{-1}(\mathbf{q},\omega_1) = \frac{-2R_{\mathbf{GG'}}(\mathbf{q})}{\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})}, \tag{2.18}$$

$$\epsilon^{-1}(\mathbf{q},\omega_2) = \frac{-2R_{\mathbf{GG'}}(\mathbf{q})\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})}{E_0^2+\tilde{\omega}^2_{\mathbf{GG'}}(\mathbf{q})}, \tag{2.19}$$

one obtains the plasmon frequency and the spectral function:

$$\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q}) = E_0\sqrt{\frac{\epsilon^{-1}(\mathbf{q},\omega_2)}{\epsilon^{-1}(\mathbf{q},\omega_1)-\epsilon^{-1}(\mathbf{q},\omega_2)}}, \tag{2.20}$$

$$R_{\mathbf{GG'}}(\mathbf{q}) = -\frac{\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})}{2}\epsilon^{-1}(\mathbf{q},\omega_1). \tag{2.21}$$

The Plasmon Pole Approximation is valid for systems where the dielectric response is dominated by its main plasmon excitation. Then, the overall shape of $\epsilon^{-1}(\omega)$ will be determined by a single resonance at the plasmon frequency and all other details of its structure will be averaged out in the frequency integration for the self-energy in Eq. (2.1). This is illustrated in Fig. 2.1.

2.1.3 Static COHSEX

A static approximation assumes that the main contributions to the GW self-energy (2.2) arise from terms, where $\omega - \varepsilon^s_{m\mathbf{k}-\mathbf{q}}$ is small compared to the energy of the main excitation in the screened potential, that is essentially the plasmon energy [8]. By setting $\omega - \varepsilon^s_{m\mathbf{k}-\mathbf{q}} = 0$, the Coulomb hole and screened exchange parts of the self-energy Σ_{xc} become frequency-independent and read:

$$\Sigma^{\mathrm{COH}}_{n\mathbf{k}} = \frac{1}{2\Omega}\sum_{\mathbf{GG'}}\sum_{\mathbf{q}}\sum_{m}^{\mathrm{all}}\overline{W}_{\mathbf{GG'}}(\mathbf{q},0)\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G})\rho^{n\mathbf{k}*}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G'}), \tag{2.22}$$

$$\Sigma^{\mathrm{SEX}}_{n\mathbf{k}} = -\frac{1}{\Omega}\sum_{\mathbf{GG'}}\sum_{\mathbf{q}}\sum_{m}^{\mathrm{occ}}W_{\mathbf{GG'}}(\mathbf{q},0)\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G})\rho^{n\mathbf{k}*}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G'}), \tag{2.23}$$

respectively. This is known as the static COHSEX approximation. In real space representation, they are given as:

$$\Sigma^{\mathrm{COH}} = \frac{1}{2}\delta(\mathbf{r}-\mathbf{r'})\left(W(\mathbf{r},\mathbf{r'};\omega=0)-V(\mathbf{r},\mathbf{r'})\right), \tag{2.24}$$

$$\Sigma^{\mathrm{SEX}} = -\sum_{j}^{\mathrm{occ}}\phi^*_j(\mathbf{r})\phi_j(\mathbf{r'})W(\mathbf{r},\mathbf{r'};\omega=0), \tag{2.25}$$

from which the interpretation as static Coulomb hole and screened exchange becomes clear.

The QP spectrum is evaluated as:

$$\varepsilon^{\mathrm{QP}}_{n\mathbf{k}} = \varepsilon^s_{n\mathbf{k}} + \left\langle n\mathbf{k}\middle|\Sigma^{\mathrm{SEX}}+\Sigma^{\mathrm{COH}}-V_{xc}\middle|n\mathbf{k}\right\rangle. \tag{2.26}$$

2.1.4 Divergence of the screened potential

The head ($\mathbf{G} = \mathbf{0}$ and $\mathbf{G}' = \mathbf{0}$) and wings ($\mathbf{G} = \mathbf{0}$ or $\mathbf{G}' = \mathbf{0}$) of the screened potential (2.4) diverge as $1/q^2$ and $1/q$, respectively, in the long wavelength limit $q \to 0$. However, for an infinitely dense k-point sampling, these divergencies are lifted in the calculation of the self-energy (2.2). This can be seen by replacing the sum over $\mathbf{q}$ by an integral over the volume of the 1. Brillouin zone:

$$\sum_{\mathbf{q}} \to \frac{\Omega}{(2\pi)^3} \int_{\Omega_{\mathrm{BZ}}} d\mathbf{q} = \frac{\Omega}{(2\pi)^3} \int dq\, 4\pi q^2. \tag{2.27}$$

Assuming $\epsilon^{-1}_{\mathbf{00}}(\mathbf{q} \to 0)$ to be isotropic, $\overline{W}_{\mathbf{00}}(\mathbf{q} = \mathbf{0})$ can be found by integrating the divergent part $4\pi/q^2$ over a small sphere with volume $\Omega'_{\mathrm{BZ}} = \Omega_{\mathrm{BZ}}/N_{\mathbf{k}}$ (this corresponds to a radius $q' = (6\pi^2/\Omega)^{1/3}$). This yields:

$$\overline{W}_{\mathbf{00}}(\mathbf{q} = \mathbf{0}, \omega) = \frac{2\Omega}{\pi} \left(\frac{6\pi^2}{\Omega}\right)^{1/3} \left[\epsilon^{-1}_{\mathbf{00}}(\mathbf{q} \to \mathbf{0}, \omega) - 1\right]. \tag{2.28}$$

and similar for the wings:

$$\overline{W}_{\mathbf{G0}}(\mathbf{q} = \mathbf{0}, \omega) = \frac{1}{|\mathbf{G}|} \frac{\Omega}{\pi} \left(\frac{6\pi^2}{\Omega}\right)^{2/3} \epsilon^{-1}_{\mathbf{G0}}(\mathbf{q} \to \mathbf{0}, \omega), \tag{2.29}$$

with the dielectric matrix taken in the optical limit.

Alternatively, these values can also be obtained by numerical averaging on a very fine q'-point grid around the Γ-point.

2.1.5 Truncation of the Coulomb potential

In supercell calculations for systems which are infinite and periodic in two dimensions (2D systems), the long range Coulomb interaction can be cut off along the non-periodic direction, z, in order to avoid artificial image charge effects from neighboring cells [29]:

$$\tilde{v}^{2D}(\mathbf{r}) = \frac{\theta(R - |r_z|)}{|\mathbf{r}|}, \tag{2.30}$$

where θ is the step function and R the truncation length. Fourier transformation to reciprocal space yields:

$$\tilde{v}^{2D}(\mathbf{G}) = \frac{4\pi}{\mathbf{G}^2} \left[1 + e^{-G_{\parallel}R} \left(\frac{G_z}{G_{\parallel}} \sin(G_z R) - \cos(|G_z|R)\right)\right], \tag{2.31}$$

where $G_{\parallel}$ and G_z are the parallel and perpendicular components of $\mathbf{G}$, respectively. For $R = L_z/2$, where L_z is the length of the unit cell in the non-periodic direction, this becomes [30]:

$$\tilde{v}^{2D}(\mathbf{G}) = \frac{4\pi}{\mathbf{G}^2} \left(1 - e^{-G_{\parallel}R} \cos(|G_z|R)\right). \tag{2.32}$$

For $G_{\parallel} \to 0$, the expression of Eq. (2.31) and thereby also Eq. (2.32) is not well defined. It is therefore replaced by numerical averaging on a fine uniform q'-point grid around the Γ-point over a small volume Ω'_{BZ}:

$$\tilde{v}^{2\mathrm{D}}(G_{\parallel} = 0) = \frac{1}{\Omega'_{\mathrm{BZ}}} \int_{\Omega'_{\mathrm{BZ}}} d\mathbf{q}'\, \tilde{v}^{2\mathrm{D}}(G_z + \mathbf{q}'). \tag{2.33}$$

The truncated Coulomb potential is used both for the calculation of the dielectric function and of the self-energy.

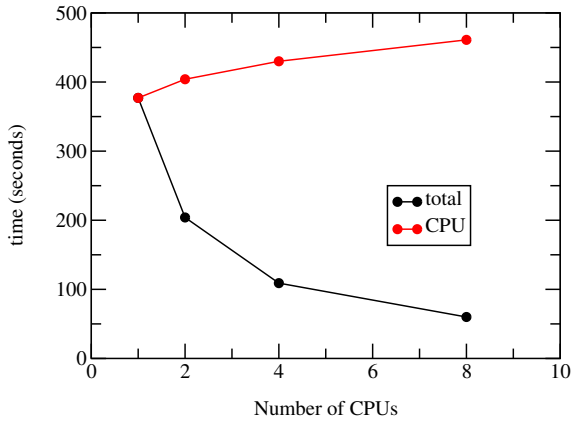


Figure 2.2: Computational time for a small bulk silicon test system with full parallelization over q points on 1, 2, 4 and 8 Intel Xeon cores. The actual time spent for the whole calculation is shown in black, while the CPU time (sum of the times spent on all CPUs) is in red. For perfect parallelization the CPU time would be constant and the total time would divide by the number of cores. Deviances are due to initialization of the calculations, postprocessing and communication between the cores.
$(2 \times 2 \times 2)$ k points, 89 plane waves and bands and 1064 frequency points were used in the calculations.

2.2 Exact exchange contributions

The exact exchange contributions are given in plane wave representation as:

$$\langle n\mathbf{k}|V_x|n\mathbf{k}\rangle = -\frac{\pi}{\Omega}\sum_{m}\sum_{\mathbf{k}'} f^s_{m\mathbf{k}'}\sum_{\mathbf{G}'}\frac{|C_{n\mathbf{k}m\mathbf{k}'}(\mathbf{G}')|^2}{|\mathbf{k}-\mathbf{k}'-\mathbf{G}'|^2}, \tag{2.34}$$

where

$$C_{n\mathbf{k}m\mathbf{k}'}(\mathbf{G}') = \sum_{\mathbf{G}} c^*_{n\mathbf{k}}(\mathbf{G})c_{m\mathbf{k}'}(\mathbf{G}+\mathbf{G}'), \tag{2.35}$$

and $c_{n\mathbf{k}}(\mathbf{G})$ are plane wave coefficients. Treatment of the divergent term $\mathbf{k} = \mathbf{k}'$ and $\mathbf{G}' = \mathbf{0}$ follows Ref. [31], while the calculation of the PAW corrections is described in Ref. [20].

2.3 Computational details

By default, the calculation of the self-energy is fully parallelized over q points. As shown for the example of 8 q points in Fig. 2.2, the parallelization is very efficient, meaning that the total computational time scales very well with the number of available cores. For every $\mathbf{q}$, the inverse dielectric function $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$ is calculated on a given frequency grid as a matrix in $\mathbf{G}$ and $\mathbf{G}'$ using the GPAW implementation of the linear density response function as described in Ref. [24], but modified for time-ordering. From this, the screened potential $\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$ is constructed. Then, Eqs. 2.2 and 2.9 are evaluated for every matrix element $|n\mathbf{k}\rangle$ as described in the previous section. For calculations including the Γ point only, that means finite systems, parallelization over bands m is used instead. Since the arrays $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\omega)$ and $\overline{W}_{\mathbf{G}\mathbf{G}'}(\omega)$ can become very large for high plane wave cutoffs, they can be split and distributed over different cores with additional frequency and plane wave parallelization. This reduces the required memory on each core.

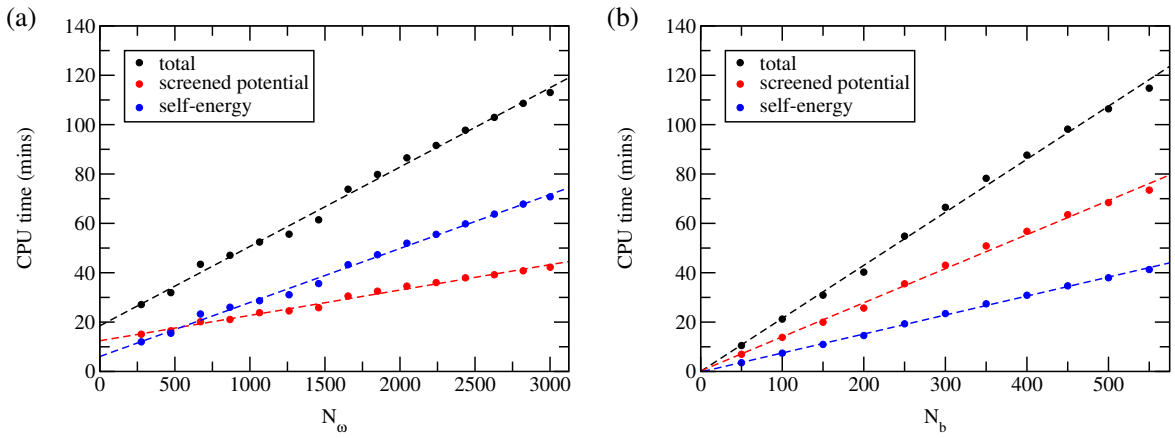


Figure 2.3: Computational time as function of the number of (a) frequency points and (b) bands for a bulk silicon test system with $(3 \times 3 \times 3)$ k points. A cutoff energy of 100 eV (corresponding to 89 plane waves) was used. For (a), 89 bands were included, while the PPA was applied for (b). The time spent on the calculation of the screened potential only (of the self-energy from the screened potential only) is shown in red (blue). Dashed lines are linear fits to the data points.

The calculation of $\Sigma_{n\mathbf{k}}$ scales as $N_\omega \times N_b \times N_k^2 \times N_G^2$ with number of frequency points, bands, k points and plane waves, respectively. This is demonstrated in Figs. 2.3 and 2.4 for a silicon bulk test system on a single 64-bit Intel Xeon core. The graphs also show that the ratio of the computational times spent on the calculation of the screened potential and the self-energy alone depends strongly on the parameters used. For a large number of k points, the computation of the screened potential becomes the bottleneck, since this has to be done seperately for every $\mathbf{q}$ in the 1. Brillouin zone and the calculation of the response function (2.6) itself involves a sum over all k points.

2.3.1 Parameters

All parameters for a GW calculation are defined in a GW object and are listed in Table 2.1.

- `file` is a GPAW file from which all wavefunctions $|n\mathbf{k}\rangle$ and energy eigenvalues $\varepsilon^s_{n\mathbf{k}}$ used as starting point as well as general informations on the system are read. It is created in a preceeding groundstate calculation.
- `nbands` is the number of bands to be included in the summations for the response function (2.6) and the self-energy (2.2).
- `bands` is a list of band indices for which the quasi-particle spectrum 2.7 should be evaluated. Often, only a few bands around the Fermi level are requested.
- `kpoints` is a list of k-point indices for which the quasi-particle spectrum 2.7 should be evaluated. This can be a line of points along a certain direction of the Brillouin zone, for example.
- `e_skn` can be defined to use self-defined starting point eigenvalues $\varepsilon^s_{n\mathbf{k}}$ different from the groundstate. This can be used to perform eigenvalue self-consistent GW calculations, for instance.

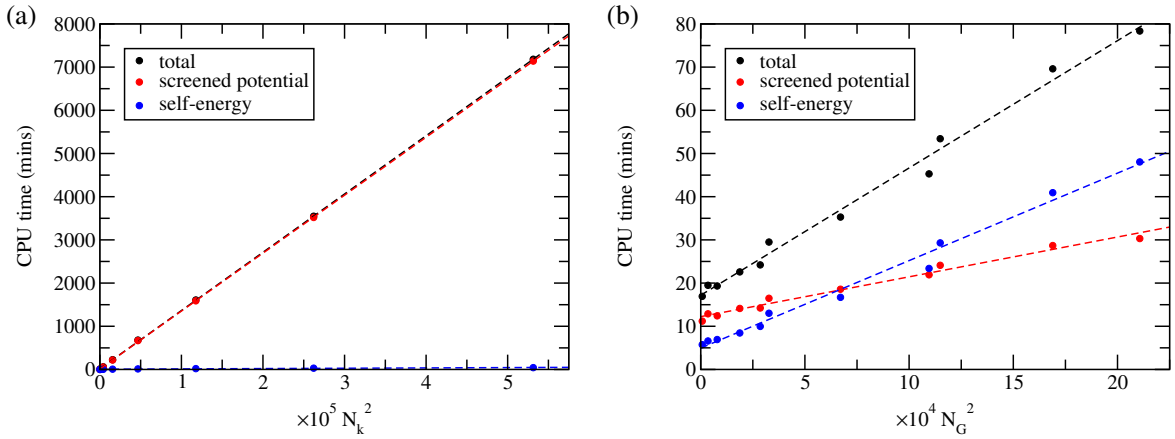


Figure 2.4: Computational time as function of the squared number of (a) k points and (b) plane waves for a bulk silicon test system. For (a), 89 bands and plane waves and for (b), 100 bands and $(3\times3\times3)$ k points were used. The PPA was applied in all calculations. The number of k points in (a) correspond to samplings of $(2\times2\times2)$ up to $(9\times9\times9)$. The number of plane waves in (b) correspond to cutoff energies from 50 up to 300 eV.

- `eshift` shifts all unoccupied bands of the starting point energy eigenvalues by the given value in eV. This corresponds to applying a constant scissors operator like the derivative discontinuity, for example.
- `w` defines the frequency grid on which $\epsilon(\omega)$ and $\overline{W}(\omega)$ are evaluated. In the static COHSEX approximation, it is simply put to $\omega = 0$, while the two values $\omega_1 = 0$ and $\omega_2 = iE_0$ are used in the PPA. For the full frequency-dependent method, a non-uniform grid is created as depicted in Fig. 2.5.
- `ecut` is the plane wave energy cutoff in eV and determines the size of the matrices $\epsilon_{\mathbf{GG'}}$ and $\overline{W}_{\mathbf{GG'}}$ (local field effects). For every $\mathbf{q}$, all plane waves with a maximum kinetic energy $(\mathbf{G}+\mathbf{q})^2/2 = E_{\text{cut}}$ are included.
- `eta` is the broadening parameter given in eV for the calculation of the response function (2.6) and in the PPA for Eq. (2.17). For the static COHSEX approximation, it is set to $\eta = 0.0001$ eV, while it is chosen accordingly to the frequency grid for the full frequency-dependent method as $\eta(\omega) = 4\Delta\omega$.
- `ppa` enables the use of the Plasmon Pole Approximation.
- `E0` defines the PPA fitting frequency.
- `hilbert_trans` can be used to switch between the two different ways of calculating the self-energy in the full frequency-dependent method, as explained in Section 2.1.1.
- `wpar` is the number of cores for parallelizing over frequencies and plane waves in the full frequency-dependent method.
- `vcut='2D'` enables use of the Coulomb truncation.
- `txt` defines the name of the file to which the output from the GW calculation is written.

Table 2.1: Parameters of the GW object. The number of bands, `nbands`, and the plane wave cutoff energy, `ecut`, are always equal in the calculation of the response function (2.6) and the self-energy (2.2).

name	type	default value
`file`	string	None
`nbands`	integer	equal to number of plane waves
`bands`	numpy.ndarray	all up to `nbands`
`kpoints`	numpy.ndarray	all irreducible k points
`e_skn`	numpy.ndarray	None
`eshift`	float	None
`w`	numpy.ndarray	None
`ecut`	float	150 eV
`eta`	float	0.1 eV
`ppa`	boolean	False
`E0`	float	27.2114 eV
`hilbert_trans`	boolean	False
`wpar`	integer	1
`vcut`	string	None
`txt`	string	None

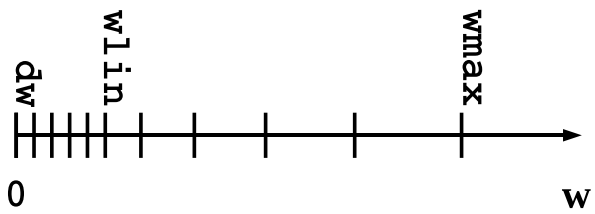


Figure 2.5: For the full frequency-dependent method, a non-uniform grid is defined by `w = [wlin, wmax, dw]`. It is linear in the lower part up to `wlin` with a constant grid spacing `dw`. Above `wlin`, the grid spacing increases linearly up to the maximum frequency `wmax`.

Two functions can be used from the GW object:

- `get_exact_exchange(ecut=None, communicator=world, file='EXX.pckl')` calculates the exact exchange and exchange-correlation contributions and stores the required matrix elements $\langle n\mathbf{k}|V_x|n\mathbf{k}\rangle$ and $\langle n\mathbf{k}|V_{xc}|n\mathbf{k}\rangle$ for later use.
- `get_QP_spectrum(exxfile='EXX.pckl', file='GW.pckl')` performs the actual GW calculation and adds the different contributions for the QP spectrum together.

Further details are documented on the GPAW homepage [32].

2.3.2 Convergence

In principal, all GW calculations need to be checked carefully for convergence with respect to all parameters used. The broadening parameter η and the fitting frequency E_0 for the PPA,

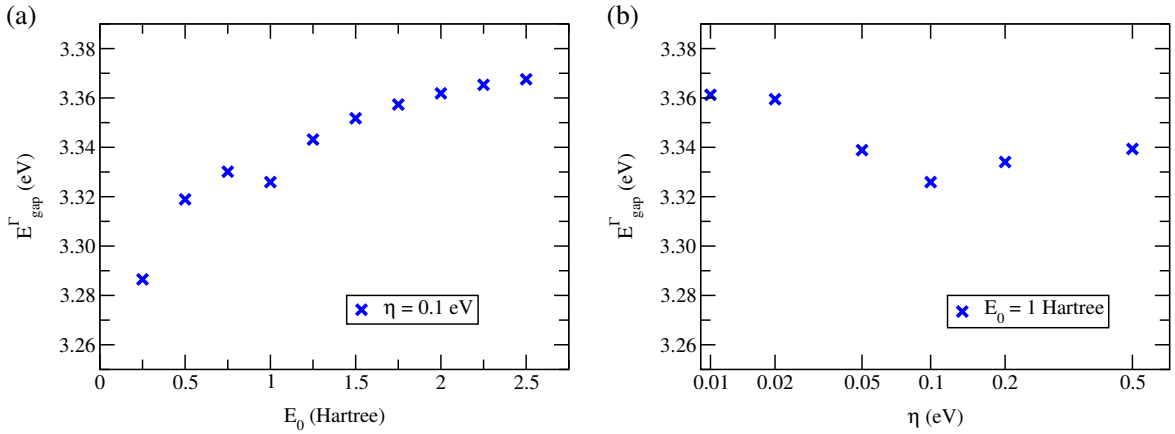


Figure 2.6: Dependence of the direct band gap at the Γ point on the (a) fitting parameter E_0 and (b) broadening η in the PPA for a bulk silicon test system with $(3\times3\times3)$ k points. A cutoff energy of 100 eV and 89 bands were used. All calculations were performed with the LDA functional as starting point, that means G_0W_0@LDA.

however, are often kept at their default values of 0.1 eV and 1 Hartree, respectively, since results are rather insensitive to variations around them. This is illustrated in Fig. 2.6 for the case of the direct band gap of silicon.

For the full frequency-dependent method, results have to be converged with respect to the frequency grid used, e.g. the density and the total number of frequency points. This is shown in Fig. 2.7 (a) for the dependence of the Γ-point band gap on the linear grid spacing $\Delta\omega$ and the frequency ω_{lin} up to which the grid is linear. The maximum frequency is kept constant at 150 eV. In general, ω_{max} only needs to be slightly larger than the largest energy difference $\varepsilon^s_{n\mathbf{k}} - \varepsilon^s_{m\mathbf{k}-\mathbf{q}}$ that occurs in the summation (2.2), as can be seen in Fig. 2.7 (b). The frequency grid should reflect the spectral structure which exhibits in principal very sharp and irregular features for low energies, while it is more broad and smooth in the high range. Well converged results are usually found for $\Delta\omega = 0.05$ eV and $\omega_{\text{lin}} = \omega_{\text{max}}/3$, which results in a few thousand frequency points in practice. Choosing a nonuniform grid in this way may increase the computational speed significantly without any loss of accuracy.

Much more care is to be taken for the convergence with respect to the number k points and the plane wave cutoff. This already holds for the exact exchange contributions, as demonstrated in Fig. 2.8, which shows the Hartree-Fock band gap. The HF bandstructure was obtained non-selfconsistently from LDA wavefunctions and eigenvalues as:

$$\varepsilon^{\text{HF}}_{n\mathbf{k}} = \varepsilon^s_{n\mathbf{k}} + \langle n\mathbf{k}|V_x - V_{xc}|n\mathbf{k}\rangle\,. \tag{2.36}$$

Due to the long-range nature of the exchange potential, a high number of k points is required in order to obtain well-converged results. However, the k-point dependence of the GW self-energy is less severe, since the screened interaction is more short-ranged. As shown in Fig. 2.9 (a), the GW band gap converges much faster with respect to k points, while the dependence on the plane wave cutoff energy is similar. Furthermore, the curves showing the dependence on the cutoff energy only differ by a vertical offset for different k-point samplings. That means, that results converge independently with respect to these two parameters.

On the other, hand it becomes clear from Fig. 2.9 (b) that the convergence of the band gap with respect to the number of bands is not independent from E_{cut}. A too low plane wave energy

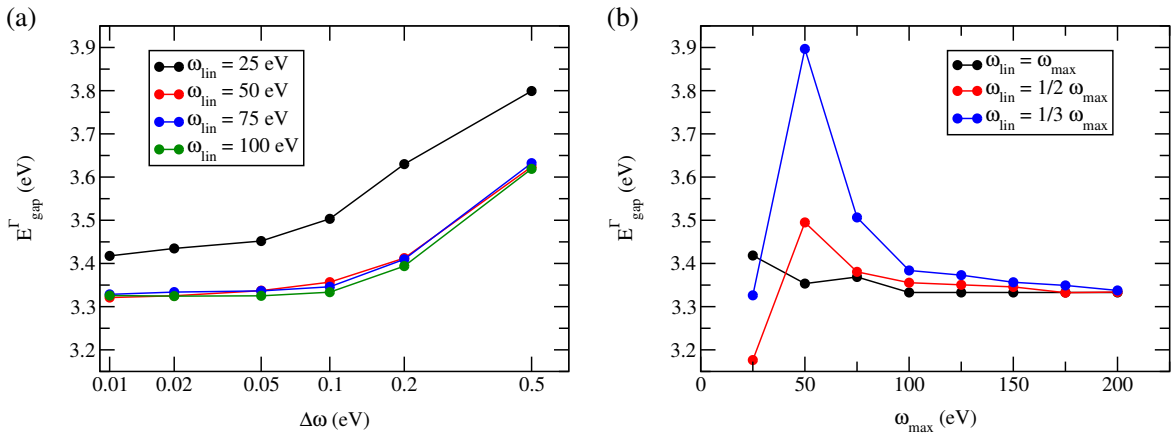


Figure 2.7: Convergence of the direct band gap at the Γ-point with respect to the frequency grid for the silicon test system with $(3 \times 3 \times 3)$ k points, 100 eV plane wave cutoff energy and 89 bands. For (a), the maximum frequency is 150 eV and for (b), the linear grid spacing is 0.1 eV. E^{Γ}_{gap} is found to be well converged (within 20 meV) for $\Delta\omega = 0.1$ eV and $\omega_{lin} = 50$ eV. This corresponds to 548 frequency points. ω_{max} hardly effects the results as long as it is larger than 100 eV (the energy difference between the highest and the lowest band).

cutoff may lead to a wrong value, which seems converged with respect to the number of bands, N_b. Therefore, N_b should always be adapted to E_{cut}.

These observations allow for a general strategy for convergence tests: One series of calculations with varying E_{cut} for a low k-point sampling and another series with increasing number of k points for a fixed (low) value of the cutoff energy. It is convenient to check the convergence for the non-selfconsistent HF bandstructure, which is usually fast and easy to do. Thereby, the computational efforts can be minimized. From these results, the 'optimal' parameters for the actual GW calculation can be determined. The number of bands included in the evaluation of the self-energy should be chosen so that the energy of the highest band is close to the plane wave cutoff energy. This is the default option. The use of the Plasmon Pole approximation is about 5-20 times faster than the full frequency dependent method. Its quality, however, needs to be checked for every system.

These observations only serve as a rough guideline. For different materials, the convergence behavior can change significantly. They will be one central topic in the following chapters.

To conclude this chapter, a typical input script and the corresponding output are shown in Figs. 2.10 and 2.11. The direct QP band gap of bulk silicon can be read off from the last lines of the output as 3.28 eV, which is very close to the experimental value of 3.40 eV [34] and in good agreement with other implementations [23, 35]. In order to determine the indirect gap, a finer k-point sampling should be used.

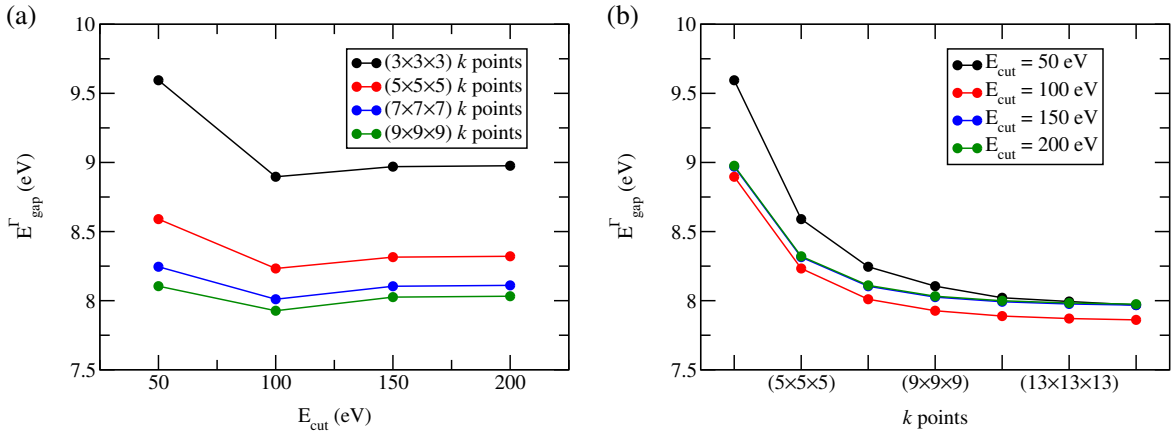


Figure 2.8: Hartree-Fock Γ-point band gap of bulk silicon as function of the (a) plane wave cutoff energy and (b) k-point sampling. The calculations were performed non-selfconsistently from LDA wavefunctions and eigenvalues. Good convergence is reached for $E_{\text{cut}} = 150\,\text{eV}$, while a k-point sampling of at least $(9 \times 9 \times 9)$ is required.

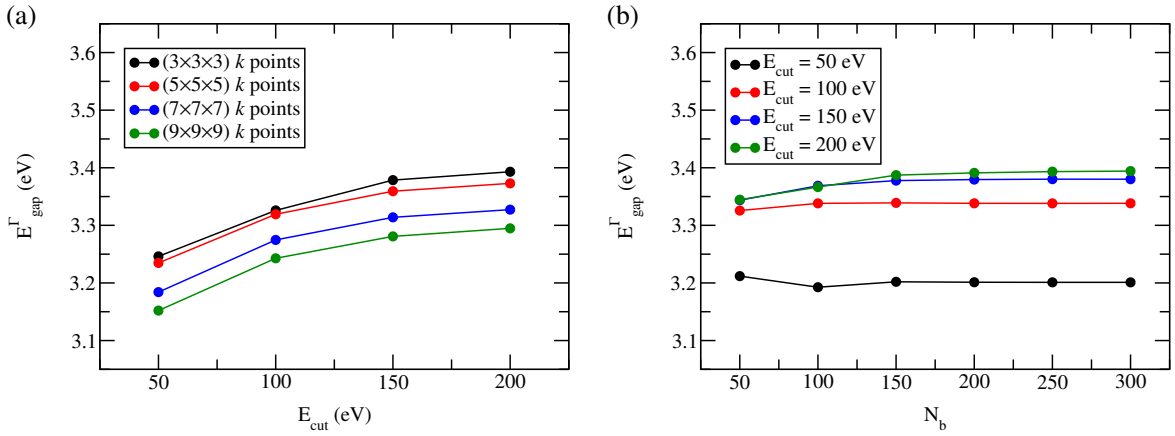


Figure 2.9: G_0W_0 Γ-point band gap of bulk silicon as function of (a) plane wave cutoff energy and (b) number of bands. For (a) the number of bands equal the number of plane waves corresponding to E_{cut}, while $(3 \times 3 \times 3)$ k points were used for (b). All calculations were performed with LDA wavefunctions and eigenvalues as starting point. The exact exchange contributions were determined seperately with a higher, fixed value of E_{cut}. That means, that the curves shown depend on the correlation part of the self-energy only (for a given k-point sampling). In comparison to Fig. 2.8, the scale on the y-axis is much smaller.

```
import numpy as np
from ase.structure import bulk
from gpaw import GPAW, FermiDirac
from gpaw.wavefunctions.pw import PW
from gpaw.response.gw import GW

a = 5.431
atoms = bulk('Si', 'diamond', a=a)

calc = GPAW(mode=PW(200),
            kpts=(9,9,9),
            xc='LDA',
            eigensolver='cg',
            occupations=FermiDirac(0.001),
            txt='Si_groundstate_k9.txt')

atoms.set_calculator(calc)
atoms.get_potential_energy()

calc.diagonalize_full_hamiltonian()
calc.write('Si_groundstate_k9.gpw','all')

gw = GW(file='Si_groundstate_k9.gpw',
        nbands=None,
        bands=np.array([2,3,4,5]),
        kpoints=None,
        ecut=150.,
        ppa=True,
        txt='Si_GW_k9_ecut150.out')

gw.get_exact_exchange()

gw.get_QP_spectrum()
```

Figure 2.10: Example script for a GW calculation in GPAW for bulk silicon. A plane wave basis up to a kinetic energy of 200 eV and the LDA functional is used for the groundstate. The G_0W_0 bandstructure is evaluated for all k points in the irreducible Brillouin zone for a $(9 \times 9 \times 9)$ k-point sampling and the two highest valence and two lowest conduction bands. The plane wave cutoff is 200 eV (as given by the groundstate calculation) for the exact exchange contributions and 150 eV for the self-energy.

```
GPAW version 0.9.1.10481
------------------------------------------------
GW calculation started at:
Tue Aug 20 00:10:28 2013
------------------------------------------------
Use eigenvalues from the calculator.
------------------------------------------------
calculating Exact exchange and E_XC
Use planewave ecut from groundstate calculator:  200.0 eV
------------------------------------------------
non-selfconsistent HF eigenvalues are (eV):
[[[   4.16253746   4.16271098  12.19495873  12.19495872]
   ...
  [   2.42342032   2.4234209   10.83576913  13.33986919]]]
Lowest eigenvalue (spin=0) :  -6.831460 eV
Highest eigenvalue (spin=0):  148.298072 eV

Plane wave ecut (eV)          :  150.0
Number of plane waves used    :  169
Number of bands               :  169
Number of k points            :  729
Number of IBZ k points        :  35
Number of spins               :  1

Use Plasmon Pole Approximation
imaginary frequency (eV)     :  27.21
broadening (eV)              :  0.10

Coulomb interaction cutoff :  None

Calculate matrix elements for k = :
[ 0.  0.  0.]
 ...
[ 0.44444444 0.44444444 0.44444444]

Calculate matrix elements for n = :
[2 3 4 5]

calculating Self energy
   Finished iq 0 in 0:24:23, estimated 18:41:31 left.
   ...
   Finished iq 45 in 18:25:56, estimated 0:15:34 left.
W_wGG takes 14:59:26
Self energy takes 3:26:31
```

```
reading Exact exchange and E_XC from file
-------------------------------------------------
Kohn-Sham eigenvalues are (eV):
[[[  5.13518868  5.1351951   7.66534713  7.66534713]
   ...
  [  3.96136002  3.96136002  6.58726608  8.45939817]]]

Occupation numbers are:
[[[ 2.00000000e+00 2.00000000e+00 7.44015195e-44 7.44015195e-44]
   ...
  [ 2.00000000e+00 2.00000000e+00 7.44015195e-44 7.44015195e-44]]]

Kohn-Sham exchange-correlation contributions are (eV):
[[[-13.52403727 -13.52403428 -11.78313739 -11.78313739]
   ...
  [-13.20283986 -13.20283986 -12.62234213 -10.9711573 ]]]

Exact exchange contributions are (eV):
[[[-14.49668849 -14.49651839  -7.25352579  -7.2535258 ]
   ...
  [-14.74077956 -14.74077899  -8.37383909  -6.09068628]]]

Self energy contributions are (eV):
[[[ 0.43563965  0.43563883 -4.09443116 -4.09443151]
   ...
  [ 0.93440746  0.93438013 -3.82047071 -4.44281029]]]

Renormalization factors are:
[[[ 0.77244275  0.77244309  0.77209811  0.77209854]
   ...
  [ 0.76569843  0.76569572  0.78009198  0.77452521]]]

GW calculation finished in 18:57:06
-------------------------------------------------
Quasi-particle energies are (eV):
[[[  4.72037799   4.72051267   8.00134912   8.00134904]
   ...
  [  3.49923634   3.49921748   6.92117067   8.79837744]]]
```

Figure 2.11: (Abridged) output from the example script. The calculation ran on two 8-core HP DL160 G6 nodes with two 64-bit Intel Nehalem Xeon X5570 quad-core CPUs each running at 2.93 GHz [33]. Results are sorted by spins (blocks), k points (rows) and bands (columns).

CHAPTER 3

SOLIDS

Any new implementation needs to be tested thoroughly and compared with other exisiting codes. The GW method has been known for more than 50 years and has been applied to real systems since the late 1980's. Up to today, a large number of results from GW calculations for simple materials have been well established in literature. Only about 10 years ago, progress in the development of computational resources made it possible to investigate more complex structures and perform calculations for a broader range of systems. But even though the GW method is now a standard tool in many electronic structure codes and its advantages and drawbacks have been intensively discussed, it has not been used in systematic studies in the same way as traditional methods, namely DFT, have. This is mostly due to its immense computational requirements. However, as more powerful supercomputers become available, this is starting to change and numerous applications are gaining interest.

This chapter starts with a discussion of results for a number of simple semiconductors and insulators – systems which have been extensively studied both in theory and by experiment – and an assessment of the different approximations. Band gaps obtained at the G_0W_0@LDA level are in very good agreement with results from literature and experimental data. The Plasmon Pole approximation is found to perform very well, whereas the static COHSEX fails completely. LDA as a standard DFT functional drastically underestimates band gaps, due to two main problems: First, it contains large self-interaction errors. These can be reduced if hybrid functionals are used instead. Hartree-Fock on the other hand is self-interaction free, but completely neglects correlation effects and thereby overshoots gaps. Secondly, density functional theory suffers in general from the so-called band gap problem. Here, the band gap is defined as the difference between Kohn-Sham energies at the conduction band minimum and the valence band maximum, which is not equal to the fundamental or quasiparticle band gap. The difference is given by the derivative discontinuity, Δ_{xc} [4].

Of special interest is therefore a comparison with the GLLBSC potential [36, 37], which is non-local and allows for the calculation of Δ_{xc}. With only slight additional computational effort, it cures some of the main deficiencies of DFT.

Investigation of the quasiparticle bandstructure illustrates that the main effect of the GW

approximation lies in an almost constant shift of the occupied and unoccupied bands, similar to the application of a scissors operator.

Following is a G_0W_0 study of gold as one example of a metallic system, where the band-structure is effected in an inhomogeneous way.

Furthermore, the GW method is applied to a series of layered perovskites, which are potential new candidates for photocatalytic water splitting. In order for a semiconductor to be capable of converting sunlight into electrical energy, one requirement is that its electronic band gap lies within a certain energy window. The G_0W_0 calculations help to confirm results from a vast screening study, performed with the GLLBSC potential. In this way, efficiency and accuracy can be combined in the search for new materials.

Finally, a comparison of calculated band gaps with G_0W_0, GLLBSC and the hybrid functional HSE06 is made for 20 randomly chosen materials with different geometry and chemical properties, for which no experimental data is available, providing an estimate of the accuracy of the different methods.

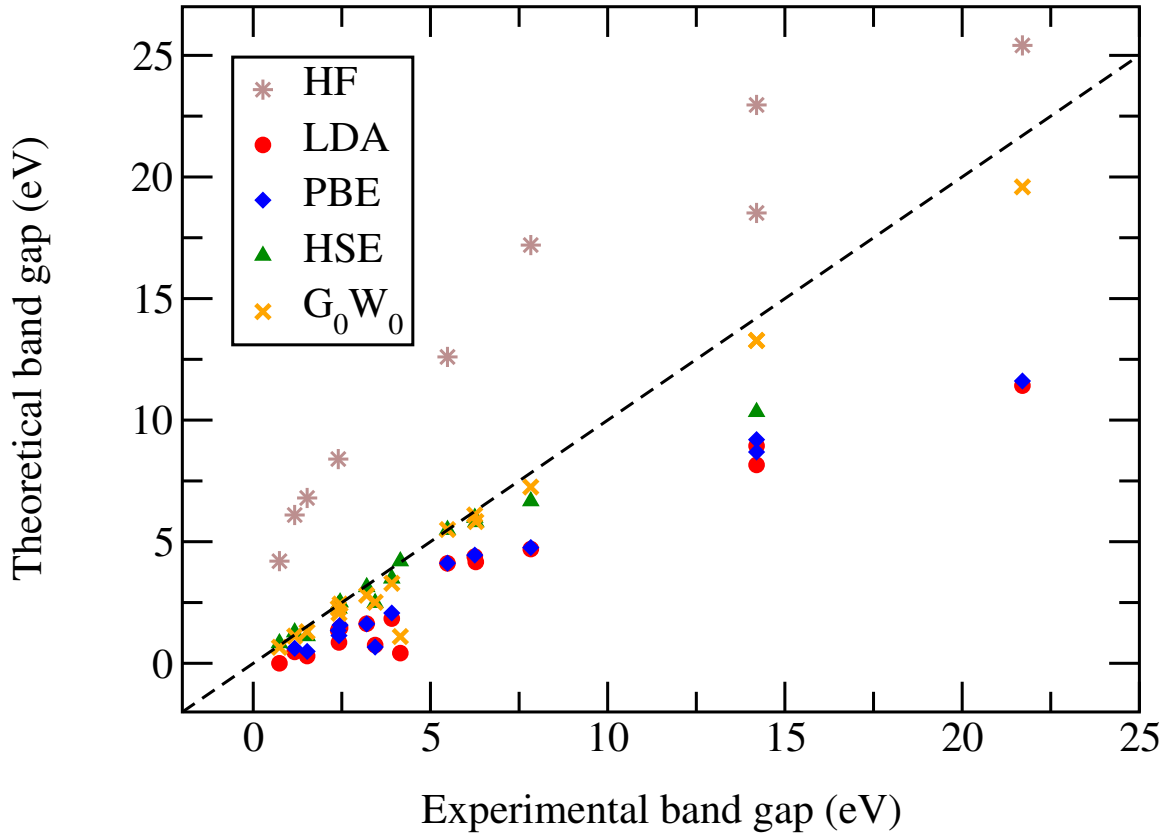


Figure 3.1: Calculated vs. experimental band gaps for Ne, Ar, LiF, MgO, AlN, BN, C, NiO, ZnS, ZnO, GaN, AlP, CdS, SiC, GaAs, Si and Ge (from right to left in descending order). Data taken from Refs. [38] and [39].

3.1 Semiconductors and insulators

3.1.1 Band gaps

The GW approximation is a true *ab-initio* method and therefore universally applicable for electronic structure calculations. This is in contrast to DFT, in which a thoughtful choice of the functional has to be made, depending on the kind of system and the properties one is interested in. Most functionals are designed to meet special requirements, often by fitting them to experimental data of a test set. This leaves them with a big uncertainty when applied to new materials and errors are often hard to estimate. On the other hand, the GW approximation is based on fundamental physical observations. As pointed out in Sec. 1.4, its validity is well justified for weakly correlated systems, in which correlation is dominated by screening effects. Screening is expected to play a particularly large role for systems with a small band gap, while it diminishes for large gaps. Therefore, GW acts as an intermediate between many of the DFT functionals, which overestimate screening and thereby underestimate band gaps, and Hartree-Fock theory, which does not contain screening at all. HF band gaps are typically found to be much too large. An interesting alternative is given by the hybrid functionals, in which a fraction of the DFT exchange is replaced by exact exchange and which thus balance between the two

Table 3.1: Crystal structures for the ten semiconductors and insulators studied.

	structure	lattice constant in Å
Si	diamond	5.431
InP	zincblende	5.869
GaAs	zincblende	5.650
AlP	zincblende	5.451
ZnO	zincblende	4.580
ZnS	zincblende	5.420
C	diamond	3.567
BN	zincblende	3.615
MgO	rocksalt	4.212
LiF	rocksalt	4.024

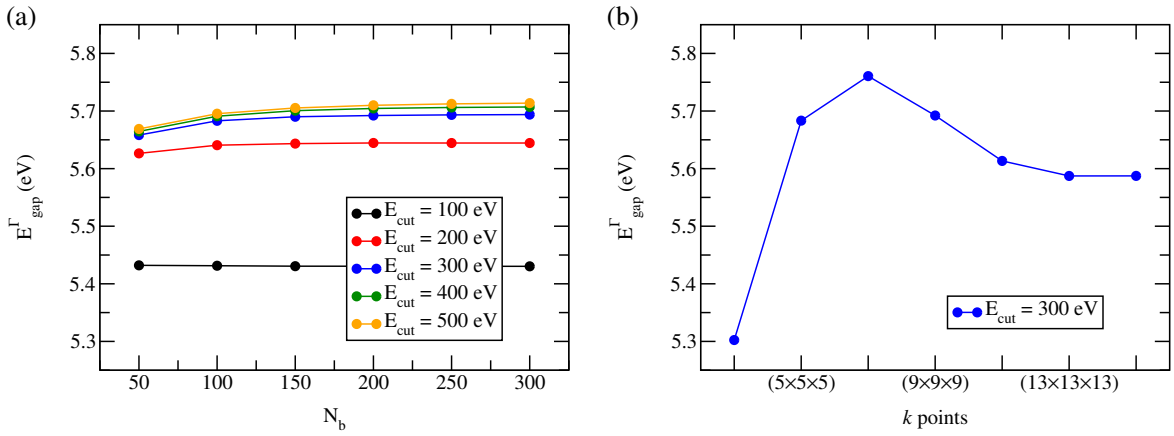


Figure 3.2: Convergence of the direct QP band gap of diamond with respect to (a) plane wave cutoff energy and number of bands for a $(9\times9\times9)$ k-point sampling and (b) number of k points for 200 bands.

opposites. Typical results from literature are shown in Fig. 3.1 for Hartree-Fock, the local density approximation (LDA), the generalized gradient approximation PBE functional by Perdew, Burke and Ernzerhof [40], the range-seperated hybrid HSE functional by Heyd, Scuseria and Ernzerhof [41] and G_0W_0@PBE. They clearly illustrate the above mentioned problems of DFT and HF.

As a first test application for the GW implementation in GPAW, ten different semiconductors and insulators with band gaps ranging from 1 to 15 eV of both direct and indirect nature have been chosen. Their geometric structures are listed in Table 3.1. For the groundstate calculations and evaluation of the exact exchange contributions, a plane wave basis set with a kinetic energy up to 800 eV was used. All GW calculations were performed with LDA wavefunctions and eigenvalues as starting point, that means G_0W_0@LDA. Convergence with respect to k points, plane wave cutoff energy and number of bands was checked for all systems individually. This is shown for the direct band gap of C as one example in Fig. 3.2. For all systems, convergence within around 20 meV is reached for $E_{\mathrm{cut}} = 200 - 300$ eV and a few hundred empty bands. A k-point sampling of $(9\times9\times9)$ was found sufficient for InP, GaAs, ZnO, ZnS, MgO and LiF,

Table 3.2: Band gaps in eV. The type of gap is indicated in the last column. The last row gives the mean absolute error compared to experiment. Experimental data is taken from Ref. [53], except for InP, which is taken from Ref. [44]. Note that the experimental value for ZnO refers to the wurtzite structure. The band gap of zincblende ZnO, however, is supposed to be very similar (see text).

				G_0W_0@LDA					
	LDA	HF@LDA	PBE0	COHSEX	PPA	dyn	GLLBSC	experiment	
Si	0.48	5.26	1.70	0.56	1.09	1.13	1.06	1.17	indirect
InP	0.48	5.51	1.92	1.99[a]	1.38	1.36	1.53	1.42	direct
GaAs	0.38	5.46	1.82	3.77[b]	1.76	1.75	1.07	1.52	direct
AlP	1.47	7.15	2.92	1.88	2.38	2.42	2.78	2.45	indirect
ZnO	0.60	10.42[c]	3.07	0.10	2.20	2.24	2.32	3.44	direct
ZnS	1.83	9.43	3.94	1.52	3.28	3.32	3.65	3.91	direct
C	4.12	11.83	6.01	6.51	5.59	5.66	5.50	5.48	indirect
BN	4.41	13.27	6.62	7.08	6.30	6.34	6.78	6.25	indirect
MgO	4.59	14.84	7.28	10.30	7.44	7.61	8.30	7.83	direct
LiF	8.83	21.86	12.25	16.02	13.64	13.84	14.93	14.20	direct
MAE	2.05	5.74	0.56	1.59	0.35	0.31	0.41	.	

[a] COHSEX predicts an indirect band gap of 1.73 eV.
[b] COHSEX predicts an indirect band gap of 1.07 eV.
[c] HF predicts an indirect band gap of 9.73 eV.

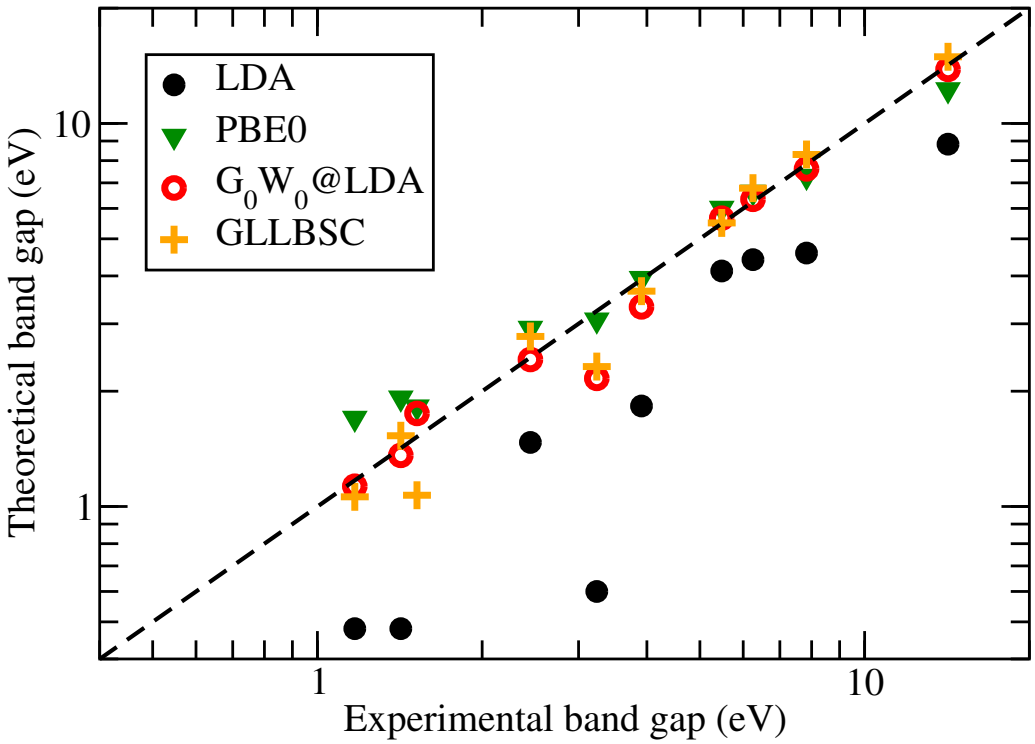


Figure 3.3: Calculated band gaps for the different methods described in the text

which all have a direct band gap at the Γ-point, whereas $(15\times15\times15)$ k points were used for Si, AlP, C, BN, which have an indirect band gap. The finer k-point sampling also ensures a good resolution of the bandstructure to determine the valence band maximum and conduction band minimum.

A frequency grid with typically 1000 to 3000 grid points was used for the full frequency dependent method. The broadening parameter η was set to 0.2 eV in the Plasmon Pole approximation and 10^{-4} eV in the static COHSEX approximation.

Additional calculations were done with non-selfconsistent HF from LDA orbitals and energies and with PBE0, a hybrid method based on the PBE functional, in which 1/4 of the PBE exchange contributions are non-selfconsistently replaced by exact exchange. Further comparisons were made with self-consistent GLLBSC calculations [37], which use the PBEsol correlation potential [42] and an efficient orbital-dependent approximation to the exact exchange optimized effective potential [43].

The calculated band gaps are summarized in Table 3.2 and Fig. 3.3 and compared to experimental reference data. In all cases, LDA drastically underestimates the band gap by up to a factor of 4, as for GaAs. It performs poorly for both the systems with low and with large gaps with a mean absolute error (MAE) of 2 eV and a mean relative error (MRE) of around 50 %. Even larger errors are found for Hartree-Fock with 5.7 eV (180 %) on average. The highest deviations occur for Si, InP and GaAs. Screening is particularly important in these systems with small gaps. The relative error decreases from 350 % for Si to 50 % for LiF. The results for PBE0 lie inbetween these two extremes, as expected for a hybrid method. Quantitative agreement, however, is still poor, in particular for systems with small band gaps. The static COHSEX approximation seems to give a slight improvement over LDA with 1.5 eV MAE. However, some results are even qualitatively wrong. For example, it predicts indirect band gaps for InP and

Table 3.3: Comparison of G_0W_0 band gaps. Details about methodological differences are described in the text.

	present	Ref. [39]	Ref. [45]	experiment
Si	1.13	1.12	0.90	1.17
InP	1.36	.	1.25	1.42
GaAs	1.75	1.30	1.31	1.52
AlP	2.42	2.44	2.15	2.45
ZnO	2.24	2.12	.	3.44
ZnS	3.32	3.29	3.24	3.91
C	5.66	5.50	5.49	5.48
BN	6.34	6.10	.	6.25
MgO	7.61	7.25	6.77	7.83
LiF	13.84	13.27	.	14.20
MAE	0.31	0.43	0.38	

GaAs, whereas ZnO becomes almost metallic. Also for ZnS, the gap is even smaller than with LDA. This means that the inclusion of only static screening is clearly insufficient. On the other hand, G_0W_0 results are in very good agreement with experiment when dynamical screening is taken into account, both with PPA and full frequency dependence with mean errors of 0.35 eV (9,5 %) and 0.31 eV (8,6 %), respectively. For all systems studied, the PPA turns out to be an excellent approximation to the frequency dependence of the dielectric function with obtained band gaps deviating by 0.2 eV at most for LiF.

Table 3.3 shows a comparison with two similar GW studies. In Ref. [39], the full frequency-dependent G_0W_0 was used with PBE wavefunctions and eigenvalues as starting point. The implementations are to a large extent similar. One difference worth mentioning is the inclusion of core-valence interaction in the exchange part [23], which is not present in this work. The results agree very well with an average absolute deviation of 0.21 eV. The only notable differences are for GaAs and the two materials with the widest gaps, MgO and LiF, probably due to the choice of a different starting point. The band gaps of Ref. [45] were obtained with full frequency-dependent G_0W_0@LDA and are systematically smaller by 0.31 eV on average. That implementation uses a mixed basis set in an all-electron linear muffin-tin orbital framework, which is fundamentally different from the PAW method. Better quantitative agreement can therefore not be expected. Other recent works [46–48] report band gaps in the same range as the values presented here.

Overall, the G_0W_0 approximation gives very good band gaps for most of the semiconductors and insulators in this study. Noticeable deviations from the experimental values are only found for LiF and ZnO where the calculated gaps are about 0.4 and 1.2 eV too small, respectively. For LiF, results are very sensitive to changes in the lattice constant. A slightly smaller lattice constant of 3.972 Å, which is the experimental value corrected for zero-point anharmonic expansion effects [49], yields a 0.4 eV larger band gap.

The quasiparticle band gap of ZnO has been intensively discussed in literature. Calculated G_0W_0 are typically 1 eV too low [50–53], both in the zincblende and the wurtzite structure. The results presented here are for the zincblende structure only, although this phase is not found to be

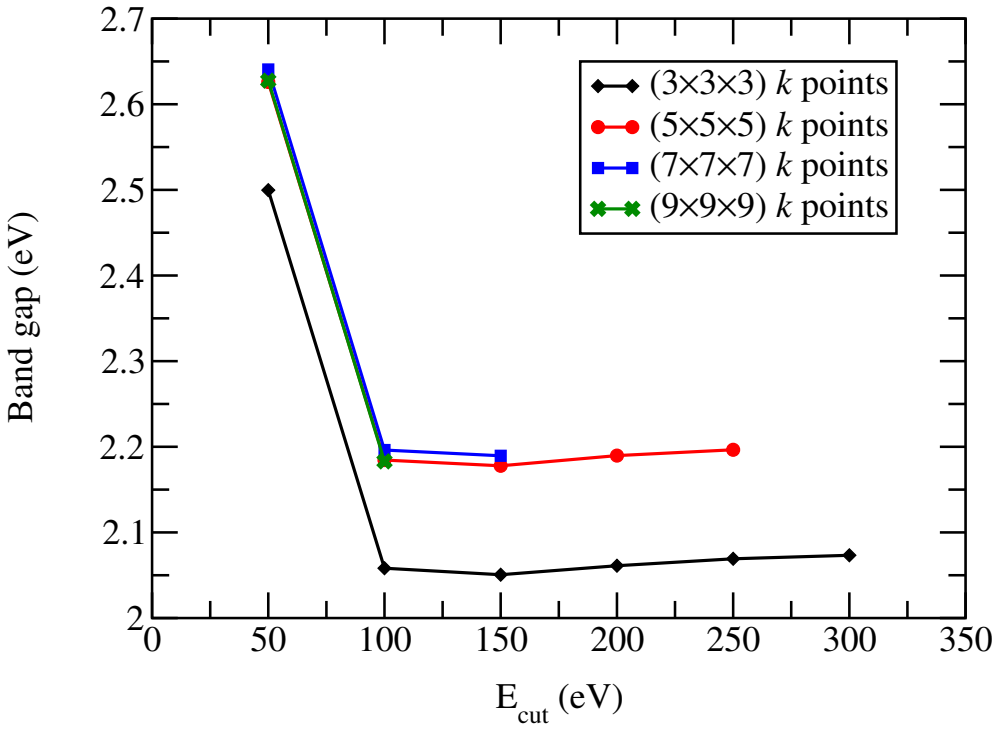


Figure 3.4: Convergence of the band gap of zinc oxide for G_0W_0@LDA with the plasmon pole approximation. The number of bands is chosen equally to the number of plane waves corresponding to the respective cutoff energy, i. e. 300 eV equal ~ 1100 plane waves and bands.

stable in experiment [54]. It was chosen for consistency with Ref. [39]. Additional calculations for the wurtzite structure gave around only 0.1 eV smaller gaps, both with LDA, G_0W_0@LDA and GLLBSC. Thus, all conclusions made here are expected to hold for the wurtzite structure as well. In Refs. [55] and [56], the low value of the QP band gap was attributed to a very slow convergence of the self-energy with respect to the number of bands. Ref. [55] finds a converged band gap of 3.4 eV for 3000 bands using pseudo-potentials, whereas the all-electron calculations of Ref. [56] yield only 2.99 eV after extrapolation to infinite number of bands. Both studies used the LDA functional as starting point. In the present work, good convergence is already reached for $E_{\text{cut}} = 100$ eV and around 200 bands, as can be seen in Fig. 3.4. A major technical difference between the current implementation and the one of Ref. [55] is the model used for the Plasmon Pole approximation, namely the method by Godby and Needs in contrast to the one by Hybertsen and Louie [35]. The latter fits the model dielelectric function only to the static limit, but requires Johnson's f-sum rule [57] to be fulfilled. However, as pointed out in Ref. [58], this can lead to a wrong convergence behavior and too large band gaps for ZnO as compared to evaluating the frequency dependence explicitely. This is consistent with the results presented here, where the PPA reproduces the results from the full frequency-dependent method. The calculations in Ref. [56] in a full-potential linearized augmented-plane-wave method on the other hand depend strongly on the special set of basis functions chosen. In Ref. [59], an effective-energy technique was applied, which avoids the summations over an in principle infinite number of empty states. In this way, a band gap of 2.56 eV was obtained. The drastic underestimation of the band gap of ZnO thus seems to be a systemtatic error of

G_0W_0 on top of LDA or GGA. In Ref. [60], this was explained by a wrong positioning of the *d*-bands. With an onsite DFT+U correction, a band gap of 3.30 eV was found, in very good agreement with experiment. Good results were also achieved using HSE as starting point, yielding 3.22 eV [60] and 3.14 eV [61] and eigenvalue self-consistent GW calculations with 3.20 eV [39].

The orbital-dependent GLLBSC potential gives overall good band gaps with a mean average error of 0.41 eV compared to experiments and close to the GW results for most systems. Only for GaAs, the band gap is found considerably lower, while it is too large for MgO and LiF. As with GW, the band gap of ZnO is underestimated by more than 1 eV. In all GLLBSC calculations, the derivative discontinuity is added explicitely to the eigenvalues of the unoccupied bands. This shift is around 50 % of the Kohn-Sham band gap for the investigated structures and has therefore an essential impact on the bandstructure.

3.1.2 Bandstructures

The bandstructure of diamond is shown in Fig. 3.5 (a) for LDA and G_0W_0@LDA. With both methods, the valence band maximum is at the Γ point and the conduction band minimum close to the X point along the Γ-X direction giving an indirect band gap of 4.12 eV and 5.66 eV, respectively. With the GW approximation, the valence bands are shifted down in energy by around 0.7 eV. The opposite holds for the conduction bands, which are shifted up in energy by roughly the same value. Thus, the effect of the GW self-energy is an opening of the band gap. These shifts are almost constant across the whole Brillouin zone, leaving the shape of the bands unchanged. This is comparable to applying a scissors operator.

The same holds for Si, as shown in Fig. 3.5 (b), with a constant downshift of the occupied bands of around 0.5 eV, and a constant upshift of the unoccupied bands of around 0.3 eV. This leads to an increase of the indirect band gap, which is located between the Γ- and a point close to X.

For GaAs, as shown in Fig. 3.5 (c), the main effect of the GW approximation can be seen in the conduction bands, whereas the valence bands remain nearly unchanged. Unlike for the previous two examples, the shift of the unoccupied bands to higher energies is not entirely constant but varies between 0.4 and 1.0 eV. The largest change occurs at the Γ-point, giving rise to a significant increase of the direct band gap. Thereby, the calculated G_0W_0 gap is exceptionally high compared to experiment and previous studies.

Overall, for a number of simple semiconductors and insulators, the self-energy can be approximated by a simple scissors operator to correct the Kohn-Sham energy spectrum in an easy and efficient way [62].

3.2 Metals

In noble metals, many-body correlation effects typically influence the bandstructure in a non-trivial way, which cannot be accounted for by a simple scissors operator. Self-energy corrections are found to be band and *k*-point dependent and may not even be of the same sign among occupied and unoccupied bands [63]. A correct quasiparticle description is thus essential in order to determine the electronic structure correctly, in particular the position and width of the *d* bands and energies of interband transitions [64, 65].

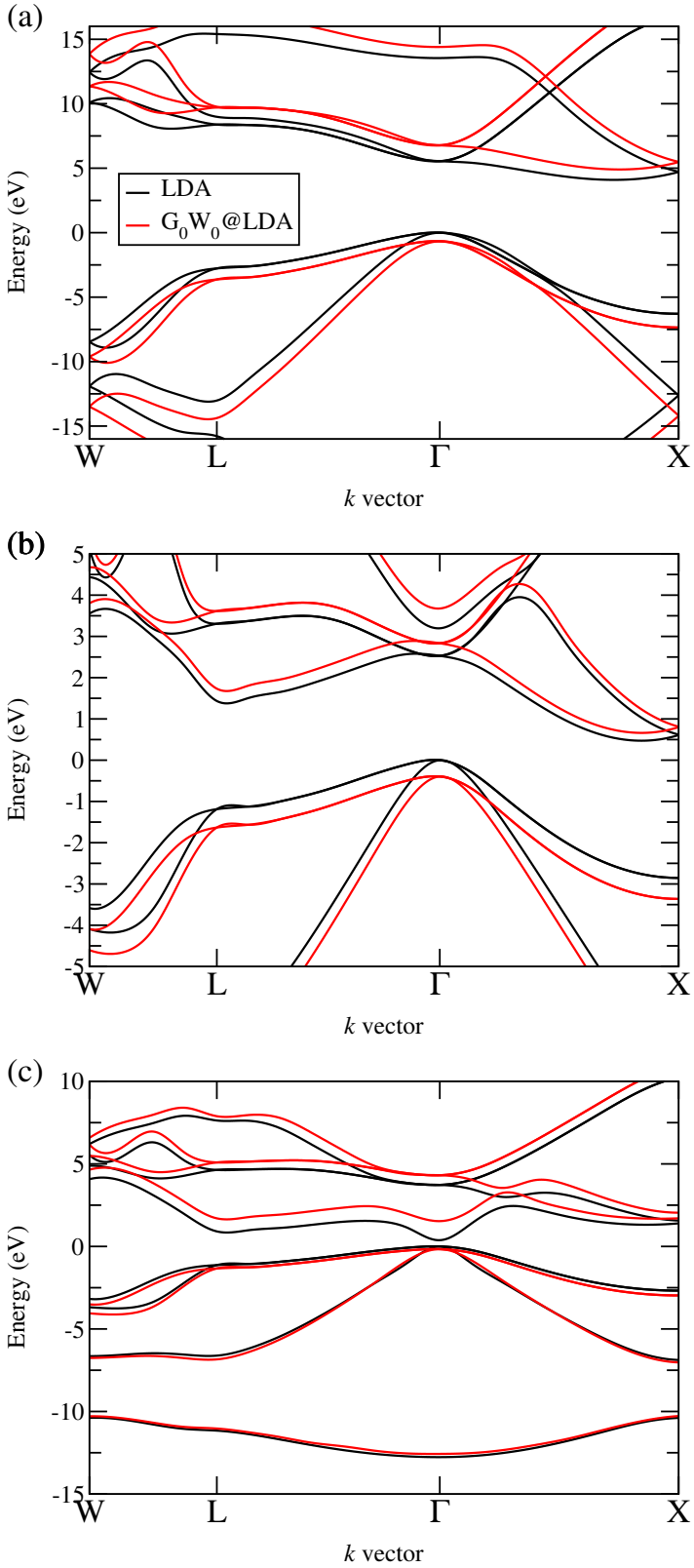


Figure 3.5: Bandstructure of (a) diamond, (b) Si and (c) GaAs, interpolated by splines from a $(15 \times 15 \times 15)$ k-point sampling. Bands are aligned to the respective LDA valence band maximum.

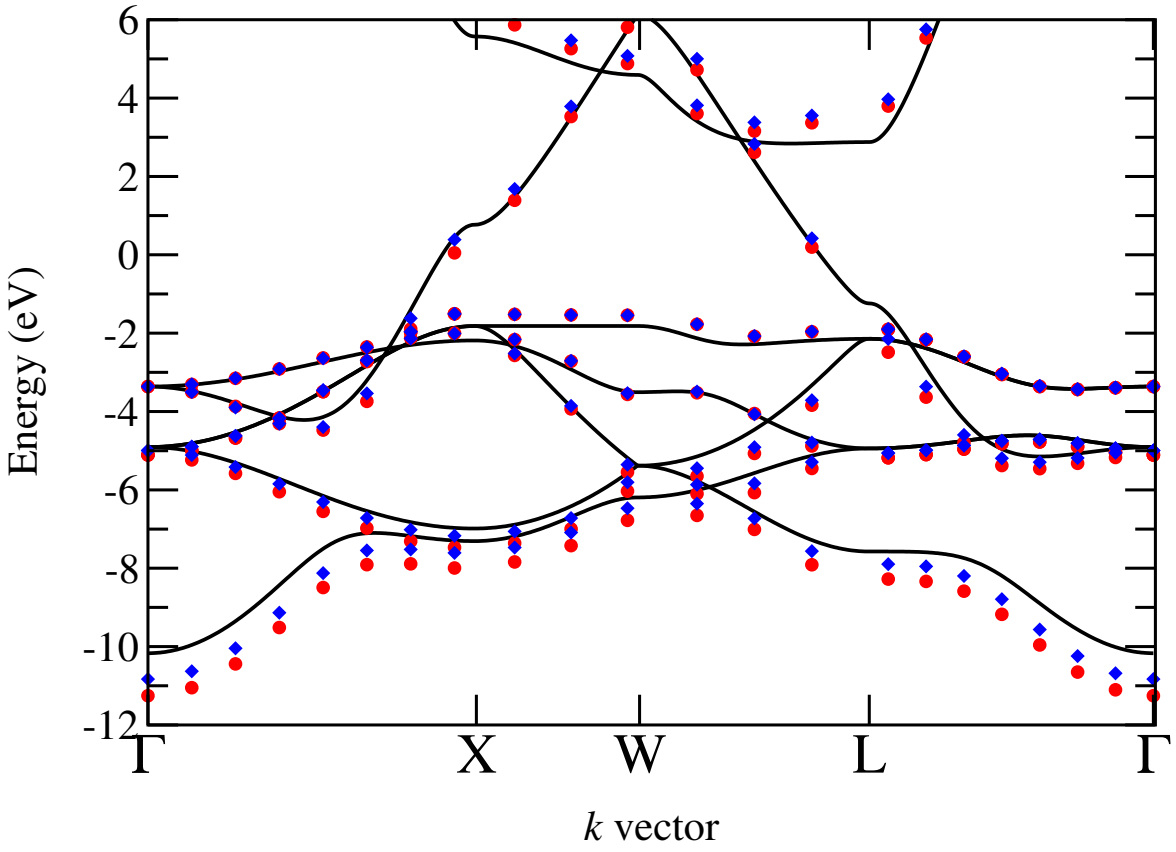


Figure 3.6: Calculated bandstructure of fcc gold with LDA (black lines) and G_0W_0@LDA with PPA (red dots) and full frequency dependence (blue diamonds). The highest occupied bands have 5 *d* character, whereas the lowest unoccupied band arises from 6 *sp* electrons. The Fermi level is set to 0.

Fig. 3.6 shows the Kohn-Sham and quasiparticle bandstructure of gold in the fcc structure with a lattice constant of 4.079 Å, obtained with the LDA functional and with G_0W_0@LDA, respectively. Only 5 *d* and 6 *sp* states were considered for valence electrons. With the GW approximation, the *d*-band width increases by up to 0.5 eV, since the upper bands are moved up in energy while the lower bands are moved down. Larger deviations up to 1 eV are found for the low-lying *s* and the unoccupied *s*-*p* band. However, the effects are significantly smaller when the full frequency dependent method is used instead of the Plasmon Pole approximation. It is clear that the shifts in energy are not constant throughout the Brillouin zone but vary for different *k* points and bands.

The effect of including 5 *sp* semicore states is shown in Fig. 3.7 with PBE wavefunctions and eigenvalues as starting point. Chosing the PBE instead of the LDA functional does not change the Kohn-Sham bandstructure. The semicore states, which lie around 50 eV deeper in energy than the valence states, have no impact on the electronic structure around the Fermi level within DFT. On the other hand, significant changes can be seen within the GW approximation, in particular a downshift of the *d* bands when the semicore states are included. These shifts arise mainly from the exact exchange contributions.

These results agree well with the calculations of Ref. [66] using pseudopotentials. In that study, it was shown that QP self-consistent GW calculations are necessary in order to bring

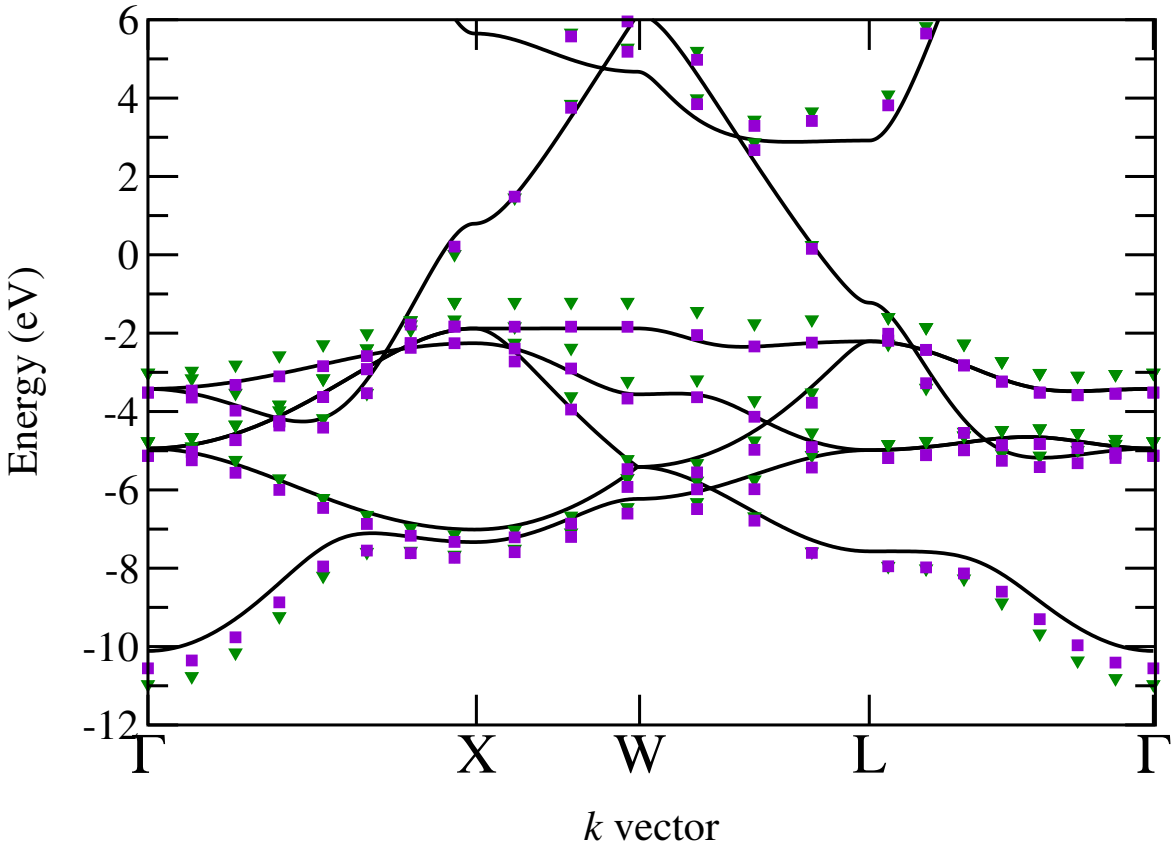


Figure 3.7: PBE (black lines) and G_0W_0@PBE bandstructures without (green triangles) and with (purple squares) inclusion of 5 *sp* semicore states. GW calculations were performed with the Plasmon Pole approximation.

the position of the *d* bands in better quantitative agreement with experiments. Also, spin-orbit coupling and relativistic effects should be considered.

3.3 Perovskites

Perovskites are a group of structures of the same kind as $CaTiO_3$ with the general chemical formula ABX_3, where A and B is a large and a small cation, respectively. X is an anion, commonly oxygen, nitrogen or halogens. In extensive screening studies, new potential candidates for one- and two-photon water-splitting could be identified among 19.000 oxides, oxynitrides, oxysulfides, oxyfluorides and oxyfluoronitrides in the cubic perovskite structure [67, 68]. One important requirement for a material to be a good photocatalyst, that means to be able to convert sunlight into chemical energy, is that the electronic band gap lies within the visible light range and is well positioned with respect to the redox potential of water. For the one-photon water-splitting process, this implies a gap between 1.5 and 3 eV with a valence band edge above 1.23 eV and a conduction band edge below 0 eV. Furthermore, high charge carrier mobility and chemical and structural stability are needed. In these studies, the GLLBSC potential was used for the calculation of the electronic bandstructure, after being tested on a set of 40 metal oxides

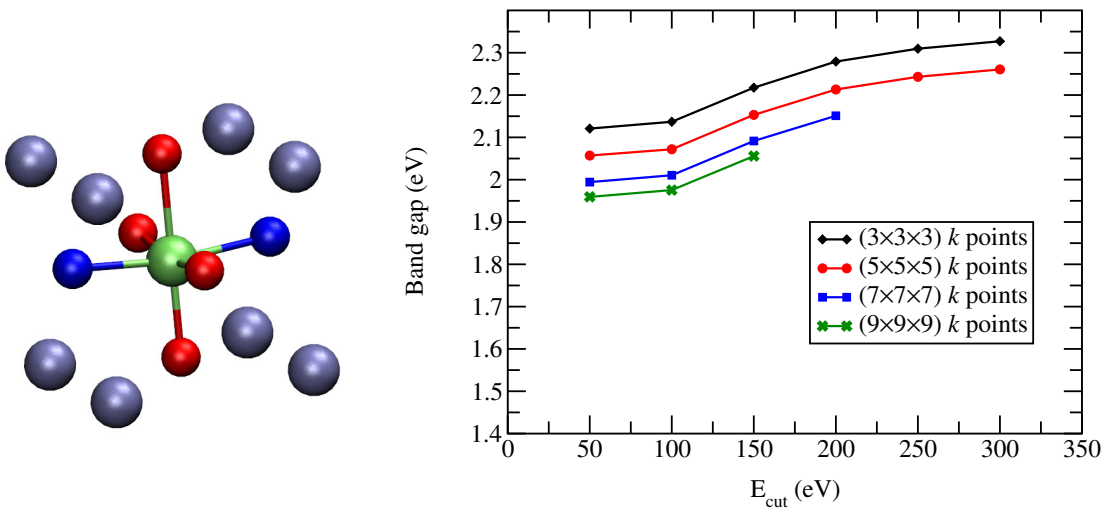


Figure 3.8: Convergence of the band gap of $LaTiO_2N$ for G_0W_0@LDA in the plasmon pole approximation with respect to the plane wave cutoff energy and number of bands, which is varied with the number of plane waves corresponding to E_{cut}.

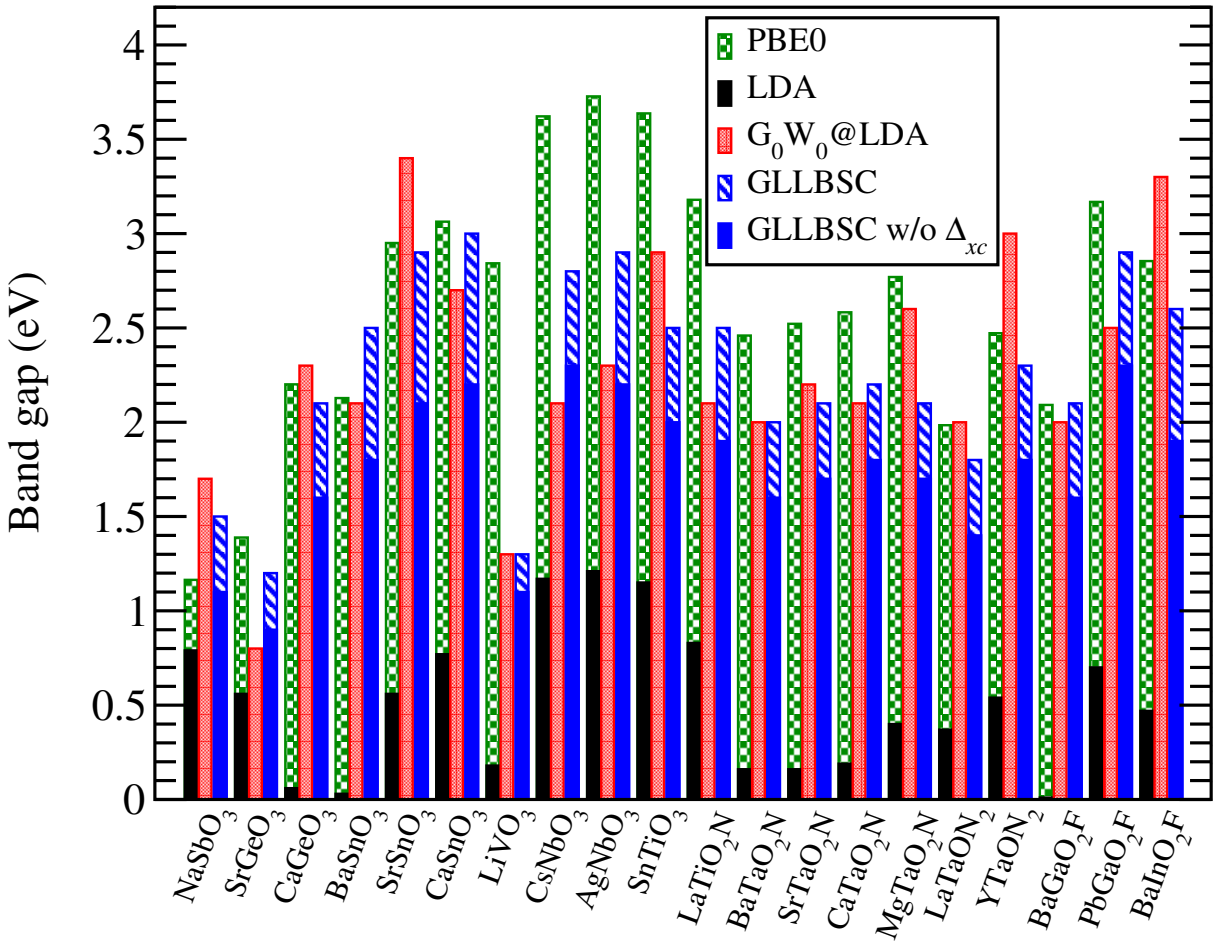


Figure 3.9: Calculated band gaps for 20 cubic perovskites. GLLBSC results are shown with and without explicit addition of the derivative discontinuity, Δ_{xc}.

for which a good overall agreement with experimental data was found [67].

In order to further assess the reliability of this method, additional GW calculations were performed for 20 cubic perovskites for which the desired electronic properties had been predicted by GLLBSC. Throughout the study, a $(7 \times 7 \times 7)$ k-point sampling, 150 eV plane wave cutoff energy and 150 bands have been used. A careful examination of the convergence of the band gap was done for $LaTiO_2N$ as one example and is illustrated in Fig. 3.8. The chosen parameters gave a good balance between efficiency and accuracy. Furthermore, the PPA was compared to the full frequency dependent method for this system, giving almost identical values for the band gap. Since the number of atoms and the symmetry group are identical for all structures, the same set of parameters was used consistently. Also, the number of valence electrons is similar in all cases. The expected accuracy is around 0.1 eV.

Fig. 3.9 shows the calculated band gaps from LDA, PBE0, G_0W_0 and GLLBSC, respectively. GLLBSC results are shown with and without the derivative discontinuity, Δ_{xc}, added to the Kohn-Sham gap. In all cases, Δ_{xc} increases the gap by around 30 % and gives therefore an essential contribution to the total gap. The total gaps from GLLBSC agree well with the quasiparticle gaps from G_0W_0 with a mean average error of 0.3 eV. In comparison, the PBE0 results deviate much stronger and are considerably higher in some cases. LDA gaps are drastically lower, indicating a very poor description of the electronic structure for all systems in this study.

In conclusion, the results suggest that the GLLBSC potential can provide a cheap way to determine electronic band gaps of novel materials, that are possible candidates for photocatalytic water splitting, with much better accuracy than convential functionals. The G_0W_0 approximation as a much more advanced method, on the other hand, is orders of magnitude slower and can therefore not be used for screening purposes. It can, however, help to confirm the outcomes of GLLBSC calculations for systems which seem especially promising.

3.4 The Materials Project database

The Materials Project database [69] contains results from computational studies of the electronic structure of materials from the ICSD database [70], which is the most complete repository for experimental data of crystal structures. It serves as a complementary source of information for materials, where no or only insufficient experimental results for the electronic structure is available and helps to analyse and compare theoretical findings. High-throughput materials design has become feasible in the last decade due to the rapid increase in computational resources. However, this is still out of reach for many-body methods like the GW approximation and therefore, cheaper methods have to be used instead, generally at the cost of reduced accuracy.

Here, the electronic structure of 20 randomly chosen compounds from the Materials Project database is calculated with G_0W_0@LDA and compared to LDA, GLLBSC and HSE06, a range-separated hybrid functional [71, 72].

Quasiparticle gaps were obtained in the G_0W_0 approximation in a plane wave representation using LDA wavefunctions and eigenvalues as input. The initial Kohn-Sham states and energies have been calculated in a plane wave basis with kinetic energies up to 600 eV. The same value is used for determining the exact exchange contributions. The GW self-energy has been carefully converged with respect to k points, number of bands and plane wave cutoff energy for each material individually. Typically, a $(7 \times 7 \times 7)$ k-point sampling, 100 - 200 eV energy cutoff and unoccupied bands up to the same energy (a few hundred bands in total) were found to be sufficient in order to converge band gaps within less than 0.1 eV. Both, the Plasmon Pole

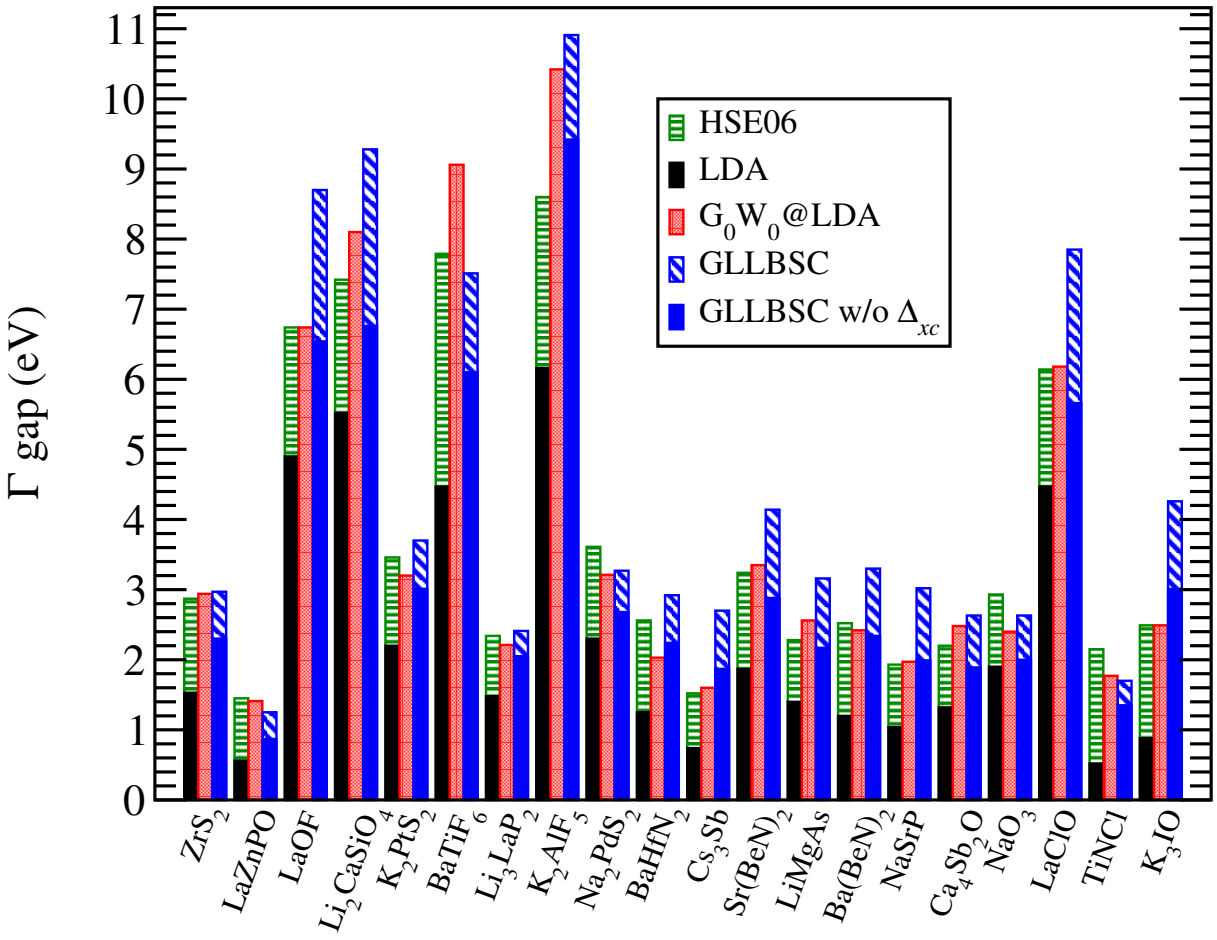


Figure 3.10: Γ point band gaps for 20 structures from the Materials Project database. The results shown for G_0W_0 were obtained with the full frequency dependent method.

Table 3.4: Mean absolute errors in eV of the Γ-point band gap of the materials from Fig. 3.10 for the different methods compared to each other.

			GLLBSC		G_0W_0	
	LDA	PBE	w/o Δ_{xc}	incl. Δ_{xc}	@LDA	HSE06
LDA	.	0.15	1.07	2.13	1.56	1.43
PBE	0.15	.	0.93	1.98	1.41	1.28
GLLBSC w/o Δ_{xc}	1.07	0.93	.	1.01	0.56	0.53
GLLBSC incl. Δ_{xc}	2.13	1.98	1.01	.	0.75	0.86
G_0W_0@LDA	1.56	1.41	0.56	0.75	.	0.34
HSE06	1.43	1.28	0.53	0.86	0.34	.

Table 3.5: Mean relative errors of the Γ-point band gap of the materials from Fig. 3.10 for the different methods compared to each other.

			GLLBSC		G_0W_0	
	LDA	PBE	w/o Δ_{xc}	incl. Δ_{xc}	@LDA	HSE06
LDA	.	0.09	0.35	0.51	0.43	0.43
PBE	0.10	.	0.30	0.47	0.38	0.38
GLLBSC w/o Δ_{xc}	0.67	0.50	.	0.25	0.15	0.16
GLLBSC incl. Δ_{xc}	1.26	1.02	0.34	.	0.23	0.26
G_0W_0@LDA	0.88	0.69	0.18	0.17	.	0.08
HSE06	0.90	0.71	0.21	0.19	0.08	.

approximation and the explicit frequency dependence of the dielectric function, $\epsilon(\omega)$, have been used, yielding almost identical results (within 50 meV) for all materials, except for NaO_3, where the full frequency dependent method gives a ~ 0.8 eV lower Γ point gap.

For all materials in this study, comparison between the different methods is shown by means of the direct Γ-point gap, in order to avoid the need for interpolation of the bandstructure in case that the minimum of the conduction band is not located at a high symmetry point in the Brillouin zone. The results are illustrated in Fig. 3.10. Mean absolute and mean relative errors for all methods compared to each other are given in Tab. 3.4 and 3.5, respectively. A very good agreement is found between G_0W_0 and HSE06 with a mean absolute error of 0.33 eV and a mean relative error of 0.08. Typically, HSE06 underestimates band gaps of semiconductors and insulators [72] as it is also the case with G_0W_0@LDA (see previous sections). In comparison, the GLLBSC results are much larger on average when the derivative discontinuity is included, while the agreement with HSE06 and G_0W_0 is reasonably well when Δ_{xc} is not added to shift the unoccupied bands. Since GLLBSC + Δ_{xc} often overestimates band gaps [67], the right values are expected to lie somewhere in between.

The bandstructures of ZrS_2 and $BaHfN_2$ are shown as two examples in Figs. 3.11 and 3.12. For ZrS_2, no considerable differences in the two highest valence and the two lowest conduction bands can be seen for GLLBSC, G_0W_0@LDA and HSE06 and the band gap at the Γ point is thus almost the same with all three methods. This is clearly different for $BaHfN_2$: While the valence bands are similar, large deviations are found for the conduction bands. Throughout the Brillouin zone, the position of the second unoccupied band is affected much stronger by the different methods than the position of the first one. In particular, the order of bands is reversed at the Γ point for G_0W_0 compared to GLLBSC, while they are almost degenerate for HSE06.

In the GLLBSC method, the derivative discontinuity, Δ_{xc}, is added as as constant shift to all unoccupied bands. As illustrated for $BaHfN_2$, this might result obviously in a deficient description of the conduction states. This might be one reason for the large discrepancies of calculated gaps for many materials of this study.

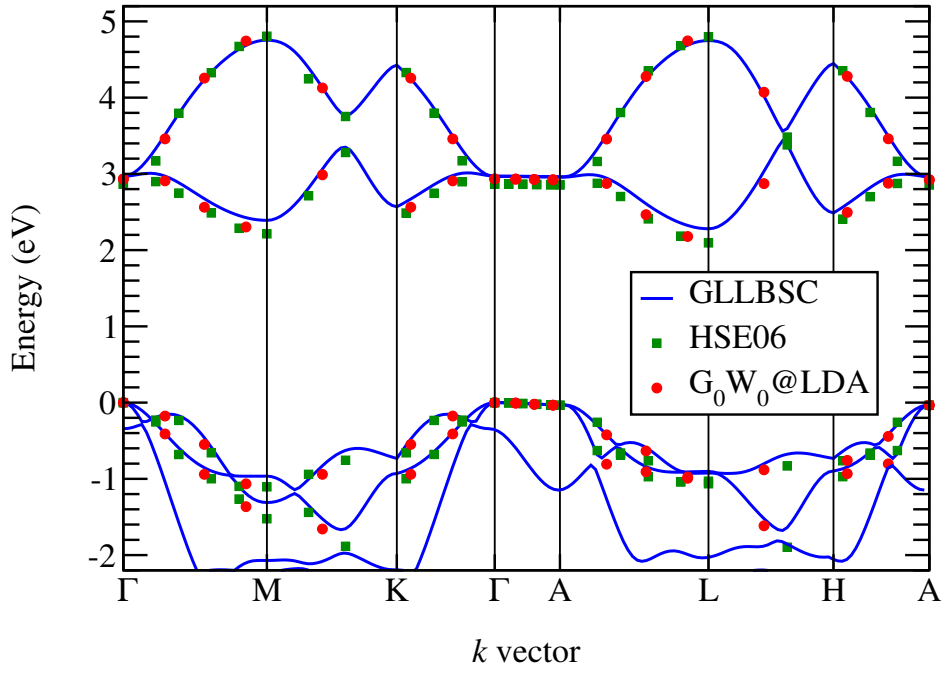


Figure 3.11: Bandstructure of ZrS_2. Bands have been aligned to the respective valence band maxima.

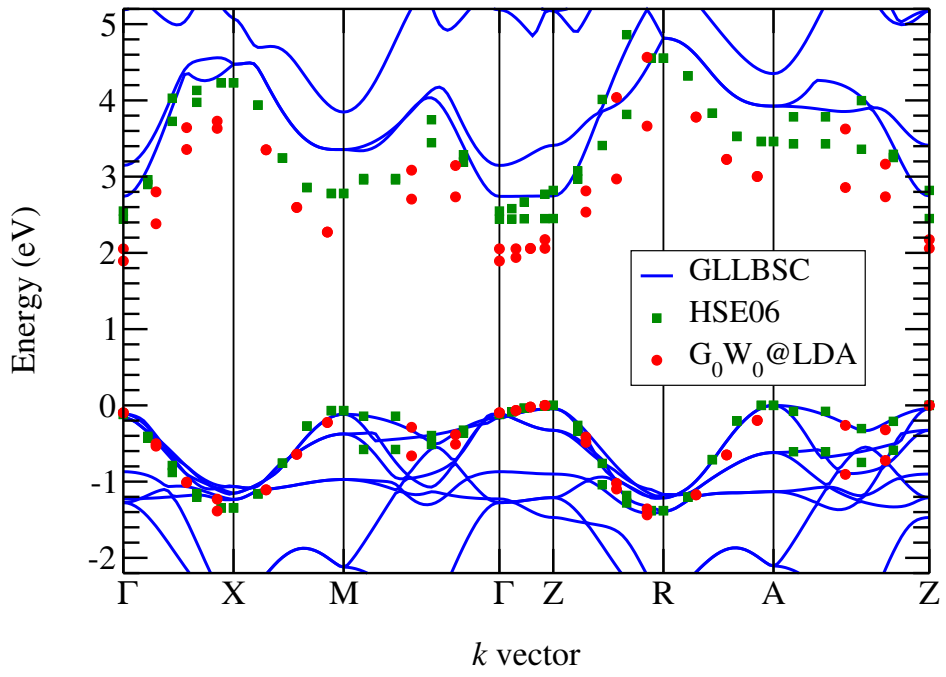


Figure 3.12: Bandstructure of $BaHfN_2$. Bands have been aligned to the respective valence band maxima.

CHAPTER 4

2D MATERIALS

2-dimensional (2D) materials are characterized by non-periodicity in one of the spatial directions. This includes not only flat monolayers, but also surfaces and interfaces, for example. The physics in these systems is often completely different than in 3 dimensions.

Dynamical screening plays a particularly important role at interfaces between semiconductors, where changes in the charge density lead to the formation of an interface dipole [73], and between metals and an insulators, where it takes the form of an image charge effect [74, 75].

One model system for studying the impact of screening and exploiting the advantages of the GW approximation over DFT is graphene/hexagonal-boron nitride (*h*-BN). Recently, it has been shown that *h*-BN serves as an ideal substrate for graphene, due to its great stability, low roughness and high charge carrier mobility [76]. Furthermore, these and similar graphene-based 2D heterostructures have opened up new possibilities for the design of novel electronic and photonic devices with unique physical properties [77–80].

In this chapter, first the Kohn-Sham and quasiparticle bandstructures of freestanding sheets of graphene and *h*-BN are discussed. Compared to LDA, G_0W_0 increases the Fermi velocity and the band gap, respectively. For boron nitride adsorbed on graphene, screening effects are illustrated by varying the distance between the two slabs and by changing the number of graphene layers. As a technical aspect, it is shown that truncating the Coulomb interaction in the perpendicular direction is crucial for GW calculations of 2-dimensional systems.

Another 2D material, which has gained much attention in the past years is single layer MoS_2 [81–84]. Whereas strong photoluminescence suggests a direct band gap in experiment, there is disagreement even on the type of gap in theory and among various GW studies, in particular. Here, a very careful analysis of the dependence of the quasiparticle gap on the cell size and the number of k points in G_0W_0 calculations is given. The results explain why previous studies found reasonable band gaps, even though a Coulomb truncation had been neglected in the calculations. The fully converged G_0W_0@LDA band gap is indirect with a value of 2.58 eV when the experimental lattice constant is used. Applying strain leads to a transition to a direct gap.

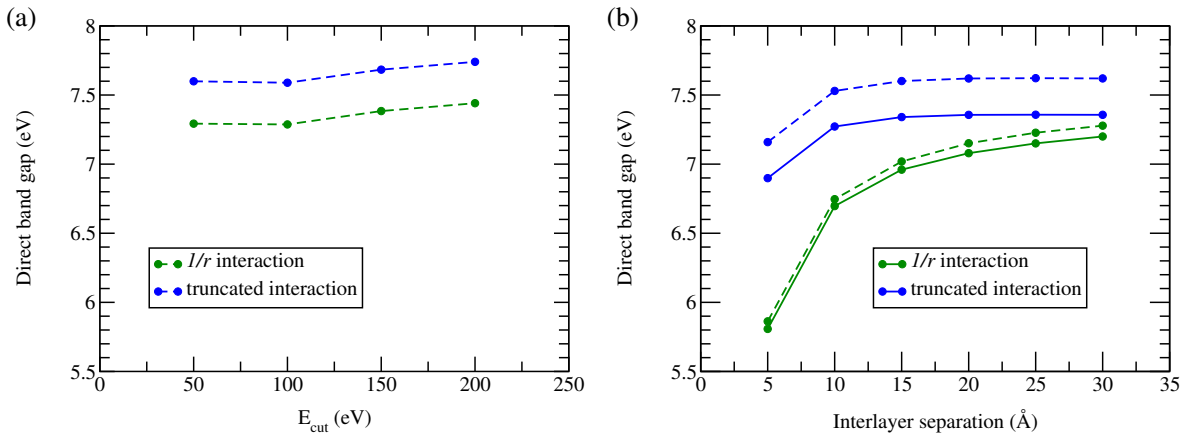


Figure 4.1: Convergence of the direct G_0W_0 band gap of a freestanding *h*-BN sheet with respect to (a) plane wave cutoff energy and corresponding number of bands and (b) interlayer separation with and without truncating the Coulomb interaction in the direction perpendicular to the slab. Dashed lines are for a $(15\times15\times1)$ and straight lines for a $(45\times45\times1)$ *k*-point sampling. The interlayer separation is defined as the distance between the centers of the slabs in neighboring supercells. Even much higher interlayer distances than the ones shown here would be required for the curve with the untruncated interaction to reach the results obtained with the Coulomb truncation.

4.1 Graphene / hexagonal-boron nitride

Graphene and hexagonal-boron nitride both form perfectly flat honeycomb structures with very similar lattice constants, which in these studies is set to the experimental value for *h*-BN of 2.503 Å [85] for both materials. In addition to the parameters discussed in the previous chapters for bulk structures, GW results need to be converged with respect to the cell size for two-dimensional materials, which defines the distance between layers in neighboring supercells. For *h*-BN, the convergence behaviour of the direct band gap at the K point, calculated with G_0W_0@LDA, as a function of the plane wave cutoff energy and the interlayer separation is shown in Fig. 4.1. For the untruncated $(1/r)$ interaction, a numerical integration around each $\mathbf{q}$ point has been used for the head of the screened potential:

$$\begin{aligned} W_{\mathbf{00}}(\mathbf{q}) &= \frac{1}{\Omega_\mathbf{q}} \int_{\Omega_\mathbf{q}} d\mathbf{q}'\, V(\mathbf{q}')\epsilon^{-1}_{\mathbf{00}}(\mathbf{q}') \\ &\approx \frac{1}{\Omega_\mathbf{q}} \epsilon^{-1}_{\mathbf{00}}(\mathbf{q}) \int_{\Omega_\mathbf{q}} d\mathbf{q}'\, V(\mathbf{q}'), \end{aligned} \tag{4.1}$$

where $\Omega_\mathbf{q}$ is a small volume around $\mathbf{q}$.

Without truncating the Coulomb interaction, the value of the band gap depends strongly on the separation and increases drastically between 5 and 15 Å. And even for 30 Å, it is far from being converged. Using a truncation of the Coulomb potential, as described in Sec. 2.1.5, corresponds to extending the interlayer separation to infinity, and thus makes the calculations largely independent of the cell size. This results in, that with the truncation, convergence is already reached for a separation of 20 Å. The calculated gap is significantly higher than without use of the truncation, even for very large separations. The very slow convergence of the quasiparticle gap and the necessity of correcting for the spurious long-range interaction in periodic supercell calculations for slabs and surfaces has previously been discussed in Refs. [86–88]. From Fig.

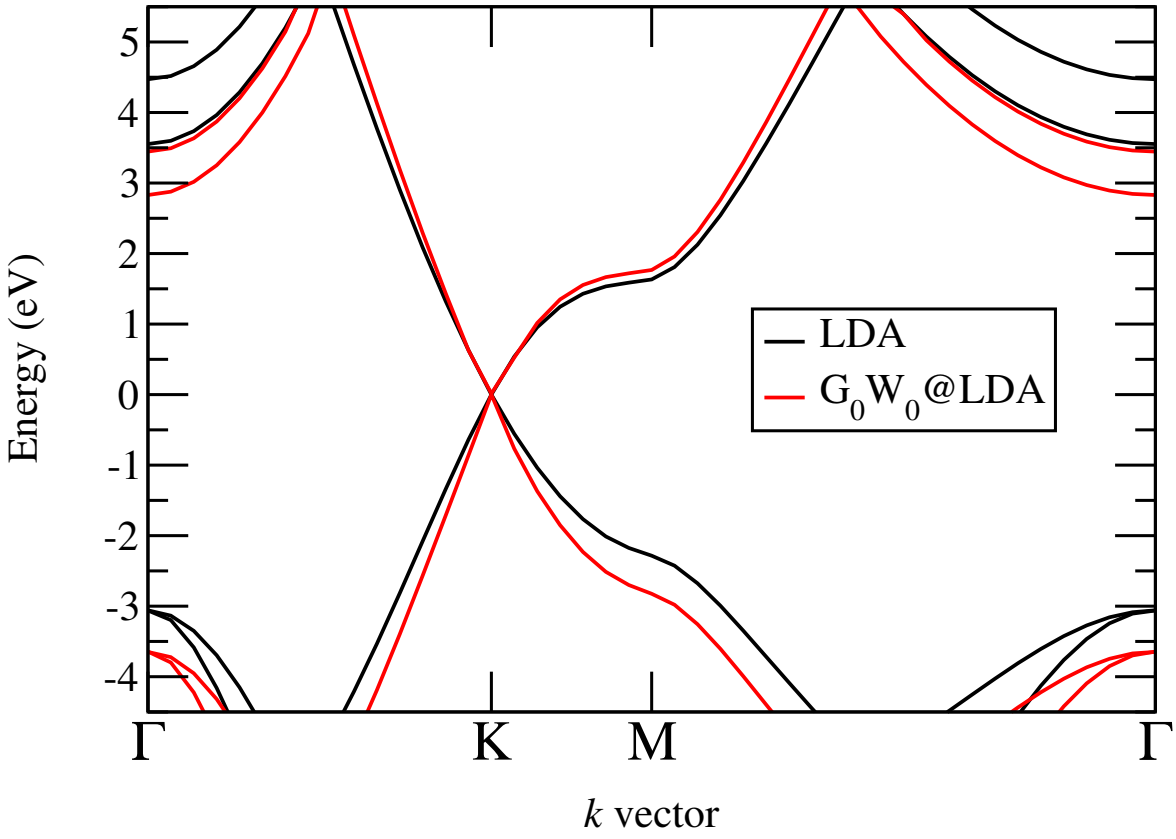


Figure 4.2: Bandstructure of graphene. The zero of energy is set to the Fermi level. Not shown are the results with GLLBSC since they lie almost completely on top of LDA.

4.1 (a), one can see that the use of the Coulomb truncation does not affect the convergence with respect to the plane wave energy cutoff and number of bands, which again is in all calculations set equal to the number of plane waves according to E_{cut}, other than by a constant shift. This is not the case for the k-point dependence (see Fig. 4.1 (b)). A more detailed discussion of the convergence of the band gap of a 2D material with respect to the k-point sampling and the interlayer separation with and without truncating the Coulomb interaction will be given in the next section.

All further calculations are performed with 30 Å interlayer separation (unless otherwise stated), Coulomb truncation and $(45 \times 45 \times 1)$ k-point sampling. The bandstructure of a free-standing graphene sheet is plotted in Fig. 4.2. The typical Dirac cone can be seen at the K point. LDA and GLLBSC yield almost identical curves, whereas G_0W_0 affects in particular strongly the slope of the valence and conduction band at the Fermi level. This gives a large reduction of the Fermi velocity, which can be derived as:

$$v_F = \frac{a}{2h}\frac{\Delta E}{\Delta k}, \tag{4.2}$$

where a is the lattice constant, $h = 4.1357 \cdot 10^{-15}$ eVs and k given in units of the reciprocal lattice vectors. The calculated values are $0.87 \cdot 10^6$ m/s (LDA), $0.87 \cdot 10^6$ m/s (GLLBSC) and $1.17 \cdot 10^6$ m/s (G_0W_0@LDA), respectively. The latter is in excellent agreement with corresponding results

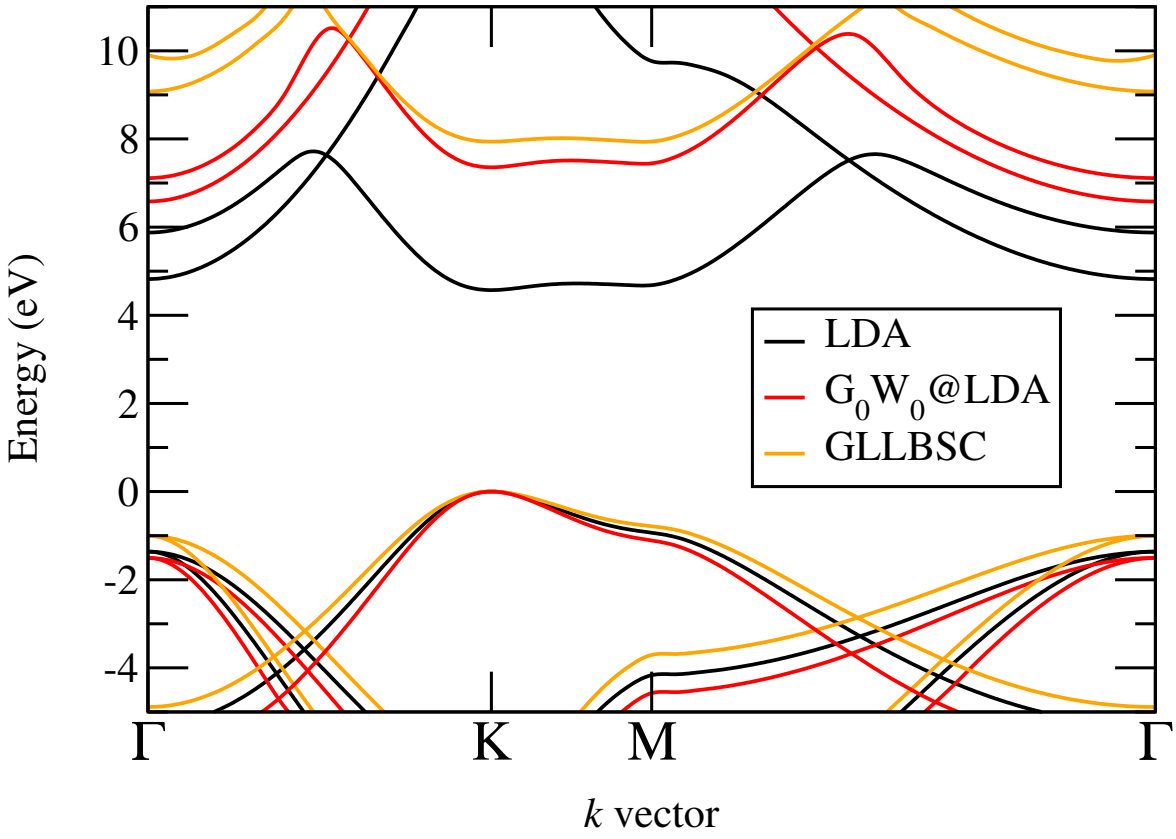


Figure 4.3: Bandstructure of *h*-BN, aligned to the top of the LDA valence band.

from previous calculations [89, 90] and state-of-the-art experiments [91].

For *h*-BN, the calculated valence band maximum lies with all three methods at the K point, as shown in Fig. 4.3. However, LDA and GLLBSC predict a direct band gap of 4.57 eV and 7.94 eV, respectively, whereas an indirect gap of 6.58 eV is found with G_0W_0@LDA. The direct transition at the K point is 7.37 eV. This is a result of a very inhomogeneous shift of the conduction bands in particular, which are moved up in energy by around 3 eV at the K point, but only by around 1.5 eV at the Γ point, compared to LDA. For GLLBSC, the effect is opposite with a larger upshift at the Γ point. In a previous study [92], 6.37 and 6.0 eV were obtained within the G_0W_0 approximation for the lowest indirect and direct quasiparticle gap, respectively. These calculations were, however, performed for an interlayer separation of only 13.5 Å and with a low *k*-point sampling. According to Fig. 4.1 (b), these parameters are not converged and the direct band gap is therefore around 1 eV too low.

Three different possibilites of stacking one *h*-BN and one graphene layer on top of each other are sketched in Fig. 4.4. Geometry optimization using the LDA functional yields the energetically most favorable structure to be configuration (c) with a distance of 3.18 Å between the layers, which agrees well with Ref. [93]. For this geometry, the Kohn-Sham and quasiparticle bandstructures are shown in Fig. 4.5. The different bands originating from either one of the two isolated layers can easily be identified: The highest valence and lowest conduction band still exhibit the linear dispersion at the K point typical for graphene. However, both with LDA, GLLBSC and G_0W_0, a small gap of around 50 meV is introduced, which is consistent with

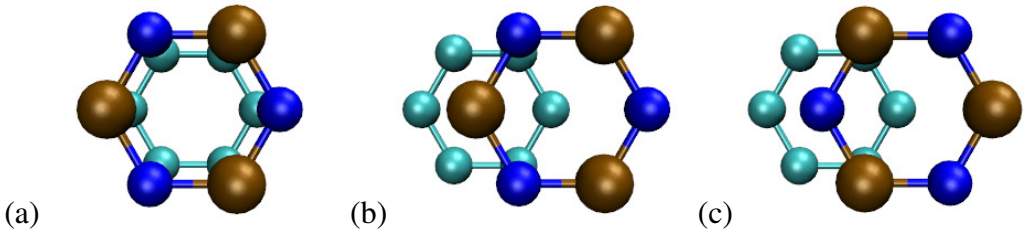


Figure 4.4: Different geometric structures for a graphene/hexagonal-boron nitride heterostructure. (a) The hexagons are placed on top of each other. (b) The N atom is placed on top of one C atom, while the B atom is centered above a graphene hexagon. (c) The B atom is placed on top of one C atom, while the N atom is centered above a graphene hexagon. A perspective view is chosen for clarity. Both lattices have the same size with a lattice constant of 2.503 Å and are separated by 3.18 Å.

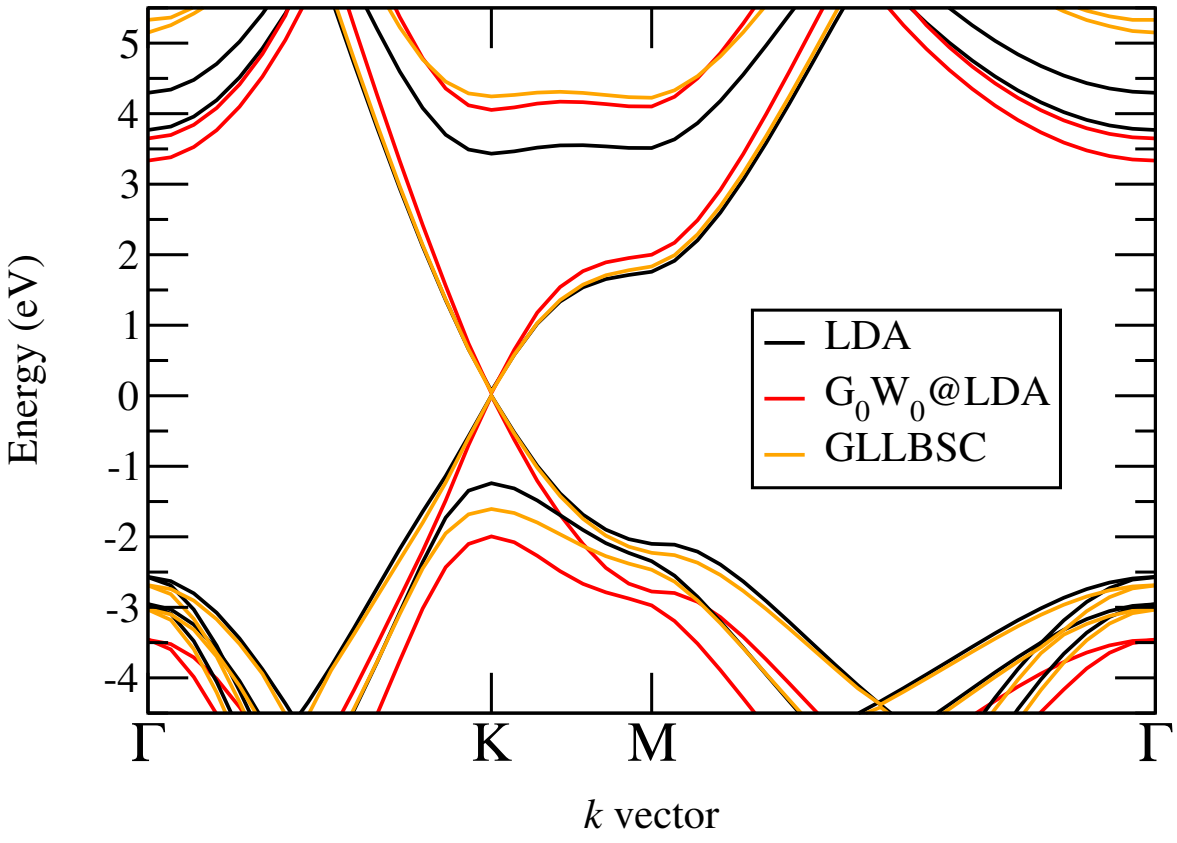


Figure 4.5: Bandstructure of graphene/*h*-BN in the configuration (c) of Fig. 4.4 for the equilibrium distance d = 3.18 Å.

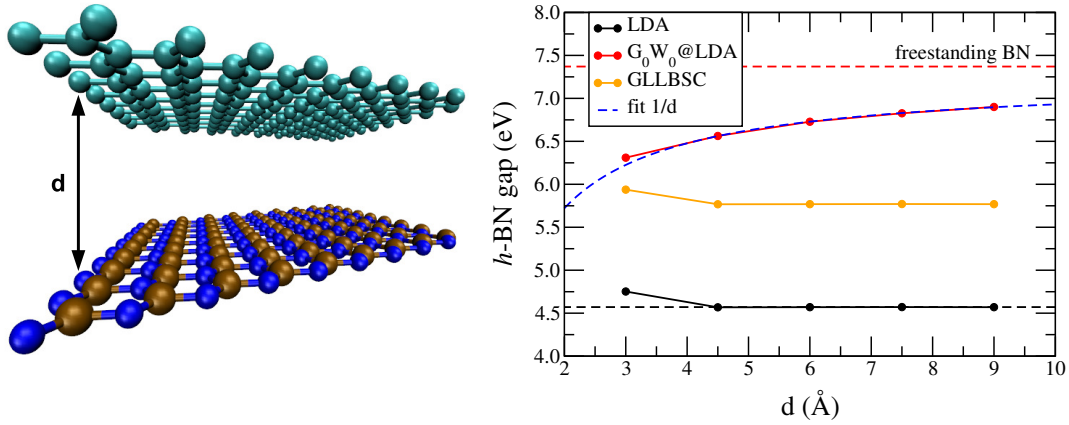


Figure 4.6: *h*-BN gap at the K point as function of the distance to the graphene sheet. For $d \to \infty$, it approaches the value of the freestanding *h*-BN monolayer, as indicated by a dashed horizontal line. The blue dashed line is a fit of the G_0W_0 results to a function $\propto 1/d$. A cell height of 60 Å was used.

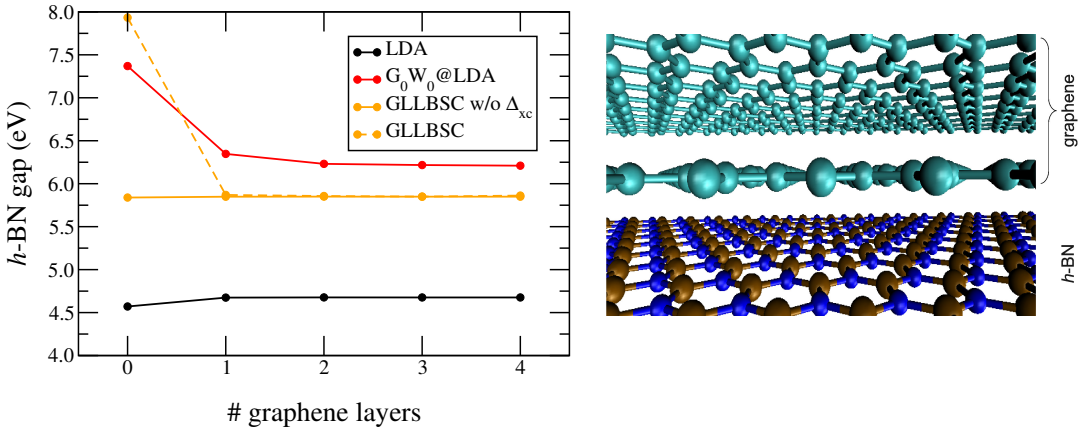


Figure 4.7: *h*-BN band gap at the K point for different numbers of adsorbed graphene layers. GLLBSC results are shown without and with explicit inclusion of the derivative discontinuity, Δ_{xc}, which becomes 0 when one or more graphene layers are present. A cell height of 60 Å was used.

previous calculations [93, 94]. Furthermore, the Fermi velocity is reduced to $0.78 \cdot 10^6$ (LDA), $0.78 \cdot 10^6$ (GLLBSC) and $1.01 \cdot 10^6$ m/s (G_0W_0@LDA), respectively. This decrease is expected when graphene is adsorbed on a substrate due to an increase of the dielectric screening [95]. The second highest occupied and unoccupied bands can be attributed to the boron nitride sheet. Especially with LDA, these bands hardly change, compared to the isolated layer, with a *h*-BN gap at the K point of 4.67 eV. With G_0W_0, on the other hand, this gap decreases to 6.35 eV. With GLLBSC, this value is 5.84 eV. These observations imply that, although there is no direct coupling between the layers, the electronic structure is effected considerably by long-ranged screening effects, which are not accounted for by a local potential.

This is further demonstrated in Fig. 4.6, where the distance between the graphene and the boron nitride sheet is varied. The value of the *h*-BN gap, which is defined as the transition energy between the second highest valence and the second lowest conduction band at the K point, remains almost constant with LDA and GLLBSC. Only for a very small distance, a

slight deviation is seen. This might be due to the formation of a chemical bond. For distances larger than 4 Å, the calculated gaps are the same as for the freestanding boron nitride, which indicates that screening is not described correctly within these methods. In contrast, the GW gap increases monotoneously with the distance between the layers and slowly approaches the value for a single *h*-BN sheet. Above 4 Å, this behaviour can be described by a function that is inversely proportional to d, which is expected within a simple picture of nonlocal screening. A classical image charge model predicts a $1/d$ dependence for the band gap reduction in a metal/semiconductor heterojunction [74, 75].

Another way of tuning the screening in a *h*-BN/graphene heterojunction is to change the number of graphene layers on which the boron nitride sheet is adsorbed. This is demonstrated in Fig. 4.7. Again, the *h*-BN gap is constant within LDA and GLLBSC. For the latter, however, there is a jump in the size of the gap when going from 0 (freestanding boron nitride) to 1 graphene layer, when the derivative discontinuity, Δ_{xc} is explicitely included. This is simply because Δ_{xc} vanishes for metallic systems, as it is the case when one or more graphene layers are present. With G_0W_0, the gap decreases by 1 eV, when adding one layer of graphene and by another 0.15 eV, when adding the second layer. This can be understood in the way that a thicker metal substrate enhances the screening which leads to a reduction of the gap. For more than two graphene layers, the GW gap does not change any further. This would presumably correspond to a system with a single boron nitride sheet adsorbed on bulk graphite.

4.2 Molybdenum disulfide

The quasiparticle bandstructure of monolayer molybdenum disulfide (MoS_2) has been studied intensively in literature over the past few years, both with one-shot and self-consistent GW methods. However, there remains ambiguity on the size and even on the type of the band gap. Reported values lie in the range from 2.6 to 3 eV and both direct and indirect gaps have been found with G_0W_0. A summary is given in Table 4.1 and will be discussed in detail at the end of this section.

Throughout the present study, the 4 *s* and 4 *p* semicore states of molybdenum were included. LDA and exact exchange calculations were done in a plane wave basis with kinetic energies up to 400 eV. A plane wave cutoff energy of 50 eV and 200 bands were used for the calculation of the dielectric function and the GW self-energy, which was found sufficient to converge band gaps to within 10 meV. Results shown here were all obtained within the Plasmon Pole approximation. Quasiparticle energies calculated with the full frequency dependent method differed by only around 0.1 eV.

First, the bandstructure is presented in Fig. 4.8 for the experimental lattice constant for bulk MoS_2 of 3.16 Å. The LDA band gap is direct at the K point with a value of 1.77 eV. A second minimum in the conduction band lies along the Γ-K direction, yielding a smallest indirect gap of 1.83 eV. This order is changed for G_0W_0@LDA which gives 2.77 and 2.58 eV for the direct and indirect transition energies, respectively.

When comparing these values to previous results, one has to consider that the gaps depend strongly on the number of k points, the interlayer distance, L, as defined in Fig. 4.9, and whether or not a truncation scheme for the Coulomb interaction is applied. This is shown in Fig. 4.10. Following conclusions can be drawn:

- For the $1/r$ interaction, enlarging L leads to a strong increase of the gaps.

Table 4.1: Calculated G_0W_0 band gaps, used parameters and comparison with literature.

Ref.	starting point	a (Å)	number of k points	layer separation (Å)	E_{gap} (eV)	
					direct	indirect
This work	LDA	3.16	$45 \times 45 \times 1$	23 (truncated V)	2.77	2.58
This work	LDA	3.19	$45 \times 45 \times 1$	23 (truncated V)	2.65	2.57
This work	LDA	3.255	$45 \times 45 \times 1$	23 (truncated V)	2.41	2.51
Ref. [96]	LDA	3.15	$18 \times 18 \times 1$	24	2.41	$\sim$ 2.40
Ref. [97]	PBE	3.16	$12 \times 12 \times 1$	19	$\sim$ 2.60	2.49
Ref. [97]	PBE	3.19	$12 \times 12 \times 1$	19	2.50	$\sim$ 2.55
Ref. [97]	PBE	3.255	$12 \times 12 \times 1$	19	2.19	2.19
Ref. [98]	LDA	3.16	$8 \times 8 \times 2$	19	2.96	–
Ref. [99]	PBE	3.18	$12 \times 12 \times 1$	20+1/L extrapolation	2.97	3.26
Ref. [99]	PBE	3.18	$12 \times 12 \times 1$	20	$\sim$ 2.60	$\sim$ 2.85
Ref. [100]	LDA	3.11	$12 \times 12 \times 1$	13	2.57	–
Ref. [101]	HSE	3.18	$6 \times 6 \times 1$	15	2.82	$\sim$ 3.00
Ref. [102]	PBE	3.19	$15 \times 15 \times 1$	15	2.66	–

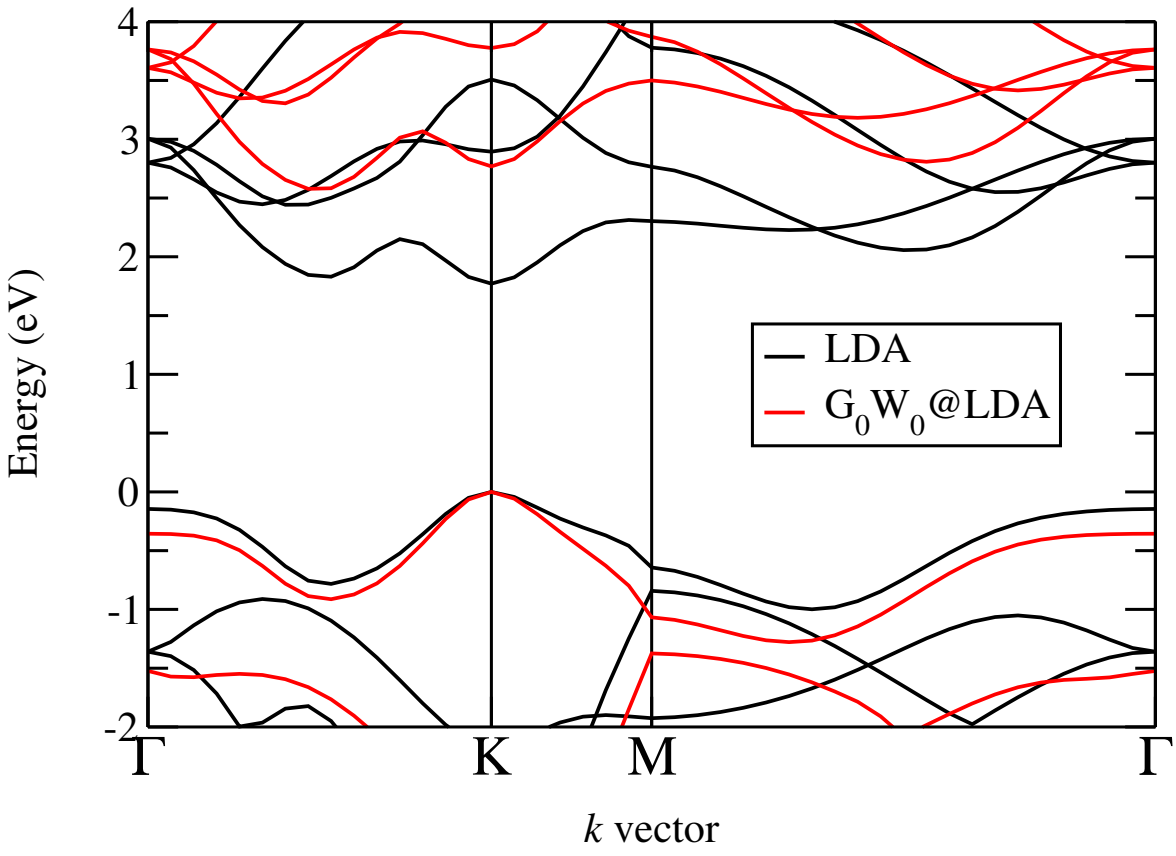


Figure 4.8: Bandstructure of monolayer MoS_2 obtained from a $(45 \times 45 \times 1)$ k-point sampling, 23 Å interlayer separation and use of the truncated Coulomb interaction. The top of the valence band has been set to 0.

- For the truncated interaction, an increase of the number of k-points gives a significant reduction of the gaps.

Since applying the truncation corresponds to $L \to \infty$, there exists a region, where these two effects cancel each other. This demonstrates that the interlayer separation and the k-point grid cannot be regarded as independent parameters. The interplay becomes clear from Fig. 4.11, where the direct band gap is shown as a function of the interlayer separation for different k-point samplings. The inverse plot allows for extrapolation of the data to infinite L. The extrapolated values are expected to converge towards the results obtained with truncation of the Coulomb interaction. Without the truncation, convergence of the gap with respect to either the number of k points or the layer separation alone gives wrong values. With $(9\times9\times1)$ k points, for example, the direct band gap goes to 3.7 eV for $L \to \infty$, which is around 1 eV too large. In contrast, a calculation with $L = 23$ Å and 21 k points in the in-plane directions seems well converged, but the obtained gap is too small. Due to these two errors, which are of different sign, the evaluated quasiparticle gaps are identical for different sets of parameters. This is where the curves cross the dashed horizontal lines. For example yield $(9\times9\times1)$ k points and $L = 23$ Å or $(15\times15\times1)$ k points and $L = 33$ Å the same result of 2.77 eV as $(45 \times 45 \times 1)$ k points and infinite layer separation. Fig. 4.12 shows all calculated results and interpolated values in a contour plot as a

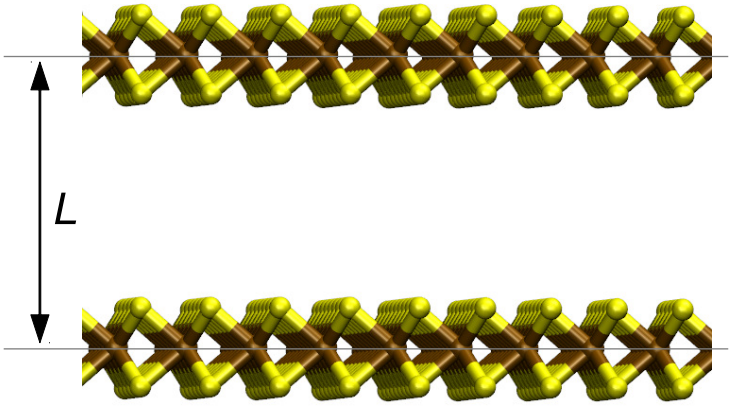


Figure 4.9: Definition of interlayer separation L for monolayer MoS_2 in a periodic supercell.

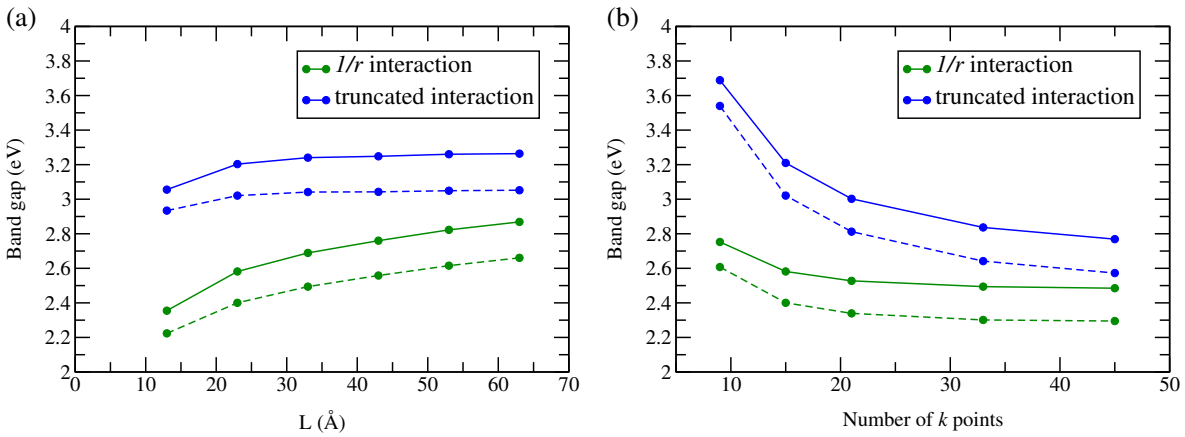


Figure 4.10: Direct (straight lines) and indirect (dashed lines) G_0W_0 band gap of monolayer MoS_2 as function of (a) interlayer separation for a k-point sampling of $(15 \times 15 \times 1)$ and (b) number of k points (in one of the in-plane directions) for $L = 23$ Å.

function of $1/L$ and number of k points.

The band gaps reported in previous studies were all obtained with low k-point samplings, small interlayer separations and without Coulomb truncation. All parameters and calculated values are summarized in Table 4.1. From the previous discussion of the convergence behaviour, the rather good agreement between those results and this work seems coincidental. Furthermore, almost all of these calculations were done with a single k point in the z direction (perpendicular to the layer). For small unit cells, this is not sufficient in order to obtain converged results. Additional calculations demonstrated that the gap size increases by around 0.2 - 0.3 eV when 3 k points are used along the z-axis for $L = 13$ Å, for example.

Another important issue, one has to keep in mind when comparing the data, is the effect of strain. As pointed out in Ref. [97], even a slight increase of the lattice constant, a, leads to significant changes in the electronic bandstructure. The present calculations predict a reduction of the direct band gap by around 0.1 eV for $a = 3.19$ Å, which corresponds to 1 % strain. The indirect gap, on the other hand, is almost unaffected. For $a = 3.255$ Å, the direct gap is further reduced, leading to a transition from an indirect to a direct band gap material.

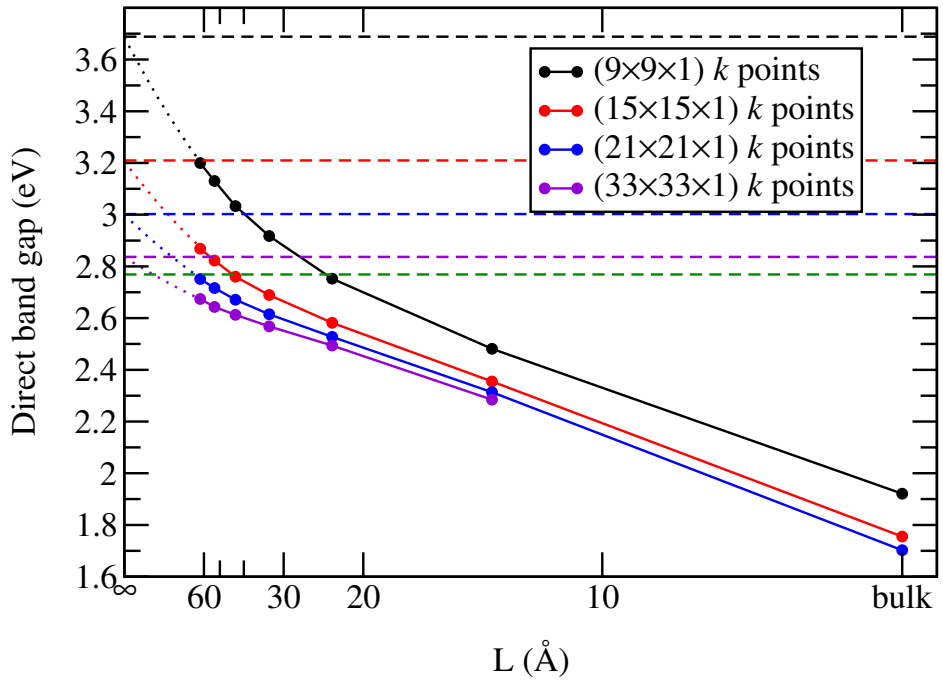


Figure 4.11: Direct G_0W_0 band gap of monolayer MoS_2 for different k-point samplings and interlayer separations without use of the truncated Coulomb interaction. Dotted lines serve as a guide for the eye to extrapolate to infinite L. The dashed horizontal line indicate the values obtained with the truncated Coulomb interaction. Also shown are results for $L = 6.145$ Å, which corresponds to bulk MoS_2.

All in all, the careful examination of the dependence of the electronic structure on the different parameters, as presented in Fig. 4.11 and Table 4.1, shows that the current implementation reproduces most of the results for the direct band gap reported in literature, given the same set of parameters. The fact that they are in good agreement with the fully converged values is mainly due to a fortunate cancellation of errors. For the experimental lattice constant of 3.16 Å, the G_0W_0@LDA band gap is found to be indirect with a value of around 2.6 eV. Choosing the PBE functional as starting point is not expected to give noticeable changes. On the other hand, both partially self-consistent GW_0 [97] and quasiparticle self-consistent GW [98] calculations have yielded direct band gaps of $2.75 - 2.80$ eV.

Experimentally, single-layered MoS_2 is a direct gap semiconductor [83] with a band gap of around 1.9 eV [82]. It is known that there are strong excitonic effects in this system, which lead to the formation of bound electron-hole pairs [96]. The experimentally measured value corresponds then to the optical gap, which is smaller than the quasiparticle gap by the exciton binding energy.

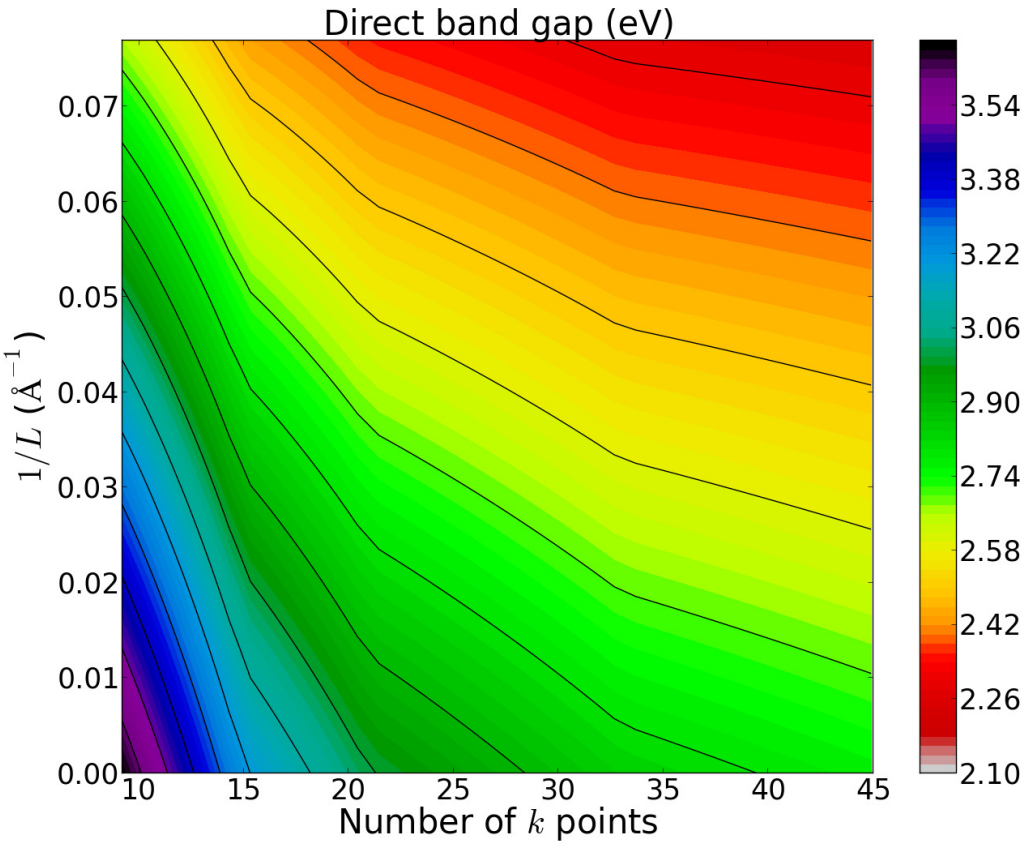


Figure 4.12: Contour plot of the direct G_0W_0 band gap as a function of the inverse interlayer distance and number of k points in one of the in-plane directions with the full $1/r$ interaction. Contour lines are separated by 0.1 eV. Interpolation from splines was used.

CHAPTER 5

MOLECULES

Molecules can be regarded as 0-dimensional systems, since their spatial extent is finite in all directions. This means, that in all electronic structure codes which require periodic boundary conditions, calculations have to be performed in sufficiently large supercells in order to circumvent artificial effects from interactions between molecules in neighboring cells. The electronic structure of finite systems consists of a discrete spectrum of well-seperated energy levels. This would correspond to a completely flat bandstructure, which can be restricted to the Γ-point only. Instead of Bloch waves, the eigenfunctions are given as molecular orbitals.

The energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is typically on the size of several eV for an isolated molecule in the gas phase. From a physicist's point of view, they are large gap insulators. Screening is therefore expected to play a minor role and HF methods usually describe the electronic structure reasonably well. Most DFT functionals on the other hand, tend to underestimate the gaps dramatically. Furthermore, unoccupied states are in principal not accessible with DFT, and only the HOMO energy can be interpreted in a physically meaningful way as the negative ionization potential (IP), given the correct functional [4].

Adsorbing a molecule on a metal surface leads to an alignment of the molecular orbitals with respect to the Fermi energy of the metal and to a decrease of the HOMO-LUMO gap. This is a direct consequence of dynamical screening and can therefore not be described with DFT [103–106]. Instead, the correct physics is captured by GW as quasiparticle theory, which enables quantitatively accurate modeling of electron transport in molecular junctions [107–111], for example.

This chapter gives a discussion of the calculated IPs of 32 small molecules. G_0W_0 results are in much better agreement with experiment than various DFT based methods, but still around 0.5 eV too low on average. This systematic deviation can be explained by a suboptimal choice of the starting point, namely LDA. The problem of extreme slow convergence of the HOMO energy with respect to the plane wave cutoff energy and number of bands is resolved by applying a simple extrapolation scheme. A detailed analysis of the frequency-dependence of the self-energy illustrates the underlying physics.

Finally, the frontier orbitals of benzenediamine are studied as a typical example for a structure used in molecular electronics.

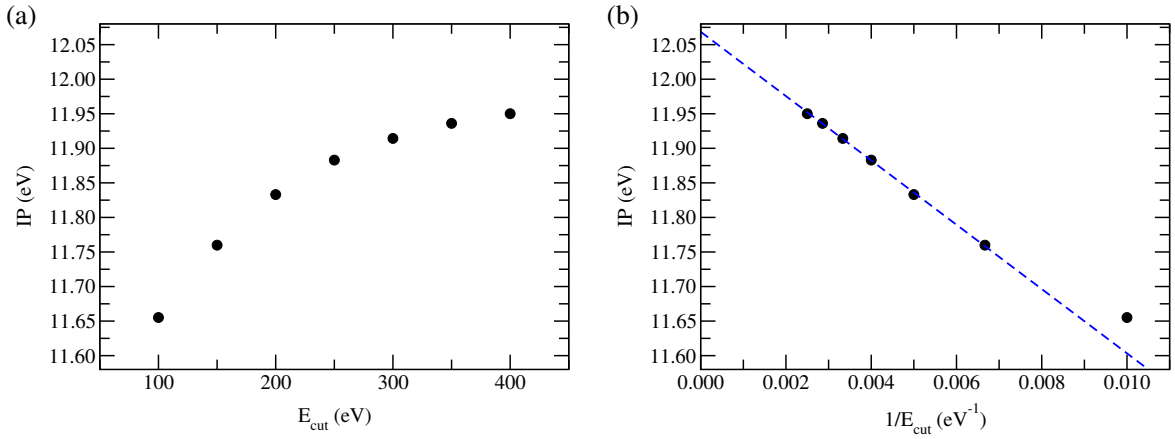


Figure 5.1: Ionization potential of water as function of (a) the plane wave cutoff energy and (b) the inverse plane wave cutoff energy for G_0W_0@LDA. The dashed line in (b) shows a linear fit of the data points with $E_{\text{cut}} > 100\,\text{eV}$ ($1/E_{\text{cut}} < 0.01\,\text{eV}^{-1}$). Extrapolation to $0\,\text{eV}^{-1}$ (infinite cutoff energy) gives a value of 12.1 eV.

5.1 Ionization potentials

In order to study the performance of the implementation for finite systems, the quasiparticle energies of 32 small molecules in the gas phase were calculated with the G_0W_0 approximation and compared to results from non-selfconsistent Hartree-Fock, LDA, PBE0 and GLLBSC. Their geometric structures were taken from the Atomic Simulation Environment database [112]. 3.5 Å of vacuum were added in each direction of the supercell. For all groundstate calculations, plane waves with kinetic energies up to 400 eV were used as basis functions. In literature, linear combinations of atomic orbitals (LCAO) are usually preferred as basis functions for molecular systems. In this way, even small basis sets are often sufficient in order to describe the spatially localized electronic states with good accuracy. With plane waves, this can only be achieved with a very high number of basis functions. On the other hand, the size of the plane wave basis can be increased freely for a systematic control over the quality of the basis, which is particularly important for GW calculations. This is very difficult to assess with LCAO.

The G_0W_0@LDA results for the ionization potential of a water molecule are shown in Fig. 5.1 (a) for different values of the plane wave cutoff energy used in the calculation of the GW self energy. The exact exchange contributions were determined with a fixed cutoff energy of 400 eV. In all cases, the number of bands in the calculation of Σ was chosen accordingly to E_{cut}. Even for 400 eV, which corresponds to more than 8000 plane waves and bands, the IP does not seem to be fully converged. This extremely slow convergence behaviour has been observed previously for a benzene molecule [113, 114]. A data fit to the linear function $\text{IP}(E_{\text{cut}}) = \text{IP}(\infty) - A/E_{\text{cut}}$ allowed for extrapolation to infinite basis size. A similar scheme has been applied for the valence band maximum in rutile TiO_2 [115]. As demonstrated in Fig. 5.1 (b), a linear function of $1/E_{\text{cut}}$ fits the data points perfectly for $E_{\text{cut}} > 100\,\text{eV}$. Its intersection with the y-axis gives an IP of 12.1 eV.

This extrapolation scheme has been applied to all 32 molecules by a linear fit of the calculated IPs for plane wave cutoff energies between 200 and 400 eV. Examples are shown in Fig. 5.2. The Plasmon Pole approximation was used in all GW calculations presented here. Additional calculations were performed with the full frequency-dependent method for comparison.

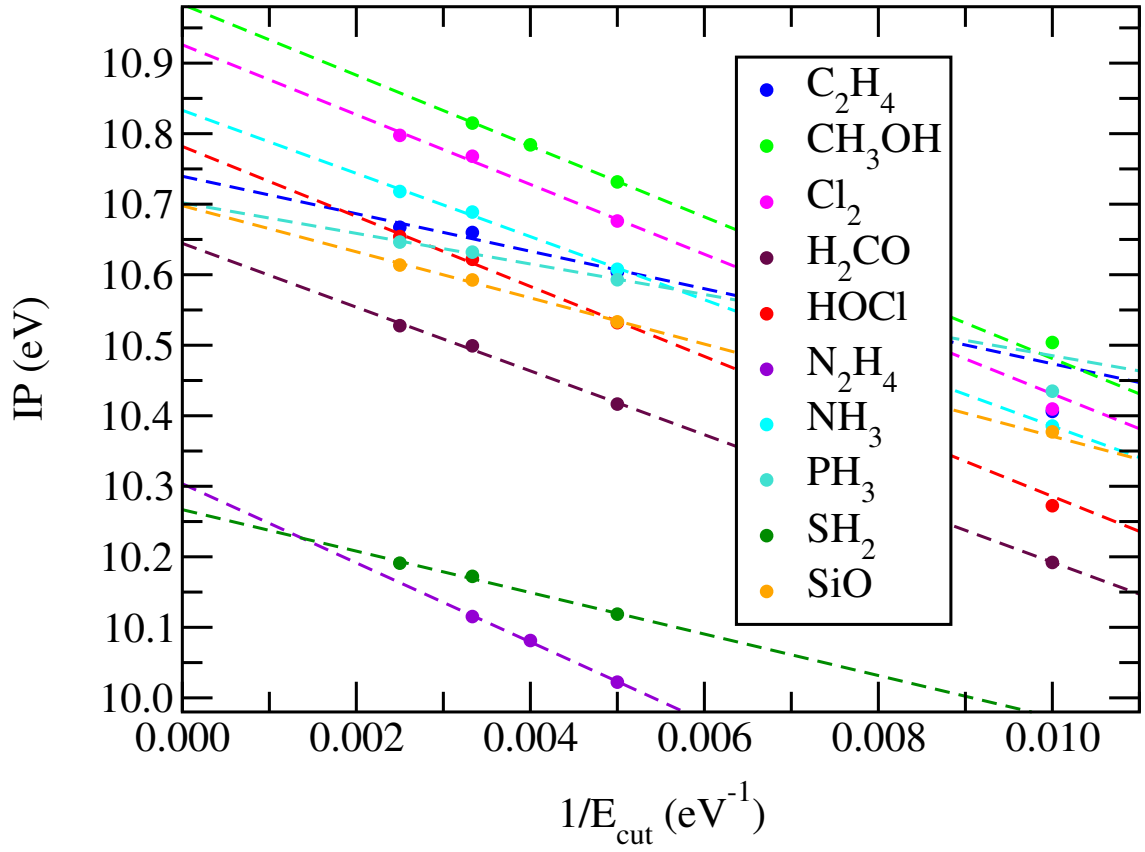


Figure 5.2: Calculated IPs with G_0W_0@LDA and linear fitting of the data points with extrapolation to infinite number of plane waves and bands ($E_{\text{cut}} = 0$).

Deviations in the QP energies were smaller than 0.1 eV.

All IPs obtained from extrapolation are listed in Table 5.1, along with results from LDA, non-selfconsistent HF on top of LDA, PBE0 with 1/4 of PBE exchange replaced non-selfconsistently by exact exchange and GLLBSC and compared to experimental data taken from Ref. [116]. LDA and PBE0 predict much too small values with mean absolute errors of 4.8 and 3.5 eV, respectively. Also GLLBSC underestimates IPs by 1.8 eV on average. This is not surprising, since it is a method designed for calculating the electronic structure of solids. HF on the other hand overestimates the IP for most molecules with a MAE of 1.1 eV. Best results are found with the GW approximation, where the MAE is less than 0.6 eV. Still, many IPs are too low and the mean signed error is −0.3 eV. This systematic error may be attributed to the use of LDA wavefunctions as starting point, which are most likely a bad representation of the true quasiparticle wavefunctions.

A comparison of the G_0W_0 results with previous studies on the same set of molecules is shown in Fig. 5.4. The calculated IPs agree overall well with Ref. [117], in which also LDA wavefunctions and eigenvalues were used as starting point, but with a Gaussian basis set. The values spread with a mean absolute error of 0.32 eV, but with similar magnitude for both positive and negative deviations, so that the mean signed error is only 0.02 eV. In comparison to the all-electron G_0W_0@PBE calculations of Ref. [118], in which numerical atomic orbitals were used as basis set, a systematic deviation is found with a MAE of 0.36 eV and a MSE of

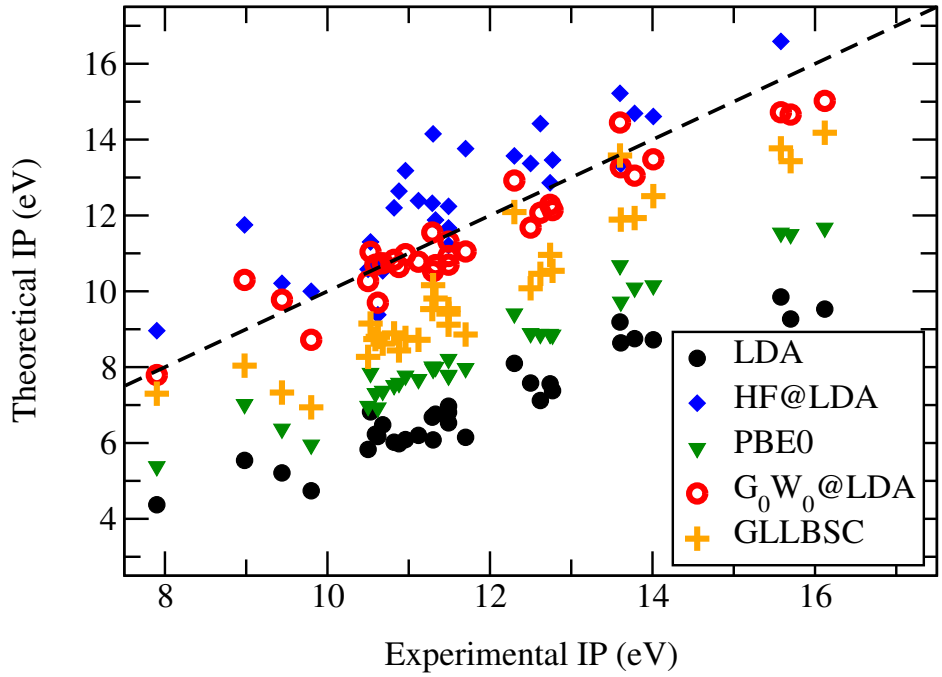


Figure 5.3: Calculated IPs for the 32 molecules listed in Table 5.1 plotted against the experimental values. The dashed line indicates equality.

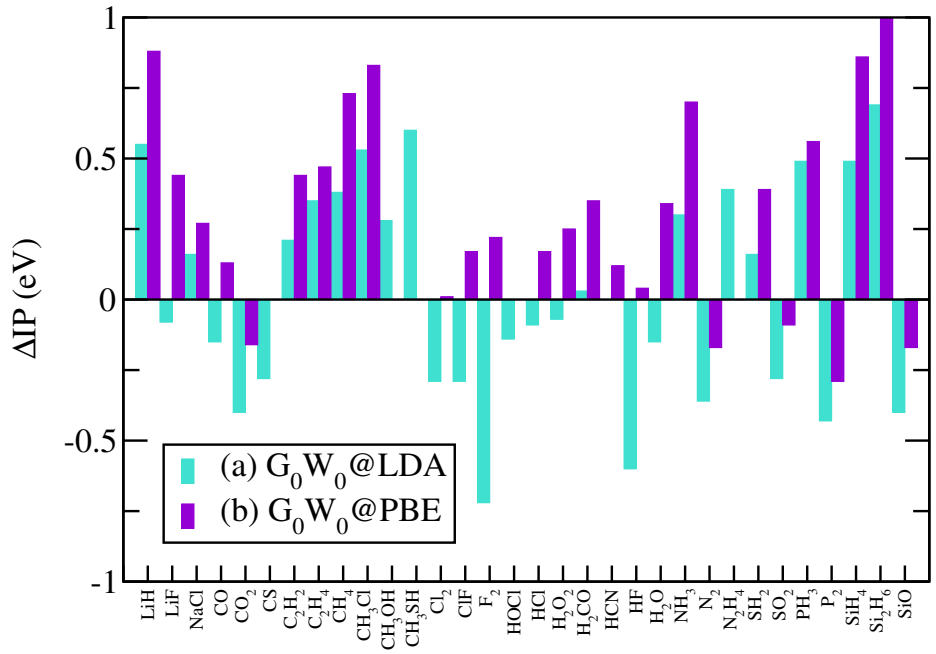


Figure 5.4: Difference in calculated IPs with respect to (a) Ref. [117] and (b) Ref. [118].

Table 5.1: Theoretical IPs calculated with different methods as described in the text. Experimental data is given in the last column. All values are in eV. The last rows shows the mean absolute error and the mean signed error with respect to experiment.

Molecule	LDA	HF@LDA	PBE0	GLLBSC	G_0W_0@LDA	experiment
LiH	4.37	8.96	5.38	7.30	7.79	7.90
LiF	6.08	14.15	7.95	10.16	10.53	11.30
NaCl	4.74	10.00	5.95	6.94	8.72	9.80
CO	8.72	14.61	10.15	12.51	13.48	14.01
CO_2	8.75	14.69	10.09	11.93	13.05	13.78
CS	6.76	11.88	8.00	9.81	10.69	11.33
C_2H_2	6.81	11.21	7.79	9.41	11.22	11.49
C_2H_4	6.48	10.54	7.37	8.62	10.74	10.68
CH_4	9.19	15.22	10.68	13.58	14.45	13.60
CH_3Cl	6.68	12.32	8.01	9.53	11.55	11.29
CH_3OH	6.09	13.18	7.77	8.77	10.98	10.96
CH_3SH	5.21	10.21	6.37	7.33	9.78	9.44
Cl_2	6.53	11.67	7.77	9.12	10.93	11.49
ClF	7.38	13.46	8.85	10.54	12.14	12.77
F_2	9.27	18.44	11.50	13.43	14.66	15.70
HOCl	6.20	12.39	7.68	8.72	10.78	11.12
HCl	7.56	12.86	8.87	10.96	12.28	12.74
H_2O_2	6.15	13.76	7.97	8.86	11.05	11.70
H_2CO	5.98	12.64	7.58	8.44	10.64	10.88
HCN	8.64	13.35	9.72	11.89	13.27	13.61
HF	9.53	18.29	11.67	14.18	15.02	16.12
H_2O	7.12	14.42	8.87	10.46	12.07	12.62
NH_3	6.02	12.20	7.52	8.89	10.83	10.82
N_2	9.85	16.59	11.54	13.77	14.72	15.58
N_2H_4	5.54	11.75	7.02	8.04	10.30	8.98
SH_2	5.83	10.58	6.97	8.27	10.27	10.50
SO_2	7.58	13.37	8.89	10.08	11.68	12.50
PH_3	6.23	10.77	7.31	8.74	10.70	10.59
P_2	6.17	9.38	6.93	8.80	9.70	10.62
SiH_4	8.10	13.57	9.41	12.09	12.92	12.30
Si_2H_6	6.82	11.30	7.84	9.15	11.04	10.53
SiO	6.97	12.24	8.21	9.53	10.70	11.49
MAE	4.84	1.11	3.46	1.83	0.56	.
MSE	-4.84	0.99	-3.46	-1.83	-0.30	.

0.30 eV. These errors are expected regarding the different starting points and basis sets. Comparing the results of Ref. [117] with Ref. [118] gives almost the same MAE and MSE (0.30 eV and 0.28 eV, respectively). Very similar deviations are also found between these results and Ref. [119], where an accurate basis set of Wannier functions augmented by numerical atomic orbitals was used. Discrepancies in the calculated values are therefore within the accuracy of different implementations.

A much better agreement with experiment was found for HF and hybrid functionals as starting point for G_0W_0 [117, 119], whereas this does not necessarily seem to be the case with self-consistent GW [118–120].

5.2 Frequency-dependence of the self-energy

For all molecules in this study, both the HOMO and the LUMO energies decrease with larger plane wave cutoff energies (meaning that the IPs increase). Since the exact exchange contributions are left unchanged, this results from the correlation part of the self-energy only. This might be surprising at first, since one could expect that a larger basis set, e.g. number of plane waves, would enhance the screening which would shift the occupied orbitals down in energy. A closer look at the frequency-dependence of the self-energy, $\Sigma(\omega)$, helps to understand this seeming conflict. This is shown for the HOMO and LUMO of the CO molecule in Fig. 5.5, as one example. The imaginary part of $\Sigma(\omega)$ is to a large extent built up by the poles of the convolution of the Greens' Function and the screened potential, $\Sigma(\omega) = \int d\omega' \, G(\omega + \omega')W(\omega')$. W itself has poles at all transition energies between occupied and unoccupied states, $\omega_{n\to n'}$, while G has poles at all energies ε_s. Thus, the poles of Σ lie between $(2\varepsilon_{\min} - \varepsilon_{\max})$ and $(2\varepsilon_{\max} - \varepsilon_{\min})$, where $\varepsilon_{\min}$ and $\varepsilon_{\max}$ are the lowest and highest input eigenvalues, respectively. As a consequence of this asymmetry, including more bands gives larger contributions to the imaginary part of $\Sigma(\omega)$ in the high energy range, whereas the lower part of the spectrum is less effected. This is clearly seen in Fig. 5.5 (a), by comparing the light and dark blue curves (for the HOMO) for positive and negative values of ω, and correspondingly the orange and red curves (for the LUMO). The real and imaginary parts of the self-energy are related by the Kramers-Kronig relation:

$$\mathrm{Re}\,\{\Sigma(\omega)\} = \frac{1}{\pi}\mathcal{P}\left\{\int d\omega' \, \frac{\mathrm{Im}\,\{\Sigma(\omega')\}}{\omega' - \omega}\right\} \tag{5.1}$$

The correlation contributions to the quasiparticle energies are in the G_0W_0 approximation given by the value of the real part of the self-energy at the starting point eigenvalues, as indicated by the dashed lines in Fig. 5.5 (b). The real part of $\Sigma(\omega)$ changes sign around the HOMO-LUMO gap, where its structure is dominated by large resonance/antiresonance peaks. As more empty states are added and the imaginary part extends to higher energies, the real part is pushed down, in particular for high ω as a consequence of the asymetric structure of $\mathrm{Im}\,\{\Sigma\}$. This lowers the antiresonance peak and thereby the correlation.

5.3 BDA

One of the most commonly studied organic molecules is benzene with its delocalized system of π electrons. For example, it is used as a model system for electronic transport in molecular junctions, where the molecule is attached to metal electrodes via anchor groups, e.g. amino

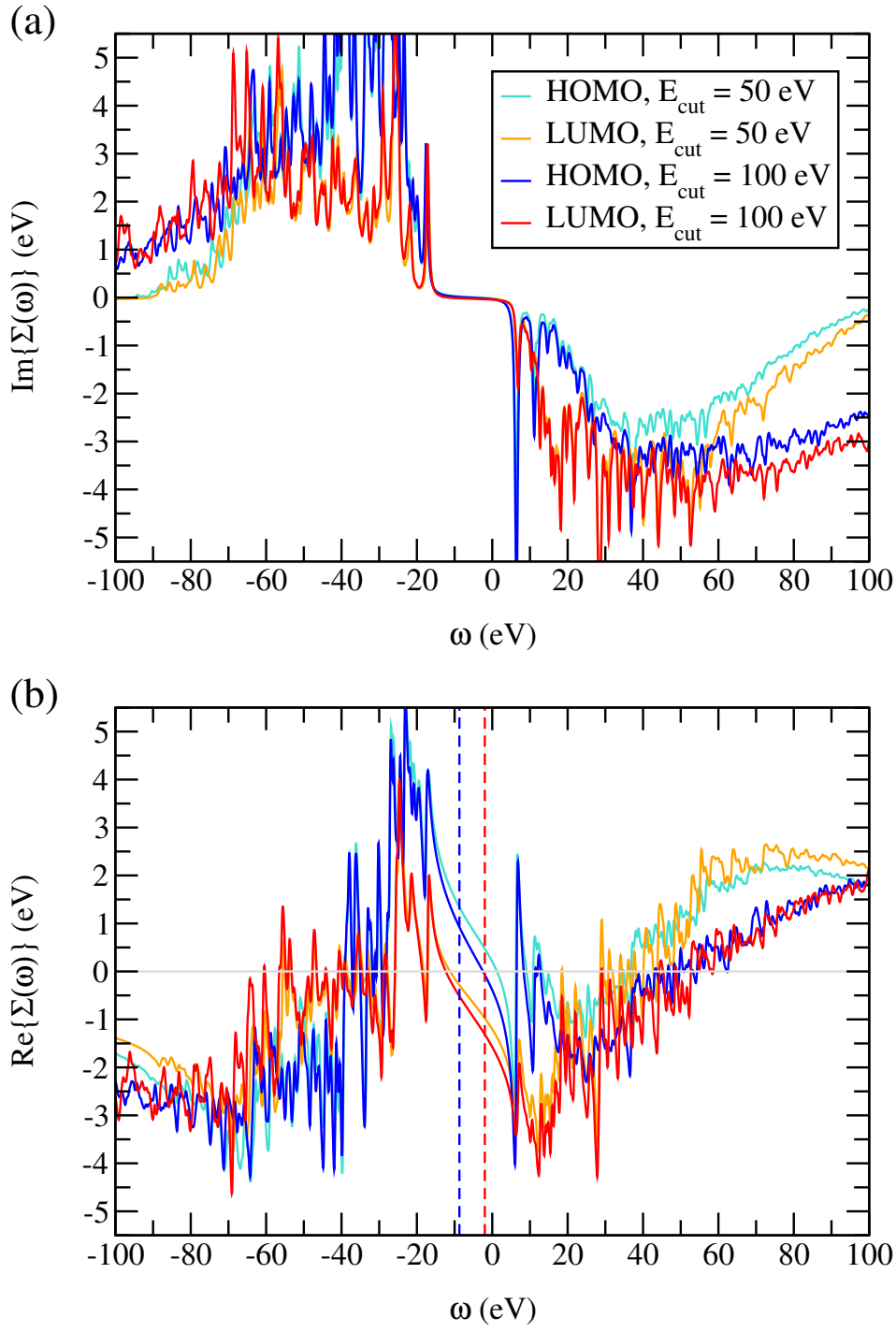


Figure 5.5: (a) Imaginary and (b) real parts of the self-energy $\langle n|\Sigma(\omega)|n\rangle$ for the HOMO and LUMO levels of the CO molecule and two different plane wave cutoff energies. Dashed lines indicate the respective LDA eigenvalues. The lowest input eigenvalue, ε_{min}, is around 30 eV, while the highest, ε_{max}, is around 50 eV and 100 eV, respectively. Thus, Im $\{\Sigma(\omega')\}$ is only non-zero in the range from -110 to 130 eV and -160 to 230 eV, respectively.

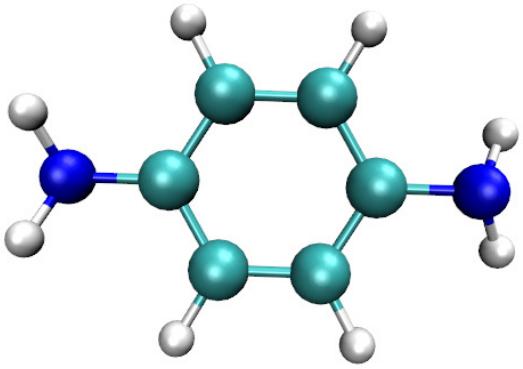


Figure 5.6: Geometric structure of benzenediamine, *para*-$C_6H_4(NH_2)_2$.

groups [121]. The transport properties are to a large extent determined by the alignment of the molecular frontier orbitals with respect to the metal Fermi level. For theoretical predictions, it is therefore essential to obtain a correct description of the electronic structure of the molecule in the gas phase at first.

The frontier orbitals of benzenediamine ($C_6H_8N_2$, as sketched in Fig. 5.6) have been calculated in the same way as described for the small molecules in Chapter 5.1, but in a larger unit cell with 5 Å of vacuum added in each direction around the molecule. This was found necessary in order to converge the DFT energies. The extrapolation scheme has been applied both to the HOMO and the LUMO. Additional HF and GW calculations have been made with PBE as starting point. All results are summarized in Table 5.2. No large differences can be seen between the results with LDA and PBE, and consequently non-selfconsistent HF and GW calculations also give very similar results for the two different starting points. G_0W_0 leads to a drastic opening of the gap, with a large downshift of the HOMO energy and an increase of the LUMO energy. With HF, these effects are even bigger.

The extrapolated IP of the BDA molecule is 7.47 eV with G_0W_0@PBE, which is in remarkably good agreement with the result of 7.3 eV reported in Ref. [122], which also used a plane wave basis, but a different extrapolation scheme, and the experimental value of 7.34 eV [123, 124].

The results of a previous study of the same molecule are presented in Table 5.3. It is important to note that a double-zeta atomic orbital basis set has been used, compared to plane waves in the present study. In principal, both methods are expected to give identical results, given that the basis sets are complete. For LCAO, this requires basis functions of very good quality, while for plane waves, the cutoff energy needs to be sufficiently high. Large differences can be seen for the Hartree-Fock results, both for the energies of the frontier orbitals and the HOMO-LUMO gap. Two reasons may be used for an explanation: First, the neglection of core-valence exchange in the current calculations. According to Ref. [110], however, these contributions are expected to be rather small, generally less than 0.4 eV for the frontier orbitals of benzene-like molecules. Therefore, the role of self-consistency is much more important, which seems to shift both the HOMO and LUMO significantly and decrease the HF gap by more than 1 eV. Also for GW, there is a large discrepancy between the orbital energies. The gap size, on the other hand, is very similar. This might be due to a cancellation of errors in the exchange and in the correlation contributions using DFT wavefunctions instead of doing self-consistent calculations. The last column of Table 5.3 states the results of PBE total energy difference calculations, which can be

Table 5.2: Frontier orbitals and HOMO-LUMO gap of the BDA molecule. All values are in eV.

	LDA	PBE	HF@LDA	HF@PBE	G_0W_0@LDA	G_0W_0@PBE
HOMO	-4.00	-4.05	-8.05	-8.04	-7.51	-7.47
LUMO	-0.88	-0.82	4.33	4.33	1.41	1.45
gap	3.12	3.23	12.38	12.37	8.92	8.92

Table 5.3: Results from Ref. [110] in eV.

	PBE	scHF	scGW	ΔE_{tot} (PBE)
HOMO	-4.1	-7.2	-6.2	-6.8
LUMO	-0.9	3.9	2.9	2.3
gap	3.2	11.1	9.1	9.1

regarded as reference values. In this method, HOMO and LUMO energies are obtained by calculating the electronic groundstate with an extra electron explicitely added to or removed from the system. In comparison, the DFT gaps are vastly underestimated which is a consequence of self-interaction errors in the functionals. These are not present within Hartree-Fock by construction. Instead, neglect of correlation, which is here equal to the lack of orbital relaxations, leads to a large overestimation of the gap [110]. In contrast, the GW approximation naturally includes these effects via the screened interaction.

BIBLIOGRAPHY

[1] V. Fock, *"Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems"*, Z. Phys. **61**, 126 (1930).

[2] P. Hohenberg and W. Kohn, *"Inhomogeneous Electron Gas"*, Phys. Rev. **136**, B864 (1964).

[3] W. Sham and L. J. Sham, *"Self-Consistent Equations Including Exchange and Correlation Effects"*, Phys. Rev. **140**, A1133 (1965).

[4] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz Jr., *"Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy"*, Phys. Rev. Lett. **49**, 1691 (1982).

[5] L. Hedin, *"New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem"*, Phys. Rev. **139**, A796 (1965).

[6] J. P. Perdew and A. Zunger, *"Self-interaction correction to density-functional approximations for many-electron systems"*, Phys. Rev. B **23**, 5048 (1981).

[7] H. Bruus and K. Flensberg, *"Many-Body Quantum Theory in Condensed Matter Physics - An Introduction"*, (Oxford University Press, Oxford, 2004).

[8] F. Aryasetiawan and O. Gunnarsson, *"The GW method"*, Rep. Prog. Phys. **61**, 237 (1998).

[9] G. Onida, L. Reining, and A. Rubio, *"Electronic excitations: density-functional versus many-body Green's function approaches"*, Rev. of Mod. Phys. **74**, 601 (2002).

[10] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, *"Quasiparticle Calculations in Solids"*, Solid State Physics: Advances in Research and Applications **54**, (2000).

[11] M. S. Hybertsen and S. G. Louie, *"Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies"*, Phys. Rev. B **34**, 5390 (1986).

[12] M. van Schilfgaarde, T. Kotani, and S. Faleev, *"Quasiparticle Self-consistent GW Theory"*, Phys. Rev. Lett. **96**, 226402 (2006).

[13] H. J. de Groot, P. A. Bobbert, and W. van Haeringen, *"Self-consistent GW for a quasi-one-dimensional semiconductor"*, Phys. Rev. B **52**, 11000 (1995).

[14] T. Kotani, M. van Schilfgaarde, S. V. Faleev, and A. Chantis, *"Quasiparticle self-consistent GW method: a short summary"*, J. Phys.: Condens Matter **19**, 365236 (2007).

[15] M. P. Surh, S. G. Louie, and M. L. Cohen, *"Quasiparticle energies for cubic BN, BP, and BAs"*, Phys. Rev. B **43**, 9126 (1991).

[16] U. von Bart and B. Holm, *"Self-consistent GW_0 results for the electron gas: Fixed screened potential W_0 within the random-phase approximation"*, Phys. Rev. B **54**, 8411 (1996).

[17] E. L. Shirley, *"Self-consistent GW and higher-order calculations of electron states in metals"*, Phys. Rev. B **54**, 7758 (1996).

[18] P. E. Blöchl, *"Projector augmented-wave method"*, Phys. Rev. B **50**, 17953 (1994).

[19] P. E. Blöchl, C. J. Först, and J. Schimpl, *"Projector Augmented Wave Method: ab-initio molecular dynamics with full wave functions"*, Bull. Mater. Sci. **26**, 33 (2003).

[20] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dulak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, et al., *"Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method"*, J. Phys.: Condens. Matter **22**, 253202 (2010).

[21] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, *"Real-space grid implementation of the projector augmented wave method"*, Phys. Rev. B **71**, 035109 (2005).

[22] A. H. Larsen, M. Vanin, J. J. Mortensen, K. S. Thygesen, and K. W. Jacobsen, *"Localized atomic basis set in the projector augmented wave method"*, Phys. Rev. B **80**, 195112 (2009).

[23] M. Shishkin and G. Kresse, *"Implementation and performance of the frequency-dependent GW method within the PAW framework"*, Phys. Rev. B **74**, 035101 (2006).

[24] J. Yan, J. J. Mortensen, K. W. Jacobsen, and K. S. Thygesen, *"Linear density response function in the projector augmented wave method: Applications to solids, surfaces, and interfaces"*, Phys. Rev. B **83**, 245122 (2011).

[25] R. W. Godby and R. J. Needs, *"Metal-Insulator Transition in Kohn-Sham Theory and Quasiparticle Theory"*, Phys. Rev. Lett. **62**, 1169 (1989).

[26] A. Fleszar and W. Hanke, *"Spectral properties of quasiparticles in a semiconductor"*, Phys. Rev. B **56**, 10228 (1997).

[27] I. Campillo, J. M. Pitarke, A. Rubio, E. Zarate, and P. M. Echenique, *"Inelastic lifetimes of Hot Electrons in Real Metals"*, Phys. Rev. Lett. **83**, 2230 (1999).

[28] A. Marini, R. Del Sole, A. Rubio, and G. Onida, *"Quasiparticle band-structure effects of the d hole lifetimes of copper within the GW approximation"*, Phys. Rev. B **66**, 161104(R) (2002).

[29] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, *"Exact Coulomb cutoff technique for supercell calculations"*, Phys. Rev. B **73**, 205119 (2006).

[30] S. Ismail-Beigi, *"Truncation of periodic image interactions for confined systems"*, Phys. Rev. B **73**, 233103 (2006).

[31] A. Sorouri, W. M. Foulkes, and N. D. Hine, *"Accurate and efficient method for the treatment of exchange in a plane-wave basis"*, J. Chem. Phys. **124**, 064105 (2006).

[32] `http://wiki.fysik.dtu.dk/gpaw/documentation/gw_theory/gw_theory.html`

[33] `http://wiki.fysik.dtu.dk/niflheim/Hardware`

[34] *"Physics of Group-IV Elements and III-V Compounds"*, Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology - New Series, Group III, Vol. 17a, edited by O. Madelung, M. Schulz, and H. Weiss, (Springer, New York, 1982).

[35] M. S. Hybertsen and S. G. Louie, *"Electron correlations in semiconductors and insulators: Band gaps and quasiparticle energies"*, Phys. Rev. B **34**, 5390 (1986).

[36] O. Gritsenko, R. van Leeuwen, E. van Lenthe, and E. J. Baerends, *"Self-consistent approximation to the Kohn-Sham exchange potential"*, Phys. Rev. A **51**, 1944 (1995).

[37] M. Kuisma, J. Ojanen, J. Enkovaara, and T. T. Rantala, *"Kohn-Sham potential with discontinuity for band gap materials"*, Phys. Rev. B **82**, 115106 (2010).

[38] F. Tran and P. Blaha, *"Accurate Band Gaps of Semiconductors and Insulators with a Semilocal Exchange-Correlation Potential"*, Phys. Rev. Lett. **102**, 226401 (2009).

[39] M. Shishkin and G. Kresse, *"Self-consistent GW calculations for semiconductors and insulators"*, Phys. Rev. B **75**, 235102 (2007).

[40] J. P. Perdew, K. Burke, and M. Ernzerhof, *"Generalized Gradient Approximation Made Simple"*, Phys. Rev. Lett. **77**, 3865 (1996).

[41] J. Heyd, G. E. Scuseria, and M. Ernzerhof, *"Hybrid functionals based on a screened Coulomb potential"*, J. Chem. Phys. **118**, 8207 (2003).

[42] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, *"Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces"*, Phys. Rev. Lett. **100**, 136406 (2008).

[43] M. Städele, J. A. Majewski, P. Vogl, and A. Görling, *"Exact Kohn-Sham exchange potential in semiconductors"*, Phys. Rev. Lett. **79**, 2089 (1997).

[44] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, *"Band parameters for III-V compound semiconductors and their alloys"*, J. Appl. Phys. **89**, 5815 (2001).

[45] T. Kotani and M. van Schilfgaarde, *"All-electron GW approximation with the mixed basis expansion based on the full-potential LMTO method"*, Solid State Comm. **121**, 461 (2002).

[46] C. Friedrich, M. Betzinger, M. Schlipf, S. Blügel, and A. Schindlmayr, *"Hybrid functionals and GW approximation in the FLAPW method"*, J. Phys.: Condens. Matter **24**, 293201 (2012).

[47] D. Waroquiers, A. Lherbier, A. Miglio, M. Stankovski, S. Poncé, M. J. T. Oliveira, M. Giantomassi, G.-M. Rignanese, and X. Gonze, *"Band widths and gaps from the Tran-Blaha functional: Comparison with many-body perturbation theory"*, Phys. Rev. B **87**, 075121 (2013).

[48] B. D. Malone and M. L. Cohen, *"Quasiparticle semiconductor band structures including spin-orbit interactions"*, J. Phys.: Condens. Matter **25**, 105503 (2013).

[49] J. Harl, L. Schimka, and G. Kresse, *"Assessing the quality of the random phase approximation for lattice constants and atomization energies of solids"*, Phys. Rev. B **81**, 115126 (2010).

[50] M. Usuda, N. Hamada, T. Kotani, and M. van Schilfgaarde, *"All-electron GW calculation based on the LAPW method: Application to wurtzite ZnO"*, Phys. Rev. B **66**, 125101 (2002).

[51] H. Dixit, R. Saniz, D. Lamoen, and B. Partoens, *"The quasiparticle band structure of zincblende and rocksalt ZnO"*, J. Phys.: Condens. Matter **22**, 125505 (2010).

[52] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, *"Combining GW calculations with exact-exchange density-functional theory: an analysis of valence-band photoemission for compound semiconductors"*, New J. Phys. **7**, 126 (2005).

[53] F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, *"Quasiparticle band structure based on a generalized Kohn-Sham scheme"*, Phys. Rev. B **76**, 115109 (2007).

[54] A. Ashrafi and C. Jagadish, *"Review of zincblende ZnO: Stability of metastable ZnO phases"*, J. Appl. Phys. **102**, 071101 (2007).

[55] B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, *"Quasiparticle Band Gap of ZnO: High Accuracy from the Conventional G^0W^0 Approach"*, Phys. Rev. Lett. **105**, 146401 (2010).

[56] C. Friedrich, M. C. Müller, and S. Blügel, *"Band convergence and linearization error correction of all-electron GW calculations: The extreme case of zinc oxide"*, Phys. Rev. B **83**, 081101(R) (2011).

[57] D. L. Johnson, *"Local field effects and the dielectric response matrix of insulators: A model"*, Phys. Rev. B **9**, 4475 (1974).

[58] M. Stankovski, G. Antonius, D. Waroquiers, A. Miglio, H. Dixit, K. Sankaran, M. Giantomassi, X. Gonze, M. Côté, and G.-M. Rignanese, *"G^0W^0 band gap of ZnO: Effects of plasmon-pole models"*, Phys. Rev. B **84**, 241201(R) (2011).

[59] J. A. Berger, L. Reining, and F. Sottile, *"Efficient GW calculations for SnO_2, ZnO, and rubrene: The effective-energy technique"*, Phys. Rev. B **85**, 085126 (2012).

[60] L. Y. Lim, S. Lany, Y. J. Chang, E. Rotenberg, A. Zunger, and M. F. Toney, *"Angle-resolved photoemission and quasiparticle calculation of ZnO: The need for d band shift in oxide semiconductors"*, Phys. Rev. B **86**, 235113 (2012).

[61] I. Abdolhosseini Sarsari, C. D. Pemmaraju, H. Salamati, and S. Sanvito, *"Many-body quasiparticle spectrum of Co-doped ZnO: A GW perspective"*, Phys. Rev. B **87**, 245118 (2013).

[62] Z. H. Levine and D. C. Allan, *"Linear Optical Response in Silicon and Germanium Including Self-Energy Effects"*, Phys. Rev. Lett. **63**, 1719 (1989).

[63] A. Marini, G. Onida, and R. Del Sole, *"Quasiparticle Electronic Structure of Copper in the GW Approximation"*, Phys. Rev. Lett. **88**, 016403 (2001).

[64] A. Marini, R. Del Sole, and G. Onida, *"First-principles calculation of the plasmon resonance and of the reflectance spectrum of silver in the GW approximation"*, Phys. Rev. B **66**, 115101 (2002).

[65] Z. Yi, Y. Ma, M. Rohlfing, V. M. Silkin, and E. V. Chulkov, *"Quasiparticle band structures and lifetimes in noble metals using Gaussian orbital basis sets"*, Phys. Rev. B **81**, 125125 (2010).

[66] T. Rangel, D. Kecik, P. E. Trevisanutto, G.-M. Rignanese, H. Van Swygenhoven, and V. Olevano, *"Band structure of gold from many-body perturbation theory"*, Phys. Rev. B **86**, 125125 (2012).

[67] I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen, *"Computational screening op perovskite metal oxides for optimal solar light capture"*, Energy Environ. Sci. **5**, 5814 (2012).

[68] I. E. Castelli, D. D. Landis, S. Dahl, K. S. Thygesen, I. Chorkendorff, T. F. Jaramillo, and K. W. Jacobsen, *"New cubic perovskites for one- and two-photon water splitting using the computational materials repository"*, Energy Environ. Sci. **5**, 9034 (2012).

[69] `http://materialsproject.org/`

[70] `http://www.fiz-karlsruhe.de/iscd_web.html`

[71] J. Heyd, G. E. Scuseria, and M. Ernzerhof, *"Hybrid functionals based on a screened Coulomb potential"*, J. Chem. Phys. **118**, 8207 (2003).

[72] A. V. Krukau, O. A. Vydrov, A. F. Izmaylov, and G. E. Scuseria, *"Influence of the exchange screening parameter on the performance of screened hybrid functionals"*, J. Chem. Phys. **125**, 224106 (2006).

[73] M. S. Hybertsen, *"Microscopic theory of heterojunction band offsets"*, Mat. Sci. Eng. B **14**, 254 (1992).

[74] J. P. A. Charlesworth, R. W. Godby, and R. J. Needs, *"First-Principles Calculations of Many-Body Band-Gap Narrowing at an Al/GaAs(110) Interface"*, Phys. Rev. Lett. **70**, 1685 (1993).

[75] J. C. Inkson, *"Many-body effects at metal-semiconductor junctions : II. The self energy and band structure distortion"*, J. Phys. C **6**, 1350 (1973).

[76] R. Decker, Y. Wang, V. W. Brar, W. Regan, H.-Z. Tsai, Q. Wu, W. Gannett, A. Zettl, and M. F. Crommie, *"Local Electronic Properties of Graphene on a BN Substrate via Scanning Tunneling Microscopy"*, Nano Lett. **11**, 2291 (2011).

[77] L. A. Ponomarenko, A. K. Geim, A. A. Zhukov, R. Jalil, S. V. Morozov, K. S. Novoselov, I. V. Grigorieva, E. H. Hill, V. V. Cheianov, V. I. Fal'ko, et al., *"Tunable metal-insulator transition in double-layer graphene heterostructures"* Nature Physics **7**, 958 (2011).

[78] H. Wang, T. Taychatanapat, A. Hsu, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, and T. Palacios, *"BN/Graphene/BN Transistors for RF Applications"* IEEE Electron Device Lett. **32**, 1209 (2011).

[79] S. J. Haigh, A. Gholinia, R. Jalil, S. Romani, L. Britnell, D. C. Elias, K. S. Novoselov, L. A. Ponomarenko, A. K. Geim, and R. Gorbachev, *"Cross-sectional imaging of individual layers and buried interfaces of graphene-based heterostructures and superlattices"* Nature Materials **11**, 764 (2012).

[80] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M. I. Katsnelson, L. Eaves, S. V. Morozov, et al., *"Field-Effect Tunneling Transistor Based on Vertical Graphene Heterostructures"* Science **335**, 947 (2012).

[81] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, *"Two-dimensional atomic crystals"*, Proc. Natl. Acad. Sci. U.S.A. **102**, 10451 (2005).

[82] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, *"Atomically thin MoS_2: A New Direct-Gap Semiconductor"*, Phys. Rev. Lett. **105**, 136805 (2010).

[83] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, *"Emerging Photoluminescence in Monolayer MoS_2"*, Nano Lett. **10**, 1271 (2010).

[84] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, *"Single-layer MoS_2 transistors"*, Nature Nanotechnology **6**, 147 (2011).

[85] E. K. Sichel, R. E. Miller, M. S. Abrahams, and C. J. Buiocchi, *"Heat capacity and thermal conductivity of hexagonal pyrolytic boron nitride"*, Phys. Rev. B **13**, 4607 (1979).

[86] L. Wirtz, A. Marini, and A. Rubio, *"Excitons in Boron Nitride Nanotubes: Dimensionality Effects"*, Phys. Rev. B **84**, 085404 (2011).

[87] C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, and M. Scheffler, *"Screening in two dimensions: GW calculations for surfaces and thin films using the repeated-slab approach"*, Phys. Rev. B **77**, 235428 (2008).

[88] H. C. Hsueh, G. Y. Guo, and S. G. Louie, *"Excitonic effects in the optical properties of a SiC sheet and nanotubes"*, Phys. Rev. B **84**, 085404 (2011).

[89] L. Yang, J. Deslippe, C.-H. Park, M. L. Cohen, and S. G. Louie, *"Excitonic Effects on the Optical Response of Graphene and Bilayer Graphene"*, Phys. Rev. Lett. **103**, 186802 (2009).

[90] P. E. Trevisanutto, C. Giorgetti, L. Reining, M. Ladisa, and V. Olevano, *"Ab Initio GW Many-Body Effects in Graphene"*, Phys. Rev. Lett. **101**, 226405 (2008).

[91] Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim, *"Experimental observation of the quantum Hall effect and Berry's phase in graphene"*, Nature **438**, 201 (2005).

[92] X. Blase, A. Rubio, S. G. Louie, and M. L. Cohen, *"Quasiparticle band structure of bulk hexagonal boron nitride and related systems"*, Phys. Rev. B **51**, 6868 (1995).

[93] G. Giovannetti, P. A. Khomyakov, G. Brocks, P. J. Kelly, and J. van den Brink, *"Substrate-induced band gap in graphene on hexagonal boron-nitride: Ab initio density functional calculations"*, Phys. Rev. B **76**, 073103 (2007).

[94] J. Yan, K. W. Jacobsen, and K. S. Thygesen, *"Optical properties of bulk semiconductors and graphene/boron nitride: The Bethe-Salpeter equation with derivative discontinuity-corrected density functional energies"*, Phys. Rev. B **86**, 045208 (2012).

[95] C. Hwang, D. A. Siegel, S.-K. Mo, W. Regan, A. Ismach, Y. Zhang, A. Zettl, and A. Lanzara, *"Fermi velocity engineering in graphene by substrate modification"*, Scientific Reports **2**, 590 (2012).

[96] A. Molina-Sánchez, D. Sangalli, K. Hummer, A. Marini, and L. Wirtz, *"Effect of spin-orbit interaction on the optical spectra of single-layer, double-layer, and bulk MoS_2"*, Phys. Rev. B **88**, 045412 (2013).

[97] H. Shi, H. Pan, Y.-W. Zhang, and B. I. Yakobsen, *"Quasiparticle band structures and optical properties of strained monolayer MoS_2 and WS_2"*, Phys. Rev. B **87**, 155304 (2013).

[98] T. Cheiwchanchamnangij and W. R. L. Lambrecht, *"Quasiparticle band structure calculation of monolayer, bilayer, and bulk MoS_2"*, Phys. Rev. B **85**, 205302 (2012).

[99] H.-P. Komsa and A. V. Krasheninnikov, *"Effects of confinement and environment on the electronic structure and exciton binding energy of MoS_2 from first principles"*, Phys. Rev. B **86**, 241201 (2012).

[100] C. Ataca and S. Ciraci, *"Functionalization of Single-Layer MoS_2 Honeycomb Structures"*, J. Phys. Chem. C **115**, 13303 (2011).

[101] A. Ramasubramaniam, *"Large excitonic effects in monolayers of molybdenum and tungsten dichalcogenides"*, Phys. Rev. B **86**, 115409 (2012).

[102] Y. Ding, Y. Wang, J. Ni, L. Shi, S. Shi, and W. Tang, *"First principles study of structural, vibrational and electronic properties of graphene-like MX_2 (M=Mo, Nb, W, Ta; X=S, Se, Te) monolayers"*, Physica B **406**, 2254 (2011).

[103] J. B. Neaton, M. S. Hybertsen, and S. G. Louie, *"Renormalization of Molecular Electronic Levels at Metal-Molecule Interfaces"*, Phys. Rev. Lett. **97**, 216405 (2006).

[104] J. M. Garcia-Lastra, C. Rostgaard, A. Rubio, and K. S. Thygesen, *"Polarization-induced renormalization of molecular levels at metallic and semiconducting surfaces"*, Phys. Rev. B **80**, 245427 (2009).

[105] K. S. Thygesen and A. Rubio, *"Renormalization of Molecular Quasiparticle Levels at Metal-Molecule Interfaces: Trends across Binding Regimes"*, Phys. Rev. Lett. **102**, 046802 (2009).

[106] C. Freysoldt, P. Rinke, and M. Scheffler, *"Controlling Polarization at Insulating Surfaces: Quasiparticle Calculations for Molecules Adsorbed on Insulator Films"*, Phys. Rev. Lett. **103**, 056803 (2009).

[107] K. S. Thygesen and A. Rubio, *"Nonequilibrium GW approach to quantum transport in nano-scale contacts"*, J. Chem. Phys. **126**, 091101 (2007).

[108] K. S. Thygesen and A. Rubio, *"Conserving GW scheme for nonequilibrium quantum transport in molecular contacts"*, Phys. Rev. B **77**, 115333 (2008).

[109] P. Darancet, A. Ferretti, D. Mayou, and V. Olevano, *"Ab initio GW electron-electron interaction effects in quantum transport"*, Phys. Rev. B **75**, 075102 (2007).

[110] M. Strange, C. Rostgaard, H. Häkkinen, and K. S. Thygesen, *"Self-consistent GW calculations of electronic transport in thiol- and amine-linked molecular junctions"*, Phys. Rev. B **83**, 115108 (2011).

[111] M. Strange and K. S. Thygesen, *"Towards quantitative accuracy in first-principles transport calculations: The GW method applied to alkane/gold junctions"*, Beilstein J. Nanotechnol. **2**, 746 (2011).

[112] `http://wiki.fysik.dtu.dk/ase/index.html`

[113] P. Umari, G. Stenuit, and S. Baroni, *"Optimal representation of the polarization propagator for large-scale GW calculations"*, Phys. Rev. B **79**, 201104(R) (2009).

[114] P. Umari, X. Qian, N. Marzari, G. Stenuit, L. Giacomazzi, and S. Baroni, *"Accelerating GW calculations with optimal polarizability basis"*, Phys. Status Solidi B **248**, 527 (2011).

[115] W. Kang and M. S. Hybertsen, *"Quasiparticle and optical properties of rutile and anastase TiO_2"*, Phys. Rev. B **82**, 085203 (2010).

[116] NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101 Release 15b, Aug 2011, Editor: Russell D. Johnson III; `http://cccbdb.nist.gov/`

[117] F. Bruneval and M. A. L. Marques, *"Benchmarking the Starting Points of the GW Approximation for Molecules"*, J. Chem. Theory Comput. **9**, 324 (2013).

[118] F. Caruso, P. Rinke, X. Ren, M. Scheffler, and A. Rubio, *"Unified description of ground and excited states of finite systems: The self-consistent GW approach"*, Phys. Rev. B **86**, 081102(R) (2012).

[119] C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, *"Fully self-consistent GW calculations for molecules"*, Phys. Rev. B **81**, 085103 (2010).

[120] A. Stan, N. E. Dahlen, and R. van Leeuwen, *"Levels of self-consistency in the GW approximation"*, J. Chem. Phys. **130**, 114105 (2009).

[121] L. Venkataraman, J. E. Klare, I. W. Tam, C. Nuckolls, M. S. Hybertsen, and M. L. Steigerwald, *"Single-Molecule Circuits with Well-Defined Molecular Conductance"*, Nano Lett. **6**, 458 (2006).

[122] S. Sharifzadeh, I. Tamblyn, P. Doak, P. T. Darancet, and J. B. Neaton, *"Quantitative molecular orbital energies within a G_0W_0 approximation"*, Eur. Phys. J. B **85**, 323 (2012).

[123] D. G. Streets, W. E. Hall, and G. P. Ceasar *"Mesomeric mixing in the π energy levels of amino-benzenes studied by photoelectron spectroscopy"*, Chem. Phys. Lett. **17**, 90 (1972).

[124] D. E. Cabelli, A. H. Cowley, and M. J. S. Dewar *"UPE studies of conjugation involving Group 5A elements. 1. Phenylphosphines"*, J. Am. Chem. Soc. **103**, 3286 (1981).

PAPERS

Paper I

"Quasiparticle GW calculations for solids, molecules, and two-dimensional materials"
Falco Hüser, Thomas Olsen, and Kristian S. Thygesen
Physical Review B **87**, 235132 (2013).

PHYSICAL REVIEW B **87**, 235132 (2013)

Quasiparticle GW calculations for solids, molecules, and two-dimensional materials

Falco Hüser,[1,*] Thomas Olsen,[1] and Kristian S. Thygesen[1,2,†]
[1]*Center for Atomic-scale Materials Design (CAMD), Department of Physics and Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
[2]*Center for Nanostructured Graphene (CNG), and Technical University of Denmark, 2800 Kgs. Lyngby, Denmark*
(Received 25 February 2013; revised manuscript received 23 May 2013; published 24 June 2013)

We present a plane-wave implementation of the G_0W_0 approximation within the projector augmented wave method code GPAW. The computed band gaps of ten bulk semiconductors and insulators deviate on average by 0.2 eV (∼5%) from the experimental values, the only exception being ZnO where the calculated band gap is around 1 eV too low. Similar relative deviations are found for the ionization potentials of a test set of 32 small molecules. The importance of substrate screening for a correct description of quasiparticle energies and Fermi velocities in supported two-dimensional (2D) materials is illustrated by the case of graphene/*h*-BN interfaces. Due to the long-range Coulomb interaction between periodically repeated images, the use of a truncated interaction is found to be essential for obtaining converged results for 2D materials. For all systems studied, a plasmon-pole approximation is found to reproduce the full frequency results to within 0.2 eV with a significant gain in computational speed. Throughout, we compare the G_0W_0 results with different exact exchange-based approximations. For completeness, we provide a mathematically rigorous and physically transparent introduction to the notion of quasiparticle states.

DOI: 10.1103/PhysRevB.87.235132 PACS number(s): 71.10.−w, 71.15.Mb, 71.20.Nr, 33.15.Ry

I. INTRODUCTION

For several decades, density functional theory (DFT)[1,2] has been the method of choice for electronic structure calculations due to its unique compromise between accuracy and efficiency. Large efforts have been made to develop better exchange-correlation (xc) functionals continuously pushing the quality of total energy calculations towards the limit of chemical accuracy. However, it is well known that the Kohn-Sham single-particle energies do not correspond to physical excitation energies, and in fact the widely used semilocal xc potentials significantly underestimate quasiparticle (QP) energy gaps.[3,4] For molecules and insulators better results can be obtained by replacing a fraction of the local exchange potential with the nonlocal Hartree-Fock exchange potential, as in the hybrid functionals. In the range-separated hybrids, the nonlocal exchange is used only for the short-range part of the potential. This improves the quality of semiconductor band structures and leads to faster convergence with k-point sampling, albeit at the cost of introducing an empirical cutoff radius. Still, the (range-separated) hybrids tend to underestimate the role of exchange in systems with weak screening, such as low-dimensional structures, and fail to account correctly for the spatial variation in the screening at metal-insulator interfaces (see below).

Many-body perturbation theory, on the other hand, offers a powerful and rigorous framework for the calculation of quasiparticle (QP) excitations. The key quantity is the electronic self-energy which is an energy-dependent and spatially nonlocal analogous of the xc potential of DFT. The self-energy can be systematically approximated by summing certain classes of perturbation terms to infinite order in the Coulomb interaction. The GW approximation[5] is the simplest approximation of this kind where the self-energy, Σ, is expanded to first order in the screened interaction. Symbolically it takes the form $\Sigma_{xc} = iGW$, where G is the Green's function and $W = \varepsilon^{-1}V$, is the screened interaction. Comparing the GW self-energy to the exchange potential, which can be written as $V_x = iGV$, we see that the GW self-energy is essentially a dynamically screened version of the exchange potential.

Apart from screening the static exchange potential, the replacement of the bare Coulomb interaction by the *dynamically* screened potential introduces correlation effects, which accounts for the interaction of an electron (or a hole) with the polarization charge that it induces in the medium. This is a highly nonlocal effect that becomes particularly evident at metal/insulator interfaces such as a molecule on a metal surface or the graphene/*h*-BN interfaces studied in the present work. For these systems, the correlation takes the form of an image charge effect that reduces the energy gap of the molecule or insulator by up to several electron volts.[6–11]

The GW approximation has been applied with great success to a broad class of materials ranging from bulk insulators, semiconductors, and metals to low-dimensional systems such as nanoclusters, surfaces and molecules (see, e.g., the reviews of Refs. 12–14). Beyond the calculation of QP energies, the GW method also serves as starting point for the calculation of optical spectra from the Bethe-Salpeter equation (BSE)[15–18] and for quantitatively accurate modeling of electron transport at metal-molecule interfaces where the alignment of the molecular energy levels with the metal Fermi level is particularly important.[19–23]

In principle, the GW self-energy should be evaluated self-consistently. However, due to the computational demands of such an approach, nonselfconsistent (G_0W_0) calculations with the initial G_0 obtained from the local density approximation (LDA) or similar, have traditionally been preferred.[24] Recently, fully self-consistent GW calculations have been performed for molecular systems yielding energies for the highest occupied orbitals with an absolute deviation from experiments of around 0.5 eV.[25,26] In comparison, the standard G_0W_0@LDA approach was found to yield slightly lower

accuracy, while better results were achieved when starting from Hartree-Fock or hybrid calculations.[25–28] For solids, earlier studies yielded contradictory conclusions regarding the accuracy of self-consistent versus nonselfconsistent GW calculations. More recently, the quasiparticle self-consistent GW method, in which the self-energy is evaluated with a self-consistently determined single-particle Hamiltonian, has been shown to yield excellent results for solids.[29–32]

On the practical side, any implementation of the GW approximation has to deal with similar numerical challenges. In addition to the already mentioned G_0W_0 approximation, it is common practice to evaluate the QP energies using first-order perturbation theory starting from the Kohn-Sham eigenvalues thereby avoiding the calculation of off-diagonal matrix elements of the self-energy. This approach is based on the assumption that the QP wave functions are similar to the Kohn-Sham wave functions. As recently shown for a metal-molecule interface this is sometimes far from being the case.[33] Another common simplification is the use of a plasmon pole approximation (PPA) for the dielectric function. The PPA leads to a considerable gain in efficiency by removing the need for evaluating the dielectric function at all frequency points and allowing the frequency convolution of G and W in the GW self-energy to be carried out analytically. In his original paper, Hedin introduced a static Coulomb hole and screened exchange (COHSEX) approximation to the full GW self-energy. The COHSEX approximation is computationally efficient and clearly illustrates the physics described by the GW approximation. However, its validity is limited to rather special cases and it should generally not be used for quantitative calculations.

In this paper we document the implementation of the G_0W_0 method in the GPAW open source electronic structure code.[34] GPAW is based on the projector augmented wave (PAW) method[35,36] and supports both real space grid and plane wave representation for high accuracy as well as numerical atomic orbitals (LCAO) for high efficiency. The G_0W_0 implementation is based on plane waves. The implementation supports both full frequency dependence (along the real axis) as well as the plasmon-pole approximation of Godby and Needs.[37] For low-dimensional systems, in particular two-dimensional (2D) systems, a truncated Coulomb interaction should be used to avoid the long-range interactions between periodically repeated unit cells. For both solids, molecules and 2D systems, we find that the PPA gives excellent results with significant reduction of the computational efforts. In contrast, the static COHSEX and the PBE0 hybrid yield unsatisfactory results. An interesting alternative to GW is offered by the local, orbital dependent potential of Gritsenko, Leeuwen, Lenthe, and Baerends with the modifications from Kuisma (GLLBSC), which explicitly adds the derivative discontinuity to the Kohn-Sham energy gap.[38] The GLLBSC band gaps for solids are found to lie on average within 0.4 eV of the G_0W_0 values but give similar accuracy when compared to experimental data. The GLLBSC ionization potentials of molecules are in average 1.5 eV below the G_0W_0 values.

The paper is organized as follows. Sec. II gives a general introduction to the theory of quasiparticle states. In Sec. III, we briefly review the central equations of the G_0W_0 method in a plane-wave basis and discuss some details of our implementation. In Sec. IV, we present results for bulk semiconductors, insulators and metals, comparing with experiments and previous calculations. The application to 2D systems is illustrated in Sec. V by the example of graphene on hexagonal boron nitride and the importance of screening effects on the QP energies is discussed. Finally, we test the implementation on finite systems by calculating the ionization potential of a set of 32 small molecules in Sec. VI.

II. QUASIPARTICLE THEORY

Quasiparticle states provide a rigorous generalization of the concept of single-particle orbitals to interacting electron systems. In this section we provide a compact, self-contained introduction to the general theory of quasiparticle states with a combined focus on physical interpretation and mathematical rigor. This presentation is completely formal; in particular we shall not discuss the physics and computation of specific self-energy approximations. Our presentation is thus complementary to most other papers on the GW method which tend to focus on the theory and derivation of the GW self-energy within the framework of many-body Green's function theory. To avoid inessential mathematical complications, we shall make the assumption that the system under consideration is finite and the relevant excitations are discrete.

A. Definition of QP energies and wave functions

We denote the N-particle many-body eigenstates and energies by $|\Psi_i^N\rangle$ and E_i^N, respectively. The occupied and unoccupied QP orbitals are denoted $|\psi_{i-}^{\mathrm{QP}}\rangle$ and $|\psi_{i+}^{\mathrm{QP}}\rangle$, respectively. These belong to the single-particle Hilbert space and are defined as:

$$\psi_{i-}^{\mathrm{QP}}(\mathbf{r})^* = \left\langle \Psi_i^{N-1} \middle| \hat{\Psi}(\mathbf{r}) \middle| \Psi_0^N \right\rangle \tag{1}$$

$$\psi_{i+}^{\mathrm{QP}}(\mathbf{r}) = \left\langle \Psi_i^{N+1} \middle| \hat{\Psi}^\dagger(\mathbf{r}) \middle| \Psi_0^N \right\rangle, \tag{2}$$

where $\hat{\Psi}(\mathbf{r})$ and $\hat{\Psi}^\dagger(\mathbf{r})$ are the field operators annihilating and creating an electron at point $\mathbf{r}$, respectively. The QP wave functions defined above are also sometimes referred to as Lehman amplitudes or Dyson orbitals.

The corresponding QP energies are defined by

$$\varepsilon_{i-}^{\mathrm{QP}} = E_0^N - E_i^{N-1} \tag{3}$$

$$\varepsilon_{i+}^{\mathrm{QP}} = E_i^{N+1} - E_0^N. \tag{4}$$

They represent the excitation energies of the $(N \pm 1)$-particle system relative to E_0^N and thus correspond to electron addition and removal energies. It is clear that $\varepsilon_{i+}^{\mathrm{QP}} > \mu$ while $\varepsilon_{i-}^{\mathrm{QP}} \leqslant \mu$ where μ is the chemical potential. Having noted this, we can in fact drop the $+/-$ subscripts on the QP states and energies. We shall do that in most of the following to simplify the notation.

The fundamental energy gap is defined as

$$E_{\mathrm{gap}} = \varepsilon_{0+}^{\mathrm{QP}} - \varepsilon_{0-}^{\mathrm{QP}} \tag{5}$$

$$= E_0^{N+1} + E_0^{N-1} - 2E_0^N. \tag{6}$$

We note that $E_{\rm gap}$ can also be expressed within the framework of Kohn-Sham (KS) theory as

$$E_{\rm gap} = \varepsilon^{\rm KS}_{N+1} - \varepsilon^{\rm KS}_{N} + \Delta_{xc}, \tag{7}$$

where $\varepsilon^{\rm KS}_n$ are the (exact) Kohn-Sham energies and Δ_{xc} is the derivative discontinuity.[39]

B. Interpretation of QP wave functions

Since the many-body eigenstates of an interacting electron system are not Slater determinants, the notion of single-particle orbitals is not well defined *a priori*. For weakly correlated systems we can, however, expect that the single-particle picture applies to a good approximation. To make this precise we ask to what extent the state $|\Psi^{N+1}_i\rangle$ can be regarded as a single-particle excitation from the ground state, i.e., to what extent it can be written on the form $c^\dagger_\phi|\Psi^N_0\rangle$ when ϕ is chosen in an optimal way. It turns out that the optimal ϕ is exactly the QP orbital. This statement follows simply from the observation[80]

$$\langle\phi|\psi^{\rm QP}_{i+}\rangle = \langle\Psi^{N+1}_i|\hat{c}^\dagger_\phi|\Psi^N_0\rangle, \tag{8}$$

for any orbital ϕ. Similarly, $|\psi^-_i\rangle$ is the orbital that makes $\hat{c}_\phi|\Psi^N_0\rangle$ the best approximation to the excited state $|\Psi^{N-1}_i\rangle$. Consequently, the QP wave function $\psi^{\rm QP}_{i\pm}$ is the single-particle orbital that best describes the state of the extra electron/hole in the excited state $|\Psi^{N\pm1}_i\rangle$.

From Eq. (8) it follows that the norm of a QP orbital is a measure of how well the true excitation can be described as a single-particle excitation. Precisely,

$$\|\psi^{\rm QP}_{i+}\| = \max_\phi\{\langle\Psi^{N+1}_i|\hat{c}^\dagger_\phi|\Psi^N_0\rangle,\ \|\phi\| = 1\rangle\}, \tag{9}$$

and similarly for the norm of $\psi^{\rm QP}_{i-}$.

The definition (1) implies a one-to-one correspondence between QP states and the excited many-body states $|\Psi^{N\pm1}_i\rangle$. Obviously, most of the latter are not even approximately of the single-particle type. These are characterized by a vanishing (or very small) norm of the corresponding QP orbital. In case of noninteracting electrons the QP states have norms 1 or 0. The former correspond to single excitations (Slater determinants) of the form $c^\dagger_n|\Psi^N_0\rangle$ while the latter correspond to multiple particle excitations, e.g., $c^\dagger_n c^\dagger_m c_k|\Psi^N_0\rangle$. Strictly speaking the term quasiparticle should be used only for those $|\psi^{\rm QP}_i\rangle$ whose norm is close to 1. The number of such states and whether any exists at all, depends on the system. For weakly correlated systems, one can expect a one-to-one correspondence between the QP states with norm $\sim$1 and the single-particle states of some effective noninteracting Hamiltonian, at least for the low-lying excitations.

C. Quasiparticle equation and self-energy

Below we show that QP states fulfill a generalized eigenvalue equation known as the QP equation, and we derive a useful expression for the norm of a QP state in terms of the self-energy.

The QP states and energies are linked to the single-particle Green's function via the Lehmann spectral representation[40]

$$G(z) = \sum_i \frac{|\psi^{\rm QP}_i\rangle\langle\psi^{\rm QP}_i|}{z - \varepsilon^{\rm QP}_i}, \tag{10}$$

where z is a complex number and it is understood that the sum runs over both occupied and unoccupied QP states. It follows that $G(z)$ is analytic in the entire complex plane except for the real points $\varepsilon^{\rm QP}_i$, which are simple poles. We note in passing that $G(z)$ equals the Fourier transform of the retarded (advanced) Green's function in the upper (lower) complex half plane.

The Green's function also satisfies the Dyson equation

$$G(z) = [z - H_0 - \Sigma_{xc}(z)]^{-1}, \tag{11}$$

where H_0 is the noninteracting part of the Hamiltonian including the Hartree field and Σ_{xc} is the exchange-correlation self-energy. The Dyson equation can be derived using many-body perturbation theory or it can simply be taken as the definition of the self-energy operator.

In the case where $\varepsilon^{\rm QP}_i$ belongs to the discrete spectrum, $\psi^{\rm QP}_i$ and $\varepsilon^{\rm QP}_i$ are solutions to the QP equation

$$\left[H_0 + \Sigma_{xc}\left(\varepsilon^{\rm QP}_i\right)\right]\left|\psi^{\rm QP}_i\right\rangle = \varepsilon^{\rm QP}_i\left|\psi^{\rm QP}_i\right\rangle. \tag{12}$$

This follows from the residue theorem by integrating the equation $[z - H_0 - \Sigma_{xc}(z)]G(z) = 1$ along a complex contour enclosing the simple pole $\varepsilon^{\rm QP}_i$.

The operator $[H_0 + \Sigma_{xc}(z)]$ is non-Hermitian and is diagonalized by a set of nonorthogonal eigenvectors,

$$[H_0 + \Sigma_{xc}(z)]|\psi_n(z)\rangle = \varepsilon_n(z)|\psi_n(z)\rangle. \tag{13}$$

Using these eigenvectors, the GF can be expressed in an alternative spectral form

$$G(z) = \sum_n \frac{|\psi_n(z)\rangle\langle\psi^n(z)|}{z - \varepsilon_n(z)}. \tag{14}$$

where $\{\psi^n(z)\}$ is the dual basis of $\{\psi_n(z)\}$, which by definition satisfies $\langle\psi_n(z)|\psi^m(z)\rangle = \delta_{nm}$.[81] We shall take the functions $\psi_n(z)$ to be normalized which also fixes the normalization of the dual basis.

In general, the vectors $\psi_n(z)$ do not have any physical meaning but are pure mathematical objects. An exception occurs for $z = \varepsilon^{\rm QP}_i$ where one of the vectors $\psi_n(\varepsilon^{\rm QP}_i)$ coincide with the QP orbital $\psi^{\rm QP}_i$ (except for normalization). We shall denote that vector by $\psi_i(\varepsilon^{\rm QP}_i)$, i.e.,

$$\left|\psi_i\left(\varepsilon^{\rm QP}_i\right)\right\rangle = \left|\psi^{\rm QP}_i\right\rangle\Big/\left\|\psi^{\rm QP}_i\right\|. \tag{15}$$

By equating the matrix element $\langle\psi^i(z)|G(z)|\psi_i(z)\rangle$ evaluated using the two alternative spectral representations Eq. (10) and Eq. (14), and integrating along a contour enclosing the pole $\varepsilon^{\rm QP}_i$, we obtain

$$\left\langle\psi^i\left(\varepsilon^{\rm QP}_i\right)\middle|\psi^{\rm QP}_i\right\rangle\left\langle\psi^{\rm QP}_i\middle|\psi_i\left(\varepsilon^{\rm QP}_i\right)\right\rangle = \frac{1}{1 - \varepsilon'_i\left(\varepsilon^{\rm QP}_i\right)}, \tag{16}$$

where the prime denotes the derivative with respect to z. This result follows by application of the residue theorem. Using

Eq. (15) it follows that the norm of the QP states is given by

$$\left\|\psi_i^{\mathrm{QP}}\right\|^2 = \left\langle\psi_i\left(\varepsilon_i^{\mathrm{QP}}\right)\middle|1-\Sigma'_{xc}\left(\varepsilon_i^{\mathrm{QP}}\right)\middle|\psi_i\left(\varepsilon_i^{\mathrm{QP}}\right)\right\rangle^{-1} \quad (17)$$

$$\equiv Z_i, \quad (18)$$

where we have used the Hellman-Feynman theorem to differentiate $\varepsilon_i(z) = \langle\psi_i(z)|H_0 + \Sigma_{xc}(z)|\psi_i(z)\rangle$.

D. Linearized QP equation

Given a self-energy operator, one must solve the QP equation to obtain the QP states and energies. This is complicated by the fact that the self-energy must be evaluated at the QP energies, which are not known *a priori*. Instead, one can start from an effective noninteracting Hamiltonian (in practice often the Kohn-Sham Hamiltonian),

$$[H_0 + V_{xc}]\left|\psi_i^s\right\rangle = \varepsilon_i^s\left|\psi_i^s\right\rangle, \quad (19)$$

and treat $\Sigma_{xc}(\varepsilon_i^{\mathrm{QP}}) - V_{xc}$ using first-order perturbation theory. Thus we write $\varepsilon_i^{\mathrm{QP}} = \varepsilon_i^s + \varepsilon_i^{(1)}$ with

$$\varepsilon_i^{(1)} = \left\langle\psi_i^s\middle|\Sigma_{xc}\left(\varepsilon_i^{\mathrm{QP}}\right) - V_{xc}\middle|\psi_i^s\right\rangle \quad (20)$$

$$= \left\langle\psi_i^s\middle|\Sigma_{xc}\left(\varepsilon_i^s\right) + \left(\varepsilon_i^{\mathrm{QP}} - \varepsilon_i^s\right)\Sigma'_{xc}\left(\varepsilon_i^s\right) - V_{xc}\middle|\psi_i^s\right\rangle. \quad (21)$$

Rearranging this equation yields

$$\varepsilon_i^{\mathrm{QP}} = \varepsilon_i^s + Z_i^s\cdot\left\langle\psi_i^s\middle|\Sigma_{xc}\left(\varepsilon_i^s\right) - V_{xc}\middle|\psi_i^s\right\rangle, \quad (22)$$

where

$$Z_i^s = \left\langle\psi_i^s\middle|1-\Sigma'_{xc}\left(\varepsilon_i^s\right)\middle|\psi_i^s\right\rangle^{-1} \quad (23)$$

approximates the true QP norm.

If $Z_i^s \ll 1$ we can conclude that ψ_i^s is not a (proper) QP state. There can be two reasons for this: (i) the electrons are strongly correlated and as a consequence the QP picture does not apply, or (ii) ψ_i^s is not a good approximation to the true QP wave function ψ_i^{QP}. While (i) is rooted in the physics of the underlying electron system, reason (ii) merely says that the Kohn-Sham orbital does not describe the true many-body excitations well. For an example where the QP picture is completely valid, i.e., all the QP states have norms very close to 1 or 0, but where simple noninteracting orbitals do not provide a good approximation to them, we refer to Ref. 33.

III. G_0W_0 APPROXIMATION

The self-energy of the GW approximation is given as a product of the (time-ordered) Green's function and screened Coulomb potential and can be split into an exchange and a correlation part, $\Sigma_{\mathrm{GW}} = V_x + \Sigma_c$, where V_x is the nonlocal Hartree-Fock exchange potential. The correlation contribution (which we from now on refer to as the self-energy $\Sigma = \Sigma_c$) is then evaluated by introducing the difference between the screened and the bare Coulomb potential $\overline{W} = W - V$,

$$\Sigma(\mathbf{r}t,\mathbf{r}'t') = iG(\mathbf{r}t,\mathbf{r}'t')\overline{W}(\mathbf{r}t,\mathbf{r}'t'), \quad (24)$$

which becomes a convolution in frequency domain

$$\Sigma(\mathbf{r},\mathbf{r}';\omega) = \frac{i}{2\pi}\int d\omega'\, G(\mathbf{r},\mathbf{r}';\omega+\omega')\overline{W}(\mathbf{r},\mathbf{r}';\omega'). \quad (25)$$

In this way, the exchange and the correlation contributions can be treated separately at different levels of accuracy. Additionally, the screened Coulomb potential approaches the bare one for large frequencies, so that $\overline{W}$ vanishes in this limit making the frequency integration numerically stable.

In the present G_0W_0 approach, the self-energy is constructed from Kohn-Sham wave functions $|n\mathbf{k}\rangle$ and eigenvalues $\varepsilon^s_{n\mathbf{k}}$, where n and $\mathbf{k}$ denote band and k-point index, respectively. Throughout this paper, spin indices are suppressed in order to simplify the notation.

Using the spectral representation for the Green's function in this basis and Fourier transforming to reciprocal space, the diagonal terms of the self-energy read[41]

$$\Sigma_{n\mathbf{k}} \equiv \langle n\mathbf{k}|\Sigma(\omega)|n\mathbf{k}\rangle = \frac{1}{\Omega}\sum_{\mathbf{GG}'}\sum_{\mathbf{q}}^{1.\mathrm{BZ}}\sum_{m}^{\mathrm{all}}\frac{i}{2\pi}\int_{-\infty}^{\infty}d\omega'\,\overline{W}_{\mathbf{GG}'}(\mathbf{q},\omega') \times\frac{\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G})\rho^{n\mathbf{k}*}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}')}{\omega+\omega'-\varepsilon^s_{m\,\mathbf{k}-\mathbf{q}}+i\eta\,\mathrm{sgn}\left(\varepsilon^s_{m\,\mathbf{k}-\mathbf{q}}-\mu\right)}, \quad (26)$$

where m runs over all bands, $\mathbf{q}$ covers the differences between all k points in the first Brillouin zone. The infinitesimal $\eta\to 0^+$ ensures the correct time ordering of the Green's function, $\Omega = \Omega_{\mathrm{cell}}N_{\mathbf{k}}$ is the total crystal volume, and μ is the chemical potential. The pair density matrix elements are defined as:

$$\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}) \equiv \langle n\mathbf{k}|e^{i(\mathbf{q}+\mathbf{G})\mathbf{r}}|m\,\mathbf{k}-\mathbf{q}\rangle. \quad (27)$$

The potential $\overline{W}_{\mathbf{GG}'}(\mathbf{q},\omega)$ is obtained from the symmetrized, time-ordered dielectric function in the random phase approximation (RPA)

$$\overline{W}_{\mathbf{GG}'}(\mathbf{q},\omega) = \frac{4\pi}{|\mathbf{q}+\mathbf{G}|}\left(\epsilon^{-1}_{\mathbf{GG}'}(\mathbf{q},\omega)-\delta_{\mathbf{GG}'}\right)\frac{1}{|\mathbf{q}+\mathbf{G}'|}. \quad (28)$$

The calculation of the dielectric function in the GPAW code is described in Ref. 42.

The quasi-particle spectrum is then calculated with Eq. (22) using first-order perturbation theory in $(\Sigma_{\mathrm{GW}} - V_{xc})$, where V_{xc} is the Kohn-Sham exchange-correlation potential

$$\varepsilon^{\mathrm{QP}}_{n\mathbf{k}} = \varepsilon^s_{n\mathbf{k}} + Z^s_{n\mathbf{k}}\,\mathrm{Re}\langle n\mathbf{k}|\Sigma\left(\varepsilon^s_{n\mathbf{k}}\right) + V_x - V_{xc}|n\mathbf{k}\rangle, \quad (29)$$

with a renormalization factor given by

$$Z^s_{n\mathbf{k}} = \left(1-\mathrm{Re}\langle n\mathbf{k}|\Sigma'\left(\varepsilon^s_{n\mathbf{k}}\right)|n\mathbf{k}\rangle\right)^{-1}, \quad (30)$$

where the derivative of the self-energy with respect to the frequency is calculated analytically from Eq. (26). The calculation of the exact exchange potential within GPAW is described in Ref. 34 using the plane-wave expressions of Ref. 43.

As discussed in the previous section, this first-order approach, i.e., using only the diagonal terms of the self-energy, is based on the assumption that the true QP wave functions and energies are similar to the Kohn-Sham wave functions and energies. To proceed beyond this approximation one must evaluate also the off-diagonal terms of the self-energy and invoke (partial) self-consistency. This is, however, beyond the scope of the present work. Similarly, the effect of electron-electron interactions on the QP lifetimes, which in principle can be deduced from the imaginary part of the GW self-energy, will not be considered in this study.

A. Frequency grid

For a fully frequency-dependent GW calculation, the dielectric matrix and thus the screened potential is evaluated on a user-defined grid of real frequencies and the integration in Eq. (26) is performed numerically. The frequency grid is chosen to be linear up to ω_{lin} with a spacing of $\Delta\omega$, which typically is set to 0.05 eV. Above ω_{lin} the grid spacing grows linearly up to a maximum frequency, ω_{max}. In practice we set ω_{max} to equal the maximum transition energy and $\omega_{\text{lin}} \approx (1/4)\cdot\omega_{\text{max}}$ which results in a few thousand frequency points. Compared to a fully linear grid, the use of a nonuniform grid gives a computational speedup of around a factor 2–3 without any loss of accuracy. The broadening parameter η is set to $4\Delta\omega$ to ensure a proper resolution of all spectral features.

B. Plasmon-pole approximation

In the plasmon-pole approximation (PPA), the frequency dependence of the dielectric function $\epsilon^{-1}_{\mathbf{GG'}}(\mathbf{q},\omega)$ is modeled as a single-pole approximation

$$\varepsilon^{-1}_{\mathbf{GG'}}(\mathbf{q},\omega) = R_{\mathbf{GG'}}(\mathbf{q})\left(\frac{1}{\omega-\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})+i\eta} - \frac{1}{\omega+\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})-i\eta}\right). \tag{31}$$

The plasmon frequency $\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})$ and the (real) spectral function $R_{\mathbf{GG'}}(\mathbf{q})$ are determined by fitting this function to the dielectric matrix given at the frequency points $\omega_1 = 0$ and $\omega_2 = iE_0$:

$$\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q}) = E_0\sqrt{\frac{\varepsilon^{-1}_{\mathbf{GG'}}(\mathbf{q},\omega_2)}{\varepsilon^{-1}_{\mathbf{GG'}}(\mathbf{q},\omega_1)-\varepsilon^{-1}_{\mathbf{GG'}}(\mathbf{q},\omega_2)}}, \tag{32}$$

$$R_{\mathbf{GG'}}(\mathbf{q}) = -\frac{\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})}{2}\varepsilon^{-1}_{\mathbf{GG'}}(\mathbf{q},\omega_1). \tag{33}$$

Using the relation

$$\lim_{\eta\to 0^+}\frac{1}{x\pm i\eta} = \mathcal{P}\left\{\frac{1}{x}\right\} \mp i\pi\delta(x), \tag{34}$$

where $\mathcal{P}$ denotes the Cauchy principal value, the spectral function of the screened potential, $\text{Im}\{\overline{W}_{\mathbf{GG'}}(\mathbf{q},\omega)\}$, is simply a δ function at the plasmon frequencies $\pm\tilde{\omega}_{\mathbf{GG'}}(\mathbf{q})$. Similarly, the relation (34) can be used in Eq. (26) allowing the GW self-energy to be evaluated analytically.

The PPA is expected to be a good approximation, when the overall structure of the dielectric function is dominated by a single (complex) pole. The true dielectric function will show variations on a finer scale. However, these are averaged out by the frequency integration in Eq. (26). In practice, we set the free parameter, E_0, to 1 Hartree in all our calculations and we find results to be insensitive to variations of around 0.5 Hartree.

C. Static COHSEX

By setting $\omega - \varepsilon_{m\,\mathbf{k}-\mathbf{q}} = 0$ in Eq. (26), the self-energy becomes frequency independent and can be split into two parts, named Coulomb hole and screened exchange.[44] The first term arises from the poles of the screened potential and describes the local interaction of an electron with its induced charge

$$\Sigma^{\text{COH}} = \tfrac{1}{2}\delta(\mathbf{r}-\mathbf{r}')[W(\mathbf{r},\mathbf{r}';\omega=0) - V(\mathbf{r},\mathbf{r}')]. \tag{35}$$

The plane-wave expression for a matrix element on a Bloch state $|n\mathbf{k}\rangle$ becomes

$$\Sigma^{\text{COH}}_{n\mathbf{k}} = \frac{1}{2\Omega}\sum_{\mathbf{GG'}}\sum_{\mathbf{q}}\sum_{m}^{\text{all}}\overline{W}_{\mathbf{GG'}}(\mathbf{q},0)\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G})\rho^{n\mathbf{k}*}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}'). \tag{36}$$

The second term originates from the poles of the Green's function and is identical to the exchange term in Hartree-Fock theory with the Coulomb kernel replaced by the screened interaction

$$\Sigma^{\text{SEX}} = -\sum_{j}^{\text{occ}}\phi^*_j(\mathbf{r})\phi_j(\mathbf{r}')W(\mathbf{r},\mathbf{r}';\omega=0), \tag{37}$$

which yields the matrix element

$$\Sigma^{\text{SEX}}_{n\mathbf{k}} = -\frac{1}{\Omega}\sum_{\mathbf{GG'}}\sum_{\mathbf{q}}\sum_{m}^{\text{occ}}W_{\mathbf{GG'}}(\mathbf{q},0)\rho^{n\mathbf{k}}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G})\rho^{n\mathbf{k}*}_{m\mathbf{k}-\mathbf{q}}(\mathbf{G}'). \tag{38}$$

The quasiparticle energies are then given as

$$\varepsilon^{\text{QP}}_{n\mathbf{k}} = \varepsilon^{s}_{n\mathbf{k}} + \langle n\mathbf{k}|\Sigma^{\text{SEX}} + \Sigma^{\text{COH}} - V_{xc}|n\mathbf{k}\rangle. \tag{39}$$

D. Coulomb divergence

For $\mathbf{q}\to 0$, the head, $\overline{W}_{\mathbf{00}}(\mathbf{q})$, and wings, $\overline{W}_{\mathbf{G0}}(\mathbf{q})$, $\overline{W}_{\mathbf{0G'}}(\mathbf{q})$, of the screened potential diverge as $1/q^2$ and $1/q$, respectively. These divergences are, however, integrable. In the limit of a very fine k-point sampling we have $\sum_{\mathbf{q}}\to\frac{\Omega}{(2\pi)^3}\int dq\,4\pi q^2$, and thus we can replace the $\mathbf{q}=0$ term in the q sum of Eq. (26) by an integral over a sphere in reciprocal space with volume $\Omega_{\text{BZ}}/N_{\mathbf{k}}$. The head and wings of the screened potential then take the form

$$\overline{W}_{\mathbf{00}}(\mathbf{q}=0,\omega) = \frac{2\Omega}{\pi}\left(\frac{6\pi^2}{\Omega}\right)^{1/3}\left[\varepsilon^{-1}_{\mathbf{00}}(\mathbf{q}\to 0,\omega)-1\right], \tag{40}$$

$$\overline{W}_{\mathbf{G0}}(\mathbf{q}=0,\omega) = \frac{1}{|\mathbf{G}|}\frac{\Omega}{\pi}\left(\frac{6\pi^2}{\Omega}\right)^{2/3}\varepsilon^{-1}_{\mathbf{G0}}(\mathbf{q}\to 0,\omega), \tag{41}$$

with the dielectric function evaluated in the optical limit.[42]

E. Coulomb truncation

In order to avoid artificial image effects in supercell calculations of systems, which are nonperiodic in one direction (2D systems), we follow Ref. 45 and cut off the Coulomb interaction by a step function in the nonperiodic direction (z axis)

$$\tilde{v}^{2D}(\mathbf{r}) = \frac{\theta(R-|r_z|)}{|\mathbf{r}|}, \tag{42}$$

where R is the truncation length. In reciprocal space, this becomes

$$\tilde{v}^{2D}(\mathbf{G}) = \frac{4\pi}{\mathbf{G}^2}\left[1+e^{-G_\parallel R}\left(\frac{G_z}{G_\parallel}\sin(G_z R) - \cos(|G_z|R)\right)\right], \tag{43}$$

where $G_\parallel$ and G_z are the parallel and perpendicular components of $\mathbf{G}$, respectively. By setting R to half the length of the unit cell in z direction, this simplifies to[46]

$$\tilde{v}^{2D}(\mathbf{G}) = \frac{4\pi}{\mathbf{G}^2}(1 - e^{-G_\parallel R}\cos(|G_z|R)). \quad (44)$$

Since Eq. (43) and thereby Eq. (44) are not well defined for $G_\parallel \to 0$, we have to evaluate these terms by numerical integration

$$\tilde{v}^{2D}(G_\parallel = 0) = \frac{1}{\Omega'}\int_{\Omega'} d\mathbf{q}'\, \tilde{v}^{2D}(G_z + \mathbf{q}'), \quad (45)$$

where Ω' is a small BZ volume around $G_\parallel = 0$. This integral is well defined and converges easily for a fine grid $\mathbf{q}'$ not containing the Γ point.

We mention that other methods have been applied to correct for the spurious long-range interaction in GW calculations for surfaces.[47,48]

F. Computational details

The calculation of one matrix element of the self-energy of Eq. (26) scales as $N_\omega \times N_b \times N_k^2 \times N_G^2$ with number of frequency points, bands, k points, and plane waves, respectively. The code is parallelized over $\mathbf{q}$ vectors. For calculations including the Γ point only, i.e., isolated systems, full parallelization over bands is used instead. Therefore, the computational time scales linearly with the number of cores. The screened potential $\overline{W}_{\mathbf{GG}'}(\mathbf{q},\omega)$ is evaluated separately for every $\mathbf{q}$ as an array in $\mathbf{G}$, $\mathbf{G}'$ and ω. For large numbers of plane waves and frequency points, this array can be distributed onto different cores, thus reducing the memory requirement on every core.

In practice, the use of the plasmon-pole approximation gives a computational speedup of a factor of 5–20 on average compared to a full frequency calculation. For both methods (PPA and full frequency integration), the computational time spent on the evaluation of the dielectric matrix and on the calculation of the quasiparticle spectrum from the screened potential is comparable.

IV. SOLIDS

As a first application, we calculate the band structures of ten simple semiconductors and insulators ranging from Si to LiF thus covering a broad range of band gap sizes of both direct and indirect nature. We compare the different approximation schemes within nonselfconsistent GW, namely (i) full frequency dependence, (ii) plasmon-pole approximation, and (iii) static COHSEX. In all these cases the self-energy is calculated with orbitals and single-particle energies obtained from an LDA calculation, i.e., G_0W_0@LDA. In addition we perform nonselfconsistent Hartree-Fock (HF), as well as PBE0 hybrid calculations in both cases using LDA orbitals. Finally, we compare to self-consistent GLLBSC[38,49] calculations. The GLLBSC is based on the PBEsol correlation potential and uses an efficient approximation to the exact exchange optimized effective potential which allows for explicit evaluation of the derivative discontinuity, Δ_{xc}. We have recently applied the GLLBSC in computational screening studies of materials for photocatalytic water splitting.[50,51] Here we present a systematic assessment of its performance by comparing to experiments and GW results for various types of systems. The bulk structures and the used lattice constants are listed in Table I.

TABLE I. Geometric structures.

	Structure	Lattice constant in Å
Si	Diamond	5.431
InP	Zincblende	5.869
GaAs	Zincblende	5.650
AlP	Zincblende	5.451
ZnO	Zincblende	4.580
ZnS	Zincblende	5.420
C	Diamond	3.567
BN	Zincblende	3.615
MgO	Rocksalt	4.212
LiF	Rocksalt	4.024

All calculations were performed with the GPAW code, which is based on the projector augmented wave method and supports both real space and plane-wave representations. In the present work only the plane-wave basis set has been used. The same set of parameters is used for the calculation of the dielectric matrix and the self-energy. For all GW calculations, convergence with respect to the plane wave cutoff, number of unoccupied bands and k points has been tested carefully, together with the size of the frequency grid for the full frequency calculations. As an example, Fig. 1 shows the dependence of the G_0W_0 band gap of zinc oxide on the plane-wave cutoff and the number of k points. For cutoff energies above 100 eV (corresponding to around 200 plane waves and bands), the value of the band gap is converged to within 0.02 eV, whereas increasing the number of k points results in a constant shift. For all the solids we have investigated, the band gap is well converged with $E_{\text{cut}} =$ 200–300 eV and a few hundred empty bands. For materials with direct band gaps ($9 \times 9 \times 9$) k points were found to be sufficient, whereas for AlP, BN, C, Si, and ZnS, which have indirect gaps, ($15 \times 15 \times 15$) k points were used in order to clearly resolve the conduction band minimum.

The results for the band gaps are summarized in Fig. 2 and Table II along with experimental data. The last row shows

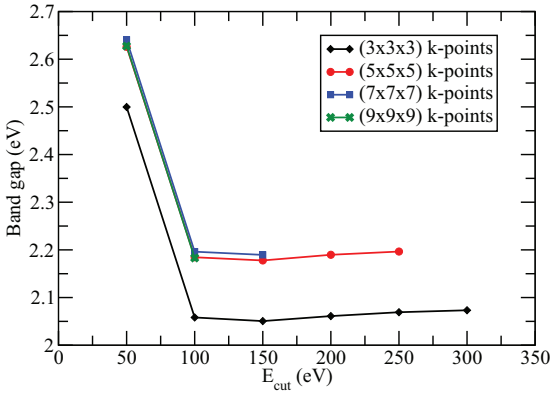


FIG. 1. (Color online) Convergence of the band gap of zinc oxide for G_0W_0@LDA with the plasmon-pole approximation. The number of bands is chosen equally to the number of plane waves corresponding to the respective cutoff energy; for example 300 eV corresponds to ∼1100 plane waves and bands.

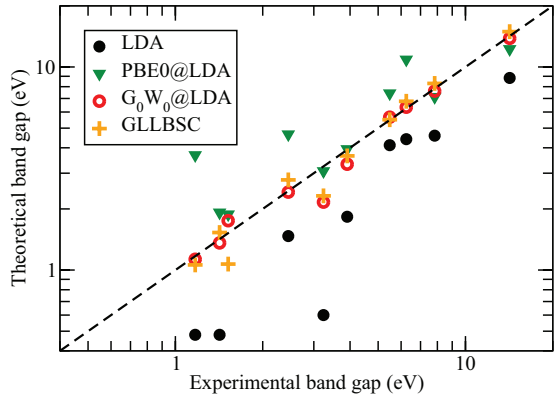


FIG. 2. (Color online) Comparison of calculated and experimental band gaps for the solids listed in Table I. The numerical values are listed in Table II. A logarithmic scale is used for better visualization. G_0W_0@LDA refers to the fully frequency-dependent nonselfconsistent GW based on LDA. The PBE0 results are obtained nonselfconsistently using LDA orbitals.

the mean absolute errors (MAE) of each method relative to experiment.

As expected LDA predicts much too small band gaps with relative errors as large as 400 % in the case of GaAs. In contrast HF greatly overestimates the band gap for all systems yielding even larger relative errors than LDA and with absolute errors exceeding 7 eV. The failure of HF is particularly severe for systems with narrow band gaps like Si and InP where the relative error is up to 500% whereas the error for the large gap insulator LiF is 50%. This difference can be understood from the relative importance of screening (completely neglected in HF) in the two types of systems. The PBE0 results lie in between LDA and HF with band gaps lying somewhat closer to the experimental values, however, still significantly overestimating the size of the gap for systems with small to intermediate band gap.

The inclusion of static screening within the COHSEX approximation significantly improves the bare HF results. However, with a MAE of 1.59 eV, the results are still unsatisfactory and there seems to be no systematic trend in the deviations from experiments, except for a slightly better performance for materials with larger band gaps. We mention that a detailed discussion of the drawbacks of COHSEX and how to correct its main deficiencies can be found in Ref. 52. In Ref. 53, the static COHSEX approximation was explored as a starting point for G_0W_0 calculations and compared to quasiparticle self-consistent GW calculations. However, no systematic improvement over the LDA starting point was found.

Introducing dynamical screening in the self-energy brings the band gaps much closer to the experimental values. The G_0W_0 calculations with the PPA and full frequency dependence yield almost identical results, with only small deviations of about 0.2 eV for the large band gap systems LiF and MgO, where the fully frequency-dependent method performs slightly better.

Our results agree well with previous works for G_0W_0 calculations using LDA[54] and PBE[29] as starting points with mean absolute errors of 0.31 and 0.21 eV in comparison, respectively. Compared to Ref. 29, the only significant deviations can be seen for GaAs and the wide gap systems, where our calculated band gaps are somewhat larger. We expect that this is due to the difference between LDA and PBE as starting point. The values reported in Ref. 54 are all smaller than ours. A more detailed comparison is, however, complicated because of the differences in the implementations: Ref. 54 uses a mixed basis set in an all-electron linear muffin-tin orbital (LMTO) framework. We note that for LiF, the calculated band gap is

TABLE II. Band gaps in eV. The type of gap is indicated in the last column. The last row gives the mean absolute error compared to experiment. Experimental data is taken from Ref. 59. Note that the experimental data for ZnO refers to the wurtzite structure. We find the calculated band gap to be around 0.1 eV smaller in the zincblende than in the wurtzite structure for both LDA, G_0W_0 and GLLBSC. Experimental gap for InP taken from Ref. 63.

				G_0W_0@LDA					
	LDA	HF@LDA	PBE0@LDA	COHSEX	PPA	dyn	GLLBSC	Experiment	
Si	0.48	5.26	3.68	0.56	1.09	1.13	1.06	1.17	Indirect
InP	0.48	5.51	1.92	1.99[a]	1.38	1.36	1.53	1.42	Direct
GaAs	0.38	5.46	1.88[b]	3.77[c]	1.76	1.75	1.07	1.52	Direct
AlP	1.47	7.15	4.66	1.88	2.38	2.42	2.78	2.45	Indirect
ZnO	0.60	10.42[d]	3.07[e]	0.10	2.20	2.24	2.32	3.44	Direct
ZnS	1.83	9.43	3.94[f]	1.52	3.28	3.32	3.65	3.91	Direct
C	4.12	11.83	7.42	6.51	5.59	5.66	5.50	5.48	Indirect
BN	4.41	13.27	10.88	7.08	6.30	6.34	6.78	6.25	Indirect
MgO	4.59	14.84	7.12	10.30	7.44	7.61	8.30	7.83	Direct
LiF	8.83	21.86	12.25	16.02	13.64	13.84	14.93	14.20	Direct
MAE	2.05	5.74	1.52	1.59	0.35	0.31	0.41		

[a]COHSEX predicts an indirect band gap of 1.73 eV.
[b]PBE0 predicts an indirect band gap of 1.79 eV.
[c]COHSEX predicts an indirect band gap of 1.07 eV.
[d]HF predicts an indirect band gap of 9.73 eV.
[e]PBE0 predicts an indirect band gap of 2.83 eV.
[f]PBE0 predicts an indirect band gap of 3.80 eV.

strongly dependent on the lattice constant. With only a slightly smaller lattice constant of 3.972 Å, which is the experimental value corrected for zero-point anharmonic expansion effects,[55] the quasiparticle gap increases by 0.4 eV.

One well-known problematic case for the GW approximation is ZnO (both in the zincblende and the wurtzite structure). The calculated band gap in the present study at the G_0W_0@LDA level is about 1 eV too low, which is consistent with other previous G_0W_0 studies.[56–59] Recent G_0W_0 calculations employing pseudopotentials and the PPA[60] as well as all-electron G_0W_0[61] have attributed this discrepancy to a very slow convergence of the band gap with respect to the number of bands. This is, however, not in agreement with our PAW-based calculations, which are well converged with a cutoff energy of 100 eV and around 200 bands. We note that semicore d states of zinc are explicitly included in our calculations. The large differences of the results and the convergence behavior compared to Ref. 60 are most likely due to the use of different models for the plasmon-pole approximation. As discussed in Ref. 62, the use of a model dielectric function, which fulfills Johnson's f-sum rule (as the PPA of Hybertsen and Louie)[44] leads to a very slow convergence of the band gap of ZnO with respect to the number of plane waves and unoccupied bands and gives a result, which is 1 eV higher than obtained with the fully frequency-dependent method. With the PPA of Godby and Needs on the other hand, results converge considerably faster and agree remarkably well with the frequency-dependent method.

Our results are consistent with Ref. 29 who attributed the underestimation of the gap to the starting point (PBE in their case) and also showed that the eigenvalue-sc GW method yields a band gap of 3.20 eV in very good agreement with experiment.

The band gaps denoted GLLBSC in Table II have been obtained as the self-consistently determined Kohn-Sham band gap of a GLLBSC calculation with the estimated derivative discontinuity Δ_{xc} added. Compared to G_0W_0, this approach yields a slightly lower accuracy compared to experiment. On the other hand, the much lower computational cost of the GLLBSC (which is comparable to LDA) makes this method very attractive for band structure calculations of large systems.

We conclude that even single-shot GW calculations with the plasmon pole approximation reproduce the experimental results to 0.2 eV for most of the semiconductors. The largest deviations are observed for ZnO and LiF where the computed band gaps are around 1 and 0.5 eV too small, respectively. Both of these systems have strong ionic character and LDA is presumably not a good starting point—in particular the LDA wave functions might be too delocalized. In such cases, a different starting point based on, e.g., a hybrid or LDA + U might yield better results although a systematic improvement seems difficult to achieve in this way.[29]

In Fig. 3, we compare the band structure of diamond obtained with the LDA and G_0W_0@LDA approximation. The valence band maximum occurs at the Γ point and the conduction band minimum is situated along the Γ–X direction, resulting in an indirect band gap of 4.1 and 5.7 eV, respectively. We can see that the main effect of the G_0W_0 approximation lies in an almost constant shift of the LDA bands: Occupied bands are moved to lower energies, whereas the unoccupied bands are shifted up. This is thus an example where the effect of G_0W_0 is well described by a simple scissors operator.

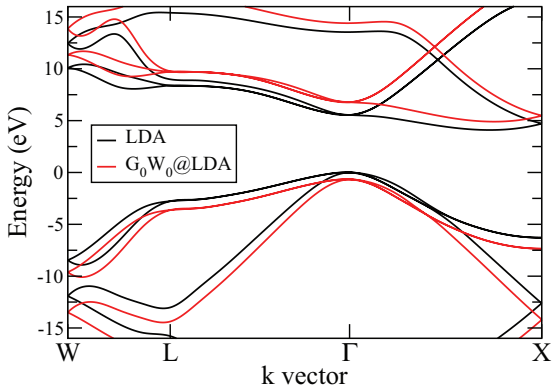


FIG. 3. (Color online) Band structure of diamond calculated with LDA (black) and G_0W_0 (red). The bands have been interpolated by splines from a (15 × 15 × 15) k-point sampling. The band gap is indirect between the Γ point and close to the X point with a value of 4.12 eV and 5.66 eV for LDA and G_0W_0, respectively.

Finally, we present the calculated band structure of gold in Fig. 4 as one example for a metallic system. The lattice parameter used for the fcc structure is 4.079 Å. The effect of GW is a small broadening of the occupied d bands, with the top being shifted slightly up and the bottom down in energy. The change in the low-lying s band and the unoccupied s-p band are significantly larger and inhomogeneous. Our band structure agrees well with the calculations of Ref. 64 with use of the plasmon-pole approximation and exclusion of 5s and 5p semicore states. In Ref. 64 it was also shown that QP self-consistent GW approximation shifts the d band down by 0.4 eV relative to PBE in good agreement with experiments.

V. 2D STRUCTURES

In this section we investigate the quasiparticle band structure of a two-dimensional structure composed of a single layer of hexagonal-boron nitride (h-BN) adsorbed on N layers

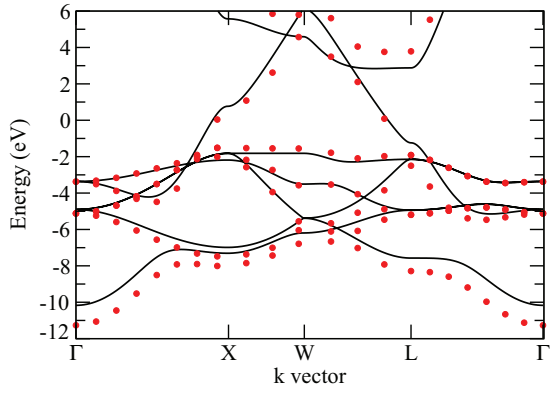


FIG. 4. (Color online) Band structure of fcc gold calculated with LDA (black lines) and G_0W_0@LDA with PPA (red dots). (45 × 45 × 45) and (15 × 15 × 15) k points have been used for LDA and GW, respectively. The bands are aligned to the respective Fermi level.

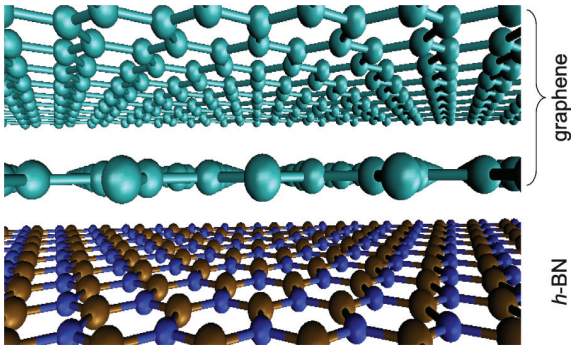


FIG. 5. (Color online) Schematic picture of the N-graphene/h-BN interface.

of graphene (as sketched in Fig. 5 for $N = 2$). Such 2D heterostructures have recently attracted much attention due to their unique physical properties and potential application in the next-generation electronic and photonic devices.[65–68]

Since graphene and h-BN are hexagonal structures with almost the same lattice constant, h-BN serves as a perfect substrate for graphene.[69] Based on LDA total energy calculations we find the most stable structure to be the configuration with one carbon over the B atom and the other carbon centered above a h-BN hexagon [equivalent to configuration (c) of Ref. 70] with a layer separation of 3.18 Å. The lattice constant is set to 2.5 Å for both lattices. The calculations are performed in the same way as described in the previous section with a k-point sampling of (45 × 45) in the in-plane direction. Also for this system we have found that the PPA yields almost identical results to the full frequency G_0W_0 and therefore all calculations presented in this section have been performed with the PPA.

The importance of truncating the Coulomb potential in order to avoid spurious interaction between neighboring supercells is shown in Fig. 6 for the direct gap at the K point for a freestanding boron nitride monolayer. Without truncation, the gap converges very slowly with the cell size and is still 0.3 eV below the converged value for 30 Å of vacuum. Applying the truncated Coulomb potential, the band

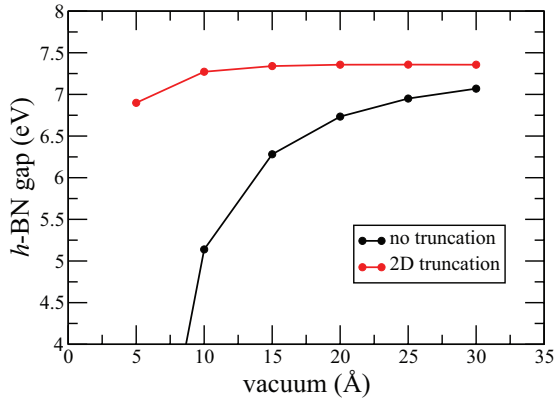


FIG. 6. (Color online) Direct G_0W_0 band gap at the K point for a freestanding h-BN sheet as function of the vacuum used to separate layers in neighboring supercell with and without use of the Coulomb truncation method as described in Sec. III E.

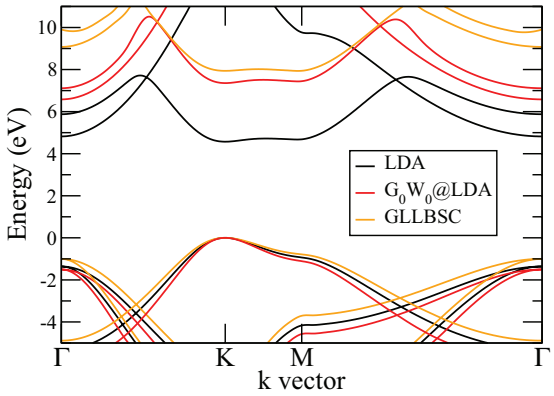


FIG. 7. (Color online) Band structure for a freestanding h-BN sheet. The band gap is direct at the K point with LDA (4.57 eV) and GLLBSC (7.94 eV) and changes to indirect between the K and the Γ point for G_0W_0 (7.37 eV).

gap is clearly converged already for 10 Å vacuum. These observations are consistent with recent G_0W_0 calculations for a SiC sheet, where the same trends were found.[71]

First, we summarize the most important features of the band structure calculations for the freestanding h-BN as shown in Fig. 7. LDA predicts a direct band gap at the K point of 4.57 eV and an indirect K–Γ transition of 4.82 eV. With GLLBSC, the bands are shifted significantly in energy. However, the shift is not constant for the different bands, resulting in a larger increase of the gap at the Γ point than at the K point. This yields 7.94 eV and 9.08 eV for the direct and indirect transition, respectively. The opposite is the case for G_0W_0@LDA calculations, which predict an indirect band gap of 6.58 eV and a direct transition at the K point of 7.37 eV. These values are 0.6 and 1.0 eV larger than the ones reported in Ref. 72 which were obtained from pseudopotential-based G_0W_0@LDA calculations. We note, however, that the amount of vacuum used in Ref. 72 was only 13.5 Å, which is not sufficient according to our results.

For the freestanding graphene (not shown), we find from the slope of the Dirac cone at the K point the Fermi velocity to be 0.87×10^6 m/s, 0.87×10^6 m/s, and 1.17×10^6 m/s with LDA, GLLBSC, and G_0W_0, respectively. This is in good agreement with previous G_0W_0 calculations, which obtained 1.15×10^6 m/s (Ref. 73) and 1.12×10^6 m/s (Ref. 74), respectively, and accurate magnetotransport measurements, which yielded 1.1×10^6 m/s (Ref. 75).

The band structure of graphene on a single h-BN sheet is shown in Fig. 8. At a qualitative level the band structure is similar to a superposition of the band structures of the isolated systems. In particular, due to the limited coupling between the layers, the bands closest to the Fermi energy can clearly be attributed to the different layers: At the K point, the linear dispersion of the graphene bands is maintained and the second highest valence and second lowest conduction band belong to the h-BN. However, there are important quantitative changes. First, the slope of the Dirac cone is reduced, giving a Fermi velocity of 1.01×10^6 m/s (0.78×10^6 m/s) with G_0W_0 (LDA). Exactly at the K point both LDA and G_0W_0 predict a small gap of 50 meV. Moreover, at the K point, the h-BN gap obtained with G_0W_0 is reduced from 7.37 eV for the

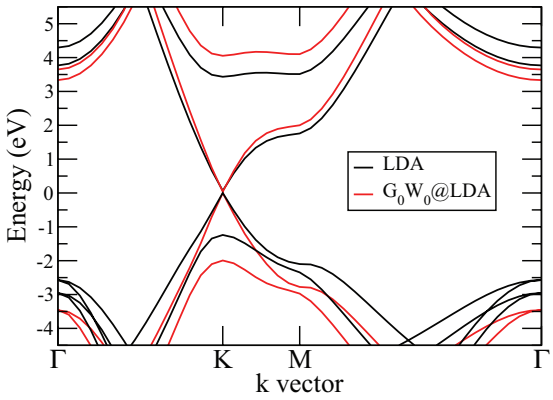


FIG. 8. (Color online) LDA and G_0W_0@LDA band structure for a graphene/boron nitride double layer structure. Only the two highest valence bands and the two lowest conduction bands are shown.

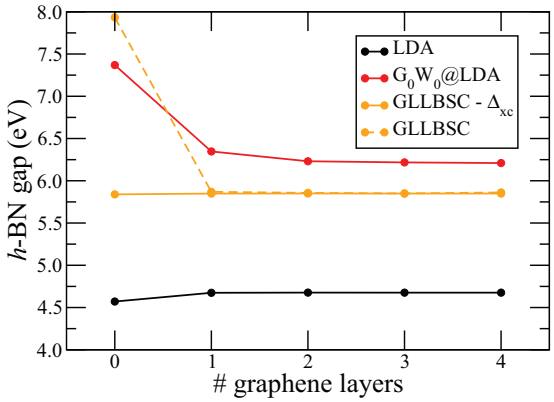


FIG. 10. (Color online) h-BN gap at the K point for different number of adsorbed graphene layers. GLLBSC results are plotted without and with the derivative discontinuity Δ_{xc}.

isolated sheet to 6.35 eV. In contrast the LDA gap is almost the same (4.67 eV) as for the isolated h-BN.

To further illustrate the importance of screening effects, we calculate the dependence of the h-BN gap with respect to the distance between the two layers. From Fig. 9, we can see that for LDA the gap is almost constant at the value of the freestanding boron nitride. For GLLBSC, the gap is around 1.2 eV larger but it does not change with the interlayer distance either. In contrast, GW predicts an increase of the gap with increasing distance and slowly approaches the value of the isolated system. The distance dependence of the gap is well fitted by $1/d$ as expected from a simple image charge model. Only for small distances, the results deviate from the $1/d$ dependence, most likely due to the formation of a chemical bond between the layers. We mention that the band gap closing due to substrate screening has been observed in previous GW studies of metal/semiconductor interfaces[6,7] as well as for molecules on metal surfaces.[8–11]

In Fig. 10, the size of the h-BN gap is shown for a varying number of graphene layers in a h-BN/N-graphene heterostructure. While LDA predicts a constant band gap of h-BN, G_0W_0 predicts a slight decrease of the gap with increasing number

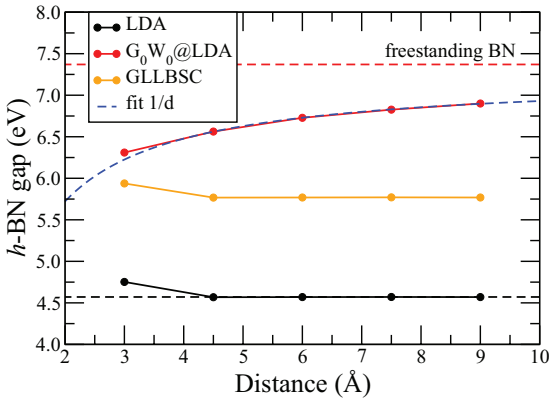


FIG. 9. (Color online) The band gap of h-BN at the K point as function of the distance to the graphene sheet (see inset). Dashed horizontal lines indicate the values for the freestanding h-BN, corresponding to $d \to \infty$.

of graphene layers due to enhanced screening. Additionally, we show the results for GLLBSC with and without the derivative discontinuity Δ_{xc} added to the Kohn-Sham gap. Due the construction of the GLLBSC, Δ_{xc} vanishes when one or more graphene layers are present because the system becomes (almost) metallic. Thus the GLLBSC gap becomes independent of the number of graphene layers, but is still close to the G_0W_0 result.

VI. MOLECULES

In this section, we present G_0W_0 calculations for a set of 32 small molecules. Recently a number of high-level GW studies on molecular systems have been published.[25–28] These studies have all been performed with localized basis sets and have explored the consequences of many of the commonly made approximations related to self-consistency, starting point dependence in the G_0W_0 approach, and treatment of core electrons. Here we use the more standard G_0W_0@LDA method and apply a plane-wave basis set. This is done in order to benchmark the accuracy of this scheme but also to show the universality of the present implementation in terms of the types of systems that can be treated.

Our calculations are performed in a supercell with 7 Å distance between neighboring molecules in all directions. As pointed out in the previous sections, careful convergence tests are crucial in order to obtain accurate results with GW. For a plane-wave basis we have found that this is particularly important for molecules, as demonstrated in Fig. 11 for water. Here, we plot the calculated ionization potential as a function of the inverse plane-wave cutoff. Again, for each data point, the number of bands is set equal to the number of plane waves corresponding to the cutoff. Even for $E_{\text{cut}} = 400$ eV ($1/E_{\text{cut}} = 0.0025\ \text{eV}^{-1}$ and corresponding to more than 8000 bands), the IP is not fully converged. However, for a cutoff larger than 100 eV, the IP grows linearly with $1/E_{\text{cut}}$ and this allows us extrapolate to the infinite cutoff (and number of empty bands) limit.[76,77] In this case the converged ionization potential is 12.1 eV, which is about 0.5 eV smaller than the experimental value. For all the molecules we have extrapolated the IP to infinite plane wave cutoff based on G_0W_0 calculations at cutoff energies 200–400 eV. Furthermore, as found for the solids

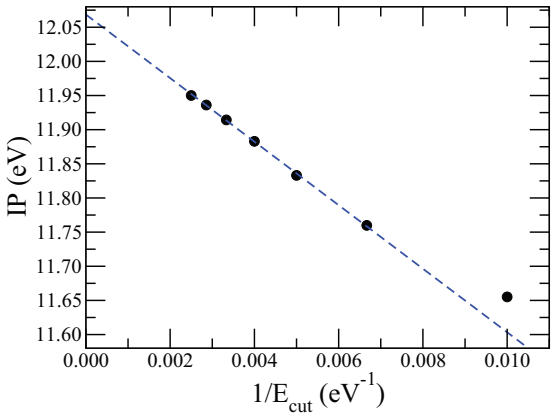


FIG. 11. (Color online) Convergence of the Ionization Potential for H_2O with respect to the plane wave cutoff for G_0W_0@LDA. The dashed line shows a linear fit of the points with $E_{cut} > 100$ eV ($1/E_{cut} < 0.01$ eV^{-1}). The IP is given as the negative HOMO energy.

and the 2D systems, the plasmon-pole approximation and the fully frequency-dependent GW calculations yield very similar results with typically 0.05 to 0.1 eV smaller IPs for the latter.

The results for all molecules are summarized and compared in Fig. 12. The LDA, PBE0 and GLLBSC calculations underestimate the IP with mean absolute errors (MAE) of 4.8, 3.5, and 2.0 eV, respectively. The opposite trend is observed for (nonselfconsistent) Hartree-Fock, which systematically overestimates the IP due to complete lack of screening. The MAE found for HF is 1.1 eV. We note that for an exact functional, according to the ionization-potential theorem, the Kohn-Sham energy of the highest occupied molecular orbital (HOMO) from DFT should be equal to the negative ionization potential.[39]

The G_0W_0 results are typically around 0.5 eV smaller than the experimental IPs, although there are a few exceptions where the calculated ionization potential is too large, and with a MAE of 0.56 eV. Recently, very similar studies have been reported for G_0W_0@LDA[27] with Gaussian basis sets and G_0W_0@PBE[28] in an all-electron framework using numerical atomic orbitals. Although there are differences of up to

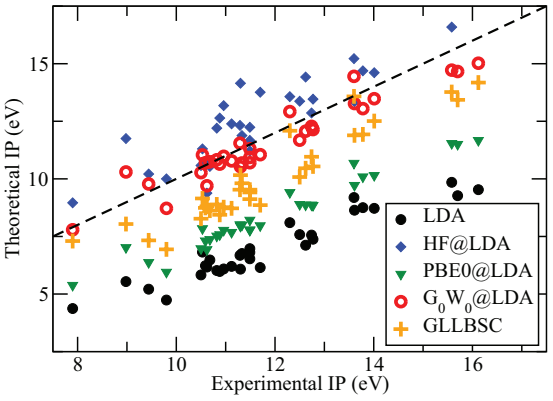


FIG. 12. (Color online) Comparison of theoretical and experimental ionization potentials. The G_0W_0 results are obtained by applying the extrapolation scheme as explained in the text. Corresponding values are listed in Table III.

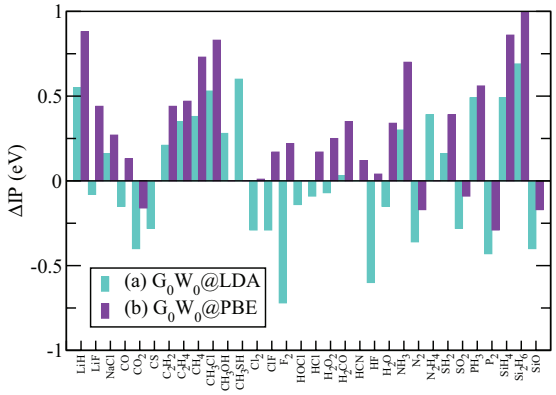


FIG. 13. (Color online) Deviations for the ionization potentials obtained with G_0W_0@LDA compared to (a) Ref. 27 and (b) Ref. 28. The mean deviations are 0.02 and 0.30 eV, respectively.

0.5 eV (both positive and negative), we find reasonable overall agreement with 0.32 eV MAE relative to Ref. 27. The mean signed error (MSE) is only 0.02 eV. Compared to Ref. 28, our results are systematically smaller with a MAE of 0.36 eV and a MSE of 0.30 eV. This is within the range of the accuracy of the different implementations, e.g., basis set, the PPA and the frozen core approximation applied in our calculations and the differences between LDA and PBE as starting points. A graphical comparison with these studies is shown in Fig. 13.

For detailed discussions of the role of self-consistency and other approximations we refer to Refs. 25,27,28, and 78.

VII. CONCLUSION

We have presented a plane-wave implementation of the single-shot G_0W_0 approximation within the GPAW projector augmented wave method code. The method has been applied to the calculation of quasiparticle band structures and energy levels in bulk crystals, 2D materials, and molecules, respectively. Particular attention has been paid to the convergence of the calculations with respect to the plane-wave cutoff and the number of unoccupied bands. While for all extended systems the value of the band gap was found to be converged at around 200 eV, the ionization potentials of the molecules required significantly higher cutoffs. In these cases, the data points were fit linearly to $1/E_{cut}$, allowing to extrapolate to infinite number of bands. For all calculations, the plasmon-pole approximation and the use of full frequency dependence of the dielectric function and the screened potential give very similar results. With these two observations, the computational demands can be drastically reduced without losing accuracy.

For the bulk semiconductors, we found good agreement with experimental results with a mean absolute error (MAE) of 0.2 eV. However, in the special case of zinc oxide and for the large gap insulators, the calculated band gaps were underestimated by 0.5–1 eV. These errors are most likely due to the lack of self-consistency and/or the quality of the LDA starting point used in our calculations. Similar conclusions apply to the 32 small molecules where the ionization potentials obtained from G_0W_0@LDA were found to underestimate the experimental values by around 0.5 eV on average. The important role of screening for the quasiparticle band structure

TABLE III. Calculated and experimental ionization potentials. All energies are in eV. Last row shows the mean absolute error (MAE) with respect to experiments. Experimental data taken from Ref. 79.

Molecule	LDA	HF@LDA	PBE0@LDA	GLLBSC	G_0W_0@LDA	Experiment
LiH	4.37	8.96	5.38	7.30	7.79	7.90
LiF	6.08	14.15	7.95	10.16	10.53	11.30
NaCl	4.74	10.00	5.95	6.94	8.72	9.80
CO	8.72	14.61	10.15	12.51	13.48	14.01
CO_2	8.75	14.69	10.09	11.93	13.05	13.78
CS	6.76	11.88	8.00	9.81	10.69	11.33
C_2H_2	6.81	11.21	7.79	9.41	11.22	11.49
C_2H_4	6.48	10.54	7.37	8.62	10.74	1 0.68
CH_4	9.19	15.22	10.68	13.58	14.45	13.60
CH_3Cl	6.68	12.32	8.01	9.53	11.55	11.29
CH_3OH	6.09	13.18	7.77	8.77	10.98	10.96
CH_3SH	5.21	10.21	6.37	7.33	9.78	9.44
Cl_2	6.53	11.67	7.77	9.12	10.93	11.49
ClF	7.38	13.46	8.85	10.54	12.14	12.77
F_2	9.27	18.44	11.50	13.43	14.66	15.70
HOCl	6.20	12.39	7.68	8.72	10.78	11.12
HCl	7.56	12.86	8.87	10.96	12.28	12.74
H_2O_2	6.15	13.76	7.97	8.86	11.05	11.70
H_2CO	5.98	12.64	7.58	8.44	10.64	10.88
HCN	8.64	13.35	9.72	11.89	13.27	13.61
HF	9.53	18.29	11.67	14.18	15.02	16.12
H_2O	7.12	14.42	8.87	10.46	12.07	12.62
NH_3	6.02	12.20	7.52	8.89	10.83	10.82
N_2	9.85	16.59	11.54	13.77	14.72	15.58
N_2H_4	5.54	11.75	7.02	8.04	10.30	8.98
SH_2	5.83	10.58	6.97	8.27	10.27	10.50
SO_2	7.58	13.37	8.89	10.08	11.68	12.50
PH_3	6.23	10.77	7.31	8.74	10.70	10.59
P_2	6.17	9.38	6.93	8.80	9.70	10.62
SiH_4	8.10	13.57	9.41	12.09	12.92	12.30
Si_2H_6	6.82	11.30	7.84	9.15	11.04	10.53
SiO	6.97	12.24	8.21	9.53	10.70	11.49
MAE	4.84	1.11	3.46	1.83	0.56	

was illustrated by the case of a 2D graphene/boron-nitride heterojunction. For this system, we found a truncation of the Coulomb potential to be crucial in periodic supercell calculations.

The G_0W_0 results were compared to band structures obtained with Hartree-Fock, the PBE0 hybrid, and the GLLBSC potential. While Hartree-Fock and PBE0 yield overall poor results, the computationally efficient GLLBSC results were found to be in surprisingly good agreement with G_0W_0 for the band gaps of semiconductors, while the ionization potentials of molecules were found to be 1.5 eV lower on average.

ACKNOWLEDGMENTS

We would like to thank Jun Yan and Jens Jørgen Mortensen for useful discussions and assistance with the coding. The authors acknowledge support from the Danish Council for Independent Research's Sapere Aude Program through Grant No. 11-1051390. The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation. The Catalysis for Sustainable Energy (CASE) initiative is funded by the Danish Ministry of Science, Technology and Innovation.

*falco.hueser@fysik.dtu.dk
†thygesen@fysik.dtu.dk

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[3] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).
[4] F. Bechstedt, F. Fuchs, and G. Kresse, Phys. Status Solidi B **246**, 1877 (2009).
[5] L. Hedin, Phys. Rev. **139**, A796 (1965).
[6] J. P. A. Charlesworth, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **70**, 1685 (1993).
[7] J. C. Inkson, J. Phys. C **6**, 1350 (1973).
[8] J. B. Neaton, M. S. Hybertsen, and S. G. Louie, Phys. Rev. Lett. **97**, 216405 (2006).
[9] J. M. Garcia-Lastra, C. Rostgaard, A. Rubio, and K. S. Thygesen, Phys. Rev. B **80**, 245427 (2009).

[10]K. S. Thygesen and A. Rubio, Phys. Rev. Lett. **102**, 046802 (2009).
[11]C. Freysoldt, P. Rinke, and M. Scheffler, Phys. Rev. Lett. **103**, 056803 (2009).
[12]W. G. Aulbur, L. Jönsson, and J. W. Wilkins, Solid State Phys. **54**, 1 (2000).
[13]F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[14]G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[15]S. Albrecht, L. Reining, R. Del Sole, and G. Onida, Phys. Rev. Lett. **80**, 4510 (1998).
[16]E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951).
[17]M. Rohlfing and S. G. Louie, Phys. Rev. B **62**, 4927 (2000).
[18]J. Yan, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **86**, 045208 (2012).
[19]K. S. Thygesen and A. Rubio, J. Chem. Phys. **126**, 091101 (2007).
[20]K. S. Thygesen and A. Rubio, Phys. Rev. B **77**, 115333 (2008).
[21]P. Darancet, A. Ferretti, D. Mayou, and V. Olevano, Phys. Rev. B **75**, 075102 (2007).
[22]M. Strange, C. Rostgaard, H. Häkkinen, and K. S. Thygesen, Phys. Rev. B **83**, 115108 (2011).
[23]M. Strange and K. S. Thygesen, Beilstein J. Nanotechnol. **2**, 746 (2011).
[24]G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. B **25**, 2867 (1982).
[25]C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **81**, 085103 (2010).
[26]X. Blase, C. Attaccalite, and V. Olevano, Phys. Rev. B **83**, 115103 (2011).
[27]F. Bruneval and M. A. L. Marques, J. Chem. Theory Comput. **9**, 324 (2013).
[28]F. Caruso, P. Rinke, X. Ren, M. Scheffler, and A. Rubio, Phys. Rev. B **86**, 081102(R) (2012).
[29]M. Shishkin and G. Kresse, Phys. Rev. B **75**, 235102 (2007).
[30]S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).
[31]M. P. Surh, S. G. Louie, and M. L. Cohen, Phys. Rev. B **43**, 9126 (1991).
[32]T. Kotani, M. van Schilfgaarde, S. V. Faleev, and A. Chantis, J. Phys.: Condens. Matter **19**, 365236 (2007).
[33]M. Strange and K. S. Thygesen, Phys. Rev. B **86**, 195121 (2012).
[34]J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dulak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen *et al.*, J. Phys.: Condens. Matter **22**, 253202 (2010).
[35]P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[36]P. E. Blöchl, C. J. Först, and J. Schimpl, Bull. Mater. Sci. **26**, 33 (2003).
[37]R. W. Godby and R. J. Needs, Phys. Rev. Lett. **62**, 1169 (1989).
[38]O. Gritsenko, R. van Leeuwen, E. van Lenthe, and E. J. Baerends, Phys. Rev. A **51**, 1944 (1995).
[39]J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, Jr., Phys. Rev. Lett. **49**, 1691 (1982).
[40]H. Bruus and K. Flensberg, *Many-Body Quantum Theory in Condensed Matter Physics - An Introduction* (Oxford University Press, Oxford, 2004).
[41]M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006).
[42]J. Yan, J. J. Mortensen, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **83**, 245122 (2011).
[43]A. Sorouri, W. M. Foulkes, and N. D. Hine, J. Chem. Phys. **124**, 064105 (2006).
[44]M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[45]C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, Phys. Rev. B **73**, 205119 (2006).
[46]S. Ismail-Beigi, Phys. Rev. B **73**, 233103 (2006).
[47]C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, R. W. Godby, and M. Scheffler, Comput. Phys. Commun. **176**, 1 (2007).
[48]C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, and M. Scheffler, Phys. Rev. B **77**, 235428 (2008).
[49]M. Kuisma, J. Ojanen, J. Enkovaara, and T. T. Rantala, Phys. Rev. B **82**, 115106 (2010).
[50]I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen, Energy Environ. Sci. **5**, 5814 (2012).
[51]I. E. Castelli, D. D. Landis, S. Dahl, K. S. Thygesen, I. Chorkendorff, T. F. Jaramillo, and K. W. Jacobsen, Energy Environ. Sci. **5**, 9034 (2012).
[52]W. Kang and M. S. Hybertsen, Phys. Rev. B **82**, 195108 (2010).
[53]F. Bruneval, N. Vast, and L. Reining, Phys. Rev. B **74**, 045102 (2006).
[54]T. Kotani and M. van Schilfgaarde, Solid State Commun. **121**, 461 (2002).
[55]J. Harl, L. Schimka, and G. Kresse, Phys. Rev. B **81**, 115126 (2010).
[56]M. Usuda, N. Hamada, T. Kotani, and M. van Schilfgaarde, Phys. Rev. B **66**, 125101 (2002).
[57]H. Dixit, R. Saniz, D. Lamoen, and B. Partoens, J. Phys.: Condens. Matter **22**, 125505 (2010).
[58]P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. **7**, 126 (2005).
[59]F. Fuchs, J. Furthmüller, F. Bechstedt, M. Shishkin, and G. Kresse, Phys. Rev. B **76**, 115109 (2007).
[60]B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **105**, 146401 (2010).
[61]C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B **83**, 081101(R) (2011).
[62]M. Stankovski, G. Antonius, D. Waroquiers, A. Miglio, H. Dixit, K. Sankaran, M. Giantomassi, X. Gonze, M. Côté, and G.-M. Rignanese, Phys. Rev. B **84**, 241201(R) (2011).
[63]I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, J. Appl. Phys. **89**, 5815 (2001).
[64]T. Rangel, D. Kecik, P. E. Trevisanutto, G.-M. Rignanese, H. Van Swygenhoven, and V. Olevano, Phys. Rev. B **86**, 125125 (2012).
[65]L. A. Ponomarenko, A. K. Geim, A. A. Zhukov, R. Jalil, S. V. Morozov, K. S. Novoselov, I. V. Grigorieva, E. H. Hill, V. V. Cheianov, V. I. Fal'ko *et al.*, Nature Phys. **7**, 958 (2011).
[66]H. Wang, T. Taychatanapat, A. Hsu, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, and T. Palacios, IEEE Electron Device Lett. **32**, 1209 (2011).
[67]S. J. Haigh, A. Gholinia, R. Jalil, S. Romani, L. Britnell, D. C. Elias, K. S. Novoselov, L. A. Ponomarenko, A. K. Geim, and R. Gorbachev, Nature Mater. **11**, 764 (2012).
[68]L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M. I. Katsnelson, L. Eaves, S. V. Morozov *et al.*, Science **335**, 947 (2012).
[69]R. Decker, Y. Wang, V. W. Brar, W. Regan, H.-Z. Tsai, Q. Wu, W. Gannett, A. Zettl, and M. F. Crommie, Nano Lett. **11**, 2291 (2011).
[70]G. Giovannetti, P. A. Khomyakov, G. Brocks, P. J. Kelly, and J. van den Brink, Phys. Rev. B **76**, 073103 (2007).

[71]H. C. Hsueh, G. Y. Guo, and S. G. Louie, Phys. Rev. B **84**, 085404 (2011).
[72]X. Blase, A. Rubio, S. G. Louie, and M. L. Cohen, Phys. Rev. B **51**, 6868 (1995).
[73]L. Yang, J. Deslippe, C.-H. Park, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **103**, 186802 (2009).
[74]P. E. Trevisanutto, C. Giorgetti, L. Reining, M. Ladisa, and V. Olevano, Phys. Rev. Lett. **101**, 226405 (2008).
[75]Y. Zhang, Y.-W. Tan, H. L. Stormer, and P. Kim, Nature (London) **438**, 201 (2005).
[76]W. Kang and M. S. Hybertsen, Phys. Rev. B **82**, 085203 (2010).
[77]P. Umari, X. Qian, N. Marzari, G. Stenuit, L. Giacomazzi, and S. Baroni, Phys. Status Solidi B **248**, 527 (2011).
[78]A. Stan, N. E. Dahlen, and R. van Leeuwen, J. Chem. Phys. **130**, 114105 (2009).
[79]NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101, Release 15b, Aug 2011, edited by Russell D. Johnson III; http://cccbdb.nist.gov/
[80]With the use of $c_\phi^\dagger = \int d\mathbf{r}\, \phi^*(\mathbf{r})\hat{\Psi}^\dagger(\mathbf{r})$.
[81]The dual basis functions are in fact the eigenvectors of the adjoint operator $[H_0 + \Sigma(z)]^\dagger$.

Paper II

"Stability and bandgaps of layered perovskites for one- and two-photon water splitting"
Ivano E. Castelli, Juan María García-Lastra, Falco Hüser, Kristian S. Thygesen, and Karsten W. Jacobsen
New Journal of Physics **15**, 105026 (2013).

New Journal of Physics
The open access journal for physics

Stability and bandgaps of layered perovskites for one- and two-photon water splitting

Ivano E Castelli[1,3], **Juan María García-Lastra**[1,2], **Falco Hüser**[1], **Kristian S Thygesen**[1] **and Karsten W Jacobsen**[1]

[1] Center for Atomic-Scale Materials Design, Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
[2] Nano-Bio Spectroscopy Group and ETSF Scientific Development Center, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, E-20018 San Sebastian, Spain
E-mail: ivca@fysik.dtu.dk

New Journal of Physics **15** (2013) 105026 (14pp)
Received 11 June 2013
Published 28 October 2013
Online at http://www.njp.org/
doi:10.1088/1367-2630/15/10/105026

Abstract. Direct production of hydrogen from water and sunlight requires stable and abundantly available semiconductors with well positioned band edges relative to the water red-ox potentials. We have used density functional theory (DFT) calculations to investigate 300 oxides and oxynitrides in the Ruddlesden–Popper phase of the layered perovskite structure. Based on screening criteria for the stability, bandgaps and band edge positions, we suggest 20 new materials for the light harvesting photo-electrode of a one-photon water splitting device and 5 anode materials for a two-photon device with silicon as photo-cathode. In addition, we explore a simple rule relating the bandgap of the perovskite to the number of octahedra in the layered structure and the B-metal ion. Finally, the quality of the GLLB-SC potential used to obtain the bandgaps, including the derivative discontinuity, is validated against G_0W_0@LDA gaps for 20 previously identified oxides and oxynitrides in the cubic perovskite structure.

[3] Author to whom any correspondence should be addressed

Contents

1. Introduction

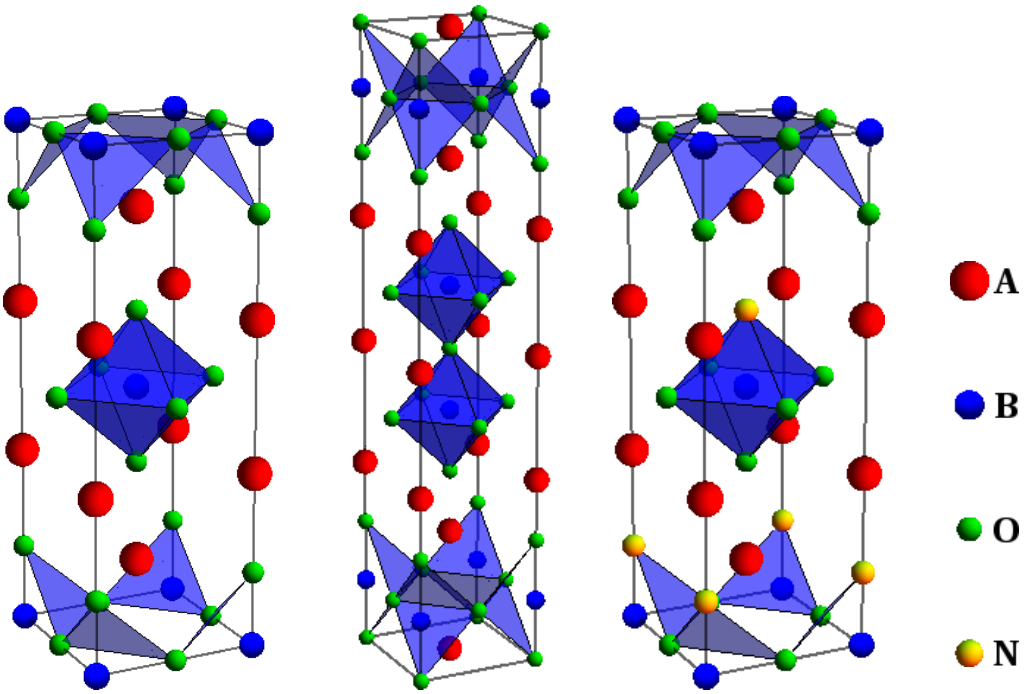
The direct conversion of solar light into chemical fuels through electrochemical reactions represents a clean, sustainable and potentially cheap alternative to fossil fuels. The simplest reaction of this kind is the water splitting (WS) reaction in which water is split into hydrogen and oxygen. In the most basic photo-electrochemical device, each photon is harvested by a single semiconductor and the created electron–hole pair is used to evolve hydrogen and oxygen. The maximum efficiency of such a one-photon device is only around 7% when overpotential of the reactions and losses are accounted for [1]. Significantly higher energy conversion efficiencies of up to 27% [2] can be achieved using a combination of two or more semiconductors with appropriately aligned band edges.

First attempts of photocatalytic WS were made in the 1970s using TiO_2 as the light harvesting and hydrogen evolving material [3]. Due to the large bandgap of TiO_2 and its poor catalytic properties the devices had extremely low efficiencies. Since then several materials have been proposed as light harvesting photo-electrodes for WS both in the UV and in the visible range [4]. Still, the efficiencies of the WS devices are significantly lower than the alternative combination of electrolysis driven by standard photovoltaic cells.

The search for new materials can be guided by *ab initio* quantum mechanical calculations, avoiding expensive 'trial and error' experimental processes. Recently, high-throughput materials design of stable binary and ternary alloys [5], carbon capture and storage [6], batteries [7], photovoltaic [8, 9] and WS materials [2, 10] have been reported. A number of databases have been also implemented to store and to analyze the huge amount of computed data generated. Some examples are the Materials Project database[4], the AFLOWLIB consortium [5] and the Computational Materials Repository[5].

In previous studies [2, 10], we have investigated the cubic perovskite structure and proposed 20 materials for the one-photon WS process and 12 others for the anode in a two-photon device with a Si cathode. Several of those were unknown in the WS community, but some of them have been already successfully tested. In particular the oxynitrides, like $BaTaO_2N$, $SrTaO_2N$, $CaTaO_2N$, $LaTiO_2N$ and $LaTaON_2$, gives good results in term of oxygen and/or hydrogen evolution in presence of sacrificial agents [11].

[4] Materials Project—A Materials Genome Approach (http://materialsproject.org/).
[5] Computational Materials Repository (documentation: https://wiki.fysik.dtu.dk/cmr/ and database: https://cmr.fysik.dtu.dk/).



(a) A_2BO_4 (b) $A_3B_2O_7$ (c) A_2BO_3N

Figure 1. Crystal structure of the Ruddlesden–Popper phase with $n = 1$ (a) and $n = 2$ (b). Panel (c) shows the $n = 1$ structure when one oxygen is replaced by nitrogen. (a) A_2BO_4, (b) $A_3B_2O_7$ and (c) A_2BO_3N.

The perovskite is a very versatile structure with materials exhibiting a large variety of properties and applications [12]. In addition to the standard cubic and low-symmetry perovskites, two cubic perovskites with general formula ABO_3 can be combined together in the so-called double perovskite [13]. The double perovskite structure was recently explored with the focus of finding new materials for WS and engineering of the bandgap by combining two perovskites with different electronic properties [14, 15].

In the present work, we investigate the layered perovskite structure. A layered perovskite is composed of two-dimensional (2D) slabs of ABO_3 cubic perovskite separated by a motif of metal atoms. There are several phases of layered perovskites which differ in the thickness and the relative displacement of the cubic perovskite slabs and in the motifs. The phase studied here is the Ruddlesden–Popper with general formula $A_{n+1}B_nO_{3n+1}$, where $n = 1, \ldots, \infty$ is the number of BO_6 octahedra forming the 2D slabs and the upper limit for n correspond to the cubic phase. Figure 1 shows the structures under investigation: we consider the cases of $n = 1$ and $n = 2$ (A_2BO_4 and $A_3B_2O_7$, respectively) and one oxygen replacement in favor of nitrogen in the case of $n = 1$ (A_2BO_3N).[6] There are several distinct oxygen sites that can be replaced by nitrogen. To avoid large distortions in the octahedron, we replace the oxygen between the A- and B-atoms leaving the xy-plane of the octahedron unchanged. This unit cell does not allow

[6] The phases with $n = 1$ and 2 permit us to investigate the layered perovskite structures at a reasonable computational cost. The $n = 3$ structure is described by a 34 atoms unit cell, which is too demanding for a screening project.

to investigate possible symmetry lowering in the structure (i.e. octahedra tilting and Jahn–Teller distortions). In order to do that a larger unit cell is required, which would increase dramatically computational time, making the screening unaffordable. In any case we should remark that the use of a cubic symmetry is well justified since this symmetry lowering usually disappears at high temperatures. Indeed in many cases perovskites recover a cubic-like symmetry at room temperature [16, 17].

2. Method

In the one-photon WS device, one material is responsible for the evolution of both oxygen and hydrogen. First a photon is harvested creating an electron–hole pair. Next, the electron and hole reach two different points on the surface where the evolution of O_2 and H_2 takes place. Any material that should function as photo-electrode for WS should meet a number of criteria: (i) chemical and structural stability; (ii) a bandgap in the visible range; (iii) well positioned band edges with respect to the red-ox levels of water; and (iv) high mobility for electrons and holes. In addition, low cost and non-toxicity are required. Here we focus on the criteria (i)–(iii).

The stability of a material is evaluated with respect to around 2000 reference systems. The reference systems constitute stable phases of the possible materials into which the considered layered perovskite can be decomposed, and are taken from the experimental ICSD database[7] and the Materials Project database (see footnote 4). A material is considered stable only if the energy difference between the layered perovskite phase and the most stable alternative combination of reference systems is below 0.2 eV atom^{-1}. We use this finite threshold energy to avoid the exclusion of potential candidates due to inaccuracies in the calculations [18] and to take into account metastability of the investigated structure. Each combination is fully relaxed using the revised Perdew–Burke–Ernzerhof (RPBE) functional [19] implemented in the density functional theory (DFT)-GPAW code [20, 21].

As an example, the stability of the Ba_2TaO_3N compound is given by

$$\Delta E = \mathrm{Ba_2\,TaO_3N} - \min_{c_i}\Big\{c_1\mathrm{Ba_2} + c_2\,\mathrm{Ta_2} + c_3\mathrm{BaO} + c_4\mathrm{BaO_2} + c_5\mathrm{Ta_2O_5} + c_6\mathrm{BaN_6} + c_7\,\mathrm{Ta_3N_5} + c_8\mathrm{Ba_2\,TaN_3} + c_9\mathrm{BaTaO_2N} + c_{10}\mathrm{N_2} + c_{11}\mathrm{O}\Big\}, \quad (1)$$

where c_i are positive and sum up to assure the correct stoichiometry of the A_2BO_3N layered structure and the chemical formulae indicate the DFT total energies of the references. All the references are in their solid state phase except for N_2 and O which are in the gas phase. Note that the energy of oxygen is calculated from H_2 and H_2O due to the well known problems associated with the DFT description of the O_2 triplet ground state.

We note that the present stability analysis does not include corrosion of the materials. Extended stability analysis including this effect were recently found to be of some importance [22, 23]. However, we leave this for a future study.

It is well known that the Kohn–Sham eigenvalues from DFT systematically underestimate bandgaps of extended semiconductors due to the approximate nature of the exchange-correlation functionals and the missing derivative discontinuity [24]. On the other hand, many-body methods like the GW approximation give better bandgaps, but are computationally too expensive to be used in a screening project of several hundred materials. Here, we used the Gritsenko, van Leeuwen, van Lenthe and Baerends potential (GLLB) improved for solids (-SC)

[7] ICSDWeb (www.fiz-karlsruhe.de/icsd_web.html).

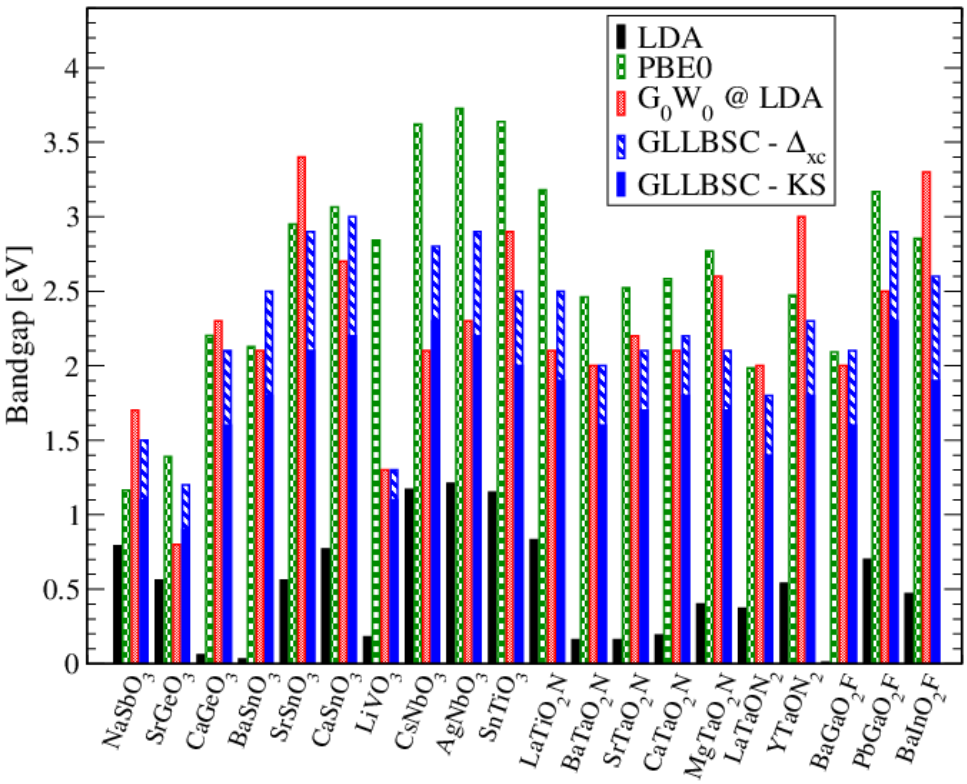


Figure 2. The bandgaps of the 20 identified candidates for one-photon WS calculated with the GLLB-SC potential (in blue) G_0W_0@LDA (in red), LDA (in black) and PBE0 (in green). The GLLB-SC bandgaps are obtained adding the derivative discontinuity (Δ_{xc}, in dashed blue) to the Kohn–Sham bandgap (KS, in solid blue). The discontinuities are usually proportional to the KS bandgaps. The agreement between the GLLB-SC potential and G_0W_0 is very good.

potential [25, 26], which includes an explicit estimation of the derivative discontinuity. In a previous work [10], the GLLB-SC bandgaps of around 40 metal oxides were compared to experiments. On average the GLLB-SC bandgaps were found to lie within 0.5 eV of the experimental gaps.

Figure 2 shows the bandgaps of the 20 earlier identified cubic perovskites for one-photon WS [2] here calculated using different methods and xc-functionals. Standard DFT functionals, like local density approximation (LDA) and PBE (not shown in the figure), seriously underestimate the bandgap, while the hybrid PBE0 tends to overestimate the gaps. We have performed G_0W_0 calculations using a recent implementation in the GPAW code [27]. In these calculations the G_0W_0 self-energy is evaluated using LDA energies and wave functions and the frequency dependence of the dielectric function is fitted to a plasmon pole model. We use a 150 eV plane wave cut off for the representation of the dielectric matrix and include empty bands up to the same energy cut off. A $7 \times 7 \times 7$ k-point mesh is used for both DFT and G_0W_0 calculations[8]. The GLLB-SC and G_0W_0 bandgaps are quite similar with a mean

[8] Convergence with respect to k points, plane wave cutoff and bands as well as validity of the plasmon pole approximation have been carefully tested for one of the structures. With the chosen parameters, we expect GW bandgaps to be converged within around 0.1 eV for all materials.

absolute difference of 0.3 eV. This is consistent with the results of a similar comparison made for a set of ten well characterized simple semiconductors and insulators [27]. Based on this we conclude that the bandgaps obtained with the GLLB-SC are sufficiently accurate for the purpose of materials screening.

In addition to the bandgap, the position of the conduction and valence band edges with respect to the water red-ox levels is crucial: the energy of electrons at the surface should be above the hydrogen evolution potential and the energy of holes must be below the oxygen evolution level. It is not trivial to calculate the band edge positions at a semiconductor–water interface from first principles. Although some methods have been proposed [28, 29], they are computationally rather demanding and not suited for screening studies. Instead we obtain the position of the band edges simply using an empirical equation [30, 31] that gives the center of the bandgap in terms of the geometrical average of the electronegativities in the Mulliken scale of the pure chemical elements, χ_M, forming the compound. The band edge positions are then calculated by adding and subtracting half of the bandgap, E_{gap}. In formula, for the A_2BO_4 layered perovskite:

$$E_{VB,\,CB} = E_0 + (\chi_A^2 \chi_B \chi_O^4)^{1/7} \pm E_{gap}/2\,, \tag{2}$$

where E_0 is the difference between the normal hydrogen electrode (NHE) and the vacuum ($E_0 = -4.5$ eV). More details about the method and its validation are available in [2, 10].

It has been recently shown for the cubic perovskite structure [32], that the search for new materials can be guided using chemical-based rules. The three chemical rules used here are:

- *Valence balance rule*: in a realistic material, the sum of the possible oxidation states of its elements must be zero.
- *Even–odd electrons rule*: the number of electrons per unit cell should be an even number. A material with an odd number of electrons has partially occupied bands at the Fermi level and is thus a metal. There are some exceptions to this rule, but they correspond to magnetic and/or strongly correlated materials which require more advanced theoretical descriptions than the one used here.
- *Ionic radii of the atoms*: the metals that can occupy the A- and B-ion position in the perovskite depend on the ionic radii of the neutral atoms. Usually alkali and alkaline earth metals occupy the A-ion side, and post-transition metals the B-ion position [12]. We implement this rule by having separate allowed chemical elements for the A- and B-sites[9].

The use of these rules drastically reduce the number of calculations from around 8000 to 300 different materials.

The screening parameters for the one-photon WS device are summarized in table 1. In addition to stability and a bandgap in the visible range, the band edges should straddle the red-ox levels of water plus the required overpotentials (0.1 eV for hydrogen and 0.4 eV for oxygen [33]).

The ideal efficiency of the one-photon device can be up to 7% [1]. Higher efficiencies can be achieved using a two-photon, or tandem, device, consisting of two semiconductors forming a pn-junction. Electrons generated in the cathode move to the surface and evolve hydrogen

[9] The metals used in the A-ion position are: Li, Na, Mg, K, Ca, Ga, Ge, Rb, Sr, Y, Ag, Cd, In, Sn, Cs, Ba, La, Tl and Pb. The B-metals are Al, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, Zr, Nb, Mo, Ru, Rh, Pd, In, Sn, Sb, Te, Hf, Ta, W, Re, Os, Ir, Pt and Bi.

Table 1. Screening parameters (in eV) used for the one- and two-photon WS devices. The red-ox levels of water with respect to the NHE are 0 and 1.23 eV for the hydrogen and oxygen evolution, respectively. The criterion for the position of the band edges include also the overpotentials for hydrogen (0.1 eV) and for oxygen (0.4 eV) [33]. Silicon is used as cathode in the two-photon WS device ($VB^{Si}_{edge} = 0.86$ eV).

Criterion	One-photon WS	Two-photon WS
Stability (ΔE)	0.2 eV atom^{-1}	0.2 eV atoms^{-1}
Bandgap (E_{gap})	$1.7 \leqslant E_{gap} \leqslant 3$	$1.3 \leqslant E_{gap} \leqslant 3$
Band edges	$VB_{edge} > 1.6$	$VB^{anode}_{edge} > 1.6$
(VB_{edge}, CB_{edge})	$CB_{edge} < -0.1$	$CB^{cathode}_{edge} < -0.1$
		$CB^{anode}_{edge} < VB^{cathode}_{edge}$

while the holes move toward the pn-junction and recombine with electrons from the anode. Similarly, the holes generated in the anode move toward the surface and evolve oxygen. Thus it takes two photons to generate one H_2. The efficiency of the device depends strongly on the relative sizes of the bandgaps of the two materials. Assuming ideal band edge line up, the best performance (around 25%) is obtained by combining two materials with bandgaps of 1.1 eV and 1.7, respectively [2]. The sum of the two bandgaps should not be smaller than 2.8 eV: 1.23 eV is the energy required per electron to split water and approximately 0.5 eV is required to account for the overpotentials of the oxidation and reduction reactions. In addition, the relevant quasi-Fermi levels are located ≈ 0.25 eV below (above) the conduction (valence) bands[10]. Finally, the quasi-Fermi level corresponding to the conduction band of the cathode should be at least 0.1 eV above the quasi-Fermi level corresponding to the valence band of the anode to ensure efficient charge recombination at the interface. As it turns out, silicon is an ideal photo-cathode for WS since it has a bandgap of 1.1 eV well positioned with respect to the reduction potential of hydrogen and because of the mature fabrication technology. A more detailed explanation of the two-photon device is available in [2]. In the case of two-photon devices we therefore limit our search to photo-anode materials assuming a silicon cathode and use the screening criteria summarized in table 1.

3. Trends in stability and bandgaps

The simplest layered perovskite is the Ruddlesden–Popper phase. The two phases studied here have the general formula A_2BO_4 and $A_3B_2O_7$. The former is composed of slabs of cubic perovskite ABO_3 offset by a translation of $(1/2, 1/2)$ with respect to each other, terminated by oxygens to complete the octahedra, and separated by a layer of A metal atoms. In the latter structure each slab of perovskite is formed by two octahedra BO_6 units. The unit cells

[10] The quasi-Fermi level describes the population electrons and holes in a semiconductor when their populations are not in equilibrium, for example when the semiconductor is under illumination. When an electron–hole pair is created, the density of both electrons and holes are above their equilibrium values and the populations of the carriers cannot be described by a single Fermi level. One of the consequences of the quasi-Fermi levels is that the effective energy of the electrons and holes does not correspond anymore to the conduction and valence band but the quasi-Fermi levels have to be considered.

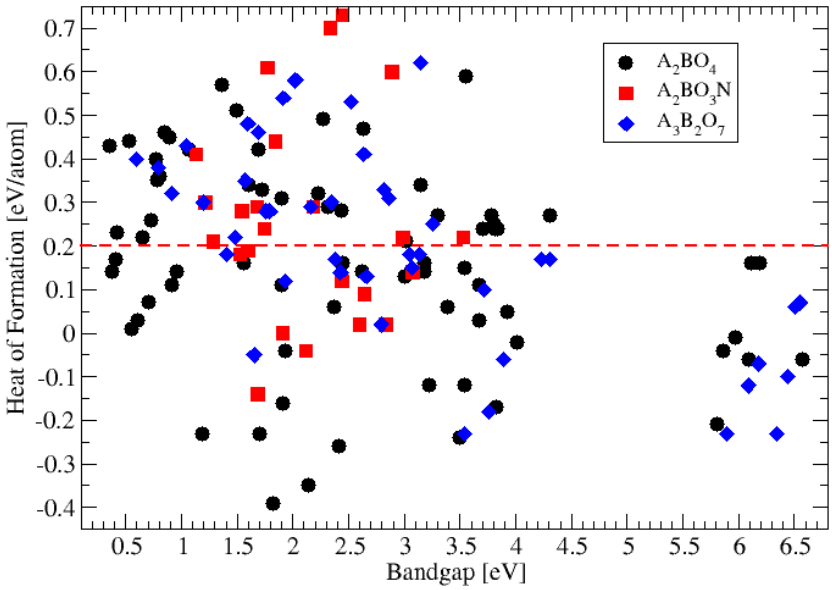


Figure 3. Calculated heat of formation plotted versus bandgap for the A_2BO_4 (black circles), A_2BO_3N (red squares) and $A_3B_2O_7$ (blue diamonds). The oxides show high stability and a wide range of bandgaps. The oxynitrides are overall less stable and have smaller bandgaps than the oxides.

of the two phases contain 14 and 24 atoms, respectively. The A-ion position is generally occupied by an alkali or alkali-earth metal, and the B-ion by a transition or post-transition metal. Materials of the form A_2BO_3N, in which an oxygen atom has been replaced by nitrogen, have been considered. In general, oxides tend to have valence band edges that are somewhat deep compared to the oxidation potential of water. Nitrogen is slightly less electronegative compared to oxygen and the valence bands of the oxynitrides are consequently shifted up compared to the oxides [34]. The smaller electronegativity of nitrogen leads also to a reduction in the size of the bandgaps. In addition, a nitrogen replacement breaks the symmetry of the cubic perovskite and creates an excess of charge in one of the corners of the octahedron. This leads to a reduction in the stability of the obtained materials, as shown in figure 3.

Figure 3 shows the calculated heat of formation for the A_2BO_4, A_2BO_3N and $A_3B_2O_7$ perovskites plotted against the GLLB-SC bandgap. The bandgaps of both types of oxides range from 0 to above 6 eV. The oxides are in general highly stable especially the materials with larger bandgaps. As expected, the oxynitrides are less stable and show smaller gaps. Despite of this, around 8 oxynitrides are found to fulfill the criteria on bandgap and stability for the one-photon WS device. A further nitrogen replacement is likely to lower the bandgaps and reduce the stability even further [2], and therefore has not been investigated here.

All the bandgaps of the stable A_2BO_4 layered perovskites are plotted in figure 4. There is some degree of correlation between the chemical elements and the size of the bandgaps. First of all it can be seen that the gap is mainly determined by the B-ion while the A-ion has less influence. Hf and Zr in the B-ion position generate large bandgap insulators with gaps above 6 eV. Ti, Ge and Sn lead to structures with gaps above 3 eV while W and Zn produce structures with gaps around 2 eV.

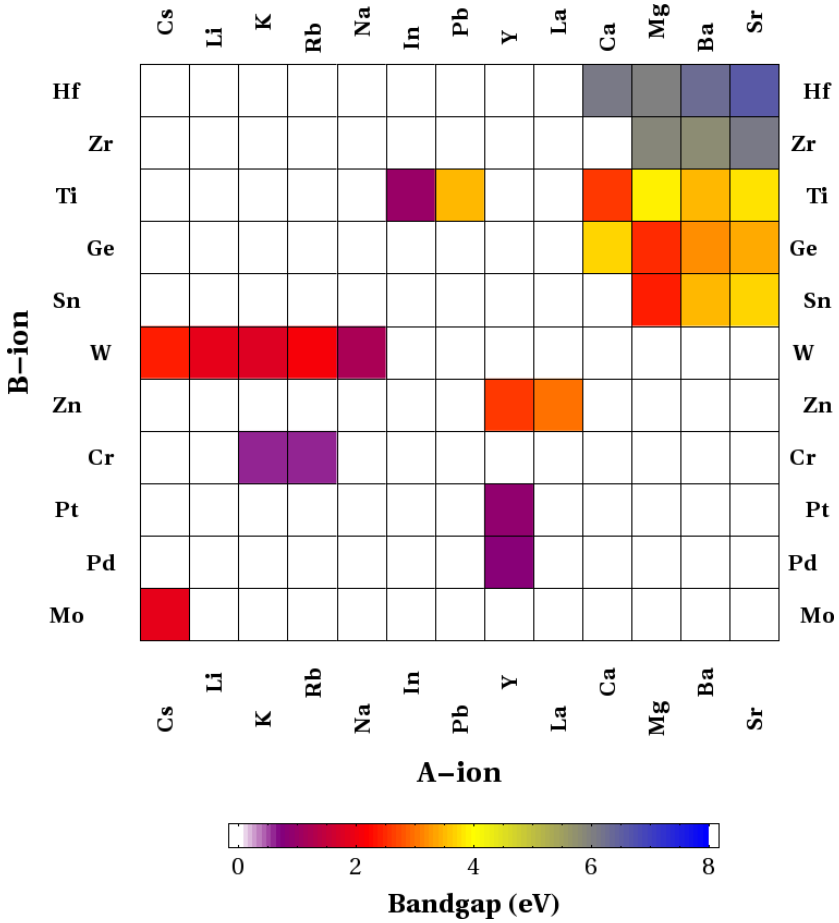


Figure 4. Bandgaps of the stable compounds in the A_2BO_4 structure. The gaps range from 0 to more than 6 eV. The chemical elements are sorted for similarity so that atoms that generate layered perovskites with similar bandgaps are close together.

Most of the oxides investigated have valence bands formed by the 2p levels of the O^{2-} ions. Since the nature of the bonds in these materials is highly ionic, this implies that the top of the valence band is roughly the same for all the oxides[11]. The conduction bands are mainly composed of the lowest unoccupied molecular orbitals (LUMO) of the A and B metal ions. This means that the size of the bandgap should correlate with the lowest lying LUMO level of the A and B cations. The relevant LUMO level is, however, not that of the cations in vacuum, but rather a cation sitting in the electrostatic field from the rest of the lattice, V_R. To analyze this, we follow an Ewald–Evjen scheme, assuming that all the ions keep their nominal charges. This scheme is described elsewhere in the literature [35].

Table 2 shows how the LUMO energies change drastically passing from the free cations to the cations under the effect of V_R. V_R destabilizes much more the energies of the B-site ions than

[11] There are some exceptions to these rule, e.g. those compounds that contain carbon group ions with valence 2+ (Ge^{2+}, Sn^{2+} and Pb^{2+}), in which the valence band is a mix between the 2p levels of the O^{2-} ions and the s levels of the carbon group ion.

Table 2. LUMO energies (in eV) of several B^{4+} ions in vacuum and in the presence of the crystal electrostatic field V_R, respectively. V_R is calculated for $ASnO_4$ and $CaBO_4$ for the A^{2+} and B^{4+} ions, respectively, (similar results are obtained for other lattices). The LUMO energies of the A^{2+} ions in vacuum and under the effect of V_R are for Ca^{2+} (s): −14.65 and −0.58; for Sr^{2+} (s): −13.58 and −1.02; and for Cd^{2+} (s): −20.35 and −4.91. The bandgaps of the compounds containing these ions are also shown for comparison.

			Bandgap		
B-ion	In vacuum	Including V_R	Ca_2BO_4	Sr_2BO_4	Cd_2BO_4
Ti^{4+} (d)	−56.13	−3.35	4.01	3.82	0.78
Zr^{4+} (d)	−42.30	−1.24	6.19	6.09	1.90
Hf^{4+} (d)	−39.90	−1.10	6.11	6.57	1.60
Ge^{4+} (s)	−52.28	−3.64	3.67	3.39	0.36
Sn^{4+} (s)	−45.45	−3.82	3.70	3.67	0.54

the ones of the A-site ions. This is due to the short distance between the O^{2-} and A ions (usually in a range from 1.9 to 2.3 Å) in comparison with the B–O^{2-} distances (in a range from 2.5 to 3.0 Å). The table shows the correlation between the lowest LUMO energy from A and B (under the effect of V_R) ions and the bandgap for the particular case of $(A^{2+})_2B^{4+}(O^{2-})_4$ perovskites. The LUMO of the B^{4+} ions is more stable than the LUMO of the A^{2+} ions when the A position is occupied by an alkaline-earth ion (in the table only the results for Ca^{2+} and Sr^{2+} are shown, but similar features are observed for Mg^{2+} and Ba^{2+}). Thus, in these cases the bandgap is correlated with the LUMO of the B ion. This explains why Zr^{4+} and Hf^{4+} compounds have similar gaps, whereas Ti^{4+} shows a smaller gap. This could be understood even just by looking at the energies of the ions in vacuum, since the d-levels of Ti^{4+} are much deeper in energy that those of Zr^{4+} and Hf^{4+}. This difference is strongly reduced when V_R is considered due to the smaller ionic radius of Ti^{4+} (10 pm smaller than Zr^{4+} or Hf^{4+} ionic radii [36]), which has O^{2-} ions closer. In the case of the carbon group (Ge^{4+} and Sn^{4+}), their compounds have similar bandgaps due to a compensation of effects: Ge^{4+} s level in vacuum is lower in energy than Sn^{4+} s level (by ≈ 7 eV), but Ge^{4+} ionic radii is 16 pm smaller than that of Sn^{4+}.

When Cd^{2+} occupies the A site, its LUMO (the s level) is more stable than the LUMO of B^{4+} ions, contrary to the case of an alkaline-earth on A position. This leads to compounds with a conduction band dominated by the Cd^{2+} s level and consequently to a small bandgap. A similar investigation has been performed also for the $A_3B_2O_7$ layered perovskites with similar results. Since nitrogen has an electronegativity very close to the one of oxygen, we expect that these results are still valid for the oxynitrides. The main difference will be that the valence band is formed by an hybridization of the oxygen and nitrogen 2p orbitals.

A couple of A- and B-ions can be used for both the A_2BO_4 and the $A_3B_2O_7$ perovskites. For these cases, we have investigated the changes in the size of the bandgaps depending on the number of octahedra forming the 2D slab. There is a weak trend that correlates the bandgaps to the B-ions: when the B-ion position is occupied by a p-metal (e.g. Ge and Sn) the gaps are generally reduced when the slab thickness is increased, while for d-metals (e.g. Ti, Zr and Hf), the gaps seem to increase with slab thickness. In average, for the p-metals the bandgaps are reduced by 0.6 eV from the $n = 1$ to $n = 2$ structures and by 0.4 eV from the $n = 1$ to $n = \infty$

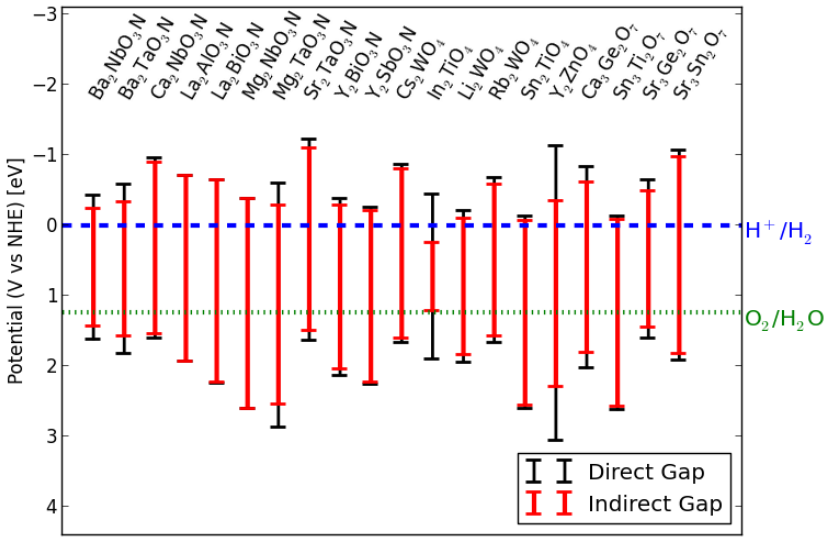


Figure 5. The identified candidates in the Ruddlesden–Popper phases for one-photon WS. The figure shows the red-ox levels of water and the calculated band edges for the indirect (red) and direct (black) bandgaps.

structures. For the d-metals the bandgaps are reduced by 0.03 and 0.2 eV, respectively. There are some exceptions to this trend, like $Sr_{n+1}Ti_nO_{3n+1}$. Lee *et al* [37] have recently shown that bandgaps of the $Sr_{n+1}Ti_nO_{3n+1}$ Ruddlesden–Popper phase is reduced with an increase of the thickness and we have seen the same effect. In general, these trends could be used to tune the bandgap by increasing or reducing the number of octahedra within the layers.

4. Candidates for water splitting

The screening criteria of table 1 have been applied to the calculated structures. Out of the 300 investigated materials 20 fulfill the criteria for one-photon WS, see figure 5. Ba_2TaO_3N and $Sr_3Sn_2O_7$ are known to exist experimentally in the layered perovskite structure. Ba_2TaO_3N has also recently been proposed as a WS material by Wu *et al* [38] who used a computational screening procedure similar to the one proposed here. Y_2BiO_3N and the other $A_3B_2O_7$ compounds are already known in other stoichiometries and most of the A_2BO_4 materials and Sr_2TaO_3N have been synthesized in other crystal structures with the same stoichiometry as the layered perovskite. It might not be an issue that some compounds are known in another crystal structure. In fact, layered perovskites can often be grown epitaxially even if they are not globally stable. Eight of the identified perovskites have not been investigated experimentally to our knowledge. In particular, the oxynitrides are interesting candidates for WS. In fact, the cubic perovskites ABO_2N with A = Ba, Sr and Ca, and B = Ta and Nb are known to evolve hydrogen and/or oxygen in the presence of a sacrificial agent [11, 39]. We expect that the layered perovskites containing these elements conserve the good properties in terms of activity already present in the cubic phase.

For the two-photon device with a silicon cathode, our screening identifies five layered perovskites as candidates for the anode material. All five are experimentally known in other structures/stoichiometry, but none of them has been used so far as photocatalyst.

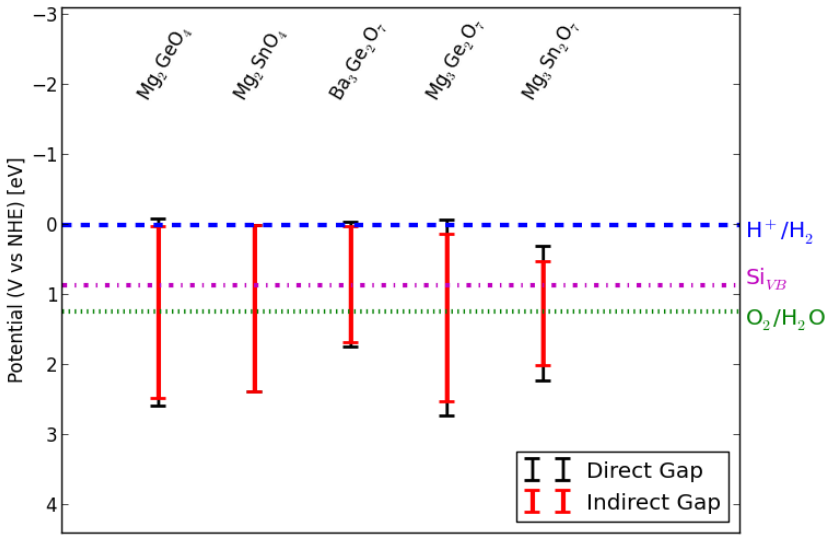


Figure 6. The identified candidates in the Ruddlesden–Popper phases for two-photon WS device. The figure shows the red-ox levels of water and the valence band of silicon. The calculated band edges for the indirect (red) and direct (black) bandgaps are drawn.

5. Conclusions

With the aim of identifying stable and abundantly available semiconductors for light harvesting photo-electrodes for WS, we have screened 300 oxides and oxynitrides in the layered perovskite structure with stoichiometry A_2BO_4, $A_3B_2O_7$ and A_2BO_3N. The stability and bandgaps were calculated using DFT and the band edge alignment relative to the water red-ox potentials was estimated using an empirical formula. The accuracy of the calculated bandgaps, which were obtained with the GLLB-SC potential, was validated by comparing with state of the art G_0W_0 calculations for 20 oxides in the cubic perovskite structure.

We have identified 20 candidate materials for one-photon WS, and additional 5 materials for photo-anode in a two-photon device with a silicon cathode. A few of these materials are already experimentally known, but none of them have been used so far for photoelectrocatalysis.

The trends in stability and bandgap have been studied in some detail. We have found that the bandgap of the layered perovskites is mainly determined by the position of the most stable LUMO of the A- and B-ion. Furthermore, our results indicate that the bandgap can be tuned to some extent by varying the number of octahedra forming the layered structure. Specifically, for a d-metal (p-metal) at the B-ion position the effect is to increase (decrease) the gap with the number of octahedra within a layer.

Acknowledgments

The authors acknowledge support from the Catalysis for Sustainable Energy (CASE) initiative funded by the Danish Ministry of Science, Technology and Innovation and from the Center on Nanostructuring for the Efficient Energy Conversion (CNEEC) at Stanford University, an Energy Frontier Research Center founded by the US Department of Energy, Office of Science,

Office of Basic Energy Sciences under award number DE-SC0001060. JMGL acknowledges support from the Spanish Ministry of Economy and Competitiveness under project numbers FIS2009-07083, FIS2010-21282-C02-01 and FIS2012-30996 and through Ramon y Cajal grant number RYC-2011-07782.

References

[1] Weber M R and Dignam M J 1986 Splitting water with semiconducting photoelectrodes-efficiency considerations *Int. J. Hydrog. Energy* **11** 225–32
[2] Castelli I E, Landis D D, Thygesen K S, Dahl S, Chorkendorff Ib, Jaramillo T F and Jacobsen K W 2012 New cubic perovskites for one- and two-photon water splitting using the computational materials repository *Energy Environ. Sci.* **5** 9034–43
[3] Fujishima A and Honda K 1972 Electrochemical photolysis of water at a semiconductor electrode. *Nature* **238** 37–8
[4] Kudo A and Miseki Y 2009 Heterogeneous photocatalyst materials for water splitting. *Chem. Soc. Rev.* **38** 253–78
[5] Curtarolo S *et al* 2012 Aflow: an automatic framework for high-throughput materials discovery *Comput. Mater. Sci.* **58** 218–26
[6] Lin *et al* 2012 *In silico* screening of carbon-capture materials *Nature Mater.* **11** 633–41
[7] Ceder G, Chiang Y -M, Sadoway D R, Aydinol M K, Jang Y -I and Huang B 1998 Identification of cathode materials for lithium batteries guided by first-principles calculations *Nature* **392** 694
[8] Hachmann J, Olivares-Amaya R, Atahan-Evrenk S, Amador-Bedolla C, Sanchez-Carrera R S, Gold-Parker A, Vogt L, Brockway A M and Aspuru-Guzik A 2011 The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid *J. Phys. Chem. Lett.* **2** 2241–51
[9] d'Avezac M, Luo J -W, Chanier T and Zunger A 2012 Genetic-algorithm discovery of a direct-gap and optically allowed superstructure from indirect-gap Si and Ge semiconductors *Phys. Rev. Lett.* **108** 027401
[10] Castelli I E, Olsen T, Datta S, Landis D D, Dahl S, Thygesen K S and Jacobsen K W 2012 Computational screening of perovskite metal oxides for optimal solar light capture *Energy Environ. Sci.* **5** 5814
[11] Yamasita D, Takata T, Hara M, Kondo J N and Domen K 2004 Recent progress of visible-light-driven heterogeneous photocatalysts for overall water splitting *Solid State Ion.* **172** 591–5
[12] Ishihara T 2009 *Perovskite Oxide for Solid Oxide Fuel Cells* (Berlin: Springer)
[13] Kobayashi K-I, Kimura T, Sawada H, Terakura K and Tokura Y 1998 Room-temperature magnetoresistance in an oxide material with an ordered double-perovskite structure *Nature* **395** 677–80
[14] Berger R F and Neaton J B 2012 Computational design of low-band-gap double perovskites *Phys. Rev.* B **86**
[15] Castelli I E, Thygesen K S and Jacobsen K W 2013 Bandgap engineering of double perovskites for one- and two-photon water splitting *MRS Online Proc. Libra.* **1523**
[16] Megaw H D 1945 Crystal structure of barium titanate *Nature* **155** 484–5
[17] Møller C K 1957 A phase transition in cæsium plumbochloride *Nature* **180** 981–2
[18] Honkala K, Hellman A, Remediakis I N, Logadottir A, Carlsson A, Dahl S, Christensen C H and Nørskov J K 2005 Ammonia synthesis from first-principles calculations *Science* **307** 555–8
[19] Hammer B, Hansen L B and Nørskov J K 1999 Improved adsorption energetics within density-functional theory using revised Perdew–Burke–Ernzerhof functionals *Phys. Rev.* B **59** 7413–21
[20] Mortensen J J, Hansen L B and Jacobsen K W 2005 Real-space grid implementation of the projector augmented wave method *Phys. Rev.* B **71** 35109
[21] Enkovaara J *et al* 2010 Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method *J. Phys.: Condens. Matter* **22** 253202
[22] Persson K A, Waldwick B, Lazic P and Ceder G 2012 Prediction of solid-aqueous equilibria: scheme to combine first-principles calculations of solids with experimental aqueous states *Phys. Rev.* B **85** 235438

[23] Castelli I E, Thygesen K S and Jacobsen K W 2013 Calculated Pourbaix diagrams of cubic perovskites for water splitting: a critical stability analysis *Top. Catalysis* accepted
[24] Godby R W, Schlüter M and Sham L J 1986 Accurate exchange-correlation potential for silicon and its discontinuity on addition of an electron *Phys. Rev. Lett.* **56** 2415–8
[25] Gritsenko O, van Leeuwen R, van Lenthe E and Jan Baerends E 1995 Self-consistent approximation to the Kohn–Sham exchange potential *Phys. Rev.* A **51** 1944
[26] Kuisma M, Ojanen J, Enkovaara J and Rantala T T 2010 Kohn–Sham potential with discontinuity for band gap materials *Phys. Rev.* B **82** 115106
[27] Hüser F, Olsen T and Thygesen K S 2013 Quasiparticle GW calculations for solids, molecules and two-dimensional materials *Phys. Rev.* B **87** 235132
[28] Wu Y, Chan M K Y and Ceder G 2011 Prediction of semiconductor band edge positions in aqueous environments from first principles *Phys. Rev.* B **83** 235301
[29] Georg Moses P and Van deWalle C G 2010 Band bowing and band alignment in ingan alloys *Appl. Phys. Lett.* **96** 021908
[30] Butler M A and Ginley D S 1978 Prediction of flatband potentials at semiconductor–electrolyte interfaces from atomic electronegativities *J. Electrochem. Soc.* **125** 228–32
[31] Xu Y and Schoonen M A A 2000 The absolute energy positions of conduction and valence bands of selected semiconducting minerals *Am. Mineral.* **85** 543–56
[32] Jain A, Castelli I E, Hautier G, Bailey D H and Jacobsen K W 2013 Performance of genetic algorithms in search for water splitting perovskites *J. Mater. Sci.* **48** 6519–34
[33] Trasatti S 1990 *Croat. Chem. Acta* **63** 313–29
[34] Aguiar R, Logvinovich D, Weidenkaff A, Rachel A, Reller A and Ebbinghaus S G 2008 The vast colour spectrum of ternary metal oxynitride pigments *Dyes Pigments* **76** 70–5
[35] Piken A G and van Gool W 1968 *Ford Tech. Report* SL 68-10
[36] Shannon R D 1976 Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides *Acta Crystallogr.* A **32** 751–67
[37] Lee C-H *et al* 2013 Effect of reduced dimensionality on the optical band gap of $SrTiO_3$ *Appl. Phys. Lett.* **102** 122901
[38] Wu Y, Lazic P, Hautier G, Persson K and Ceder G 2013 First principles high throughput screening of oxynitrides for water-splitting photocatalysts *Energy Environ. Sci.* **6** 157–68
[39] Siritanaratkul B, Maeda K, Hisatomi T and Domen K 2011 Synthesis and photocatalytic activity of perovskite niobium oxynitrides with wide visible-light absorption bands *ChemSusChem* **4** 74–8

Paper III

"On the convergence of many-body excited state calculations for monolayer MoS_2"
Falco Hüser, Thomas Olsen, and Kristian S. Thygesen
accepted for Physical Review B.

On the convergence of many-body excited state calculations for monolayer MoS_2

Falco Hüser* and Thomas Olsen
Center for Atomic-scale Materials Design (CAMD), Department of Physics
Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

Kristian S. Thygesen†
Center for Atomic-scale Materials Design (CAMD), Department of Physics
Technical University of Denmark, 2800 Kgs. Lyngby, Denmark and
Center for Nanostructured Graphene (CNG)
Technical University of Denmark,
2800 Kgs. Lyngby, Denmark
(Dated: October 29, 2013)

We present first principles many-body calculations of the dielectric constant, quasiparticle band structure and optical absorption spectrum of monolayer MoS_2. As the separation between the repeated layers is increased, the dielectric function of the layer develops a strong q-dependence around $q = 0$. This implies that denser k-point grids are required to converge the band gap and exciton binding energies when large supercells are used. In the limit of inifinite layer separation, here obtained using a truncated Coulomb interaction, a 45×45 k-point grid is needed to converge the G_0W_0 band gap and exciton energy to within 0.1 eV. We provide an extensive comparison with previous studies and explain agreement and variations in the results. It is demonstrated that too coarse k-point sampling and the interactions between the repeated layers have opposite effects on the band gap and exciton energy, leading to a (partial) fortuitous error cancellation in the previously published results.

PACS numbers: 71.20.Nr, 71.35.-y, 73.22.-f, 78.20.Bh, 78.60.Lc
Keywords: MoS2

I. INTRODUCTION

Atomically thin two-dimensional (2D) materials such as graphene, hexagonal boron nitride, and transition metal dichalcogenides (TMDC) possess unique electronic and optical properties including high intrinsic carrier mobilities,[1–3] tunable band gaps,[4,5] and strong light-matter interactions.[6–9] These features, combined with the possibility of engineering their electronic properties further via strain, alloying or stacking, make the 2D materials ideal as building blocks for new opto-electronic structures and devices with minimal sizes and performances surpassing present technologies.

After the intense focus on graphene, the TMDCs are now attracting increasing interest.[10] This stems mainly from the greater variation in their electronic properties including both semiconducting and metallic behavior. So far, the most intensively studied single-layer TMDC is the semiconductor MoS_2. Nanostructured forms of MoS_2 have previously been explored as potential catalysts for desulferization of crude oil and more recently for (photo)-electrochemical hydrogen evolution.[11–13] Bulk MoS_2 is composed of two-dimensional sheets held together by weak van der Waals forces and individual sheets can be isolated by exfoliation techniques similar to those used to produce graphene.[1] Single layers of MoS_2 therefore comprise highly interesting two-dimensional systems with a finite band gap and have recently been proposed for nano-electronics applications.[2]

The optical properties of MoS_2 have been thoroughly studied experimentally.[14–19] The absorption spectrum shows two distinct low energy peaks at 1.88 eV and 2.06 eV, which are denoted by A and B, respectively,[20] and derive from direct transitions between a split valence band and the conduction band at the K point of the Brillouin zone. Their Rydberg satellites, Zeeman splitting, and dependence on crystal thickness have been investigated in detail.[17] Recently, the quantum yield of luminescence from MoS_2 was shown to increase dramatically when the sample thickness was changed from a few layers to a monolayer[7,8] indicating a transition to a direct band gap in the single layer.

In the past couple of years a number of theoretical studies of the electronic band structure and optical excitations in monolayer MoS_2 have been published.[4,21–26] These studies are based on many-body perturbation theory in the GW approximation (mainly the non-selfconsistent G_0W_0 approach) for the band structure and the Bethe-Salpeter equation (BSE) with a statically screened electron-hole interaction for the optical excitations. As is standard practice, the calculations have been performed on a supercell geometry where the MoS_2 layers have been separated by $10 - 20$ Å vacuum and the Brillouin Zone (BZ) sampled on grids ranging from 6×6 to 15×15. With these parameters G_0W_0 band gaps in the range $2.6 - 3.0$ eV, and G_0W_0-BSE exciton binding energies of $0.6 - 1.1$ eV, have been reported. Moreover, both direct[21–25] and indirect[4] band gaps have been found at the G_0W_0 level, while only direct gaps have been obtained with self-consistent GW[21] and GW_0.[4,26] When comparing these values, it should be kept in mind that both size and nature of the band gap of MoS_2 depends

sensitively on the in-plane lattice parameter, a.[4]

One of the most fundamental quantities describing the electronic structure of a material is the dielectric function. The dielectric properties of atomically thin 2D materials are quite different from their 3D counterparts.[27] For example plasmons in 2D metals have acoustic dispersion relations ($\omega_p(q) \to 0$ as $q \to 0$), and screening is generally much weaker leading to strong exciton binding energies in 2D semiconductors. Reported static dielectric constants for monolayer MoS_2 obtained using the supercell approach lie in the range $4.2 - 7.6$ (for in-plane polarization).[21,24,28] These values have been used to rationalize the exciton binding energy in MoS_2 using the simple Mott-Wannier model.

In this paper, we present an in-depth study of the dielectric function, quasiparticle band structure and excitonic states in monolayer MoS_2. We focus on separating the spurious interlayer screening from the intrinsic intralayer screening in supercell geometries, and the consequences that the physics of screening in 2D has for the convergence of many-body excited state calculations. The 3D macroscopic dielectric constant, as used for solids, converges to 1 for all q vectors in the limit of infinite separation of the layers and is thus meaningless for a 2D material. We use an alternative approach to calculate the dielectric constant by averaging the total field over the material rather than the supercell. This 2D dielectric constant shows strong q-dependence for small wave vectors and becomes exactly 1 for $q = 0$. This property has important consequences for the k-point convergence of many-body calculations.

In general, the use of a truncated Coulomb interaction is essential to avoid interlayer screening which decays slowly with the layer separation, L. The interlayer screening yields too large dielectric constant for wave vectors $q < 1/L$. As a consequence, the G_0W_0 band gaps and exciton energies are 0.5 eV too low on average for layer separations of around 20 Å. For larger layer separations, the strong q-dependence of the dielectric constant for small q implies that a k-point grid of at least 45×45 is required to converge band gaps and exciton energies to 0.1 eV. For k-point grids below 15×15 the band gap is at least 0.5 eV too large in the limit $L \to \infty$. Thus the effect of interlayer screening and too coarse k-point grids partially cancel out leading to reasonable values for the band gap and exction binding energy with underconverged parameters as applied in previous studies.

The paper is organized as follows. In Sec. II we present G_0W_0 band structures and study the convergence of the gap with respect to interlayer separation and k-point sampling. In Sec. III we show calculations for the 2D dielectric constant and explain the origin of the slow k-point convergence of the band gap. In Sec. IV we present many-body calculations of the lowest excitons and analyse their convergence with layer separation and k-point sampling. Our conclusions are given in Sec. V.

II. QUASIPARTICLE BAND STRUCTURE

In this section we demonstrate that GW band structures for monolayer MoS_2 converge extremely slow with respect to the interlayer separation. In order to obtain well converged results (within 0.1 eV), the use of a truncated Coulomb interaction is inevitable, along with a k-point grid of around 45×45. Previously reported calculations with the full Coulomb interaction have employed only separation between 10 and 20 Å and used from 6×6 to 12×12 k points. The resulting band structures are, however, somewhat saved by a fortunate error cancellation between the two effects.

A. Computational details

All our calculations have been performed with the projector augmented wave method code GPAW.[29,43] The Kohn-Sham wave functions and energies of monolayer MoS_2 were calculated in the local density approximation (LDA) using a plane wave basis with cut-off energy 400 eV. The $4s$ and $4p$ semicore electrons of Mo were explicitly included in all calculations. Unless otherwise stated the calculations have been performed for the experimental lattice constant of 3.16 Å. One-shot G_0W_0 calculations were performed using the LDA wave functions and eigenvalues to obtain the G_0W_0@LDA quasiparticle energies. A plane wave cut off of 50 eV and 200 bands were used for the dielectric function, screened interaction and GW self-energy. Convergence with respect to these parameters has been checked very carefully. With these values band gaps were found to be converged within around 10 meV. The plasmon pole approximation for the dielectric function was found to yield QP energies within 0.1 eV of those obtained from full frequency dependence and was consequently used in all calculations. To avoid interaction between the periodically repeated MoS_2 sheets, we have applied a truncated Coulomb interaction of the form $v_c(\mathbf{r}) = (1/r)\theta(R_c - z)$, following Refs. 30 and 31. For details on the implementation of the GW method in the GPAW code we refer to Ref. 32. We note that we have used a numerical averaging of the head of the screened potential $W_{\mathbf{00}}(\mathbf{q})$ for all wavevectors $\mathbf{q}$ in the Brillouin zone (similar to Eq. 9 in Sec. III C). This was found to be crucial in all calculations with the full $1/r$ Coulomb interaction.

B. Results

The band structure calculated using 45×45 k points and the truncated Coulomb interaction is shown in Fig. 1. At the LDA level, we find a direct band gap at the K point of 1.77 eV while the smallest indirect gap of 1.83 eV occurs from Γ to a point along the Γ-K direction. In contrast, G_0W_0 predicts an indirect gap of 2.58 eV and at direct gap at K of 2.77 eV.

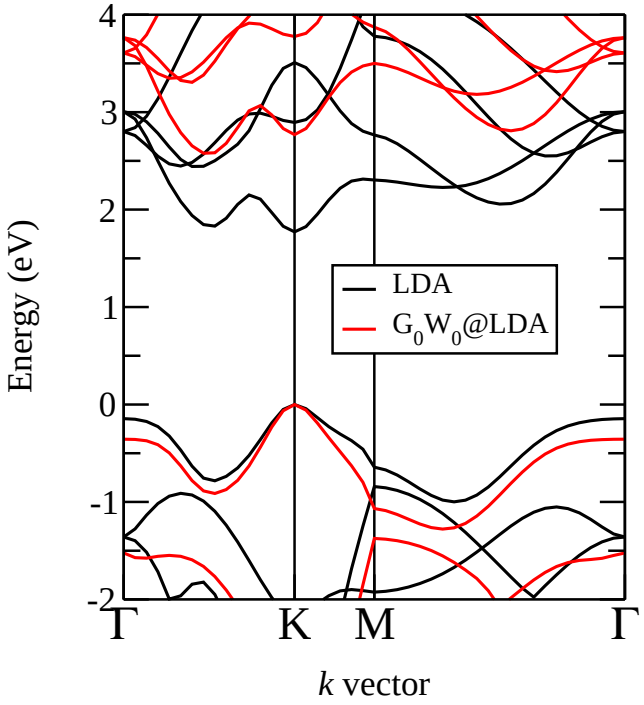


FIG. 1: (Color online). Bandstructure of monolayer MoS_2 calculated with LDA and G_0W_0@LDA using 45×45 k points and a truncated Coulomb interaction to avoid interaction between periodically repeated layers. The valence band tops have been aligned.

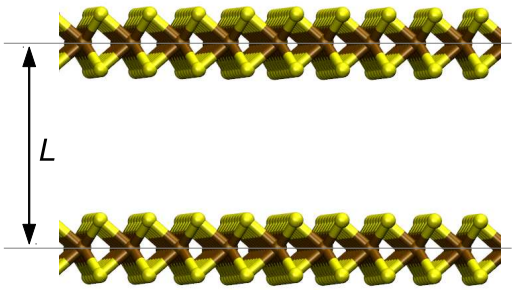


FIG. 2: Definition of the interlayer separation, L.

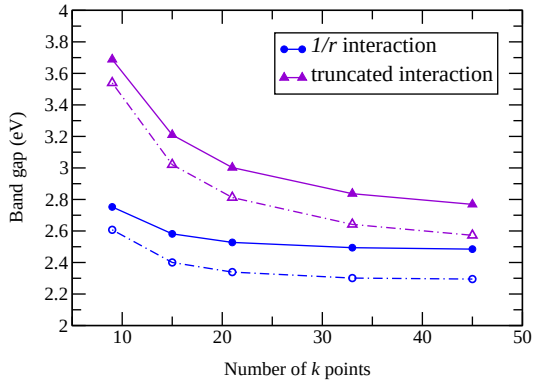


FIG. 3: (Color online). Direct (full symbols) and indirect (open symbols) G_0W_0 band gaps as function of the number of k points in one of the in-plane directions for a layer separation of $L = 23$ Å.

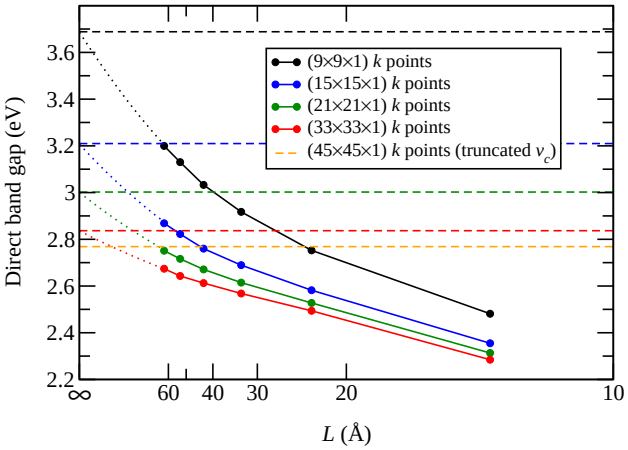


FIG. 4: (Color online). Direct G_0W_0 band gap plotted as a function of interlayer distance for different k-point samplings with the full $1/r$ interaction. Dotted lines serve as a guide for the eye to extrapolate for $L \to \infty$. They were obtained by fitting all values for $L > 30$ Å, including the results with the Coulomb truncation, to a quadratic function. Dashed horizontal lines indicate the calculated values with the truncated Coulomb interaction.

In Fig. 3 we show the convergence of both the direct and the indirect band gap with respect to the k-point grid for a fixed interlayer separation of 23 Å (see Fig. 2 for the definition of L). It is clear that a very dense k-point grid is needed in order to obtain well-converged results with the truncated Coulomb interaction. For 45×45 k points, band gaps are converged within less than 0.1 eV, while this is already the case for 15×15 k points with the bare Coulomb interaction. However, the calculated values are too low. The slow convergence with respect to k points when the truncation is used will be discussed in detail in Sec. III C.

We see that results do not converge independently with respect to the number of k points and the interlayer separation. In Fig. 4, we plot the L-dependence of the direct band gap for different k-point samplings with the bare interaction. The k-point dependence becomes much stronger for large L. For $L \to \infty$, the values are expected to converge to the results calculated with the truncation (indicated by dotted lines). They seem to exhibit a linear $1/L$ behaviour only for $L > 50$ Å or very dense k-point samplings. Fig. 5 shows all results and interpolated values in a contourplot as a function of $1/L$ and number of k points. The effects of using more k points and increasing L are of different sign and partially cancel each other. This is the reason, why different choices of the two parameters yield the same results. Especially, the

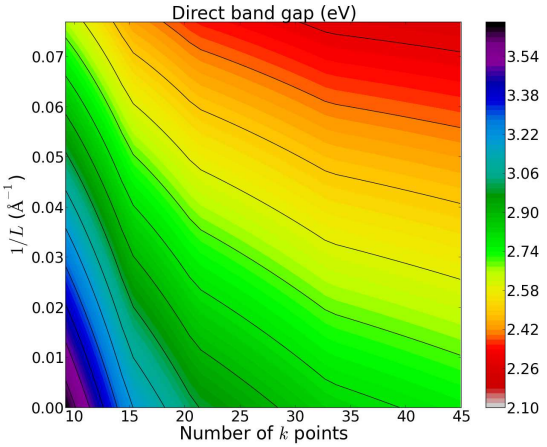


FIG. 5: (Color online). Contour plot of the direct G_0W_0 band gap as a function of the inverse interlayer distance and number of k points in one of the in-plane directions with the full $1/r$ interaction. Contour lines are separated by 0.1 eV. Interpolation from splines was used.

band gaps calculated with 9×9 k points and $L = 23$ Å and 15×15 k points and $L = 43$ Å are the same as with 45×45 k points and infinite L. This seems, however, conincidental and we do not expect it to be the case for other systems.

We note that all calculations have been performed with a single k point in the direction perpendicular to the layer. This is, however, insufficient for small interlayer distances. For $L = 13$ Å, we find an increase of the band gap of around 0.2 - 0.3 eV when at least 3 k points are used, for example. For $L > 20$ Å or use of the truncation, this effect is negligible.

C. Comparison with previous work

In table I we show our converged results obtained with the truncated Coulomb interaction and 45×45 k points together with previous G_0W_0 results from the literature. For each reference we show the values used for the lattice constant, the interlayer separation and the k-point sampling. It can be seen that all the previous calculations have used small layer separations and no truncation method. As pointed out in the preceding discussion, this gives a fast k-point convergence. A properly converged calculation, however, requires larger separations and thereby more k points. But as a consequence of a cancellation of errors, a calculation with 19 Å layer separation and 12×12 k points yields almost the same band gaps as our converged result (within 0.15 eV). We are thus led to conclude that the reasonable agreement between our results and previous ones are to a large extent fortuitous.

Furthermore, the effect of strain can have a large impact on the MoS_2 band gap. As demonstrated in Ref. 4, using 12×12 k points and 19 Å layer separation, the G_0W_0 band gap for the experimental lattice constant of 3.160 Å is indirect. With a lattice constant of 3.190 Å, corresponding to 1% strain, the gap changes to direct. The lowering of the direct band gap becomes even more pronounced for larger lattice constants. As can be seen from the table our converged results predict the same trend, in particular the decrease of the direct gap as function of strain, with our values for the direct gap being generally 0.2 eV larger. We note that for 3.255 Å, the smallest indirect transition occurs from the Γ point at the valence band to the K point at the conduction band. This is also in agreement with Ref. 4. In the partially self-consistent GW_0 calculations of Ref. 26, the opposite trend was found, namely a transition from a direct to indirect band gap for ~ 5 % strain. However, a layer separation of only 12 Å and less than 9 k points in the in-plane directions were used.

In Ref. 22, the band gap was determined by extrapolating from $L = 20$ Å to infinite layer separation, under the assumption that the gap scales linearly with the inverse distance between the layers. The obtained values for the direct and indirect band gaps are ~ 3.0 and ~ 3.3 eV, respectively. This is consistent with our findings using the truncated Coulomb interaction, the same lattice constant of 3.18 Å and the same (under-converged) k-point grid of 12×12 as in Ref. 22.

From our studies, we conclude that the G_0W_0@LDA band gap of monolayer MoS_2 is indirect with a value of 2.6 eV while the direct gap is 2.8 eV, when the experimental lattice constant of 3.16 Å is used. The question of how well the one-shot G_0W_0@LDA approximation describes the true electronic structure of this system remains open. Partially self-consistent GW_0[4] and fully self-consistent GW[21] calculations have been shown to consistently yield direct band gaps of $2.75 - 2.80$ eV.

III. STATIC SCREENING

In this section we present a detailed investigation of the (static) dielectric properties of monolayer MoS_2. This serves a dual purpose. First, it illustrates the origin of the slow convergence of the GW results presented in the previous section (and the BSE results presented in the next section). Secondly, it shows that the usual definition of the macroscopic dielectric constant of a periodic solid is not meaningful when applied to a 2D system represented in a periodic supercell. We discuss the difference between screening in 2D and 3D which becomes particularly pronounced in the $q \to 0$ limit with large consequences for the calculation of optical excitations with static screening of the electron-hole interaction (see next section).

TABLE I: Calculated G_0W_0 band gaps obtained in present work and compared with previous results from the literature. All our calculations have been performed using a truncated Coulomb interaction.

Ref.	starting point	a (Å)	number of k-points	layer separation (Å)	$E_{\rm gap}$ (eV) direct	$E_{\rm gap}$ (eV) indirect
This work	LDA	3.16	$45\times45\times1$	23 (truncated v_c)	2.77	2.58
This work	LDA	3.19	$45\times45\times1$	23 (truncated v_c)	2.65	2.57
This work	LDA	3.255	$45\times45\times1$	23 (truncated v_c)	2.41	2.51
Ref. 25	LDA	3.15	$18\times18\times1$	24	2.41	~ 2.40
Ref. 4	PBE	3.16	$12\times12\times1$	19	~ 2.60	2.49
Ref. 4	PBE	3.19	$12\times12\times1$	19	2.50	~ 2.55
Ref. 4	PBE	3.255	$12\times12\times1$	19	2.19	2.19
Ref. 21	LDA	3.16	$8\times8\times2$	19	2.96	–
Ref. 22	PBE	3.18	$12\times12\times1$	$20+1/L$ extrapolation	2.97	3.26
Ref. 22	PBE	3.18	$12\times12\times1$	20	~ 2.60	~ 2.85
Ref. 23	LDA	3.11	$12\times12\times1$	13	2.57	–
Ref. 24	HSE	3.18	$6\times6\times1$	15	2.82	~ 3.00
Ref. 33	PBE	3.19	$15\times15\times1$	15	2.66	–

A. 3D macroscopic dielectric constant

The microscopic dielectric function determines the relation between a weak external potential and the total potential in the material,

$$V_{\rm tot}(\mathbf{r}) = \int d\mathbf{r}'\, \epsilon^{-1}(\mathbf{r},\mathbf{r}')V_{\rm ext}(\mathbf{r}'). \quad (1)$$

For a periodic system the dielectric function can be conveniently expressed in plane waves

$$\epsilon^{-1}(\mathbf{r},\mathbf{r}') = \sum_{\mathbf{G}\mathbf{G}'}\sum_{\mathbf{q}} e^{i(\mathbf{G}+\mathbf{q})\mathbf{r}}\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})e^{-i(\mathbf{G}'+\mathbf{q})\mathbf{r}'}, \quad (2)$$

where $\mathbf{G}$ is a reciprocal lattice vector, $\mathbf{q}$ belongs to the 1. BZ. Within the random phase approximation (RPA) we have

$$\epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = \delta_{\mathbf{G}\mathbf{G}'} - v_c(\mathbf{q}+\mathbf{G})\chi^0_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega), \quad (3)$$

where χ^0 is the non-interacting density response function. Here, v_c can be the Fourier representation of either the full or the truncated Coulomb interaction. For the calculations in this section we have used a 50 eV cutoff for the reciprocal lattice vectors to account for local field effects. The non-interacting response function, χ^0, was constructed from local density approximation (LDA) wave functions and energies including states up to 50 eV above the Fermi level. All calculations were performed with the projector augmented wave method code GPAW. Details on the implementation of the dielectric function in the GPAW code can be found in Ref. 34.

It follows from Eq. (2) that the total potential resulting from a plane wave external potential $V_0e^{i\mathbf{q}\cdot\mathbf{r}}$ has the form

$$V_{\rm tot}(\mathbf{r}) = \widetilde{V}_{\mathbf{q}}(\mathbf{r})e^{i\mathbf{q}\cdot\mathbf{r}} \quad (4)$$

where $\widetilde{V}_{\mathbf{q}}(\mathbf{r})$ is a lattice periodic function. We thus define the macroscopic dielectric constant as

$$\frac{1}{\epsilon_M(\mathbf{q})} \equiv \frac{\langle\widetilde{V}_{\mathbf{q}}\rangle_\Omega}{V_0} = \epsilon^{-1}_{\mathbf{00}}(\mathbf{q}), \quad (5)$$

where $\langle\ldots\rangle_\Omega$ denotes a spatial average over a unit cell. Note that in general $\epsilon_M(\mathbf{q},\omega) \neq \epsilon_{\mathbf{00}}(\mathbf{q},\omega)$ because of local field effects.[35,36]

To explicitly demonstrate that Eq. (5) does not apply to low-dimensional materials, we have calculated the macroscopic dielectric constant as a function of the layer separation, L. The results are shown in Fig. 6 for different values of the in-plane momentum transfer q. We also show the dielectric constant corresponding to polarization orthogonal to the layer. Clearly the macroscopic dielectric constant approaches unity for all q-vectors in the limit of large interlayer separation. This occurs because the total field is averaged over an increasingly larger vacuum region.

Previously reported values for the macroscopic dielectric constant of monolayer MoS_2 lie in the range $4-8$.[21,24,28] In these calculations the MoS_2 layers were separated by $10-20$ Å vacuum. As can be seen from $\epsilon_\parallel(q=0)$ in Fig. 6 this is consistent with our results. However, it should also be clear that numbers depend on the distance between layers and in fact are not meaningful.

B. 2D macroscopic dielectric constant

For a 2D material, the average of the total potential in the definition of the macroscopic dielectric constant must be confined to the region of the material. Since Eq. (4) still holds for a 2D material when $\mathbf{q}$ is confined to the plane of the material, we average the in-plane coordinates ($\mathbf{r}_\parallel$) over the unit cell area A and the out-of-plane

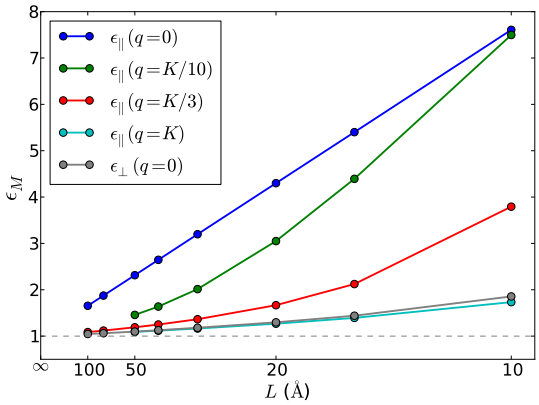


FIG. 6: (Color online). The 3D static macroscopic dielectric constant $1/\epsilon^{-1}_{\mathbf{00}}(\mathbf{q})$ of monolayer MoS_2 as a function of the interlayer separation, L. $\epsilon_{\parallel}$ is the dielectric constant with polarization parallel to the monolayer and $\epsilon_{\perp}$ is the dielectric constant for polarization orthogonal to the layer.

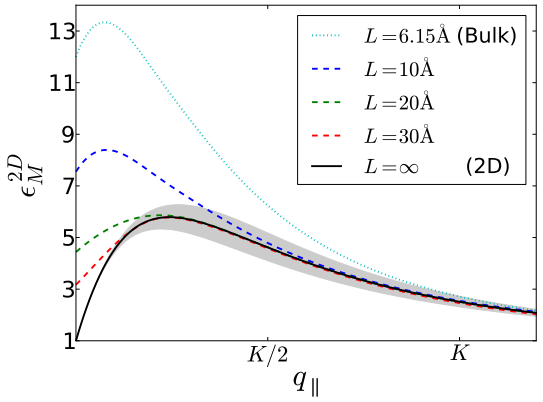


FIG. 7: (Color online). Static macroscopic dielectric constant for a single layer of MoS_2 calculated along the Γ-K line. The calculations are performed using Eq. (6) with the microscopic dielectric constant, $\epsilon^{-1}_{\mathbf{GG'}}(\mathbf{q})$, evaluated from Eq. (3) with either the bare Coulomb interaction (dotted and dashed lines) or truncated Coulomb interaction (full black line). The grey area represents the result obtained when the averaging region perpendicular to the layer, d, is varied by ±10%. The dotted line corresponds to a layer separation of 6.15 Å and thus coincide with the dielectric constant of bulk MoS_2. The curves have been interpolated from a 32×32 q-point mesh.

coordinate (z) from $z_0 - d/2$ to $z_0 + d/2$ where z_0 denotes the center of the material and d its width. The 2D macroscopic dielectric constant then becomes

$$\frac{1}{\epsilon^{2D}_M(\mathbf{q}_{\parallel})} \equiv \frac{\langle \widetilde{V}_{\mathbf{q}} \rangle_{A,d}}{V_0} = \frac{2}{d}\sum_{G_{\perp}} e^{iG_{\perp}z_0}\frac{\sin(G_{\perp}d/2)}{G_{\perp}}\epsilon^{-1}_{\mathbf{G0}}(\mathbf{q}_{\parallel}), \quad (6)$$

where the sum is over all $\mathbf{G}$ with $\mathbf{G}_{\parallel} = \mathbf{0}$. In this work we have taken $d = 6.15$ Å corresponding to the interlayer separation in bulk MoS_2. We shall return to the problem of chosing d below.

The results for the static dielectric constant evaluated from Eq. (6) using the bare Coulomb interaction is shown in Fig. 7 for four different layer separations. The result for $L = d = 6.15$ Å coincides with the 3D dielectric constant of bulk MoS_2 given by Eq. (5). The result obtained with the truncated Coulomb interaction is shown in black; it represents the case of infinite layer separation. Before discussing the results, it is instructive to consider the potential arising from a 2D charge density fluctuation of the form,

$$n(\mathbf{r}) = n_0 e^{i\mathbf{q}_{\parallel}\cdot\mathbf{r}_{\parallel}}\delta(z), \quad (7)$$

The corresponding potential follows from Poisson's equation[44]

$$\phi(\mathbf{r}) = \frac{n_0}{q_{\parallel}} e^{-i\mathbf{q}_{\parallel}\cdot\mathbf{r}_{\parallel}} e^{-q_{\parallel}|z|}. \quad (8)$$

It follows that the potential perpendicular to the layer falls off exponentially over a characteristic distance of $1/q_{\parallel}$. This explains why in general $\epsilon^{2D}_M(\mathbf{q}_{\parallel})$ coincides with the isolated layer result for $q_{\parallel} \gtrsim 1/L$.

The variation of ϵ^{2D}_M when the parameter d is changed by ±10% is indicated by the shaded region in Fig. 7. To the left of the maximum, $\epsilon^{2D}_M(\mathbf{q}_{\parallel})$ is insensitive to d since the induced potential is more or less constant over the averaging region. To the right of the maximum, the variation in $\epsilon^{2D}_M(\mathbf{q}_{\parallel})$ follows the ±10% variation in d. This is because for these wave vectors the induced potential has essentially vanished at the borders of the averaging region. In general, increasing (decreasing) d will decrease (increase) $\epsilon^{2D}_M(\mathbf{q}_{\parallel})$ in the large wave vector region.

Another characteristic feature of the potential in Eq. (8) is the $1/q_{\parallel}$ scaling which should be compared with the $1/q^2$ form of the Coulomb potential from a 3D charge oscillation. Since the non-interacting response function, $\chi^0_{\mathbf{00}}(\mathbf{q})$, scales as $\sim q^2$ for $q \to 0$ for both 2D and 3D systems, it follows from Eq. (3) that $\epsilon^{2D}_M(0) = 1$, while this is in general not the case in 3D. In our calculations, the effect of interlayer interactions is eliminated by using a truncated Coulomb interaction of the form $v_c(\mathbf{r}) = (1/r)\theta(R_c - z)$. For $q_z = 0$ and in the limit of small $q_{\parallel}$, the Fourier representation of the truncated Coulomb interaction becomes $v^{2D}(\mathbf{q}) = \frac{4\pi R_c}{|\mathbf{q}|}$, i.e. it scales as $1/q$ as the potential from the 2D charge density wave ensuring the correct limit $\epsilon^{2D}_M(0) = 1$.

Finally, we note that previous studies[37,38] have employed a strict 2D model for the dielectric function in the small q limit of the form $\epsilon(q_{\parallel}) = 1 + \alpha q_{\parallel}$. This form is convenient as it leads to an analytical expression for the screened interaction in 2D.[27] Our definition differs by being a 3D (or quasi 2D) quantity valid for general $q_{\parallel}$.

C. Screened interaction

In Fig. 8 we show $\epsilon^{-1}_{\mathbf{00}}$ as a function of $q_{\parallel}$ evaluated with and without the truncated Coulomb interaction. For small q, the two curves differ significantly due to the long range nature of the induced potential (8). At large q ($\sim K/2$), the induced potential decays within the cutoff range for the truncated Coulomb interaction and therefore no difference can be seen between the two methods. We emphasize that neither of the dielectric constants shown in the figure can be interpreted as a dielectric constant of monolayer MoS_2, since they give the average potential over the supercell and not over the MoS_2 layer. In particular their value will be highly dependent on the size of the unit cell (in the limit of infinite layer separation both will equal 1 for all q). Nevertheless, this quantity is a crucial ingredient of both the GW self-energy and the BSE kernel as it provides the screening of the divergent term of the Coulomb interaction.

For $q = 0$ the Coulomb kernel diverges and we approximate $W(q = 0)$ by the integral

$$W_{\mathbf{00}}(\mathbf{q}=0) = \frac{1}{\Omega_\Gamma}\int_{\Omega_\Gamma} d\mathbf{q}\, v_c(\mathbf{q})\epsilon^{-1}_{\mathbf{00}}(\mathbf{q}) \approx \frac{1}{\Omega_\Gamma}\epsilon^{-1}_{\mathbf{00}}(\mathbf{q}=0)\int_{\Omega_\Gamma} d\mathbf{q}\, v_c(\mathbf{q}), \qquad (9)$$

where Ω_{BZ} is the Brillouin zone volume and Ω_Γ is a small volume containing $\mathbf{q} = 0$. In isotropic systems $\epsilon^{-1}_{\mathbf{00}}(\mathbf{q})$ is usually constant in the vicinity of $\mathbf{q} = 0$ and the approximation works well. However, when ϵ^{-1} is evaluated with the truncated Coulomb interaction, $\epsilon^{-1}_{\mathbf{00}}$ acquires much more structure for small q as can be seen from Fig. 8. Thus, for coarse k-point samplings we will underestimate the Γ-point screening since we simply use $\epsilon^{-1}_{\mathbf{00}}(\mathbf{q} = 0) = 1$.

The linear behavior of the screened interaction for small q suggests that a better approximation for $W_{\mathbf{00}}(\mathbf{q} = 0)$ would be

$$W_{\mathbf{00}}(\mathbf{q}=0) = \frac{1}{\Omega_\Gamma}\int_{\Omega_\Gamma} d\mathbf{q}\, v_c(\mathbf{q})\Big[1 + \mathbf{q}\cdot\nabla_{\mathbf{q}}\epsilon^{-1}_{\mathbf{00}}(\mathbf{q})\Big|_{\mathbf{q}=0}\Big]. \qquad (10)$$

Since the dielectric matrix in RPA is $\epsilon_{\mathbf{GG'}}(\mathbf{q}) = 1 - v^c(\mathbf{q})\chi^0_{\mathbf{GG'}}(\mathbf{q})$, we can derive an analytic expression for the first order Taylor expansion in q, and its inverse. These quantities can be evaluated with vanishing additional cost, but we will leave the assessment of this approximation to future work.

IV. OPTICAL ABSORPTION SPECTRUM

In this section we present many-body calculations of the optical absorption spectrum of monolayer MoS_2 by solving the Bethe-Salpeter Equation (BSE) under the standard assumption of static screening of the electron-hole interaction. As for the GW band gap, we find that the use of a truncated Coulomb interaction is essential

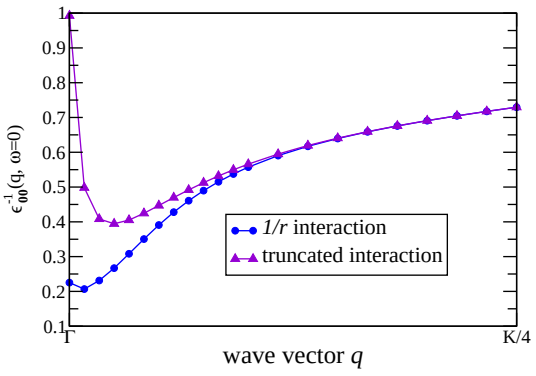


FIG. 8: (Color online). The 3D static inverse dielectric constant $\epsilon^{-1}_{\mathbf{00}}(\mathbf{q})$ of monolayer MoS_2 calculated in the RPA for different values of in-plane momentum transfer q along the Γ-K direction. The separation between layers is $L = 20$ Å. Note that neither of the quantities can be interpreted as the macroscopic dielectric constant of the monolayer (this quantity is the black curve in Fig. 7).

to avoid interlayer screening and obtain well converged exciton binding energies. Furthermore, the very strong q-dependence of the 2D static dielectric function around $q = 0$, leads to very slow k-point convergence for the exciton binding energy.

In order to obtain an accurate absorption spectrum including excitonic effects we calculate the response function from the Bethe-Salpeter Equation (BSE). Using the standard assumption of a static dielectric screening of the electron-hole (e-h) interaction, the BSE[39] can be recast as an effective two-particle Hamiltonian,[40] which is diagonalized on a basis of electron-hole pairs. In this way the excitonic eigenstates can be expressed as a linear combination of single-particle transitions

$$|\lambda\rangle = \sum_{vck} A^\lambda_{vck}|vck\rangle, \qquad (11)$$

where v, c, and k denote valence band, conduction band and Brillouin zone wave vector, respectively. The absorption spectrum is proportional to the imaginary part of the macroscopic dielectric function, which in the Tamm-Dancoff approximation can be written

$$\epsilon_2(\omega) = 2\pi \lim_{\mathbf{q}\to 0} v_c(\mathbf{q}) \sum_\lambda \delta(\omega - E_\lambda) \times \left|\sum_{vc\mathbf{k}} A^\lambda_{vck}\langle v\mathbf{k}-\mathbf{q}|e^{-i\mathbf{q}\cdot\mathbf{r}}|c\mathbf{k}\rangle\right|^2, \qquad (12)$$

where E_λ are the eigenvalues associated with $|\lambda\rangle$.

In all calculations we have included a single valence band and a single conduction band in the BSE Hamiltonian. We have tested that the first excitonic peak is completely unaffected if we instead include 6 valence bands and 4 conduction bands. This is also expected since the

highest (lowest) valence (conduction) band is well isolated from the remaining bands at K where the exciton is centered, see Fig. 1. For the screening we have included 65 bands in the evaluation of the response function, which is sufficient for converged results. Increasing the number of bands to 300 affects the position of the first exciton by less than 10 meV. The plane wave cutoff for the response function (local field effects) was set to 50 eV and we checked that the excitonic binding energy changed by less than 10 meV when increasing the cutoff to 200 eV. The dependence on k-point sampling and interlayer separation will be examined below. Details on the implementation of the BSE method in the GPAW code can be found in Ref. 5.

A. Convergence tests

In the lower panel of Fig. 9, we show the exciton binding energy as a function of interlayer separation calculated for different k-point samplings using the bare Coulomb interaction and the truncated Coulomb interaction. With the bare Coulomb interaction, the obtained results are far from convergence, even for $L = 50$ Å. The dependencies on the layer separation and number of k points is very similar as for the quasiparticle gap discussed in Sec. II B, even on a quantitative level. Therefore, the optical gap, which is given by the difference of the QP gap and the exciton binding energy, is almost indenpedent of L and whether or not the truncation method is used, as shown in the upper panel. This is consistent with the observations in Ref. 22.

The convergence of the binding energy with respect to the k-point sampling is plotted in Fig. 10 for an interlayer separation of 20 Å. The truncated Coulomb kernel gives a much slower convergence with respect to the number of k-points than the bare Coulomb interaction. However, it should be clear from Fig. 9 that the binding energy obtained with the bare Coulomb interaction converges to a value which is highly dependent on the interlayer separation. In Ref. 25, convergence was found with 18×18 k points, but for a layer separation of only 24 Å. The obtained exciton binding energy was around 0.2 eV. According to our results, this is much too weak due to interlayer screening.

The slow k-point convergence observed when using the truncated Coulomb interaction is related to the q-dependence of the screening in two-dimensional systems. As demonstrated by Eq. (9) and Fig. 8 (blue curve), a too low k-point sampling leads to an underestimation of the screening in the vicinity of $q = 0$ and thus an overestimation of the exciton binding energy.

B. Results

From the convergence tests described above we conclude that the BSE calculations are (nearly) converged if

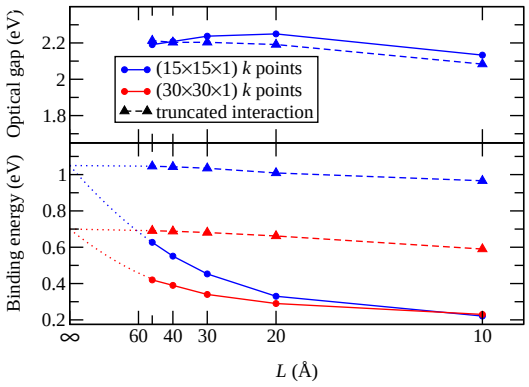


FIG. 9: (Color online). Optical gap and binding energy of the lowest exciton in monolayer MoS_2 as a function of interlayer separation calculated from the BSE and the G_0W_0 quasiparticle gap. Results with the full $1/r$ Coulomb interaction (full lines) and the truncated interaction (dashed lines) are shown. Dotted lines give an estimation for extrapolation to infinite L.

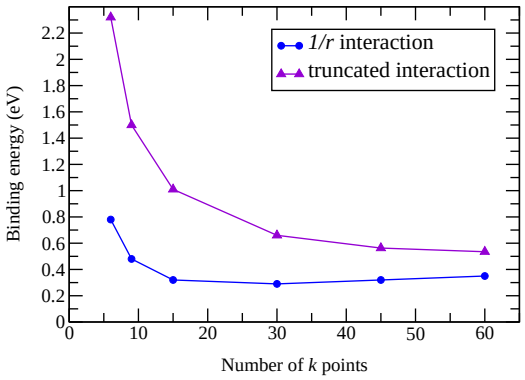


FIG. 10: (Color online). Binding energy of the lowest exciton in monolayer MoS_2 as a function of k-point sampling for a supercell with a layer separation of $L = 20$ Å.

we use a truncated Coulomb interaction and a 45×45 k-point sampling. With these setting we have calculated the BSE spectrum on top of a G_0W_0 quasiparticle band structure obtained with the same parameters. The resulting absorption spectrum is shown in Fig. 11. We also show an RPA calculation, i.e. neglecting electron-hole interactions in the BSE, performed on top of the same G_0W_0 band structure for comparison. With electron-hole interaction included, we obtain an exciton binding energy of ~ 0.6 eV, whereas RPA does not show an exciton peak and simply gives an absorption edge at the band gap.

Experimentally, the absorption spectrum of single layer MoS_2 exhibits a spin-orbit split peak around 1.9 eV.[7] Since we have not included spin-orbit coupling in our calculations, the spectrum Fig. 11 only shows a single peak at low energies. However, it has previ-

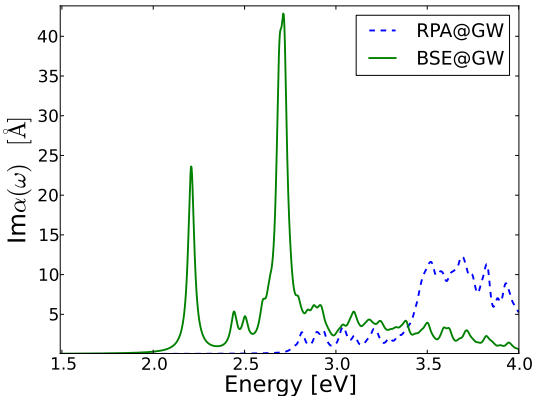


FIG. 11: (Color online). Absorption spectrum of single layer MoS_2 calculated with the RPA and BSE using the G_0W_0 quasiparticle band structure. The calculation has been performed with a truncated Coulomb interaction to avoid interactions between repeated layers and with a 45×45 k-point grid.

ously been shown[4,24] that the spin-orbit coupling does not have a large effect on the exciton binding energy and only results in a split excitonic peak. The main peak in the BSE@G_0W_0 spectrum is situated at 2.2 eV which is 0.3 eV higher than the experimental value. At present we cannot say if this is due to an insufficient description of the quasiparticle gap within G_0W_0 or underestimation of the exciton binding energy in BSE with a static electron-hole interaction.

From the above discussion it should be clear that it is extremely challenging to converge the exciton binding energy with respect to interlayer separation and k points. In general, the optical gap is much easier to converge with respect to interlayer separation than either the quasiparticle gap or the exciton binding energy.[22,41] Nevertheless, for many physical applications it is of importance to obtain accurate values for both the quasiparticle gap and the exciton binding energy separately. In Ref. 22 the exciton binding energy was obtained by $1/L$ extrapolation of the quasiparticle gap calculated in a range of interlayer separations between 10 and 20 Å and assuming the same dependence for the exciton binding energy. Our results indicate that one should be cautious with such extrapolations. This is because the screening at different q-points has a very different dependence on interlayer separation, which results in different convergence behavior at different k-point samplings (see Fig. 9 full lines). The extrapolation procedure may therefore not give reliable results, since higher k-point samplings are required at larger interlayer separation. We are aware that the convergence issues may depend a lot on the implementation of the BSE method. However, we have previously performed the same calculations with YAMBO[42] code, which produced very similar convergence behavior for quasiparticle gap and exciton binding energy (also using truncated Coulomb cutoff and 45×45 k-point sampling).

V. CONCLUSIONS

We have presented a careful investigation of the dielectric properties, band gap and excitonic states in a two-dimensional semiconductor exemplified by monolayer MoS_2. We have demonstrated that the "standard" macroscopic dielectric constant used for solids is not applicable (meaningless) to supercells describing the 2D material as an infinite array of parallel sheets, and therefore replaced it by a 2D version in which the induced field is averaged over the extent of the material rather than over the entire supercell. We showed that the effect of interlayer screening leads to underestimation of the band gap and exciton binding energy by up to more than 0.5 eV for layer separations < 30 Å. The reason for this is that interlayer screening affects $\epsilon(q)$ for $q < 1/L$ where L is the distance between layers in the supercell. Since it is the small q limit of $\epsilon(q)$ that is most important for the screened interaction $W(q) = \epsilon^{-1}(q)/q^2$, the effect cannot be neglected. Here we have circumvented the problem by using a truncated Coulomb interaction that explicitly cuts off the interaction between neighboring layers.

The properly defined 2D dielectric function $\epsilon_M^{2D}(q)$ has a very sharp wave vector dependence for small q and satisfies $\epsilon_M^{2D}(0) = 1$ in general. This has the consequence that quasiparticle- and optical excitations obtained from the GW and Bethe-Salpeter Equation, respectively, require much denser k-point grids than experience from 3D systems would suggest. For MoS_2 we find that a precision of 0.2 eV requires k-point grids of at least 30×30. Interestingly, the effect of interlayer screening and too small k-point grids have opposite effects on the band gap and exciton energies leading to fortuitous error cancellation. Our calculations applying the truncated Coulomb interaction and 45×45 k points give G_0W_0@LDA gaps of 2.77 eV (direct) and 2.58 eV (indirect) and binding energy of the lowest exciton of 0.55 eV. This places the lowest exciton at $\sim$ 2.2 eV which is 0.3 eV higher than the experimental result. This difference may be due to the G_0W_0@LDA approximation or the use of static screening in the BSE.

VI. ACKNOWLEDGEMENT

KST acknowledges support from the Danish Council for Independent Research's Sapere Aude Program through grant no. 11-1051390. The Center for Nanostructured Graphene (CNG) is sponsored by the Danish National Research Foundation, Project DNRF58.

* Electronic address: falco.hueser@fysik.dtu.dk
† Electronic address: thygesen@fysik.dtu.dk

[1] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, Proc. Natl. Acad. Sci. U.S.A. **102**, 10451 (2005).
[2] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, Nature Nanotechnology **6**, 147 (2011).
[3] K. Kaasbjerg, K. S. Thygesen, and K. W. Jacobsen, PHYSICAL REVIEW B **85**, 115317 (2012).
[4] H. Shi, H. Pan, Y.-W. Zhang, and B. I. Yakobsen, Phys. Rev. B **87**, 155304 (2013).
[5] J. Yan, K. W. Jacobsen, and K. S. Thygesen, PHYSICAL REVIEW B **86**, 045208 (2012).
[6] L. Britnell, R. M. Ribeiro, A. Eckmann, R. Jalil, B. D. Belle, A. Mishchenko, Y. J. Kim, R. V. Gorbachev, T. Georgiou, S. V. Morozov, et al., SCIENCE **340**, 1311 (2013).
[7] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Phys. Rev. Lett. **105**, 136805 (2010).
[8] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, Nano Lett. **10**, 1271 (2010).
[9] M. Bernardi, M. Palummo, and J. C. Grossman, Nano Lett. (to be published).
[10] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, Nature Nanotechnology **7**, 699 (2012).
[11] M. V. Bollinger, J. Lauritsen, K. W. Jacobsen, J. K. Nørskov, S. Helveg, and F. Besenbacher, Phys. Rev. Lett. **87**, 196803 (2001).
[12] T. F. Jaramillo, K. P. Jørgensen, J. Bonde, J. H. Nielsen, S. Horch, and I. Chorkendorff, Science **317**, 100 (2007).
[13] X. Zong, H. Yan, G. Wu, G. Ma, F. Wen, L. Wang, and C. Li, J. Am. Chem. Soc. **130**, 7176 (2008).
[14] R. F. Frindt and A. D. Yoffe, Proc. R. Soc. Lond. A **273**, 69 (1963).
[15] B. L. Evans and P. A. Young, Proc. R. Soc. Lond. A **284**, 402 (1965).
[16] J. A. Wilson and A. D. Yoffe, Adv. Phys. **18**, 193 (1969).
[17] R. A. Neville and B. L. Evans, Phys. Stat. Sol. B **73**, 597 (1976).
[18] G. L. Frey, S. Elani, M. Homyonfer, Y. Feldman, and R. Tenne, Phys. Rev. B **57**, 6666 (1998).
[19] J. P. W. Newcomer and G. A. Samara, J. Appl. Phys. **81**, 7934 (1997).
[20] R. Coehoorn, C. Haas, and R. A. de Groot, Phys. Rev. B **35**, 6203 (1987).
[21] T. Cheiwchanchamnangij and W. R. L. Lambrecht, Phys. Rev. B **85**, 205302 (2012).
[22] H.-P. Komsa and A. V. Krasheninnikov, Phys. Rev. B **86**, 241201 (2012).
[23] C. Ataca and S. Ciraci, The Journal of Physical Chemistry C **115**, 13303 (2011), http://pubs.acs.org/doi/pdf/10.1021/jp2000442, URL `http://pubs.acs.org/doi/abs/10.1021/jp2000442`.
[24] A. Ramasubramaniam, Phys. Rev. B **86**, 115409 (2012).
[25] A. Molina-Sánchez, D. Sangalli, K. Hummer, A. Marini, and L. Wirtz, Phys. Rev. B **88**, 045412 (2013), URL `http://link.aps.org/doi/10.1103/PhysRevB.88.045412`.
[26] H. J. Conley, B. Wang, J. I. Ziegler, R. F. Haglund, S. T. Pantelides, and K. I. Bolotin, Nano Lett. **13**, 3626 (2013).
[27] L. V. Keldysh, JETP Lett. **29**, 658 (1979).
[28] A. Molina-Sanchez and L. Wirtz, Phys. Rev. B **84**, 155413 (2011).
[29] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Duak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, et al., Journal of Physics: Condensed Matter **22**, 253202 (2010), URL `http://stacks.iop.org/0953-8984/22/i=25/a=253202`.
[30] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, Phys. Rev. B **73**, 205119 (2006).
[31] S. Ismail-Beigi, Phys. Rev. B **73**, 233103 (2006).
[32] F. Hüser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **87**, 235132 (2013), URL `http://link.aps.org/doi/10.1103/PhysRevB.87.235132`.
[33] D. Y., Y. Wang, J. Ni, L. Shi, S. Shi, and W. Tang, Physica B **406**, 2254 (2011).
[34] J. Yan, J. J. Mortensen, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **83**, 245122 (2011), URL `http://link.aps.org/doi/10.1103/PhysRevB.83.245122`.
[35] S. ADLER, PHYSICAL REVIEW **126**, 413+ (1962), ISSN 0031-899X.
[36] N. Wiser, Phys. Rev. **129**, 62 (1963), URL `http://link.aps.org/doi/10.1103/PhysRev.129.62`.
[37] P. Cudazzo, I. V. Tokatly, and A. Rubio, Phys. Rev. B **84**, 085406 (2011), URL `http://link.aps.org/doi/10.1103/PhysRevB.84.085406`.
[38] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, Phys. Rev. B **88**, 045318 (2013), URL `http://link.aps.org/doi/10.1103/PhysRevB.88.045318`.
[39] E. E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951).
[40] G. Strinati, Phys. Rev. B **29**, 5718 (1984).
[41] L. Wirtz, A. Marini, and A. Rubio, Phys. Rev. Lett. **96**, 126104 (2006).
[42] A. Marini, C. Hogan, M. Grüning, and D. Varsano, Comp. Phys. Comm. **180**, 1392 (2009).
[43] The `gpaw` code is available as a part of the CAMPOS software: `www.camd.dtu.dk/Software`
[44] This is most easily seen by performing a 3D Fourier transformation of δn, then multiplying by $1/q^2$ and Fourier transforming back to real space.