Ph.D. Thesis
Doctor of Philosophy

DTU Physics
Department of Physics

First Principles Calculations of Electronic Excitations in 2D Materials

Filip Anselm Rasmussen

Kongens Lyngby 2016

DTU Physics
Department of Physics
Technical University of Denmark

Fysikvej, Building 307
2800 Kongens Lyngby, Denmark
Phone +45 4525 3344
info@fysik.dtu.dk
www.fysik.dtu.dk

Summary

Since the first reported synthesization of graphene - an atomically thin carbon material - in 2004 there has been a surge of research in discovering other novel two-dimensional materials. The reason is clear: two-dimensional materials are thought to be able to lead to new fast and low-power ultra-thin electronics and high efficiency solar cells. Contrary to many other nano-materials, methods for large scale fabrication and patterning have already been demonstrated and the first real technological applications have already be showcased. Still the technology is very young and the number of well-studied 2D materials are few. However as the list of 2D materials is growing it is necessary to investigate their fundamental structural, electronic and optical properties. These are determined by the atomic and electronic structure of the materials that can quite accurately predicted by computational quantum mechanics methods.

One of these methods, Density Functional Theory (DFT), has been bery successful at determining structural properties of 2D materials. It is however well-known that it less accurate when it comes to predicting the energy levels of excited states that are important in order to determine electronic transport, optical and chemical properties. On the other hand it has shown to be a great starting point for a systematic pertubation theory approach to obtain the so-called quasiparticle spectrum. In the *GW* approximation one considers the considers the potential from a charged excitation as if it is being screened by the electrons in the material. This method has been very successful for calculating quasiparticle energies of bulk materials but results have been more varying for 2D materials. The reason is that the 2D confined electrons are less able to screen the added charge and some of the numerical methods that are efficient for bulk systems become invalid.

In this thesis I describe the study of a set of novel 2D materials and establish their electronic and optical properties using DFT and the *GW* approximation while taking the reduced screening properly into account as well as taking regard to other numerical problems that have often been neglected. Secondly I show how one can efficiently take the 2D nature into account in the *GW* approximation and thereby make future calculations require much less computational resources.

Resumé

Siden grafen - et atomart tyndt kulstofmateriale - blev syntetiseret første gang i 2004 har der været en bølge af forskning i opdagelsen af nye todimensionale materialer. Årsagen er klar: todimensionale materialer menes at være i stand til at lede til nye hurtige og strømbesparende elektronikkomponenter og højeffektive solceller. Modsat mange andre nanomaterialer er metoder til fabrikation og modifikation på stor skala allerede blevet demonstreret og de første rigtige teknologiske anvendelser er allerede blevet fremvist. Teknologien er dog stadig ung og antallet ag velstuderede 2D-materialer er få. Som listen af 2D materialer gror er det nødvendigt at undersøge deres fundamentale strukturelle, electriske og optiske egenskaber. Disse er bestemt af den atomare og elektroniske struktur af materialerne og kan forholdsvis præcist forudsiges af computerbaserede kvantemekaniske metode.

En af disse metoder, tæthedfunktionalteori (DFT), har været utrolig successfuld til at bestemme strukturelle egenskaber af 2D-materialer. Men det er velkendt at den er mindre præcis når det kommer til at forudsige energiniveauerne for exciterede tilstande, der er vigtige for at kunne bestemme kemiske, optiske og ledningsmæssige egenskaber. På den anden side har den vist sig at være godt startpunkt for en systematisk pertubationsteoretisk fremgangsmåde til at beregne det såkaldte kvasipartikelspektrum. I GW-approksimationen betragter man potentialet fra en laded excitation som var det skærmet af alle elektronerne i materialet. Denne metode har været særdeles effektiv til at beregne kvasipartikelenergier for tre-dimensionale materialer men resultaterne har været mere blandede for 2D-materialer. Årsagen er at de 2D-indesluttede elektroner er mindre i stand til effektivt at skærme den tilføjede ladning og derved bliver nogle af de numeriske metoder, der er effektive for 3D systemer, ubruglige.

In denne afhandling beskriver jeg et studie af en gruppe af nye 2D materialer og bestemmer deres elektriske og optiske egenskaber ved anvendelse af DFT og GW-approksimationen, hvor jeg samtidigt tager ordentligt højde for den reducerede skærmning samt andre numeriske problemer, der ofte ellers er blevet negligeret. Derudover beskriver jeg en metode hvorleded man effektivt kan tage 2D-egenskaberne med i betragtning i GW-approksimationen hvorved fremtidige beregninger vil kræve langt færre computerresourcer.

Preface

This thesis is submitted in candidacy for the Ph.D. degree in physics from the Technical University of Denmark. The work has been carried out at the Department of Physics in the period from October 2012 to December 2015 under the supervision of Professor Kristian S. Thygesen. Funding has been provided by the Center of Nanostructured Graphene through a grant from the Danish National Research Council; Project DNRF58.

First and foremost I would like to thank my supervisor Kristian Thygesen for his guidance and support. I am happy that he gave me the opportunity to be a part of the computational physics research that I have come to appreciate so much.

I would very much like to thank Falco Hüser whose *GW* code I apparently became the heir to but a special appreciation goes to Jens Jørgen Mortensen for completely rewriting the code introducing improvements that made this project possible. He has also been a great aid when one had technical problems or was unsure about the technical foundations. I would also like to thank Thomas Olsen, Simone Latini, Morten Gjerding, Kirsten Andersen, Mohnish Pandey and Chengjun Jin for collaboration and many fruitful discussions that have benefited my project. Many thanks also go to the current and previous people in the office and at CAMD and Department of Physics for being good friends and making work a joyful experience: Per Schmidt, Korina Kuhar, Ivano Castelli, Kristian Ørnsøe, Chris Patrick, Martin Hangaard, Juan Maria García Lastra and Marta Majkut. I would also like to thank Ole Holm Nielsen for keeping Niflheim running and for many great stories and Marianne for her kind help with the administrative tasks.

Filip Anselm Rasmussen
Kongens Lyngby, December 2015

Contents

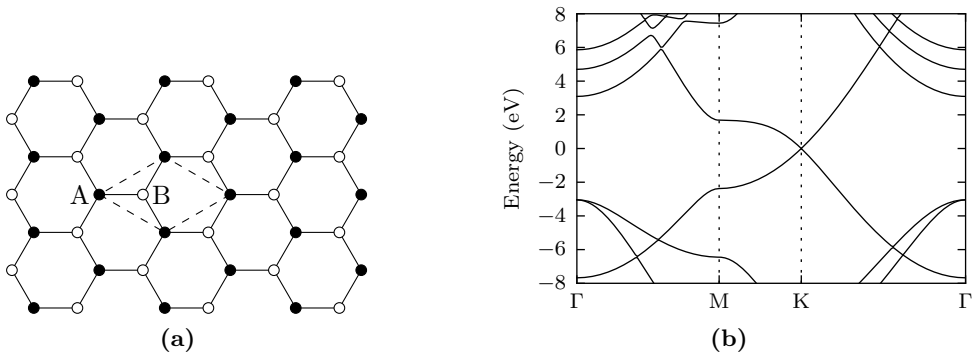
CHAPTER 1

Introduction

For a long time scientists and engineers have mostly been interested in the properties of bulk materials; most of the world we see is made up of matter where the tiniest features do not matter much. For most technological purposes a metal, for instance, can be described as a homogeneous solid whose electronic properties can more or less be determined from its conductivity which takes a constant value throughout the material. This means that many electronic devices can be modeled by use of simple relations like Ohm's and Kirchoff's laws. Even more complex geometries can be modeled using continuous versions of these laws. Of course the basic properties are directly related to the fundamental atomic structure of the material, but for most practical applications this can be taken to be similar everywhere in the material. That these basic properties for materials can be predicted and calculated reliably just from the atomic structure has been a big triumph for condensed matter physics and especially the ab-initio community. However with the invention of the transistor it was recognized that the physics of interfaces can also lead to interesting and useful phenomena. The advent of semiconductor technology has lead industries to consider ever smaller feature sizes of their devices, yet it has still been possible to predict their behavior from the properties of the bulk materials they consists of. However the continuous requirement for faster and more efficient electronics will soon push the requirements for the feature sizes below the limit where bulk properties give an adequate description and one will have to use methods where the features at atomic scales are properly included. This may also very well lead to completely new kinds of devices that require very different kinds of materials than those used in the semiconductor industry today. In fact, instead of trying to mold or shape the material in a way so that it fits within the current technology one may use the inherent microscopic features of another material in a different way so that one ends up with the desired result.

The first of such naturally microstructured materials that were discovered were the carbon nanotubes, which saw a great deal of research interest up through the 1990s and they were promised great technological success due to their unique properties: high tensile strength and varying electronic properties; from metallic to insulating, that can to a high degree be controlled. Despite this their widespread technological breakthrough has not happened yet, mainly due to challenges in scaling up the methods for controlled growth and fabrication of nanotube devices. However, the research in carbon nanotubes made scientists look into the basic ingredient for making carbon

nanotubes, namely carbon or more precise its bulk phase allotrope graphite. Graphite is basically a stack of single sheets of carbon atoms arranged in a two-dimensional honeycomb lattice. The sheets are only weakly bound and can be made to slide among themselves when a relatively small force is applied. This had been known for a long time and is the reason why graphite is used in pencils and as a dry lubricant. While carbon nanotubes were made by single or few layers of graphite rolled up into a tube, single layers of graphite were not thought to be thermodynamically stable[60, 76, 61]. However, in 2004 Geim and Novoselov reported that it was indeed possible to isolate single stable layers of graphite, dubbed graphene, and that these layers had extraordinary properties, for instance they found an extremly high charge carrier mobility[79]. Graphene is special in that the lattice structure consists of two identical hexagonal sublattices, see Figure 1.1(a) which results in a band crossing just at the Fermi level as seen in Figure 1.1(b). Around this point called the Dirac point the dispersion is approximately linear which means that the low energy charged states effectively acts as massless fermions, which results in the high mobility and various other special properties like room temperature quantum Hall effect[82, 125, 80] and so-called Klein tunnelling[50]. While few layer graphite had been studied for decades and the theory of graphene had actually been developed to describe bulk graphite[24, 23], the discovery of single layer graphene and subsequent exploration of its properties proved to be a breakthrough which lead to the Nobel Prize in 2010. Due to the high carrier mobility graphene was initially thought to be ideal for fast and efficient low-loss electronics and their seminal paper Geim and Novoselov actually illustrated its use in a field effect transistor. However, graphene is a semimetal and the lack of a band gap limits the on-off or switching ratio which makes it unuseful for todays high frequency electronics. A lot of subsequent research has then focused on ways to introduce a band gap in graphene without decreasing its extraordinary transport



(a) **(b)**

Figure 1.1: **(a)** The lattice structure of graphene consists of two identical hexagonal sublattices A and B. **(b)** Band structure of monolayer graphene calculated with DFT using the PBE exchange-correlation functional.

characteristic. Several methods were suggested and attempted including functionalizing the graphene sheets with chemical adsorbates, contructing nanoribbons[33], modulating the microscopic lattice structure by use of specific substrates[126] and even modifying the lattice structure on the nanoscale by making holes or other patterns[89]. Besides the technical challenges faced in these methods they all also appear to trade off some of the properties that made graphene extraordinary in the first place. So while the quest for engineering graphene continued some people started to think that maybe there exists other two-dimensional materials with properties similar to graphene but which are naturally gapped.

Another was quickly realized to exist in single layers[81]: hexagonal boron nitride (h-BN) has like graphene a layered honeycomb structure and was already known to also form nanotubes[103, 15]. But unlike graphene the diatomic unit cell break the sublattice symmetry and opens a band gap -- in fact a large band gap. h-BN is an insulator and is thus unuseful for most electronic applications in e.g. field effect transistors or solar cells, but it has also shown to be quite inert and has proven to be quite good to protect single layers of other materials from the environment. However along with h-BN it was shown to be possible to exfoliate single layers of other wellknown layered materials[81]. One of the most notable examples is MoS_2 -- a material that like graphene and h-BN has a honeycomb-like crystal structure but with a three atom unit cell. It was shown that MoS_2 has a sizeable band gap while still yielding very high mobilities[6, 7] but it was also found that contrary to bulk MoS_2, single layers had a direct optical band gap opening possibilities for photonic devices[75]. MoS_2 is part of a family of layered materials known as transition-metal dichalcogenides for several of which it was found to be possible to fabricate monolayers, notably $MoSe_2$, WS_2 and WSe_2[16]. These turned out to have a range of interesting properties like large spin-orbit splittings[128] and exceptionally large exciton binding energies[99]. These interesting discoveries has ensured a massive research effort in these materials but many properties are still not fully enlightened and more interesting materials may await discovery.

But what makes two-dimensional materials so special? From a fundamental point of view two-dimensional systems may have properties that are vastly different than three-dimensional systems. For a long time quasi-two-dimensional systems like gallium arsenide heterostructures have been studied as an approximation to a two-dimensional electron gas (2DEG). Such a systems have quantized Landau levels when subject to a magnetic field and under certain circumstances such a system can exhibit the Fractional Quantum Hall Effect. Many high temperature superconductors like the copper oxides and iron pnictides also have layered structure where two-dimensional effects may play a big part of the explanation. While this is leads to interesting effects from a fundamental point of view, materials like graphene and transition metal dichalcogenides are tecnologically appealing because they allow for ultrathin, fast and efficient electronic devices. Besides the intrinsic properties the fact that the electrons are confined to move in a plane also means that electrons become inefficient at screening an external applied electric field, which make them very desirable for Field Effect Transistors. It has also been shown that several layers

of different 2D materials can be stacked to form vertical devices with completely new functionality[22, 28].

In order to guide future experimental, technological and theoretical effort it is necessary to get an accurate and throrough understanding of the basic electronic and optical properties of novel two-dimensional materials. There has initially been quite a bit of confusion and disagreement about what quantities of these properties for some materials. For instance the band gap of MoS_2 was initially estimated by computer calculations to be a semiconductor with a direct gap of 1.8 eV[71, 64] which seemed to fit quite well with an experimentally observed value of 1.9 eV inferred from photoluminescence. However, it has later been discovered that this agreement was more due to a fortunate coincidence of error cancellation. It is well known that the band gaps obtained directly from Density Functional Theory used the the initial calculations are typically grossly underestimated and at the same the low screening capability of two-dimensional materials means that the binding energies of optical excitonic states are much higher than in typical bulk systems, where they can often be ignored. This means that the first photoluminescence experiments measured the excitonic band gap and not the electronic band gap.

Several computational studies were done in order to find new two-dimensional materials and elucidate their properties, but in order to get accurate estimates of the electronic and optical band gap one has to employ methods that go beyond standard Density Functional Theory. One way is to use many-body pertubation theory, where the *GW* of Hedin has had great success in predicting accurate band gaps. While this method had also early on been used on 2D materials it was shown that in order to get accurate results one has to properly take into account the weak screening present inherent in these systems[42]. Thus many studies have failed to provide reproducible results when calculations were properly converged.

It has been the aim of this project to study a group of two-dimensional materials, some of which have been studied before, others are new and unstudied. Emphasis has been on obtained accurate and predictable results where the 2D nature and low screening is correctly taken into account and calculations are ensured to be converged and reliable. To do this both Density Functional Theory and the *GW* approximation has been used and estimates of exciton binding energies have been obtained with a simple Mott-Wannier model. It is also shown how convergence of the *GW* calculations can be made much more efficient by directly incorporating the special behavior of dielectric screening in 2D materials. The final goal is to start the creation of a database of reliable physical properties of 2D materials.

The thesis is organized as follows:

- Chapter 2 introduces the basic theoretical framework including the quantum many-body problem in condensed matter physics and Density Functional Theory.
- Chapter 3 goes through the the theory of linear response and many-body pertubation thery. It also described the *GW* approximation and briefly explains how it is implemented in practice.

- Chapter 4 explains how 2D systems are modelled computationally and introduces a new method for including the 2D dielectric screening in GW calculations.
- Chapter 5 goes through the computational procedure and results of a study of 216 2D transistion-metal dichalcogenides and -oxides.

It should be noted that the work described in Chapter 5 has been carried out before that described in Chapter 4. This means that the methods developed in Chapter 4 unfortunately were not used for obtaining the results in Chapter 5, but the order has been chosen like this in this thesis for a more coherent description.

CHAPTER 2

Density Functional Theory

2.1 The many-body atomic problem

The starting point for all calculations of all systems at the atomic level, be it calculations to determine the atomic structure of a material or molecule or electronic, optical or other physical properties, is the fundamental laws of quantum mechanics. In quantum mechanics the different particles like the electrons or the nucleons are described by their corresponding wavefunctions. At a fundamental level we best describe the world around us with the Standard Model. In this model all matter consists of certain fermionic elementary particles, either quarks or leptons (e.g. the electron), that interact via mediating bosonic particles like photons and gluons. While this model has proven extremely accurate using the full description to calculate anything but two interacting particles is incredibly hard and trying to describe a full atomic system consisting of many electrons and composite nuclei is basically impossible. Thus in atomic and condensed matter physics we try to simplify the model so to only describe the phenomena we are interested in, which are mainly effects detectable by optical or electronic experiments carried out at relatively low temperatures. The first approximation we do is to neglect subatomic particles and their interactions, since they are almost only responsible for nuclear decay, fission and fusion processes. Instead we typically describe atomic nuclei as different kinds of particles. Using this approach the total state of a system with N_a nuclei and N_e electrons can described by a wavefunction,

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_\mathrm{e}}; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_\mathrm{a}}), \tag{2.1}$$

where $\mathbf{r}_i$ denotes the position of an electron and $\mathbf{R}_i$ the position of an atomic nuclei. The wave function is the solution to the many-body Schrödinger equation given by

$$\hat{H}\Psi = E\Psi, \tag{2.2}$$

where the total Hamiltonian for an atomic system consisting of a number of ionic atom cores and electrons is given by

$$\hat{H} = (\hat{T}_\mathrm{ion} + \hat{V}_\mathrm{ion-ion}) + (\hat{T}_\mathrm{el} + \hat{V}_\mathrm{el-el}) + \hat{V}_\mathrm{el-ion} + \hat{V}_\mathrm{ext}. \tag{2.3}$$

Here $\hat{T}$ is the kinetic energy operator and $\hat{V}$ is the potential energy operator for the respective sub-systems. $\hat{V}_{\mathrm{ion-ion}}$ represents the Coulombic repulsion between the positively charged atomic cores, $\hat{V}_{\mathrm{el-el}}$ represents the Coulombic repulsion between the negatively charged electrons and $\hat{V}_{\mathrm{el-ion}}$ represents the mutually attractive forces between ions and electrons. In the first quantization picture the terms are given by

$$\hat{T}_{\mathrm{ion}} = \sum_a \left(-\frac{-\hbar^2}{2M_{\mathrm{ion},a}} \nabla_a^2 \right), \qquad V_{\mathrm{ion-ion}} = \frac{1}{2} \frac{1}{4\pi\varepsilon_0} \sum_{aa'} \frac{Z_a Z_{a'} e_0^2}{|\mathbf{R}_a - \mathbf{R}_a|}. \tag{2.4}$$

Here the subscripts a and a' refer to the ions present in the material under study. The electronic part takes the form

$$\hat{T}_{\mathrm{el}} = \sum_i \left(-\frac{\hbar^2}{2m} \nabla_i^2 \right) \qquad \hat{V}_{\mathrm{el-el}} = \frac{1}{2} \frac{1}{4\pi\varepsilon_0} \sum_{ij} \frac{e_0^2}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{2.5}$$

The part that couples the electrons and the ions is

$$\hat{V}_{\mathrm{el-ion}} = \frac{1}{4\pi\varepsilon_0} \sum_{i,a} \frac{(-Z_a e_0^2)}{|\mathbf{r}_i - \mathbf{R}_a|}. \tag{2.6}$$

One is usually interested in calculating several things: the atomic and electronic structure of a crystal or molecule, physical properties of a material like optical absorption spectrum, dielectric constant, electronic or thermal conductances and so on.

For calculating the atomic structure one is actually interested in finding the groundstate $|\Phi_0\rangle$ of the material characterized by having the lowest possible energy

$$E_0 = \langle \Phi_0 | \hat{H} | \Phi_0 \rangle, \quad E_0 \leq E_i = \langle \Phi_i | \hat{H} | \Phi_i \rangle \quad \forall \quad i, \tag{2.7}$$

which is already an extremely complicated problem since the Hilbert space is enourmous. Thus a large part of the research in condensed matter physics has been devoted to developing various approximation schemes that can simplify this problem to one that is more tangible.

In order to simplify the atomic problem an important observation is that the mass of the electron are orders of magnitude smaller than that of the nucleii. This means that for many purposes the dynamics of the electrons occur at much smaller timescales than for the nulcei. This leads to the Born-Oppenheimer approximation[9], where the total wave function is assumed to be a product of one governing the electrons and one governing the nuclei,

$$\Psi(\mathbf{r}, \mathbf{R}) = \psi(\mathbf{r}; \mathbf{R}) \chi(\mathbf{R}), \tag{2.8}$$

where $\mathbf{r} \equiv \{\mathbf{r}_i\}$ and $\mathbf{R} \equiv \{\mathbf{R}_j\}$. In this way the electron wavefunction are governed by an effective Schrödinger equation where the elctron only feel the potential from the nuclei from their average position,

$$\hat{H}_e \psi(\mathbf{r}; \mathbf{R}) = E_e(\mathbf{R}) \psi(\mathbf{r}; \mathbf{R}), \tag{2.9}$$

with the electronic Hamiltonian given by

$$\hat{H} = \sum_{i=1}^{N_e} \left(-\frac{1}{2}\nabla_i^2 - \sum_{a=1}^{N_a} \frac{Z_a}{|\mathbf{r}_i - \mathbf{R}_a|} \right) + \frac{1}{2} \sum_{i,j \neq i}^{N_e} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{2.10}$$

In Equation (2.10) an throughout the rest of this thesis I will use atomic units ($\hbar = m = e = 1/4\pi\varepsilon_0 = 1$) unless otherwise stated. We see that the nuclear coordinates simply enter the Schrödinger equation as parameters and that the electron-nuclear interaction can simply be thought as an external potential. When a solution to the electronic Schrödinger equation has been found one can the then take the electron density as input into a Schrödinger equation governing the nuclei. In the simplest approximation we can take the nuclei to be classical particles in which case the total energy is

$$E_{\text{tot}}(\mathbf{R} = E_e(\mathbf{R}) + V_{\text{nn}}(\mathbf{R}), \quad V_{\text{nn}}(\mathbf{R}) = \sum_{I=1}^{N_a} \sum_{J \neq I}^{N_a} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}. \tag{2.11}$$

Optimizing the nuclear positions such that the total energy is a minimum is called *structural relaxation* and several methods for doing so have been developed so that atomic structures of materials can be predicted efficiently. The corresponding time-independent problem can also be solved, and in this case the classical nuclei approximation leads to a set of Newtonian equations of motion for the nuclear positions and solution of these is commonly called *molecular dynamics*.

2.2 Density Functional Theory

As we saw in last section, when we use the Born-Oppenheimer approximation, the first step is always to consider a configuration of the average nuclear positions and the solve the electronic N-body Schrödinger equation, Equation (2.8). Even though this is already a drastic simplification of the full atomic problem, the equation constitutes a second order partial differential equation in $3N_e$ variables which does only have closed form solutions in a few special systems with a small number of electrons. It is in principle possible to directly solve the equation numerically to any precision by directly diagonalizing the Hamiltonian in chosen basis, but the scaling of this method with the number of electrons makes this approach intangible for any but the smallest molecules or atoms. One of the first successful methods for solving the many-electron Schrödinger equation is the Hartree-Fock approximation[34, 20], where the the wave function is restricted to consist of a single Slater-determinant of single particle wave functions. The result is a variational procedure where the single particle wave functions are optimized until self-consistensy. While this approach has proven to be quite accurate for atoms and small molecules, it computational cost it quite high and it does not have the same success for larger molecules and solids, for which it underestimates the correlation energy.

In 1964 Hohenberg and Kohn[41] noted that the external potential including the nuclear potentials completely determines the many-electron Hamiltonian and thus also completely determines the solutions. These on the other hand also completely determine the electron density $n(\mathbf{r})$, which is a local quantity that only depends on a single coordinate. They then showed that the opposite is also true; that there is a one to one correspondence between the external potential and the ground state density. This observation lead to the realization that all properties of the system can be expressed as functionals of the ground state density. This has lead to the formulation of Density Functional Theory that is a computational scheme that instead of dealing with the complicated many-body wave function uses the electron density as the basic ingredient.

2.2.1 The Hohenberg-Kohm theorems

The basis for Density Functional Theory are two statements that are now known as the Hohenberg-Kohm theorems. The first one says:

> The external potential $v_{\text{ext}}(\mathbf{r})$ is a unique functional of the density $n(\mathbf{r})$, apart from a trivial additive constant.

The proof is based on showing that two different external potentials that are assumed to yield the same ground state densities will lead to a contradiction and therefore the theorem has to be true.

Now due to previous theorem the ground state is uniquely determined by its density $n(\mathbf{r})$, which means that it can be considered a functional of this density $|\Psi[n]\rangle$. Given an external potential $v_{\text{ext}}(\mathbf{r})$ we now define the energy functional as

$$E_v[n] = F[n] + \int n(\mathbf{r}) v_{\text{ext}}(\mathbf{r})\, \mathrm{d}\mathbf{r}, \tag{2.12}$$

where

$$F[n] = \langle \Psi[n] | \hat{T} + \hat{V}_{e-e} | \Psi[n] \rangle. \tag{2.13}$$

One can then show that the energy functional, Equation (2.12), is minimized for the ground state density $n_0(\mathbf{r})$ and that the energy is the ground state energy,

$$E_0 = \min_n E_v[n] \quad \leftrightarrow \quad n(\mathbf{r}) = n_0(\mathbf{r}). \tag{2.14}$$

Since the functional $F[n]$ does not explicitly depend on the external potential but only on the density through the kinetic energy and the electron-electron interactions it is in principle possible to define it independently of the system. Thus finding the ground state density and energy will then only be a matter of minimizing the energy functional, Equation (2.12). In order to devise explicit methods for minimizing the energy functional and finding the ground state density we note that minimization requirement, Equation (2.14), can be restated using the variational principle,

$$\delta \left\{ E_v[n] - \mu \left(n(\mathbf{r})\, \mathrm{d}\mathbf{r} - N \right) \right\} = 0, \tag{2.15}$$

where μ is a Lagrange multiplier introduced to contrain the minimization to densities which have the right number of electrons, $\int n(\mathbf{r})\,\mathrm{d}\mathbf{r} = N$. This is equivalent to the statement that the functional derivative should be zero and we therefore have

$$\frac{\delta E_v}{\delta n(\mathbf{r})} = \frac{\delta F}{\delta n(\mathbf{r})} + v_{\mathrm{ext}}(\mathbf{r}) = \mu. \tag{2.16}$$

Doing the minimization with respect to all densities does not necessarily lead to the ground state density of the real system, because their are no restrictions on the density except that it should give the correct number of electrons, $\int n(\mathbf{r})\,\mathrm{d}\mathbf{r}$. However, all such densities are not physically sound, since for a true system it should be possible to contruct the density from an antisymmetric wave function. The search should therefore be done only within densities that fulfill this requirement[69], also called *constrained search*.

2.2.2 Kohn-Sham equations

The Hohenberg-Kohn theorems shows that it is possible to reduce the problem of finding the $3N$ dimensional many-electron wave function $\Psi(\{\mathbf{r}_i\})$ that minimizes $\langle\Psi\rangle\hat{H}|\Psi\rangle$ to what seems to be much simpler problem of finding the 3-dimensional density $n(\mathbf{r})$ that minimizes the functional $E_v[n]$. But while the solution of the original many-body Schrödinger equation can at least be done straightforwardly, the other approach is somewhat mysterious since we have not yet specified the exact functional form of $F[n]$ or provided any way of constructing it -- we just know that it in principle exists. A crucial step forward came in 1965 when Kohn and Sham showed that a practical scheme for computing $F[n]$ can be constructed[54] and what came to be known as Density Functional Theory was born.

The idea is to consider an *auxilary* system consisting of non-interaction particles subject to an *effective* external potential $v_{\mathrm{eff}}(\mathbf{r})$ *such that the ground state density is equal to the ground state density of the fully interacting system.* The Hamiltonian for this system is simply

$$\hat{H}_{\mathrm{aux}} = \sum_{i=1}^{N}\left[-\frac{1}{2}\nabla_i^2 + v_{\mathrm{eff}}(\mathbf{r}_i)\right]. \tag{2.17}$$

Since there are no electron-electron interactions, $\hat{V}_{e-e} = 0$, the N-electron ground state wave function consists of a single Slater determinant of single particle wave functions $\psi_i(\mathbf{r})$ found from solving the corresponding single-particle Hamiltonian called the Kohn-Sham equation,

$$\left[-\frac{1}{2}\nabla^2 + v_{\mathrm{eff}}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}). \tag{2.18}$$

The ground state density for such a non-interacting system can be written

$$n(\mathbf{r}) = \sum_{i}^{\mathrm{occ}} |\psi_i(\mathbf{r})|^2, \tag{2.19}$$

and the functional $F[n]$ reduces to the kinetic energy functional given by

$$F_{\text{aux}}[n] = T_{\text{aux}}[n] = \langle \Psi[n] | \hat{T} | \Psi[n] \rangle = -\frac{1}{2} \sum_{i=1}^{\text{occ}} \int \psi_i(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) \, \mathrm{d}\mathbf{r}, \tag{2.20}$$

where the single particle wave functions are implicitly functionals of the density. We now choose to redfine the density functional of the original interacting system in terms of the kinetic energy of the auxilary system,

$$F[n] = T_{\text{aux}}[n] + \frac{1}{2} \int \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}\mathbf{r} \, \mathrm{d}\mathbf{r}' + E_{\text{xc}}[n], \tag{2.21}$$

where the second term is the Hartree energy, i.e. the electrastatic energy from a classical charge distribution given by $n(\mathbf{r})$, and the last term, $E_{\text{xc}}[n]$ is a new functional that contains all the unknown contributions and is called the exchange-correlation energy functional. While this separation may seem arbitrary it has turned out to be quite clever. The reason is that two first terms, the kinetic energy and the electrostatic energy, most often contains by far the largest contribution to the ground state energy and errors in the approximation of $E_{\text{xc}}[n]$ are then guaranteed to be relatively small. It should also be noted that the wave functions $\psi_i(\mathbf{r})$ of the auxilary system do not have any physical interpretation as real orbitals of the system and the kinetic energy from $T_{\text{aux}}[n]$ is not the kinetic energy of the real system, which would have to be calculated from the real many-body wave function. What has been done now is effectively moving all the complicated many-body interactions from the functional $F[n]$ of the original system into the as of yet unknown functional $E_{\text{xc}}[n]$ which should however be of smaller size. To make any further progress one would have to come up with ways of approximating the exchange-correlation functional. If we can find such an expression the Euler-Lagrance equation for the energy functional, Equation (2.16), takes the form

$$\frac{\delta T_{\text{aux}}[n]}{\delta n(\mathbf{r})} + v_{\text{ext}}(\mathbf{r}) + \int \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}\mathbf{r}' + \frac{\delta E_{\text{xc}}[n]}{\delta n(\mathbf{r})} = \mu. \tag{2.22}$$

Similarly the Euler-Lagrange equation for the auxilary system is just

$$\frac{\delta T_{\text{aux}}[n]}{\delta n(\mathbf{r})} + v_{\text{eff}}(\mathbf{r}) = \mu_{\text{aux}}. \tag{2.23}$$

Since the interacting system and the auxilary system have the same ground state density and ground state energy, their chemical potential should be equal $\mu = \mu_{\text{aux}}$. We can then insert Equation (2.23) into Equation (2.22) and we see that the effective potential is given by

$$v_{\text{eff}}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + v_H[n](\mathbf{r}) + v_{\text{xc}}[n](\mathbf{r}), \tag{2.24}$$

where

$$v_H[n](\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}\mathbf{r}' \tag{2.25}$$

is the Hartree potential corresponding to the classical repulsive potential from the electron density and

$$v_{\mathrm{xc}}[n](\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}[n]}{\delta n(\mathbf{r})} \tag{2.26}$$

is called the exchange-correlation potential. The problem of solving a complicated many-electron Schrödinger equation has now been reduced to solving a much simpler single-electron Schrödinger equation, Equation (2.18), of an auxilary system. However, the problem requires us to know the effective potential, Equation (2.24), which in turn depends on the solutions to the single-particle Schrödinger equation through the density, Equation (2.19). Thus the solution has to be done iteratively by starting with a guess of the real ground state density, calculating the effective potential, solving the Kohn-Sham Schrödinger equation, calculating a new density, etc. This should be repeated until self-consistency, i.e. until the input density equals the output density.

2.2.3 Exchange-Correlation functionals

The formulation of Density Functional Theory so far has been *exact*, but it relies on the existence of an exact functional $E_{\mathrm{xc}}[n]$, which includes all the exchange and correlation contributions to the energy from the real many-body wave function. This functional is not known and in order to make the procedure practical one has to find approximations for this functional that are both relatively easy to evaluate and yield accurate result for a wide variety of systems. In order to make any approximations to this functional it is worth examining where this energy contribution really comes from. First we note that the Hartree-Fock approximation can be formulated in a way similar to DFT. In Hartree-Fock the many-body wave function is approximated by a single Slater determinant and the kinetic energy is therefore directly given by Equation (2.20) and the total energy can be written

$$E_{\mathrm{HF}} = T + V_{\mathrm{ext}} + E_H + E_{\mathrm{exx}}, \tag{2.27}$$

where

$$E_H = \int n(\mathbf{r}) v_H(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{2.28}$$

is the Hartree energy and

$$E_{\mathrm{exx}} = \frac{1}{2} \sum_i^{\mathrm{occ}} \sum_j^{\mathrm{occ}} \int \int \frac{\psi_i^*(\mathbf{r}) \psi_j^*(\mathbf{r}') \psi_i(\mathbf{r}') \psi_j(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} \tag{2.29}$$

is the Fock exchange energy also called *exact exchange*. If one considers the exact exchange energy as a functional of the density then HF-theory can be thought of as an approximation to DFT with $E_{\mathrm{xc}}[n] = E_{\mathrm{exx}}$. This tells us that we could in pratice split the exchange-correlation functional into two parts: $E_{\mathrm{xc}} = E_x + E_c$, where the exchange part could be calculated exactly as in Hartree-Fock and the correlation part

is left to approximate. However, the real many-body wave function is a superposition of Slater determinants so it is clear that the exact functional $E_{\mathrm{xc}}[n]$ should somehow include the exchange contributions from all Slater determinants. The error we make by assuming a single Slater determinant has to exactly be included in the correlation contribution and it is thus clear that the exact exchange energy is not guaranteed to be a better approximation to the xc-functional than the correlation-functional. Thus it should be best to assume that exchange and correlation energies are of similar magnitude and find an approximation to the combined xc-functional.

The first approximation for the xc-functional was described in the original paper by Kohn and Sham and is now known as the Local Density Approximation (LDA)[54]. In this approximation $E_{\mathrm{xc}}[n]$ is taken as the exchange and correlation energy of a homogeneous electron gas with density $n(\mathbf{r})$. The homogeneous electron gas, as its name implies, is homogeneous so the density is constant. So in order to contruct an xc-functional that gives an estimate of the exchange-correlation energy even for systems with spacially varying densities we instead consider the system as *locally* having the exchange-correlation energy of a homogeneous electron gas with a density given by the local density $n(\mathbf{r})$. Thus the functional may be written

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[n] = \int n(\mathbf{r})\epsilon_{\mathrm{xc}}^{\mathrm{HEG}}\big(n(\mathbf{r})\big)\,\mathrm{d}\mathbf{r}, \tag{2.30}$$

where $\epsilon_{\mathrm{xc}}^{\mathrm{HEG}}(n)$ is the exchange-correlation energy *density* of the homogeneous electron gas with a uniform density n. Thus it is just a simply function of the loca density and not a functional. The homogeneous electron gas exchange-correlation energy density is known to within numerical accuracy so the LDA functional is exact for the homogeneous electron and is expected to be a good approximation for system that are similar to a homogeneous electron, i.e. where the electron density varies slowly in space. Even though the LDA is assumed only to be good for systems with near homogeneous density is has been quite successful for calculating many properties for many, even quite inhomogeneous, crystalline solid state system. Molecular systems are another story. They do not quite fit the homogeneous electron gas description and the LDA has never really given the accuracy needed to predict chemical properties satisfactory.

An extension of the approach used in the LDA to improve its accuracy for more inhomogeneous densities have been provided by Generalized Gradient Approximations[62, 8, 93, 95]. The idea is to consider the exchange-correlation energy to be equal to the one given by LDA in case of a homogeneous density but then include the effect of inhomogeneities through an expansion in the gradient of the density,

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[n] = \int n(\mathbf{r})\epsilon_{\mathrm{xc}}^{\mathrm{HEG}}\big(n(\mathbf{r})\big)F_{\mathrm{xc}}\big(n(\mathbf{r}), \nabla n(\mathbf{r}), \nabla^2 n(\mathbf{r}), \dots\big). \tag{2.31}$$

where F_{xc} is the so-called enhancement factor that modifies the local LDA contribution to the energy. Attempts at doing a systematic expansion the the density gradients leads to divergent terms that are only cancelled by going to infinite order

in the expansion[74]. Hoever, it was shown that one could still construct functionals of the type Equation (2.31) with a finite order of the gradient, requiring that the functional satisfies some physically motivated requirements. Since the any such construction is not unique this has lead to a whole family of functionals known as Generalized Gradient Approximations. A particular simple one, now called PBE, was provided by Perdew, Burke and Ernzerhof[91] and it has proven quite successful in calculating many properties, most notably lattice constants.

2.3 Charged states and the band gap problem

A condensed matter system has a rich spectrum of excitations that give all the different materials their unique properties when interacting with an external pertubation. One kind of excitation is the addition or removal of an electron, whose dynamics are directly related to the electronic transport properties of the material. Both for solid state systems as well as molecules it is often of interest to know the energy to add an electron to the system, the so-called electron affinity (EA), $E_\mu^+ = E_{N+1,\mu} - E_{N,0}$, as well as the energy it requires to remove an electron from the system, the ionization potential (IP), $E_\nu^- = E_{N-1,\nu} - E_{N,0}$. An equally important quantity is the *band gap*, defined as the smallest energy difference between a state with an electron added and a state with an electron removed,

$$E_{\text{gap}} = \min_{\mu,\nu}\{E_\mu^+ - E_\nu^-\} = \min_{\mu,\nu}\{E_{N+1,\mu} - E_{N-1,\nu}\}. \tag{2.32}$$

So the big question arises: How do we determine these values? The first attempts at solving this problem within atomic physics was by the use of Hartree-Fock theory where the ground state is approximated as a single Slater determinant and electron-electron interactions are included by a self-consistently determined the mean-field Hartree potential. The outcome of a self-consistent Hartree-Fock calculation, like DFT, is a set of eigenfunction $\psi_i(\mathbf{r})$ and eigenvalues ϵ_i, which can be interpretated as molecular orbitals and corresponding energies. The ground state energy is then constructed as

$$E_N^{\text{HF}} = \sum_{n=0}^{N} \epsilon_n^{\text{HF}}, \tag{2.33}$$

where ϵ_n^{HF} is the n'th lowest eigenvalue. In the *frozen orbital approximation* we assume that the Hartree potential does not change much upon addition or removal of an electron so the electron affinity and ionization potential corresponds to the energy of the lowest unoccupied molecular orbital (LUMO) and minus the energy of the highest occupied molecular orbital (HOMO), i.e.

$$E_{\text{EA}}^{\text{HF}} = E_{N+1}^{\text{HF}} - E_N^{\text{HF}} = \epsilon_{N+1}^{\text{HF}} \tag{2.34}$$

$$E_{\text{IP}}^{\text{HF}} = E_{N-1}^{\text{HF}} - E_N^{\text{HF}} = -\epsilon_N^{\text{HF}}, \tag{2.35}$$

which is the well-known Koopman's theorem[55]. One can also consider a generalized version of Koopman's theorem where the energies of the higher lying molecular orbitals correspond to higher energy charged excitations, so that a whole spectrum can be obtained from the eigenvalues. Koopman's theorem was initially a success and has been used for a long time to estimate IP's and AE's for molecules, but it is nevertheless a pretty rough approximation. Firstly, the frozen orbital approximation may not be valid for many systems as we would imagine that there would always be some relaxation of the molecular orbitals when charge is added or removed. Moreover, the Hartree-Fock theory assumes a single Slater determinant ground state and completely neglect electron correlation contributions to the energy. The first problem can be alleviated by calculating the self-consistent HF ground state energy of charged states directly in which case there is no need to use Koopman's theorem. While exact within HF theory this method is still leaving out possibly important electron correlations and is restricted to only determine the lowest energy charged excitations. Furthermore it turns out that while HF theory and Koopman's theorem works quite well for small molecules they fail quite horribly when applied to extended condensed matter systems. The reason is that for solid state systems the electron density is very delocalized and the correlation and exchange energies are of more similar magnitude, which means that we can't neglect the electron correlations. Density Functional Theory, which given the right density functional, is formally exact in determining the ground state density and energy may thus be a better way to determine the EA and IP than Hartree-Fock. This means that *exact* DFT will find the correct EA and IP when calculating the ground state energies of a system with N, $N-1$ and $N+1$ electrons. No matter how much we wish for it, we still don't have a closed expression for the exact density functional and we are therefore forced to use approximate functionals like the Local Density Approximation, Generalized Gradient approximations or any of the many other functionals that now exist. Using DFT we may like in HF theory try and apply Koopman's theorem to predict the EA and IP, but it turns out that even though that DFT includes electron correlation contributions to the energy for molecules it most often performs worse than HF[97, 32]. The reason is that contrary to HF theory which directly includes the exchange energy, DFT functionals often treat approximations to the exchange-correlation energy as a whole and thus may underestimate the exchange energy. This means for instance that one introduces spurious self-interaction errors[90], where the density from a single electron leads to a finite exchange-correlation contributions due to the functional being only dependent on the absolute electron density and not the total number of electron. However, DFT is still regularly being used to estimate EA's and IP's even though its limitations are obvious. Furthermore, a lot of effort has been invested in the development of new functionals that better overcome these problems, e.g. hybrid functionals that include a certain part of exact exchange energy.

Like with HF theory one can use DFT to directly find the self-consistent ground-state energies of the charged states. This should be more accurate than using Koopman's theorem which uses the frozen orbitals of the neutral system. This method may yield quite accurate values of the ionization potential for many molecules, but

is more problematic for determining the electron affinity. The reason is that the self-interaction error inherent in many XC-functionals work to delocalize the electron density and in some cases the total effective potential turns out to be positive in a large enough region in space to make bound solutions to the Schodinger equation impossible[68, 51].

So far I have only considered isolated molecules which can be treated in a closed non-periodic unit cell. When trying to transfer the previous mentioned methods to extended solids described by periodic unit cell we run into a number of problems. Firstly, the total effective potential is taken to be periodic which means that it doesn't necessarily go to zero on the boundary. In fact the absolute level of the effective potential is often chosen such that it integrates to zero over the unit cell, but one can add any constant potential without changing the physics. This means that for bulk solids we can't determine the absolute values of the EA and IP, while relative quantities like the band gap do still make sense. In order to determine the absolute values of the EA and IP we would have to simulate a surface, so that the effective potential is allowed to attain its zero-value in vacuum. Now, instead of calculating the EA and IP directly we can consider their relative difference: the band gap. In periodic structures the solutions to the DFT effective single-particle Hamiltonian are Bloch waves with corresponding eigenvalues that yield a certain band structure. Blindly applying Koopmans' theorem we may consider these eigenvalues as the energies of excited states and we can then determine the band gap as the difference in energy between the conduction band minimum (CBM) and valence band maximum (VBM). The results are not convincing; typically the band gap is grossly underestimated[92]. One could think that the problem stems from the lack of relaxing the wave functions upon addition or removal of electrons, so one could try and do calculations for these cases separately. However it is not really possible do calculations on states $|N \pm 1\rangle$, because in an extended system N is strictly speaking infinite. By considering the system as open and connected to a resevoir with the contraint that $\int n(\mathbf{r})\,\mathrm{d}\mathbf{r} = N+\delta N$, where $n(\mathbf{r})$ is the electron density, N is the total number of electron and $0 \leq \delta N \leq 1$ is a fractional charge, the state of the system is a statistical mixture of $|N\rangle$ and $|N+1\rangle$. The lowest average energy is then[94]

$$E = (1 - \delta N)E_N + \delta N E_{N+1}, \tag{2.36}$$

where E_N and E_{N+1} are the ground state energies of the N and $N+1$ electron system, respectively. It is clear that if the system is gapped, the total energy in Equation (2.36) will be a continuous linear function in the fractional charge δN, with discontinuities at integer δN as seen in Figure 2.1. In DFT the ground state energy depends on the fractional charge through the electron density $n(\mathbf{r})$. Since all other terms in the expression for the total energy depend on the position of the density and not on the total number of electrons it can be shown[105, 92] that the discontinuity in the total energy can only come from the exchange-correlation energy functional

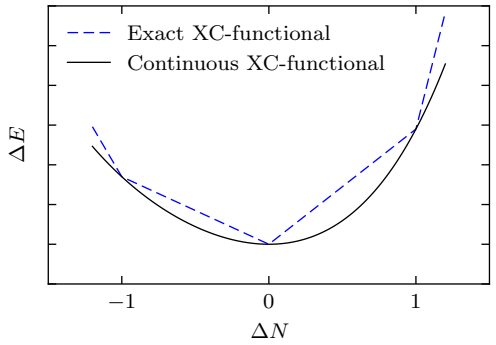


Figure 2.1: Example of the difference in groundstate energy as a function of fractional added charge ΔN for exact DFT and with a functional that it continuous in the electron density.

$E_{\mathrm{xc}}[n(\mathbf{r})]$. In fact we have

$$\left.\frac{\delta E_{\mathrm{xc}}}{\delta n(\mathbf{r})}\right|_{N+\delta} - \left.\frac{\delta E_{\mathrm{xc}}}{\delta n(\mathbf{r})}\right|_{N-\delta} = \Delta_{\mathrm{xc}}, \tag{2.37}$$

with Δ_{xc} normally just called the *derivative discontinuity*. Using Janak's theorem[47],

$$E_{N+1,i} - E_N = \int_0^1 \epsilon_i(f)\,\mathrm{d}f, \tag{2.38}$$

where $\epsilon_i(f)$ is the eigenvalue of the i'th state self-consistently determined with fractional occupancy f, the band gap, Equation (2.32), can be shown to be

$$E_{\mathrm{gap}} = \min_{\mu,\nu}\{\epsilon_{N+1,\mu} - \epsilon_{N,\nu}\} + \Delta_{\mathrm{xc}}. \tag{2.39}$$

Approximate XC-functionals like LDA and Generalized Gradient approximations are however continuous in total electron charge like shown in Figure 2.1 and therefore they do not have any derivative discontinuity, which then explains why DFT using these kind of functionals underestimate the band gap. In order to get a good estimate of the EA, IP and the band gap it is therefore necessary to find a way to calculate the derivative discontinuity or use a completely different method like from many-body pertubation theory.

2.3.1 Derivative discontinuity with the GLLB-SC functional

It is possible to model approximate XC-functionals that also include an estimate of the derivative discontinuity. A general idea is to relax the contraint that the XC-functional whould be fully determined by the electron density $n(\mathbf{r})$ alone and instead

seek to contruct an optimized local effective potential (OEP) by solving an integral equation that depends on the orbital wavefunctions $\psi_(\mathbf{r})$ their occupation numbers f_i and eigenvalues ϵ_i. The earliest such methods only aimed at determining an exchange potential[106, 114, 57] but it has later been extended to also include approximations to the correlation potential[29, 13, 56]. In this project I have used the so-called GLLB-SC functional[59] that is baed on the GLLB functional[30] which models the KLI[57] exchange potential so that the computationally expensive evaluation of an integral equation is avoided. In GLLB-SC the total effective potential is approximated as

$$v_{\text{GLLB-SC}}(\mathbf{r}) = 2\epsilon_{\text{XC}}^{\text{PBEsol}}(\mathbf{r}) + v_{\text{c,resp}}^{\text{PBEsol}}(\mathbf{r}) + \sum_{i}^{\text{occ.}} K_x \sqrt{\epsilon_r - \epsilon_i} \frac{|\psi_i(\mathbf{r})|^2}{n(\mathbf{r})}, \tag{2.40}$$

where $\epsilon_{\text{XC}}^{\text{PBEsol}}(\mathbf{r})$ is the exchange-correlation energy density and $v_{\text{c,resp}}^{\text{PBEsol}}(\mathbf{r}) = v^{\text{PBEsol}}(\mathbf{r}) - \epsilon^{\text{PBEsol}}(\mathbf{r})$ is the linear correlation response potential, both from the PBEsol XC-functional[118]. In this formula $K_x \approx 0.382\,\text{a u}$ is a prefactor of unit energy$^{1/2}$ found from the homogeneous electron gas and ϵ_r is a reference eigenvalue taken as that corresponding to the highest occupied orbital, $\epsilon_r = \epsilon_{\text{HOMO}}$. One of the great successes of this method is that it gives a simple expression for the derivative discontinuity,

$$\Delta_{\text{xc}} \approx \sum_{i}^{N} K_x \left(\sqrt{\epsilon_{N+1} - \epsilon_i} - \sqrt{\epsilon_N - \epsilon_i} \right) \int \frac{|\psi_{N+1}(\mathbf{r})|^2 |\psi_i(\mathbf{r})|^2}{n(\mathbf{r})}, \tag{2.41}$$

which can be straightforwardly implemented.

CHAPTER 3

Many-Body Pertubation Theory

In this chapter I will briefly go through the basic concepts of many-body pertubation theory and show how this can be used to calculate electronic and optical properties of a system when it is excited by an external potential or electrons are added or removed. It will be shown how the linear response of the system can be described by correlation and response functions trough use of the Kubo formula and the polarizability and dielectric function will be introduced. I will introduce the concept of the Green's function and introduce the concept of self-energy and show how this can be calculated from Hedins *GW* theory. Finally I will show how the *GW* theory can be used to calculate energies of charged excited states, which turned out to be difficult task to do within DFT.

The concept of many-body pertuabtion theory is best described in the second quantization formulation of quantum mechanics. In this a N-electron state $|\Phi_N\rangle$ is given in terms of the occupation numbers of single-particle states

$$|\Phi_N\rangle = |\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\rangle. \tag{3.1}$$

The quantum field operator $\hat{\Psi}^\dagger(\mathbf{r})$ creates an electron in position state $\mathbf{r}$,

$$\hat{\Psi}^\dagger(\mathbf{r})|\mathbf{r}_1, \ldots, \mathbf{r}_N\rangle = |\mathbf{r}_1, \ldots, \mathbf{r}, \ldots, \mathbf{r}_N\rangle \tag{3.2}$$

and its adjoint $\hat{\Psi}(\mathbf{r})$ annihilates an electron at position $\mathbf{r}$,

$$\hat{\Psi}(\mathbf{r})|\mathbf{r}_1, \ldots, \mathbf{r}, \ldots, \mathbf{r}_N\rangle = |\mathbf{r}_1, \ldots, \mathbf{r}_N\rangle. \tag{3.3}$$

The quantum field operators satisfy the canonical anti-commutation relations

$$\left\{\hat{\Psi}^\dagger(\mathbf{r}), \hat{\Psi}(\mathbf{r}')\right\} = \delta(\mathbf{r} - \mathbf{r}'), \quad \left\{\hat{\Psi}(\mathbf{r}), \hat{\Psi}(\mathbf{r}')\right\} = 0, \tag{3.4}$$

The vacuum state $|0\rangle$ is defined by $\hat{\Psi}(\mathbf{r})|0\rangle = 0 \ \forall \ \mathbf{r}$ and a general N-body can be therefore be created from the vacuum by application of several creation operators,

$$|\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\rangle = \hat{\Psi}^\dagger(\mathbf{r}_N) \cdots \hat{\Psi}^\dagger(\mathbf{r}_2)\hat{\Psi}^\dagger(\mathbf{r}_1)|0\rangle. \tag{3.5}$$

The anticommutation relations, Equation (3.4), ensures that two electrons cannot exist in the same state. The many-body wave function can be written

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \langle 0|\hat{\Psi}(\mathbf{r}_N)\cdots\hat{\Psi}(\mathbf{r}_2)\hat{\Psi}(\mathbf{r}_1)|N, \Psi\rangle, \tag{3.6}$$

and like-wise, the anti-commutation relations ensures that the wave function is anti-symmetric upon interchange of two coordinates.

Using the quantum field operators the many-body electronic Hamiltonian introduced in Equation (2.10) can be written

$$\hat{H} = \hat{h}_0 + \hat{v}_{e-e}, \tag{3.7}$$

where the non-interacting Hamiltonian is given by

$$\hat{h}_0 = \int \hat{\Psi}^\dagger(\mathbf{r}) \left[-\frac{1}{2}\nabla^2 + v_{\text{ext}}(\mathbf{r})\right] \hat{\Psi}(\mathbf{r})\, \mathrm{d}\mathbf{r} \tag{3.8}$$

and the electron interactions are governed by

$$\hat{v}_{e-e} = \iint \frac{\hat{\Psi}^\dagger(\mathbf{r})\hat{\Psi}^\dagger(\mathbf{r}')\hat{\Psi}(\mathbf{r}')\hat{\Psi}(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}'. \tag{3.9}$$

This methos makes is easy to treat a general system with any number of electrons, as the total number only enters through the states $|N, i\rangle$ and at the same time it ensures the right particle statistics when developing pertubation theory.

3.1 Linear response and correlation functions

We are often interested in knowing how a system responds when we interact with it.

Consider a system in equilibrium described by a time-independent Hamiltonian H_0 and assume that it is in its ground state $|\Psi_0\rangle$, which is known. Know we want to instead study the system under influence of an external time-dependent potential $V(t)$, so that the total Hamiltonian is

$$H(t) = H_0 + V(t). \tag{3.10}$$

When $V(t)$ is small compared to H_0 we usually deal with this problem in the interaction picture where the states and operators evolve in time according to

$$\begin{aligned} |\psi_{\text{I}}(t)\rangle &= e^{iH_0 t}|\psi_{\text{S}}(t)\rangle \\ \hat{A}_{\text{I}}(t) &= e^{iH_0 t}\hat{A}e^{-iH_0 t}, \end{aligned} \tag{3.11}$$

where $|\psi_{\text{S}}(t)\rangle = e^{-iHt}|\psi(t_0)\rangle$ is the time-dependent state in the Schrödinger picture. In the interaction picture it is seen that the time-dependent Schrödinger equation can simply be written

$$i\frac{\partial}{\partial t}|\psi_{\text{I}}(t)\rangle = \hat{V}(t)|\psi_{\text{I}}(t)\rangle. \tag{3.12}$$

Since the Hamiltonian is the generator of small time-evolutions knowing the state at t_0 we can find the state at time t by

$$|\psi_{\mathrm{I}}(t)\rangle = \hat{U}(t,t_0)|\psi_{\mathrm{I}}(t_0)\rangle, \tag{3.13}$$

where $\hat{U}(t,t')$ is the unitary time evolution operator. From Eq. (3.12) we see that the time-evolution operator is governed by the differential equation

$$i\frac{\partial}{\partial t}\hat{U}(t,t_0) = \hat{V}(t)\hat{U}(t,t_0), \tag{3.14}$$

which with the obvious condition $\hat{U}(t_0,t_0) = 1$ can be solved by integration to yield the integral equation

$$\hat{U}(t,t_0) = 1 + \frac{1}{i}\int_{t_0}^{t} \mathrm{d}t'\hat{V}(t')\hat{U}(t',t_0). \tag{3.15}$$

This can rarely be solved directly but inserting it into itself we can solve it iteratively to obtain

$$\begin{aligned}\hat{U}(t,t_0) &= 1 + \frac{1}{i}\int_{t_0}^{t}\mathrm{d}t_1\hat{V}(t_1) + \frac{1}{i^2}\int_{t_0}^{t}\mathrm{d}t_1\hat{V}(t_1)\int_{t_0}^{t_1}\mathrm{d}t_2\hat{V}(t_2) + \cdots \\ &= \mathrm{T}_t\left\{e^{-i\int_{t_0}^{t}\mathrm{d}t'\hat{V}(t')}\right\},\end{aligned} \tag{3.16}$$

where $\mathrm{T}_t\{\cdot\}$ is time-ordering operator that orders products of operators so that those that are taken at the latest times are at the left. This expression is much easier to deal with as it can be approximated by systematic expansion of the exponential function.

Now if $\langle A\rangle$ was the expectation value of the observable A without the external potential we are now interested knowing how the external potential changes the expectation value after the pertubation has been turned on. Using the time-evolution in the interaction picture just described we have

$$\langle\hat{A}(t)\rangle = \langle\psi_I(t)|\hat{A}_I(t)|\psi_I(t)\rangle = \langle\psi_I(t_0)|\hat{U}^\dagger(t,t_0)\hat{A}_I(t)\hat{U}(t,t_0)|\psi_I(t_0)\rangle. \tag{3.17}$$

Assuming the pertubation to be small we may insert the expression for the time-evolution operator, Equation (3.16), and expand to first order, $\hat{U}(t,t_0) \approx 1 - i\int_{t_0}^{t} V(t')$. Inserting this expression in Equation (3.17) and keeping only terms to linear order in $\hat{V}(t)$ we arrive at the Kubo formula[58]:

$$\langle A(t)\rangle \approx \langle\hat{A}\rangle_0 - i\int_{t_0}^{t}\langle[\hat{A}_I(t), V(t')]\rangle_0, \tag{3.18}$$

where $\langle\cdot\rangle_0 = \langle\psi_I(t_0)|\cdot|\psi_I(t_0)\rangle$ corresponds to the expectation value with respect to the state before the pertubation was applied, i.e. wrt. $\hat{H}_0$. The change in the expectation value may be written

$$\delta\langle\hat{A}(t)\rangle = \langle\hat{A}\rangle - \langle A\rangle_0 = \int_{t_0}^{\infty} C^r_{AV}(t,t')\,\mathrm{d}t', \tag{3.19}$$

where

$$C^r_{AV}(t,t') = -i\theta(t-t')\langle[\hat{A}(t),\hat{V}(t')]\rangle_0 \tag{3.20}$$

is called the *retarded correlation function.*

3.2 Polarizability and dielectric function

When considering an electronic system with density $n(\mathbf{r})$ and applying a small external potential $v_{\text{ext}}(\mathbf{r})$ electrons will move around and the system will become polarized so that the new total density is

$$n_{\text{tot}}(\mathbf{r}) = n_0(\mathbf{r},t) + n_{\text{ind}}(\mathbf{r},t), \tag{3.21}$$

where $n_{\text{ind}}(\mathbf{r},t)$ is the density induced by the external potential. The external potential can be though as leading to a small pertubation to the Hamiltonian:

$$\hat{V}(t) = \theta(t-t_0)\int \hat{n}(\mathbf{r})v_{\text{ext}}(\mathbf{r},t)\,\mathrm{d}\mathbf{r}, \tag{3.22}$$

where $\hat{n}(\mathbf{t})$ is the density operator and it is assumed that the pertubation has been adiabatically turned on in the infinite past, $t_0 \to -\infty$. Using the Kubo formula we the see that the induced density $n_{\text{ind}}(\mathbf{r},t)$ may be written

$$n_{\text{ind}}(\mathbf{r},t) = \int\int_{t_0}^{\infty} \chi^r(\mathbf{r},t;\mathbf{r}',t')v_{\text{ext}}(\mathbf{r}',t')\,\mathrm{d}t'\,\mathrm{d}\mathbf{r}', \tag{3.23}$$

where $\chi^r(\mathbf{r},t;\mathbf{r}',t')$ is the retarded density-density response function and is given by

$$\chi^r(\mathbf{r},t;\mathbf{r}',t') = -i\theta(t-t')\langle[\hat{n}(\mathbf{r},t),\hat{n}(\mathbf{r}',t')]\rangle_0. \tag{3.24}$$

The density-density response function is also called the electric susceptibility, polarizability or simply the density response function and I will use these names interchangebly to complete the confusion.

If the external potential oscillates periodically in time: $v_{\text{ext}}(\mathbf{r},r) = v_{\text{ext}}(\mathbf{r},\omega)e^(i\omega t)$, the polarizability will only depend on relative time $t-t'$ and it is therefore simpler to work with its Fourier transform $\chi^r(\mathbf{r},\mathbf{r}',\omega)$. In case of a periodic non-interacting system, the ground state will be a single Slater determinant of Bloch wave functions $\psi_{n\mathbf{k}}(\mathbf{r})$ and using the Kubo formula one can show that the corresponding non-interacting polarizability is given by

$$\chi^0(\mathbf{r},\mathbf{r}';\omega) = \sum_{\mathbf{k},\mathbf{k}'}^{\text{BZ}}\sum_{n,n'}(f_{n\mathbf{k}} - f_{n'\mathbf{k}'})\frac{\psi^*_{n\mathbf{k}}(\mathbf{r})\psi_{n'\mathbf{k}'}(\mathbf{r})\psi_{n'\mathbf{k}'}(\mathbf{r}')\psi^*_{n\mathbf{k}}(\mathbf{r}')}{\omega+\epsilon_{n\mathbf{k}}-\epsilon_{n'\mathbf{k}'}+i\eta}, \tag{3.25}$$

where $f_{n\mathbf{k}}$ is the occupation number of the state $(n,\mathbf{k})$ with corresponding energy $\epsilon_{n\mathbf{k}}$ and $\eta \to 0^+$ is a small positive number that should eventually be taken to 0.

Another important and related correlation function is the dielectric function which relates the external potential to the total potential

$$v_{\text{ext}}(\mathbf{r},t) = \int\int \varepsilon(\mathbf{r},t;\mathbf{r}',t')v_{\text{tot}}(\mathbf{r}',t')\,\mathrm{d}t'\,\mathrm{d}\mathbf{r}'. \tag{3.26}$$

Noting that the total potential may be written as the Coulomb potential from the total charge density plus the external potential,

$$v_{\text{tot}}(\mathbf{r},t) = \int v(\mathbf{r},t;\mathbf{r}',t')n_{\text{tot}}(\mathbf{r}',t')\,\mathrm{d}t'\,\mathrm{d}\mathbf{r}' + v_{\text{ext}}(\mathbf{r},t), \tag{3.27}$$

we see that the inverse dielectric function $\varepsilon^{-1}(\mathbf{r},t;\mathbf{r},t')$ may be found from a functional derivative and written in terms of the polarizability defined in Equation (3.23):

$$\begin{aligned}\varepsilon^{-1}(\mathbf{r},t;\mathbf{r}',t') = \frac{\delta v_{\text{tot}}(\mathbf{r},t)}{\delta v_{\text{ext}}(\mathbf{r}',t')} &= \delta(\mathbf{r}-\mathbf{r}')\delta(t-t') \\ &+ \iint v(\mathbf{r},t;\mathbf{r}'',\mathbf{t}'')\chi(\mathbf{r}'',t'';\mathbf{r}',t')\,\mathrm{d}t''\,\mathrm{d}\mathbf{r}''.\end{aligned} \tag{3.28}$$

This means that if the polarizability is known, the dielectric function can easily be found if the above integral an inversion can be performed.

3.3 Green's functions and the self-energy

We are now going to introduce the many-body Green's function which forms the basis of many-body pertubation theory. I am going to follow the notion of Hedin[38] and simplify the notation by denoting a space-time coordinate $(\mathbf{r}_1,t_1)$ by a number (1). The time-ordered Green's function is given by

$$G(1,2) = -i\langle N,0|\hat{T}\big\{\hat{\psi}(1)\hat{\psi}^\dagger(2)\big\}|N,0\rangle, \tag{3.29}$$

where $|N,0\rangle$ is the ground state of the N-electron Hamiltonian. The operator $\hat{T}$ is the time-ordering operator introduced in Equation (3.16), but one has to take into account that when two fermionic field operators are interchanged the result get a factor of -1,

$$\hat{T}\big\{\hat{\psi}(1)\hat{\psi}^\dagger(2)\big\} = \begin{cases}\hat{\psi}(1)\hat{\psi}^\dagger(2) & t_1 > t_2, \\ -\hat{\psi}^\dagger(2)\hat{\psi}(1) & t_2 > t_1.\end{cases} \tag{3.30}$$

The Green's function, Equation (3.29), is also called the propagator because it described the probability amplitude of an electron added at $(\mathbf{r}_2,t_2)$ and removed at $(\mathbf{r}_1,t_1)$. The Green's function is fundamental since it can be shown that the ground state expectation value of *any* single-particle operator can be determined from it. As an example the ground state electron density is given by

$$n(\mathbf{r}_1,t_1) = \langle N,0|\hat{\psi}(\mathbf{r}_1,t_1)\hat{\psi}(\mathbf{r}_1,t_1)|N,0\rangle = -iG(1,1^+), \tag{3.31}$$

where 1^+ means $t_1 + \delta$, where a infinitesimal positive time has been introduced in order to get the right time-ordering.

The starting point for developing a pertubative expansion expression for the Green's function is the equation of motion for the field operator

$$i\frac{\partial\hat{\psi}}{\partial t} = [\hat{\psi}, \hat{H}] = \hat{h}_0(\mathbf{r}_1)\hat{\psi}(1) + \int v(12)\hat{\psi}^\dagger(2)\hat{\psi}(2)\hat{\psi}(1)\,\mathrm{d}(2) \tag{3.32}$$

where $\hat{H}$ is the Hamiltonian given in Equation (3.7). We will now first consider the Green's function to the non-interacting system $G_0(12)$, which is seen to be governed by the equation of motion

$$\left[i\frac{\partial}{\partial t_1} - \hat{h}_0(\mathbf{r}_1)\right] G_0(12) = \delta(12). \tag{3.33}$$

The integral equation of this defines the inverse non-interacting Green's function $G_0^{11}(12)$, which may informally written

$$G_0^{-1}(31) = \left[i\frac{\partial}{\partial t_1} - \hat{h}_0(\mathbf{r}_1)\right]^{-1}. \tag{3.34}$$

Using Equation (3.32) we may similarly try to find an equation of motion for the full interacting Green's function, but differentiating with respect to time gives

$$\begin{aligned}\left[i\frac{\partial}{\partial t_1} - \hat{h}_0(\mathbf{r}_1)\right] G(12) &= \delta(12) \\ &\quad - i\int v(1^+3)\langle N,0|\hat{T}\{\hat{\psi}^\dagger(3)\hat{\psi}(3)\hat{\psi}(1)\hat{\psi}^\dagger(2)\}|N,0\rangle\,\mathrm{d}(3).\end{aligned} \tag{3.35}$$

The last term can be rewritten in terms of a so-called two-particle Green's function given by

$$G(1234) = (-i)^2\langle N,0|\hat{T}\{\hat{\psi}(1)\hat{\psi}(2)\hat{\psi}^\dagger(3)\hat{\psi}^\dagger(4)\}|N,0\rangle. \tag{3.36}$$

Identifying the square paranthesis on the left in Equation (3.35) as the inverse non-interacting Green's function we may act from the left with the non-interacting Green's function and integrate to get

$$G(12) = G_0(12) - i\iint G_0(13)v(3^+4)G(3424^+)\,\mathrm{d}(34) \tag{3.37}$$

In order to proceed further one could derive the equation of motion for the two-particle Green's function which would turn out to contain reference to a three-particle Green's function. One can continue in this way an generate an infinite series of coupled differential equations that all have to be solved in order to calculate the one-particle Green's function. Wick's theorem shows that this infinite series can be written solely in terms of combinations of the single-particle non-interacting Green's function. This

procedure is known as many-body pertubation theory and all the terms in the series can efficiently be evaluated using Feynman diagrams. One could in principle just truncate the sum and only retain a few terms, but as is well-known many of the terms turn out to diverge and the divergencies are only cancelled when going to infinite order.

Instead we now introduce the self-energy $\Sigma(12)$ by rewriting Equation (3.35) in the form of the so-called Dyson equation

$$G(12) = G_0(12) + \iint G_0(13)\Sigma(34)G(42)\,\mathrm{d}(34). \tag{3.38}$$

With this definition it is will be possible with an approximation to the self-energy to solve Equation (3.38) by direct inversion or self-consistently to obtain a closed expression for the full interacting Green's function. The self-energy also has a clear physical interpretation that will be mentioned later in the description of quasiparticles. In order to calculate the self-energy one has to compare Equation (3.38) with the pertubative expansion of Equation (3.35) whereby one can identify the self-energy in terms of Feynman diagrams as

$$\Sigma(34) \equiv \left\{\begin{matrix}\text{The sum of all irreducible diagrams in } G(12) \text{ without} \\ \text{the two external fermion lines } G_0(13) \text{ and } G_0(42)\end{matrix}\right\} \tag{3.39}$$

Still we have not given any expression of the self-energy or how to otherwise calculate it, but this will be described in the next section.

3.4 Hedin's equations and GW approximation

The problem with pertubation theory is that the bare Coulomb potential $v(\mathbf{r}, \mathbf{r}')$ is strongly repulsive and diverges when $\mathbf{r}_2 \to \mathbf{r}_1$. However, as we saw in section Section 3.2 the system will respond to an external field so as to screen it. This also means that it screens the Coulomb potentials from its constituent electrons, and therefore Hedin showed that[37] instead of doing pertubation theory in the base Coulomb potential it may be easier to work with the so-called *screened potential* given by

$$W(12) = \int v(13)\varepsilon^{-1}(32)\,\mathrm{d}(3). \tag{3.40}$$

The screened potential should have a less severe divergence and a series expansion in this is thus expected to converge much faster.

The method of Hedin is to introduce to the Hamiltonian, Equation (3.7), a small external potential $\phi(1)$ and then use functional derivatives to obtain the key quantitites. In the end the external potential will be taken to zero. We first note that the external potential will induce a density $n_{\mathrm{ind}}(2)$ which will lead to an induced potential $\phi_{\mathrm{ind}}(2)$. The total potential is thus $\phi_{\mathrm{tot}}(1) = \phi(1) + \phi_{\mathrm{ind}}(1)$. Next we note

that the polarizbility defined in Equation (3.23) may be written in terms of functional derivatives as

$$\chi(12) = \frac{\delta n(1)}{\delta \phi(2)} = -i\frac{\delta G(11^+)}{\delta \phi(2)}. \tag{3.41}$$

Similarly we now define the so-called *irreducible* polarizability as the change in the density due to the change in the total potential

$$P(12) = \frac{\delta n(1)}{\delta \phi_{\text{tot}}(2)} = -i\frac{\delta G(11^+)}{\delta \phi_{\text{tot}}(2)}. \tag{3.42}$$

One can show that the full polarizability and the reduced polarizability are related by a Dyson equation[10]

$$\chi(12) = P(12) + \iint P(13)v(34)\chi(52)\,\mathrm{d}(34). \tag{3.43}$$

From the definition of the dielectric function, Equation (3.26), we have

$$\varepsilon^{-1}(12) = \frac{\delta \phi_{\text{tot}}(1)}{\delta \phi(2)} = \delta(12) + \int v(13)\chi(32)\,\mathrm{d}(3) \tag{3.44}$$

and using the Dyson equation Equation (3.43) one can show that the screened potential may be written

$$W(12) = v(12) + \iint v(13)P(34)W(42)\,\mathrm{d}(34), \tag{3.45}$$

which again constitutes a new self-consistent equation. The underlying principle for finding an expression for the self-energy is the observation that the functional derivative of the Green's function with respect to the external field may be written[37]

$$\frac{\delta G(12)}{\delta \phi(3)} = G(12)G(33^+) - G(1323^+). \tag{3.46}$$

The first term contains the density $n(1) = -iG(11^+)$ and the second term is seen to be the two-particle Green's function that also appears in the equation of motion in Equation (3.35). We may thus rewrite Equation (3.35) in the form

$$\left[i\frac{\partial}{\partial t_1} - \hat{h}_0(\mathbf{r}_1)\right] G(12) = \delta(12) - iG(12)\int v(13)G(33^+)\,\mathrm{d}(3) \tag{3.47}$$

$$+ i\int v(13)\frac{\delta G(12)}{\delta \phi(3)}\,\mathrm{d}(3) \tag{3.48}$$

$$= \delta(12) + v_H(1)G(12) + \int \Sigma^{\text{xc}}(13)G(32)\,\mathrm{d}(3), \tag{3.49}$$

where by comparison with the equation of motion Equation (3.35) or similarly the Dyson equation in Equation (3.38) we have split the total selfenergy into a Hartree-term and an exchange-correlation self-energy:

$$\Sigma(12) = v_H(1)\delta(12) + \Sigma^{\text{xc}}(12). \tag{3.50}$$

While the Hartree potential can easily be obtained from the Green's function all the complicated correlation effects are hidden in the exchange-correlation self-energy that we still do not have an expression for. However, Hedin showed that one can construct a set of coupled equations that together can be solved iteratively to get the exchange-correlation energy. Today these equations are colloquially known has *Hedin's equations* and are given by

$$G(12) = G_0(12) + \int G_0(13)[v_H(3)\delta(34) + \Sigma^{\text{xc}}(34)]G(42)\,\mathrm{d}(34), \tag{3.51}$$

$$\Sigma^{\text{xc}}(12) = i\int W(1^+3)G(14)\Gamma(42;3)\,\mathrm{d}(34), \tag{3.52}$$

$$W(12) = v(12) + \int W(13)P(34)v(42)\,\mathrm{d}(34), \tag{3.53}$$

$$P(12) = -i\int G(23)\Gamma(34;1)G(42)\,\mathrm{d}(34), \tag{3.54}$$

$$\Gamma(12;3) = \delta(12)\delta(13) + \int \frac{\delta\Sigma^{\text{xc}}(12)}{\delta G(45)}G(46)G(75)\Gamma(67;3)\,\mathrm{d}(4567). \tag{3.55}$$

In principle these can be used to generate the infinite series that is standard in pertubation theory, but their true strength lie in the way the quantities are related. As is shown in Figure 3.1 when just a rough estimate for one of the quantities is known another quantity can be calculated from it at the same level of approximation and this can be repeated until all quantities are known. This will then lead to a better approximation for the first quantity and the circle can be repeated over until self-consistency. While this procedure sounds straightforward it is nevertheless complicated by the functional derivative that appears in equation for the so-called vertex function $\Gamma(12;3)$.

The simplest approximation to a solution to Hedins equations is the GW approximation where the obly the lowest order contribution to the vertex function is retained

$$\Gamma(12;3) \approx \delta(12)\delta(13). \tag{3.56}$$

This simple choice leads to an expression for the irreducible polarizability given by

$$P(12) = -iG(21^+)G(12), \tag{3.57}$$

which turns out to be exactly the same as that obtained by the Random Phase Approximation (RPA). Finally we see that the self-energy is

$$\Sigma^{\text{xc}}(12) = -iG(12)W(1^+2), \tag{3.58}$$

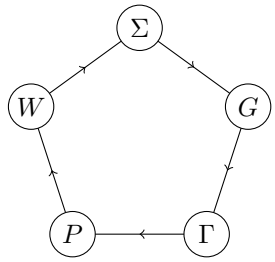


Figure 3.1: Diagram showing the recursive interrelation of the quantities in Hedin's equations.

from which the approximation has its name. Typically one starts by approximating the Green's function $G(12)$ by its non-interacting counterpart $G_0(12)$, which can usually be found exactly or within numerical accuracy. From this one can then calculate the corresponding non-interacting irreducible polarizability $P(12)$ and the thereby the screened potential $W_0(12)$ and one obtains the self-energy $\Sigma^{\text{xc}} = -iG_0W_0$. In principle one could iterate this procedure but it has turned out that doing so may not always be a good approach due to the neglect of higher order contributions in the vertex function $\Gamma(12;3)$ - so-called vertex-corrections.

3.5 Quasi-particles

While we have now introduced Green's functions and seen how they can be used to calculate expectation values of single particle observables and also higher order correlation function we still have not shown any practical scheme for doing so. In order to do so we note that the charged excitation energies may be written

$$\epsilon_i = \begin{cases} E_i^{N+1} - E_0^N & \epsilon_i > \mu \\ E_0^N - E_i^{N-1} & \epsilon_i > \mu, \end{cases} \tag{3.59}$$

where E_0^N is the ground state energy of the electron system with N electron, $E_i^{N\pm1}$ is the energy of the i'th excited state with $N \pm 1$ electrons and μ is the chemical potential. This notation makes it possible to treat both positively and negatively charged states at the same time. We note that by comparison to the definition of the electron affinity and ionization potential in Equations (2.34) and (2.35) we have

$$E_{\text{EA},i} = \epsilon_i, \epsilon_i > \mu \tag{3.60}$$

$$E_{\text{IP},i} = -\epsilon_i, \epsilon_i < \mu. \tag{3.61}$$

We now introduce the so-called Lehmann amplitudes or quasi-particle wave functions by

$$\psi_i(\mathbf{r}) \equiv \begin{cases} \langle N,0|\hat{\psi}(\mathbf{r})|N+1,i\rangle & \epsilon_i > \mu \\ \langle N-1,i|\hat{\psi}(\mathbf{r})|N,0\rangle & \epsilon_i < \mu. \end{cases} \tag{3.62}$$

By inserting the completeness relation $\sum_i |N \pm 1, i\rangle\langle N \pm 1, i|$ in definition of the time-ordered Green's function in Equation (3.29) and Fourier transforming to the frequency domain, noting that the result only depends on time differences $t - t'$, we get

$$G(\mathbf{r}, \mathbf{r}'; \omega) = \sum_i \frac{\psi_i(\mathbf{r})\psi^*(\mathbf{r}')}{\omega - \epsilon_i + i\eta\,\mathrm{sgn}(\epsilon_i - \mu)}, \tag{3.63}$$

where η is a small positive number that has been introduced to ensure that the Fourier transform converges should be taken to zero in the end. From Equation (3.63) it is clear that the full spectrum of the excitation energies of the charged states can be identified from the poles of the Green's function. Moreover, inserting of the Lehmann representation of the Greens function, Equation (3.63), in the equation of motion in Equation (3.49) yields an integral equation for the quasiparticle wave functions and energies, commonly called the quasiparticle equation:

$$\left[\hat{h}_0(\mathbf{r}) + v_H(\mathbf{r})\right]\psi_i(\mathbf{r}) + \int \Sigma^{\mathrm{xc}}(\mathbf{r}, \mathbf{r}'; \epsilon_i)\psi_i(\mathbf{r}')\,\mathrm{d}\mathbf{r}' = \epsilon_i\psi_i(\mathbf{r}). \tag{3.64}$$

The origin of the name ``quasiparticle equation'' should be clear: it resembles a Schrödinger equation for a single electron moving in the attractive nuclear potential (included as $v_{\mathrm{ext}}(\mathbf{r})$ in $\hat{h}_0(\mathbf{r})$) and the repulsive effective mean field of all the other electrons, $v_H(\mathbf{r})$, but also subject to a non-local potential $\Sigma^{\mathrm{xc}}(12)$ arising from the fact that the system is in fact a many-body system. However, while the exact many-body states $|N \pm 1, i\rangle$ are orthogonal, the quasiparticle wave functions $\psi_i(\mathbf{r}$ need not constitute an orthogonal set and the self-energy operator is in general not necessarily Hermitian. This means that the single-particle expectation value of the self-energy need not be a real number and in fact the imaginary part is related to the lifetime of such a state.

Normally we will use the Kohn-Sham wave functions and eigenvalues from a DFT calculation to initially approximate the quasiparticle wave functions and energies. In fact the Kohn-Sham equation, Equation (2.18), already constitutes a quasiparticle equation with a local Hermitian approximation to the self-energy given by $\Sigma^{\mathrm{xc}}_{\mathrm{DFT}}(\mathbf{r}, \mathbf{r}'; \omega) = v_{\mathrm{xc}}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r})$. While in Section 2.3 I described how Kohn-Sham eigenvalues obtained from typical DFT schemes fails to predict the right quasiparticle energies, it has been shown that the Kohn-Sham wave functions in general come quite close the quasiparticle wave functions[45, 44, 70]. This means that the self-energy should be more or less diagonal in a basis consisting of the Kohn-Sham wave functions and it should therefore be possible to use these as trial wave functions in the solution of Equation (3.64).

Using a Kohn-Sham wave function $\psi_i^{\mathrm{KS}}(\mathbf{r})$ and as a trial wave function, the expectation value of Equation (3.64) is

$$\epsilon_i^{\mathrm{QP}} = \epsilon_i^{\mathrm{KS}} - \langle\psi_i^{\mathrm{KS}}|v_{\mathrm{xc}}(\mathbf{r})|\psi_i^{\mathrm{KS}}\rangle + \langle\psi_i|\hat{\Sigma}^{\mathrm{xc}}(\epsilon_i^{\mathrm{QP}})|\psi_i^{\mathrm{KS}}\rangle. \tag{3.65}$$

If we assume that Kohn-Sham eigenvalue ϵ_i^{KS} is close to the real quasiparticle energy ϵ_i^{QP} then the self-energy at the quasiparticle energy may be obtained from an expansion around the energy ϵ_i^{KS},

$$\Sigma_i(\epsilon_i^{\mathrm{QP}}) \approx \Sigma_i^{\mathrm{xc}}(\epsilon_i) + (\epsilon_i^{\mathrm{QP}} - \epsilon_i^{\mathrm{KS}})\frac{\partial\Sigma_i^{\mathrm{xc}}(\omega)}{\partial\omega}\bigg|_{\omega=\epsilon_i}, \tag{3.66}$$

where we have used the definition $\Sigma_i^{\mathrm{xc}}(\omega) = \langle\psi_i^{\mathrm{KS}}|\hat{\Sigma}^{\mathrm{xc}}(\omega)|\psi_i^{\mathrm{KS}}\rangle$. We may then estimate the roots of Equation (3.65) by the Newton-Raphson method and we get

$$\epsilon_i^{\mathrm{QP}} = \epsilon_i^{\mathrm{KS}} + Z_i\big[\Sigma_i^{\mathrm{xc}}(\epsilon_i) - \epsilon_i^{\mathrm{xc}}\big], \tag{3.67}$$

where $\epsilon_i^{\mathrm{xc}} = \langle\psi_i^{\mathrm{KS}}|v_{\mathrm{xc}}(\mathbf{r})|\psi_i^{\mathrm{KS}}\rangle$ and Z_i is called the renormalization factor and is given by

$$Z_i = \left[1 - \frac{\partial\Sigma_{ii}^{\mathrm{xc}}(\omega)}{\partial\omega}\bigg|_{\omega=\epsilon_i}\right]^{-1}. \tag{3.68}$$

While the renormalization factor is actually just a result of the Newton-Rapshon method it can also be shown that it corresponds to the norm of the wave function $\psi_i(\mathbf{r})$ and thus gives a measure of how well the single-particle wave function approximates that of a true quasiparticle state[43]. If $Z_i \ll 1$, the wave function is not a good description of a quasi particle state but if $Z_i \approx 1$ it is very close to a real quasi particle wave function.

3.6 Implementation in GPAW

In order to estimate quasiparticle energies from Kohn-Sham wave functions and eigenvalues by the method described in previous section we need an expression for the self-energy. As already mentioned the simplest approximation is to use the GW approximation given in Equation (3.58). In the frequency domain this takes the form

$$\Sigma^{\mathrm{xc}}(\mathbf{r},\mathbf{r}';\omega) = \frac{i}{2\pi}\int G(\mathbf{r},\mathbf{r}';\omega+\omega')W(\mathbf{r},\mathbf{r}';\omega')e^{i\omega'\delta}\,\mathrm{d}\omega', \tag{3.69}$$

where a small positive number δ has been introduced in order to ensure that the Fourier transformation converges. The starting point for approach is G_0W_0, i.e. we will initially use the non-interacting Green's function G_0 and the screened potential in the simplest approximation, W_0. The screened potential, Equation (3.40), depends

on the dielectric function which from its definition in Equation (3.26) can be related to the irreducible polarizability $P(12)$ by

$$\begin{aligned}\varepsilon(12) &= \frac{\delta v_{\text{ext}}(1)}{\delta v_{\text{tot}}(2)} = \frac{\delta}{\delta v_{\text{tot}}(2)}\left[v_{\text{tot}}(1) - \int v(13)n(3)\,\mathrm{d}(3)\right] \\ &= \delta(12) - \int v(13)P(32)\,\mathrm{d}(3). \end{aligned} \tag{3.70}$$

In real space and frequency domain this corresponds to

$$\varepsilon(\mathbf{r},\mathbf{r}',\omega) = \delta(\mathbf{r}-\mathbf{r}') - \int v(\mathbf{r}.\mathbf{r}'')P(\mathbf{r}'',\mathbf{r}';\omega)\,\mathrm{d}\mathbf{r}''. \tag{3.71}$$

In the standard implementation in GPAW it is assumed that the system is periodic in all directions and all properties should therefore be invariant under translation with an arbitrary lattice vector R and for a correlation function like the dielectric function this means that $\varepsilon(\mathbf{r}+\mathbf{R},\mathbf{r}'+\mathbf{R};\omega) = \varepsilon(\mathbf{r},\mathbf{r}';\omega)$. For extended systems it is favorable to work in reciprocal space and here the dielectric function therefore takes the form

$$\varepsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega) = \delta_{\mathbf{G}\mathbf{G}'} - \sum_{\mathbf{G}_1} v_{\mathbf{G}\mathbf{G}_1}(\mathbf{q})P_{\mathbf{G}_1\mathbf{G}'}(\mathbf{q};\omega), \tag{3.72}$$

where $\mathbf{q}$ is a point in the first Brillouin zone. In the usual 3D implementation the Coulomb potential in reciprocal space is given by

$$v_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) = \frac{4\pi\delta_{\mathbf{G}\mathbf{G}'}}{|\mathbf{q}+\mathbf{G}|^2}, \tag{3.73}$$

so that the sum over $\mathbf{G}$-vectors could have been left out in Equation (3.72), but as I will show in the next chapter this need not apply to lower-dimensional systems. However, assuming that the Coulomb potential is diagonal in $\mathbf{G}$-vectors it is preferable to work with a *symmetrized* dielectric function given by

$$\tilde{\varepsilon}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) \equiv \frac{\sqrt{v_{\mathbf{G}}(\mathbf{q})}}{\sqrt{v_{\mathbf{G}'}(\mathbf{q})}}\varepsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega), \tag{3.74}$$

with a corresponding inverse:

$$\tilde{\varepsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) \equiv \frac{\sqrt{v_{\mathbf{G}}(\mathbf{q})}}{\sqrt{v_{\mathbf{G}'}(\mathbf{q})}}\varepsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega). \tag{3.75}$$

In terms of the symmetrized dielectric function the Fourier transform of the screened potential is

$$W_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = v_{\mathbf{G}}(\mathbf{q})\varepsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = \sqrt{v_{\mathbf{G}}(\mathbf{q})}\tilde{\varepsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)\sqrt{v_{\mathbf{G}'}(\mathbf{q})}. \tag{3.76}$$

An expression for the dielectric function can directly be found from Equation (3.72) which however depends on the irreducible polarizability $P_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$ which in the

G_0W_0 approximation is given by the RPA expression as given in Equation (3.57). Inserting the Lehmann representation from Equation (3.63) we obtain an expression for the irreducible polarizability in real space and frequency given by[5]

$$P(\mathbf{r},\mathbf{r}';\omega) = 2\sum_i \sum_j (f_i - f_j)\frac{\psi_i(\mathbf{r})\psi_j^*(\mathbf{r})\psi_i^*(\mathbf{r}')\psi_j(\mathbf{r}')}{\omega + \epsilon_i - \epsilon_j + i\eta\,\mathrm{sgn}(\epsilon_i - \epsilon_j)}, \tag{3.77}$$

where f_i is 0 if $\epsilon_i > \mu$ and 1 if $\epsilon_i < \mu$ and the factor 2 in front accounts for spin degeneracy. Now if we use DFT wave functions as an approximation for the quasiparticle wave functions the Green's functions are approximated by the DFT non-interacting Green's functions $G(12) \approx G_0^{\mathrm{KS}}(12)$ and the RPA irreducible polarization $P(12)$ turns out to be equivalent to the non-interacting polarizability $\chi^0(12)$. In terms of KS wave functions this takes the form

$$\chi^0(\mathbf{r},\mathbf{r}';\omega) = 2\sum_{\mathbf{k},n}\sum_{\mathbf{k}',n'} (f_{n\mathbf{k}} - f_{n'\mathbf{k}'})\frac{\psi_{n\mathbf{k}}(\mathbf{r})\psi_{n'\mathbf{k}'}^*(\mathbf{r})\psi_{n\mathbf{k}}^*(\mathbf{r}')\psi_{n'\mathbf{k}'}(\mathbf{r}')}{\omega + \epsilon_{n\mathbf{k}} - \epsilon_{n'\mathbf{k}'} + i\eta\,\mathrm{sgn}(\epsilon_{n\mathbf{k}} - \epsilon_{n'\mathbf{k}'})}, \tag{3.78}$$

where $f_{n\mathbf{k}}$ is the occupation number of the single-particle state $|n\mathbf{k}\rangle$. Fourier transforming Equation (3.78) one has[1, 121]

$$\chi^0_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\omega) = \frac{2}{V}\sum_{\mathbf{k}}\sum_{n,n'} (f_{n\mathbf{k}} - f_{n'\mathbf{k}-\mathbf{q}})\frac{M_{\mathbf{G}}^{nn'}(\mathbf{k},\mathbf{q})\left[M_{\mathbf{G}'}^{nn'}(\mathbf{k},\mathbf{q})\right]^*}{\omega + \epsilon_{n\mathbf{k}} - \epsilon_{n'\mathbf{k}-\mathbf{q}} + i\eta\,\mathrm{sgn}(\epsilon_{n\mathbf{k}} - \epsilon_{n'\mathbf{k}'})}, \tag{3.79}$$

where $M_{\mathbf{G}}^{nn'}(\mathbf{k},\mathbf{q})$ is called an oscillator matrix element and is a shorthand notation for

$$M_{\mathbf{G}}^{nn'}(\mathbf{k},\mathbf{q}) = \langle n'\mathbf{k}-\mathbf{q}|e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}|n\mathbf{k}\rangle. \tag{3.80}$$

The details of the implementation and evaluation of Equation (3.79) in GPAW is described in Ref. 122. Worth mentioning is that the evaluation of an oscillator matrix element includes both a contributions from the pseudo wave functions and an augmentation part that includes the corrections to the pseudo part from within the augmentation spheres. Since we are employing the frozen core approximation only a certain number of occupied valence or semi-core states are included and the core sattes are thus excluded from the sum over bands in Equation (3.79).

When the irreducible polarizability has been calculated through Equation (3.79) one can then calculate the symmetrized dielectric function trough Equations (3.72) and (4.42) and the inverse can be found numerically by direct matrix inversion. In order to calculate the self-energy the screened potential given in Equation (3.40) is split up into two:

$$W(12) = v(12) + \overline{W}(12), \tag{3.81}$$

where $v(12)$ is the bare Coulomb potential and $\overline{W}(12)$ is the dynamically screened potential given by

$$\overline{W}(12) = \int v(13)[\varepsilon^{-1}(32) - \delta(32)]\,\mathrm{d}(3). \tag{3.82}$$

There are two reasons for this: First of all the GW self-energy is then also seen to be split up into two:

$$\Sigma^{\mathrm{xc}}(12) = iG(12)v(12) + iG(12)\overline{W}(12) = \Sigma^x(12) + \Sigma^c(12), \tag{3.83}$$

where $\Sigma^x(12)$ is the static exchange self-energy, whose contribution to the quasiparticle energy in Equation (3.64) will be seen to be equal to the Fock exchange energy, and $\Sigma^c(12)$ is the dynamical correlation self-energy. Secondly one finds that the division in Equation (3.83) allows for different numerical treatments to be employed, which ultimately leads to greater numerical accuracy.

I will now start by considering the exchange part. Using the Lehmann representation of the DFT non-interacting Green's function $G_0^{\mathrm{DFT}}(12)$ and the fact that the Coulomb potential does not depend on time on can show that

$$\langle n\mathbf{k}|\hat{\Sigma}^x|n'\mathbf{k}\rangle = -\frac{1}{V}\sum_{m}^{\mathrm{occ}}\sum_{\mathbf{q}}^{\mathrm{BZ}}\sum_{\mathbf{G}} v_{\mathbf{G}}(\mathbf{q})\left[M_{\mathbf{G}}^{mn}(\mathbf{k},\mathbf{q})\right]^* M_{\mathbf{G}}^{mn'}(\mathbf{k},\mathbf{q}). \tag{3.84}$$

The fact that the exchange self-energy is static and that the sum over all quasiparticle states in the Greens function has been replaced by a sum over only occupied states ($\epsilon_i < \mu$) makes actual calculations quite easy. However, the expression still contains a sum over in principle an infinite amount of $\mathbf{G}$-vectors that in practice is truncated. The convergence with respect to this cutoff is not always fast and one may often need to use a very high cutoff. The expression also contains a $\mathbf{q}$-point integration over the Brillouin zone, but the integrand which contains the Coulomb potential $v_{\mathbf{G}}(\mathbf{q}) = 4\pi/|\mathbf{q}+\mathbf{G}|^2$ which is singular for $|\mathbf{q}+\mathbf{G}| \to \mathbf{0}$. Mathematically the integral should not be a problem since it the singularity is integrable in three dimensions, but in practice the integrand is only evaluated numerically on a set of $\mathbf{q}$-points and if the $\mathbf{q} = \mathbf{0}$ term is left out the integral converges very slowly with the size of the $\mathbf{q}$-point sampling. This problem was recognized early on and the first solution was to evaluate the contribution to the integral from the $\mathbf{q} = \mathbf{0}$ term analytically by approximating it as[45]

$$\int_{\Omega_\Gamma} \frac{f(\mathbf{q})}{|\mathbf{q}|^2}\,\mathrm{d}\mathbf{q} \approx f(\mathbf{q}=\mathbf{0})\int_{\Omega_\Gamma}\frac{1}{|\mathbf{q}|^2}\,\mathrm{d}\mathbf{q} \tag{3.85}$$

If the small volume around $\mathbf{q} = \mathbf{0}$, Ω_Γ, is taken to be spherical the integral is readily integrated analytically. Several other methods have been developed to provide better approximations to this integral for instance auxiliary functions[31, 119, 12] or spherical real space truncation of the Coulomb potential[110]. However, so far the best overall performance has been obtained by a method which relies on a modified Coulomb potential that is truncated in real space at the so-called Wigner-Seitz cell[112] and this is also the method employed in GPAW.

Next up is the correlation self-energy. Inserting the DFT non-interacting Green's

functions as replacement for the true Green's functions leads to the expression

$$\begin{aligned}\langle n\mathbf{k}|\hat{\Sigma}^c(\omega)|n'\mathbf{k}\rangle = \frac{i}{V}\sum_{\mathbf{q}}^{\mathrm{BZ}}\sum_{\mathbf{G}\mathbf{G}'}\frac{1}{2\pi}\int_{-\infty}^{\infty}\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega') \\ \times\sum_{m}\frac{\left[M_{\mathbf{G}}^{mn}(\mathbf{k},\mathbf{q})\right]^* M_{\mathbf{G}'}^{mn'}(\mathbf{k},\mathbf{q})}{\omega+\omega'-\epsilon_{m\mathbf{k}+\mathbf{q}}+i\eta\,\mathrm{sgn}(\epsilon_{m\mathbf{k}+\mathbf{q}}-\mu)}\,\mathrm{d}\omega'.\end{aligned} \tag{3.86}$$

This expression is much more complex than the one for the exchange part and its calculation can be very computationally demanding. Several schemes have been developed to try and simplify its evaluation. First of all there exists several ways of approximating the dielectric function through so-called Plasmon Pole Models so that the frequency integration can be carried out analytically[26, 84, 44, 73, 19]. If higher accuracy is needed one can also directly calculate the dielectric function at a number of real and/or imaginary frequencies and carry out the integral using either contour deformation[65] or analytical continuation techniques[27]. However in this project I have used a direct integration over real frequencies made efficient by the use of Hilbert transforms[77, 108, 122]. Besides the frequency integration the evaluation of Equation (3.86) is hampered by the sum over an infinite amount of bands and **G**-vectors. Typically these sums have been truncated at rather arbitrarily chosen cutoff, but it has been found that in reality the self-energy converges quite slowly with the number of bands and plane waves used, especially when semicore states are included[115]. In fact it has been observed that the truncation error scales as $1/N$, where N is the number of basis functions used[49, 107] which has made it possible to extrapolate the calculated self-energies to an infinite basis set[117, 43]. The $1/N$ scaling of the error has recently been explained theoretically and approximate methods of avoiding it without extrapolation has also been suggested[52]. Lastly one notices that the correlation self-energy, Equation (3.86), like the exchange part also contains a **q**-point integration over the singular Coulomb potential. However, the this Coulomb potential is effectively screened since the most singular part has been subtracted, and the divergence is less severe such that the simple integration method presented in Equation (3.85) can be used without too much trouble. However, as we shall see in next chaper this fact changes completely for two-dimensional systems. Futher details on the implementation of *GW* in GPAW can be found in Ref. 43.

CHAPTER 4

Calculating the electronic structure for 2D systems

When studying materials we are often first interested in the properties of its most pure form since even in the real world where any sample may be mixed with other materials, contain different kinds of disorder, impurities etc., these may often be a good approximation to the real properties or at least at good starting point for further investigation. This means that we usually start by studying the smallest possible system that captures the physics of the entire extended system. For bulk materials this usually means studying the unit cell, though in order to capture anti-ferromagnetic ordering or other complex magnetic structures one has to consider larger supercells. For 2D systems this is not quite as straightforward as the real unit cell is really infinite in the perpendicular direction. To simulate a system which has an infinite non-periodic extend is of course not practically possible so we need some way of dealing with this. For ground state properties of atoms or molecules it is common to use a finite simulation cell and fixing the wave function to be zero on the boundary which can conveniently be done in a real space basis. When simulating extended solids and expecially when calculating excited state properties using pertubation theory methods it is often more convenient to work in a plane wave basis. When studying systems with reduced dimensionality like 2D systems one could work in a plane wave basis in the periodic direction and a real space basis with zero boundary conditions in the non-periodic direction. This requires a special implementation so is not that practical. Instead we will choose to work in a standard planewave basis which is periodic in all directions. This, however gives some problems that are treated in this chapter.

4.1 Plane waves for 2D systems

As mentioned in the introduction when simulating 2D systems we have to find some way to describe wave functions and other properties in the non-periodic direction. Our starting point is that we will consider the system using a finite unit cell with a length L in the non-periodic direction, here taken as the z-direction. If the cell is only repeated in-plane the solutions to the single particle Schrödinger equation are wave function of the Bloch form,

$$\psi_{n\mathbf{k}_\parallel}(\mathbf{r}_\parallel, z) = e^{i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel} u_{n\mathbf{k}_\parallel}(\mathbf{r}_\parallel, z), \tag{4.1}$$

where $\mathbf{k}_\parallel$ is a wave vector in the two-dimensional Brillouin zone and $u_{n\mathbf{k}_\parallel}(\mathbf{r}_\parallel + \mathbf{R}_\parallel, z) = u_{n\mathbf{k}_\parallel}(\mathbf{r}_\parallel, z)$ is function that is periodic in-plane. The Fourier transform of the periodic part of the wave function is simply,

$$u(\mathbf{G}) = \frac{1}{V} \int_\Omega u(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} \, \mathrm{d}\mathbf{r}, \tag{4.2}$$

where Ω denotes the unit cell, V is it's volume and $\mathbf{G} = (\mathbf{G}_\parallel, G_\perp)$ is a reciprocal lattice vector satisfying $e^{i\mathbf{G}_\parallel \cdot \mathbf{R}_\parallel} = 1$ and $G_\perp = \frac{2\pi n}{L}$, with n an integer. Note that this Fourier transform is well defined even in the case where the system is not periodic in the z-direction as long as the function has the boundary condition $u(\mathbf{r}_\parallel, 0) = u(\mathbf{r}_\parallel, L)$. When solving the single particle Schrödinger equation we rewrite all the contributions to the Hamiltonian $H_{\mathbf{G}\mathbf{G}'}(\mathbf{k})$ in terms of their Fourier components which is easily done for 3D systems since the only troublesome term is external potential which is naturally periodic in all directions and thus satisfy the periodic boundary conditions necessary for the Fourier transform to converge everywhere. Specifically we have

$$V_\mathbf{G} = \frac{1}{V} \int_\Omega V(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} \tag{4.3}$$

with $V(\mathbf{r}+\mathbf{R}) = V(\mathbf{r})$. However, for 2D systems the external potential is not periodic in all directions and also need not be symmetric and the requirement $V(\mathbf{r}_\parallel, 0) = V(\mathbf{r}_\parallel, L)$ is in general not satisfied. Assuming that the 2D system is located in the center of the unit cell, then in the limit $L \to \infty$ we expect both $V(0) \to 0$ and $V(L) \to 0$ so that the boundary condition is asymptotically fulfilled. In reality we are using a finite L so to make our 2D system satisfy the boundary conditions we need to introduce some small error that fixes the potential. One way is to smoothly adjust the value of the potential near the boundary so that it satisfies the boundary conditions. Another way is to take the potential to also be periodic in the out-of-plane direction creating an artificial periodic structure. The last method is what we will do since this is basically equivalent to the first method if L is large enough, since the Coloumb potential from the atom cores drops off as $1/z$.

So we assume that if the length, L, of the unit cell in the non-periodic direction is large enough it is fair approximation to take the Fourier components of the external potential to be equal to those of the equivalent unit cell that is periodic in all directions.

This means that the Hamiltonian also becomes completely identical to an equivalent system periodic in all directions. The solutions to the corresponding Schrödinger equation has the general Bloch solutions,

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}), \tag{4.4}$$

where $\mathbf{r}$ is a vector in all of space, $\mathbf{k}$ is reciprocal vector in a 3D Brillouin zone and the function $u_{n\mathbf{k}}(\mathbf{r})$ is periodic in all directions. The solutions to the 2D system correspond to the subset which has *exactly* $k_z = 0$ in which case we recover Equation (4.2). The real system is not periodic in the out-of-plane direction so $k_z \neq 0$ is not really defined. It is possible to model the 2D system by considering it as a real 3D crystal periodic also in the z-direction but with layers interacting so weakly that the solutions will in principle be without dispersion in that direction. In this case the Brillouin zone is three-dimensional and one can have properties that are non-zero for $k_z \neq 0$ corresponding to oscillations with wavelengths larger than L. When L is large these components will be small, but I will still say that they are artificial and should be neglected. Doing so effectively leads to only considering solutions with $k_z = 0$ and thus reducing the Brillouin zone to an exact 2D region.

The use of a finite unit cell in the out-of-plane direction has some simple consequences when dealing with Fourier transformed quantities specifically we have

$$f(\mathbf{k}_\parallel, G_\perp) = \frac{1}{L} \int \int_0^L e^{i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel} e^{iG_\perp z} f(\mathbf{r}_\parallel, z)\, \mathrm{d}z\, \mathrm{d}\mathbf{r}_\parallel \tag{4.5}$$

$$f(\mathbf{r}_\parallel, z) = \frac{1}{(2\pi)^2} \sum_{G_z} \int e^{-i\mathbf{k}_\parallel \cdot \mathbf{r}} e^{-iG_\perp z} f(\mathbf{k}_\parallel, G_z)\, \mathrm{d}\mathbf{k}_\parallel, \tag{4.6}$$

where the Fourier transform in the z-direction has been replaced by a Fourier series. Most of the properties calculated in any plane wave implementation assumes that the system is periodic in all directions and we don't want to change this and lucklily we don't have to. Since we have assumed all properties to have periodic boundary conditions the usual 3D periodic Fourier transform will only have non-zero components at $k_z = G_z$, where it will be sharply peaked δ-function. If we by $f^{\mathrm{3D}}(\mathbf{k}_\parallel, k_z)$ denote the weight of this distribution the 2D plane wave expansion coefficients, Equation (4.5), are related to the standard 3D equivalents by

$$f^{\mathrm{2D}}(\mathbf{k}_\parallel, G_z) = \frac{1}{2\pi} f^{\mathrm{3D}}(\mathbf{k}_\parallel, k_z = G_z). \tag{4.7}$$

For a two-point function that is periodic in-plane, i.e. $f(\mathbf{r} + \mathbf{R}, z; \mathbf{r}' + \mathbf{R}, z') =$

$f(\mathbf{r}, z; \mathbf{r}', z')$, we have Fourier transform relations:

$$f_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \mathbf{q}') = (2\pi)^2 \delta(\mathbf{q} - \mathbf{q}') f_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) \tag{4.8}$$

$$\begin{aligned} f_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) = \frac{1}{L^2}\frac{1}{A}\int_{\mathcal{A}}\int\int_0^L\int_0^L & e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r})} e^{-iG_z z} \\ & \times f(\mathbf{r}, z; \mathbf{r}', z') e^{i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}'} e^{iG'_z z'} \,\mathrm{d}z \,\mathrm{d}z' \,\mathrm{d}\mathbf{r} \,\mathrm{d}\mathbf{r}' \end{aligned} \tag{4.9}$$

$$f(\mathbf{r}, z; \mathbf{r}', z') = \frac{1}{(2\pi)^2} \sum_{\mathbf{G}\mathbf{G}'} \int_{\mathrm{BZ}} e^{i(\mathbf{q}_\parallel + \mathbf{G}_\parallel)\cdot\mathbf{r}} e^{iG_z z} f_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) e^{-i(\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}'} e^{-iG_z z'} \,\mathrm{d}\mathbf{q} \tag{4.10}$$

If we have a convolution in real space:

$$f(\mathbf{r}, z; \mathbf{r}', z') = \int\int_0^L g(\mathbf{r}, z; \mathbf{r}_1, z_1) h(\mathbf{r}_1, z_1; \mathbf{r}', z') \,\mathrm{d}\mathbf{r}_1 \,\mathrm{d}z_1 \tag{4.11}$$

its Fourier transform is correspondingly

$$f_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \mathbf{q}) = \frac{L}{(2\pi)^2} \sum_{\mathbf{G}_1} \int_{\mathrm{BZ}} g_{\mathbf{G}\mathbf{G}_1}(\mathbf{q}, \mathbf{q}_1) h_{\mathbf{G}_1\mathbf{G}'}(\mathbf{q}_1, \mathbf{q}') \,\mathrm{d}\mathbf{q}_1. \tag{4.12}$$

An elementary quantity for all response properties is the density response function or polarizability $\chi(\mathbf{r}, \mathbf{r}', \omega)$. The starting point is the non-interacting irreducible polarization $\chi^0(\mathbf{r}, \mathbf{r}'; \omega)$ for which an expression is given in Equation (3.78). As all response function this is periodic in plane, $\chi^0(\mathbf{r} + \mathbf{R}, z; \mathbf{r}' + \mathbf{R}, z') = \chi^0(\mathbf{r}, z; \mathbf{r}', z')$ so its Fourier transform is given by

$$\begin{aligned} \chi^{0,2\mathrm{D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}; \omega) = \frac{1}{L^2}\frac{1}{A}\sum_{\mathbf{k}}\sum_{n,n'} & \frac{f_{n\mathbf{k}} - f_{n'\mathbf{k}-\mathbf{q}}}{\omega + \epsilon_{n\mathbf{k}} - \epsilon_{n'\mathbf{k}-\mathbf{q}} + i\eta\,\mathrm{sgn}(\epsilon_{n\mathbf{k}} - \epsilon_{n'\mathbf{k}'})} \\ & \times M^{nn'}_{\mathbf{G}}(\mathbf{k}, \mathbf{q}) \left[M^{nn'}_{\mathbf{G}'}(\mathbf{k}, \mathbf{q})\right]^* \end{aligned}. \tag{4.13}$$

Except for a factor of $1/L$ this is similar to the expression for 3D systems, Equation (3.79), except that the sum over $\mathbf{k}$-points is only over the 2D Brillouin zone. When doing calculation we will use the 3D expression but only include $\mathbf{k}$-points in-plane and we therefore have the simple relation

$$\chi^{0(2\mathrm{D})}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}; \omega) = \frac{1}{L}\chi^{0(3\mathrm{D})}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}; \omega). \tag{4.14}$$

In most cases the Fourier transforms of 2D response functions are related to their 3D counterparts by a simple factor $1/L$ so that the values become independent of the choice of unit cell height. However care has to be taken when the expressions involve an integral over the Brillouin zone as the implementation assumes that the system is periodic in all directions.

4.2 Truncated Coulomb potential

In the previous section we discussed how we deal with functions and response functions in 2D using a plane wave basis and ended up with an expression for the Fourier transform of 2D non-interacting density response function. We are most often interested in more realistic interacting system in which case we need the full response function, which is related to the non-interacting response function by the Dyson equation, Eq. (3.43),

$$\chi(\mathbf{r},\mathbf{r}') = \chi^0(\mathbf{r},\mathbf{r}') + \int\int \chi^0(\mathbf{r},\mathbf{r}_1)K(\mathbf{r}_1,\mathbf{r}_2)\chi(\mathbf{r}_2,\mathbf{r}')\,\mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2. \tag{4.15}$$

Using the definition Equation (4.5) the 2D Fourier transform of this expression is

$$\begin{aligned}\chi_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\mathbf{q}') = \chi^0_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\mathbf{q}') + \frac{L^2}{(2\pi)^4}\int_{\mathrm{BZ}}\mathrm{d}\mathbf{q}_1\int_{\mathrm{BZ}}\mathrm{d}\mathbf{q}_2\sum_{\mathbf{G}_1\mathbf{G}_2}\chi^0_{\mathbf{G}\mathbf{G}_1}(\mathbf{q},\mathbf{q}_1)\\ \times K^{\mathrm{2D}}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q}_1,\mathbf{q}_2)\chi_{\mathbf{G}_2\mathbf{G}'}(\mathbf{q}_2,\mathbf{q}')\end{aligned} \tag{4.16}$$

where

$$\begin{aligned}K^{\mathrm{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\mathbf{q}') = \frac{1}{L^2}\int\mathrm{d}\mathbf{r}_\parallel\int\mathrm{d}\mathbf{r}'_\parallel\int_0^L\mathrm{d}z\int_0^L\mathrm{d}z'\\ \times e^{-i(\mathbf{q}+\mathbf{G}_\parallel)\cdot\mathbf{r}_\parallel}e^{-iG_z z}K(\mathbf{r}_\parallel,z;\mathbf{r}'_\parallel,z')e^{i(\mathbf{q}'+\mathbf{G}'_\parallel)\cdot\mathbf{r}'_\parallel}e^{iG'_z z'}\end{aligned} \tag{4.17}$$

Note that the integrals over the z-direction in Eq. 4.17 only extends over the unit cell height, which is why we will call this for the 2D truncated kernel. In RPA we have $K(\mathbf{r},\mathbf{r}') = v(\mathbf{r},\mathbf{r}') = 1/|\mathbf{r}'-\mathbf{r}|$, so we may just define the 2D truncated Coulomb interaction as

$$v^{\mathrm{2D}}(\mathbf{r},z;\mathbf{r}',z') = \frac{\theta(z)\theta(L-z)\theta(z')\theta(L-z')}{\sqrt{(\mathbf{r}'-\mathbf{r})^2+(z'-z)^2}} \tag{4.18}$$

with the Fourier transform given by

$$v^{\mathrm{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q};\mathbf{q}') = \frac{1}{L^2}\iint\int_0^L\int_0^L\frac{e^{-i(\mathbf{q}+\mathbf{G}_\parallel)\cdot\mathbf{r}_\parallel}e^{-iG_z z}e^{i(\mathbf{q}'+\mathbf{G}'_\parallel)\cdot\mathbf{r}'_\parallel}e^{iG'_z z'}}{\sqrt{(\mathbf{r}'-\mathbf{r})^2+(z'-z)^2}}\,\mathrm{d}\mathbf{r}_\parallel\,\mathrm{d}\mathbf{r}'_\parallel\,\mathrm{d}z\,\mathrm{d}z'. \tag{4.19}$$

Due to the in-plane symmetry of the Coulomb potential it's Fourier represenetation is diagonal in $\mathbf{q}_\parallel$ and $\mathbf{q}'_\parallel$ and it therefore simplifies to

$$v^{\mathrm{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) = \frac{\delta_{\mathbf{G}_\parallel,\mathbf{G}'_\parallel}}{L^2}\int\int_0^L\int_0^L\frac{e^{-i(\mathbf{q}+\mathbf{G}_\parallel)\cdot\mathbf{r}_\parallel}e^{-iG_z z}e^{iG'_z z'}}{\sqrt{\mathbf{r}_\parallel^2+(z'-z)^2}}\,\mathrm{d}z'\,\mathrm{d}z\,\mathrm{d}\mathbf{r}_\parallel \tag{4.20}$$

$$= \frac{1}{L^2}\int\int_0^L e^{-i(G_z-G'_z)Z}\int_{-\frac{L}{2}+|Z-\frac{L}{2}|}^{\frac{L}{2}+|Z-\frac{L}{2}|}\frac{e^{-i(\mathbf{q}+\mathbf{G}_\parallel)\cdot\mathbf{r}_\parallel}e^{-\frac{i}{2}(G_z+G'_z)u}}{\sqrt{\mathbf{r}_\parallel^2+u^2}}\,\mathrm{d}u\,\mathrm{d}Z\,\mathrm{d}\mathbf{r}_\parallel \tag{4.21}$$

where in the last line the integrations over z and z' have been substituted by integrated over the mean and relative coordinates, $Z = (z + z')/2$ and $u = z' - z$, respectively. Due to the finite integration limits this integral does not have a closed analytical expression and we are therefore required to either solve it using numerical Fourier transformation methods or by introducing further approximations.

A simple approximation is to instead of truncating the Coulomb potential at the unit cell border, truncate it at some relative distance R_c between two particles,

$$v^{\text{trunc}}(\mathbf{r}, z; \mathbf{r}', z') = \frac{\theta(R - |z' - z|)}{\sqrt{(\mathbf{r}' - \mathbf{r})^2 + (z' - z)^2}}. \tag{4.22}$$

The Coulomb potential thus only depends on the relative positions of the particles and a simple closed analytical expression for the 3D Fourier transform can be found[46, 102], given by

$$v^{\text{trunc}}(\mathbf{q}) = \frac{4\pi}{q^2}\left[1 + e^{-|\mathbf{q}_\parallel|R}\left(\frac{q_z}{|\mathbf{q}_\parallel|}\sin(q_z R) - \cos(q_z R)\right)\right], \tag{4.23}$$

where q_z here denotes the z-component of a general vector in a 3D reciprocal space. Since this is based on Coulomb potential that is independent of the absolute position and extend of the unit cell, it is well-defined for all values of q_z and we could in principle choose any value of the real space cut-off R. For instance we see that we recover the usual untruncated 3D Coulomb potential $v(q) = \frac{4\pi}{q^2}$ when $q_\parallel R \gg 1$, i.e. when the wavelength becomes much smaller than the real-space cut-off. However it appears that the potential diverges when $q_\parallel \to 0$ for *any* $q_z \neq 0$. The divergencies are integrable but it is still unsatisfying and hard to deal with numerically. It is seen that the divergencies for all but $q_z = 0$ can be removed by choosing $R = L/2$, such that $\sin(q_z L/2) = 0$ and $\cos(q_z L/2) = 1$, and we have

$$v^{\text{trunc}}_{\mathbf{G}}(\mathbf{q}) = \frac{4\pi}{|\mathbf{q} + \mathbf{G}|^2}\left[1 - e^{-(|\mathbf{q}| + G_\parallel)L/2}\cos(G_z L/2)\right], \tag{4.24}$$

where we let $\mathbf{q}$ denote only a point in the in-plane Brillouin zone and $\mathbf{G}$ any reciprocal lattice vector. In the special case $\mathbf{G} = \mathbf{0}$ this reduces to

$$v^{\text{2D}}_{\mathbf{0}}(\mathbf{q}) = \frac{4\pi}{|\mathbf{q}|^2}(1 - e^{-|\mathbf{q}|L/2}). \tag{4.25}$$

We see that in the small wavelength limit we recover the 3D like behavior of the Coulomb potential, $v^{\text{2D}}_{\mathbf{0}}(q) \approx \frac{4\pi}{q^2}$ for $q \gg L/2$, while in the long wavelength limit, $q \ll L/2$ we have $v^{\text{2D}}_{\mathbf{0}}(q) \approx \frac{2\pi}{q}$, which diverges much slower than in the 3D case.

We note that the truncation is equivalent to replacing the integration limits of the integration over the relative z-coordinate u in Equation (4.21) with $-L/2$ and $L/2$, respectively. Thus the truncated Coulomb potential can be seen as an approximation to the real 2D Coulomb potential and we thus note that we have

$$v^{\text{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) \approx \frac{\delta_{\mathbf{G}\mathbf{G}'}}{L} v^{\text{trunc}}_{\mathbf{G}}(\mathbf{q}). \tag{4.26}$$

Using this approximation we can then rewrite Equation (4.16) into the form

$$\chi^{\mathrm{2D}}_{\mathbf{GG}'}(\mathbf{q}) = \chi^{0(\mathrm{2D})}_{\mathbf{GG}'}(\mathbf{q}) + L\sum_{\mathbf{G}_1} \chi^{0(\mathrm{2D})}_{\mathbf{GG}_1}(\mathbf{q}) v^{\mathrm{trunc}}_{\mathbf{G}_1}(\mathbf{q}) \chi^{\mathrm{2D}}_{\mathbf{G}_1\mathbf{G}'}(\mathbf{q}). \tag{4.27}$$

We note that if we replace the 2D non-interacting response function with the 3D equivalent, $\chi^{0(\mathrm{2D})}_{\mathbf{GG}'}(\mathbf{q}) \to \frac{1}{L}\chi^{0(\mathrm{3D})}_{\mathbf{GG}'}(\mathbf{q})$, we recover the usual Dyson equation for a 3D system with the Coloumb kernel replaced by the truncated potential.

The truncated Coulomb potential, Equation (4.22), allows for really simple expressions an a minimum of changes in the practical implementation since it fits nicely with the plane wave basis. However, it is not completely without problems. As illustrated in Figure 4.1 the truncated Coulomb potential will introduce and interaction between charge densities in artificial periodic images of the unit cell. The charge density in the center of the unit cell will not interact with neighboring cells, but so will the charge density further away from the center, and the charge just on the unit cell border interacts equally with the original cell and an artificial one. This of course is not a problem if the unit cell is big enough and the density drops off fast enough so that the interacting densities of two cells are negligible. If we let the region of non-vanishing density by characerized by a length h, then the amount of ''vacuum´´ on each side of the layer is $l = (L-h)/2$ and if we don't want the regions of non-vanishing densities to interact it means that the unit cell height should be chosen such that

$$L > 2h. \tag{4.28}$$

For thin materials this is not really a tough constraint, but it is clear that with materials of increasing size the amount of vacuum must also be increased drastically even though the electron density is likely to vanish into the vacuum region at the same rate as for a thin material. Since the number of **G**-vectors scale with the unit cell size the above criterion means that thick materials or multilayers may need an unfavorable

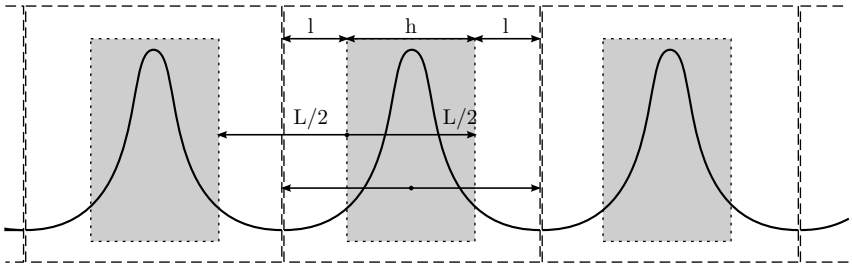


Figure 4.1: Illustration of the interaction with artificial periodic images introduced by the truncated Coulomb potential. The region of non-vanishing density is shaded in grey and has width h, the unit cell height is L and the amount of vacuum is thus $l = (L-h)/2$.

large number of **G**-vectors making calculations very expensive. This problem can in theory be allieviated by implementation of the correct Fourier transform of the 2D Coulomb potential given in Equation (4.21), but it is not known how to do this in practice.

4.3 Dielectric function for 2D systems

As already shown in the introduction; due to the vanishing electron density in the perpendicular direction the dielectric properties of 2D materials are qualitatively different than for bulk systems, that have a higher degree of isotropy. In general the dielectric function gives the relation between an external potential and the total field,

$$v^{\text{ext}}(\mathbf{r}, z; \omega) = \int \int_0^L \varepsilon(\mathbf{r}z, \mathbf{r}'z') v^{\text{tot}}(\mathbf{r}', z'; \omega) \, \mathrm{d}z' \, \mathrm{d}\mathbf{r}'. \tag{4.29}$$

Since the dielectric function is lattice periodic the 2D Fourier transform of Equation (4.29) can be written

$$v^{\text{ext}}_{\mathbf{G}}(\mathbf{q}, \omega) = L \sum_{\mathbf{G}'} \varepsilon^{\text{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) v^{\text{tot}}_{\mathbf{G}'}(\mathbf{q}, \omega), \tag{4.30}$$

which means that $\varepsilon^{\text{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega)$ can be thought of as response density per z-distance. Often we are going to compare the dielectric function to 3D bulk systems and in that case we are interested in the dielectric response over all of space and thus the relevant quantity is $\varepsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = L\varepsilon^{\text{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega)$, as seen by Equation (4.30). To make the connection with the results from 3D systems I will mostly use the 3D dielectric function $\varepsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega)$ and will specifically mark the 2D Fourier transformed dielectric by a superscript ``2D''.

In Section 3.4 it was shown that the dielectric function can be calculated within RPA from the non-interacting density response function by

$$\varepsilon^{\text{RPA}}(\mathbf{r}, \mathbf{r}'; \omega) = \delta(\mathbf{r} - \mathbf{r}') - \int v(\mathbf{r}, \mathbf{r}_1) \chi^0(\mathbf{r}_1, \mathbf{r}'; \omega). \tag{4.31}$$

Using the truncated potential, Equation (4.26), the 2D Fourier transform of the RPA dielectric function is

$$\varepsilon^{\text{RPA(2D)}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) = \frac{1}{L} \delta_{\mathbf{G}\mathbf{G}'} - v^{\text{2D}}_{\mathbf{G}}(\mathbf{q}) \chi^{0(\text{2D})}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}), \tag{4.32}$$

which is equal to $\frac{1}{L}\varepsilon^{\text{RPA(3D)}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})$ with $\mathbf{q}$ lying in the $q_z = 0$ plane of the 3D Brillouin zone. The dielectric function has a qualitively different behavior in the long wavelength limit $q \to 0$ than for bulk systems. To see this we note that the irreducible polarizability contains a product of oscillator matrix elements, $\langle n\mathbf{k}|e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}|m\mathbf{k}+\mathbf{q}\rangle$, which in the case of $\mathbf{G} = \mathbf{0}$ long wavelength limit has the leading order[122]

$$\langle n\mathbf{k}|e^{i\mathbf{q}\cdot\mathbf{r}}|m\mathbf{k}+\mathbf{q}\rangle \approx -i\mathbf{q} \cdot \frac{\langle n\mathbf{k}|\boldsymbol{\nabla}|m\mathbf{k}\rangle}{\epsilon_{n\mathbf{k}} - \epsilon_{m\mathbf{k}}} \quad \text{for} \quad \mathbf{q} \to \mathbf{0}. \tag{4.33}$$

For a non-metal it can therefore be shown that the head ($\mathbf{G} = \mathbf{G}' = \mathbf{0}$) and wings ($\mathbf{G} = 0$ or $\mathbf{G}' = \mathbf{0}$) of the irreducible polarizability will have the following lowest order expression

$$\chi^0_{\mathbf{00}}(\mathbf{q},\omega) = \mathbf{q}\cdot\mathsf{P}(\omega)\mathbf{q} = |\mathbf{q}|^2\hat{\mathbf{q}}\cdot\mathsf{P}(\omega)\hat{\mathbf{q}} \tag{4.34}$$

$$\chi^0_{\mathbf{G0}}(\mathbf{q},\omega) = \mathbf{q}\cdot\mathbf{p}_{\mathbf{G}}(\omega) = |\mathbf{q}|\hat{\mathbf{q}}\cdot\mathbf{p}_{\mathbf{G}}(\omega) \tag{4.35}$$

$$\chi^0_{\mathbf{0G}'}(\mathbf{q},\omega) = \mathbf{q}\cdot\mathbf{p}^*_{\mathbf{G}'}(\omega) = |\mathbf{q}|\hat{\mathbf{q}}\cdot\mathbf{p}^*_{\mathbf{G}}(\omega), \tag{4.36}$$

where P is a tensor and $\mathbf{p}$ is a vector that depends on a reciprocal lattice vector. It is thus understood that $\mathbf{q}\cdot\mathsf{P}(\omega)\mathbf{q} = \sum_{ij} q_i P_{ij}(\omega) q_j$ and $\mathbf{q}\cdot\mathbf{p}_{\mathbf{G}}(\omega) = \sum_i q_i p_i(\mathbf{G},\omega)$. Inserting these expressions in Eq. 4.32 the head and wings of the dielectric function are

$$\varepsilon^{\mathrm{RPA}}_{\mathbf{00}}(\mathbf{q}\to\mathbf{0},\omega) = 1 - 4\pi(1-e^{-|\mathbf{q}|L/2})\hat{\mathbf{q}}\cdot\mathsf{P}(\omega)\hat{\mathbf{q}} \tag{4.37}$$

$$\varepsilon^{\mathrm{RPA}}_{\mathbf{G0}}(\mathbf{q}\to\mathbf{0},\omega) = -v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{q})|\mathbf{q}|\hat{\mathbf{q}}\cdot\mathbf{p}_{\mathbf{G}}(\omega) \tag{4.38}$$

$$\varepsilon^{\mathrm{RPA}}_{\mathbf{0G}'}(\mathbf{q}\to\mathbf{0},\omega) = -\frac{4\pi(1-e^{-|\mathbf{q}|L/2})}{|\mathbf{q}|}\hat{\mathbf{q}}\cdot\mathbf{p}^*_{\mathbf{G}'}(\omega). \tag{4.39}$$

We immediately see that in the long wavelength limit $\mathbf{q}\to\mathbf{0}$ the head of the dielectric function $\varepsilon^{\mathrm{RPA}}_{\mathbf{00}}(\mathbf{q}) \to 1$, which is very different from the 3D case, where it attains a finite value > 1.

Often we are interested in the inverse of the dielectric function, defined by

$$\int \mathrm{d}\mathbf{r}''\varepsilon^{-1}(\mathbf{r},\mathbf{r}'')\varepsilon(\mathbf{r}'',\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}'). \tag{4.40}$$

In Fourier space this is equivalent to

$$\sum_{\mathbf{G}''}\varepsilon^{-1}_{\mathbf{G}\mathbf{G}''}(\mathbf{q})\varepsilon_{\mathbf{G}''\mathbf{G}'}(\mathbf{q}) = \delta_{\mathbf{G}\mathbf{G}'}, \tag{4.41}$$

and the inverse dielectric function $\varepsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q})$ can in practice be found by a matrix inversion. We would however like to have an expression for the $\mathbf{q}\to\mathbf{0}$ limit for inverse dielectric function. In order to facilitate this and make it easier to exploit symmetries to reduce the computational load in *GW* calculations we will often work with the *symmetrized* dielectric function given by

$$\begin{aligned}\tilde{\varepsilon}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) &= \frac{\sqrt{v_{\mathbf{G}'}(\mathbf{q})}}{\sqrt{v_{\mathbf{G}}(\mathbf{q})}}\varepsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)\\ &= \delta_{\mathbf{G}\mathbf{G}'} - \sqrt{v_{\mathbf{G}}(\mathbf{q})}\chi^0_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)\sqrt{v_{\mathbf{G}'}(\mathbf{q})}.\end{aligned} \tag{4.42}$$

The standard inverse dielectric function is then related to the inverse of the inverse symmetrized dielectric function by

$$\varepsilon^{-1}_{\mathbf{G}\mathbf{G}'} = \frac{\sqrt{v_{\mathbf{G}'}(\mathbf{q})}}{\sqrt{v_{\mathbf{G}}(\mathbf{q})}}\tilde{\varepsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega), \tag{4.43}$$

which means that we can easily find one given the other. Inserting the lowest order expansion of $\chi_{\mathbf{GG}}(\mathbf{q})$, Equations (4.34) to (4.36), in Equation (4.42) we see that the head and wings of the symmetrized RPA dielectric function are

$$\begin{aligned}\tilde{\varepsilon}^{\mathrm{RPA}}_{\mathbf{00}}(\mathbf{q}\to\mathbf{0},\omega) &= 1 - v^{\mathrm{2D}}(\mathbf{q})|\mathbf{q}|^2\hat{\mathbf{q}}\cdot\mathsf{P}(\omega)\hat{\mathbf{q}} \\ &= 1 - 4\pi(1-e^{-|\mathbf{q}|L/2})\hat{\mathbf{q}}\cdot\mathsf{P}(\omega)\hat{\mathbf{q}} \qquad (4.44)\\ \tilde{\varepsilon}^{\mathrm{RPA}}_{\mathbf{G0}}(\mathbf{q}\to\mathbf{0},\omega) &= -\sqrt{v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{0})}\sqrt{v^{\mathrm{2D}}_{\mathbf{0}}(\mathbf{q})}|\mathbf{q}|\hat{\mathbf{q}}\cdot\mathbf{p}_{\mathbf{G}}(\omega) \qquad (4.45)\\ &= -\sqrt{v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{0})}\sqrt{4\pi(1-e^{-|\mathbf{q}|L/2})}\hat{\mathbf{q}}\cdot\mathbf{p}_{\mathbf{G}}(\omega) \qquad (4.46)\\ \tilde{\varepsilon}^{\mathrm{RPA}}_{\mathbf{0G}}(\mathbf{q}\to\mathbf{0},\omega) &= \left[\tilde{\varepsilon}^{\mathrm{RPA}}_{\mathbf{G0}}(\mathbf{q}\to\mathbf{0},\omega)\right]^* \qquad (4.47)\end{aligned}$$

As with the standard dielectric function we have $\tilde{\varepsilon}_{\mathbf{00}}(\mathbf{0}) = 1$ and $\tilde{\varepsilon}_{\mathbf{G0}}(\mathbf{0}) = 0$. To obtain the analytic behavior of the inverse symmetrized dielectric function at small $\mathbf{q}$ we write the dielectric function as a block matrix in the $\mathbf{G},\mathbf{G}'$ components (ignoring function parameters),

$$\tilde{\boldsymbol{\varepsilon}} = \begin{pmatrix} H & \mathbf{v} \\ \mathbf{w} & \mathbf{B}\end{pmatrix}, \qquad (4.48)$$

where $H = \tilde{\varepsilon}_{\mathbf{00}}(\mathbf{q},\omega)$, $w_{\mathbf{G}} = \tilde{\varepsilon}_{\mathbf{G0}}(\mathbf{q},\omega)$, $v_{\mathbf{G}'} = \tilde{\varepsilon}_{\mathbf{0G}'}(\mathbf{q},\omega)$ and $B_{\mathbf{GG}'} = \tilde{\varepsilon}_{\mathbf{GG}'}(\mathbf{q},\omega)$, with $\mathbf{G},\mathbf{G}' \neq \mathbf{0}$. Using the rule of inversion of block matrices we then have[96]

$$\boldsymbol{\varepsilon}^{-1} = \begin{pmatrix} (H-\mathbf{v}\mathbf{B}^{-1}\mathbf{w})^{-1} & -(H-\mathbf{v}\mathbf{B}^{-1}\mathbf{w})^{-1}\mathbf{v}\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{w}(H-\mathbf{v}\mathbf{B}^{-1}\mathbf{w})^{-1} & \mathbf{B}^{-1}+\mathbf{B}^{-1}\mathbf{w}(H-\mathbf{v}\mathbf{B}^{-1}\mathbf{w})^{-1}\mathbf{v}\mathbf{B}^{-1}\end{pmatrix}. \qquad (4.49)$$

From this we see that

$$\begin{aligned}\tilde{\varepsilon}^{-1}_{\mathbf{00}} &= \left(\tilde{\varepsilon}_{\mathbf{00}} - \sum_{\mathbf{G},\mathbf{G}'\neq\mathbf{0}} \tilde{\varepsilon}_{\mathbf{0G}} B^{-1}_{\mathbf{GG}'}\tilde{\varepsilon}_{\mathbf{G}'\mathbf{0}}\right)^{-1} \qquad (4.50)\\ \tilde{\varepsilon}^{-1}_{\mathbf{G0}} &= -\tilde{\varepsilon}^{-1}_{\mathbf{00}}\sum_{\mathbf{G}'\neq\mathbf{0}} B^{-1}_{\mathbf{GG}'}\tilde{\varepsilon}_{\mathbf{G}'\mathbf{0}} \qquad (4.51)\\ \tilde{\varepsilon}^{-1}_{\mathbf{0G}} &= -\tilde{\varepsilon}^{-1}_{\mathbf{00}}\sum_{\mathbf{G}'\neq\mathbf{0}} \tilde{\varepsilon}_{\mathbf{0G}'} B^{-1}_{\mathbf{G}'\mathbf{G}} \qquad (4.52)\end{aligned}$$

$$\tilde{\varepsilon}^{-1}_{\mathbf{GG}'} = B^{-1}_{\mathbf{GG}'} - \tilde{\varepsilon}^{-1}_{\mathbf{00}}\left(\sum_{\mathbf{G}''\neq\mathbf{0}} B^{-1}_{\mathbf{GG}''}\tilde{\varepsilon}_{\mathbf{G}''\mathbf{0}}\right)\left(\sum_{\mathbf{G}''\neq\mathbf{0}} \tilde{\varepsilon}_{\mathbf{0G}''}B^{-1}_{\mathbf{G}''\mathbf{G}'}\right). \qquad (4.53)$$

We can now find the analytic behavior in the long wavelength limit by inserting Equations (4.37), (4.46) and (4.47). To take out the $\mathbf{q}$-dependence and simplify the notation we define

$$\begin{aligned}\mathbf{a}_{\mathbf{G}}(\omega) &= -\sum_{\mathbf{G}'\neq\mathbf{0}} B^{-1}_{\mathbf{GG}'}\sqrt{v^{\mathrm{2D}}_{\mathbf{G}'}(\mathbf{0})}\mathbf{p}_{\mathbf{G}'}(\omega) \qquad (4.54)\\ \mathsf{A}(\omega) &= -\mathsf{P}(\omega) - \sum_{\mathbf{G}\neq\mathbf{0}}\sqrt{v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{0})}\mathbf{p}^*_{\mathbf{G}}(\omega)\otimes\mathbf{a}_{\mathbf{G}}(\omega), \qquad (4.55)\end{aligned}$$

where $\otimes$ is the outer product, meaning that we have $(\mathbf{p}^*_{\mathbf{G}} \otimes \mathbf{a}_{\mathbf{G}})_{ij} = p^*_i(\mathbf{G})a_j(\mathbf{G})$. We can then write

$$\tilde{\varepsilon}^{-1}_{\mathbf{00}}(\mathbf{q} \to \mathbf{0}, \omega) = \frac{1}{1 + 4\pi(1 - e^{-|\mathbf{q}|L/2})\hat{\mathbf{q}} \cdot \mathsf{A}(\omega)\hat{\mathbf{q}}} \tag{4.56}$$

$$\tilde{\varepsilon}^{-1}_{\mathbf{G0}}(\mathbf{q} \to \mathbf{0}, \omega) = -\frac{\sqrt{4\pi(1 - e^{-|\mathbf{q}|L/2})}\hat{\mathbf{q}} \cdot \mathbf{a}_{\mathbf{G}}(\omega)}{1 + 4\pi(1 - e^{-|\mathbf{q}|L/2})\hat{\mathbf{q}} \cdot \mathsf{A}(\omega)\hat{\mathbf{q}}} \tag{4.57}$$

$$\tilde{\varepsilon}^{-1}_{\mathbf{0G}}(\mathbf{q} \to \mathbf{0}, \omega) = -\frac{\sqrt{4\pi(1 - e^{-|\mathbf{q}|L/2})}\hat{\mathbf{q}} \cdot \mathbf{a}^*_{\mathbf{G}}(\omega)}{1 + 4\pi(1 - e^{-|\mathbf{q}|L/2})\hat{\mathbf{q}} \cdot \mathsf{A}(\omega)\hat{\mathbf{q}}} \tag{4.58}$$

$$\tilde{\varepsilon}^{-1}_{\mathbf{GG'}}(\mathbf{q} \to \mathbf{0}, \omega) = B^{-1}_{\mathbf{GG'}}(\omega) - \frac{4\pi(1 - e^{-|\mathbf{q}|L/2})(\hat{\mathbf{q}} \cdot \mathbf{a}_{\mathbf{G}}(\omega))(\hat{\mathbf{q}} \cdot \mathbf{a}^*_{\mathbf{G'}}(\omega))}{1 + 4\pi(1 - e^{-|\mathbf{q}L/2})\hat{\mathbf{q}} \cdot \mathsf{A}(\omega)\hat{\mathbf{q}}}. \tag{4.59}$$

Not surprisingly we have for $\mathbf{q} = \mathbf{0}$ that $\tilde{\varepsilon}^{-1}_{\mathbf{00}}(\mathbf{0}) = 1$, $\tilde{\varepsilon}^{-1}_{\mathbf{G0}}(\mathbf{0}) = \tilde{\varepsilon}^{-1}_{\mathbf{0G}}(\mathbf{0}) = 0$. The low-order expansion can (and will) be used not only to determine the exact $\mathbf{q} = \mathbf{0}$ limit of the inverse dielectric function but also the value at finite but small $\mathbf{q}$. The check the validty of the expansion we have performed calculations of the full irreducible polarizability, $\chi_{\mathbf{GG'}}(\mathbf{q}, \omega)$, for a dense sampling of $\mathbf{q}$-points along a path in the Brillouin zone for a few materials. We start by studying a monolayer of h-BN as it is a relatively simple material that is cheap to calculate due to its low number of valence electrons (= 8) and small unit cell. Specifically we have calculated the irreducible polarizability based on a DFT calculation done with a very dense $\mathbf{k}$-grid of $90 \times 90 \times 1$ points. In Figure 4.2 the calculated polarizability, dielectric function and inverse dielectric function is shown for a range of $\mathbf{q}$-points along the $\Gamma \to M$ direction in the Brillouin zone and compared to the corresponding functions based on the lowest order expansion at $\mathbf{q} \to \mathbf{0}$. we see that the lowest order expansion for the head of the static irreducible polarizability for h-BN seems to be an almost spot-on approximation up to about 20 % to 30 % of the $\Gamma \to M$ distance. The linear approximation to the wings is good for a distance up to $\sim 0.1|\overrightarrow{\Gamma M}|$ but then deviates considerably. The accuracy of the expansions are directly carried over to the dielectric function: The linear expansion of the truncated Coulomb potential is enough for getting the slope at $\mathbf{q} = \mathbf{0}$ right. However, the approximations seem to be quite worse for the inverse dielectric function; for the head it still does well up to about $0.1|\overrightarrow{\Gamma M}|$, while for the wing it deviates already at $\sim 0.05|\overrightarrow{\Gamma M}|$ and overestimates the maximal value by $\sim 30\,\%$. The situation appears to be even worse for monolayer MoS_2, see Figure 4.3, where for the case shown, the low order expansion of the wing of the inverse dielectric function even has the wrong sign. The expansion for the head of the inverse dielectric function, however, still appears to be a quite good approximation for small $\mathbf{q}$. The reason why the low order expansion leads to bad descriptions of the inverse functions is because of local field effects; when inverting the matrix the head and wings contain contributions from all $\mathbf{G}$-components of the dielectric function. As we see from Figure 4.2 and Figure 4.3, the head of the irreducible polarizability is well described by an expansion to second order, but the wings, which we only expand to first order, are not quite so. The component shown, $\mathbf{G} = (-1, 0, 0)$ is even quite

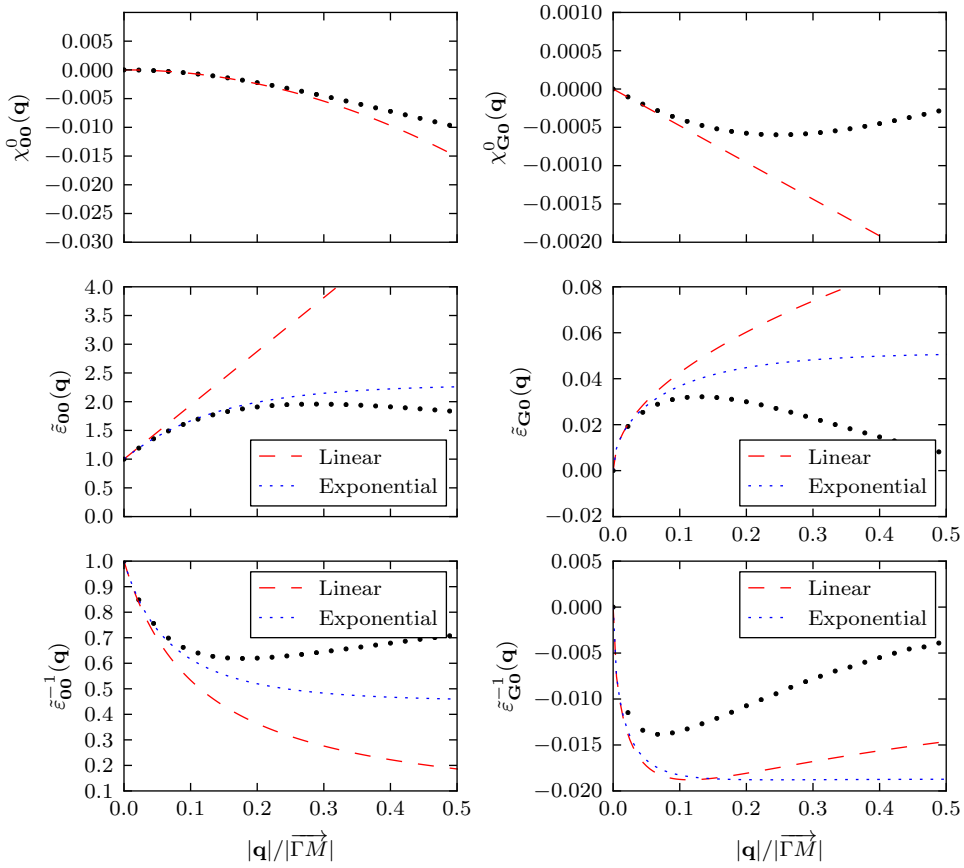


Figure 4.2: Comparison of the small $\mathbf{q}$ expansion and full numerically calculated value of the head ($\mathbf{G} = \mathbf{G}' = \mathbf{0}$) and wings ($\mathbf{G}' = \mathbf{0}$) of (a) and (b): the density response function, (c) and (d): the symmetrized dielectric function and (e) and (f): the inverse symmetrized dielectric function. In all cases we consider the static, $\omega = 0$, value and for the wings we consider the component with $\mathbf{G} = (-1, 0, 0)$, in scaled coordinates respective to the reciprocal cell.

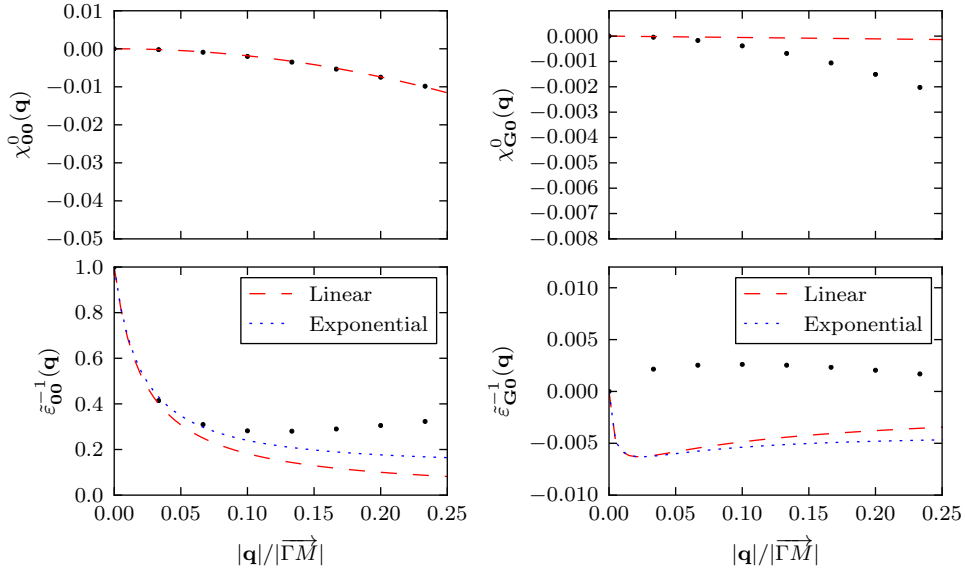
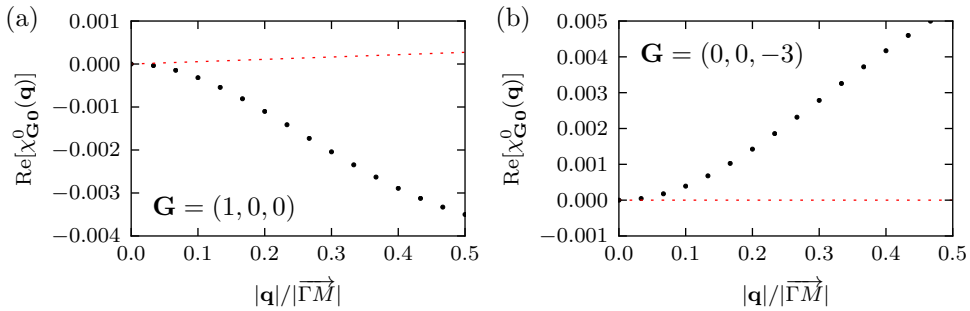


Figure 4.3: Irreducible polarizability and inverse symmetrized dielectric function for monolayer 2H-MoS_2.

fortunate chosen in that regard, and we see that if we choose another **G**-vector, see Figure 4.4(b), we get results where the linear approximation is very bad. For MoS_2 the wing component for the component with $\mathbf{G} = (1, 0, 0)$ is seen to have a small but positive slope, but the actual results are negative. If we consider a component with a **G**-vector pointing in the z-direction we see that the slope is practically zero and the calculated values instead follow a parabolic trend.



hej

((b)) Irreducible polarizability of MoS_2 for two different G-vectors

The problem seems to be that the linear expansion in the wings is in general only valid for extremely small values of $\mathbf{q}$; the high-order prefactors are typically much larger than for the linear term. For the components with $\mathbf{G}$-vectors only along the z-direction this is actually pretty clear why as the unit cell is typically much larger in the out-of-plane direction than in the in-plane directions which means that the approximation $|\mathbf{q}+\mathbf{G}_z| \approx |\mathbf{G}_z|$ is only valid when $|\mathbf{q}| \ll 2\pi/L$. The description should be greatly improved by going to one order higher in the expansion, which in practice could be done by calculating $\langle \mathbf{k}n|e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}|\mathbf{k}+\mathbf{q}m\rangle$ using $\mathbf{k}\cdot\mathbf{p}$ pertubation theory. The wings of the inverse dielectric function is seen to be several orders of magnitude smaller than the head component so for the many purposes a more accurate treatment of the wings in the long wavelength limit is not needed and the implementation of higher order corrections has therfore not been pursued.

4.3.1 The macroscopic dielectric function

When considering experiments that measure macroscopic quantities, like e.g. the light absorption or electron energy loss spectrum (EELS) of a bulk crystal, the response from the microscopic spatial variations are typically averaged out and give and just give an overall contribution to the signal. It is therefore conventient to define *macroscopic* quantities that only carry information about the effect...

When writing a field in terms of its Fourier transform we can write it as a sum of waves with periodicity both larger and smaller than the unit cell size

$$v(\mathbf{r},\omega) = \int_{\mathrm{BZ}} \mathrm{d}\mathbf{q} \sum_{\mathbf{G}} v(\mathbf{q}+\mathbf{G},\omega) e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} \tag{4.60}$$

$$= \int_{\mathrm{BZ}} \mathrm{d}\mathbf{q} e^{-i\mathbf{q}\cdot\mathbf{r}} v(\mathbf{r};\mathbf{q},\omega), \tag{4.61}$$

where $v(\mathbf{r};\mathbf{q},\omega) = \sum_{\mathbf{G}} v(\mathbf{q}+\mathbf{G},\omega)e^{-i\mathbf{G}\cdot\mathbf{r}}$ is a function that has the periodicity of the unit cell, i.e. $v(\mathbf{r}+\mathbf{R};\mathbf{q},\omega) = v(\mathbf{r};\mathbf{q},\omega)$. When considering most experiments like optical absorption, electron energy loss etc. the rapidly varying changes in the field are typically averaged out and it therefore often suffices to consider the *macroscopic* quantity defined as the average of $v(\mathbf{r},\mathbf{q},\omega)$ averaged over the unit cell

$$v_m(\mathbf{q}) = \frac{1}{V}\int_{\Omega} \mathrm{d}\mathbf{r}' v(\mathbf{r}+\mathbf{r}';\mathbf{q},\omega) \tag{4.62}$$

$$= \frac{1}{V}\int_{\Omega} \mathrm{d}\mathbf{r}' \sum_{\mathbf{G}} v_{\mathbf{G}}(\mathbf{q},\omega) e^{-i\mathbf{G}\cdot(\mathbf{r}+\mathbf{r}')} \tag{4.63}$$

$$= \sum_{\mathbf{G}} v_{\mathbf{G}}(\mathbf{q}) \frac{1}{V}\int_{\Omega} \mathrm{d}\mathbf{r}'' e^{-i\mathbf{G}\cdot(\mathbf{R}+\mathbf{r}'')} \tag{4.64}$$

$$= v_{\mathbf{0}}(\mathbf{q}) \tag{4.65}$$

where we have used that the position can be written in terms of a lattice vector plus a position within a unit cell, $\mathbf{r} = \mathbf{R}+\tilde{\mathbf{r}}$ and the integral over the unit cell is thus seen

to yield a kronecker delta $\frac{1}{V}\int_\Omega \mathrm{d}\mathbf{r} e^{-i\mathbf{G}\cdot\mathbf{r}} = \delta_{\mathbf{G},\mathbf{0}}$. The macroscopic quantity is thus obtained by simply taking the $\mathbf{G} = \mathbf{0}$ component of the Fourier transform.

When considering the dielectric properties of a material we are interested in the behavior of the total field, $v^{\mathrm{tot}}(\mathbf{q}+\mathbf{G},\omega)$ due to an external macroscopic field $v^{\mathrm{ext}}(\mathbf{q})$, which follows from the dielectric function via the relation

$$v^{\mathrm{ext}}(\mathbf{q}) = \sum_{\mathbf{G}'} \varepsilon_{\mathbf{0}\mathbf{G}'}(\mathbf{q},\omega) v^{\mathrm{tot}}_{\mathbf{G}'}(\mathbf{q},\omega). \tag{4.66}$$

By inverting the dielectric function we then have

$$v^{\mathrm{tot}}_{\mathbf{G}}(\mathbf{q},\omega) = \varepsilon^{-1}_{\mathbf{G}\mathbf{0}}(\mathbf{q},\omega) v^{\mathrm{ext}}(\mathbf{q},\omega) \tag{4.67}$$

and the macroscopic total potential is found by taking the $\mathbf{G} = \mathbf{0}$ component and we have

$$v^{\mathrm{tot}}_{m}(\mathbf{q},\omega) = \varepsilon^{-1}_{\mathbf{0}\mathbf{0}}(\mathbf{q},\omega) v^{\mathrm{ext}}(\mathbf{q},\omega). \tag{4.68}$$

We use Eq. (4.68) as a basis for defining the *macroscopic dielectric function*,

$$\varepsilon_m(\mathbf{q},\omega) = \frac{1}{\varepsilon^{-1}_{\mathbf{0}\mathbf{0}}(\mathbf{q},\omega)}. \tag{4.69}$$

The macroscopic dielectric function is somewhat ill-defined for 2D systems because the unit cell length in the non-periodic direction is arbitrarily chosen. This means that the usual macroscopic quantities will decrease with increasing unit cell size, so that their values also become completely arbitrary. One way to solve this issue is to average only over a small region containing the important physics of system. If we consider a 2D system periodic in the x, y-direction centered in a unit cell that extends in the z-direction and we choose the averaging region, Ω_{2D} to be within a distance of $\pm d/2$ of the center of the system, located at z_0, we define the 2D macroscopic field corresponding to the total microscopic field $v(\mathbf{q}+\mathbf{G})$

$$v_m(\mathbf{r}_\parallel,\mathbf{q}) = \frac{1}{V_{\mathrm{2D}}}\int_{\Omega_{\mathrm{2D}}} \mathrm{d}\mathbf{r}' v(\mathbf{r}+\mathbf{r}') \tag{4.70}$$

$$= \sum_{\mathbf{G}} v_{\mathbf{G}}(\mathbf{q},\omega)\frac{1}{A}\int_{\mathcal{A}} \mathrm{d}\mathbf{r}' e^{-i\mathbf{G}_\parallel\cdot(\mathbf{r}_\parallel+\mathbf{r}'_\parallel)}\frac{1}{d}\int_{-d/2}^{d/2}\mathrm{d}z' e^{-iG_z(z_0+z')} \tag{4.71}$$

$$= \frac{2}{d}\sum_{\mathbf{G}} e^{iG_z z_0}\frac{\sin(G_z d/2)}{G_z}\delta_{\mathbf{G}_\parallel,\mathbf{0}} v_{\mathbf{G}}(\mathbf{q},\omega). \tag{4.72}$$

We see that in the case where the averaging region is the entire unit cell, $d = L$, we have $v^m(\mathbf{r},\mathbf{q} = v_{\mathbf{0}}(\mathbf{q}$ and we recover the 3D macroscopic average.

In a similar fashion as fro the 3D case we now go on to define a macroscopic dielectric function in terms of the macroscopic total potential generated from a macroscopic external potenal, Eq. (4.67),

$$v^{\mathrm{tot}}_m(\mathbf{q},\omega) = \frac{2}{d}\sum_{\mathbf{G}} e^{iG_z z_0}\frac{\sin(G_z d/2)}{G_z}\delta_{\mathbf{G}_\parallel,\mathbf{0}}\varepsilon^{-1}_{\mathbf{G},\mathbf{0}}(\mathbf{q},\omega) v^{\mathrm{ext}}(\mathbf{q},\omega). \tag{4.73}$$

We thus define the 2D macroscopic dielectric function, $\varepsilon_m^{\mathrm{2D}}(\mathbf{q},\omega)$, from the relation

$$\frac{1}{\varepsilon_m^{\mathrm{2D}}(\mathbf{q},\omega)} = \frac{2}{d}\sum_{\mathbf{G}} e^{iG_z z_0}\frac{\sin(G_z d/2)}{G_z}\delta_{\mathbf{G}_\parallel,\mathbf{0}}\varepsilon^{-1}_{\mathbf{G0}}(\mathbf{q},\omega). \tag{4.74}$$

The special behavior of the microscopic dielectric function in the long wavelength limit, $\mathbf{q}\to\mathbf{0}$, carries over to the macroscopic quantity. We therefore use the relation between the inverse dielectric function and its symmetrized counterpart, Equation (4.43), and insert the small $\mathbf{q}$ expressions, Equations (4.56) to (4.59), into Equation (4.74) and see that we have

$$\varepsilon_m^{\mathrm{2D}}(\mathbf{q}\to\mathbf{0},\omega) = \frac{1+4\pi(1-e^{-|\mathbf{q}|L/2})\hat{\mathbf{q}}\cdot\mathsf{A}(\omega)\hat{\mathbf{q}}}{1-|\mathbf{q}|\hat{\mathbf{q}}\cdot\mathbf{c}_m(\omega)}, \tag{4.75}$$

with

$$\mathbf{c}_m(\omega) = \frac{d}{2}\sum_{\mathbf{G}\neq\mathbf{0}} e^{iG_z z_0}\frac{\sin(G_z d/2)}{G_z}\delta_{\mathbf{G}_\parallel,\mathbf{0}}\mathbf{c}_{\mathbf{G}}(\omega). \tag{4.76}$$

Here $\mathsf{A}(\omega)$ is the given by Equation (4.55) and $\mathbf{c}_{\mathbf{G}}(\omega)$ is similar to Equation (4.54),

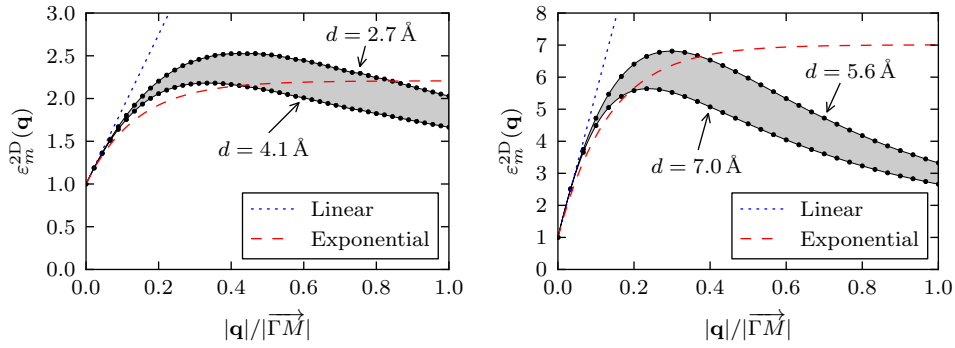
$$\mathbf{c}_{\mathbf{G}}(\omega) = \sum_{\mathbf{G}'\neq\mathbf{0}} B^{-1}_{\mathbf{G}\mathbf{G}'} v^{\mathrm{2D}}_{\mathbf{G}'}(\mathbf{0})\mathbf{p}_{\mathbf{G}'}(\omega), \tag{4.77}$$

where B^{-1} is the inverse of the body of the (unsymmetrized) dielectric function. From Equation (4.75) we see that the macroscopic dielectric function is 1 for $\mathbf{0}$ with an initial slope of

$$\left.\frac{\mathrm{d}\varepsilon_m^{\mathrm{2D}}(\mathbf{q},\omega)}{\mathrm{d}q}\right|_{q=0} = 2\pi L\hat{\mathbf{q}}\cdot\mathsf{A}(\omega)\hat{\mathbf{q}} + \hat{\mathbf{q}}\cdot\mathbf{c}_m(\omega). \tag{4.78}$$

We notice that the only difference in the initial slope of 2D and 3D macroscopic dielectric function, with the latter being defined by Equation (4.69), is the addition of the term $\hat{\mathbf{q}}\cdot\mathbf{c}_m(\omega)$, which includes the local field effects from only averaging over a part of the unitcell.

In Figure 4.5 the 2D macroscopic dielectric function is shown for monolayer h-BN and 2H-MoS_2 for different values of the averaging region height d corresponding to regions which contain between 98 % and 99.8 % of the electron density. It is clear that the initial slope is not affected by the choice of averaging region as was also found in Ref. [42]. The absolute values of the macroscopic dielectric function varies somewhat with the choice of averaging region, but considering that differences correspond to a variation in the total vacuum electron density of a factor of 10, it is really not that sensitive. The observation that the slope does not vary with the averaging region is also observed from the calculated values of $\mathbf{c}_m(\omega)$, which are in the order of $\sim 10^{-19}$Å and thus can safely be ignored in practice. The low order expansion of the macroscopic dielectric function, Equation (4.75), is therefore surprisingly completely insensitive to



(a) h-BN **(b)** MoS_2

Figure 4.5: Static macroscopic dielectric function of (a) h-BN and (b) 2H-MoS_2 calculated using two different values for the averaging region size d. The smallest d is chosen so that 98 % of the total valence electron density is contained within the averaging region, while the larges d corresponds to a region which contains 99.8 % of the total valence electron density. The dashed lines corresponds to the linearized an full (expontential) low order expansion for $\mathbf{q} \to \mathbf{0}$ as given by Equation (4.75).

the averaging region and as is seen in Figure 4.5 it fails to describe much more than the initial slope. Again, the reason for this is the failure of the first order expansion of the wings of the irreducible polarization for **G**-vectors in the z-direction as described earlier.

4.4 The GW self energy

A large part of this project has been devoted to calculating quasiparticle energies of 2D materials using the GW approximation, so it clear that we must study how the implementation of this method can be changed to fit into the 2D formalism. As noted in Section 3.5 the central part in calculating GW quasiparticle energies is the self-energy $\Sigma = iGW$. For technical reasons we split the self-energy into an exchange part $\Sigma^x = iGV$ and a dynamical correlation part $\Sigma^c = iG\overline{W}$, with the dynamical screened potenial given by $\overline{W} = (\varepsilon^{-1} - 1)V$. When calculating the exchange self-energy one has to treat the divergency of the Coulomb potential, which as explained in Section 3.5 can be done using a modified Coulomb potential truncated in real space at the Wigner-Seitz cell. The advantage of this method, besides a fast-converging regularization of the Coulomb potential, is that it inherently truncates the potential in the non-periodic direction in way similar to that described in Section 4.2 and is thus directly applicable to 2D systems. So we only have to consider the treatment of

the correlation self-energy which in real space is given by

$$\Sigma^c(\mathbf{r}, z; \mathbf{r}', z'; \omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\mathbf{r}, z; \mathbf{r}', z'; \omega + \omega') W(\mathbf{r}, z; \mathbf{r}', z'; \omega') \, \mathrm{d}\omega'. \tag{4.79}$$

Following the procedure described in Section 3.6 we can find the matrix elements of the self-energy in a quasi-particle basis by inserting the 2D Fourier transform of the screened potential and the Green's function in it's Lehmann representation and we obtain

$$\begin{aligned} \langle n\mathbf{k}|\Sigma^x|n'\mathbf{k}\rangle = \frac{i}{(2\pi)^2} \int_{\mathrm{BZ}} \mathrm{d}\mathbf{q} \sum_{\mathbf{G}\mathbf{G}'} \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega' \overline{W}^{\mathrm{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) \\ \times \sum_m \frac{[M^{mn}_{\mathbf{G}}(\mathbf{k}, \mathbf{q})]^* M^{mn'}_{\mathbf{G}'}(\mathbf{k}, \mathbf{q})}{\omega + \omega' - \epsilon_{m\mathbf{k}+\mathbf{q}} + i\eta \, \mathrm{sgn}(\epsilon_{m\mathbf{k}+\mathbf{q}} - \mu)}. \end{aligned} \tag{4.80}$$

The central part in the implementation of this equation is the evaluation of the $\mathbf{q}$-point integral over the Brillouin zone. The dynamical part of the screened potential, given by

$$\overline{W}^{\mathrm{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = L \sum_{\mathbf{G}''} v_{\mathbf{G}\mathbf{G}'}(\mathbf{q}) (\varepsilon^{-1(\mathrm{2D})}_{\mathbf{G}''\mathbf{G}'}(\mathbf{q}, \omega) - \delta_{\mathbf{G}''\mathbf{G}'}), \tag{4.81}$$

which by using the 2D truncated Coulomb potential and the 3D inverse dielectric function may be written

$$\overline{W}^{\mathrm{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = \frac{1}{L} v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{q}) (\varepsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) - \delta_{\mathbf{G}\mathbf{G}'}). \tag{4.82}$$

We can then rewrite Equation (4.80) by in terms of $\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) = L\overline{W}^{\mathrm{2D}}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega)$, such that we can use the usual 3D Fourier transforms most commonly implemented. We note that the expression, Equation (4.80) is similar to the expression for the self-energy of a 3D system, Equation (3.86), except that the $\mathbf{q}$-point integration is over a 2D Brillouin zone. This integral has to be evaluated carefully since the screened potential in the integrand contains the Coulomb potential, $v_{\mathbf{G}}(\mathbf{q})$, which diverges at $\mathbf{q} = \mathbf{0}$. The standard way to perform the $\mathbf{q}$-point integration is by a standard quadrature method where the $\mathbf{q} = \mathbf{0}$ term is left out and treated specially, so that the integral may be written

$$\int_{\mathrm{BZ}} \mathrm{d}\mathbf{q} \mathcal{S}(\mathbf{q}) \to \frac{(2\pi)^2}{A} \frac{1}{N_{\mathbf{q}}} \sum_{\mathbf{q}_n \neq \mathbf{0}} \mathcal{S}(\mathbf{q}_n) + \int_{\Omega_0} \mathrm{d}\mathbf{q} \mathcal{S}(\mathbf{q}), \tag{4.83}$$

where $\mathcal{S}(\mathbf{q})$ denotes the entire integrand, $(2\pi^2)/A$ is the reciprocal space area of the 2D Brillouin zone and Ω_0 is a small region around $\mathbf{q} = \mathbf{0}$. In the usual implementation[44, 108] we assume that for $\mathbf{q} \to \mathbf{0}$ everything but the screened potential is approximately constant in $\mathbf{q}$ and we can therefore take the exact $\mathbf{q} = \mathbf{0}$ and reduce the Γ-region integral to one over the screened potential,

$$\int_{\Omega_0} \mathrm{d}\mathbf{q} \overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) \mathcal{G}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega) \approx \mathcal{G}_{\mathbf{G}\mathbf{G}'}(\mathbf{0}, \omega) \int_{\Omega_0} \mathrm{d}\mathbf{q} \overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}, \omega). \tag{4.84}$$

Writing the dynamical part of the screened potential in terms of the symmetrized inverse dielectric function as introduced in Equation (4.43) we have

$$\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = \sqrt{v_{\mathbf{G}}(\mathbf{q})}\left[\tilde{\varepsilon}^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) - \delta_{\mathbf{G}\mathbf{G}'}\right]\sqrt{v_{\mathbf{G}'}(\mathbf{q})}. \tag{4.85}$$

For 3D dielectric materials the inverse symmetrized dielectric function goes to a finite constant > 1 for $\mathbf{q} \to \mathbf{0}$, so using just this constant in the integration Equation (4.84) we simply have

$$\int_{\Omega_0} \mathrm{d}\mathbf{q}\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = (\varepsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{0},\omega) - \delta_{\mathbf{G}\mathbf{G}'})\int_{\Omega_0} \mathrm{d}\mathbf{q}\sqrt{v_{\mathbf{G}}(\mathbf{q})v_{\mathbf{G}'}(\mathbf{q})}. \tag{4.86}$$

In this case the resulting integral over the symmetrized Coulomb potential can be solved analytically for a spherical volume or using approximate or numerical method for other geometries. However, it is clear that the dielectric properties of 2D dielectric materials causes the dynamical part of the screened potential to have a $\mathbf{q}$-dependent behavior that is qualitatively different than for 3D systems. The reasons is that in order to not introduce artificial interactions when simulating an isolated 2D system using a finite unit cell we replace the Coulomb potential with a truncated one, which according to Equations (4.56) to (4.57) means that for $\mathbf{q} = \mathbf{0}$ we have $\varepsilon^{-1}_{\mathbf{0}\mathbf{0}}(\mathbf{0},\omega) = 1$ and $\varepsilon^{-1}_{\mathbf{G}\mathbf{0}}(\mathbf{0},\omega) = 0$. If we just apply the 3D method for integrating the $\mathbf{q} = \mathbf{0}$ region by using Equation (4.86) we get the result that the head and wings of this quantity is exactly 0, which is very different from the 3D case. It turns out that this conclusion is also false and it will now be shown how one can correct this mistake.

To sort out the problem we return to the screened potential Equation (4.85) and investigate how this behaves in the long wavelength limit by inserting the low order expansion approximations for the inverse symmetrized dielectric function, Equations (4.56) to (4.59), and see that the head and wings are given by

$$\begin{aligned}
\overline{W}_{\mathbf{0}\mathbf{0}}(\mathbf{q}\to\mathbf{0},\omega) &= v^{\mathrm{2D}}_{\mathbf{0}}(\mathbf{q})\left[\tilde{\varepsilon}^{-1}_{\mathbf{0}\mathbf{0}}(\mathbf{q},\omega) - 1\right] \\
&= -\left(\frac{4\pi(1-e^{-|\mathbf{q}|L/2})}{|\mathbf{q}|}\right)^2 \frac{\hat{\mathbf{q}}\cdot\mathsf{A}(\omega)\hat{\mathbf{q}}}{1+4\pi(1-e^{-|\mathbf{q}|L/2})\hat{\mathbf{q}}\cdot\mathsf{A}(\omega)\hat{\mathbf{q}}} \qquad (4.87)\\
\overline{W}_{\mathbf{G}\mathbf{0}}(\mathbf{q}\to\mathbf{0},\omega) &= \sqrt{v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{0})}\tilde{\varepsilon}^{-1}_{\mathbf{G}\mathbf{0}}(\mathbf{q},\omega)\sqrt{v^{\mathrm{2D}}_{\mathbf{0}}(\mathbf{q})} \\
&= -\frac{4\pi(1-e^{-|\mathbf{q}L/2})}{|\mathbf{q}|}\frac{\sqrt{v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{0})}\hat{\mathbf{q}}\cdot\mathbf{a}_{\mathbf{G}}(\omega)}{1+4\pi(1-e^{-|\mathbf{q}|L/2})\hat{\mathbf{q}}\cdot\mathsf{A}(\omega)\hat{\mathbf{q}}} \qquad (4.88)
\end{aligned}$$

and similarly for the body:

$$\begin{aligned}
\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q}\to\mathbf{0},\omega) = \sqrt{v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{0})}\Bigg[&B^{-1}_{\mathbf{G}\mathbf{G}'} - \delta_{\mathbf{G}\mathbf{G}'} \\
&- \frac{4\pi(1-e^{-|\mathbf{q}|L/2})(\hat{\mathbf{q}}\cdot\mathbf{a}_{\mathbf{G}}(\omega))(\hat{\mathbf{q}}\cdot\mathbf{a}^*_{\mathbf{G}'}(\omega))}{1+4\pi(1-e^{-|\mathbf{q}|L/2})\hat{\mathbf{q}}\cdot\mathsf{A}(\omega)\hat{\mathbf{q}}}\Bigg]\sqrt{v^{\mathrm{2D}}_{\mathbf{G}'}(\mathbf{0})}.
\end{aligned} \tag{4.89}$$

Carefully taking the $\mathbf{q} \to \mathbf{0}$ limit we have

$$\overline{W}_{\mathbf{00}}(\mathbf{q} \to \mathbf{0}, \omega) = -(2\pi L)^2 \hat{\mathbf{q}} \cdot \mathsf{A}(\omega) \hat{\mathbf{q}} \tag{4.90}$$

$$\overline{W}_{\mathbf{G0}}(\mathbf{q} \to \mathbf{0}, \omega) = -2\pi L \sqrt{v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{0})} \hat{\mathbf{q}} \cdot \mathbf{a}_{\mathbf{G}}(\omega), \tag{4.91}$$

which are seen to neither diverge nor be exactly 0 and we therefore see that the 3D method for integrating the region around $\mathbf{q} = \mathbf{0}, \textit{Equation}$ (4.86), fails. In figure Figure 4.6

4.4.1 Implementation

The exact integral of the low order expansion does not have an analytical expression. The simplest method for doing so is to evaluate the function on a regular grid inside the region and summing the contributions up as in a standard quadrature method. When doing so a natural question arises: what should the density of this grid be? We note that the function both depends strongly on the A tensor and the height of the unit cell L, so q-point density that is considered reasonable for one material may not be sufficient for another with different A and L. To find a natural $\mathbf{q}$-point density we start by introducing the dimensionless quantity $\mathbf{x} = \mathbf{q}L/2$ and write the low order expansion of the head of the screened potential scaled with respect to its value at $\mathbf{q} = 0$,

$$\tilde{w}(\mathbf{x}) = \frac{\overline{W}_{\mathbf{00}}(2\mathbf{x}/L)}{(2\pi L)^2 \hat{\mathbf{x}} \cdot \mathsf{A}\hat{\mathbf{x}}} = -\left(\frac{1 - e^{-|\mathbf{x}|}}{|\mathbf{x}|}\right)^2 \frac{1}{1 + 4\pi \hat{\mathbf{x}} \cdot \mathsf{A}\hat{\mathbf{x}}(1 - e^{-|\mathbf{x}|})}. \tag{4.92}$$

This function no longer contains any reference to the height of the unit cell and makes it easy to compare screened potential since the only parameter is A. In Figure 4.7 is shown $\tilde{w}(x)$ shown for an isotropic system with different values of $A = \hat{\mathbf{x}} \cdot \mathsf{A}\hat{\mathbf{x}}$ and we

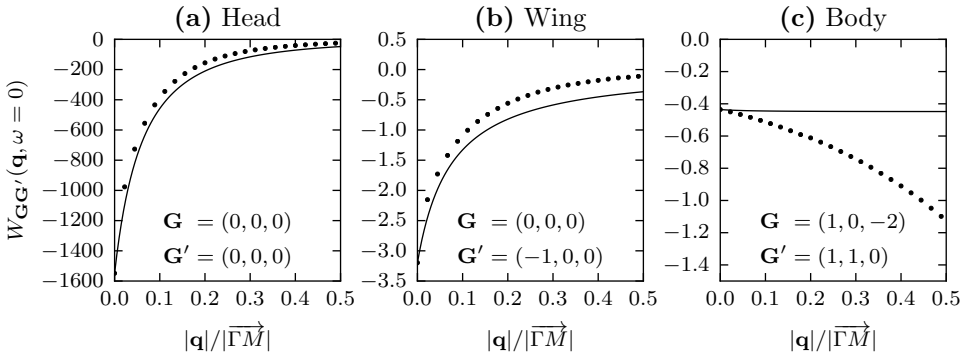


Figure 4.6: Head, wing- and body component of the screened potential.

note that for increasing values of A the function becomes more and more "pointy" at $x = 0$, which could make a numerical integration on a regular grid problematic.

To evaluate this simple integration scheme we need and accurate reference value. To find this we consider an isotropic system and consider a disk shaped integration region, $\mathcal{C}$, of radius $x_{\text{rad}} = 1$ such that we can use polar coordinates and the integral may be written

$$\begin{aligned}\int_{\mathcal{C}} \tilde{w}(\mathbf{x})\,\mathrm{d}\mathbf{x} &= \int_0^{2\pi} \mathrm{d}\phi \int_0^1 \mathrm{d}x\, x\tilde{w}(x) \\ &= -2\pi \int_0^1 \left(\frac{1-e^{-x}}{x}\right)^2 \frac{1}{1+4\pi A(1-e^{-x})}.\end{aligned} \tag{4.93}$$

While this integral still does not have a solution that can be written in terms on known analytical functions it can efficiently be evaluated using accurate numerical methods. To evalute the integral using a regular grid, we calculate the function values on a $N \times N$ Monkhorst-Pack grid centered on $\mathbf{x} = \mathbf{0}$, set the values on points with $|\mathbf{x}_i| > 1$ to 0. In Figure 4.8 is shown how the regular grid integration compares to the method where Equation (4.93) is calculated using an accurate numerical method with absolute tolerance 1×10^{-10}. We see that for $A < 1$ the integration is quickly converged to within 1 % for grid sizes less than 100×100 while for larger values of A it requires grid sizes in excess of 100×100. It is also clear that for $N \gg 1$ the integration error scales approximately as $1/N^2$. However, for $A > 0$, we see that the error scales linearly with A using even very dense grids. The reason is that for large values of A, the function becomes so sharp at $x = 0$, that only very fine grids can resolve it.

In order to make the convergence with respect to grid points better when $A > 0$ we

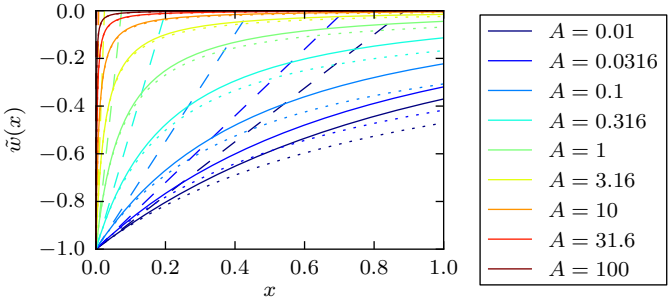


Figure 4.7: Head of the scaled screened potential $\tilde{w}(x)$ as a function of $x = \frac{L}{2}q$ for an isotropic systems with different values of A. The full lines are the values from the full $\tilde{w}(x)$, the dashed lines correspond to a linear expansion and the dotted lines correspond to an approximating function obtained from the first order Taylor expression of $1/\tilde{w}(x)$.

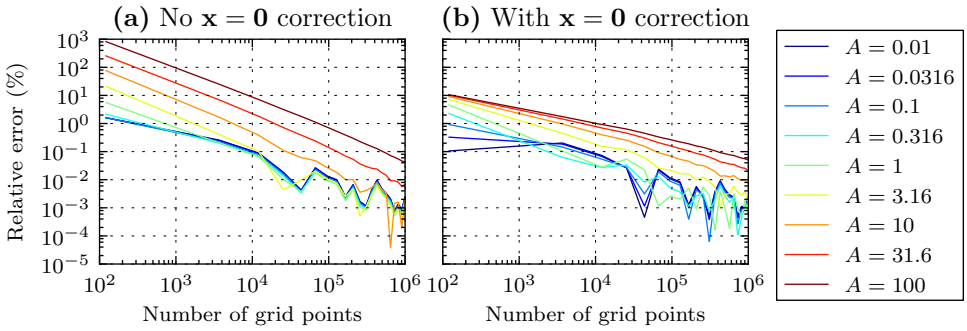


Figure 4.10: Relative numerical error when integrating the scaled screened potential in a disk of unit radius on a regular grid of size $N \times N$ compared to a numerical integration in polar coordinates using an accurate adaptive step size method with absolute tolerance of 1×10^{-10}. (a) Shows the error when no correction to the $\mathbf{x} = \mathbf{0}$ term is used and (b) shows the error when the $\mathbf{x} = \mathbf{0}$ term has been calculated by an analytical expression using an approximate function, Equation (4.96).

may try to find an approximate expression for $\tilde{w}(x)$ that can be integrated analytically for the small region around $x = 0$. The simplest such approach is to just use the first-order Taylor expansion,

$$\tilde{w}(x) \approx -1 + (4\pi A + 1)x, \tag{4.94}$$

which is shown in Figure 4.7. This is seen to only be a good approximation for $x \ll 1$ and it can be shown that going to second order only slightly improves the accuracy for $x \ll 0$, while it fails miserably at values of x above some value (the top point of the parabola). We are instead interested in some function that need not reproduce the actual value of $\tilde{w}$ with perfect precision but whose integral over some finite region comes close to the actual value. Since for the values of A where $\tilde{w}(x)$ is valid, $\tilde{w}(x)$ never becomes exactly 0, it has an inverse $1/\tilde{w}(x)$ which is analytical and differentiable. Its inverse is a smooth approximately linear function and its first order Taylor expansion is thus valid in a quite large region. We can thus approximate $\tilde{w}(x)$ using the first order Taylor expansion of it's inverse,

$$\tilde{w}(x) \approx \frac{-1}{1 + (1 + 4\pi A)x}. \tag{4.95}$$

From Figure 4.7 we see that this approach is much better than the simple linear expansion and it only gets more accurate for higher values of A, where $\tilde{w}$ has more of a $1/x$ behavior. We are also so lucky that using this approximation the integral

Equation (4.93) for isotropic systems has an analytical expression:

$$\begin{aligned}\int_{\mathcal{C}} \tilde{w}(\mathbf{x})\,\mathrm{d}\mathbf{x} &\approx -2\pi \int_0^{x_{\mathrm{rad}}} \frac{x\,\mathrm{d}x}{1+(4\pi A+1)x} \\ &= \frac{\log(4\pi A x_{\mathrm{rad}} + x_{\mathrm{rad}} + 1) - x_{\mathrm{rad}}(4\pi A+1)}{(4\pi A+1)^2}.\end{aligned} \tag{4.96}$$

To exploit this for real systems we use the value calculated by Equation (4.96) for the grid region around $x = 0$, where the radius is chosen so that the area of the disk region mathes that of the grid region, i.e. $x_{\mathrm{rad}} = \sqrt{\Omega_{x=0}/\pi}$. Also, since Equation (4.96) is only valid for isotropic systems we use the rotationally averaged value of $\hat{\mathbf{x}} \cdot \mathsf{A}\hat{\mathbf{x}}$, which with $\mathsf{A} = \begin{pmatrix} A_{xx} & A_{xy} \\ A_{xy} & A_{yy} \end{pmatrix}$ and $\hat{\mathbf{x}}(\phi) = (\cos(\phi), \sin(\phi))$, is

$$\langle A \rangle = \frac{1}{2\pi} \int_0^{2\pi} \hat{\mathbf{x}}(\phi) \cdot \mathsf{A}\hat{\mathbf{x}}(\phi)\,\mathrm{d}\phi = \frac{1}{2}(A_{xx} + A_{yy}). \tag{4.97}$$

In Figure 4.9 we show the results when using this value for the $x = 0$ region. We see that this has a tremendous effect on the convergence for $A > 1$ -- we see that for all values of A the integral is converged to within 10 % accuracy at 10×10 grid points and 1 % at 100×100 grid points and the linear scaling of the error with A is also gone, which is a big improvement.

I have implemented the method just described in GPAW and to test it's performance I have carried out calculations of the GW self-energy of a few 2D systems using different grid sizes for the $\mathbf{k}$-sampling and thus different grid for the evaluation of the $\mathbf{q}$-integral over the Brillouin zone in Equation (4.80). As mentioned earlier in the standard 3D implementation will lead to zero contribution to the integral from the $\mathbf{q} = \mathbf{0}$ when used with the 2D truncated Coulomb potential, so since this has been the standard implementation in most GW codes I will use this as a reference. In order to compare values that are meaningful for typical GW calculations I calculate the correlation self-energy contribution to the quasi-particle energies, Equation (3.67), ie.

$$\Sigma^c_{n\mathbf{k}} = Z\langle n\mathbf{k}|\Sigma^c(\epsilon^{\mathrm{DFT}}_{n\mathbf{k}})|n\mathbf{k}\rangle. \tag{4.98}$$

First I consider the two isotropic systems of monolayer h-BN and 2H-MoS_2 which have quite different dielectric properties. h-BN is a large gap insulator and its dielectric function has a small slope for $\mathbf{q} \to \mathbf{0}$ as seen in Figure 4.5 while MoS_2 has a much steeper slope of the dielectric function. This means that we expect the dynamical part of the screened potential for MoS_2 to be much larger and much sharper at $\mathbf{q} = \mathbf{0}$ than for h-BN. The typical regular grid integration method where we neglect the $\mathbf{q} = \mathbf{0}$ term is then expected to converge somewhat slower for MoS_2 than for h-BN. This is confirmed by my results shown in Figure 4.11 where it is seen that for large $\mathbf{k}$-point samplings the self-energy converges as $1/N_{\mathbf{q}}$, where $N_{\mathbf{q}}$ is the total number of $\mathbf{q}$-points in the grid, just as it is expected. However, h-BN is converged to within about 0.1 eV already at 24×24 $\mathbf{q}$-points, while this is first the case for MoS_2

using at 36×36 $\mathbf{q}$-points. Therefore in general this method is very sensitive to the dielectric properties of the material and one $\mathbf{k}$-point sampling that leads to a certain accuracy for one material may give a worse accuracy for another material. On the other hand, the method described in this section is seen to converge much, much faster, with accuracy within 0.1 eV already by using coarse $\mathbf{k}$-point sampling of 6×6 and basically accuracies in the meV range for $\mathbf{k}$-samlings of 18×18 - something which would probably have taken grids larger than 100×100 using the standard method.

Since the method also explicitly takes any dielectric anisotropy in the $\mathbf{q} \to \mathbf{0}$ limit into account I also test the method on the highly anisotropic system of monolayer black phosphorous (phosphorene). In Figure 4.12(a) the macroscopic dielectric function for phosphorene is shown along to perpendicular high-symmetry directions and it is clear that the initial slope at $\mathbf{q} = \mathbf{0}$ vary quite a lot. In Figure 4.12(b) the convergence of the correlation self-energy with $\mathbf{k}$-point sampling is shown. The method is seen to perform slightly worse for an anisotropic materials like phosphorene than for the isotropic materials of h-BN and MoS_2 for a small number of $\mathbf{k}$-points but for samplings with more than 100 points they perform similarly. I suppose the slightly worse performance at very small samplings is either due to the use of the average slope for the $\mathbf{q} = \mathbf{0}$ term or that the analytical expansion simply fails to describe the anisotropic behavior far out in the Brillouin zone.

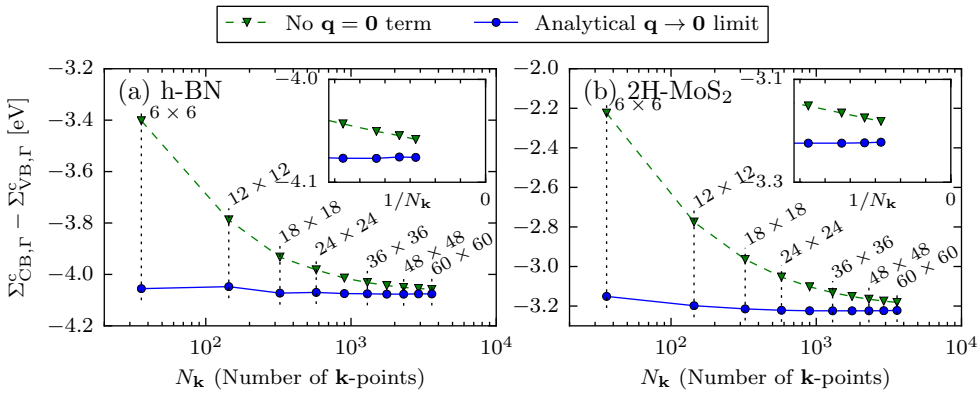
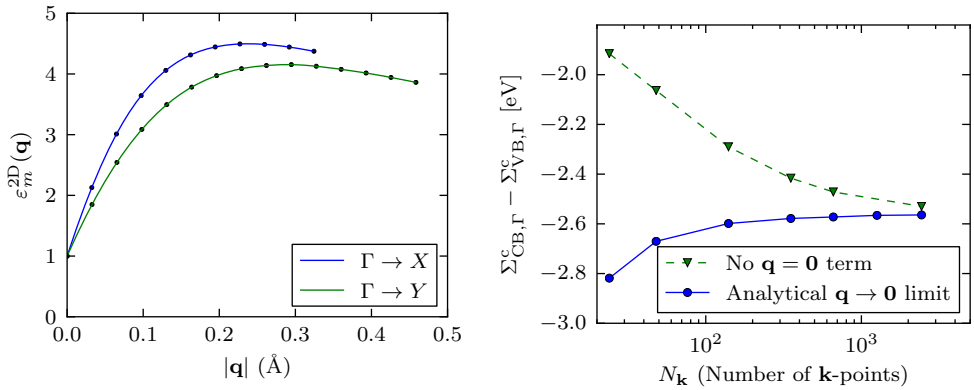


Figure 4.11: Convergence of the correlation self-energy contribution to the quasiparticle band gap at the Γ-point with respect to the $\mathbf{k}$-point grid size. (a) is for h-BN and (b) is for MoS_2. The green triangles correspond to the method, where the $\mathbf{q} = \mathbf{0}$ term in the numerical Brillouin zone integration has been neglected and the blue circles correspond to the method developed in this section.



(a) **(b)**

Figure 4.12: **(a)** Macroscopic dielectric function for phosphorene along the $\Gamma \to X$ and $\Gamma \to Y$ directions, respectively. **(b)** Convergence of the correlation self-energy for phosphorene with respect to $\mathbf{k}$-point sampling.

CHAPTER 5

Transition metal dichalcogenides and -oxides

Some of the first two-dimensional materials to be synthesized besides graphene and h-BN were single layers of the transition-metal dichalcogenides MoS_2 and WS_2. These are part of a group of various well studied bulk materials with the chemical formula MX_2, with M a transition metal and X a chalcogen or oxygen atom, that are known to have a wide variety of properties like unconventional superconductivity[113], exceptional lubrication[67] and possibility to catalyze the hydrogen evolution reacion[39]. The transition-metal dichalcogenides and -oxides come in a variety of crystal structures but an especially interesting one is a layered hexagonal lattice structure where the transition metal occupies a trigonal prismatic coordination sphere and is thus surrounded by six chalcogen or oxygen ligands. The structure comes in two principal variants: 2H, which has the D_{6h} point-group symmetry, and 1T, which has the D_{3d} point-group symmetry, see Figure 5.1. Each atom in the unit cell inherits the hexagonal crystal structure, but in the 2H structure the chalcogen or oxygen atoms share the same in-plane coordinate and seen from above it forms a honeycomb-like structure, while in the 1T structure the chalcogen/oxygen atoms are located in two different high-symmetry positions. Due to the weak interlayer interaction it was shown that it was possible to synthesize single stable layers of some of these materials by mechanical or chemical exfoliation[6, 16].

These structures were early on given much interest, but initial studies were limitied to a few well-known materials. Based on initial experimental success it was hypothesized that a whole family of similar materials might be realizable. A study on a large group of transition metals paired with all possible combinations of oxygen and chalcogens in these structure were originally carried out in Ref. 4 in order to find stable 2D compounds. In total they considered 11 transition metals and the four group 16 elements of O, S, Se and Te yielding a total of 88 structures of which they found 52 stable compounds. While their stability study was extensive using both structure optimization, calculation of phonon spectra and finite temperature molec-

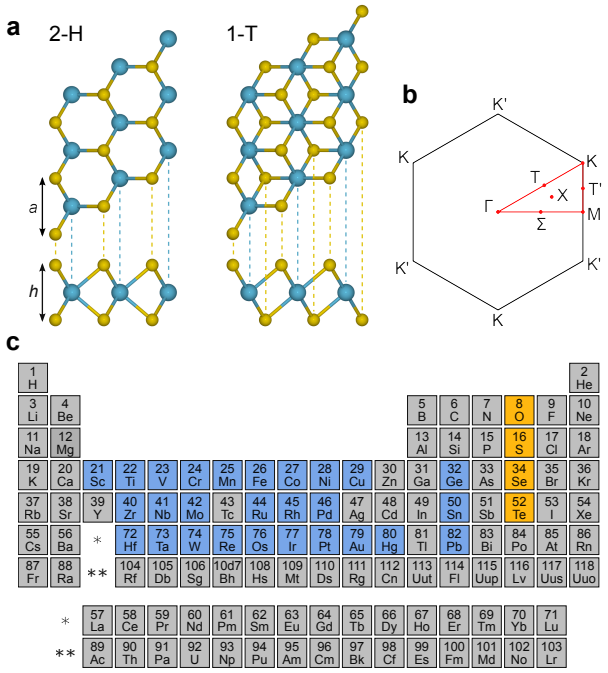


Figure 5.1: **(a)** Structure of the 2H and 1T phases of the transition metal dichalcogenides and -oxides. a denotes the in-plane lattice constant and h is the vertical distance between the chalcogen/oxygen atoms. **(b)** In-plane Brillouin zone of the hexagonal lattice structure with labels of the high-symmetry and general **k**-points. **(c)** Periodic table of the elements with the 27 metals and metalloids considered in this study colored blue and the 4 group 16 elements colored yellow.

ular dynamics, we felt that their report on the electronic structure of the materials was inadequate. For instance while they actually provided electronic band gaps calculated with DFT for all stable compounds they only provided *GW* results for a small subset of the materials. And at the same time the procedure for the calculated *GW* band gaps was known to be inaccurate for 2D materials due to the low screening and inadequate convergence of the used **k**-point grid. Therefore we set forth to produce our own study involving an even larger set of potential 2D metal dichalcogenides and -oxides ensuring properly calculated electronic properties. We chose a set of materials that at the same time were under study for their use as catalysts for hydrogen evolution[88, 86] and it included 24 transition metals and 3 metalloid, see Figure 5.1. Some of these compounds were identical to those studied in Ref. 4, but the focus on this project was accurate electronic properties and not stabilities.

5.1 Structure optimization and stability

The first step is as always to get the atomic structure of the materials. Though many of the materials are known to exist in bulk form in nature and their lattice structures have been measured by crystallographic methods and are listed in various databases. However, we are interested in the properties of the monolayers, and due to the lack of neighboring layers, their crystal structure might be somewhat different than their bulk counterparts. An extensive stability study on some of the transition metal dichalcogenides and oxides had already been carried out in Ref. [4], and we could in principle have used the optimized geometries from that study in our calculations, but since the we considered a somewhat larger class of materials, we decided to use the same method of structure optimization to make results consistent.

To relax the structures we started with unit cell of the somewhat arbitrarily chosen in-plane lattice constant of 3.1 Å and a vertical distance between the oxygen or chalchogen atoms of 2.5 Å and a unit cell height corresponding to an interlayer distance of 20 Å. The relaxation calculations were doen with fully spin-polarized DFT using the PBE XC-functional, a plane wave basis with an energy cut-off of 750 eV and a Γ-centered **k**-point grid of size $18 \times 18 \times 1$. The unit cell was relaxed, keeping the relative positions of the atoms fixed, using the ASE's StrainFilter object until the maximum component of the stress tensor multiplied by the unit cell volume was below 0.01 eV. Hereafter the atomic positions were relaxed undtil the maximum force on any atom was below 0.01 eV/Å. These two steps were then repeated until the optimization criteria for both the unit cell and atomic positions were met.

With the optimized structures we can then get out the total ground state energy, E^{tot}. This can be used to calculate the heat of formation,

$$\Delta H(\mathrm{MX_2}) = E_{\mathrm{tot}}(\mathrm{MX_2}) - E_{\mathrm{ref}}(\mathrm{M}) - 2E_{\mathrm{ref}}(\mathrm{X}), \tag{5.1}$$

where $E_{\mathrm{tot}}(\mathrm{MX_2})$ is the PBE total energy of the monolayer while $E_{\mathrm{ref}}(\mathrm{M})$ and $E_{\mathrm{ref}}(\mathrm{X})$ are the reference energies of the metal and chalcogen/oxygen (X), respectively. The reference energies are the total energies of the individual elements in their reference state at room temperature, i.e. for oxygen and chalcogen that is their molecular gaseous phase and for the metals, it is their typical pure solid phase. The reference energies were likewise calculated with DFT using PBE XC-functional in a manner described in Refs. [88, 86]. Heat of formation calculated this way has however shown not always be that close to the experimentally measured values and there are various reasons to this. One simple, but maybe not so satisfying, solution is to use elemental phase reference energies that are obtained by fitting calculated heat of formations of a large set of materials to their experimentally measured values[111, 85]. This method has proven to in general give values for the heat of formation that are closer to experiments, though the underlying reasons for the discrepancies are not considered.

Figure 5.2 shows the calculated heat of formations for all 216 compounds calculated both with PBE reference energies and the fitted elemental phase reference energies (FERE). In general, the oxides have the highest stability followed by the sulphides, selenides, and tellurides in that order. Furthermore, the stability decreases

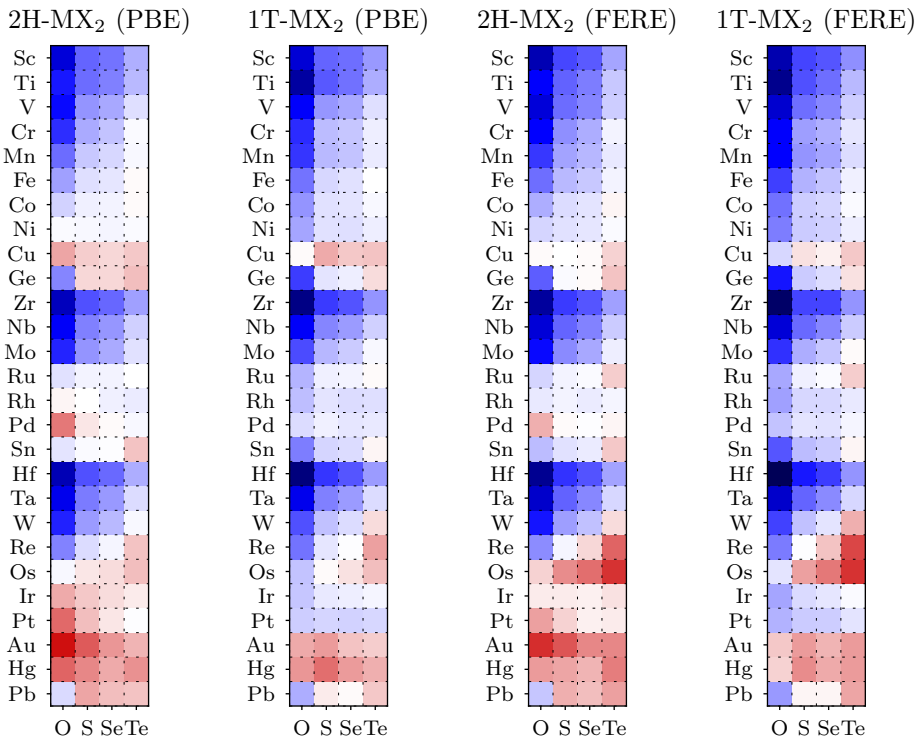


Figure 5.2: Calculated heat of formation for all monolayers in the 2H and 1T phases. In general, the oxides have the highest stability followed by the sulphides, selenides, and tellurides in that order. Furthermore, the stability decreases as the metal ion goes through the transition metal series. 369.0pt

as the metal ion goes through the transition metal series. We see that most of the materials have negative heat of formation but filtering out materials with a heat of formation above 0.1 eV, so that even a few materials that may be energetically unfavorable are included for the benefit of the doubt, reduces the original 216 compounds to a subset of 171. For comparison with previous studies we note that the 52 monolayer MX2 compounds found to be stable based on the LDA calculations of Ref. [4] form a subset of the stable materials identified in the present work. From the structure optimization we obtain the equilibrium in-plane lattice constant a and layer thickness h calculated as the vertial distance between the oxygen/chalcogen atoms, see Figure 5.1. We can also get the DFT band gap which determines if the material metallic or insulating and also the magnetic properties can be estimated. Out of the 171 compounds with near negative heat of formation 57 are found to be semiconduct-

ing and their properties found from the structure optimization are listed in Table 5.1.

While the heat of formation is a natural descriptor for whether the material will be possible to synthesize a negative heat of formation is not a sufficient condition for the stability of a material. The material may be mechanical unstable and thus like to form another lower energy structure. It is possible to carry out a more rigourous stability study considering both phonon spectra and investigating the high temperature stability by molecular dynamics as was done in Ref. [4] and one could also take into account more reference phases[14] but this has not been pursued in this study. Instead we can try and infer the possible stable materials from those already existing in bulk form in nature. Such a study has been done[66] by searching the Inorganic Crystal Structure Database (ICSD) for known layered bulk materials and 46 existing TMDs were identified in this way. However we were not able to identify from this study whether the found structures came in the 1T or 2H phase, but assuming them to be equally stable, which according to Figure 5.2 appears to be a reasonable assumption, 76 out of the 171 materials with negative heat of formation, are already known as layered bulk materials. The materials known in bulk phase are labelled by an asterix (*) in Table 5.1. However, none of the transition metal oxides considered in this study were found in bulk form in the database search. This could indicate that they could be very hard to synthesize. It is very likely that there exists non-layered bulk phases of higher stability. On the other hand bulk structures of TMO layers intercalated with metal ions have been known to exist for a long time[21] and are recently studied for their unconventional superconductivity[113] and possible use in batteries[35]. It might be that monolayer TMOs can be meta-stable and stabilized by interaction with a substrate or encapsulation between other monolayers.

5.2 Electronic structure

With the ground state properties for a large variety of materials determines we now turn our attention to the somewhat smaller group of the 51 non-magnetic semiconductors. While the metals and semimetal magnets are interesting in their own right, conventional semiconductors are of most current interest for optical and electronic applications. I therefore started by calculating the Kohn-Sham band structure for all compounds using both the LDA, PBE and GLLB-SC[59] xc-functionals. As has been shown previously, due to the fact that the 2H structure lacks an inversion centre, some of these materials, notably MoS_2 and WS_2, show very large spin-orbit splittings at the valence band maximum. While GPAW does not currently support spinors and the inclusion of the spin-orbit term in the self-consistent cycle it is possible to calculate the spin-orbit effects as a post-processing step by calculating the spin-orbit interaction as a pertubation to the self-consistently determined Kohn-Sham Hamiltonian. Since the spin-orbit interaction is neglible far away from the cores only the contribution from within the PAW spheres are included. The implementation of the spin-orbit interaction in GPAW will be described elsewhere but is similar to that of

name	a (Å)	h (Å)	$E_\mathrm{f}^\mathrm{PBE}$ (eV)	$E_\mathrm{f}^\mathrm{FERE}$ (eV)	μ (μ_B)
2H-CrO_2	2.63	2.34	−1.64	−1.99	0.0
2H-CrS_2*	3.05	2.95	−0.662	−0.892	0.0
2H-$CrSe_2$*	3.21	3.15	−0.474	−0.65	0.0
2H-$CrTe_2$*	3.47	3.41	−0.051	−0.104	0.0
2H-GeO_2	2.81	2.32	−0.969	−1.28	0.0
1T-GeO_2	2.9	1.96	−1.53	−1.84	0.0
1T-GeS_2	3.44	2.8	−0.222	−0.416	0.0
2H-HfO_2	3.12	2.34	−2.71	−3.08	0.0
1T-HfO_2	3.25	1.95	−3.27	−3.64	0.0
2H-HfS_2*	3.54	3.14	−1.37	−1.62	0.0
1T-HfS_2*	3.65	2.9	−1.59	−1.83	0.0
2H-$HfSe_2$*	3.68	3.36	−1.17	−1.36	0.0
1T-$HfSe_2$*	3.77	3.16	−1.34	−1.53	0.0
2H-$HfTe_2$*	3.91	3.7	−0.656	−0.723	0.0
1T-MnO_2	2.89	1.93	−1.58	−2	3.0
2H-MoO_2	2.82	2.45	−1.73	−1.94	0.0
2H-MoS_2*	3.18	3.13	−0.842	−0.93	0.0
2H-$MoSe_2$*	3.32	3.34	−0.663	−0.698	0.0
2H-$MoTe_2$*	3.55	3.61	−0.237	−0.149	0.0
1T-NiO_2	2.84	1.91	−0.716	−1.01	0.0
1T-NiS_2	3.35	2.35	−0.248	−0.424	0.0
1T-$NiSe_2$	3.54	2.49	−0.251	−0.374	0.0
1T-PbO_2	3.39	2.14	−0.641	−0.8	0.0
1T-PbS_2	3.85	3.09	0.069	0.031	0.0
1T-PdO_2	3.09	1.96	−0.272	−0.482	0.0
1T-PdS_2*	3.55	2.49	−0.125	−0.214	0.0
1T-$PdSe_2$*	3.73	2.63	−0.206	−0.242	0.0
1T-$PdTe_2$*	4.02	2.76	−0.177	−0.09	0.0
1T-PtO_2	3.14	1.9	−0.405	−0.612	0.0
1T-PtS_2*	3.57	2.46	−0.332	−0.418	0.0
1T-$PtSe_2$*	3.75	2.62	−0.364	−0.397	0.0
1T-$PtTe_2$*	4.02	2.77	−0.321	−0.23	0.0
2H-ScO_2	3.22	2.07	−2.37	−2.74	1.0
2H-ScS_2	3.79	2.72	−1.21	−1.46	1.0
2H-$ScSe_2$	3.95	2.94	−1.1	−1.29	1.0
2H-SnO_2	3.09	2.46	−0.225	−0.54	0.0
1T-SnO_2	3.22	2	−1.01	−1.33	0.0
2H-SnS_2*	3.61	3.23	−0.048	−0.241	0.0
1T-SnS_2*	3.7	2.96	−0.333	−0.527	0.0
1T-$SnSe_2$*	3.86	3.19	−0.285	−0.425	0.0
2H-TiO_2	2.88	2.26	−1.83	−2.02	0.0
1T-TiO_2	2.99	1.94	−2.91	−3.1	0.0
2H-TiS_2*	3.34	3.02	−1.16	−1.23	0.0
2H-$TiSe_2$*	3.49	3.24	−1	−1.02	0.0
2H-$TiTe_2$*	3.74	3.58	−0.544	−0.441	0.0
2H-VSe_2*	3.34	3.2	−0.699	−0.956	1.0
2H-VTe_2*	3.6	3.5	−0.263	−0.397	1.0
2H-WO_2	2.83	2.48	−1.74	−1.85	0.0
2H-WS_2*	3.19	3.15	−0.783	−0.776	0.0
2H-WSe_2*	3.32	3.36	−0.547	−0.487	0.0
2H-ZrO_2	3.14	2.33	−2.65	−2.96	0.0
1T-ZrO_2	3.26	1.93	−3.18	−3.49	0.0
2H-ZrS_2*	3.57	3.14	−1.37	−1.55	0.0
1T-ZrS_2*	3.68	2.9	−1.55	−1.47	0.0
2H-$ZrSe_2$*	3.7	3.37	−1.2	−1.33	0.0
1T-$ZrSe_2$*	3.79	3.16	−1.34	−1.47	0.0
2H-$ZrTe_2$*	3.92	3.73	−0.739	−0.746	0.0

Table 5.1: Ground state properties of the 57 semiconductors including in-plane lattice constant a, layer thickness h, heats of formation E_f calculated with PBE and using Fitted Elemental Phase Reference energies (FERE), and magnetic moments μ. Materials found to exist in bulk in the ICSD are marked with an *[66].

Refs. 40 and 17. Since the implementation in GPAW was rather new I decided to test the result by comparting to other DFT software packages. In Figure 5.3 I show the band structure for 2H-$MoSe_2$ calculated using DFT with the LDA xc-functional and spin-orbit interaction included from three different software packages: GPAW, ELK[18] and QuantumEspresso[25]. Even though ELK is an all-electron code and QuantumEspresso uses a fully self-consistent unconstrained non-collinear spin implementation, the end results are very similar to those obtained from GPAW. We checked the differences for a total of 10 materials and the agreement between the different DFT packages were in general less than 0.02 eV. We therefore believe that the spin-orbit interaction in GPAW gives results that are of similar accuracy of other well-established codes.

5.3 GW quasiparticle energies

As discussed in Section 2.3 DFT does in general not yield accurate band gaps and ionization potentials. One way to improve this is to use a xc-functional that allows for the calculation of the derivative discontinuity such as the GLLB-SC functional, see Section 2.3.1. While this has also been done we also want to compare with the reliable method of the *GW* approximation described in Section 3.5. I have therefore calculated the full quasiparticle band structure using the G_0W_0 approximation implemented in GPAW as described in Section 3.6. For obtaining the non-interacting polarizability and Green's function G_0 I used LDA wave functions obtained from an exact diagonalization of the Kohn-Sham Hamiltonian with a plane wave cut-off of 600 eV and 30×30 k-points. The dense **k**-point sampling was needed in order to obtain

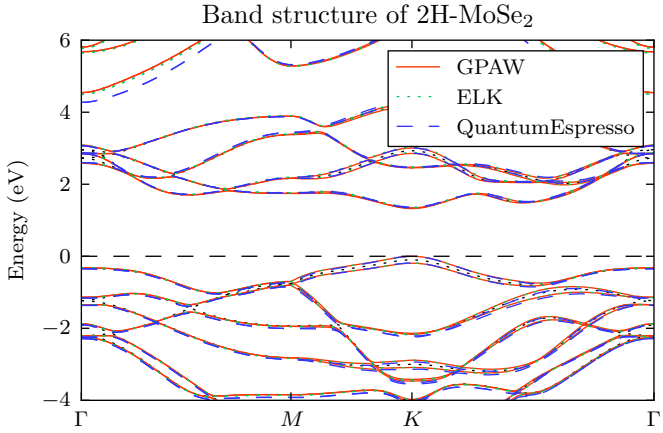


Figure 5.3: Kohn-Sham band structure with spin-orbit interaction of 2H-$MoSe_2$ calculated using three different DFT packages, all using the LDA XC functional.

results that were adequately converged since no $\mathbf{q} = 0$ correction to the self-energy integration as described in had been implemented at the time when these calculations were performed. Also it should be noted that instead of the 2D truncated Coulomb potential given in Equation (4.24) another similar truncated Coulomb potential was used, namely one that also truncates the interaction in-plane at the Wigner-Seitz cell[112]. The same truncated potential had been used with great success for the exchange self-energy so it was therefore easy to implement for the correlation self-energy as well. However, this method does not properly deal with the decreased screening in the long wavelength limit, $\mathbf{q} \to \mathbf{0}$ and thus has the same convergence behaviour as the standard 2D truncated Coulomb potential where the $\mathbf{q} = \mathbf{0}$ contribution to the self-energy integral has been left out. From Figures 4.11 and 4.12(b) it is therefore estimated that the G_0W_0 band gaps calculated using this method will only be converged to within 0.1 eV. On the other hand, the use of a dense $\mathbf{k}$-point sampling has allowed for high resolution band structures to be obtained so that it has been possible to calculate effective masses without the need to rely on other techniques like Wannier interpolation.

The plane wave cut-off used to construct the screened interaction and self-energy was varied between 150 eV and 500 eV and extrapolated to infinite cut-off energy as described below. For all calculations, the number of unoccupied orbitals used to construct the screened interaction and GW self-energy was set equal to the number of plane waves. The frequency dependence was represented on a non-linear grid from 0 eV to the energy of the highest transition included in the basis with a gradually increasing grid spacing starting at 0.1 eV and reaching 0.2 eV at $\omega = 15$ eV. The frequency grid typically contained 300 to 350 grid points. The PAW potentials applied in this work include semi-core states, i.e. atomic states down to at least 1 Hartree below vacuum, while deeper lying states are included in the frozen core. The frozen core states are included in the exchange contribution to the GW corrections.

The *GW* self-energy contains a sum over all occupied and unoccupied bands as well as a sum over all $\mathbf{G}$-vectors. These sums are in practice truncated at some finite number and this naturally leads to a truncation error. The error in the self-energy directly leads to an error in the QP energies. Often one is interested in relative energies like the band gap and in some cases the error in both values cancel. For this reason, many had previously assumed that finite plane-wave cut-off and number of bands could be considered to yield converged results. In general this is not the case. In Figure 5.4 we see that for MoS_2 both the valence band and conduction band seem to scale with the same slope with increasing number of plane waves included which means that the band gap is already converged with a cut-off of 150 eV. This is on the other hand not the case for 2H-TiO_2, where the valence band extrapolated to infinite cut-off changes much more than the conduction band. This leads to a slow convergence of the band gap with planewave cut-off. However as we clearly see, the error appears to scale linearly with $1/N_{\mathbf{G}}$, an observation that has also recently been explained theoretically[52]. This means that it is possible to extrapolate the energies to infinite cut-off as long as we know that the QP energies are in the linear regime in $1/N_{\mathbf{G}}$[117, 43, 49]. In practice this was done by calculating the QP energies using an

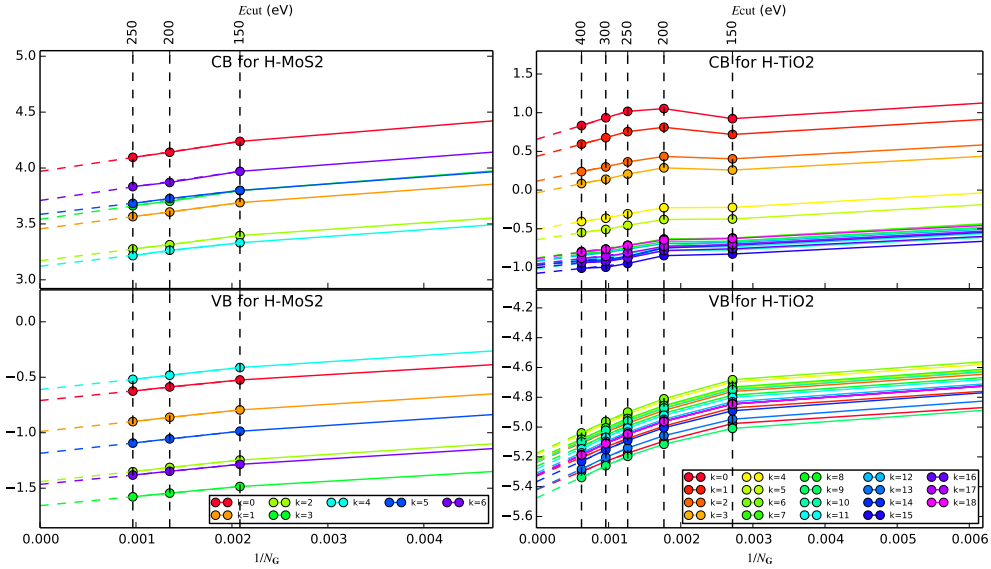


Figure 5.4: QP energies for H-MoS_2 and H-TiO_2 for the valence band and conduction band at various **k**-points calculated using different number plane waves, $N_{\mathbf{G}}$, determined by the plane wave energy cut-off E_{cut}.

initial plane wave cut-off of 150 eV and then progressively calculating new QP energies using higher and higher cut-offs and plotting the energies as in Figure 5.4. When a satisfactory linear relationship was realized a linear fit was made through the points that showed the linear behavior and the QP energy at $1/N_{\mathbf{G}} = 0$ was extrapolated. For most materials only 3 to 4 points between 150 eV and 250 eV was needed in order to complete the extrapolation, but for a few; most notably Ti and Ni compounds, cut-offs in the range 300 eV to 500 eV were needed to get a linear relationship. The reason for this is most likely due to the localized 3d states of these elements which makes them hard to describe with a planewave basis[3, 107, 72, 63]. Even when using an adequate planewave cut-off the QP energies might be incorrectly determined within the PAW formalism due to an incomplete basis set used for constructing the PAW corrections[52]. A simple method that should improve the accuracy with respect to this last issue is to use norm-conserving PAW setups[52], but this has not been pursued here.

Since the extrapolation procedure is quite computationally expensive if done at 30×30 **k**-points I developed another method for speeding up the calculations. The approach is based on the fact that the convergence with respect to plane waves is independent of the convergence with respect to **k**-points as seen in Figure 5.5. To obtain results converged with respect to both **k**-points and plane waves I have therefore performed the $N_{\mathbf{G}}$-extrapolation for a coarse **k**-point sampling of 12×12 and

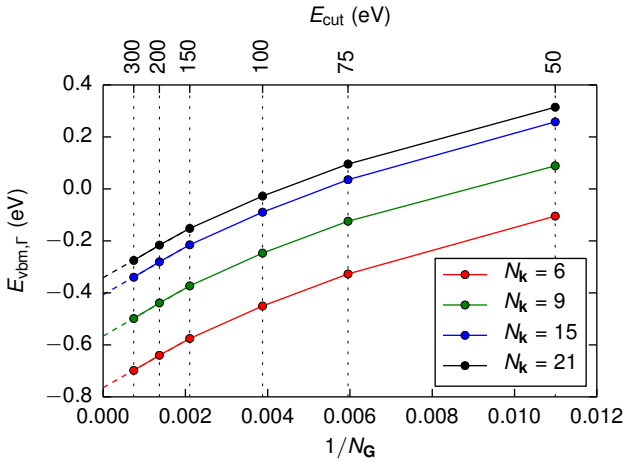


Figure 5.5: G_0W_0 quasi-particle energy of the valence band at the Γ point of monolayer 2H-MoS_2 as a function of $1/N_{\mathbf{G}}$, where $N_{\mathbf{G}}$ is the number of plane waves. The different lines correspond to different $\mathbf{k}$-point samplings ($N_k \times N_k \times 1$). The dashed lines shows the extrapolation to infinite plane wave cut-off.

thereby obtain the so-called extrapolation shifts from the calculation at 150 eV cut-off to the extrapolated values at infinite cut-off. These shifts are then transferred to a calculation done with 30×30 $\mathbf{k}$-points, but in order to do so they have to be interpolated to the fine grid. In Ref. 52 they found that the truncation error, i.e. the extrapolation shifts $\Delta\epsilon_{n\mathbf{k}}$, are basically determined by the quasiparticle state density $\rho_{n\mathbf{k}}$. Therefore we developed an interpolation scheme where the extrapolation shift, $\Delta\epsilon^{\text{fine}}_{n\mathbf{k}}$, for a state in the dense $\mathbf{k}$-point calculation is calculated by a weighted average of the extrapolations shifts, $\Delta\epsilon^{\text{coarse}}_{n\mathbf{k}'}$, from the nearest $\mathbf{k}'$-points in the coarse $\mathbf{k}$-point calculation. The weights used in the average are given by the overlap between the DFT state densities: $w_{n\mathbf{k},n\mathbf{k}'} = \langle \rho^{\text{fine}}_{n\mathbf{k}}, \rho^{\text{coarse}}_{n\mathbf{k}'} \rangle$, where $\rho_{n\mathbf{k}}(\mathbf{r}) = |\psi_{n\mathbf{k}}(\mathbf{r})|^2$. This procedure has been crucial in obtaining QP energies that were both converged with respect to the $\mathbf{k}$-point sampling and plane wave basis. With the method for integrating improving the $\mathbf{k}$-point convergence described in Section 5.3 the requirement for dense $\mathbf{k}$-point samplings is avoided and one can perform the plane wave extrapolation without the need for this interpolation procedure.

5.4 Strain effects on the band structure

While DFT underestimates the band gaps it usually gives quite good predictions for the structural properties. Still it is quite well-known that the PBE xc-functional gives errors on lattice constants in the order of 2 % and at the same time the band gap may also very well depend on the lattice constant. Thus our expensive GW estimates of

the band gap may not be anymore precise than the error from inaccurately determined lattice constants. To investigate this I calculated the change in LDA band gap when the in-plane lattice constants was varied within $\pm 2\,\%$.

From Figure 5.6 we see that a change of the lattice parameter within the considered range can actually produce quite drastic changes in the band gap. For example, in the case of 2H-MoS2 a change in the lattice constant from the PBE value (3.18 Å) to the experimental value (3.16 Å) changes the band gap by around 0.1 eV. It was also found[100] that the LDA gap changes from indirect to direct under 1 % compressive strain. A few other direct gap materials are seen to develop an indirect gap when strained. Thus we conclude that both the size and nature of the band gap of the monolayers can depend delicately on the lattice constant.

From Figure 5.6 we also notice that some materials have a positive change in band gap under tensile strain while other show the opposite trend. To understand this behaviour I analyzed the projected density (see Supplementary information of Ref. 100) of states and found that the materials can be roughly divided into two classes according to the nature of the wave functions around the band gap. For the materials with group 6 metals (Cr, Mo and W), the valence and conduction band states are bonding/anti-bonding combinations of the metal d-states and oxygen/chalcogen p-states and in their equilibrium lattice constant they have direct band gaps. For these materials one finds that increasing the M-X binding distance by stretching weakens the hybridization and reduces the bonding/anti-bonding gap. The other class is TMDs with metals from group 4, 10 and 14 (Ti, Zr, Hf, Ni, Pd, Pt, Ge, Sn, Pb). For these materials, the valence band states have primarily chalcogen p-character while the conduction band is either metal-d (group 4), chalcogen-p (group 10) or metal-s and chalcogen-p (group 14). For these materials the size of the gap is determined by the width of the conduction band and the chalcogen valence band. One then finds that tensile strain will cause the states to become more localized, narrowing each of the bands which will open the gap. We also found that as a consequence of the

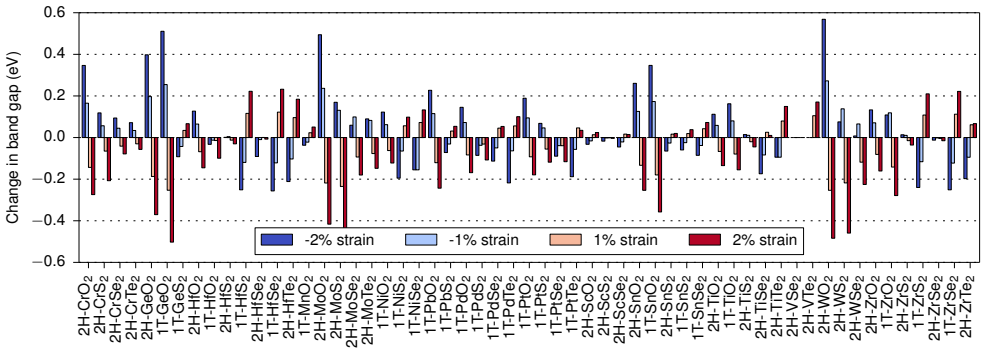


Figure 5.6: Change in the LDA band gap when the in-plane lattice constant is varied between $-2\,\%$ and $2\,\%$.

decoupled bands, these materials also all systematically have indirect band gaps[100]. Recently this has effect has also been studied as a way to determine if the band gap is protected from the defect states[87].

5.5 Effective masses

Effective masses describe the approximate dispersion of the low energy quasiparticle excitations. These states often determine contribute by the largest amount to low power electronic properties and are thus important for establishing electronic device properties. Since the QP energies have been calculated on a dense **k**-point grid it is possible to directly get the real QP effective masses by fitting a parabola around the valence band maximum (VBM) and conduction band minimum (CBM). If $\mathbf{k} = (k_x, k_y)$ is a point in the vicinity of the VBM or CBM denoted by $\mathbf{k}_0$, then the band structure approximately behaves as

$$E(\mathbf{k}) = \frac{\hbar^2}{2m_\mathrm{e}} \tilde{\mathbf{k}}^\mathrm{T} \mathbf{A} \tilde{\mathbf{k}}, \tag{5.2}$$

where $\tilde{\mathbf{k}} = \mathbf{k} - \mathbf{k}_0$ is the in plane k-point measured from the band extremum. The eigenvalues of the matrix $\mathbf{A}$ yield the inverse effective masses in the direction of smallest and largest curvature, i.e. the inverse of $\mathbf{A}$ is the effective mass tensor. If the CBM or VBM is located at one of the high symmetry points of the BZ (the Γ or K point) the effective masses will naturally be isotropic. However, for band extrema located at other points this is generally not the case. To obtain the effective mass tensor I fitted the QP band structure of the 19 nearest points to the VBM and CBM to Equation (5.2) using least-squares.

One thing to note is that in some cases this method gives negative effective masses, which sound contradicting as the curvature should not be able to be negative at an extremum. The reason for this is the fitting method; if the band is highly anisotropic at the extremum so that it is very flat in one direction but has a high dispersion in the perpendicular direction, the fitting algorithm is more likely to yield a parabola which runs thorugh the center of the distribution and comes with negative curvature in the flat direction. I tried countering this by weighting the points closest to extremum higher in the least squares algorithm, but some materials still yielded negative masses. One can effectively think off these as having infinite mass along that direction.

In Figure 5.7(a) and Figure 5.7(b) the effective electron and hole masses along the two directions of highest and lowest curvature are shown. Points falling on the diagonal line correspond to isotropic masses. The effective electron masses lie in the range 0.1 to $10m_\mathrm{e}$ with roughly an equal number being light ($m_\mathrm{e}^* < m_\mathrm{e}$) and heavy ($m_\mathrm{e}^* > m_\mathrm{e}$). Hole masses are similar, although they seem to be slightly larger than for electrons. In accordance with the discussion in section , we see that only the materials with direct gaps (group 6 metals) have both isotropic electron and hole masses. For other materials the masses can be quite anisotropic and we would also expect the masses to depend sensitively on the lattice constant. I would also like to note that I

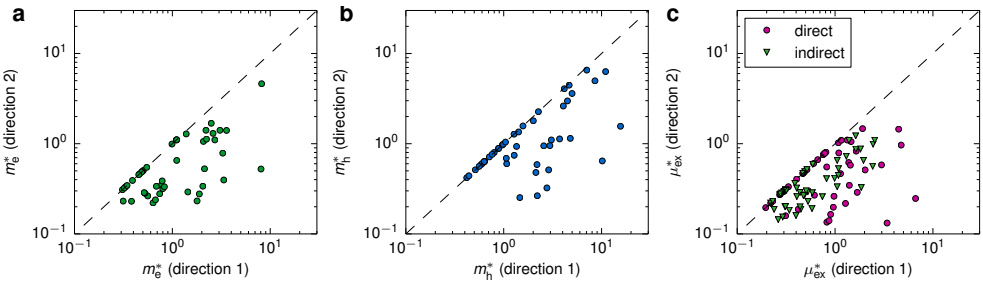


Figure 5.7: Effective electron (**a**), hole (**b**) and exciton (**c**) masses (in units of m_e) along the two principal directions obtained. The masses are calculated from the *GW* band structures including spin-orbit interaction. Points on the dashed lines correspond to isotropic masses.

have only considered the effective masses of the valence band and conduction band. In cases where the VBM or CBM is degenerate and there therefore will be both light and heavy electrons/holes I have only calculated the mass of the heavy charges.

Lastly I have also calculated the exciton mass, which in general is given calculated from the dispersion of the conduction and valence band difference:

$$E_{\mathrm{ex}}(\mathbf{k}) = E_{\mathrm{cb}}(\mathbf{k}) - E_{\mathrm{vb}}(\mathbf{k}) \approx \frac{\hbar^2}{2m_{\mathrm{e}}} \tilde{\mathbf{k}}^{\mathrm{T}} \mathbf{M}_{\mathrm{ex}} \tilde{\mathbf{k}}, \tag{5.3}$$

where $\mathbf{M}_{\mathrm{ex,dir}}$ is the inverse direct exciton mass tensor and $\tilde{\mathbf{k}}$ is point measured from the location of the optical/vertical band gap. Excitons can also form by excitation of an electron from the VBM to the CBM and in this case the (indirect) exciton mass is given by

$$\mu_{\mathrm{ex,ind}}^{-1} = (m_e^*)^{-1} + (m_h^*)^{-1}. \tag{5.4}$$

If the material has a direct gap the direct an indirect mass are necessarily equal: $\mu_{\mathrm{ex,dir}} = \mu_{\mathrm{ex,ind}}$. Comparison of calculated exciton masses can be seen in Figure 5.7(c).

5.6 Absolute band positions

The band gap is the most important propery for many purposes, but for other applications it is also crucial to know the absolute positions of the band edges; to so-called ionization potential and electron affinity. As previously explained these are hard to obtain computationally for bulk materials since the nuclear potential is taken to be periodic whereby the knowledge of the absolute value is lost. While it is possible to infer the values from thick slabs by referring the band energies to the asymptotic value of the Hartree potential in the vacuum region it may still be hampered by the fact

that the Hartree potential depends on the surface dipoles (on both sides of the slab) which makes the problem highly surface dependent and complicates the comparison with experiments. On the other hand, the monolayers in this study do not have those problems and we can thus easily obtain the absolute band positions.

In Figure 5.8 the positions of the valence band maximum (VBM) and conduction band minimum (CBM) relative to the vacuum level are shown for the different oxides and chalcogenides at both the LDA and G_0W_0 level. As a significant part of the GLLB-SC band gap comes from the derivative discontinuity which applies to the fundamental gap rather than the individual band energies, the GLLB-SC eigenvalues cannot directly be used to obtain the absolute band edge positions. For all materials, the effect of the G_0W_0 correction is to shift the conduction band up and the valence band down with respect to the LDA values. In fact, the corrections of the VBM and CBM are rather symmetric meaning that the band gap centre is largely unaffected by the GWcorrection (see below).

It has been suggested that 2D semiconductors could be used for photo-catalytic water splitting. This is mainly motivated by their excellent light absorption, large specific surface area, and readily tuneable electronic properties[124, 109]. The equilibrium potentials for the hydrogen- and oxygen evolution reactions at pH 7 are indicated by dashed green lines in Figure 5.8. Materials with CBM above the standard hydrogen electrode (SHE) at $-4.03\,\mathrm{eV}$ relative to vacuum (at pH 7) , could in principle be used to evolve hydrogen at the cathode of a photo-catalytic water splitting device[116]. Likewise materials with VBM below the oxygen evolution potential (1.23 eV below the SHE) could in principle be used a photo anode in the water splitting reaction. In practice, the CBM/VBM should lie a few tenths of an eV above/below the redox potentials to account for the intrinsic energy barriers of the water-splitting reactions

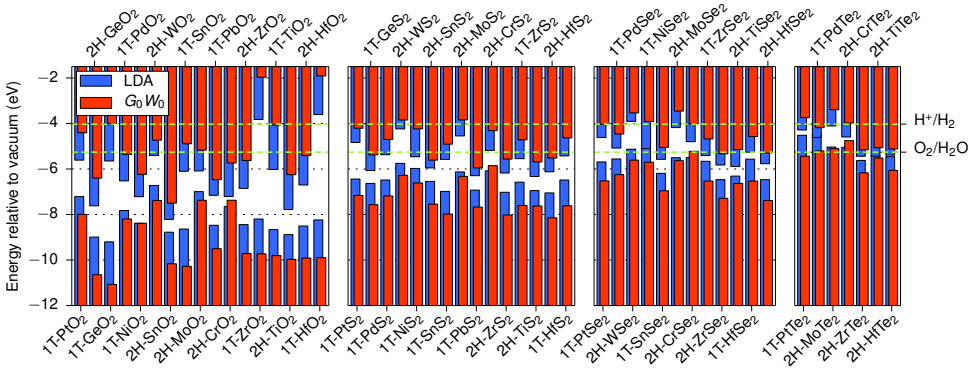


Figure 5.8: Position of the valence band maximum and conduction band minimum relative to the vacuum level (set to zero) for LDA and G_0W_0. In both cases spin-orbit splitting of the bands has been taken into account. The hydrogen and oxygen evolution potentials at pH 7 are shown by green dashed lines.

reactions[78]. As can be seen a number of the TMD monolayers qualify as potential water splitting photo-electrodes based on their energy level positions. A very citical issue, however, is the stability of the materials under the highly oxidizing reaction conditions. A possible solution to this problem could be to protect the photo-electrode from direct contact with the water by a transparent and highly stable thin film, which in practice means an oxide material.

In Ref. [11] a simple empirical relation was observed between the band gap center of a semiconductor and the electronegativities of the constituent atoms,

$$E_{\text{center}} = -[\chi(\text{M})\chi(\text{X})^2]^{1/3}, \tag{5.5}$$

where χ(M) and χ(X) are the electronegativity of the metal and oxygen/chalcogen on the Mulliken scale, respectively. In Figure 5.9 we compare the band gap centers obtained from *GW* with those obtained from LDA and calculated with Eq. Equation (5.5), where experimentally obtained values of the electronegativities[98] have been used. The band gap centers from LDA and *GW* agree quite well showing a mean absolute deviation from the *GW* values of only 0.2 eV. While it is known that the Kohn-Sham band gap center is formally exact within DFT[92], it is somewhat surprising that the LDA performs that well. While the empirical formula is able to describe the qualitative trends of the gap centres the quantitative values deviate significantly from the ab-initio results, with a mean absolute difference from the *GW* result of 0.9 eV and a mean relative deviation of 14 %. We ascribe a large part of this deviation to originate from dipole fields formed due to the positively charged metal ions and negatively charge chalcogens/oxygens which will increase potential outside the monolayer and thereby down shift the bands -- an effect not accounted for by the empirical formula. Since the size of the dipoles is determined by the amount of charge transfer, the deviation between Eq. Equation (5.5) and the ab-initio results is expected to correlate with the difference in electronegativity between the metal atom and chalcogen/oxygen atoms. From the inset of Figure 5.9 we see that this indeed is the case: For materials with larger difference in electronegativity between the atomic species ($\Delta\chi$) the band gap center given by Eq. Equation (5.5) generally deviates more from the *GW* results.

While it is important to establish the intrinsic properties of the 2D materials in their isolated form, practical applications as well as most experimental setups, involve heterostructures where the 2D materials are stacked into van der Waals heterostructure or simply lie on a substrate. In such systems the alignment of the bands at the heterostructure interfaces become crucial. Due to the weak interaction between 2D semiconductors it is reasonable to expect that the band alignment at the interface between two different 2D can be obtained by aligning the band edges of the isolated systems relative to a common vacuum level. This is equivalent to disregarding effects of band hybridization and the formation of interface dipoles due to charge redistribution. Verifying this assumption from first-principles calculations is, however, difficult due to the lattice mismatch between different 2D materials.

To provide an overview of the band edge positions of the 51 monolayers, we show in Fig. 5.10(a) the CBM plotted against the VBM obtained from *GW*. To

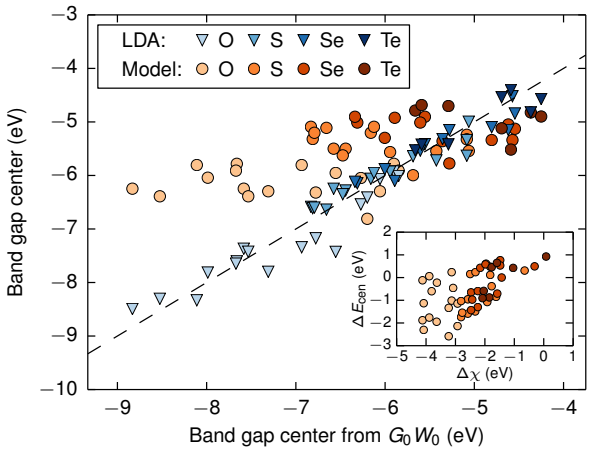


Figure 5.9: Comparison of the absolute band gap centers (relative to vacuum) obtained from GW, LDA, and the empirical formula Eq. Equation (5.5). Inset: The difference in the band gap centers from $GW$and Eq. Equation (5.5), $\Delta E_{\text{cen}} = E_{\text{cen}}^{\text{GW}} - E_{\text{cen}}^{\text{Model}}$, compared to the difference in the electronegativities of the metal and oxygen/chalcogen atom, $\Delta\chi = \chi(\text{M}) - \chi(\text{X})$.

illustrate the use of such a diagram we have highlighted 2H-MoS_2 and indicated regions corresponding to different band alignments with MoS_2. The possible band alignments are: Straddling gap (type I), staggered gap (type II), and broken gap (type III). For many applications, e.g. tandem photovoltaic devices or creation of long lived indirect excitons, a type II band alignment is preferred. We have highlighted a few materials that are expected to form type II band alignment with MoS_2. The detailed band alignments for these materials are shown in Fig. 5.10(b).

5.7 Exciton binding energies

One of the most characteristic features of atomically thin 2D semiconductors is the large binding energy of excitons[120, 99, 53]. The reason for this is the reduced screening due to the lower dimension which yields a stronger attraction between electrons and holes (see discussion in previous section). The conventional method for calculating exciton binding energies from first principles is the Bethe-Salpeter equation (BSE). The BSE is computationally highly demanding and not suited for large-scale studies like the present. Instead we use a recently developed 2D Mott-Wannier model for excitons that only needs the exciton effective mass and the quasi-2D dielectric function as input. In real space the model takes the form of a 2D

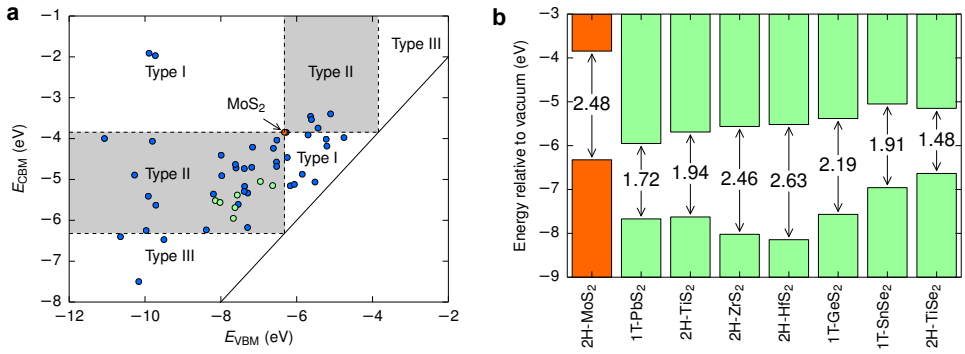


Figure 5.10: Band alignment diagram. (**a**) Conduction band minimum E_{CBM} plotted against the valence band maximum E_{VBM} for the 51 monolayers. The band edges relative to vacuum are obtained from GW. As an example we have highlighted 2H-MoS2 (orange dot) and indicated the regions corresponding to the different types of band alignment: Straddling gap (type I), staggered gap (type II), and broken gap (type III). A few selected materials that will form type-II heterostructures with MoS2 have been highlighted in green. (**b**) Absolute band edge positions and band gaps of 2H-MoS2 and the selected materials highlighted in (**a**).

Schrödinger equation,

$$\left[-\frac{1}{2\mu_{\mathrm{ex}}}\nabla^2_{\mathrm{2D}} + W(\mathbf{r})\right]\psi(\mathbf{r}) = E_{\mathrm{b}}\psi(\mathbf{r}), \tag{5.6}$$

where μ_{ex} is the effective exciton mass and $W(r)$ is the $1/r$ Coulomb interaction between the electron and the hole screened by the non-local $\epsilon^{\mathrm{2D}}_{\mathrm{M}}$. The model has been benchmarked against full BSE calculations for 2H-MoS_2 and 2H-WS_2 and the results were found to deviate by less than 0.1 eV.

The four basic assumptions behind the Mott-Wannier exciton model are: (i) Isotropic exciton masses, (ii) parabolic band structures close to the fundamental gap, (iii) the exciton is well described by transitions between the valence and conduction band only, and (iv) the valence and conduction band wave functions are uniformly distributed over the layer, i.e. their profile along z can be approximated by a step function. While the dielectric functions were found to be very nearly isotropic for all materials, this is not the case for the exciton masses, see Figure 5.7 (c). While it is possible to modify the model to allow for anisotropic masses we here limit ourselves to the materials with isotropic exciton masses. The exciton binding energies obtained from the model are shown as the dark region on the top of the bars in Figure 5.11. The total height of the bar represents the GWcalculated QP gap. For direct (indirect) band gap materials we have used the direct (indirect) exciton mass in the model.

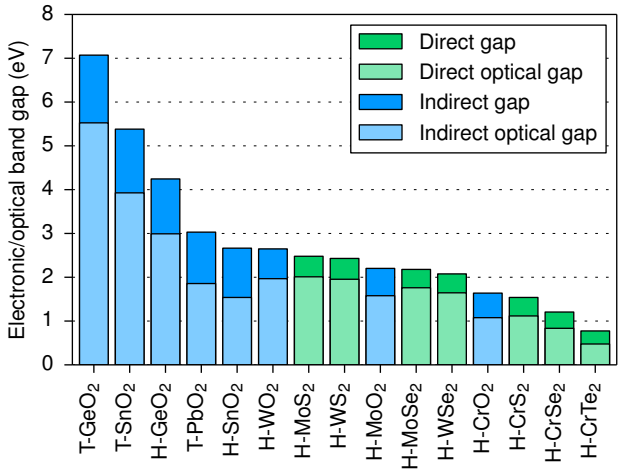


Figure 5.11: *GW* band gaps (total bar height) and exciton binding energies (darker topmost part of the bar). The exciton binding energy was obtained from a quasi-2D Mott-Wannier model. Only materials with isotropic exciton masses are shown. The green and blue bars refer to indirect and direct band gaps and excitons, respectively.

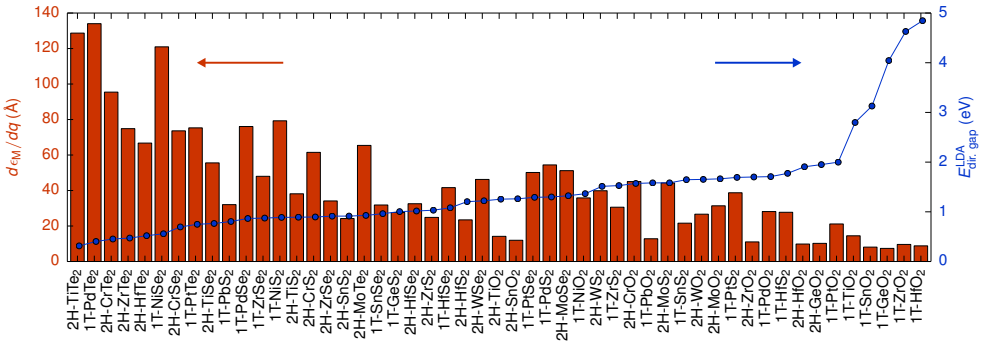


Figure 5.12: The slope of the static quasi-2D dielectric function, $\epsilon^{2D}_{M}(q)$, evaluated at $q = 0$. The materials are ordered according to their LDA direct band gap.

In accordance with earlier experimental and theoretical studies we find strong exciton binding energies on the order of 20-30% of the band gap. In general materials with larger QP band gaps have more strongly bound excitons. This follows from the correlation between the size of the band gap and the dielectric function in Figure 5.12: Larger band gap implies a smaller dielectric function and thus a stronger electron-hole interaction. In Table 5.2 we compare our calculated exciton binding energies with optical data from experiments. We find good agreement for MoS2, MoSe2 and WSe2

while the agreement is less satisfactory for MoTe2 and WS2. It should be noted, however, that the experimental exciton binding energy for MoTe2 was obtained as the difference between the calculated GW band gap and the position of the optical photoluminiscence peak. Thus inaccuracies in the GW band gap as well as substrate effects on the measured photoluminescence peak could explain the disagreement.

Table 5.2: Exciton binding energies in eV calculated from the Mott-Wannier model compared to experimental values.

name	E_b (model)	E_b (exp.)
2H-MoS2	0.47	0.55[53]
2H-MoSe2	0.42	0.5[53]
2H-MoTe2	0.36	0.6[104][a]
2H-WS2	0.48	0.66[123], 0.71[127]
2H-WSe2	0.43	0.38[53], 0.37[36]

[a] The exciton binding energy is obtained from subtracting the energy of the measured exciton photoluminiscence peak from our calculated GW band gap.

CHAPTER 6

Conclusion

In this thesis I have looked at the issues arising when calculating properties of 2D materials using first principles methods. It has been shown that advanced methods like the *GW* approximation, which is usually regarded at providing state-of-the-art benchmarks results, can be decieving when applied at unchartred territory. Its computational requirements are so large that it can be tempting to skip a careful convergence investigation and apply parameters that have been working for other systems. Aware of this we started an ambitious project of charting nearly all the electronic properties of all the transition metal dichalcogenides and -oxides carefully trying to avoid all the pitfalls known to us. This has resulted in a database of more than 171 2D materials for which properties like lattice constant, magnetic moments, heats of formation, electronic band gaps and other electronic properties are known. The accurate G_0W_0 based calculations of the quasiparticle band gap and absolute band positions allows one to predict the band alignment of 2D heterostructures and thus the basic functionality of 2D semiconductor devices. We have also calculated the effective masses and two-dimensional dielectric function and used these quantities in a Mott-Wannier model to estimate the exciton binding energies, which have shown good agreement with experiments. While the results for each material might be interesting in its own right the force has also been the structure of the collection, which have already inspired follow-up studies[2, 83, 87].

The challenge of doing converged G_0W_0 calculations for more than 50 compounds turned out to be harder than initially thought. Besides the complications of storing large amounts of complex data in an easily accessible way the sheer computational requirements forced us to implement new parallelization schemes in the code and develop the plane wave extrapolation/k-point interpolation procedure. In the process I have learned that it is not only important to know how to apply a method but often it is required that you know exactly how the almost every part of the machinery works. This lead me to realize that instead of a convergence by brute force approach may be circumvented by investigating the cause of the problem and use physical arguments to avoid it alltogether. This lead to the development of the analytical treatment of the long wavelength limit of the screened potential. While maybe not a completely new idea, it is still something which will make future *GW* calculations of 2D materials much less of a hassle.

Bibliography

[1] Stephen L. Adler. "Quantum Theory of the Dielectric Constant in Real Solids". In: *Physical Review* 126.2 (April 15, 1962), pages 413–420.

[2] Kirsten Andersen, Simone Latini, and Kristian S. Thygesen. "Dielectric Genome of van der Waals Heterostructures". In: *Nano Letters* 15.7 (July 8, 2015), pages 4616–4621.

[3] F. Aryasetiawan and O. Gunnarsson. "3*d* semicore states in ZnSe, GaAs, and Ge". In: *Physical Review B* 54.24 (December 15, 1996), pages 17564–17567.

[4] C. Ataca, H. Şahin, and S. Ciraci. "Stable, Single-Layer MX2 Transition-Metal Oxides and Dichalcogenides in a Honeycomb-Like Structure". In: *The Journal of Physical Chemistry C* 116.16 (April 26, 2012), pages 8983–8999.

[5] Wilfried G. Aulbur, Lars Jönsson, and John W. Wilkins. "Quasiparticle Calculations in Solids". In: *Solid State Physics*. Edited by Henry Ehrenreich and Frans Spaepen. Volume Volume 54. Academic Press, 1999, pages 1–218.

[6] Anthony Ayari et al. "Realization and electrical characterization of ultrathin crystals of layered transition-metal dichalcogenides". In: *Journal of Applied Physics* 101.1 (January 1, 2007), page 014507.

[7] Wenzhong Bao et al. "High mobility ambipolar MoS2 field-effect transistors: Substrate and dielectric effects". In: *Applied Physics Letters* 102.4 (January 28, 2013), page 042104.

[8] A. D. Becke. "Density-functional exchange-energy approximation with correct asymptotic behavior". In: *Physical Review A* 38.6 (September 1, 1988), pages 3098–3100.

[9] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Annalen der Physik* 389.20 (January 1, 1927), pages 457–484.

[10] Henrik Bruus and Karsten Flensberg. *Many-Body Quantum Theory in Condensed Matter Physics: An Introduction*. OUP Oxford, September 2, 2004. 456 pages.

[11] M. A. Butler and D. S. Ginley. "Prediction of Flatband Potentials at Semiconductor-Electrolyte Interfaces from Atomic Electronegativities". In: *Journal of The Electrochemical Society* 125.2 (February 1, 1978), pages 228–232.

[12] Pierre Carrier, Stefan Rohra, and Andreas Görling. “General treatment of the singularities in Hartree-Fock and exact-exchange Kohn-Sham methods for solids”. In: *Physical Review B* 75.20 (May 29, 2007), page 205126.

[13] Mark E. Casida. “Generalization of the optimized-effective-potential model to include electron correlation: A variational derivation of the Sham-Schlüter equation for the exact exchange-correlation potential”. In: *Physical Review A* 51.3 (March 1, 1995), pages 2005–2013.

[14] Ivano E. Castelli et al. “Computational screening of perovskite metal oxides for optimal solar light capture”. In: *Energy & Environmental Science* 5.2 (January 25, 2012), pages 5814–5819.

[15] Nasreen G. Chopra et al. “Boron Nitride Nanotubes”. In: *Science* 269.5226 (August 18, 1995), pages 966–967.

[16] Jonathan N. Coleman et al. “Two-Dimensional Nanosheets Produced by Liquid Exfoliation of Layered Materials”. In: *Science* 331.6017 (February 4, 2011), pages 568–571.

[17] Andrea Dal Corso. “Projector augmented wave method with spin-orbit coupling: Applications to simple solids and zincblende-type semiconductors”. In: *Physical Review B* 86.8 (August 23, 2012), page 085135.

[18] Kay Dewhurst et al. *ELK code*. Version 2.3.22. visited February 2, 2015. May 22, 2014.

[19] G. E. Engel and Behnam Farid. “Generalized plasmon-pole model and plasmon band structures of crystals”. In: *Physical Review B* 47.23 (June 15, 1993), pages 15931–15934.

[20] V. Fock. “Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems”. In: *Zeitschrift für Physik* 61.1 (January 1930), pages 126–148.

[21] Claude Fouassier, Claude Delmas, and Paul Hagenmuller. “Evolution structurale et proprietes physiques des phases AXMO2 (A = Na, K; M = Cr, Mn, Co) (x 1)”. In: *Materials Research Bulletin* 10.6 (June 1, 1975), pages 443–449.

[22] Guanhui Gao et al. “Artificially Stacked Atomic Layers: Toward New van der Waals Solids”. In: *Nano Letters* 12.7 (July 11, 2012), pages 3518–3525.

[23] A. K. Geim. “Graphene prehistory”. In: *Physica Scripta* 2012 (T146 2012), page 014003.

[24] A. K. Geim and K. S. Novoselov. “The rise of graphene”. In: *Nature Materials* 6.3 (March 2007), pages 183–191.

[25] Paolo Giannozzi et al. “QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials”. In: *Journal of Physics: Condensed Matter* 21.39 (2009), page 395502.

[26] R. W. Godby and R. J. Needs. “Metal-insulator transition in Kohn-Sham theory and quasiparticle theory”. In: *Physical Review Letters* 62.10 (March 6, 1989), pages 1169–1172.

[27] R. W. Godby, M. Schlüter, and L. J. Sham. “Self-energy operators and exchange-correlation potentials in semiconductors”. In: *Physical Review B* 37.17 (June 15, 1988), pages 10159–10175.

[28] Yongji Gong et al. “Vertical and in-plane heterostructures from WS2/MoS2 monolayers”. In: *Nature Materials* advance online publication (September 28, 2014).

[29] Tobias Grabo and E. K. U. Gross. “Density-functional theory using an optimized exchange-correlation potential”. In: *Chemical Physics Letters* 240.1 (June 23, 1995), pages 141–150.

[30] Oleg Gritsenko et al. “Self-consistent approximation to the Kohn-Sham exchange potential”. In: *Physical Review A* 51.3 (March 1, 1995), pages 1944–1954.

[31] F. Gygi and A. Baldereschi. “Self-consistent Hartree-Fock and screened-exchange calculations in solids: Application to silicon”. In: *Physical Review B* 34.6 (September 15, 1986), pages 4405–4408.

[32] Sebastien Hamel et al. “Kohn–Sham orbitals and orbital energies: fictitious constructs but good approximations all the same”. In: *Journal of Electron Spectroscopy and Related Phenomena*. Determination of cross-sections and momentum profiles of atoms, molecules and condensed matter 123.2 (May 2002), pages 345–363.

[33] Melinda Y. Han et al. “Energy Band-Gap Engineering of Graphene Nanoribbons”. In: *Physical Review Letters* 98.20 (May 16, 2007), page 206805.

[34] D. R. Hartree. “The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 24.1 (January 1928), pages 89–110.

[35] Ivana Hasa et al. “A Comparative Study of Layered Transition Metal Oxide Cathodes for Application in Sodium-Ion Battery”. In: *ACS Applied Materials & Interfaces* 7.9 (March 11, 2015), pages 5206–5212.

[36] Keliang He et al. “Tightly Bound Excitons in Monolayer WSe2”. In: *Physical Review Letters* 113.2 (July 10, 2014), page 026803.

[37] Lars Hedin. “New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem”. In: *Physical Review* 139.3 (August 2, 1965), A796–A823.

[38] Lars Hedin and Stig Lundqvist. “Effects of Electron-Electron and Electron-Phonon Interactions on the One-Electron States of Solids”. In: *Solid State Physics*. Edited by David Turnbull and Henry Ehrenreich Frederick Seitz. Volume Volume 23. Academic Press, 1970, pages 1–181.

[39] Berit Hinnemann et al. “Biomimetic Hydrogen Evolution: MoS2 Nanoparticles as Catalyst for Hydrogen Evolution”. In: *Journal of the American Chemical Society* 127.15 (April 1, 2005), pages 5308–5309.

[40] D. Hobbs, G. Kresse, and J. Hafner. “Fully unconstrained noncollinear magnetism within the projector augmented-wave method”. In: *Physical Review B* 62.17 (November 1, 2000), pages 11556–11570.

[41] P. Hohenberg and W. Kohn. “Inhomogeneous Electron Gas”. In: *Physical Review* 136.3 (November 9, 1964), B864–B871.

[42] Falco Huser, Thomas Olsen, and Kristian S. Thygesen. “How dielectric screening in two-dimensional crystals affects the convergence of excited-state calculations: Monolayer MoS2”. In: *Physical Review B* 88.24 (December 26, 2013), page 245309.

[43] Falco Hüser, Thomas Olsen, and Kristian S. Thygesen. “Quasiparticle GW calculations for solids, molecules, and two-dimensional materials”. In: *Physical Review B* 87.23 (June 24, 2013), page 235132.

[44] Mark S. Hybertsen and Steven G. Louie. “Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies”. In: *Physical Review B* 34.8 (October 15, 1986), pages 5390–5413.

[45] Mark S. Hybertsen and Steven G. Louie. “First-Principles Theory of Quasiparticles: Calculation of Band Gaps in Semiconductors and Insulators”. In: *Physical Review Letters* 55.13 (September 23, 1985), pages 1418–1421.

[46] Sohrab Ismail-Beigi. “Truncation of periodic image interactions for confined systems”. In: *Physical Review B* 73.23 (June 28, 2006), page 233103.

[47] J. F. Janak. “Proof that $\frac{\partial E}{\partial n_i} = \epsilon$ in density-functional theory”. In: *Physical Review B* 18.12 (December 15, 1978), pages 7165–7168.

[48] Chengjun Jin, Filip A. Rasmussen, and Kristian S. Thygesen. “Tuning the Schottky Barrier at the Graphene/MoS2 Interface by Electron Doping: Density Functional Theory and Many-Body Calculations”. In: *The Journal of Physical Chemistry C* 119.34 (August 27, 2015), pages 19928–19933.

[49] Wei Kang and Mark S. Hybertsen. “Quasiparticle and optical properties of rutile and anatase TiO_2”. In: *Physical Review B* 82.8 (August 12, 2010), page 085203.

[50] M. I. Katsnelson, K. S. Novoselov, and A. K. Geim. “Chiral tunnelling and the Klein paradox in graphene”. In: *Nature Physics* 2.9 (September 2006), pages 620–625.

[51] Min-Cheol Kim, Eunji Sim, and Kieron Burke. “Communication: Avoiding unbound anions in density functional calculations”. In: *The Journal of Chemical Physics* 134.17 (May 7, 2011), page 171103.

[52] Jiří Klimeš, Merzuk Kaltak, and Georg Kresse. “Predictive GW calculations using plane waves and pseudopotentials”. In: *Physical Review B* 90.7 (August 14, 2014), page 075125.

[53] A. R. Klots et al. “Probing excitonic states in suspended two-dimensional semiconductors by photocurrent spectroscopy”. In: *Scientific Reports* 4 (October 16, 2014).

[54] W. Kohn and L. J. Sham. “Self-Consistent Equations Including Exchange and Correlation Effects”. In: *Physical Review* 140.4 (November 15, 1965), A1133–A1138.

[55] T Koopmans. “Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms”. In: *Physica* 1.1 (1934), pages 104–113.

[56] Takao Kotani. “An optimized-effective-potential method for solids with exact exchange and random-phase approximation correlation”. In: *Journal of Physics: Condensed Matter* 10.41 (1998), page 9241.

[57] J. B. Krieger, Yan Li, and G. J. Iafrate. “Construction and application of an accurate local spin-polarized Kohn-Sham potential with integer discontinuity: Exchange-only theory”. In: *Physical Review A* 45.1 (January 1, 1992), pages 101–126.

[58] Ryogo Kubo. “Statistical-Mechanical Theory of Irreversible Processes. I. General Theory and Simple Applications to Magnetic and Conduction Problems”. In: *Journal of the Physical Society of Japan* 12.6 (June 15, 1957), pages 570–586.

[59] M. Kuisma et al. “Kohn-Sham potential with discontinuity for band gap materials”. In: *Physical Review B* 82.11 (September 7, 2010), page 115106.

[60] L. D. Landau. “Zur Theorie der phasenumwandlungen II”. In: *Phys. Z. Sowjetunion* 11 (1937), pages 26–35.

[61] Lev Davidovich Landau and Evgeny Mikhailovich Lifshitz. *Statistical Physics, Third Edition, Part 1: Volume 5*. 3 edition. Butterworth-Heinemann, January 15, 1980. 544 pages.

[62] David C. Langreth and M. J. Mehl. “Beyond the local-density approximation in calculations of ground-state electronic properties”. In: *Physical Review B* 28.4 (August 15, 1983), pages 1809–1834.

[63] Stephan Lany. “Band-structure calculations for the 3*d* transition metal oxides in *GW*”. In: *Physical Review B* 87.8 (February 13, 2013), page 085112.

[64] S. Lebègue and O. Eriksson. “Electronic structure of two-dimensional crystals from *ab initio* theory”. In: *Physical Review B* 79.11 (March 11, 2009), page 115409.

[65] S. Lebègue et al. “Implementation of an all-electron GW approximation based on the projector augmented wave method without plasmon pole approximation: Application to Si, SiC, AlAs, InAs, NaH, and KH”. In: *Physical Review B* 67.15 (April 30, 2003), page 155208.

[66] S. Lebègue et al. “Two-Dimensional Materials from Data Filtering and Ab Initio Calculations”. In: *Physical Review X* 3.3 (July 8, 2013), page 031002.

[67] Changgu Lee et al. “Frictional Characteristics of Atomically Thin Sheets”. In: *Science* 328.5974 (April 2, 2010), pages 76–80.

[68] Donghyung Lee, Filipp Furche, and Kieron Burke. “Accuracy of Electron Affinities of Atoms in Approximate Density Functional Theory”. In: *The Journal of Physical Chemistry Letters* 1.14 (July 15, 2010), pages 2124–2129.

[69] Mel Levy. “Electron densities in search of Hamiltonians”. In: *Physical Review A* 26.3 (September 1, 1982), pages 1200–1208.

[70] Je-Luen Li, G.-M. Rignanese, and Steven G. Louie. “Quasiparticle energy bands of NiO in the GW approximation”. In: *Physical Review B* 71.19 (May 26, 2005), page 193102.

[71] Tianshu Li and Giulia Galli. “Electronic Properties of MoS2 Nanoparticles”. In: *The Journal of Physical Chemistry C* 111.44 (November 1, 2007), pages 16192–16196.

[72] Linda Y. Lim et al. “Angle-resolved photoemission and quasiparticle calculation of ZnO: The need for d band shift in oxide semiconductors”. In: *Physical Review B* 86.23 (December 11, 2012), page 235113.

[73] Wolfgang von der Linden and Peter Horsch. “Precise quasiparticle energies and Hartree-Fock bands of semiconductors and insulators”. In: *Physical Review B* 37.14 (May 15, 1988), pages 8351–8362.

[74] SHANG-kENG MA and KEITH A. BRUECKNER. “Correlation Energy of an Electron Gas with a Slowly Varying High Density”. In: *Physical Review* 165.1 (January 5, 1968), pages 18–31.

[75] Kin Fai Mak et al. “Atomically Thin MoS2: A New Direct-Gap Semiconductor”. In: *Physical Review Letters* 105.13 (September 24, 2010), page 136805.

[76] N. D. Mermin. “Crystalline Order in Two Dimensions”. In: *Physical Review* 176.1 (December 5, 1968), pages 250–254.

[77] T. Miyake and F. Aryasetiawan. “Efficient algorithm for calculating noninteracting frequency-dependent linear response functions”. In: *Physical Review B* 61.11 (March 15, 2000), pages 7172–7175.

[78] J. K. Nørskov et al. “Origin of the Overpotential for Oxygen Reduction at a Fuel-Cell Cathode”. In: *The Journal of Physical Chemistry B* 108.46 (November 1, 2004), pages 17886–17892.

[79] K. S. Novoselov et al. “Electric Field Effect in Atomically Thin Carbon Films”. In: *Science* 306.5696 (October 22, 2004), pages 666–669.

[80] K. S. Novoselov et al. “Room-Temperature Quantum Hall Effect in Graphene”. In: *Science* 315.5817 (March 9, 2007), pages 1379–1379.

[81] K. S. Novoselov et al. “Two-dimensional atomic crystals”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.30 (July 26, 2005), pages 10451–10453.

[82] K. S. Novoselov et al. “Two-dimensional gas of massless Dirac fermions in graphene”. In: *Nature* 438.7065 (November 10, 2005), pages 197–200.

[83] Thomas Olsen et al. “Simple Screened Hydrogen Model of Excitons in Two-Dimensional Materials”. Submitted to Phys. Rev. B. December 2015.

[84] A. Oschlies, R. W. Godby, and R. J. Needs. “GW self-energy calculations of carrier-induced band-gap narrowing in n -type silicon”. In: *Physical Review B* 51.3 (January 15, 1995), pages 1527–1535.

[85] Mohnish Pandey and Karsten W. Jacobsen. “Heats of formation of solids with error estimation: The mBEEF functional with and without fitted reference energies”. In: *Physical Review B* 91.23 (June 3, 2015), page 235201.

[86] Mohnish Pandey et al. “Correction to “Two-Dimensional Metal Dichalcogenides and Oxides for Hydrogen Evolution: A Computational Screening Approach””. In: *The Journal of Physical Chemistry Letters* 6.14 (July 16, 2015), pages 2669–2670.

[87] Mohnish Pandey et al. “Defect Tolerance in Monolayer Transition Metal Dichalcogenides”. Submitted to Nano Letters. December 2015.

[88] Mohnish Pandey et al. “Two-Dimensional Metal Dichalcogenides and Oxides for Hydrogen Evolution: A Computational Screening Approach”. In: *The Journal of Physical Chemistry Letters* 6.9 (May 7, 2015), pages 1577–1585.

[89] Thomas G. Pedersen et al. “Graphene Antidot Lattices: Designed Defects and Spin Qubits”. In: *Physical Review Letters* 100.13 (April 3, 2008), page 136804.

[90] J. P. Perdew and Alex Zunger. “Self-interaction correction to density-functional approximations for many-electron systems”. In: *Physical Review B* 23.10 (May 15, 1981), pages 5048–5079.

[91] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized Gradient Approximation Made Simple”. In: *Physical Review Letters* 77.18 (October 28, 1996), pages 3865–3868.

[92] John P. Perdew and Mel Levy. “Physical Content of the Exact Kohn-Sham Orbital Energies: Band Gaps and Derivative Discontinuities”. In: *Physical Review Letters* 51.20 (November 14, 1983), pages 1884–1887.

[93] John P. Perdew et al. “Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation”. In: *Physical Review B* 46.11 (September 15, 1992), pages 6671–6687.

[94] John P. Perdew et al. “Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy”. In: *Physical Review Letters* 49.23 (December 6, 1982), pages 1691–1694.

[95] John P. Perdew et al. “Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation”. In: *Physical Review B* 48.7 (August 15, 1993), pages 4978–4978.

[96] Robert M. Pick, Morrel H. Cohen, and Richard M. Martin. “Microscopic Theory of Force Constants in the Adiabatic Approximation”. In: *Physical Review B* 1.2 (January 15, 1970), pages 910–920.

[97] Peter Politzer and Fakher Abu-Awwad. “A comparative analysis of Hartree-Fock and Kohn-Sham orbital energies”. In: *Theoretical Chemistry Accounts* 99.2 (April 1998), pages 83–87.

[98] M. V. Putz, N. Russo, and E. Sicilia. “About the Mulliken electronegativity in DFT”. In: *Theoretical Chemistry Accounts* 114.1 (September 2005). WOS:000232053800006, pages 38–45.

[99] Ashwin Ramasubramaniam. “Large excitonic effects in monolayers of molybdenum and tungsten dichalcogenides”. In: *Physical Review B* 86.11 (September 6, 2012), page 115409.

[100] Filip A. Rasmussen and Kristian S. Thygesen. “Computational 2D Materials Database: Electronic Structure of Transition-Metal Dichalcogenides and Oxides”. In: *The Journal of Physical Chemistry C* 119.23 (June 11, 2015), pages 13169–13183.

[101] Filip A. Rasmussen and Kristian S. Thygesen. “Efficient *GW* calculations on two-dimensional materials using accurate long wavelength limit of the screened potential”. In: *arXiv:1511.00129 [cond-mat]* (October 31, 2015). Submitted to Phys. Rev. B. arXiv: `1511.00129`.

[102] Carlo A. Rozzi et al. “Exact Coulomb cutoff technique for supercell calculations”. In: *Physical Review B* 73.20 (May 26, 2006), page 205119.

[103] Angel Rubio, Jennifer L. Corkill, and Marvin L. Cohen. “Theory of graphitic boron nitride nanotubes”. In: *Physical Review B* 49.7 (February 15, 1994), pages 5081–5084.

[104] Claudia Ruppert, Ozgur Burak Aslan, and Tony F. Heinz. “Optical Properties and Band Gap of Single- and Few-Layer MoTe2 Crystals”. In: *Nano Letters* 14.11 (November 12, 2014), pages 6231–6236.

[105] L. J. Sham and M. Schlüter. “Density-Functional Theory of the Energy Gap”. In: *Physical Review Letters* 51.20 (November 14, 1983), pages 1888–1891.

[106] R. T. Sharp and G. K. Horton. “A Variational Approach to the Unipotential Many-Electron Problem”. In: *Physical Review* 90.2 (April 15, 1953), pages 317–317.

[107] Bi-Ching Shih et al. “Quasiparticle Band Gap of ZnO: High Accuracy from the Conventional G^0W^0 Approach”. In: *Physical Review Letters* 105.14 (September 27, 2010), page 146401.

[108] M. Shishkin and G. Kresse. “Implementation and performance of the frequency-dependent GW method within the PAW framework”. In: *Physical Review B* 74.3 (July 6, 2006), page 035101.

[109] Arunima K. Singh et al. “Computational Screening of 2D Materials for Photocatalysis”. In: *The Journal of Physical Chemistry Letters* 6.6 (March 19, 2015), pages 1087–1098.

[110] James Spencer and Ali Alavi. “Efficient calculation of the exact exchange energy in periodic systems using a truncated Coulomb potential”. In: *Physical Review B* 77.19 (May 23, 2008), page 193110.

[111] Vladan Stevanović et al. “Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies”. In: *Physical Review B* 85.11 (March 7, 2012), page 115104.

[112] Ravishankar Sundararaman and T. A. Arias. “Regularization of the Coulomb singularity in exact exchange by Wigner-Seitz truncated interactions: Towards chemical accuracy in nontrivial systems”. In: *Physical Review B* 87.16 (April 17, 2013), page 165122.

[113] Kazunori Takada et al. “Superconductivity in two-dimensional CoO2 layers”. In: *Nature* 422.6927 (March 6, 2003), pages 53–55.

[114] James D. Talman and William F. Shadwick. “Optimized effective atomic central potential”. In: *Physical Review A* 14.1 (July 1, 1976), pages 36–40.

[115] Murilo L. Tiago, Sohrab Ismail-Beigi, and Steven G. Louie. “Effect of semicore orbitals on the electronic band gaps of Si, Ge, and GaAs within the GW approximation”. In: *Physical Review B* 69.12 (March 19, 2004), page 125212.

[116] S. Trasatti. “The absolute electrode potential: an explanatory note (Recommendations 1986)”. In: *Pure and Applied Chemistry* 58.7 (January 1, 1986).

[117] P. Umari et al. “Accelerating GW calculations with optimal polarizability basis”. In: *physica status solidi (b)* 248.3 (March 1, 2011), pages 527–536.

[118] Yue Wang and John P. Perdew. “Correlation hole of the spin-polarized electron gas, with exact small-wave-vector and high-density scaling”. In: *Physical Review B* 44.24 (December 15, 1991), pages 13298–13307.

[119] B. Wenzien, G. Cappellini, and F. Bechstedt. “Efficient quasiparticle band-structure calculations for cubic and noncubic crystals”. In: *Physical Review B* 51.20 (May 15, 1995), pages 14701–14704.

[120] Ludger Wirtz, Andrea Marini, and Angel Rubio. “Excitons in Boron Nitride Nanotubes: Dimensionality Effects”. In: *Physical Review Letters* 96.12 (March 30, 2006), page 126104.

[121] Nathan Wiser. “Dielectric Constant with Local Field Effects Included”. In: *Physical Review* 129.1 (January 1, 1963), pages 62–69.

[122] Jun Yan et al. “Linear density response function in the projector augmented wave method: Applications to solids, surfaces, and interfaces”. In: *Physical Review B* 83.24 (June 24, 2011), page 245122.

[123] Ziliang Ye et al. “Probing excitonic dark states in single-layer tungsten disulphide”. In: *Nature* 513.7517 (September 11, 2014), pages 214–218.

[124] Te-Fu Yeh et al. “Nitrogen-Doped Graphene Oxide Quantum Dots as Photocatalysts for Overall Water-Splitting under Visible Light Illumination”. In: *Advanced Materials* 26.20 (May 1, 2014), pages 3297–3303.

[125] Yuanbo Zhang et al. “Experimental observation of the quantum Hall effect and Berry's phase in graphene”. In: *Nature* 438.7065 (November 10, 2005), pages 201–204.

[126] S. Y. Zhou et al. “Substrate-induced bandgap opening in epitaxial graphene.” In: *Nature Materials* 6.11 (November 2007), pages 916–916.

[127] Bairen Zhu, Xi Chen, and Xiaodong Cui. “Exciton Binding Energy of Monolayer WS2”. In: *arXiv:1403.5108 [cond-mat]* (March 20, 2014). arXiv: 1403.5108.

[128] Z. Y. Zhu, Y. C. Cheng, and U. Schwingenschlögl. “Giant spin-orbit-induced spin splitting in two-dimensional transition-metal dichalcogenide semiconductors”. In: *Physical Review B* 84.15 (October 14, 2011), page 153402.

Papers

Paper I

Filip A. Rasmussen and Kristian S. Thygesen. "Computational 2D Materials Database: Electronic Structure of Transition-Metal Dichalcogenides and Oxides". In: *The Journal of Physical Chemistry C* 119.23 (June 11, 2015), pages 13169–13183

THE JOURNAL OF PHYSICAL CHEMISTRY C

Article

pubs.acs.org/JPCC

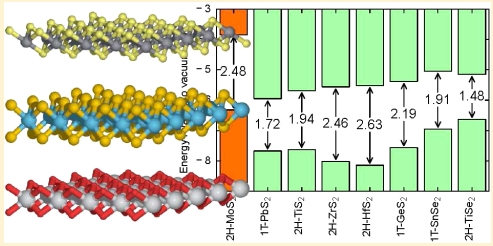
Computational 2D Materials Database: Electronic Structure of Transition-Metal Dichalcogenides and Oxides

Filip A. Rasmussen and Kristian S. Thygesen*

Center for Atomic-scale Materials Design (CAMD), Department of Physics, and Center for Nanostructured Graphene (CNG), Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

Supporting Information

ABSTRACT: We present a comprehensive first-principles study of the electronic structure of 51 semiconducting monolayer transition-metal dichalcogenides and -oxides in the 2H and 1T hexagonal phases. The quasiparticle (QP) band structures with spin−orbit coupling are calculated in the G_0W_0 approximation, and comparison is made with different density functional theory descriptions. Pitfalls related to the convergence of GW calculations for two-dimensional (2D) materials are discussed together with possible solutions. The monolayer band edge positions relative to vacuum are used to estimate the band alignment at various heterostructure interfaces. The sensitivity of the band structures to the in-plane lattice constant is analyzed and rationalized in terms of the electronic structure. Finally, the q-dependent dielectric functions and effective electron and hole masses are obtained from the QP band structure and used as input to a 2D hydrogenic model to estimate exciton binding energies. Throughout the paper we focus on trends and correlations in the electronic structure rather than detailed analysis of specific materials. All the computed data is available in an open database.



1. INTRODUCTION

Atomically thin two-dimensional (2D) materials, such as graphene, hexagonal boron-nitride, and the transition-metal dichalcogenides (TMDs) are presently being intensively researched because of their unique optoelectronic properties. The TMDs with the chemical formula MX_2 (X = S, Se, Te; M = transition metal) represent a particularly interesting class of 2D materials comprising both semiconductors and metals.[1] For example, MoS_2, $MoSe_2$, WS_2, and WSe_2 were shown to undergo a transition from indirect to direct band gap materials when their thickness is thinned down to a single layer.[2−6] Together with their strong interaction with light[7,8] and relatively high charge carrier mobilities,[9,10] this has opened up the possibility of using few-layer TMDs in a range of applications including ultrathin field effect transistors,[11] photo detectors,[12−15] light-emitting diodes,[16] and solar cells.[17,18] Furthermore, the lack of inversion symmetry in the 2H monolayer structures leads to a spin−orbit driven splitting of the valence band which in turn allows for valley-selective excitation of charge carriers.[19−22] Adding to this the possibility of tuning the electronic properties by strain,[23] dielectric screening,[24] electrostatic gating,[25,26] nanostructuring,[27] or by combining individual 2D materials into van der Waals heterostructures,[28,29] it is clear that monolayer TMDs hold great potential both as a platform for fundamental physics and as building blocks for nanoscale device applications.

To date, optoelectronic research in monolayer TMDs has mainly focused on the Mo- and W-based compounds which have (optical) band gaps in the range of 1.6−2.0 eV,[2,3,5,6,30] significantly larger than the ideal values for both photovoltaics and transistor applications.[31] In order to advance the usage of 2D materials from the level of fundamental research to real applications, it is essential to enlarge the space of available 2D building blocks beyond the handful of presently considered materials. To this end, not only the band gaps but also the absolute band edge positions, effective masses, and dielectric function will be of key importance for predicting the usefulness of a given 2D material.

The fact that the interlayer bonding in bulk TMDs is of very similar strength (around 20 meV/Å^2)[32] indicates that exfoliation of single layers should be feasible for many different TMDs. Indeed, liquid exfoliation of nanosheets of $TaSe_2$, $NbSe_2$, $NiTe_2$, and $MoTe_2$ has been demonstrated.[33] In this regard, it is interesting to note that more than 40 TMDs are already known in the bulk form and could form the basis for new 2D materials.[34] The stability of such 2D monolayers under ambient conditions is a critical issue, but it could be alleviated by encapsulation in protective layers as recently demonstrated for MoS_2 in hexagonal boron-nitride.[35]

In a previous work, Ataca et al. performed an extensive stability analysis of 88 monolayer TMDs and TMOs using density functional theory (DFT) in the local density approximation (LDA) and identified 52 stable compounds including both

Received: March 27, 2015
Revised: April 30, 2015
Published: April 30, 2015

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

metals and semiconductors.[36] While stability was their main focus, they also calculated the LDA band structures of the stable compounds and a few selected compounds using the GW_0 approximation. They concluded, surprisingly, that the LDA provides good agreement with existing experiments while GW_0 significantly overestimates the band gap. This false conclusion is based on the common confusion between the optical and the quasiparticle (QP) band gaps. The former is probed in optical experiments and is lower than the QP gap by the exciton binding energy. It is one of the characteristic features of the atomically thin semiconductors that exciton binding energies are very large (on the order of 1 eV). This leads to pronounced differences between the QP and optical spectra, both of which are well-reproduced by many-body calculations applying the GW approximation and Bethe−Salpeter equation, respectively.[37−43]

We present an extensive first-principles study of the electronic structure of a variety of monolayer TMDs and TMOs in the 2H and 1T structures based on 27 different metals. For reference, the atomic structures of the 2H and 1T phases are shown in Figure 1a, the corresponding Brillouin zone (BZ) with the special *k*-

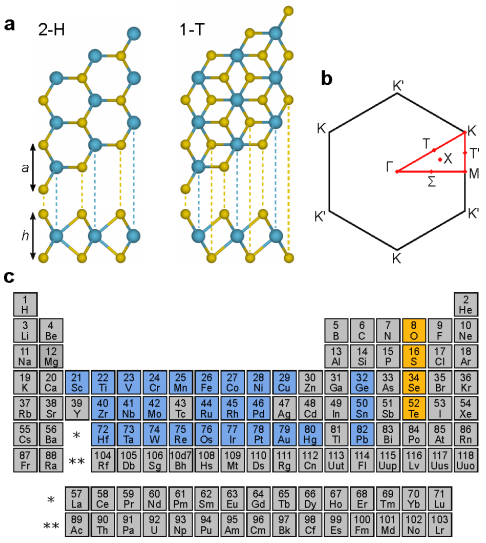


Figure 1. (a) Crystal structure of the monolayer transition-metal oxides and dichalcogenides in the 2H and 1T phases, respectively. Here, *a* denotes the in-plane hexagonal unit cell lattice constant and *h* is the vertical distance between the oxygen or chalcogen atoms. (b) In-plane Brillouin zone of the hexagonal unit cell with high-symmetry points and other *k*-points indicated. (c) Periodic table of the elements with the metals considered in this study highlighted in blue and oxygen and chalcogens highlighted in yellow.

points is shown in Figure 1b, and the elements considered are highlighted in the periodic table shown in Figure 1c. Out of 216 investigated compounds, we find 171 to be stable (defined by a negative heat of formation relative to the standard states). These results represent a consistent extension of the LDA-based stability analysis of ref 36. Out of the 171 stable monolayers, we here focus on the 51 compounds that are found to be nonmagnetic and nonmetallic. For these materials, we calculate the band structures as well as the absolute position of the valence and conduction band edges relative to vacuum using the G_0W_0 approximation with spin−orbit coupling included. Convergence of the absolute G_0W_0 quasiparticle energies is found to be particularly demanding, and we therefore discuss this issue in some detail. The G_0W_0 band gaps and band edge positions are compared to Kohn−Sham DFT using different exchange-correlation functionals. We find that the band gap is generally well-reproduced by the GLLB-SC functional while the LDA provides a surprisingly good description of the band gap center. In contrast, an empirical formula for estimating the band edge positions from the electro-negativities of the constituent atoms is found to deviate significantly from the first-principles results because of charge transfer from the metal to the oxygen or chalcogen atoms and associated electrostatic potential that lowers the electronic band energies relative to vacuum. We furthermore calculate the (static) *q*-dependent dielectric function of all the compounds and discuss some basic properties of dielectric screening in quasi 2D. The effective charge carrier masses are derived from the G_0W_0 band structures and used, together with the dielectric functions, as input to an effective 2D model for the exciton binding energies.

Overall, our results reveal a large degree of variation in the electronic properties of the investigated materials. For example, the materials MX_2 (X = S, Se, Te; M = Cr, Mo, W) have direct QP band gaps in the 0.9−2.5 eV range while all other compounds have indirect gaps in the 0.5−7.0 eV range. The band gap centers (relative to vacuum) span from −8 eV for some of the oxides to above −5 eV for the selenides and tellurides. The effective masses vary by almost 2 orders of magnitude as do the *q*-dependent dielectric functions.

All the computed data including relaxed structures, DFT and G_0W_0 band structures, absolute band edge positions, effective masses, and dielectric functions, are available online in the Computational Materials Repository (http://cmr.fysik.dtu.dk/).

2. COMPUTATIONAL METHODS

All calculations were performed using the projector augmented wave method as implemented in the GPAW code.[44] GPAW supports three different types of basis sets, namely, real space grids, numerical atomic orbitals, and plane waves. We have used the latter in the present work because excited-state calculations with GPAW are implemented only for plane waves.

2.1. Atomic Structure. The lattice constants (Figure 1a) of the 216 monolayer TMDs and TMOs were determined by a structure relaxation using the Perdew−Burke−Ernzerhof (PBE)[45] exchange-correlation (xc) functional with 750 eV plane wave cutoff, 18 × 18 × 1 Monkhorst−Pack *k*-point sampling, and 20 Å between periodically repeated layers. For both the 2H and 1T phases the lattice constant of the minimal unit cell and the vertical positions of the oxygen or chalcogen atoms were relaxed until all forces were below 0.01 eV/Å. We used the minimal unit cell and did not investigate symmetry-reducing distortions which have recently been found to occur for some 1T metallic compounds.[46−48] Because these distortions are driven by a metal-to-insulator transition (Peierls distortion), we do not expect them to be important for the semiconducting materials which are the focus of the present work.

2.2. Electronic Structure. The Kohn−Sham band structures of all compounds were calculated self-consistently using the LDA, PBE, and GLLB-SC[49] xc-functionals. Spin−orbit coupling was included in a non-self-consistent manner by diagonalizing the total Hamiltonian consisting of the spin−orbit interaction (which is applied inside the PAW spheres) and the self-consistently determined Kohn−Sham Hamiltonian. We have

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

found that the spin−orbit corrections to the band structure were unchanged (less than 0.02 eV) if we use G_0W_0 energies instead of LDA energies in the Kohn−Sham Hamiltonian. For ten representative materials, we benchmarked the spin−orbit corrected LDA band structures obtained with GPAW against the all-electron linearized augmented-plane wave ELK code[50] and found excellent agreement (difference within 0.02 eV).

The QP band structures were calculated in the G_0W_0 approximation as implemented in GPAW.[24] We used LDA wave functions obtained from an exact diagonalization of the Kohn−Sham Hamiltonian with a plane wave cutoff of 600 eV and 30 × 30 k-points as input for the G_0W_0 calculations. The plane wave cutoff used to construct the screened interaction and self-energy was varied between 150 and 500 eV and extrapolated to infinite cutoff energy as described below. For all calculations, the number of unoccupied orbitals used to construct the screened interaction and GW self-energy was set equal to the number of plane waves. The frequency dependence was represented on a nonlinear grid from 0 eV to the energy of the highest transition included in the basis with a gradually increasing grid spacing starting at 0.1 eV and reaching 0.2 eV at ω = 15 eV. The frequency grid typically contained 300 to 350 grid points. The PAW potentials applied in this work include semicore states, i.e., atomic states down to at least 1 hartree below vacuum, while deeper lying states are included in the frozen core. The frozen core states are included in the exchange contribution to the GW corrections.

In ref 43, we demonstrated the importance of using a truncated Coulomb interaction in GW calculations of 2D materials. In the present study we have used the Wigner−Seitz truncation scheme.[51] For a representative set of materials we have checked that the QP band gaps (and more generally the absolute band edge positions) change by less than 0.1 eV when the k-point grid is increased from 30 × 30 to 45 × 45. We note in passing that most previous GW calculations for 2D systems have applied significantly smaller k-point grids.[36,37,52] As explained in ref 43, the physical reason for the slow convergence with k-points is the strong q-dependence of the dielectric function of a 2D semiconductor. While $\epsilon(q)$ for a 3D semiconductor tends smoothly to constant value for $q \to 0$, $\epsilon(q) = 1 + O(q)$ for a 2D system (Figure 2). As a consequence, a denser k-point grid is required to capture the variation in $\epsilon(q)$ around q = 0. For example, the G_0W_0 band gap of 2H-MoS_2 is reduced by 0.4 eV when increasing the k-point grid from 15 × 15 to 30 × 30. Because the strong variation in $\epsilon(q)$ is limited to a small region around q = 0, it is sufficient to sample the screened interaction $W(q)$ on a fine grid in this region while a coarser sampling may be used in the remaining part of the BZ (such nonuniform sampling was, however, not used in the present work). We stress that these facts apply only to isolated 2D semiconductors, which in practice means when a truncated Coulomb interaction is used. Only then is $\epsilon(q) = 1 + O(q)$. If instead the full $1/r$ Coulomb interaction is used, the calculations converge much faster but to a wrong value depending on the interlayer distance.[43] The dielectric function and Figure 2 will be discussed in more depth in section 3.6.

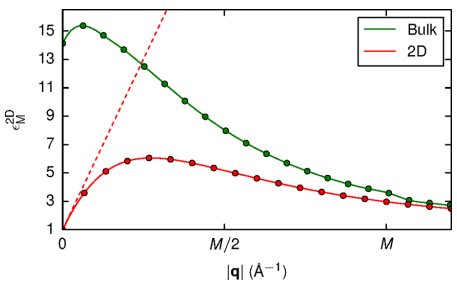


Figure 2. Static quasi-2D dielectric function of 2H-MoS_2 along the Γ → M direction. For comparison, the macroscopic dielectric function of bulk MoS_2 is also shown. The slope of the 2D dielectric function is indicated by a dashed line.

Finally, we discuss the convergence of the G_0W_0 energies with respect to the number of plane waves, N_G, used to represent the screened interaction and self-energy. It has previously been found that the GW corrections for bulk semiconductors and insulators follow a $1/N_G$ scaling[53] which makes it possible to extrapolate the QP energies to the infinite basis set limit. From our calculations with varying cutoff energy from 150 eV and in some cases up to 500 eV, we observe the following: (i) The extrapolation procedure is essential and can correct QP energies obtained with 150 eV cutoff by up to 0.5 eV. (ii) The slope of the extrapolation curve can be different for different states (bands and k-points), but generally shows a decrease as a function of N_G. (iii) The band gap tends to converge faster than the absolute band energies. In ref 53, it was also shown that the lack of norm conservation of the PAW potentials can affect the convergence of the GW energies as N_G is increased. The effect is larger for more localized states, particularly the 3d states, where the violation of norm conservation can be significant. While it is possible to construct norm-conserving PAW potentials, we have not pursued this in the present work.

Performing the extrapolation to infinite cutoff for 30 × 30 k-points is computationally demanding. Fortunately, we have found that the extrapolation is rather insensitive to the k-point mesh. This is shown in Figure 3 for the case of 2H-MoS_2.

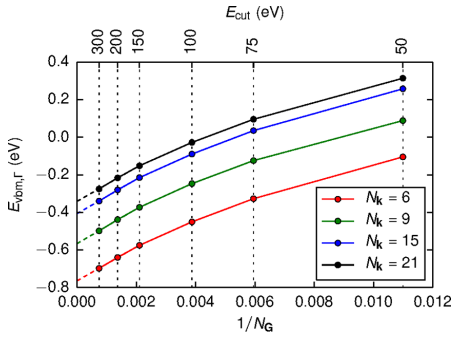


Figure 3. G_0W_0 quasiparticle energy of the valence band at the Γ-point of monolayer 2H-MoS_2 as a function of $1/N_G$, where N_G is the number of plane waves. The different lines correspond to different k-point samplings ($N_k \times N_k \times 1$). The dashed lines show the extrapolation to infinite plane wave cutoff.

Changing the k-point mesh simply shifts the entire N_G-curves but do not affect their form. To obtain results converged with respect to both k-points and plane waves, we have therefore performed the N_G-extrapolation for a coarse k-point sampling of 12 × 12 and corrected the band energies by the difference between a 12 × 12 and 30 × 30 calculation at 150 eV plane wave cutoff. In doing this we must interpolate band energies from the coarse to the fine k-

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

point grid. The GW correction to an energy, $\varepsilon_{\mathbf{k}n}$, on the fine grid is obtained as a weighted average of the corrections obtained at the nearest points of the coarse k-point grid, $\varepsilon_{\mathbf{k}'m}$, with weights determined by the overlap of the LDA state densities, $w_{\mathbf{k}n,\mathbf{k}'m} = \langle \rho^{\text{fine}}_{\mathbf{k}n}, \rho^{\text{coarse}}_{\mathbf{k}'m} \rangle$, where $\rho_{\mathbf{k}n}(\mathbf{r}) = |\psi_{\mathbf{k}n}(\mathbf{r})|^2$. Because the G_0W_0 shift depends crucially on the shape of the wave function, this approach is essential, in particular when interpolating the G_0W_0 corrections close to band crossings.

3. RESULTS

In this section we present the main results of our electronic structure calculations. To limit the presentation, we have chosen to focus on the *trends* in electronic structure observed across the investigated materials rather than giving in-depth analysis of particular materials. However, because all the data is available in the database, it is straightforward for the interested reader to obtain the entire set of computed data.

3.1. Stability. The heat of formation of the 216 monolayer TMDs and TMOs in the relaxed structure was calculated from

$$\Delta H(\mathrm{MX_2}) = E_{\text{tot}}(\mathrm{MX_2}) - E_{\text{ref}}(\mathrm{M}) - 2E_{\text{ref}}(\mathrm{X}) \quad (1)$$

where $E_{\text{tot}}(\mathrm{MX_2})$ is the PBE total energy of the monolayer and $E_{\text{ref}}(\mathrm{M})$ and $E_{\text{ref}}(\mathrm{X})$ are the reference energies of the metal and chalcogen/oxygen (X), respectively. For the latter, we use the fitted elemental phase reference energies (FERE).[54]

Figure 4 shows the calculated heat of formations for all 216 compounds. Most of the materials have negative heat of formation; in fact, by requiring a heat of formation below 0.1 eV (to allow for uncertainties in the calculation methods) we obtain 171 stable compounds. The heat of formation of the stable semiconductors together with relaxed lattice constants (a), distance between outermost chalcogen or oxygen atoms (h) (see Figure 1), and the final magnetic moments are given in Table 1. In general, the oxides have the highest stability followed by the sulfides, selenides, and tellurides in that order. Furthermore, the stability decreases as the metal ion goes through the transition-metal series. For comparison with previous studies we note that the 52 monolayer MX_2 compounds found to be stable based on the LDA calculations of ref 36 form a subset of the stable materials identified in the present work.

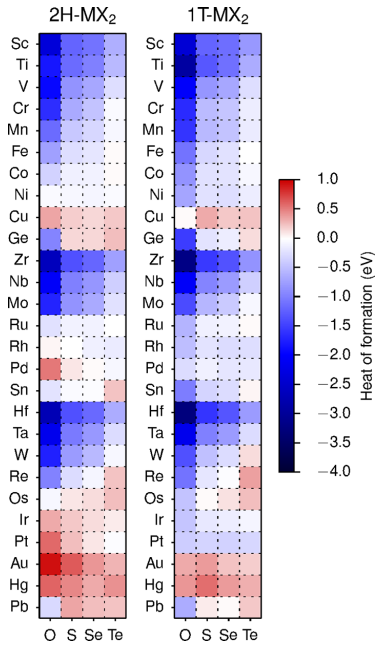


Figure 4. Calculated heat of formation for all monolayers in the 2H and 1T phases. In general, the oxides have the highest stability followed by the sulfides, selenides, and tellurides, in that order. Furthermore, the stability decreases as the metal ion goes through the transition-metal series.

While the heat of formation is a natural descriptor for whether the material will be possible to synthesize, we stress that mechanical instabilities or competing phases of lower energy have not been taken into account. While it is possible to account for both effects, e.g., by carrying out molecular dynamics simulations[36] and including a larger pool of reference systems,[55] we have not pursued this further. Lebègue et al.[34] searched the Inorganic Crystal Structure Database (ICSD) for known layered bulk materials. They identified 46 TMDs but did not specify whether the bulk materials were known in the 1T or 2H phase. Assuming the 1T and 2H phases to be equally stable (it follows from Figure 4 that this is a reasonable assumption), 76 out of the 171 materials with negative heat of formation are already known as layered bulk materials. These materials are marked by an asterix in Table 1. The fact that Lebègue et al. did not identify the layered bulk form of any of the TMOs investigated here indicates that these structures could be very challenging to synthesize, presumably because of the existence of nonlayered bulk phases of higher stability. On the other hand, the monolayer TMOs might be metastable or could be stabilized by interaction with a substrate. Encapsulation of the monolayers, as recently demonstrated for MoS_2 in hexagonal boron-nitride,[35] could be a way to prevent the material from reacting with other chemical species.

3.2. Band Gaps. For all the stable and nonmagnetic semiconductors we have performed G_0W_0 calculations following the procedure described in section 2.2. The G_0W_0 corrections have been evaluated for the 10 bands closest to the Fermi energy, and spin−orbit coupling has been included non-pertubatively. The spin−orbit splittings of the valence/conduction bands for materials where these are nonvanishing are reported in Table 2. Unless otherwise stated, all the results presented in this work include spin−orbit interactions. As an example, we show the G_0W_0 and LDA band structures of 2H-WSe_2 in Figure 5. For this particular material we find a direct G_0W_0 gap of 2.08 eV and a 0.45 eV splitting of the valence band at the K-point.

We refrain from providing a detailed comparison with previous literature values for the QP band gaps. For MoS_2, such a comparison was made in ref 43. However, the fact that most previous GW calculations for 2D TMDs have used rather coarse k-point grids and have not employed a truncated Coulomb interaction (both of which have a significant effect on the calculated gap[43]), the importance of spin−orbit interaction which is not always included, as well as the sensitivity of the gap to the in-plane lattice constant (see section 3.4), makes a general comparison difficult. We believe that the QP band structures of the present work are the most carefully converged G_0W_0 calculations reported for 2D TMOs and TMDs to date. There are only a few experimental reports on QP band gaps in freestanding TMD monolayers. For 2H-MoS_2, our G_0W_0 gap of 2.48 eV agrees well with the 2.5 eV reported in ref 15 based on

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

Table 1. Relaxed In-Plane Lattice Constant (a), Distance between Chalcogen/Oxygen Atoms (h), Formation Energies from PBE (E_f^{PBE}) and Using the Fitted Elemental Phase Reference Energies (E_f^{FERE}), and Total Magnetic Moment (μ)[a]

name	a (Å)	h (Å)	E_f^{PBE} (eV)	E_f^{FERE} (eV)	μ (μ_B)	name	a (Å)	h (Å)	E_f^{PBE} (eV)	E_f^{FERE} (eV)	μ (μ_B)
2H-CrO_2	2.63	2.34	−1.64	−1.99	0.0	1T-PtS_2*	3.57	2.46	−0.332	−0.418	0.0
2H-CrS_2*	3.05	2.95	−0.662	−0.892	0.0	1T-$PtSe_2$*	3.75	2.62	−0.364	−0.397	0.0
2H-$CrSe_2$*	3.21	3.15	−0.474	−0.65	0.0	1T-$PtTe_2$*	4.02	2.77	−0.321	−0.23	0.0
2H-$CrTe_2$*	3.47	3.41	−0.051	−0.104	0.0	2H-ScO_2	3.22	2.07	−2.37	−2.74	1.0
2H-GeO_2	2.81	2.32	−0.969	−1.28	0.0	2H-ScS_2	3.79	2.72	−1.21	−1.46	1.0
1T-GeO_2	2.9	1.96	−1.53	−1.84	0.0	2H-$ScSe_2$	3.95	2.94	−1.1	−1.29	1.0
1T-GeS_2	3.44	2.8	−0.222	−0.416	0.0	2H-SnO_2	3.09	2.46	−0.225	−0.54	0.0
2H-HfO_2	3.12	2.34	−2.71	−3.08	0.0	1T-SnO_2	3.22	2	−1.01	−1.33	0.0
1T-HfO_2	3.25	1.95	−3.27	−3.64	0.0	2H-SnS_2*	3.61	3.23	−0.048	−0.241	0.0
2H-HfS_2*	3.54	3.14	−1.37	−1.62	0.0	1T-SnS_2*	3.7	2.96	−0.333	−0.527	0.0
1T-HfS_2*	3.65	2.9	−1.59	−1.83	0.0	1T-$SnSe_2$*	3.86	3.19	−0.285	−0.425	0.0
2H-$HfSe_2$*	3.68	3.36	−1.17	−1.36	0.0	2H-TiO_2	2.88	2.26	−1.83	−2.02	0.0
1T-$HfSe_2$*	3.77	3.16	−1.34	−1.53	0.0	1T-TiO_2	2.99	1.94	−2.91	−3.1	0.0
2H-$HfTe_2$*	3.91	3.7	−0.656	−0.723	0.0	2H-TiS_2*	3.34	3.02	−1.16	−1.23	0.0
1T-MnO_2	2.89	1.93	−1.58	−2	3.0	2H-$TiSe_2$*	3.49	3.24	−1	−1.02	0.0
2H-MoO_2	2.82	2.45	−1.73	−1.94	0.0	2H-$TiTe_2$*	3.74	3.58	−0.544	−0.441	0.0
2H-MoS_2*	3.18	3.13	−0.842	−0.93	0.0	2H-VSe_2*	3.34	3.2	−0.699	−0.956	1.0
2H-$MoSe_2$*	3.32	3.34	−0.663	−0.698	0.0	2H-VTe_2*	3.6	3.5	−0.263	−0.397	1.0
2H-$MoTe_2$*	3.55	3.61	−0.237	−0.149	0.0	2H-WO_2	2.83	2.48	−1.74	−1.85	0.0
1T-NiO_2	2.84	1.91	−0.716	−1.01	0.0	2H-WS_2*	3.19	3.15	−0.783	−0.776	0.0
1T-NiS_2	3.35	2.35	−0.248	−0.424	0.0	2H-WSe_2*	3.32	3.36	−0.547	−0.487	0.0
1T-$NiSe_2$	3.54	2.49	−0.251	−0.374	0.0	2H-ZrO_2	3.14	2.33	−2.65	−2.96	0.0
1T-PbO_2	3.39	2.14	−0.641	−0.8	0.0	1T-ZrO_2	3.26	1.93	−3.18	−3.49	0.0
1T-PbS_2	3.85	3.09	0.069	0.031	0.0	2H-ZrS_2*	3.57	3.14	−1.37	−1.55	0.0
1T-PdO_2	3.09	1.96	−0.272	−0.482	0.0	1T-ZrS_2*	3.68	2.9	−1.55	−1.47	0.0
1T-PdS_2*	3.55	2.49	−0.125	−0.214	0.0	2H-$ZrSe_2$*	3.7	3.37	−1.2	−1.33	0.0
1T-$PdSe_2$*	3.73	2.63	−0.206	−0.242	0.0	1T-$ZrSe_2$*	3.79	3.16	−1.34	−1.47	0.0
1T-$PdTe_2$*	4.02	2.76	−0.177	−0.09	0.0	2H-$ZrTe_2$*	3.92	3.73	−0.739	−0.746	0.0
1T-PtO_2	3.14	1.9	−0.405	−0.612	0.0						

[a]An asterisk (*) denotes whether the material is found in bulk form according to ref 34.

photocurrent spectroscopy on suspended MoS_2. We note that the slightly smaller band gap obtained here for MoS_2 (2.48 eV) compared to our previous work[43] (2.65 eV) is mainly due to the inclusion of spin−orbit interaction in the present work. Alternatively, the QP gap can be inferred from optical absorption or photoluminescence spectra, which are experimentally simpler to obtain. However, this requires knowledge of the exciton binding energy which in turn depends on the screening from the substrate.[56]

In the last two columns of Table 3 we show the calculated direct and indirect band gaps of the 51 stable 2D semiconductors. It is well-known from photoluminescence spectroscopy that the 2H phase of monolayer MoS_2, $MoSe_2$, $MoTe_2$, WS_2, and WSe_2 have direct band gaps. This is reproduced by our G_0W_0 calculations. However, we find that the indirect gap in $MoTe_2$ is about the same size as the direct gap. The only other materials we find to have a direct band gap are CrS_2, $CrSe_2$, and $CrTe_2$ with gaps of 1.54, 1.21, and 0.77 eV, respectively. All other compounds have indirect gaps in the range of 0.5−7.0 eV.

In addition to LDA and G_0W_0, we have calculated the band gaps using the GLLB-SC functional of Kuisma et al.[49] The GLLB-SC is an orbital-dependent exact exchange-based functional which, in addition to the Kohn−Sham band gap, provides an estimate of the derivative discontinuity. The GLLB-SC has previously been shown to yield good results for the band gap of bulk semiconductors,[24,49,55] but to our knowledge it has not been previously applied to 2D materials. In Figure 6 we compare the G_0W_0 band gaps with the PBE and GLLB-SC gaps. We first note that the G_0W_0 band gaps range from 0.5 eV to almost 8 eV, with the majority of the materials lying in the 1−3 eV range. We note that the size of the band gaps is directly correlated with the heat of formation of the materials with the oxides having the largest band gaps followed by the sulfides, selenides, and tellurides in that order. As expected, the LDA gaps are significantly lower than those obtained from G_0W_0, which is consistent with the situation known from bulk materials and molecules. In contrast, except for a few outliers, the band gaps obtained with the GLLB-SC functional lie very close to the G_0W_0 values with a mean absolute error of 0.4 eV. This is consistent with the results obtained for both bulk and molecular systems[24] and supports the use of the GLLB-SC functional as viable alternative to GW in large-scale studies in which one would benefit from its low computational requirements, which are similar to those of LDA.

3.3. Absolute Band Positions. For many applications, not only the distance between the occupied and unoccupied bands, i.e., the band gap, but also the absolute position of the band edges relative to vacuum are of interest. We have calculated these by referring the band energies to the asymptotic value of the Hartree potential in the vacuum region between the layers. For bulk materials this is a difficult task as it requires the use of thick slabs to represent both the bulk interior and its surface. Moreover, the Hartree potential depends on the surface dipoles (on both sides of the slab), which makes the problem highly surface-dependent and complicates the comparison with experiments. These problems are obviously not present for the monolayers studied

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

Table 2. Spin−Orbit-Induced Splittings at the Valence Band Maximum and Conduction Band Minimum As Found in the LDA and G_0W_0 Band Structure, Respectively[a]

name	ΔE^{soc}_{vbm} (LDA)	ΔE^{soc}_{cbm} (LDA)	ΔE^{soc}_{vbm} (G_0W_0)	ΔE^{soc}_{cbm} (G_0W_0)
2H-CrS_2	0.07 (K)	0 (K)	0.07 (K)	0.01 (K)
2H-$CrSe_2$	0.09 (K)	0.02 (K)	0.1 (K)	0.02 (K)
2H-$CrTe_2$	0.12 (K)	0.02 (K)	0.13 (K)	0.03 (K)
2H-HfO_2	0 (T)	0.17 (T)	0 (T)	0.15 (T)
2H-HfS_2	0.03 (T)	0.07 (T)	0.02 (T)	0.09 (X)
2H-$HfSe_2$	0.13 (T)	0.1 (T)	0.12 (T)	0.11 (X)
2H-$HfTe_2$	0 (Γ)	0.15 (Γ)	0.48 (T)	0.18 (X)
2H-MoS_2	0.15 (K)	0 (K)	0.15 (K)	0 (K)
2H-$MoSe_2$	0.19 (K)	0.02 (K)	0.19 (K)	0.02 (K)
2H-$MoTe_2$	0.23 (K)	0.04 (K)	0.25 (K)	0.05 (X)
2H-TiO_2	0 (T)	0.02 (T)	0 (X)	0.02 (T)
2H-TiS_2	0.02 (T)	0 (T)	0 (Γ)	0 (Σ)
2H-$TiTe_2$	0 (Γ)	0 (Γ)	0.32 (T)	0 (Σ)
2H-WO_2	0 (Γ)	0.02 (Γ)	0 (Γ)	0 (K)
2H-WS_2	0.45 (K)	0.04 (K)	0.45 (K)	0.02 (K)
2H-WSe_2	0.49 (K)	0.04 (K)	0.49 (K)	0.03 (K)
2H-ZrO_2	0 (T)	0.05 (T)	0 (T)	0.05 (T)
2H-ZrS_2	0.02 (T)	0.02 (T)	0.02 (T)	0 (Σ)
2H-$ZrSe_2$	0.1 (T)	0.03 (T)	0.1 (T)	0 (Σ)
2H-$ZrTe_2$	0 (Γ)	0 (Γ)	0.28 (T)	0 (Σ)

[a]Materials with negligible spin−orbit coupling are not shown. The location of the band extremum in the BZ is indicated in parentheses (see Figure 1b). Note that this can be different in LDA and G_0W_0.

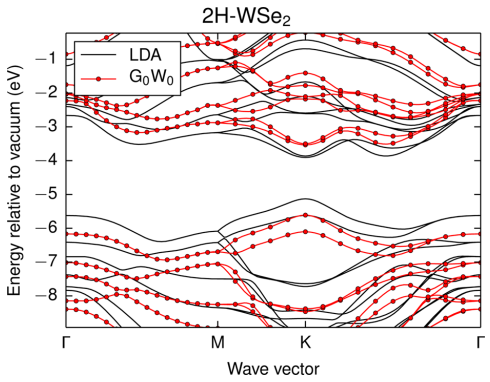


Figure 5. Band structure of 2H-WSe_2 using LDA (black) and G_0W_0 (red) and LDA projected density of states. Note the spin−orbit coupling gives rise to a splitting of the bands at various regions of the Brillouin zone. The red line connecting the G_0W_0 points is obtained from a cubic spline interpolation.

here, making them ideal as benchmarking systems for the band alignment problem.

In Figure 7, the positions of the valence band maximum (VBM) and conduction band minimum (CBM) relative to the vacuum level are shown for the different oxides and chalcogenides at both the LDA and G_0W_0 level. Because a significant part of the GLLB-SC band gap comes from the derivative discontinuity which applies to the fundamental gap rather than the individual band energies, the GLLB-SC cannot be used to obtain the absolute band edge positions. For all materials, the effect of the G_0W_0 correction is to shift the conduction band up and the valence band down with respect to the LDA values. In fact, the corrections of the VBM and CBM are rather symmetric, meaning that the band gap center is largely unaffected by the G_0W_0 correction (see below).

It has been suggested that 2D semiconductors could be used for photocatalytic water splitting. This is mainly motivated by their excellent light absorption, large specific surface area, and readily tunable electronic properties.[57,58] The equilibrium potentials for the hydrogen and oxygen evolution reactions at pH 7 are indicated by dashed green lines in Figure 7. Materials with CBM above the standard hydrogen electrode (SHE) at −4.03 eV relative to vacuum (at pH 7) could in principle be used to evolve hydrogen at the cathode of a photocatalytic water-splitting device.[59] Likewise, materials with VBM below the oxygen evolution potential (1.23 eV below the SHE) could in principle be used as a photoanode in the water-splitting reaction. In practice, the CBM/VBM should lie a few tenths of an electronvolt above/below the redox potentials to account for the intrinsic energy barriers of the water-splitting reactions.[60] As can be seen, a number of the TMD monolayers qualify as potential water-splitting photoelectrodes based on their energy level positions. A very citical issue, however, is the stability of the materials under the highly oxidizing reaction conditions. A possible solution to this problem could be to protect the photoelectrode from direct contact with the water by a transparent and highly stable thin film, which in practice means an oxide material.

In ref 61, a simple empirical relation was observed between the band gap center of a semiconductor and the electronegativities of the constituent atoms

$$E_{center} = -[\chi(M)\chi(X)^2]^{1/3} \quad (2)$$

where $\chi(M)$ and $\chi(X)$ are the electronegativity of the metal and oxygen/chalcogen on the Mulliken scale, respectively. In Figure 8, we compare the band gap centers obtained from G_0W_0 with those obtained from LDA and calculated with eq 2, where experimentally obtained values of the electronegativities[62] have been used. The band gap centers from LDA and G_0W_0 agree quite well, showing a mean absolute deviation from the G_0W_0 values of only 0.2 eV. While it is known that the Kohn−Sham band gap center is formally exact within DFT,[63] it is somewhat surprising that the LDA performs that well. While the empirical formula is able to describe the qualitative trends of the gap centers, the quantitative values deviate significantly from the ab initio results, with a mean absolute difference from the G_0W_0 result of 0.9 eV and a mean relative deviation of 14%. We ascribe a large part of this deviation to dipole fields formed because of the positively charged metal ions and negatively charged chalcogens/oxygens which will increase potential outside the monolayer and thereby down shift the bands, an effect not accounted for by the empirical formula. Because the size of the dipoles is determined by the amount of charge transfer, the deviation between eq 2 and the ab initio results is expected to correlate with the difference in electronegativity between the metal atom and chalcogen/oxygen atoms. From the inset of Figure 8 we see that this indeed is the case. For materials with larger difference in electronegativity between the atomic species ($\Delta\chi$), the band gap center given by eq 2 generally deviates more from the G_0W_0 results.

While it is important to establish the intrinsic properties of the 2D materials in their isolated form, practical applications as well as most experimental setups involve heterostructures where the 2D materials are stacked into van der Waals heterostructure or simply lie on a substrate. In such systems the alignment of the

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

Table 3. Absolute Band Edge Positions with Respect to Vacuum, Their Location in the Brillouin Zone (in Parentheses)[a] and Corresponding Band Gaps As Obtained by LDA and G_0W_0

	LDA				G_0W_0@LDA			
name	E_{vbm} (eV)	E_{cbm} (eV)	E_{gap} (eV)	E_{gap}^{direct} (eV)	E_{vbm} (eV)	E_{cbm} (eV)	E_{gap} (eV)	E_{gap}^{direct} (eV)
2H-CrO_2	−7.64 (Γ)	−7.21 (K)	0.43	1.57 (K)	−7.37 (Γ)	−5.73 (K)	1.64	2.45 (K)
2H-CrS_2	−6.08 (K)	−5.18 (K)	0.90	0.90 (K)	−5.85 (K)	−4.31 (K)	1.54	1.54 (K)
2H-$CrSe_2$	−5.50 (K)	−4.80 (K)	0.70	0.70 (K)	−5.22 (K)	−4.02 (K)	1.21	1.21 (K)
2H-$CrTe_2$	−5.05 (K)	−4.59 (K)	0.45	0.45 (K)	−4.75 (K)	−3.98 (K)	0.77	0.77 (K)
2H-GeO_2	−8.99 (K)	−7.62 (Γ)	1.37	1.95 (Γ)	−10.65 (X)	−6.40 (Γ)	4.24	4.62 (Γ)
1T-GeO_2	−9.20 (T)	−5.65 (Γ)	3.55	4.04 (Γ)	−11.07 (T)	−4.00 (Γ)	7.07	7.55 (Γ)
1T-GeS_2	−6.63 (Σ)	−6.07 (M)	0.57	1.00 (M)	−7.57 (Σ)	−5.38 (M)	2.19	2.64 (M)
2H-HfO_2	−8.51 (T)	−6.70 (T)	1.80	1.91 (T)	−9.91 (T)	−5.41 (T)	4.50	4.60 (T)
1T-HfO_2	−8.24 (T)	−3.61 (M)	4.63	4.85 (Σ)	−9.89 (X)	−1.91 (M)	7.98	8.21 (Σ)
2H-HfS_2	−7.05 (T)	−6.12 (X)	0.93	1.20 (X)	−8.14 (T)	−5.52 (X)	2.63	2.93 (X)
1T-HfS_2	−6.48 (Γ)	−5.42 (M)	1.06	1.77 (Γ)	−7.62 (Γ)	−4.63 (M)	2.98	3.97 (Γ)
2H-$HfSe_2$	−6.47 (T)	−5.77 (X)	0.70	1.02 (X)	−7.38 (T)	−5.29 (X)	2.09	2.49 (X)
1T-$HfSe_2$	−5.57 (Γ)	−5.26 (M)	0.30	1.08 (Γ)	−6.53 (Γ)	−4.58 (M)	1.96	2.95 (Γ)
2H-$HfTe_2$	−5.46 (Γ)	−5.39 (X)	0.06	0.52 (Σ)	−6.06 (T)	−5.12 (X)	0.94	1.62 (X)
2H-MoO_2	−6.99 (Γ)	−6.09 (K)	0.91	1.66 (Γ)	−7.37 (Γ)	−5.17 (K)	2.20	2.94 (Γ)
2H-MoS_2	−6.13 (K)	−4.55 (K)	1.58	1.58 (K)	−6.32 (K)	−3.84 (K)	2.48	2.48 (K)
2H-$MoSe_2$	−5.50 (K)	−4.18 (K)	1.32	1.32 (K)	−5.63 (K)	−3.46 (K)	2.18	2.18 (K)
2H-$MoTe_2$	−5.04 (K)	−4.11 (K)	0.93	0.93 (K)	−5.11 (K)	−3.40 (X)	1.71	1.72 (K)
1T-NiO_2	−8.38 (X)	−7.22 (Σ)	1.17	1.37 (Σ)	−8.38 (T)	−6.24 (Σ)	2.15	2.31 (Σ)
1T-NiS_2	−5.97 (Γ)	−5.46 (Σ)	0.51	0.89 (T′)	−6.61 (Σ)	−4.24 (Σ)	2.38	2.76 (X)
1T-$NiSe_2$	−5.10 (Γ)	−5.10 (Σ)	0.00	0.56 (T′)	−5.70 (Σ)	−3.91 (Σ)	1.79	2.24 (X)
1T-PbO_2	−8.47 (X)	−7.15 (Γ)	1.32	1.58 (Γ)	−9.50 (X)	−6.47 (Γ)	3.03	3.26 (Γ)
1T-PbS_2	−6.93 (Σ)	−6.29 (M)	0.63	0.81 (M)	−7.67 (Σ)	−5.95 (M)	1.72	1.91 (M)
1T-PdO_2	−7.82 (X)	−6.52 (Σ)	1.30	1.71 (Σ)	−8.20 (Σ)	−5.36 (Σ)	2.84	3.24 (Σ)
1T-PdS_2	−6.48 (Γ)	−5.37 (Σ)	1.11	1.30 (Σ)	−7.19 (T)	−4.70 (Σ)	2.48	2.65 (X)
1T-$PdSe_2$	−5.56 (Γ)	−5.08 (Σ)	0.48	0.87 (T′)	−6.25 (Σ)	−4.46 (Σ)	1.79	2.10 (Σ)
1T-$PdTe_2$	−4.46 (Γ)	−4.62 (Σ)	0.00	0.40 (T′)	−5.20 (Γ)	−4.18 (Σ)	1.02	1.39 (T′)
1T-PtO_2	−7.21 (Σ)	−5.61 (Σ)	1.60	2.00 (Σ)	−7.99 (Σ)	−4.41 (Σ)	3.59	4.00 (Σ)
1T-PtS_2	−6.44 (T)	−4.84 (Σ)	1.61	1.69 (Σ)	−7.16 (Σ)	−4.21 (Σ)	2.95	3.14 (T′)
1T-$PtSe_2$	−5.69 (Γ)	−4.62 (Σ)	1.07	1.29 (Σ)	−6.52 (T)	−4.04 (Σ)	2.48	2.67 (Σ)
1T-$PtTe_2$	−4.52 (Γ)	−4.29 (Σ)	0.23	0.75 (T′)	−5.44 (Γ)	−3.74 (Σ)	1.69	2.03 (T′)
2H-SnO_2	−8.78 (K)	−8.21 (Γ)	0.56	1.26 (Γ)	−10.16 (X)	−7.50 (Γ)	2.66	3.31 (Γ)
1T-SnO_2	−8.64 (X)	−6.10 (Γ)	2.54	3.13 (Σ)	−10.27 (X)	−4.89 (Γ)	5.38	5.93 (Σ)
2H-SnS_2	−6.54 (Γ)	−5.95 (M)	0.59	0.91 (Γ)	−7.54 (Γ)	−5.61 (M)	1.93	2.14 (Γ)
1T-SnS_2	−6.98 (Σ)	−5.58 (M)	1.40	1.65 (M)	−7.98 (Σ)	−4.91 (M)	3.07	3.33 (M)
1T-$SnSe_2$	−6.19 (Γ)	−5.58 (M)	0.62	0.96 (M)	−6.96 (Σ)	−5.05 (M)	1.91	2.25 (M)
2H-TiO_2	−8.88 (T)	−7.78 (T)	1.10	1.25 (X)	−9.97 (X)	−6.25 (T)	3.72	3.83 (X)
1T-TiO_2	−8.67 (X)	−6.02 (Γ)	2.65	2.80 (Γ)	−9.80 (X)	−4.07 (Σ)	5.74	5.97 (Σ)
2H-TiS_2	−6.95 (T)	−6.33 (Σ)	0.62	0.89 (X)	−7.63 (Γ)	−5.69 (Σ)	1.94	2.38 (Σ)
2H-$TiSe_2$	−6.31 (Γ)	−5.89 (Σ)	0.42	0.77 (Σ)	−6.63 (Γ)	−5.15 (M)	1.48	2.13 (M)
2H-$TiTe_2$	−5.40 (Γ)	−5.44 (Σ)	0.00	0.31 (Σ)	−5.52 (T)	−5.06 (Σ)	0.45	1.21 (T)
2H-WO_2	−6.73 (Γ)	−5.41 (K)	1.32	1.65 (Γ)	−7.38 (Γ)	−4.73 (K)	2.65	3.18 (Γ)
2H-WS_2	−5.75 (K)	−4.24 (K)	1.51	1.51 (K)	−6.28 (K)	−3.85 (K)	2.43	2.43 (K)
2H-WSe_2	−5.13 (K)	−3.91 (K)	1.22	1.22 (K)	−5.61 (K)	−3.53 (K)	2.08	2.08 (K)
2H-ZrO_2	−8.44 (T)	−6.85 (T)	1.59	1.70 (T)	−9.71 (T)	−5.63 (T)	4.08	4.19 (T)
1T-ZrO_2	−8.20 (T)	−3.82 (K)	4.37	4.63 (Γ)	−9.73 (X)	−1.97 (M)	7.76	8.25 (Σ)
2H-ZrS_2	−7.02 (T)	−6.18 (X)	0.85	1.03 (X)	−8.02 (T)	−5.56 (Σ)	2.46	2.69 (T)
1T-ZrS_2	−6.58 (Γ)	−5.55 (M)	1.03	1.53 (Γ)	−7.60 (Γ)	−4.72 (Σ)	2.88	3.61 (Γ)
2H-$ZrSe_2$	−6.47 (T)	−5.82 (X)	0.64	0.91 (X)	−7.29 (T)	−5.33 (Σ)	1.96	2.27 (X)
1T-$ZrSe_2$	−5.66 (Γ)	−5.41 (M)	0.25	0.87 (Γ)	−6.53 (Γ)	−4.68 (M)	1.85	2.63 (Γ)
2H-$ZrTe_2$	−5.62 (Γ)	−5.44 (Σ)	0.18	0.47 (Σ)	−6.17 (T)	−5.16 (Σ)	1.01	1.41 (Σ)

[a]See Figure 1b.

bands at the heterostructure interfaces becomes crucial. Because of the weak interaction between 2D semiconductors it is reasonable to expect that the band alignment at the interface between two different 2D materials can be obtained by aligning the band edges of the isolated systems relative to a common vacuum level. This is equivalent to disregarding effects of band hybridization and the formation of interface dipoles due to charge redistribution. Verifying this assumption from first-principles calculations is, however, difficult because of the lattice mismatch between different 2D materials.

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

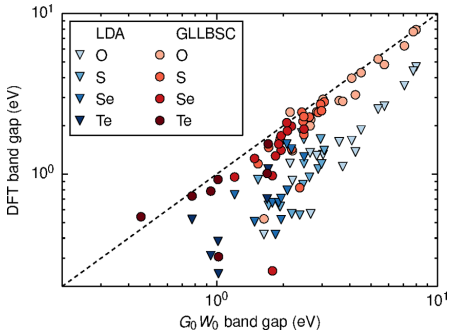


Figure 6. Computed band gaps of 51 monolayer TMDs and TMOs. The G_0W_0 band gaps are compared to the band gaps obtained from DFT with the PBE and GLLB-SC xc-functionals. The latter includes the derivative discontinuity of the xc-potential. Note the logarithmic scale.

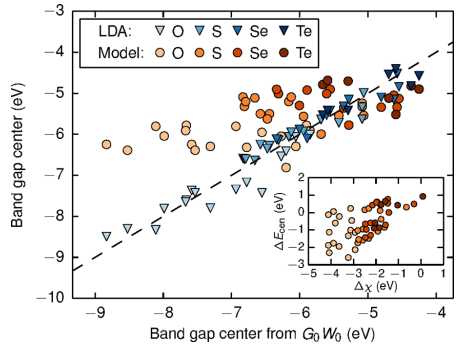


Figure 8. Comparison of the absolute band gap centers (relative to vacuum) obtained from G_0W_0, LDA, and the empirical formula eq 2. Inset: The difference in the band gap centers from G_0W_0 and eq 2, $\Delta E_{cen} = E_{cen}^{GW} - E_{cen}^{Model}$, compared to the difference in the electronegativities of the metal and oxygen/chalcogen atom, $\Delta\chi = \chi(M) - \chi(X)$.

To provide an overview of the band edge positions of the 51 monolayers, we show in Figure 9a the CBM plotted against the VBM obtained from G_0W_0. To illustrate the use of such a diagram, we have highlighted 2H-MoS_2 and indicated regions corresponding to different band alignments with MoS_2. The possible band alignments are straddling gap (type I), staggered gap (type II), and broken gap (type III). For many applications, e.g., tandem photovoltaic devices or creation of long-lived indirect excitons, a type II band alignment is preferred. We have highlighted a few materials that are expected to form type II band alignment with MoS_2. The detailed band alignments for these materials are shown in Figure 9b.

3.4. Strain Effects on the Band Structure. In the present work we have considered monolayers in their PBE relaxed geometry. Because PBE errors on lattice constants typically are around 1−2%, we have investigated how this would affect the LDA band structure. From Figure 10 we see that a change of the lattice parameter within the considered range can produce quite drastic changes in the band gap. For example, in the case of 2H-MoS_2, a change in the lattice constant from the PBE value (3.18 Å) to the experimental value (3.16 Å) changes the band gap by around 0.1 eV. From Table 4, where the band gaps are given as a function of lattice constant, we furthermore see that the LDA gap changes from indirect to direct under 1% compressive strain. A few other direct gap materials are seen to develop an indirect gap when strained. Thus, we conclude that both the size and nature of the band gap of the monolayers can depend delicately on the lattice constant.

To understand the different behavior of the band gap upon strain, we have analyzed the projected density of states (see Supporting Information). We find that the materials can be roughly divided into two classes according to the nature of the wave functions around the band gap. For the materials with group 6 metals (Cr, Mo and W), the valence and conduction band states are bonding−antibonding combinations of the metal d-states and oxygen/chalcogen p-states and in their equilibrium lattice constant they have direct band gaps. For these materials we find that increasing the lattice constant increases the M−X binding distance which weakens the hybridization and reduces the bonding−antibonding gap. The other class is TMDs with metals from group 4, 10, or 14 (Ti, Zr, Hf, Ni, Pd, Pt, Ge, Sn, Pb).

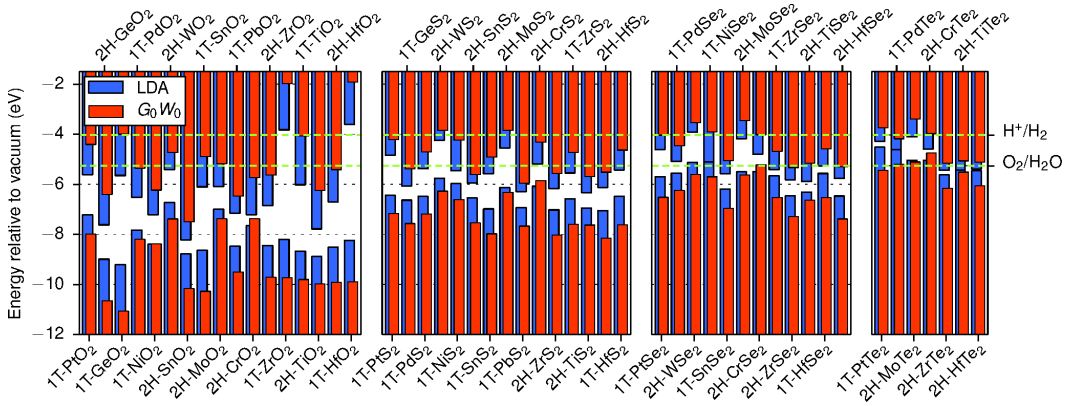


Figure 7. Position of the valence band maximum and conduction band minimum relative to the vacuum level (set to zero) for LDA and G_0W_0. In both cases, spin−orbit splitting of the bands has been taken into account. The hydrogen and oxygen evolution potentials at pH 7 are shown by green dashed lines.

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

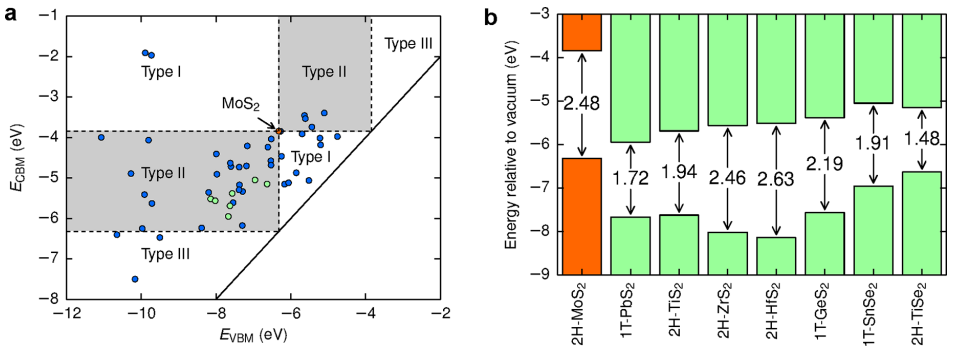


Figure 9. Band alignment diagram. (a) Conduction band minimum (E_{CBM}) plotted against the valence band maximum (E_{VBM}) for the 51 monolayers. The band edges relative to vacuum are obtained from G_0W_0. As an example, we have highlighted 2H-MoS_2 (orange dot) and indicated the regions corresponding to the different types of band alignment: straddling gap (type I), staggered gap (type II), and broken gap (type III). A few selected materials that will form type-II heterostructures with MoS_2 have been highlighted in green. (b) Absolute band edge positions and band gaps of 2H-MoS_2 and the selected materials highlighted in panel a.

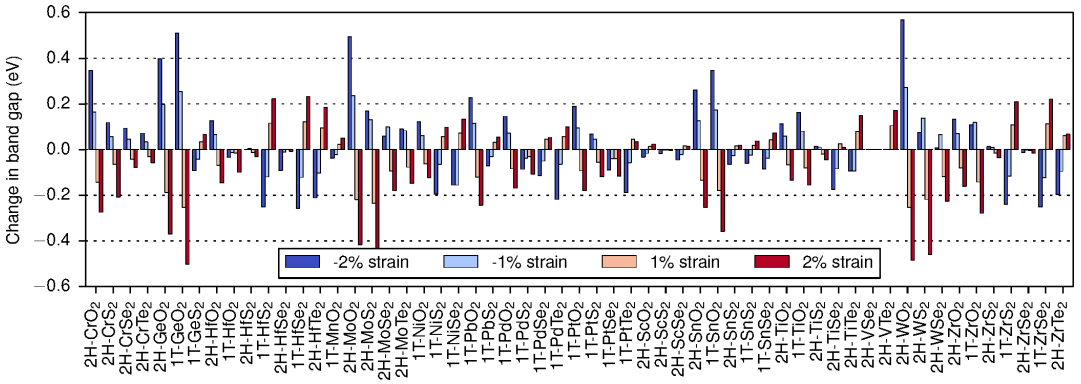


Figure 10. Change in the LDA band gap when the in-plane lattice constant is varied between −2% and +2%.

For these materials, the valence band states have primarily chalcogen p-character while the conduction band is either metal-d (group 4), chalcogen-p (group 10), or metal-s and chalcogen-p (group 14). In these cases, the gap size is controlled by the width of the conduction band and the chalcogen valence band; application of a tensile strain will cause the states to become more localized, narrowing each of the bands and thereby opening the gap. As a consequence of the decoupled bands, these also all have indirect band gaps.

3.5. Effective Masses. From the G_0W_0 band structures we have extracted the effective electron and hole masses by fitting a paraboloid to the energies of the 19 nearest k-points around the conduction band minimum (CBM) and valence band maximum (VBM) according to

$$E = \frac{\hbar^2}{2m_e}\mathbf{k}^T\mathbf{A}\mathbf{k} \tag{3}$$

where $\mathbf{k} = (k_x, k_y)$ is the in-plane k-point measured from the band extremum. The eigenvalues of the matrix $\mathbf{A}$ yield the inverse effective masses in the direction of smallest and largest curvature. If the CBM or VBM is located at one of the high-symmetry points of the BZ (the Γ or K point), the effective masses will naturally be isotropic. However, for band extrema located at other points this is generally not the case.

In panels a and b of Figure 11, we show the effective electron and hole masses along the two natural directions, respectively. Points falling on the diagonal line correspond to isotropic band masses. The effective masses are also listed in Table 5. We note that the effective electron masses lie in the 0.1−10 m_e range, with roughly an equal number being light ($m_e^* < m_e$) and heavy ($m_e^* > m_e$). The same approximately applies to the hole masses, although they seem to be generally heavier than the electrons. In accordance with the discussion in the previous section, we see that only the materials with direct gaps (group 6 metals) have both isotropic electron and hole masses. For other materials, the masses can be quite anisotropic and we would also expect the masses to depend sensitively on the lattice constant.

To estimate exciton binding energies (see section 3.7) we also evaluate the effective exciton masses defined as

$$\mu_{ex}^{-1} = m_e^{*-1} + m_h^{*-1} \tag{4}$$

We distinguish between two kinds of excitons: direct excitons that possess zero momentum and indirect gap excitons that have a finite momentum corresponding to the distance in k-space between the VBM and CBM. In Figure 11c, we plot the effective

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

Table 4. LDA Band Gaps (in Electronvolts) as Function of Strain[a]

name	−2%	−1%	0%	1%	2%
2H-CrO_2	0.77 (I)	0.59 (I)	0.42 (I)	0.28 (I)	0.15 (I)
2H-CrS_2	1.04 (D)	0.98 (D)	0.92 (D)	0.86 (I)	0.72 (I)
2H-$CrSe_2$	0.84 (D)	0.79 (D)	0.74 (D)	0.70 (D)	0.67 (D)
2H-$CrTe_2$	0.59 (D)	0.56 (D)	0.52 (D)	0.49 (D)	0.47 (D)
2H-GeO_2	1.77 (I)	1.57 (I)	1.37 (I)	1.19 (I)	1.00 (I)
2H-HfO_2	2.02 (I)	1.96 (I)	1.89 (I)	1.82 (I)	1.75 (I)
2H-HfS_2	0.96 (I)	0.96 (I)	0.96 (I)	0.94 (I)	0.93 (I)
2H-$HfSe_2$	0.71 (I)	0.79 (I)	0.80 (I)	0.80 (I)	0.79 (I)
2H-$HfTe_2$	0.10 (I)	0.21 (I)	0.31 (I)	0.41 (I)	0.50 (I)
2H-MoO_2	1.41 (I)	1.15 (I)	0.91 (I)	0.70 (I)	0.50 (I)
2H-MoS_2	1.82 (I)	1.78 (D)	1.65 (I)	1.41 (I)	1.19 (I)
2H-$MoSe_2$	1.50 (I)	1.53 (D)	1.44 (D)	1.34 (D)	1.26 (D)
2H-$MoTe_2$	1.16 (I)	1.16 (D)	1.07 (D)	1.00 (D)	0.93 (D)
2H-ScO_2	1.12 (I)	1.13 (I)	1.15 (I)	1.16 (I)	1.17 (I)
2H-ScS_2	0.48 (I)	0.50 (I)	0.50 (I)	0.50 (I)	0.49 (I)
2H-$ScSe_2$	0.31 (I)	0.34 (I)	0.36 (I)	0.38 (I)	0.37 (I)
2H-SnO_2	0.83 (I)	0.69 (I)	0.57 (I)	0.43 (I)	0.31 (I)
2H-SnS_2	0.56 (I)	0.60 (I)	0.63 (I)	0.64 (I)	0.65 (I)
2H-TiO_2	1.25 (I)	1.19 (I)	1.14 (I)	1.07 (I)	1.00 (I)
2H-TiS_2	0.64 (I)	0.64 (I)	0.63 (I)	0.61 (I)	0.58 (I)
2H-$TiSe_2$	0.33 (I)	0.42 (I)	0.51 (I)	0.53 (I)	0.52 (I)
2H-$TiTe_2$	M	M	0.09 (I)	0.17 (I)	0.24 (I)
2H-VSe_2	M	M	M	M	M
2H-VTe_2	M	M	M	0.11 (I)	0.17 (I)
2H-WO_2	1.93 (I)	1.63 (I)	1.36 (I)	1.10 (I)	0.87 (I)
2H-WS_2	1.88 (I)	1.94 (I)	1.80 (D)	1.58 (I)	1.34 (I)
2H-WSe_2	1.55 (I)	1.61 (I)	1.54 (D)	1.43 (D)	1.32 (D)
2H-ZrO_2	1.76 (I)	1.70 (I)	1.63 (I)	1.55 (I)	1.47 (I)
2H-ZrS_2	0.87 (I)	0.87 (I)	0.86 (I)	0.84 (I)	0.82 (I)
2H-$ZrSe_2$	0.69 (I)	0.70 (I)	0.70 (I)	0.70 (I)	0.69 (I)
2H-$ZrTe_2$	0.18 (I)	0.29 (I)	0.38 (I)	0.44 (I)	0.45 (I)
1T-GeO_2	4.07 (I)	3.81 (I)	3.56 (I)	3.31 (I)	3.06 (I)
1T-GeS_2	0.48 (I)	0.53 (I)	0.57 (I)	0.61 (I)	0.64 (I)
1T-HfO_2	4.63 (I)	4.65 (I)	4.66 (I)	4.65 (I)	4.56 (I)
1T-HfS_2	0.90 (I)	1.04 (I)	1.16 (I)	1.27 (I)	1.38 (I)
1T-$HfSe_2$	0.26 (I)	0.40 (I)	0.52 (I)	0.64 (I)	0.75 (I)
1T-MnO_2	0.69 (I)	0.70 (I)	0.72 (I)	0.75 (I)	0.77 (I)
1T-NiO_2	1.30 (I)	1.24 (I)	1.18 (I)	1.11 (I)	1.05 (I)
1T-NiS_2	0.35 (I)	0.48 (I)	0.55 (I)	0.60 (I)	0.65 (I)
1T-$NiSe_2$	M	M	0.16 (I)	0.23 (I)	0.29 (I)
1T-PbO_2	1.55 (I)	1.44 (I)	1.32 (I)	1.20 (I)	1.08 (I)
1T-PbS_2	0.56 (I)	0.61 (I)	0.64 (I)	0.67 (I)	0.69 (I)
1T-PdO_2	1.46 (I)	1.39 (I)	1.32 (I)	1.23 (I)	1.15 (I)
1T-PdS_2	1.09 (I)	1.13 (I)	1.17 (I)	1.14 (I)	1.06 (I)
1T-$PdSe_2$	0.55 (I)	0.61 (I)	0.66 (I)	0.71 (I)	0.72 (I)
1T-$PdTe_2$	M	0.15 (I)	0.22 (I)	0.27 (I)	0.32 (I)
1T-PtO_2	1.78 (I)	1.68 (I)	1.59 (I)	1.50 (I)	1.41 (I)
1T-PtS_2	1.73 (I)	1.71 (I)	1.66 (I)	1.61 (I)	1.54 (I)
1T-$PtSe_2$	1.20 (I)	1.25 (I)	1.29 (I)	1.25 (I)	1.17 (I)
1T-$PtTe_2$	0.50 (I)	0.63 (I)	0.69 (I)	0.74 (I)	0.73 (I)
1T-SnO_2	2.89 (I)	2.72 (I)	2.54 (I)	2.36 (I)	2.18 (I)
1T-SnS_2	1.35 (I)	1.38 (I)	1.41 (I)	1.43 (I)	1.45 (I)
1T-$SnSe_2$	0.58 (I)	0.63 (I)	0.67 (I)	0.71 (I)	0.74 (I)
1T-TiO_2	2.82 (I)	2.74 (I)	2.66 (I)	2.58 (I)	2.50 (I)
1T-ZrO_2	4.48 (I)	4.50 (I)	4.38 (I)	4.23 (I)	4.10 (I)
1T-ZrS_2	0.84 (I)	0.96 (I)	1.08 (I)	1.19 (I)	1.29 (I)
1T-$ZrSe_2$	0.17 (I)	0.30 (I)	0.42 (I)	0.53 (I)	0.64 (I)

[a]The character in the parentheses denotes whether the gap is indirect (I) or direct (D).

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

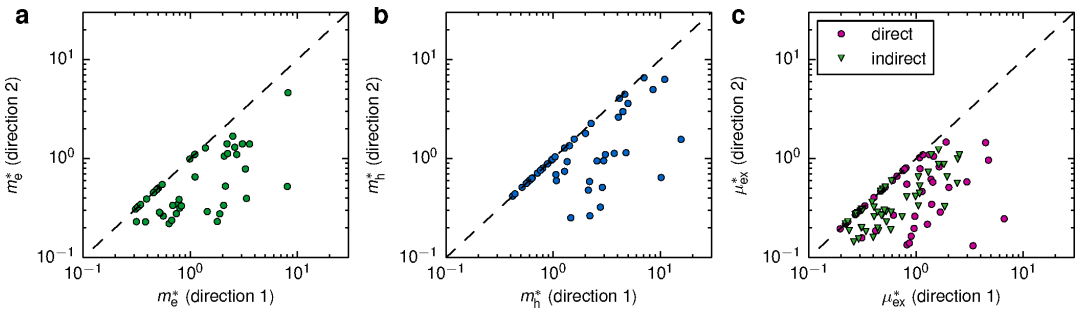


Figure 11. Effective electron (a), hole (b), and exciton (c) masses (in units of m_e) along the two principal directions obtained. The masses are calculated from the G_0W_0 band structures including spin−orbit interaction. Points on the dashed lines correspond to isotropic masses.

exciton masses along the two natural directions (we show both the direct and indirect exciton mass whether the material has direct or indirect gap).

3.6. Dielectric Function. The dielectric function is one of the most important material response functions. It relates the strength of an externally applied field to the total (screened) field in the material. In particular, it determines the strength of the electron−electron interaction and is a key ingredient in calculations of electronic states such as QP band structures and excitons.

For many purposes it is not necessary to know the precise spatial variation of the induced potentials but only its average value over a unit cell. The relation between the external potential and the averaged total potential is described by the *macroscopic* dielectric function, which can be obtained from the microscopic dielectric function according to

$$\frac{1}{\epsilon_M(\mathbf{q},\omega)} = \epsilon^{-1}_{\mathbf{G}=\mathbf{0},\mathbf{G}'=\mathbf{0}}(\mathbf{q},\omega) \tag{5}$$

Here, $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$ is the plane wave representation of the inverse microscopic dielectric function, which is a standard output of many electronic structure codes. For bulk semiconductors one usually refers to the $q = 0$ and $\omega = 0$ limit of ϵ_M as the dielectric *constant*.

In the case of a 2D material, eq 5 must be generalized as there is no natural unit cell over which to perform the average of the total field. If one restricts the averaging region to a slab of width d containing the 2D material, one arrives at the following expression:[43]

$$\frac{1}{\epsilon^{2D}_M(\mathbf{q}_{||})} = \frac{2}{d}\sum_{G_\perp} e^{iG_\perp z_0}\frac{\sin(G_\perp d/2)}{G_\perp}\epsilon^{-1}_{\mathbf{G0}}(\mathbf{q}_{||}) \tag{6}$$

We note that because of the averaging procedure, ϵ^{2D}_M takes the finite thickness of the material into account. We therefore refer to it as a quasi-2D dielectric function to distinguish it from a mathematically strict 2D quantity where the third dimension has been integrated out. As discussed in ref 43, $\epsilon^{2D}_M(\mathbf{q}_{||} = 0) = 1$, which implies that long wavelength perturbations are not screened by the 2D material at all. In particular, there is no direct analogue of the dielectric *constant* in 2D; any realistic model for screening in 2D materials must be q-dependent.

We have calculated $\epsilon^{2D}_M(q_{||})$ along the $\Gamma\rightarrow M$ and $\Gamma\rightarrow K$ directions for the 51 stable 2D semiconductors. We have found that this quantity is almost isotropic within the plane of the monolayer. The thickness of the averaging region has been set to $d = 2h$, where h is the thickness of the layer, but as shown in ref 43, the dielectric function is not very sensitive to this value; in particular, it is not very sensitive for the most important regime of $q < 1/d$.

As an example, Figure 2 shows the static macroscopic dielectric function of 2H-MoS_2. The linear increase for small q followed by a maximum and then a monotonic decrease toward 1 in the large q limit is characteristic for all 2D semiconductors. For comparison we also show the dielectric function of bulk MoS_2 for the same in-plane q vectors. To continue with the discussion in section 2.2, we note that it is the strong q-dependence of ϵ^{2D}_M for small q that is responsible for the very slow k-point convergence of the GW calculations.

To illustrate the variation in the dielectric properties of the monolayers, we show the slope of $\epsilon^{2D}_M(q)$ at $q = 0$ in Figure 12. Not surprisingly, the variation correlates well with the size of the electronic band gaps, also shown in the figure: large band gap materials have smaller dielectric function and vice versa. The slopes of the macroscopic dielectric function are listed in Table 5.

3.7. Excitons. One of the most characteristic features of atomically thin 2D semiconductors is the large binding energy of excitons.[15,41,64] The reason for this is the reduced screening due to the lower dimension which yields a stronger attraction between electrons and holes (see discussion in previous section). The conventional method for calculating exciton binding energies from first-principles is the Bethe−Salpeter equation (BSE). The BSE is computationally highly demanding and not suited for large-scale studies such as the present work. Instead, we use a recently developed 2D Mott−Wannier model for excitons that requires only the exciton effective mass and the quasi-2D dielectric function as input. In real space, the model takes the form of a 2D Schrödinger equation

$$\left[-\frac{1}{2\mu_{ex}}\nabla^2_{2D} + W(\mathbf{r})\right]\psi(\mathbf{r}) = E_b\psi(\mathbf{r}) \tag{7}$$

where μ_{ex} is the effective exciton mass and $W(r)$ is the $1/r$ Coulomb interaction between the electron and the hole screened by the nonlocal ϵ^{2D}_M. The model has been benchmarked against full BSE calculations for 2H-MoS_2 and 2H-WS_2, and the results were found to deviate by less than 0.1 eV.

The four basic assumptions behind the Mott−Wannier exciton model are (i) isotropic exciton masses, (ii) parabolic band structures close to the fundamental gap, (iii) the exciton is well-described by transitions between the valence and conduction band only, and (iv) the valence and conduction band

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

Table 5. Effective Electron and Hole Masses Together with the Direct and Indirect Exciton Masses Calculated from the G_0W_0 Quasiparticle Band Structures with Spin−Orbit Interaction Included[a,b]

name	m_e^* (m_e)	m_h^* (m_e)	μ_{ex}^{direct} (m_e)	$\mu_{ex}^{indirect}$ (m_e)	$d\epsilon_M^{2D}/dq|_{q=0}$ (Å)	E_b^{direct} (eV)	$E_b^{indirect}$ (eV)
2H-CrO_2	1.1/1.1	1.3/1.3	0.6/0.6	0.59/0.59	45	0.56	0.56
2H-CrS_2	1.1/1.1	0.88/0.88	0.49/0.49	0.49/0.49	61.5	0.42	0.42
2H-$CrSe_2$	1.1/1.1	0.97/0.97	0.52/0.52	0.52/0.52	73.6	0.37	0.37
2H-$CrTe_2$	0.99/0.99	0.89/0.89	0.47/0.47	0.47/0.47	95.4	0.30	0.30
2H-GeO_2	0.32/0.32	5/8.6	0.3/0.3	0.3/0.31	10.2	1.24	1.25
1T-GeO_2	0.34/0.34	4.1/2.6	0.81/0.81	0.32/0.3	7.39	1.97	1.55
1T-GeS_2	0.64/0.22	1.5/0.25	0.82/0.13	0.26/0.14	27.6	–	–
2H-HfO_2	2.5/1.7	4.7/4.5	1.1/1.5	1.6/1.2	9.85	–	–
1T-HfO_2	3.3/0.79	1.1/3.1	1.1/0.79	0.53/1.1	8.82	–	–
2H-HfS_2	−21/1.2	3/0.95	1.5/4.5	2.4/0.56	23.5	–	–
1T-HfS_2	1.4/0.29	0.63/0.63	0.78/0.78	0.44/0.2	27.7	0.80	–
2H-$HfSe_2$	−38/0.75	2.9/0.51	0.97/4.7	1.9/0.33	32.6	–	–
1T-$HfSe_2$	1.8/0.23	0.51/0.51	0.47/0.47	0.4/0.16	41.6	0.55	–
2H-$HfTe_2$	−48/0.46	0.93/1.4	−4.2/0.23	1.1/0.33	66.7	–	–
2H-MoO_2	0.51/0.51	0.8/0.8	0.75/0.75	0.31/0.31	31.4	0.75	0.62
2H-MoS_2	0.55/0.55	0.56/0.56	0.28/0.28	0.28/0.28	44.3	0.47	0.47
2H-$MoSe_2$	0.49/0.49	0.61/0.61	0.27/0.27	0.27/0.27	51.2	0.42	0.42
2H-$MoTe_2$	0.65/1.1	0.64/0.64	0.31/0.31	0.32/0.4	65.4	0.36	–
1T-NiO_2	1.1/2.1	4.2/33	0.62/1.4	0.87/1.8	35.8	–	–
1T-NiS_2	0.39/0.79	1.3/1.4	1.4/0.35	0.3/0.51	79.3	–	–
1T-$NiSe_2$	0.29/0.52	−58/2.9	0.26/0.98	0.29/0.44	121	–	–
1T-PbO_2	0.39/0.39	53/5	0.41/0.41	0.39/0.36	12.8	1.20	1.17
1T-PbS_2	0.83/0.33	10/0.64	0.96/0.2	0.48/0.27	32	–	–
1T-PdO_2	1.3/2.6	$6.8/1.1 \times 10^2$	0.58/1.4	1.1/2.5	28.1	–	–
1T-PdS_2	0.35/0.77	1.3/0.74	0.29/1.7	0.26/0.41	54.4	–	–
1T-$PdSe_2$	0.28/0.52	6.5/7.1	0.22/1.3	0.27/0.49	76.1	–	–
1T-$PdTe_2$	0.23/0.31	0.99/0.99	0.16/0.31	0.19/0.24	134	–	–
1T-PtO_2	1.1/2.2	1.6/16	0.47/1.1	0.65/2	21.2	–	–
1T-PtS_2	0.32/0.8	0.48/2.1	0.27/0.62	0.19/0.58	38.7	–	–
1T-$PtSe_2$	0.26/0.56	1.1/0.6	0.25/6.6	0.2/0.31	50.2	–	–
1T-$PtTe_2$	0.23/0.38	1.6/1.6	0.19/0.42	0.2/0.31	75.3	–	–
2H-SnO_2	0.31/0.31	6.3/11	0.33/0.33	0.29/0.3	12	1.16	1.13
1T-SnO_2	0.33/0.33	3/4.5	0.55/0.82	0.3/0.31	8.11	–	1.45
2H-SnS_2	0.69/0.34	2.3/2.3	0.2/0.2	0.53/0.29	24.2	0.64	–
1T-SnS_2	0.74/0.28	2.8/0.32	0.9/0.16	0.34/0.18	21.6	–	–
1T-$SnSe_2$	0.67/0.24	2.2/0.26	0.87/0.14	0.29/0.15	31.8	–	–
2H-TiO_2	1.4/2.2	5/3.6	1.5/1.9	1.1/1.4	14.2	–	–
1T-TiO_2	8.2/4.6	1.1/4.8	1.3/1.1	0.99/2.4	14.5	–	–
2H-TiS_2	−29/0.78	1/1	−21/0.83	1.1/0.45	38.1	–	–
2H-$TiSe_2$	8.1/0.52	0.63/0.63	47/0.32	0.59/0.29	55.6	–	–
2H-$TiTe_2$	3.4/0.4	1.1/0.69	1.6/0.83	0.74/0.26	129	–	–
2H-WO_2	0.45/0.45	0.76/0.76	0.78/0.78	0.28/0.28	26.7	0.84	0.68
2H-WS_2	0.46/0.46	0.42/0.42	0.22/0.22	0.22/0.22	39.9	0.48	0.48
2H-WSe_2	0.48/0.48	0.44/0.44	0.23/0.23	0.23/0.23	46.2	0.43	0.43
2H-ZrO_2	2.7/1.1	4.1/4.2	1.1/1	1.6/0.87	11	1.59	–
1T-ZrO_2	3.6/1.4	1.1/3.7	1.2/1.1	0.73/1.3	9.63	1.82	–
2H-ZrS_2	1.4/3.1	2.6/0.95	0.58/3	1.1/0.65	24.9	–	–
1T-ZrS_2	2/0.34	0.71/0.71	0.67/0.67	0.53/0.23	30.6	0.73	–
2H-$ZrSe_2$	1.3/1.4	2.2/0.59	0.51/2	0.82/0.41	34.1	–	–
1T-$ZrSe_2$	1.9/0.28	0.59/0.59	0.47/0.47	0.45/0.19	48	0.50	–
2H-$ZrTe_2$	2.1/0.53	1.8/2	3.4/0.13	0.99/0.41	74.9	–	–

[a]Negative masses occur in some directions because of bad fitting. This is usually the case if the band structure is very flat in one direction but highly varying in the other direction. Thus, negative masses generally mean that the mass in this direction is much larger than in the other direction. [b]The slope of the quasi-2D dielectric function at $q = 0$ is shown, and the exciton binding energies are obtained from a quasi-2D Mott−Wannier model.

wave functions are uniformly distributed over the layer, i.e. their profile along z can be approximated by a step function. While the dielectric functions were found to be very nearly isotropic for all materials, this is not the case for the exciton masses, see Figure 11c. While it is possible to modify the model to allow for anisotropic masses, we here limit ourselves to the materials with isotropic exciton masses. The exciton binding energies obtained from the model are shown as the dark region on the top of the

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

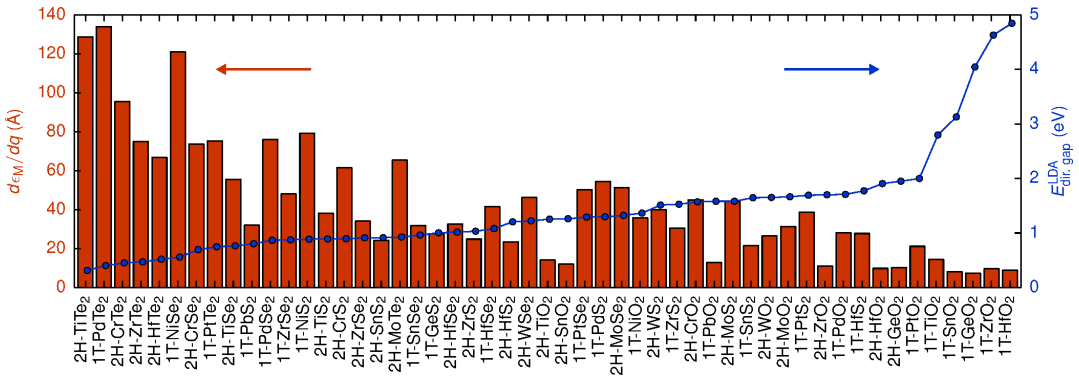


Figure 12. Slope of the static quasi-2D dielectric function, $\epsilon^{2D}_M(q)$, evaluated at $q = 0$. The materials are ordered according to their LDA direct band gap.

bars in Figure 13 (see also Table 5). The total height of the bar represents the G_0W_0 calculated QP gap. For direct (indirect)

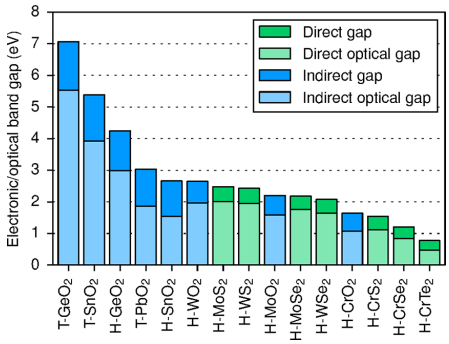


Figure 13. G_0W_0 band gaps (total bar height) and exciton binding energies (darker topmost part of the bar). The exciton binding energy was obtained from a quasi-2D Mott−Wannier model. Only materials with isotropic exciton masses are shown. The green and blue bars refer to indirect and direct band gaps and excitons, respectively.

band gap materials, we have used the direct (indirect) exciton mass in the model.

In accordance with earlier experimental and theoretical studies, we find strong exciton binding energies on the order of 20−30% of the band gap. In general, materials with larger QP band gaps have more strongly bound excitons. This follows from the correlation between the size of the band gap and the dielectric function in Figure 12: larger band gap implies a smaller dielectric function and thus a stronger electron−hole interaction. In Table 6, we compare our calculated exciton binding energies with optical data from experiments. We find good agreement for MoS_2, $MoSe_2$, and WSe_2, while the agreement is less satisfactory for $MoTe_2$ and WS_2. It should be noted, however, that the experimental exciton binding energy for $MoTe_2$ was obtained as the difference between the calculated G_0W_0 band gap and the position of the optical photoluminiscence peak. Thus, inaccuracies in the G_0W_0 band gap as well as substrate effects on the measured photoluminescence peak could explain the disagreement.

Table 6. Exciton Binding Energies (in Electronvolts) Calculated from the Mott−Wannier Model Compared to Experimental Values

name	E_b (model)	E_b (exptl)
2H-MoS_2	0.47	0.55[15]
2H-$MoSe_2$	0.42	0.5[15]
2H-$MoTe_2$	0.36	0.6[65,a]
2H-WS_2	0.48	0.66,[66] 0.71[67]
2H-WSe_2	0.43	0.38,[15] 0.37[68]

[a]The exciton binding energy is obtained by subtracting the energy of the measured exciton photoluminiscence peak from our calculated G_0W_0 band gap.

4. CONCLUSIONS

We have presented a detailed electronic structure study of 51 monolayer transition-metal dichalcogenides and -oxides. The 51 monolayers were chosen out of an initial set of 216 compounds as those having a finite band gap and a negative heat of formation. The calculated properties include the LDA band structure for in-plane lattice constants in a range around the equilibrium structure, the quasiparticle band structure at the equilibrium lattice constant evaluated in the G_0W_0 approximation and including spin−orbit coupling, the absolute positions of the conduction and valence band edges relative to vacuum, the effective electron and hole masses, and the static q-dependent dielectric functions. As an example, we showed how the computed data, in this case the effective masses and dielectric functions, can be used to obtain the lowest exciton binding energies from a 2D Mott−Wannier model.

Rather than providing a detailed account of the electronic structure of specific materials, we have chosen to focus on general trends and correlations in the electronic structure of the materials. However, because all the computed data is available in an open database, it is straightforward to retrieve and analyze data for specific materials in greater detail. We are presently working to expand the database to include other 2D materials and properties. We strongly believe that such a database will be useful both for guiding experimental efforts in the search for new 2D materials and as a platform for predicting properties of more complex materials such as van der Waals heterostructures.

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

■ ASSOCIATED CONTENT

Supporting Information

Figures showing the LDA and G_0W_0 band structures of all 51 stable nonmagnetic semiconductors and projected density of states. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.5b02950.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +45 4525 3188. E-mail: thygesen@fysik.dtu.dk.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Jens Jørgen Mortensen and Thomas Olsen for assistance with the G_0W_0 calculations and spin−orbit coupling. We acknowledge support from the Danish Council for Independent Researchs Sapere Aude Program, Grant 11-1051390. The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation, Project DNRF58.

■ REFERENCES

(1) Wang, Q. H.; Kalantar-Zadeh, K.; Kis, A.; Coleman, J. N.; Strano, M. S. Electronics and Optoelectronics of Two-Dimensional Transition Metal Dichalcogenides. *Nat. Nanotechnol.* **2012**, *7*, 699−712.

(2) Mak, K. F.; Lee, C.; Hone, J.; Shan, J.; Heinz, T. F. Atomically Thin MoS_2: A New Direct-Gap Semiconductor. *Phys. Rev. Lett.* **2010**, *105*, 136805.

(3) Splendiani, A.; Sun, L.; Zhang, Y.; Li, T.; Kim, J.; Chim, C.-Y.; Galli, G.; Wang, F. Emerging Photoluminescence in Monolayer MoS_2. *Nano Lett.* **2010**, *10*, 1271−1275.

(4) Balendhran, S.; Walia, S.; Nili, H.; Ou, J. Z.; Zhuiykov, S.; Kaner, R. B.; Sriram, S.; Bhaskaran, M.; Kalantar-zadeh, K. Two-Dimensional Molybdenum Trioxide and Dichalcogenides. *Adv. Funct. Mater.* **2013**, *23*, 3952−3970.

(5) Zeng, H.; Liu, G.-B.; Dai, J.; Yan, Y.; Zhu, B.; He, R.; Xie, L.; Xu, S.; Chen, X.; Yao, W. Optical Signature of Symmetry Variations and Spin-Valley Coupling in Atomically Thin Tungsten Dichalcogenides. *Sci. Rep* **2013**, *3*, 1608 DOI: 10.1038/srep01608.

(6) Zhang, C.; Johnson, A.; Hsu, C.-L.; Li, L.-J.; Shih, C.-K. Direct Imaging of Band Profile in Single Layer MoS_2 on Graphite: Quasiparticle Energy Gap, Metallic Edge States, and Edge Band Bending. *Nano Lett.* **2014**, *14*, 2443−2447.

(7) Britnell, L.; Ribeiro, R. M.; Eckmann, A.; Jalil, R.; Belle, B. D.; Mishchenko, A.; Kim, Y.-J.; Gorbachev, R. V.; Georgiou, T.; Morozov, S. V.; et al. Strong Light-Matter Interactions in Heterostructures of Atomically Thin Films. *Science* **2013**, *340*, 1311−1314.

(8) Bernardi, M.; Palummo, M.; Grossman, J. C. Extraordinary Sunlight Absorption and One Nanometer Thick Photovoltaics Using Two-Dimensional Monolayer Materials. *Nano Lett.* **2013**, *13*, 3664−3670.

(9) Kaasbjerg, K.; Thygesen, K. S.; Jacobsen, K. W. Phonon-Limited Mobility in *n*-Type Single-Layer MoS_2 from First Principles. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *85*, 115317.

(10) Kaasbjerg, K.; Thygesen, K. S.; Jauho, A.-P. Acoustic Phonon Limited Mobility in Two-Dimensional Semiconductors: Deformation Potential and Piezoelectric Scattering in Monolayer MoS_2 from First Principles. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 235312.

(11) Radisavljevic, B.; Radenovic, A.; Brivio, J.; Giacometti, V.; Kis, A. Single-Layer MoS_2 Transistors. *Nat. Nanotechnol.* **2011**, *6*, 147−150.

(12) Lopez-Sanchez, O.; Lembke, D.; Kayci, M.; Radenovic, A.; Kis, A. Ultrasensitive Photodetectors Based on Monolayer MoS_2. *Nat. Nanotechnol.* **2013**, *8*, 497−501.

(13) Xia, J.; Huang, X.; Liu, L.-Z.; Wang, M.; Wang, L.; Huang, B.; Zhu, D.-D.; Li, J.-J.; Gu, C.-Z.; Meng, X.-M. CVD Synthesis of Large-Area, Highly Crystalline $MoSe_2$ Atomic Layers on Diverse Substrates and Application to Photodetectors. *Nanoscale* **2014**, *6*, 8949−8955.

(14) Zhang, W.; Chuu, C.-P.; Huang, J.-K.; Chen, C.-H.; Tsai, M.-L.; Chang, Y.-H.; Liang, C.-T.; Chen, Y.-Z.; Chueh, Y.-L.; He, J.-H. Ultrahigh-Gain Photodetectors Based on Atomically Thin Graphene-MoS2 Heterostructures. *Sci. Rep.* **2014**, *4*, 3826 DOI: 10.1038/srep03826.

(15) Klots, A. R.; Newaz, A. K. M.; Wang, B.; Prasai, D.; Krzyzanowska, H.; Lin, J.; Caudel, D.; Ghimire, N. J.; Yan, J.; Ivanov, B. L. Probing Excitonic States in Suspended Two-Dimensional Semiconductors by Photocurrent Spectroscopy. *Sci. Rep.* **2014**, *4*, 6608 DOI: 10.1038/srep06608.

(16) Ross, J. S.; Klement, P.; Jones, A. M.; Ghimire, N. J.; Yan, J.; Mandrus, D. G.; Taniguchi, T.; Watanabe, K.; Kitamura, K.; Yao, W.; et al. Electrically Tunable Excitonic Light-Emitting Diodes Based on Monolayer WSe_2 p-n Junctions. *Nat. Nanotechnol.* **2014**, *9*, 268−272.

(17) Baugher, B. W. H.; Churchill, H. O. H.; Yang, Y.; Jarillo-Herrero, P. Optoelectronic Devices Based on Electrically Tunable p-n Diodes in a Monolayer Dichalcogenide. *Nat. Nanotechnol.* **2014**, *9*, 262−267.

(18) Pospischil, A.; Furchi, M. M.; Mueller, T. Solar-Energy Conversion and Light Emission in an Atomic Monolayer p-n Diode. *Nat. Nanotechnol.* **2014**, *9*, 257−261.

(19) Cao, T.; Wang, G.; Han, W.; Ye, H.; Zhu, C.; Shi, J.; Niu, Q.; Tan, P.; Wang, E.; Liu, B.; et al. Valley-Selective Circular Dichroism of Monolayer Molybdenum Disulphide. *Nat. Commun.* **2012**, *3*, 887.

(20) Zeng, H.; Dai, J.; Yao, W.; Xiao, D.; Cui, X. Valley Polarization in MoS_2Monolayers by Optical Pumping. *Nat. Nanotechnol.* **2012**, *7*, 490−493.

(21) Mak, K. F.; He, K.; Shan, J.; Heinz, T. F. Control of Valley Polarization in Monolayer MoS_2 by Optical Helicity. *Nat. Nanotechnol.* **2012**, *7*, 494−498.

(22) Xiao, D.; Liu, G.-B.; Feng, W.; Xu, X.; Yao, W. Coupled Spin and Valley Physics in Monolayers of MoS_2 and Other Group-VI Dichalcogenides. *Phys. Rev. Lett.* **2012**, *108*, 196802.

(23) Conley, H. J.; Wang, B.; Ziegler, J. I.; Haglund, R. F.; Pantelides, S. T.; Bolotin, K. I. Bandgap Engineering of Strained Monolayer and Bilayer MoS_2. *Nano Lett.* **2013**, *13*, 3626−3630.

(24) Hüser, F.; Olsen, T.; Thygesen, K. S. Quasiparticle GW Calculations for Solids, Molecules, and Two-Dimensional Materials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 235132.

(25) Liu, Q.; Li, L.; Li, Y.; Gao, Z.; Chen, Z.; Lu, J. Tuning Electronic Structure of Bilayer MoS_2 by Vertical Electric Field: A First-Principles Investigation. *J. Phys. Chem. C* **2012**, *116*, 21556−21562.

(26) Rostami, H.; Moghaddam, A. G.; Asgari, R. Effective Lattice Hamiltonian for Monolayer MoS_2: Tailoring Electronic Structure with Perpendicular Electric and Magnetic Fields. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 085440.

(27) Pedersen, T. G.; Flindt, C.; Pedersen, J.; Mortensen, N. A.; Jauho, A.-P.; Pedersen, K. Graphene Antidot Lattices: Designed Defects and Spin Qubits. *Phys. Rev. Lett.* **2008**, *100*, 136804.

(28) Gao, G.; Gao, W.; Cannuccia, E.; Taha-Tijerina, J.; Balicas, L.; Mathkar, A.; Narayanan, T. N.; Liu, Z.; Gupta, B. K.; Peng, J.; et al. Artificially Stacked Atomic Layers: Toward New van der Waals Solids. *Nano Lett.* **2012**, *12*, 3518−3525.

(29) Geim, A. K.; Grigorieva, I. V. Van der Waals Heterostructures. *Nature* **2013**, *499*, 419−425.

(30) Elías, A. L.; Perea-López, N.; Castro-Beltrán, A.; Berkdemir, A.; Lv, R.; Feng, S.; Long, A. D.; Hayashi, T.; Kim, Y. A.; Endo, M.; et al. Controlled Synthesis and Transfer of Large-Area WS_2 Sheets: From Single Layer to Few Layers. *ACS Nano* **2013**, *7*, 5235−5242.

(31) Shockley, W.; Queisser, H. J. Detailed Balance Limit of Efficiency of *p-n* Junction Solar Cells. *J. Appl. Phys.* **1961**, *32*, 510−519.

(32) Björkman, T.; Gulans, A.; Krasheninnikov, A. V.; Nieminen, R. M. Are We van der Waals Ready? *J. Phys.: Condens. Matter* **2012**, *24*, 424218.

(33) Coleman, J. N.; Lotya, M.; O'Neill, A.; Bergin, S. D.; King, P. J.; Khan, U.; Young, K.; Gaucher, A.; De, S.; Smith, R. J.; et al. Two-

Dimensional Nanosheets Produced by Liquid Exfoliation of Layered Materials. *Science* **2011**, *331*, 568−571.
(34) Lebègue, S.; Björkman, T.; Klintenberg, M.; Nieminen, R. M.; Eriksson, O. Two-Dimensional Materials from Data Filtering and Ab Initio Calculations. *Phys. Rev. X* **2013**, *3*, 031002.
(35) Cui, X.; Lee, G.-H.; Kim, Y. D.; Arefe, G.; Huang, P. Y.; Lee, C.-H.; Chenet, D. A.; Zhang, X.; Wang, L.; Ye, F. et al. Multi-terminal Electrical Transport Measurements of Molybdenum Disulphide Using Van Der Waals Heterostructure Device Platform. *arXiv:1412.5977 [cond-mat]* **2014**, arXiv: 1412.5977.
(36) Ataca, C.; Şahin, H.; Ciraci, S. Stable, Single-Layer MX_2 Transition-Metal Oxides and Dichalcogenides in a Honeycomb-Like Structure. *J. Phys. Chem. C* **2012**, *116*, 8983−8999.
(37) Shi, H.; Pan, H.; Zhang, Y.-W.; Yakobson, B. I. Quasiparticle Band Structures and Optical Properties of Strained Monolayer MoS_2 and WS_2. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 155304.
(38) Cheiwchanchamnangij, T.; Lambrecht, W. R. L. Quasiparticle Band Structure Calculation of Monolayer, Bilayer, and Bulk MoS_2. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *85*, 205302.
(39) Komsa, H.-P.; Krasheninnikov, A. V. Effects of Confinement and Environment on the Electronic Structure and Exciton Binding Energy of MoS_2 from First Principles. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *86*, 241201.
(40) Ataca, C.; Ciraci, S. Functionalization of Single-Layer MoS_2 Honeycomb Structures. *J. Phys. Chem. C* **2011**, *115*, 13303−13311.
(41) Ramasubramaniam, A. Large Excitonic Effects in Monolayers of Molybdenum and Tungsten Dichalcogenides. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *86*, 115409.
(42) Molina-Sánchez, A.; Sangalli, D.; Hummer, K.; Marini, A.; Wirtz, L. Effect of Spin-Orbit Interaction on the Optical Spectra of Single-Layer, Double-Layer, and Bulk MoS_2. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 045412.
(43) Huser, F.; Olsen, T.; Thygesen, K. S. How Dielectric Screening in Two-Dimensional Crystals Affects the Convergence of Excited-State Calculations: Monolayer MoS_2. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 245309.
(44) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dułak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; et al. Electronic Structure Calculations with GPAW: A Real-Space Implementation of the Projector Augmented-Wave Method. *J. Phys.: Condens. Matter* **2010**, *22*, 253202.
(45) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(46) Qian, X.; Liu, J.; Fu, L.; Li, J. Quantum Spin Hall Effect in Two-Dimensional Transition Metal Dichalcogenides. *Science* **2014**, *346*, 1344−1347.
(47) Tongay, S., Sahin, H., Ko, C., Luce, A., Fan, W., Liu, K., Zhou, J., Huang, Y.-S., Ho, C.-H., Yan, J. et al. Monolayer Behaviour in Bulk ReS_2 Due to Electronic and Vibrational Decoupling. *Nat. Commun.* **2014**5.
(48) Voiry, D.; Yamaguchi, H.; Li, J.; Silva, R.; Alves, D. C. B.; Fujita, T.; Chen, M.; Asefa, T.; Shenoy, V. B.; Eda, G.; et al. Enhanced Catalytic Activity in Strained Chemically Exfoliated WS_2 Nanosheets for Hydrogen Evolution. *Nat. Mater.* **2013**, *12*, 850−855.
(49) Kuisma, M.; Ojanen, J.; Enkovaara, J.; Rantala, T. T. Kohn-Sham Potential with Discontinuity for Band Gap Materials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 115106.
(50) Dewhurst, K., Sharma, S., Nordström, L., Cricchio, F., Bultmark, F., Granäs, O., Gross, H. The Elk FP-LAPW Code. http://elk.sourceforge.net/ (accessed February 2, 2015), 2014.
(51) Sundararaman, R.; Arias, T. A. Regularization of the Coulomb Singularity in Exact Exchange by Wigner-Seitz Truncated Interactions: Towards Chemical Accuracy in Nontrivial Systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 165122.
(52) Liang, Y.; Huang, S.; Soklaski, R.; Yang, L. Quasiparticle Band-Edge Energy and Band Offsets of Monolayer of Molybdenum and Tungsten Chalcogenides. *Appl. Phys. Lett.* **2013**, *103*, 042106.
(53) Klimeš, J.; Kaltak, M.; Kresse, G. Predictive GW Calculations Using Plane Waves and Pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90*, 075125.
(54) Stevanović, V.; Lany, S.; Zhang, X.; Zunger, A. Correcting Density Functional Theory for Accurate Predictions of Compound Enthalpies of Formation: Fitted Elemental-Phase Reference Energies. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *85*, 115104.
(55) Castelli, I. E.; Olsen, T.; Datta, S.; Landis, D. D.; Dahl, S.; Thygesen, K. S.; Jacobsen, K. W. Computational Screening of Perovskite Metal Oxides for Optimal Solar Light Capture. *Energy Environ. Sci.* **2012**, *5*, 5814−5819.
(56) Ugeda, M. M.; Bradley, A. J.; Shi, S.-F.; da Jornada, F. H.; Zhang, Y.; Qiu, D. Y.; Ruan, W.; Mo, S.-K.; Hussain, Z.; Shen, Z.-X.; et al. Giant Bandgap Renormalization and Excitonic Effects in a Monolayer Transition Metal Dichalcogenide Semiconductor. *Nat. Mater.* **2014**, *13*, 1091−1095.
(57) Yeh, T.-F.; Teng, C.-Y.; Chen, S.-J.; Teng, H. Nitrogen-Doped Graphene Oxide Quantum Dots as Photocatalysts for Overall Water-Splitting Under Visible Light Illumination. *Adv. Mater. (Weinheim, Ger.)* **2014**, *26*, 3297−3303.
(58) Singh, A. K.; Mathew, K.; Zhuang, H. L.; Hennig, R. G. Computational Screening of 2D Materials for Photocatalysis. *J. Phys. Chem. Lett.* **2015**, *6*, 1087−1098.
(59) Trasatti, S. The Absolute Electrode Potential: An Explanatory Note (Recommendations 1986). *Pure Appl. Chem* **1986**, *58*, 955−966.
(60) Nørskov, J. K.; Rossmeisl, J.; Logadottir, A.; Lindqvist, L.; Kitchin, J. R.; Bligaard, T.; Jónsson, H. Origin of the Overpotential for Oxygen Reduction at a Fuel-Cell Cathode. *J. Phys. Chem. B* **2004**, *108*, 17886−17892.
(61) Butler, M. A.; Ginley, D. S. Prediction of Flatband Potentials at Semiconductor-Electrolyte Interfaces from Atomic Electronegativities. *J. Electrochem. Soc.* **1978**, *125*, 228−232.
(62) Putz, M. V.; Russo, N.; Sicilia, E. About the Mulliken Electronegativity in DFT. *Theor. Chem. Acc.* **2005**, *114*, 38−45 DOI: 10.1007/s00214-005-0641-4.
(63) Perdew, J. P.; Levy, M. Physical Content of the Exact Kohn-Sham Orbital Energies: Band Gaps and Derivative Discontinuities. *Phys. Rev. Lett.* **1983**, *51*, 1884−1887.
(64) Wirtz, L.; Marini, A.; Rubio, A. Excitons in Boron Nitride Nanotubes: Dimensionality Effects. *Phys. Rev. Lett.* **2006**, *96*, 126104.
(65) Ruppert, C.; Aslan, O. B.; Heinz, T. F. Optical Properties and Band Gap of Single- and Few-Layer $MoTe_2$ Crystals. *Nano Lett.* **2014**, *14*, 6231−6236.
(66) Ye, Z.; Cao, T.; O'Brien, K.; Zhu, H.; Yin, X.; Wang, Y.; Louie, S. G.; Zhang, X. Probing Excitonic Dark States in Single-Layer Tungsten Disulphide. *Nature* **2014**, *513*, 214−218.
(67) Zhu, B.; Chen, X.; Cui, X. Exciton Binding Energy of Monolayer WS_2 *arXiv:1403.5108 [cond-mat]* **2014**, arXiv:1403.5108.
(68) He, K.; Kumar, N.; Zhao, L.; Wang, Z.; Mak, K. F.; Zhao, H.; Shan, J. Tightly Bound Excitons in Monolayer WSe_2. *Phys. Rev. Lett.* **2014**, *113*, 026803.

DOI: 10.1021/acs.jpcc.5b02950
J. Phys. Chem. C 2015, 119, 13169−13183

Paper II

Chengjun Jin, Filip A. Rasmussen, and Kristian S. Thygesen. "Tuning the Schottky Barrier at the Graphene/MoS2 Interface by Electron Doping: Density Functional Theory and Many-Body Calculations". In: *The Journal of Physical Chemistry C* 119.34 (August 27, 2015), pages 19928–19933

THE JOURNAL OF PHYSICAL CHEMISTRY C

Article

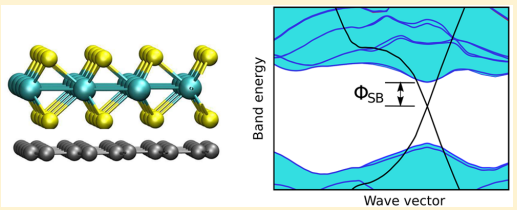
pubs.acs.org/JPCC

Tuning the Schottky Barrier at the Graphene/MoS_2 Interface by Electron Doping: Density Functional Theory and Many-Body Calculations

Chengjun Jin, Filip A. Rasmussen, and Kristian S. Thygesen*

Center for Atomic-Scale Materials Design (CAMD) and Center for Nanostructured Graphene (CNG), Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

ABSTRACT: Using ab initio calculations we investigate the energy level alignment at the graphene/MoS_2 heterostructure and the use of electron doping as a strategy to lower the Schottky barrier and achieve a low-resistance Ohmic contact. For the neutral heterostructure, density functional theory (DFT) with a generalized gradient approximation predicts a Schottky barrier height of 0.18 eV, whereas the G_0W_0 method increases this value to 0.60 eV. While the DFT band gap of MoS_2 does not change when the heterostructure is formed, the G_0W_0 gap is reduced by 0.30 eV as a result of the enhanced screening by the graphene layer. In contrast to the case of metal substrates, where the band alignment is governed by Pauli repulsion-induced interface dipoles, the graphene/MoS_2 heterostructure shows only a negligible interface dipole. As a consequence, the band alignment at the neutral heterostructure is not changed when the two layers are brought into contact. We systematically follow the band alignment as a function of doping concentration and find that the Fermi level of the graphene crosses the MoS_2 conduction band at a doping concentration of around 10^{12} cm^{-2}. The variation of the energy levels with doping concentration is shown to be mainly governed by the electrostatic potential resulting from the doping charge.



INTRODUCTION

Atomically thin two-dimensional (2D) materials such as graphene and transition-metal dichalcogenides (TMDs) are presently being intensively investigated because of their unique and easily tunable properties, which make them promising candidates for next-generation (opto)electronics devices. The TMDs, with the chemical formula MX_2 (X = S, Se, Te; M = transition metal), represent a particularly interesting class of 2D semiconductors that can be exfoliated from the bulk or grown by chemical vapor deposition, similar to the methods used to produce graphene.[1,2] However, in contrast to the semimetal graphene, several TMDs possess a finite band gap. Combined with their strong interaction with light,[3] resulting from the direct nature of the band gap and the large oscillator strength between valence-band and conduction-band states, this has recently opened the possibility of using these novel 2D materials for a range of applications including ultrathin field-effect transistors (FETs),[4] optical sensors,[3] and solar cells.[5]

With intrinsic room-temperature mobilities of around 400 cm^2 V^{-1} s^{-1} [6] and current on/off ratios exceeding 10^8,[4] monolayer MoS_2 has great potential as a channel material in ultrathin FET devices. Unfortunately, because of the Fermi level pinning at the metal/MoS_2 interface,[7] most metals form Schottky contacts to MoS_2, preventing efficient electron injection/extraction. As a consequence, the transport properties of two-terminal metal−MoS_2−metal devices are often limited by the contact resistance rather than the intrinsic resistance of the MoS_2 channel. This makes the search for low-contact-resistance materials one of the most important challenges in the field of 2D electronics.[8−11]

Recent experiments have demonstrated good electrical contacts between graphene and few-layer MoS_2.[12,13] Moreover, the tunability of the graphene work function, e.g., by electrostatic doping,[14] can be exploited to adjust the graphene Fermi level to the MoS_2 conduction band minimum (CBM) to completely remove the Schottky barrier at the interface. Supposedly, the use of work-function-tuned graphene electrodes is not limited to MoS_2 but could be used to achieve low-resistance contacts to other 2D materials, thus paving the road for all-2D integrated circuits.[15]

Previous studies have used density functional theory (DFT) to investigate the Schottky barrier of the graphene/MoS_2 heterostructure. Yu et al.[13] demonstrated that the Schottky barrier height (SBH) of 0.4 eV can be removed by decreasing the graphene work function using an electric field. Furthermore, Leong et al.[16] showed that the SBH can be reduced by using a nickel-etched graphene electrode because of its lower work function compared with graphene. In general, the relative change in the SBH is correctly described by DFT, although the absolute SBH is questionable.

Experience from metal/molecule interfaces has shown that image-charge screening can induce significant shifts in the

Received: June 11, 2015
Revised: July 30, 2015
Published: July 31, 2015

DOI: 10.1021/acs.jpcc.5b05580
J. Phys. Chem. C 2015, 119, 19928−19933

molecular energy levels when a molecule is weakly physisorbed on a metal surface.[17] Typically, this effect reduces the energy gap between the occupied and empty energy levels by an amount that scales as $1/d$, where d is the distance to the surface. Naively, one could expect that this effect would not be relevant for a 2D crystal since the potential associated with an electron/hole state delocalized over the crystal area A vanishes in the limit $A \to \infty$. However, the delocalized wave function is just a probability amplitude, and in reality the electron has a definite, albeit unknown, position within the 2D material. Consequently, the field associated with the negative point-charge electron will be screened by the substrate, and this lowers its energy. Previous *GW* calculations have indeed shown the effect for hexagonal boron nitride adsorbed on graphite,[18] and it was recently demonstrated experimentally for WSe_2 on graphite.[19] Importantly, the image-charge effect is completely missed by all of the standard exchange−correlation (xc) functionals, including hybrids.[20] Consequently, many-body methods like the *GW* approximation must be applied for predictive modeling of the energy level alignment at 2D interfaces.

In this paper we present a detailed first-principles analysis of the mechanism governing the energy level alignment at the graphene/MoS_2 interface. We find that as a result of the small wave function overlap, an insignificant dipole is formed at the interface. This is qualitatively different from metal/MoS_2, contacts where Pauli repulsion leads to the formation of significant interface dipoles that in turn govern the level alignment. Instead, we show that the level alignment at the neutral graphene/MoS_2 interface can be obtained by aligning the vacuum levels for the isolated layers and correcting the MoS_2 band edges by ±0.15 eV to account for the image-charge effect. Our G_0W_0 calculations then predict an SBH of 0.60 eV, in good agreement with recent experiments.[21,22] For the doped interface, the MoS_2 band edges and the graphene states must also be shifted to account for the change in the electrostatic potential created by the doping charge. When this is done, the band alignment for different doping concentrations can be accurately described without having to perform a calculation for the heterostructure.

COMPUTATIONAL METHODS

All of the calculations were performed with the GPAW electronic structure code.[23] DFT calculations used a plane-wave basis set with a cutoff energy of 600 eV. Unless otherwise stated, the Perdew−Burke−Ernzerhof (PBE) xc functional was employed. To compensate for the lattice mismatch between MoS_2 (a = 3.16 Å) and graphene (a = 2.46 Å), we used a supercell comprising 4 × 4 primitive MoS_2 cells and 5 × 5 primitive graphene cells (see Figure 1a). For this supercell the lattice mismatch is reduced to 1.9%. It should be noted that graphene is stretched to match the MoS_2 lattice. The resulting increase of 0.17 eV in the graphene work function due to the strain has been corrected in all the figures and tables presented in the paper. To simulate electron doping, a Jellium slab with a thickness of 1 Å was employed to work as an electrostatic gate. The Jellium slab was placed 5 Å away from the graphene sheet of the heterostructure (see Figure 2a). For the graphene/MoS_2 heterostructure, the added charge is localized on the graphene as long as the Fermi level is below the MoS_2 conduction band.

The G_0W_0 calculations[18] were performed for graphene, monolayer MoS_2, and the graphene/MoS_2 heterostructure using a minimal unit cell. For the heterostructure we used four different lattice constants ranging from 2.46 to 3.16 Å,

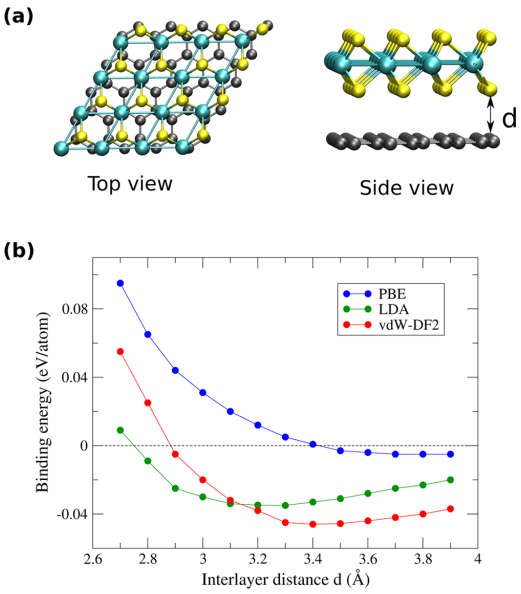


Figure 1. (a) Supercell used to model the graphene/MoS_2 heterostructure, comprising 4 × 4 primitive cells of MoS_2 and 5 × 5 primitive cells of graphene, resulting in a 1.9% lattice mismatch. The graphene has been stretched to match the MoS_2 lattice constant. (b) Binding energy curves for graphene/MoS_2 calculated using the PBE, LDA, and vdW-DF2 xc functionals.

corresponding to relaxed graphene and relaxed MoS_2, respectively. The *GW* self-energy was calculated from the PBE single-particle wave functions and energies. The plane-wave cutoff used to represent the screened interaction and self-energy was varied between 200 and 300 eV and extrapolated to infinite cutoff.[18] The frequency dependence was represented on a nonlinear grid, thereby avoiding the use of a plasmon-pole approximation. We used a truncated Coulomb interaction to avoid screening from periodically repeated layers[24] and a 30 × 30 k-point grid to sample the 2D Brillouin zone.

RESULTS AND DISCUSSION

Binding energy curves for the graphene/MoS_2 heterostructure were calculated using the PBE, local density approximation (LDA), and van der Waals density functional 2 (vdW-DF2) xc functionals (see Figure 1b). While PBE does not predict any binding, the LDA and vdW-DF2 functionals predict binding distances of 3.3 and 3.4 Å, respectively, in agreement with previous findings.[25] In this work, we used a binding distance of 3.4 Å, but our conclusions are not too sensitive to the exact binding distance.

In Figure 2a we show Δn, the change in the electron density upon formation of the heterostructure, which is given by

$$\Delta n = n[\text{graphene/MoS}_2] - n[\text{graphene}] - n[\text{MoS}_2] \quad (1)$$

The induced density is shown for doping concentrations of zero (solid line) and 3×10^{12} cm^{-2} (dashed line). The electron density is redistributed from the graphene toward the MoS_2 layer. However, the charge transfer is quite small, and the resulting surface dipole is 3 orders of magnitude smaller than

DOI: 10.1021/acs.jpcc.5b05580
J. Phys. Chem. C 2015, 119, 19928−19933

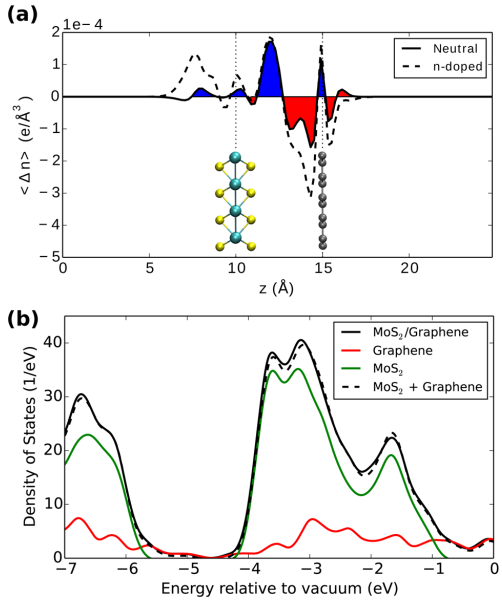


Figure 2. (a) Change in the electron density per supercell upon formation of the graphene/MoS_2 heterostructure. The induced densities for both the neutral and n-doped (3×10^{12} cm^{-2}) systems are shown. In the doped case, the Jellium slab was placed 5 Å away on the graphene side, as indicated by the gray bar. (b) Density of states (DOS) of isolated MoS_2 and graphene compared to the DOS of the graphene/MoS_2 heterostructure.

that of the Au/MoS_2 contact.[7] As a consequence, the charge redistribution at the interface does not change the electronic structure. This is evident from the density of states (DOS) shown in Figure 2b. The DOS of the heterostructure simply equals the sum of the DOS of the isolated layers. The energy scales are aligned relative to the vacuum level, which is taken as the asymptotic value of the Hartree potential in the vacuum region. As the DOS is calculated with PBE, the band gap of MoS_2 is only 1.73 eV, which is significantly smaller than the experimental quasi-particle gap of 2.5−2.7 eV.[26,27] Furthermore, PBE does not predict a change in the MoS_2 band gap when MoS_2 is combined with graphene. This reflects the inability of the PBE functional to capture the image-charge effect, as will be discussed below.

To obtain a more realistic description of the band alignment at the interface, we performed G_0W_0 band structure calculations. Because of the computational cost of such calculations, we considered lattice-matched heterostructures using four different lattice constants ranging from 2.46 Å (the graphene lattice constant) to 3.16 Å (the MoS_2 lattice constant). While the band structures vary significantly with lattice constant, we consistently observe that the MoS_2 band gap is reduced symmetrically by around 0.3 eV relative to the band gap of isolated MoS_2 with the same lattice constant. The image-charge band-gap renormalization is illustrated in Figure 3a. Taken together with the observation that the formation of the heterostructure does not lead to interface dipoles or hybridization-driven band structure effects, this implies that the band structure of the heterostructure can be obtained by superimposing the G_0W_0 band structures of the isolated layers calculated for their respective equilibrium structures, aligned relative to a common vacuum level and with the MoS_2 band edges shifted by ±0.15 eV to account for the image-charge effect.

The G_0W_0 band structure of the graphene/MoS_2 heterostructure is shown in Figure 3b, and the energy level positions aligned to the vacuum are listed in Table 1. We obtain a graphene work function of 4.29 eV using PBE, while G_0W_0 gives a work function of 4.45 eV. The latter is in perfect agreement with the experimental value.[14] For freestanding MoS_2, the G_0W_0 band gap of 2.73 eV is in good agreement with experiments,[26,27] and the value is further reduced to 2.43 eV when MoS_2 is contacted with graphene because of the enhanced screening. The distance from the graphene Dirac point to the MoS_2 conduction band minimum (i.e., the SBH) is

Table 1. Energy Level Positions Aligned to the Vacuum for the MoS_2 before and after Contacting Graphene (Gr) Calculated with PBE and G_0W_0, Along with the SBH Φ_{SB} (All Energies Are in Units of eV)

	Gr	isolated/Gr-contacted MoS_2			
	E_F	CBM	VBM	band gap	Φ_{SB}
PBE	−4.29	−4.11/−4.11	−5.84/−5.84	1.73/1.73	0.18
G_0W_0	−4.45	−3.70/−3.85	−6.43/−6.28	2.73/2.43	0.60

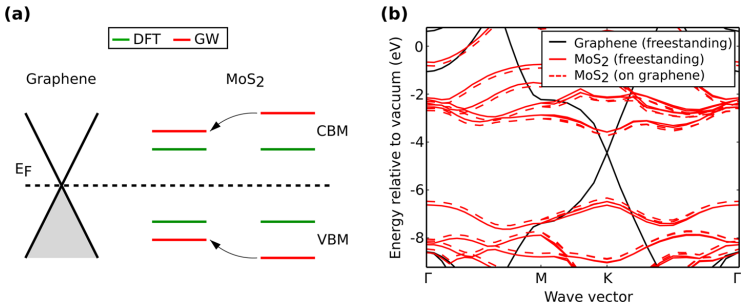


Figure 3. (a) Schematic of the image-charge band-gap renormalization. When a 2D semiconductor is placed on a (semi)metallic surface, the quasi-particle energy levels are shifted toward the Fermi level as a consequence of the enhanced screening. For MoS_2 on graphene, the band gap is reduced by 0.30 eV. (b) G_0W_0 band structures of free-standing graphene, MoS_2, and MoS_2 on graphene.

DOI: 10.1021/acs.jpcc.5b05580
J. Phys. Chem. C 2015, 119, 19928−19933

0.18 eV with PBE and 0.60 eV with G_0W_0. This is again in good agreement with the recently reported experimental value of 0.5 eV based on electron tunneling measurements on graphene/hBN/MoS_2 van der Waals heterostructures[21] and the value of 0.32 eV obtained from scanning tunneling spectroscopy of MoS_2 on graphite.[22]

Although PBE severely underestimates the band gap of MoS_2, the SBH of 0.18 eV is in reasonable agreement with G_0W_0 and experiments. We therefore performed PBE calculations to study the evolution of the band alignment as a function of doping concentration. As we will show, the change in the band alignment induced by the doping is a purely electrostatic effect that should be well-described by DFT using PBE. In the left panel of Figure 4 we show the dependence of the graphene Dirac point and Fermi level on the doping concentration for isolated graphene. In agreement with earlier work, we find that the graphene work function is highly dependent on the doping level.[28] The band edges of pristine MoS_2 are indicated by the gray-shaded regions. The right panel of Figure 4 shows the band alignment for the doped heterostructure. At doping concentrations of around 1×10^{12} cm^{-2}, the Fermi level crosses the MoS_2 CBM. After this point, any additional electrons reside in both the graphene and the MoS_2 layer.

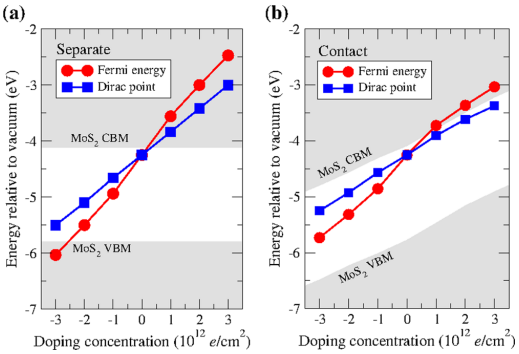


Figure 4. Positions of the Fermi level (E_F) and the Dirac point (E_{Dirac}) of graphene as functions of the doping level (a) before and (b) after contacting with MoS_2, calculated using the PBE xc functional. The MoS_2 conduction and valence bands are indicated by the gray areas.

The variation of the energy levels with doping concentration shown in Figure 4b is governed by the electrostatic potential associated with the dopant electrons. For n-type doping, the Fermi level is shifted above the Dirac point as a result of the band filling. Meanwhile, the positions of all of the graphene energy levels are shifted up by an amount given by the local electrostatic potential due to the dopant electrons at the graphene layer (see Figure 5). The shift due to the local potential is denoted by ΔV. Similarly, the MoS_2 states are shifted by the magnitude of ΔV at the position of the MoS_2 layer. The potential created by the dopant electrons in the graphene is naturally larger in the region of graphene than in the region of MoS_2. This is the reason why the Dirac point shows a slightly stronger variation with doping than the MoS_2 band edges for doping concentrations below 1×10^{12} cm^{-2}. When the Fermi level enters the MoS_2 CBM, the additional dopant electrons are shared between the graphene and MoS_2. As a consequence, the increase in the MoS_2 band edges with doping concentration becomes slightly stronger at this point, and the increase in the Dirac point becomes slightly weaker. This can be seen as small kinks in the band energies versus doping in Figure 4b.

Figure 6 shows the change in the position of the Fermi level relative to the MoS_2 CBM (i.e., the change in SBH) upon formation of the heterostructure as a function of the graphene work function (red curve). A slope of zero corresponds to perfect tunability of the level alignment, whereas a slope of 1 corresponds to complete Fermi level pinning. It should be noted that for pristine graphene the shift is essentially zero because of the negligible interface dipole, as discussed earlier in this paper. For finite doping concentrations, the change in the SBH is governed by the electrostatic potential of the dopant electrons. To show this more clearly, we have also plotted the change in the SBH after correction for the difference in the dopant-induced electrostatic potential ΔV at the graphene and MoS_2 layers (blue curve). To obtain this correction, ΔV was

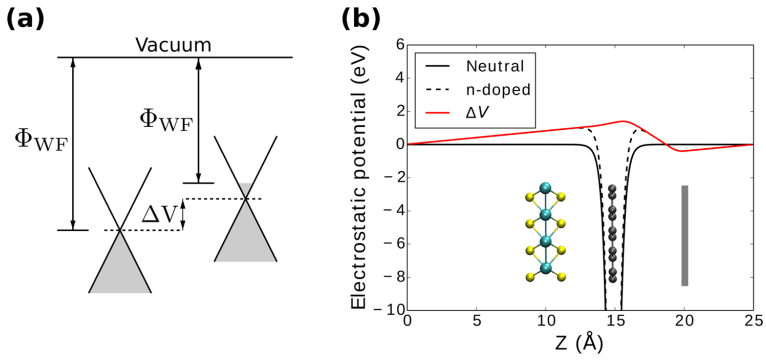


Figure 5. (a) The two contributions to the increase in work function of n-doped graphene. First, the Fermi level increases as a result of the band filling. Second, all of the energy levels are pushed up by the electrostatic potential created by the dopant charge density. (b) The electrostatic potential ΔV created by the dopant electrons. In the doped case, the Jellium slab was placed 5 Å away on the graphene side, as indicated by the gray bar. Note that the asymptotic behavior of ΔV (red curve) is a consequence of the periodic boundary conditions imposed on the supercell. We have checked that the band energies are converged with respect to the amount of vacuum included in the supercell.

DOI: 10.1021/acs.jpcc.5b05580
J. Phys. Chem. C 2015, 119, 19928−19933

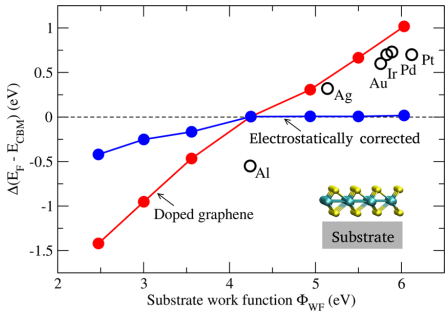


Figure 6. Change in the position of the Fermi level relative to the MoS_2 conduction band minimum, $\Delta(E_F - E_{CBM})$, before and after MoS_2 contacts the doped graphene as a function of the substrate (i.e., doped graphene) work function (red curve). Correcting the CBM of the isolated MoS_2 layer by the value of the local potential created by the doping electrons (i.e., ΔV in Figure 5b) results in the blue curve. Results for MoS_2 on different metal substrates are shown by black circles.[7]

determined from a DFT-PBE calculation on isolated doped graphene. For doping concentrations below 1×10^{12} cm^{-2}, the electrostatically corrected band edges are in perfect agreement with those obtained from the full heterostructure calculation. For higher doping concentrations, the small deviation is due to the fact that the additional dopant electrons are shared between graphene and MoS_2, as shown in Figure 2a for the n-doped case.

Finally, in Figure 6 we also show for comparison the changes in the SBH for MoS_2 on different metal surfaces.[7] Although the trend is similar to the case of doped graphene, we stress that the origin of the change in the SBH is very different: while for doped graphene the effect is a result of the electrostatic potential of the charged graphene layer, the pinning effect in the metals is due to the formation of an interface dipole stemming from the Pauli repulsion between the metal and MoS_2 wave functions.

CONCLUSIONS

We have presented a systematic study of the energy band alignment at a graphene/MoS_2 heterostructure using DFT and many-body G_0W_0 calculations. For the neutral interface, the image-charge effect reduces the MoS_2 band gap by 0.30 eV, resulting in a Schottky barrier of 0.60 eV, in good agreement with experiments. In contrast to metal/MoS_2 contacts, the graphene/MoS_2 interface presents a negligible interface dipole, implying that the band alignment at the interface can be obtained from the freestanding layers by aligning the vacuum levels. By doping of the graphene layer, the SBH can be tuned to some extent. However, despite the absence of a significant interface dipole, some degree of Fermi level pinning arises because of the electrostatic potential associated with the dopant electrons, which partially ties the energy of the MoS_2 states to the graphene Fermi level. Nevertheless, the Schottky barrier disappears for doping concentrations of around 10^{12} cm^{-2}, showing that doping control indeed provides an opportunity to realize low contact resistance for TMD-based devices.

AUTHOR INFORMATION

Corresponding Author

*E-mail: thygesen@fysik.dtu.dk.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge support from the Danish Council for Independent Research's Sapere Aude Program (Grant 11-1051390). The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation (Project DNRF58).

REFERENCES

(1) Wang, Q. H.; Kalantar-Zadeh, K.; Kis, A.; Coleman, J. N.; Strano, M. S. Electronics and Optoelectronics of Two-Dimensional Transition Metal Dichalcogenides. *Nat. Nanotechnol.* **2012**, *7*, 699−712.
(2) Rasmussen, F. A.; Thygesen, K. S. Computational 2D Materials Database: Electronic Structure of Transition-Metal Dichalcogenides and Oxide. *J. Phys. Chem. C* **2015**, *119*, 13169−13183.
(3) Britnell, L.; Ribeiro, R. M.; Eckmann, A.; Jalil, R.; Belle, B. D.; Mishchenko, A.; Kim, Y.-J.; Gorbachev, R. V.; Georgiou, T.; Morozov, S. V.; et al. Strong Light-Matter Interactions in Heterostructures of Atomically Thin Films. *Science* **2013**, *340*, 1311−1314.
(4) Radisavljevic, B.; Radenovic, A.; Brivio, J.; Giacometti, V.; Kis, A. Single-Layer MoS_2 Transistors. *Nat. Nanotechnol.* **2011**, *6*, 147−150.
(5) Bernardi, M.; Palummo, M.; Grossman, J. C. Extraordinary Sunlight Absorption and One Nanometer Thick Photovoltaics Using Two-Dimensional Monolayer Materials. *Nano Lett.* **2013**, *13*, 3664−3670.
(6) Kaasbjerg, K.; Thygesen, K. S.; Jacobsen, K. W. Phonon-Limited Mobility in n-type Single-Layer MoS_2 from First Principles. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *85*, 115317.
(7) Gong, C.; Colombo, L.; Wallace, R. M.; Cho, K. The Unusual Mechanism of Partial Fermi Level Pinning at Metal-MoS_2 Interfaces. *Nano Lett.* **2014**, *14*, 1714−1720.
(8) Kang, J.; Liu, W.; Sarkar, D.; Jena, D.; Banerjee, K. Computational Study of Metal Contacts to Monolayer Transition-Metal Dichalcogenide Semiconductors. *Phys. Rev. X* **2014**, *4*, 031005.
(9) Du, Y.; Yang, L.; Zhang, J.; Liu, H.; Majumdar, K.; Kirsch, P. D.; Ye, P. D. MoS_2 Field-Effect Transistors With Graphene/Metal Heterocontacts. *IEEE Electron Device Lett.* **2014**, *35*, 599−601.
(10) Jena, D.; Banerjee, K.; Xing, G. H. 2D Crystal Semiconductors: Intimate Contacts. *Nat. Mater.* **2014**, *13*, 1076−1078.
(11) Kappera, R.; Voiry, D.; Yalcin, S. E.; Branch, B.; Gupta, G.; Mohite, A. D.; Chhowalla, M. Phase-Engineered Low-Resistance Contacts for Ultrathin MoS_2 Transistors. *Nat. Mater.* **2014**, *13*, 1128−1134.
(12) Lee, Y. T.; Choi, K.; Lee, H. S.; Min, S.-W.; Jeon, P. J.; Hwang, D. K.; Choi, H. J.; Im, S. Graphene Versus Ohmic Metal as Source-Drain Electrode for MoS2 Nanosheet Transistor Channel. *Small* **2014**, *10*, 2356−2361.
(13) Yu, L.; Lee, Y.-H.; Ling, X.; Santos, E. J. G.; Shin, Y. C.; Lin, Y.; Dubey, M.; Kaxiras, E.; Kong, J.; Wang, H.; et al. Graphene/MoS_2 Hybrid Technology for Large-Scale Two-Dimensional Electronics. *Nano Lett.* **2014**, *14*, 3055−3063.
(14) Yu, Y.-J.; Zhao, Y.; Ryu, S.; Brus, L. E.; Kim, K. S.; Kim, P. Tuning the Graphene Work Function by Electric Field Effect. *Nano Lett.* **2009**, *9*, 3430−3434.
(15) Wang, H.; Yu, L.; Lee, Y.-H.; Shi, Y.; Hsu, A.; Chin, M. L.; Li, L.-J.; Dubey, M.; Kong, J.; Palacios, T. Integrated Circuits Based on Bilayer MoS_2 Transistors. *Nano Lett.* **2012**, *12*, 4674−4680.
(16) Leong, W. S.; Luo, X.; Li, Y.; Khoo, K. H.; Quek, S. Y.; Thong, J. T. L. Low Resistance Metal Contacts to MoS_2 Devices with Nickel-Etched-Graphene Electrodes. *ACS Nano* **2015**, *9*, 869−877.

DOI: 10.1021/acs.jpcc.5b05580
J. Phys. Chem. C 2015, 119, 19928−19933

(17) Neaton, J. B.; Hybertsen, M. S.; Louie, S. G. Renormalization of Molecular Electronic Levels at Metal-Molecule Interfaces. *Phys. Rev. Lett.* **2006**, *97*, 216405.
(18) Hüser, F.; Olsen, T.; Thygesen, K. S. Quasiparticle GW Calculations for Solids, Molecules, and Two-Dimensional Materials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 235132.
(19) Ugeda, M. M.; Bradley, A. J.; Shi, S.-F.; da Jornada, F. H.; Zhang, Y.; Qiu, D. Y.; Ruan, W.; Mo, S.-K.; Hussain, Z.; Shen, Z.-X.; et al. Giant Bandgap Renormalization and Excitonic Effects in a Monolayer Transition Metal Dichalcogenide Semiconductor. *Nat. Mater.* **2014**, *13*, 1091−1095.
(20) Garcia-Lastra, J. M.; Rostgaard, C.; Rubio, A.; Thygesen, K. S. Polarization-Induced Renormalization of Molecular Levels at Metallic and Semiconducting Surfaces. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2009**, *80*, 245427.
(21) Withers, F.; del Pozo-Zamudio, O.; Mishchenko, A.; Rooney, A. P.; Gholinia, A.; Watanabe, K.; Taniguchi, T.; Haigh, S. J.; Geim, A. K.; Tartakovskii, A. I.; et al. Light-Emitting Diodes by Band-Structure Engineering in van der Waals Heterostructures. *Nat. Mater.* **2015**, *14*, 301−306.
(22) Zhang, C.; Johnson, A.; Hsu, C.-L.; Li, L.-J.; Shih, C.-K. Direct Imaging of Band Profile in Single Layer MoS_2 on Graphite: Quasiparticle Energy Gap, Metallic Edge States, and Edge Band Bending. *Nano Lett.* **2014**, *14*, 2443−2447.
(23) Enkovaara, J.; Rostgaard, C.; Mortensen, J. J.; Chen, J.; Dułak, M.; Ferrighi, L.; Gavnholt, J.; Glinsvad, C.; Haikola, V.; Hansen, H. A.; et al. Electronic Structure Calculations with GPAW: a Real-Space Implementation of the Projector Augmented-Wave Method. *J. Phys.: Condens. Matter* **2010**, *22*, 253202.
(24) Hüser, F.; Olsen, T.; Thygesen, K. S. How Dielectric Screening in Two-Dimensional Crystals Affects the Convergence of Excited-State Calculations: Monolayer MoS_2. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *88*, 245309.
(25) Ma, Y.; Dai, Y.; Guo, M.; Niu, C.; Huang, B. Graphene Adhesion on MoS_2 Monolayer: An ab initio Study. *Nanoscale* **2011**, *3*, 3883−3887.
(26) Klots, A. R.; Newaz, A. K. M.; Wang, B.; Prasai, D.; Krzyzanowska, H.; Lin, J.; Caudel, D.; Ghimire, N. J.; Yan, J.; Ivanov, B. L.; et al. Probing Excitonic States in Suspended Two-Dimensional Semiconductors by photocurrent spectroscopy. *Sci. Rep.* **2014**, *4*, 6608.
(27) Lin, Y.; Ling, X.; Yu, L.; Huang, S.; Hsu, A. L.; Lee, Y.-H.; Kong, J.; Dresselhaus, M. S.; Palacios, T. Dielectric Screening of Excitons and Trions in Single-Layer MoS_2. *Nano Lett.* **2014**, *14*, 5569−5576.
(28) Liu, H.; Liu, Y.; Zhu, D. Chemical Doping of Graphene. *J. Mater. Chem.* **2011**, *21*, 3335−3345.

■ NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on August 6, 2015. The abstract graphic has been updated and Reference 2 has been added. The revised version was re-posted on August 19, 2015.

DOI: 10.1021/acs.jpcc.5b05580
J. Phys. Chem. C 2015, 119, 19928−19933

Paper III

Filip A. Rasmussen and Kristian S. Thygesen. "Efficient *GW* calculations on two-dimensional materials using accurate long wavelength limit of the screened potential". In: *arXiv:1511.00129 [cond-mat]* (October 31, 2015). Submitted to Phys. Rev. B. arXiv: 1511.00129

Efficient GW calculations on two-dimensional materials using accurate long wavelength limit of the screened potential

Filip A. Rasmussen and Kristian S. Thygesen
Center for Atomic-scale Materials Design (CAMD),
Department of Physics and Center for Nanostructured Graphene (CNG),
Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
(Dated: November 3, 2015)

Calculating the quasiparticle (QP) band structure of two-dimensional (2D) materials within the GW self-energy approximation has proven to be a rather demanding computational task. The main reason is that the strong $\mathbf{q}$-dependence of the 2D dielectric function around $\mathbf{q} = \mathbf{0}$ calls for a much denser sampling of the Brillouin zone than is necessary for similar 3D solids. Here we use an analytical expression for the small $\mathbf{q}$-limit of the 2D response function to perform the BZ integral over the critical region around $\mathbf{q} = \mathbf{0}$. This drastically reduces the requirements on the $\mathbf{q}$-point mesh and implies a significant computational speed-up. For example, in the case of monolayer MoS_2, convergence of the G_0W_0 band gap to within $\sim 0.1\,\mathrm{eV}$ is achieved with 12×12 $\mathbf{q}$-points rather than the 36×36 mesh required with standard techniques. We obtain a converged G_0W_0-LDA band gap of MoS_2 of $2.44\,\mathrm{eV}$ which increases to $2.75\,\mathrm{eV}$ in GW_0. The method also explicitly accounts for dielectric anistropy and we therefore also test the method on the highly anisotropic material phosphorene, a 2D alloptrope of black phosphorous, and well-converged results for the QP band gap for this material is also reported.

PACS numbers: 71.15.Dx, 71.15.Qe, 73.22.-f

I. INTRODUCTION

The past few years have witnessed an explosion in research on atomically thin two-dimensional (2D) materials. Of particular interest are the 2D semiconductors including the family of transition metal dichalcogenides, which have been found to exhibit a number of unique opto-electronic properties.[1–7] Of fundamental importance for understanding and predicting the opto-electronic properties is the electronic band structure of the material. The GW method,[8,9] introduced by Hedin[10] in 1965 and first applied to real solids in an ab-initio framework by Hybertsen and Louie[11] and Godby, Sham, and Schlüter,[12] has become the "gold standard" for calculating quasi-particle (QP) band structures. Over the years its performance has been thoroughly established for bulk materials[13–15] and more recently also for molecules.[16–19] In comparison, critical assessments of the accuracy and numerical convergence of GW calculations for 2D materials are rather scarce.[20–23] These studies have shown that (i) it is essential to use a truncated Coulomb interaction to avoid long range screening between periodically repeated layers which reduces the QP band gap, and (ii) when a truncated Coulomb interaction is used, the convergence of the GW calculation with respect to the number of $\mathbf{k}$-points becomes much slower than is the case for similar bulk systems.

The slow $\mathbf{k}$-point convergence of the GW band structure is directly related to the nature of electronic screening in 2D which is qualitatively different from the well known 3D case.[24,25] Specifically, while the dielectric function, $\varepsilon(\mathbf{q})$, of bulk semiconductors is approximately constant for small wave vectors, the dielectric function of a 2D semiconductor varies sharply as $\mathbf{q} \to \mathbf{0}$.[20,21] As a consequence, the number of $\mathbf{q}$-points required to obtain a proper sampling of the screened interaction $W(\mathbf{q})$ over the Brillouin zone (BZ) is much higher for the 2D material than what would be anticipated from the 3D case. For example, the band gap of bulk MoS_2 is converged to within $\sim 0.1\,\mathrm{eV}$ with an in-plane $\mathbf{k}$-point grid of 12×12 while the same accuracy for monolayer MoS_2 requires a grid of 36×36 when standard BZ sampling schemes are applied.

Here we show that the slow $\mathbf{k}$-point convergence of the GW self-energy in 2D materials can be alleviated by performing the BZ integral of $W(\mathbf{q})$ analytically in the critical region around $\mathbf{q} = \mathbf{0}$ where $\varepsilon(\mathbf{q})$ varies most strongly. The analytical expression for $W(\mathbf{q})$ is obtained from a lowest order expansion in $\mathbf{q}$ of the head, $\chi^0_{\mathbf{00}}(\mathbf{q})$, and wings, $\chi^0_{\mathbf{0G}}(\mathbf{q})$, of the non-interacting density response function. This simple scheme reduces the required number of $\mathbf{q}$-points by an order of magnitude without loss of accuracy.

II. GW SELF-ENERGY IN 2D

We split the GW self-energy into the exchange and correlation part, respectively. The former does not present particular problems in 2D materials and is performed using a Wigner-Seitz truncated Coulomb interaction as described elsewhere.[26] In a planewave expansion the corre-

lation part of the self-energy takes the form[27]

$$\langle n\mathbf{k}|\Sigma^{\rm c}(\omega)|n'\mathbf{k}\rangle = \frac{i}{2\pi V}\int_{\rm BZ} d\mathbf{q}\sum_{\mathbf{G}\mathbf{G}'}\int_{-\infty}^{\infty} d\omega' \overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) \times \sum_m \frac{[\rho^{m,\mathbf{k}+\mathbf{q}}_{n,\mathbf{k}}(\mathbf{G})][\rho^{m,\mathbf{k}+\mathbf{q}}_{n',\mathbf{k}}(\mathbf{G}')]^*}{\omega+\omega'-\epsilon_{m\mathbf{k}+\mathbf{q}}-i\eta\,\mathrm{sgn}(\epsilon_{m\mathbf{k}+\mathbf{q}}-\mu)}, \tag{1}$$

where the pair densities are defined as $\rho^{m\mathbf{k}+\mathbf{q}}_{n\mathbf{k}}(\mathbf{G}) = \langle n\mathbf{k}|e^{i(\mathbf{G}+\mathbf{q})\cdot\mathbf{r}}|m\mathbf{k}+\mathbf{q}\rangle$, μ is the chemical potential, and $\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$ is the dynamical part of the screened potential given by

$$\overline{W}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = v_{\mathbf{G}}(\mathbf{q})\left[\varepsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)-\delta_{\mathbf{G}\mathbf{G}'}\right], \tag{2}$$

where $v_{\mathbf{G}}(\mathbf{q}) = 4\pi/|\mathbf{G}+\mathbf{q}|^2$ is the Coulomb interaction. In most implementations the BZ integral in Eq. (1) is evaluated numerically with a standard quadrature method using a regular $\mathbf{q}$-point grid matching the $\mathbf{k}$-point grid of the ground state DFT calculation. Since the screened potential, Eq. (2), diverges at $\mathbf{q}=\mathbf{0}$ this point is handled separately, so the integral may be written

$$\int_{\rm BZ} d\mathbf{q}\mathcal{S}(\mathbf{q},\omega) \to \frac{\Omega}{N_{\mathbf{q}}}\sum_{\mathbf{q}_n\neq\mathbf{0}}\mathcal{S}(\mathbf{q}_n,\omega)+\int_{\Omega_0} d\mathbf{q}\mathcal{S}(\mathbf{q},\omega), \tag{3}$$

where $\mathcal{S}(\mathbf{q},\omega)$ denotes the entire integrand, Ω is the volume of the BZ, $N_{\mathbf{q}}$ is the total number of $\mathbf{q}$-points in the grid and Ω_0 denotes a small region around $\mathbf{q}=\mathbf{0}$. We now focus on how to calculate the contribution to the integral around the special point $\mathbf{q}=\mathbf{0}$.

Within the random phase approximation (RPA) the dielectric matrix is given by

$$\varepsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = \delta_{\mathbf{G}\mathbf{G}'} - v_{\mathbf{G}}(\mathbf{q})\chi^0_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega), \tag{4}$$

valid for all systems that extend infinitely in all directions. For a solid with a finite band gap it can be shown that the head of the non-interacting response function $\chi^0_{\mathbf{0}\mathbf{0}}(\mathbf{q},\omega)\propto q^2$.[28] Since $v_{\mathbf{0}}(\mathbf{q}) = 4\pi/q^2$ it follows that in 3D the head of the dielectric function $\varepsilon_{\mathbf{0}\mathbf{0}}(\mathbf{q},\omega)$ converges to a finite value >1 when $q\to 0$. Moreover, this value is typically a reasonable approximation to $\varepsilon_{\mathbf{0}\mathbf{0}}(\mathbf{q},\omega)$ in a relatively large region around $\mathbf{q}=\mathbf{0}$. This means that in the BZ integration in Eq. (1) around the singular point $\mathbf{G}=\mathbf{G}'=\mathbf{q}=\mathbf{0}$ all factors, except $1/q^2$, can be assumed to be constant and the integral can be performed analytically over a sphere centred at $\mathbf{q}=\mathbf{0}$,[11] computed numerically[29] or using an expansion of $W(\mathbf{q})$ in spherical harmonics.[30] (For simplicity we have neglected local field effects in the above analysis, but accounting for these do not change the conclusions).

For *GW* calculations on 2D materials performed with periodic boundary conditions, the direct use of Eq. (1) leads to significant over-screening due to the interaction between the repeated images.[21] One way of dealing with this is by subtracting the artificial contribution calculated from a classical electrostatic model.[31] A more direct way of avoiding this unwanted interaction is to truncate the Coulomb interaction in the direction perpendicular to the layers. Thus in Eqs. (4) and (2), $v_{\mathbf{G}}(\mathbf{q})$ should be replaced by[20,32]

$$v^{\rm 2D}_{\mathbf{G}}(\mathbf{q}) = \frac{4\pi}{|\mathbf{q}+\mathbf{G}|^2}\left[1-e^{-|\mathbf{q}_\parallel+\mathbf{G}_\parallel|L/2}\cos(|G_z|L/2)\right], \tag{5}$$

where L is the length of the unit cell in the non-periodic direction. In the long wavelength limit $\mathbf{G}=\mathbf{0}$, $\mathbf{q}\to\mathbf{0}$, the truncated interaction becomes $v^{\rm 2D}_{\mathbf{0}}(\mathbf{q})\approx\frac{2\pi L}{q}$ and the leading order of the head of the dielectric function thus becomes

$$\varepsilon_{\mathbf{0}\mathbf{0}}(\mathbf{q}) \approx 1+\alpha q, \tag{6}$$

where we have assumed an isotropic material for simplicity (in general α will depend on the direction of $\mathbf{q}$).

Some examples of *macroscopic* dielectric functions for a representative set of 2D semiconductors are shown in Figure 1 (see Ref. 21 for a precise definition of this quantity). The linear form (6) is clearly observed in the small $\mathbf{q}$ regime. Importantly, if we use the same strategy for evaluating the BZ integral in Eq. (1) as in 3D, i.e. approximating the integral by a mean value average on the discrete $\mathbf{q}$-point grid and replacing $\mathbf{q}=\mathbf{0}$ term of the bare Coulomb potential by its analytical integral over a small sphere, we obtain zero contribution for the $\mathbf{q}=0$ term. This is because $\varepsilon^{-1}_{00}(0)-1=0$, see Eq. (2). On the other hand, by comparing Eqs. (6) and (2), and using the asymptotic form $v^{\rm 2D}_{\mathbf{0}}(\mathbf{q})\approx\frac{2\pi L}{q}$ for the truncated Coulomb interaction, it follows that $\overline{W}_{\mathbf{0}\mathbf{0}}(\mathbf{q})\approx 2\pi L\alpha/(1+\alpha q)$ for small $\mathbf{q}$ (assuming an isotropic material).

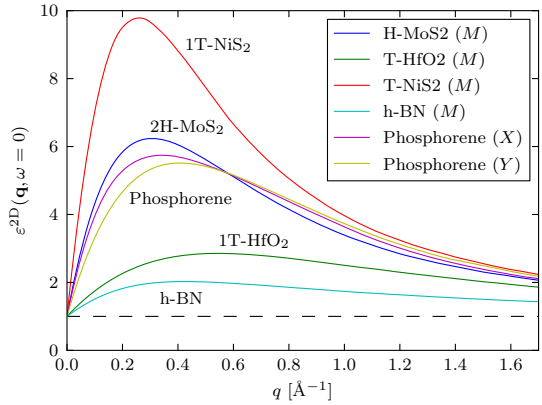


FIG. 1: (Color online) Static macroscopic dielectric function of a representative set of 2D semiconductors as a function of $\mathbf{q}$ along the $\Gamma\to M$ direction for the hexagonal structures and along the path from Γ to X or Y in the case of phosphorene.

In Appendix A we show, following an analysis similar to that of Ref. 30, that for a general non-isotropic 2D material, the small $\mathbf{q}$ limit of the head of the screened potential takes the form

$$\overline{W}_{\mathbf{00}}(\mathbf{q}) = -2\pi L \frac{\hat{\mathbf{q}} \cdot \mathsf{A}\hat{\mathbf{q}}}{1 + |\mathbf{q}|\hat{\mathbf{q}} \cdot \mathsf{A}\hat{\mathbf{q}}}, \tag{7}$$

where $\hat{\mathbf{q}} = \mathbf{q}/|\mathbf{q}|$ and A is a second rank tensor which also depends on the frequency. We see that we have exactly $\overline{W}_{\mathbf{00}}(\mathbf{q} = \mathbf{0}) = -2\pi L \hat{\mathbf{q}} \cdot \mathsf{A}\hat{\mathbf{q}}$. The expression $\hat{\mathbf{q}} \cdot \mathsf{A}\hat{\mathbf{q}}$ is closely related to the slope of the dielectric function. In addition to Eq. (7) there are similar expressions for the wings and body of the screened interaction. When calculating the screened interaction at $\mathbf{q} = \mathbf{0}$ we use an average calculated by integrating these expressions in a small surrounding region numerically on a very fine grid (typically using an in-plane $\mathbf{q}$-point density of 6.25×10^{-6} Å^2).

A. Results

To investigate how this method performs we have carried out test calculations on the three monolayers h-BN, MoS_2 and phosphorene, which have quite different dielectric functions as seen on Figure 1. h-BN is a large gap dielectric with low screening ability leading to a small slope of the dielectric function at $\mathbf{q} = \mathbf{0}$, while MoS_2 has a larger dielectric function and quite steep slope at $\mathbf{q} = \mathbf{0}$. Phosphorene has a dielectric function similar to MoS_2 in size and steepness but is anisotropic with slopes varying by $\sim 40\%$ between the two high symmetry directions, $\Gamma \to X$ and $\Gamma \to Y$. For the present calculations we use for h-BN a lattice constant of 2.504 Å and DFT calculations using PBE exchange-correlation (XC) functional. For MoS_2 we use an in-plane lattice constant of 3.184 Å and a sulphur-sulphur distance of 3.127 Å and the LDA XC functional. For phosphorene we use an in-plane unit cell of 4.611 Å by 3.308 Å, in-plane P-P-P angle of 95.9°, layer thickness of 2.114 Å and the PBE XC functional. For all calculations the amount of surrounding vacuum has been such that the interlayer distance were in the range 10 to 15Å. The eigenvalues and wavefunctions obtained from the DFT calculations were used as input in the the *GW* calculation. The DFT calculations were done using a plane wave basis with a cut-off of 600 eV, while the dielectric function and the correlation self-energy were calculated using a much smaller basis size of either 50 eV (h-BN and MoS_2) or 75 eV (phosphorene). While this cut-off might not be enough to ensure properly converged quasi-particle energies, it is adequate to describe the trends related to the $\mathbf{k}$-point sampling relevant for this study.

In Figure 2 we compare the analytical small $\mathbf{q}$ expression, Eq. (7), for the the head of the screened potential $\overline{W}_{\mathbf{00}}(\mathbf{q})$ with the numerical values obtained using a fine and coarse $\mathbf{k}$-point sampling. In all the cases the $\mathbf{q} = \mathbf{0}$ value has been set to the analytical value. We notice that the screened potential falls off quickly and thus for a coarse $\mathbf{q}$-point sampling the $\mathbf{q} = \mathbf{0}$ contribution to the integral is by far the largest and should therefore not be neglected. Similarly, just using the exact value in

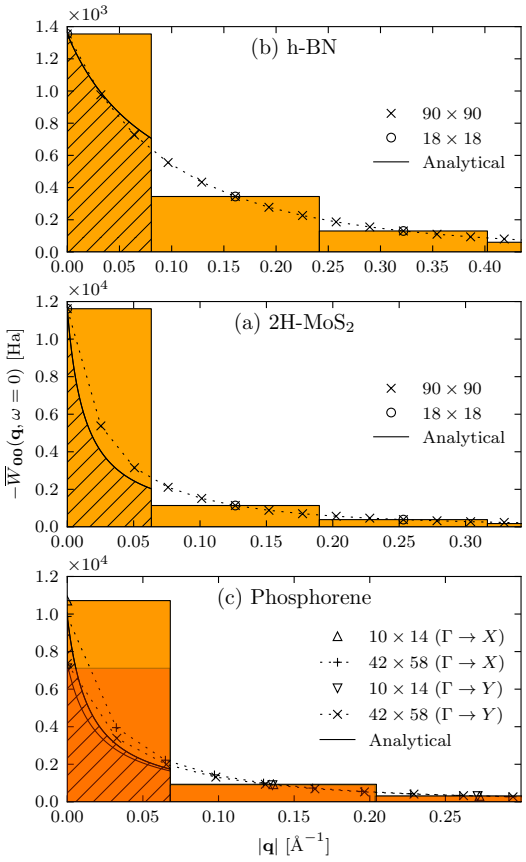


FIG. 2: (Color online) The head of the static component of the screened potential (subtracted the bare interaction) of monolayer a) h-BN b) MoS_2 and c) Phosphorene as a function of $\mathbf{q}$ along the $\Gamma \to M$ direction or $\Gamma \to X$ and $\Gamma \to Y$ in the case of phosphorene. The crosses are the numerical values obtained on a fine $\mathbf{q}$-point grid while the circles or triangles represent the values obtained on a coarse $\mathbf{q}$-point grid. The dotted line is a linear interpolation of the numerically exact values. The bars represent a simple numerical approximation to the BZ integral of $\overline{W}_{00}(\mathbf{q})$ performed on the coarse $\mathbf{q}$-point grid. The value of the screened potential for $\mathbf{q} = \mathbf{0}$ is set to the analytical result Eq. (7). The full curve, shown only inside the $\mathbf{q} = \mathbf{0}$ bar, represents the analytical small $\mathbf{q}$ approximation, Eq. (7), and the hatched area shows its contribution to the integral.

$\mathbf{q} = \mathbf{0}$ could also pose a problem as the contribution will be grossly overestimated due to the convex nature of potential. Using the analytical limit within the region around $\mathbf{q} = \mathbf{0}$ however appears quite reasonable, althpugh it clearly works better for h-BN than MoS_2 and phosphorene. We notice that the anisotropy of phosphorene makes $\overline{W}_{\mathbf{00}}(\mathbf{q})$ non-analytical at $\mathbf{q} = \mathbf{0}$ (different limit values depending on the direction of $\mathbf{q}$). For larger $\mathbf{q}$ the dielectric anisotropy becomes negligible; however, because of the relatively large weight of the $\mathbf{q} = \mathbf{0}$ contribution to the BZ integral, the anisotropy should be taken into account for accurate GW calculations.

We note that a similar approach to the treatment of the $\mathbf{q} = \mathbf{0}$ term of the screened potential was suggested in Ref. 20. That particular method was based on fitting to an empirical expression for $\varepsilon(q)$ calculated from the value at a small but finite $\mathbf{q}$. The method outlined here is different in that an analytical expression for $\overline{W}(\mathbf{q})$ obtained from a lowest order expansion of the head ($\chi^0_{\mathbf{00}}(q)$) and wings ($\chi^0_{\mathbf{0G}}(q)$) of the non-interacting density response function[33] and thus can be obtained without fitting or using empirical parameters. This also ensures that the effect of in-plane dielectric anisotropy is explicitly included, which is seen to be quite significant for a material like phosporene.

In Fig. 3 we show the correlation self-energy contribution to the G_0W_0 quasiparticle band gap of monolayer h-BN, MoS_2 and phosphorene as a function of $1/N_{\mathbf{k}}$ where $N_{\mathbf{k}}$ is the total number of $\mathbf{k}$-points in the BZ sampling (the $\mathbf{q}$ point grid for the GW integration is the same as the $\mathbf{k}$-point grid used in DFT). We compare the results obtained using two methods: i) neglecting the $\mathbf{q} = \mathbf{0}$ contribution to head and wings of the screened potential and ii) evaluating Eq. (7) numerically on a fine grid in the $\mathbf{q} = \mathbf{0}$ region. It is clear that method i) in all cases underestimates the correlation self-energy due to the underestimation of the screening; in order to get the band gap converged to within ~ 0.1 eV one would have to use a $\mathbf{k}$-point sampling of minimum 30×30 for h-BN, 36×36 for MoS_2 and 22×30 for phosphorene. We also note that for large $\mathbf{k}$-point grids the band gaps using this method converge approximately as $1/N_{\mathbf{k}}$ as the missing contribution is almost proportional to the area of the $\mathbf{q} = \mathbf{0}$ region. Clearly, the latter approach varies significantly less with the $\mathbf{k}$-point grid and in fact the gap is converged to within 0.2 eV already for a $\mathbf{k}$-point grid in the order of 6×6 and to within ~ 0.1 eV with a 12×12 grid (in the worst case). We have performed test calculations for other 2D semiconductors and obtained similar conclusions although the number of $\mathbf{k}$-points required to reach convergence within 0.1 eV following the conventional approach ($\mathbf{q} = \mathbf{0}$ term neglected) is somewhat system dependent; materials with efficient screening, e.g. MoS_2 and NiS_2, require larger $\mathbf{k}$-point grids than materials with poor screening, e.g. h-BN and HfO_2 (see Fig. 1).

As mentioned, the self-energies shown in Fig. 3 were obtained with a rather low (50 eV to 75 eV) plane wave cut-off and may not be fully converged. In Table I we report quasiparticle band gaps calculated by using self-energies obtained by a $1/N_{\text{pw}}$ extrapolation to the complete basis set limit using cut-off energies of up to 200 eV, which is needed to ensure proper convergence.[34–36] For these calculations we used 12×12 $\mathbf{k}$-points for h-BN, 18×18 $\mathbf{k}$-points for MoS_2 and 10×14 for phosphorene with the analytical integration of $W(\mathbf{q})$ around $\mathbf{q} = \mathbf{0}$.

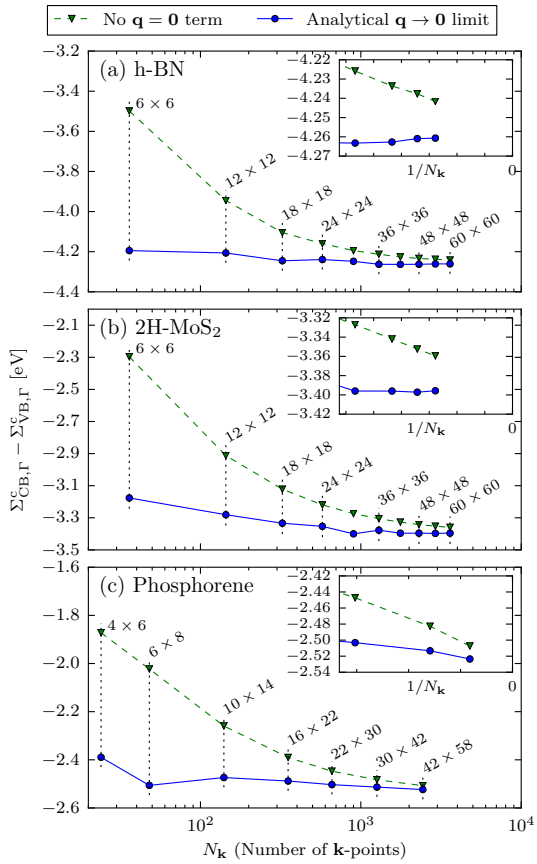


FIG. 3: (Color online) The correlation self-energy contribution to the G_0W_0 quasi-particle band gap of monolayer (a) 2H-MoS_2 (b) h-BN and (c) phosphorene, calculated using two different treatments of the $\mathbf{q} = \mathbf{0}$ term in Eq. (1). The dashed (green) line shows the contribution obtained when the head and wing elements of the $\mathbf{q} = \mathbf{0}$ term are neglected corresponding to the standard treatment used for 3D systems. The solid (blue) line shows the contribution obtained when using the analytical results, Eq. (7), to perform the integral over the $\mathbf{q} = \mathbf{0}$ element. The insets shows the results for the largest $\mathbf{k}$-point grids on a reversed linear scale in $1/N_{\mathbf{k}}$.

	DFT	G_0W_0	GW_0
h-BN (PBE)	4.54 eV	6.41 eV	6.91 eV
2H-MoS_2 (LDA)	1.66 eV	2.44 eV	2.75 eV
2H-MoS_2 (PBE)	1.66 eV	2.33 eV	-
Phosphorene (PBE)	0.89 eV	1.99 eV	2.23 eV

TABLE I: Direct band gaps calculated by DFT, G_0W_0 and GW_0. The GW calculations were performed using analytic integration of $\overline{W}(\mathbf{q})$ around $\mathbf{q} = \mathbf{0}$ and the correlation self-energy has been extrapolated to an infinite basis set. Spin-orbit interactions were not included. The following $\mathbf{k}$-point grid were used; h-BN: 12×12, 2H-MoS2: 18×18 and phosphorene: 10×14.

According to Fig. 3 this is sufficient to ensure convergence to within 0.1 eV. We note that spin-orbit interactions are not included in the reported values. Inclusion of spin-orbit interactions splits the valence band of MoS_2 at the K point by 0.15 eV thereby lowering the QP gap by around 0.07 eV.[23,37] Our converged G_0W_0@LDA band gap of 2.4 eV and G_0W_0@PBE gap of 2.3 eV are both smaller than previously reported values of 2.5 eV to 2.8 eV.[23,38] These have however been obtained without either a 2D truncation of the Coulomb potential or a proper treatment of the small $\mathbf{q}$ screening in 2D and also suffered from insufficient size of the $\mathbf{k}$-point grids. An exception is Ref. 22 that reports a G_0W_0@LDA band gap for MoS_2 of 2.70 eV using a truncated Coulomb interaction and a calculation of the screened potential at $\mathbf{q} = \mathbf{0}$ based on the method in Ref. 20. As of now we are not sure of the reason between the disagreement between that result and the ones reported here, but we note that in Ref. 22 they used a lattice constant of MoS_2 of 3.15 Å. This is 0.03 Å smaller than the one used for the calculations in this study and it is expected that the band gap increases with compressive strain,[23,39] which could explain some of the difference. Our result of 2.4 eV is also smaller than the experimental value of 2.5 eV inferred from photo current spectroscopy.[40] However, performing partially self-consistent GW_0 the band gap increases to 2.75 eV (2.68 eV including spin-orbit), which is again larger than the experimental value. The method of Ref. 20 has also been used to calculate the G_0W_0 band gap of phosphorene where a value of 2.0 eV is reported,[41] which is very close the value we obtain, despite the neglect of the dielectric anisotropy in the method used.

III. CONCLUSION

In conclusion, we have discussed the connection between the form of the $\mathbf{q}$-dependent dielectric function of a 2D semiconductor and the slow $\mathbf{k}$-point convergence of the GW band structure. We have derived an analytical expression for the $\mathbf{q} \to \mathbf{0}$ limit of the dynamical part of the screened potential of a semiconductor when a 2D truncation of the Coulomb potential is used. This method also directly accounts for any dielectric anisotropy and does not rely on any additional parameters or fitting. Using this expression we have shown that convergence of the GW self-energy with respect to the size of the $\mathbf{k}$-point grid is drastically improved. For the specific case of monolayer MoS_2, we found that the use of the analytical form alone reduces the $\mathbf{k}$-point grid required to achieve convergence of the GW self-energy contribution to the band gap to within ~ 0.1 eV from around 36×36 to 12×12 – a reduction in the number of $\mathbf{k}$-points by a factor of ~ 9. This method may therefore greatly aid in speeding up future calculations and open up the possibility of doing GW calculations for a broader range of 2D materials.

IV. ACKNOWLEDGEMENTS

We acknowledge support from the Danish Council for Independent Research's Sapere Aude Program, Grant No. 11-1051390. The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation, Project DNRF58.

Appendix A: Calculation of the $\mathbf{q} \to \mathbf{0}$ limit of the screened potential

In the following we derive the analytical form of the screened potential, Eq. (2), in the limit $\mathbf{q} \to \mathbf{0}$. We largely follow the approach of Ref. 30 where the same limit for bulk systems was considered. As explained in the main text we use a truncated Coulomb interaction of the form

$$v(\mathbf{r}_\parallel, z) = \frac{\theta(R - |z|)}{\sqrt{|\mathbf{r}_\parallel|^2 + z^2}}. \tag{A1}$$

Using this potential we effectively turn off interaction between electrons on different 2D layers of the supercell calculation. We typically choose R to be half the height of the unitcell, $R = L/2$, so that an electron in the center of the layer will not interact with electrons located in the neighboring unitcell. In the following we will work with its Fourier transform given by

$$v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{q}) = \frac{4\pi}{|\mathbf{q} + \mathbf{G}|^2} \left[1 - e^{-|\mathbf{q}+\mathbf{G}_\parallel| L/2} \cos(|G_z| L/2)\right], \tag{A2}$$

where $\mathbf{q}$ is given in-plane only. We note that in the limit $L \to \infty$ it takes the usual 3D form, $v_{\mathbf{G}}(\mathbf{q}) = \frac{4\pi}{|\mathbf{q}+\mathbf{G}|^2}$. In the long wavelength limit it has the asymptotic behavior

$$v_{\mathbf{0}}^{\mathrm{2D}}(\mathbf{q}_\parallel \to \mathbf{0}) = \frac{2\pi L}{|\mathbf{q}|}, \tag{A3}$$

diverging slower than the full Coulomb potential with profound consequences for the properties of 2D materials.

In the long wavelength limit $\mathbf{q} \to \mathbf{0}$ the non-interacting density response function or irreducible polarizability has the following behavior[33]

$$\chi^0_{\mathbf{00}'}(\mathbf{q} \to \mathbf{0}) = \mathbf{q} \cdot \mathsf{P}\mathbf{q} = |\mathbf{q}|^2 \hat{\mathbf{q}} \cdot \mathsf{P}\hat{\mathbf{q}} \tag{A4}$$

$$\chi^0_{\mathbf{G0}}(\mathbf{q} \to \mathbf{0}) = \mathbf{q} \cdot \mathbf{p}_{\mathbf{G}} = |\mathbf{q}| \hat{\mathbf{q}} \cdot \mathbf{p}_{\mathbf{G}}, , \tag{A5}$$

where P is a second rank tensor, $\mathbf{p}_{\mathbf{G}}$ is a proper vector and $\hat{\mathbf{q}} = \mathbf{q}/|\mathbf{q}|$. The density response function, and therefore also P and $\mathbf{p}_{\mathbf{G}}$, has a frequency dependence which here and through the rest of this section has been left out to simplify the notation. Within the random phase approximation the dielectric function is given by (schematically)

$$\varepsilon = 1 - v\chi^0. \tag{A6}$$

Due to technical reasons[13,42] it is easier to work with a similar symmetrized version given in Fourier space by

$$\tilde{\varepsilon}_{\mathbf{GG}'}(\mathbf{q}) = \delta_{\mathbf{GG}'} - \sqrt{v_{\mathbf{G}}(\mathbf{q})}\chi^0_{\mathbf{GG}'}(\mathbf{q})\sqrt{v_{\mathbf{G}'}(\mathbf{q})}. \tag{A7}$$

Inserting the Coulomb potential, Eq. (A2), and the expressions for the non-interacting response function Eqs. (A4)-(A5), the head and wings of the symmetrized dielectric function are

$$\begin{aligned}\tilde{\varepsilon}_{\mathbf{00}}(\mathbf{q}_\parallel \to \mathbf{0}) &= 1 - \frac{2\pi L}{|\mathbf{q}|}|\mathbf{q}|^2 \hat{\mathbf{q}} \cdot \mathsf{P}\hat{\mathbf{q}} \\ &= 1 + |\mathbf{q}| \hat{\mathbf{q}} \cdot \mathsf{U}\hat{\mathbf{q}} \end{aligned} \tag{A8}$$

$$\begin{aligned}\tilde{\varepsilon}_{\mathbf{G0}}(\mathbf{q}_\parallel \to \mathbf{0}) &= -\sqrt{v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{0})}|\mathbf{q}| \hat{\mathbf{q}} \cdot \mathbf{p}_{\mathbf{G}} \sqrt{\frac{2\pi L}{|\mathbf{q}|}} \\ &= \sqrt{|\mathbf{q}|} \hat{\mathbf{q}} \cdot \mathbf{u}_{\mathbf{G}} \end{aligned} \tag{A9}$$

with $\mathsf{U} = -2\pi L \mathsf{P}$ and $\mathbf{u}_{\mathbf{G}} = -\sqrt{2\pi L v^{\mathrm{2D}}_{\mathbf{G}}(\mathbf{0})}\mathbf{p}_{\mathbf{G}}$.

To determine the inverse dielectric function we write the dielectric function as a block matrix in the $\mathbf{G}, \mathbf{G}'$ components with head, wings and body of the form

$$\tilde{\boldsymbol{\varepsilon}} = \begin{pmatrix} H & \mathbf{w}^\dagger \\ \mathbf{w} & \mathbf{B} \end{pmatrix} \tag{A10}$$

The inverse is then given by

$$\tilde{\boldsymbol{\varepsilon}}^{-1} = \begin{pmatrix} (H - \mathbf{w}^\dagger \mathbf{B}^{-1}\mathbf{w})^{-1} & -(H - \mathbf{w}^\dagger \mathbf{B}^{-1}\mathbf{w})^{-1}\mathbf{w}^\dagger \mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{w}(H - \mathbf{w}^\dagger \mathbf{B}^{-1}\mathbf{w})^{-1} & \mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{w}(H - \mathbf{w}^\dagger \mathbf{B}^{-1}\mathbf{w})^{-1}\mathbf{w}^\dagger \mathbf{B}^{-1} \end{pmatrix} \tag{A11}$$

From this we see that

$$\tilde{\varepsilon}^{-1}_{\mathbf{00}} = \left[\tilde{\varepsilon}_{\mathbf{00}} - \sum_{\mathbf{G},\mathbf{G}' \neq \mathbf{0}} \tilde{\varepsilon}^*_{\mathbf{G0}} B^{-1}_{\mathbf{GG}'} \tilde{\varepsilon}_{\mathbf{G0}} \right]^{-1} \tag{A12}$$

$$\tilde{\varepsilon}^{-1}_{\mathbf{G0}} = \tilde{\varepsilon}^{-1}_{\mathbf{00}} \sum_{\mathbf{G}' \neq \mathbf{0}} B^{-1}_{\mathbf{GG}'} \tilde{\varepsilon}_{\mathbf{G}'\mathbf{0}} \tag{A13}$$

$$\tilde{\varepsilon}^{-1}_{\mathbf{GG}'} = B^{-1}_{\mathbf{GG}'} + \tilde{\varepsilon}^{-1}_{\mathbf{00}} \left(\sum_{\mathbf{G}'' \neq \mathbf{0}} B^{-1}_{\mathbf{GG}''} \tilde{\varepsilon}_{\mathbf{G}''\mathbf{0}} \right) \left(\sum_{\mathbf{G}'' \neq \mathbf{0}} \tilde{\varepsilon}^*_{\mathbf{G}''\mathbf{0}} B^{-1}_{\mathbf{G}''\mathbf{G}'} \right) \tag{A14}$$

Introducing the vector $\mathbf{a}_{\mathbf{G}}$ and the tensor A given by

$$\mathbf{a}_{\mathbf{G}} = \sum_{\mathbf{G}'} B^{-1}_{\mathbf{GG}'} \mathbf{u}_{\mathbf{G}'} \tag{A15}$$

$$\mathsf{A} = \mathsf{U} - \sum_{\mathbf{G} \neq \mathbf{0}} \mathbf{u}^*_{\mathbf{G}} \otimes \mathbf{a}_{\mathbf{G}}, \tag{A16}$$

where $\otimes$ denotes the tensor product, the long wavelength limit of the inverse dielectric function is seen to be given by

$$\tilde{\varepsilon}^{-1}_{\mathbf{00}}(\mathbf{q} \to \mathbf{0}) = \frac{1}{1 + |\mathbf{q}| \hat{\mathbf{q}} \cdot \mathsf{A}\hat{\mathbf{q}}} \tag{A17}$$

$$\tilde{\varepsilon}^{-1}_{\mathbf{G0}}(\mathbf{q} \to \mathbf{0}) = -\frac{\sqrt{|\mathbf{q}|} \hat{\mathbf{q}} \cdot \mathbf{a}_{\mathbf{G}}}{1 + |\mathbf{q}| \hat{\mathbf{q}} \cdot \mathsf{A}\hat{\mathbf{q}}} \tag{A18}$$

$$\tilde{\varepsilon}^{-1}_{\mathbf{GG}'}(\mathbf{q} \to \mathbf{0}) = B^{-1}_{\mathbf{GG}'} + \frac{|\mathbf{q}| (\hat{\mathbf{q}} \cdot \mathbf{a}_{\mathbf{G}})(\hat{\mathbf{q}} \cdot \mathbf{a}^*_{\mathbf{G}'})}{1 + |\mathbf{q}| \hat{\mathbf{q}} \cdot \mathsf{A}\hat{\mathbf{q}}} \tag{A19}$$

Inserting these expression in the equation for the screened potential, Eq. (2), we see that the head and

wings are given by

$$\begin{aligned}\overline{W}_{\mathbf{00}}(\mathbf{q}\to\mathbf{0}) &= v_{\mathbf{0}}^{\mathrm{2D}}(\mathbf{q})\left[\tilde{\varepsilon}^{-1}_{\mathbf{00}}(\mathbf{q})-1\right]\\ &= 2\pi L\frac{-\hat{\mathbf{q}}\cdot\mathsf{A}\hat{\mathbf{q}}}{1+|\mathbf{q}|\hat{\mathbf{q}}\cdot\mathsf{A}\hat{\mathbf{q}}} \qquad \text{(A20)}\\ \overline{W}_{\mathbf{G0}}(\mathbf{q}\to\mathbf{0}) &= \sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\tilde{\varepsilon}^{-1}_{\mathbf{G0}}(\mathbf{q})\sqrt{v_{\mathbf{0}}^{\mathrm{2D}}(\mathbf{q})}\\ &= -\sqrt{2\pi L v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})}\frac{\hat{\mathbf{q}}\cdot\mathbf{a}_{\mathbf{G}}}{1+|\mathbf{q}|\hat{\mathbf{q}}\cdot\mathsf{A}\hat{\mathbf{q}}}. \qquad \text{(A21)}\end{aligned}$$

and the body also gets a correction and becomes

$$\begin{aligned}\overline{W}_{\mathbf{GG'}}(\mathbf{q}\to\mathbf{0}) &= \sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})v_{\mathbf{G'}}^{\mathrm{2D}}(\mathbf{0})}\left[\varepsilon^{-1}_{\mathbf{GG'}}(\mathbf{q})-\delta_{\mathbf{GG'}}\right]\\ &= \sqrt{v_{\mathbf{G}}^{\mathrm{2D}}(\mathbf{0})v_{\mathbf{G'}}^{\mathrm{2D}}(\mathbf{0})}\Bigg[B^{-1}_{\mathbf{GG'}}-\delta_{\mathbf{GG'}}\\ &\qquad +\frac{|\mathbf{q}|(\hat{\mathbf{q}}\cdot\mathbf{a}_{\mathbf{G}})(\hat{\mathbf{q}}\cdot\mathbf{a}^{*}_{\mathbf{G'}})}{1+|\mathbf{q}|\hat{\mathbf{q}}\cdot\mathsf{A}\hat{\mathbf{q}}}\Bigg]. \qquad \text{(A22)}\end{aligned}$$

[1] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Phys. Rev. Lett. **105**, 136805 (2010).
[2] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, Nano Lett. **10**, 1271 (2010).
[3] H. Zeng, G.-B. Liu, J. Dai, Y. Yan, B. Zhu, R. He, L. Xie, S. Xu, X. Chen, W. Yao, and X. Cui, Sci Rep **3** (2013), 10.1038/srep01608.
[4] C. Zhang, A. Johnson, C.-L. Hsu, L.-J. Li, and C.-K. Shih, Nano Lett. **14**, 2443 (2014).
[5] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, Nature Nanotechnology **7**, 699 (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, A. N. Grigorenko, A. K. Geim, C. Casiraghi, A. H. C. Neto, and K. S. Novoselov, Science **340**, 1311 (2013).
[7] M. Bernardi, M. Palummo, and J. C. Grossman, Nano Lett. **13**, 3664 (2013).
[8] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, in *Solid State Physics*, Vol. Volume 54, edited by Henry Ehrenreich and Frans Spaepen (Academic Press, 1999) pp. 1–218.
[9] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[10] L. Hedin, Phys. Rev. **139**, A796 (1965).
[11] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985).
[12] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).
[13] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006).
[14] T. Kotani and M. van Schilfgaarde, Solid State Communications **121**, 461 (2002).
[15] A. Marini, C. Hogan, M. Grüning, and D. Varsano, Computer Physics Communications **180**, 1392 (2009).
[16] C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **81**, 085103 (2010).
[17] F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, Phys. Rev. B **88**, 075105 (2013).
[18] F. Bruneval, Phys. Rev. Lett. **103**, 176403 (2009).
[19] X. Blase, C. Attaccalite, and V. Olevano, Phys. Rev. B **83**, 115103 (2011).
[20] S. Ismail-Beigi, Phys. Rev. B **73**, 233103 (2006).
[21] F. Huser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **88**, 245309 (2013).
[22] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **115**, 119901 (2015).
[23] F. A. Rasmussen and K. S. Thygesen, J. Phys. Chem. C **119**, 13169 (2015).
[24] L. Keldysh, Jetp Lett. **29**, 658 (1979), wOS:A1979JN00600015.
[25] P. Cudazzo, I. V. Tokatly, and A. Rubio, Phys. Rev. B **84**, 085406 (2011).
[26] R. Sundararaman and T. A. Arias, Phys. Rev. B **87**, 165122 (2013).
[27] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[28] R. M. Pick, M. H. Cohen, and R. M. Martin, Phys. Rev. B **1**, 910 (1970).
[29] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, Computer Physics Communications **183**, 1269 (2012).
[30] C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, R. W. Godby, and M. Scheffler, Computer Physics Communications **176**, 1 (2007).
[31] C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, and M. Scheffler, Phys. Rev. B **77**, 235428 (2008).
[32] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, Phys. Rev. B **73**, 205119 (2006).
[33] J. Yan, J. J. Mortensen, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **83**, 245122 (2011).
[34] M. L. Tiago, S. Ismail-Beigi, and S. G. Louie, Phys. Rev. B **69**, 125212 (2004).
[35] B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **105**, 146401 (2010).
[36] J. Klimeš, M. Kaltak, and G. Kresse, Phys. Rev. B **90**, 075125 (2014).
[37] Z. Y. Zhu, Y. C. Cheng, and U. Schwingenschlögl, Phys. Rev. B **84**, 153402 (2011).
[38] A. Ramasubramaniam, Phys. Rev. B **86**, 115409 (2012).
[39] H. J. Conley, B. Wang, J. I. Ziegler, R. F. Haglund, S. T. Pantelides, and K. I. Bolotin, Nano Lett. **13**, 3626 (2013).
[40] A. R. Klots, A. K. M. Newaz, B. Wang, D. Prasai, H. Krzyzanowska, J. Lin, D. Caudel, N. J. Ghimire, J. Yan, B. L. Ivanov, K. A. Velizhanin, A. Burger, D. G. Mandrus, N. H. Tolk, S. T. Pantelides, and K. I. Bolotin, Sci. Rep. **4** (2014), 10.1038/srep06608.
[41] V. Tran, R. Soklaski, Y. Liang, and L. Yang, Phys. Rev. B **89**, 235319 (2014).
[42] A. Baldereschi and E. Tosatti, Phys. Rev. B **17**, 4710 (1978).

Paper IV

Thomas Olsen et al. “Simple Screened Hydrogen Model of Excitons in Two-Dimensional Materials”. Submitted to Phys. Rev. B. December 2015

Simple Screened Hydrogen Model of Excitons in Two-Dimensional Materials

Thomas Olsen,[1,*] Simone Latini,[1] Filip Rasmussen,[1] and Kristian S. Thygesen[1]

[1] *Center for Atomic-Scale Materials Design and Center for Nanostructured Graphene (CNG), Department of Physics, Technical University of Denmark*

We present a generalized hydrogen model for the binding energies (E_B) of excitons in two-dimensional (2D) materials that sheds light on the fundamental differences between excitons in two and three dimensions. In contrast to the well-known hydrogen model of three-dimensional (3D) excitons, the description of 2D excitons is complicated by the fact that the screening cannot be assumed to be local. We show that one can consistently define an effective 2D dielectric constant by averaging the screening over the extend of the exciton. For an ideal 2D semiconductor this leads to a simple expression for E_B that only depends on the excitonic mass and the 2D polarizability α. The model is shown to produce accurate results for 51 transition metal dichalcogenides. Remarkably, over a wide range of polarizabilities the expression becomes independent of the mass and we obtain $E_B^{2D} \approx 3/(4\pi\alpha)$, which explains the recently observed linear scaling of exciton binding energies with band gap. It is also shown that the model accurately reproduces the non-hydrogenic Rydberg series in WS_2 and can account for screening from the environment.

A striking property of two-dimensional semiconductors is the ability to form strongly bound excitons. This was initially predicted theoretically for hBN [1], graphane [2] and various transition metal dichalcogenides [3–5] and has subsequently been confirmed experimentally [6–8]. The quantum confinement of excitons in 2D comprises a tempting and intuitively appealing explanation for the large binding energies in these materials [9]. However, it is now well understood that the confinement of the local electronic environment in 2D plays a crucial role in the formation of strongly bound excitons [3, 10]. The 2D electronic system is rather poor at screening interactions and the effective Coulomb interaction between an electron and a hole is simply much stronger in 2D than in 3D.

From a first principles point of view, the treatment of excitons requires advanced computational methodology such as the Bethe-Salpeter equation [11, 12]. This approach has been applied to obtain absorption spectra for numerous insulators and usually yields very good agreement with experiments [13]. However, only systems of modest size can be treated this way and simplified models of excitons will be an inevitable ingredient in calculations of realistic systems. For example, if the effect of substrates or dielectric environment is to be included in the calculation of excitons in 2D systems [14], the computations become intractable with a standard Bethe-Salpeter approach. For three-dimensional materials the Mott-Wannier model comprises a strong conceptual and intuitive picture that provides a simple framework for calculating exciton binding energies [15]. In the center of mass frame, an excited electron-hole pair can be shown to satisfy a hydrogenic Schrödinger equation, where band structure effects are included through an excitonic effective mass μ and the dielectric screening from the environment is included through the static dielectric constant ϵ_0. The exciton binding energy in atomic units is then written as

$$E_B^{3D} = \frac{\mu}{2\epsilon_0^2}. \tag{1}$$

Thus the daunting task of solving the Bethe-Salpeter equation, has been reduced to the calculation of just two parameters: the effective mass and the static dielectric constant, both of which are easily obtained with any standard electronic structure software package. This approximation is well justified whenever the screening is local such that its Fourier transform can be approximated by a constant in the vicinity of the origin. However in highly anisotropic structures such as layered materials this assumption is expected to break down.

In 2D dielectrics, it is well known that the screening takes the form $\epsilon(\mathbf{q}) = 1 + 2\pi\alpha q$ [2], where α is the 2D polarizability. The screening is thus inherently non-local in real space, and it is not obvious if it is possible to arrive at a hydrogenic model like Eq. (1). Instead one can calculate the 2D screened potential and solve the Schrödinger equation for the electron-hole wavefunction:

$$\Big[-\frac{\nabla^2}{2\mu} + W(\mathbf{r}) \Big] \psi(\mathbf{r}) = E_n \psi(\mathbf{r}), \tag{2}$$

where $W(\mathbf{r})$ is the 2D convolution of the Coulomb interaction and $\epsilon^{-1}(\mathbf{r} - \mathbf{r}')$. This approach has previously been shown to provide good agreement with the Bethe-Salpeter equation [14, 16]. In the supplementary material we have tabulated the binding energies of 7 TMDs obtained with the Bethe-Salpeter equation and show that these agree well with results obtained from Eq. (2). However, in general the solution of Eq. (2) is a tedious task and it would be highly desirable to have an expression like Eq. (1) from which the exciton binding energy in a given material can be easily estimated and understood. To accomplish this, we calculate the average screening felt by the exciton. To this end we consider the expres-

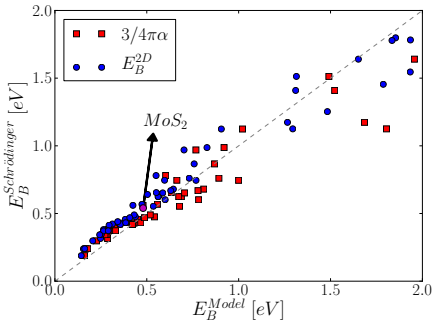


FIG. 1. (Color online) Exciton binding energies of 51 transition metal dichalcogenides calculated as the lowest eigenvalue of Eq. (2) (vertical axis) and the model result Eq. (5) (horizontal axis). We have indicated the well known example of MoS_2.

sion

$$\epsilon_{eff} = \frac{a_{eff}^2}{\pi}\int_0^{2\pi} d\theta \int_0^{1/a_{eff}} dq q \epsilon(\mathbf{q}), \qquad (3)$$

where a_{eff} is the effective Bohr radius. For the 2D hydrogen atom the Bohr radius is given by $a = \epsilon/(2\mu)$ and Eq. (4) has to be solved self-consistently for ϵ_{eff} given an expression for $\epsilon(\mathbf{q})$. In a strictly 2D system the screening is linear in q and Eq. (3) can be solved to yield

$$\epsilon_{eff} = \frac{1}{2}\Big(1 + \sqrt{1 + 32\pi\alpha\mu/3}\Big). \qquad (4)$$

Using that the hydrogenic binding energy in 2D is a factor of four larger than in 3D [9], we obtain

$$E_B^{2D} = \frac{8\mu}{\big(1 + \sqrt{1 + 32\pi\alpha\mu/3}\big)^2}. \qquad (5)$$

This is the main result of the present letter and comprises a long-sought-for 2D analog of Eq. (1).

A remarkable property of the expression (5) is the fact that it becomes independent of the effective mass if the polarizability is large. More precisely

$$E_B^{2D} \approx \frac{3}{4\pi\alpha}, \qquad 32\pi\alpha\mu/3 \gg 1. \qquad (6)$$

It may come as a surprise that the binding energy becomes independent of mass, since a large mass gives rise to a localized exciton and the binding energy typically increases with localization. This is reflected in Eq. (1), where the binding energy is seen to be proportional to the mass. However, in 2D, short range interactions are screened more effectively than long range interactions. Thus, there are two opposing effects of the exciton mass and for large polarizabilities, the binding energy becomes independent of mass. In order to assert the applicability of the expressions (5)-(6), we have calculated the indirect effective masses and static polarizabilities (in the Random Phase Approximation) of 51 semiconducting monolayers of transition metal dichalcogenides. The calculations were performed with the electronic structure code GPAW [17, 18], which is based on the projector augmented wave formalism. The calculations were converged with respect to the number of unoccupied bands included (typically on the order of 50-100). Further details on the calculations can be found in Ref. [19]. In Fig. 1 we compare the model binding energies with the full solution of Eq. (2). Using the expression (5), the agreement is seen to be on the order of 10%. For materials with anisotropic mass tensor we have used the average mass, both in the model and when solving the Schrodinger equation. With the approximated expression (6), we obtain excellent agreement for binding energies up to ~ 0.5 eV, whereas the binding energies are underestimated for strongly bound excitons.

Recently, first principles calculations have indicated that exciton binding energies in different 2D materials scale linearly with the band gaps [20]. In the present model, this behavior comes out naturally since without local field effects, the in-plane components of the polarizability in the Random Phase Approximation are given by

$$\alpha = \sum_{m,n}\int_{BZ}\frac{d\mathbf{k}}{(2\pi)^2}(f_{n\mathbf{k}} - f_{m\mathbf{k}})\frac{|\langle u_{m\mathbf{k}}|\hat{r}_\parallel|u_{n\mathbf{k}}\rangle|^2}{\varepsilon_{n\mathbf{k}} - \varepsilon_{m\mathbf{k}}}, \qquad (7)$$

and we expect that α will be roughly inversely proportional to the band gap. This is illustrated in Fig. 2 for the 51 transition metal dichalcogenides. Combining this with Eq. (6) thus gives $E_B^{2D} \propto E_{gap}$. However, in the present model the scaling originates solely from the screening and not the effective mass as previously proposed [20]. For the present set of materials, we do not observe any correlation between binding energies and effective mass. We use the LDA band gaps and not the quasiparticle gaps, which could be obtained from for example GW calculations [19], since LDA typically gives a better estimate of the two-particle excitation energies that enters the expression for α. In contrast, the use of GW gaps would underestimate the screening due to the lack of electron-hole interactions.

To validate the general applicability of the effective screening model, we now show that it can also be used to account for the entire exciton spectrum in 2D materials. In Ref. [21], the exciton spectra of graphene derivatives was predicted to deviate from the 2D Rydberg series and in Ref. [22], the exciton spectrum of WS_2 was measured and shown to deviate significantly from the Rydberg series of a 2D hydrogen model scaled by an overall screening factor. The reason is simply that the effective screening depends on the n quantum number due to the increasing spatial extend of higher lying Rydberg states. The

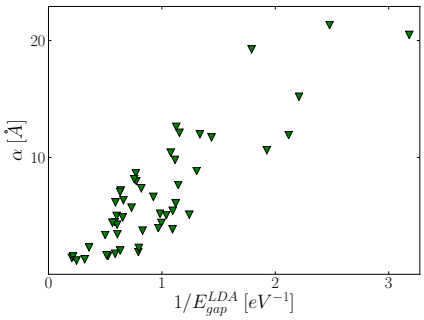


FIG. 2. (Color online) The 2D polarizability of 51 transition metal dichalcogenides shown as a function of LDA band gaps.

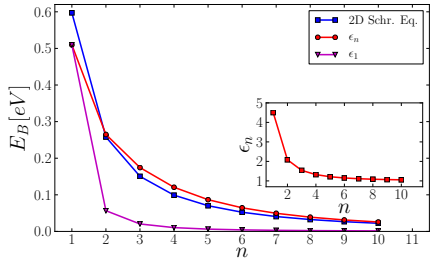


FIG. 3. (Color online) Rydberg series of a monolayer of 2H-WS_2 calculated with the generalized hydrogen model with linear screening (Eqs. (8) and (10)) and from the solution of the 2D screened Schrödinger equation (2). The results are compared with the bare hydrogen model where the effective screening obtained from the ground states is used for all states.

authors used the results to define an n-dependent effective screenings ϵ_n, which were then determined by fitting each term in the Rydberg series to a 2D hydrogen model. The Rydberg series is then given by

$$E_n^{2D} = -\frac{\mu}{2(n-\frac{1}{2})^2\epsilon_n}. \tag{8}$$

Two of the present authors have recently showed that the Rydberg series can accurately be reproduced by solving Eq. (2) with a screened 2D potential calculated from first principles [14] and we will assume that approach to be an accurate reference. Here we calculate the n-dependent effective screening from first principles by replacing a_{eff} in Eq. 3 by an n-dependent characteristic extension of the state. To this end we note that for $l = 0$, the first moment of a state with principal quantum number n in a 2D hydrogen atom is [9]

$$a_n \equiv \langle n|\hat{r}|n\rangle = [3n(n-1)+1]/(2\mu), \tag{9}$$

where $\hat{r} = \sqrt{\hat{x}^2 + \hat{y}^2}$. In terms of this, the a_{eff} defined previously is given by a_1 and E_B^{2D} is $-E_1^{2D}$. Within the linear model the effective screening for state n then becomes

$$\epsilon_n = \frac{1}{2}\left(1 + \sqrt{1 + \frac{32\pi\alpha\mu}{9n(n-1)+3}}\right). \tag{10}$$

It is straightforward to generalize these expressions to $l \neq 0$ [9], which results in a larger value of the effective radius a_{nl} and thus $\epsilon_{n,l>0} < \epsilon_{n,l=0}$. The energy is still given by Eq. (8) and at a given n, the higher angular momentum excitons will therefore have a larger binding energy, which has been observed in the case of 2H-WS_2 monolayers [6]. As a case study we consider this material and apply the linear screening model. We obtain a first principles 2D polarizability of $\alpha = 5.25$ Å and $\mu = 0.19$. In Fig. 3 we show the Rydberg series calculated with the generalized hydrogen model, which agrees very well with a full solution of Eq. (2). In contrast, the pure 2D hydrogen model with an overall effective screening is seen to significantly underestimate the binding energies at higher lying states, since the decreased screening of extended states is not taken into account. We also note that the model binding energies of the $n = 1$ state agree very well with a full solution of the Bethe-Salpeter equation which yields an exciton binding energy of 0.54 eV [23].

We now proceed to show how the effect of screening by the environment can naturally be taken into account in the present framework. It should be noted, however, that the linear model for the screening is expected to break down for systems where the vertical extend of a substrate becomes comparable to the Bohr radius of the exciton. For example, if we consider a stack of N monolayers, α will diverge in the limit of large N, since the bulk system will have $\epsilon(\mathbf{q} = \mathbf{0}) \neq 1$ [14, 24]. The linear regime will therefore only be valid at infinitesimal values of $\mathbf{q}$ when N becomes large. As an example where we expect the linear model to be applicable we consider a monolayer 2H-MoS_2 and compare the isolated layer with the two cases where it is in the vicinity of another layer of 2H-MoS_2 and in the vicinity of a metallic layer of 1T-MoS_2. In Fig 4, we show the absorption spectrum calculated from the Bethe-Salpeter equation based on Kohn-Sham eigenvalues. The BSE calculations were performed in a plane wave basis with a 2D Coulomb truncation scheme [25, 26] using a 60×60 k-point mesh. It is well known that the low energy absorption spectrum of this system exhibits a double excitonic peak due a spin-orbit split valence band [27, 28]. This facilitates the identification of the excitons in the 2H-MoS_2 layer in the vicinity of a metallic substrate with low lying excitations. We have not performed the full spinorial BSE calculations, but simply included spinorbit effects in the band structure in order to identify the excitons. In the following we

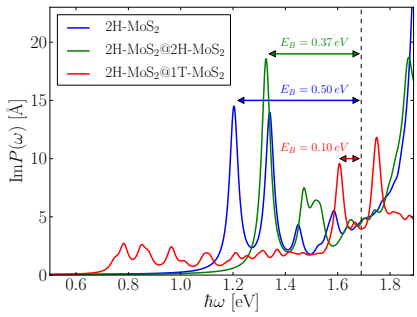


FIG. 4. (Color online) Dynamic 2D polarizability of 2H-MoS_2 in different environments calculated from the Bethe-Salpeter equation based on Kohn-Sham eigenvalues. The vertical lines at 1.7 eV marks the Kohn-Sham band gaps, which are nearly identical in the three cases.

consider the binding energies of the lowest exciton. The isolated layer exhibits an exciton bound by 0.50 eV. In the vicinity of another 2H-MoS_2 layer, the binding energy is decreased to 0.37 eV and the metallic 1T-MoS_2 decreases the binding energy to 0.10 eV. We note that the quasiparticle band structure corrections are expected to be much smaller for the case of 2H-MoS_2@1T-MoS_2 such that the actual positions of the excitons would be similar for the three cases in an optical absorption experiment. However, we have chosen to leave out the quasiparticle corrections in order to illustrate the difference in binding energies more clearly.

To apply the model we wish to calculate $\epsilon(\mathbf{q})$ for the 2H-MoS_2 layer when it is in the vicinity of a screening environment. For small q, we may still write it as $\epsilon(\mathbf{q}) = 1 + 2\pi\tilde{\alpha}q$ and we would like to extract $\tilde{\alpha}$, which is the relevant quantity for the screened hydrogen model. We calculate it by the finite difference

$$2\pi\tilde{\alpha} = \epsilon(\mathbf{q}_1) - 1, \tag{11}$$

where $\mathbf{q}_1$ is a small finite value of q. In the present case we take $\mathbf{q}_1$ as the smallest q-vector in the direction of K obtained from a 60×60 k-point grid. The 2D dielectric function is obtained from

$$\frac{1}{\epsilon(\mathbf{q})} = \frac{\langle V_{tot}(\mathbf{r}_\parallel, z_0)e^{-i\mathbf{q}\cdot\mathbf{r}}\rangle_A}{V_\mathbf{q}}, \tag{12}$$

where $V_{tot}(\mathbf{r})$ is the total potential resulting from an external perturbation $V_{ext}(\mathbf{r}) = V_\mathbf{q}e^{i\mathbf{q}\cdot\mathbf{r}}$ and $\langle\ldots\rangle_A$ denotes average over the 2D unit cell of area A. It is straightforward to relate this expression to an average over the microscopic dielectric function $\epsilon^{-1}(\mathbf{r}, \mathbf{r}')$, which can be calculated in the Random Phase Approximation by most electronic structure codes. We take z_0 to be at the center of the 2H-MoS_2 layer, but we note that $\tilde{\alpha}$ is approximately independent of the value of z_0 when z_0 chosen in any part of the central 3.0 Å of the layer. In Tab. I, we display the calculated values of $\tilde{\alpha}$ along with the exciton binding energies obtained from the model (5), the 2D Schrödinger equation (2), and the BSE calculations. As expected, the environment strongly affects the value of $\tilde{\alpha}$. In particular, the metallic 1T-MoS_2 layer significantly increases the screening, whereas the presence of another 2H-MoS_2 layer results in a less pronounced effect. We find good agreement between the simple model, the 2D Schrödinger, and the BSE calculations. We should note that the convergence of the exciton binding energies in the presence of the metallic 1T-MoS_2 layer is very slow with respect to k-point sampling and the converged result is expected to exhibit a lower binding energy than the one obtained here. Furthermore, we have not included the intraband contribution to the static screening, which, is expected to scale as $\sim 1/q$ in 2D metals. In fact it is not clear that the dynamic contributions to the screening can be neglected in the either the BSE approach or the model. On the other hand, the 1T structure is known to distort into the so-called 1T' structure, which is a topological insulator with a gap on the order 50 meV [29]. In any case, the screening is treated at the same footing in the BSE and the model calculations since the values of $\tilde{\alpha}$ were obtained by a finite difference calculation on the same k-point grid that was used in the BSE calculations. Nevertheless, the model is easily generalized to a non-linear $\epsilon(\mathbf{q})$, the only difference being that (3) should be solved numerically. We note again that the applicability of this approach is limited to cases where the linear model is expected to be agood approximation. The exact conditions under which the linear model is applicable will depend on the thickness of the substrate as well as the screening properties of the substrate. For extended substrates, the present approach may be generalized by calculating the full $\epsilon(\mathbf{q})$ and solving Eq. (3) numerically, but it is not clear that the analytical results derived from the 2D hydrogen model (8) is able to produce reliable results in this case. Alternatively one may solve a quasi-2D Schrödinger equation that incorporates the finite extend of the slab [24]. We note that the present method can be viewed as a generalized hydrogen model analogue of the approach taken by Ugeda et al [7], where the full substrate screening was taken into account when solving the Bethe-Salpeter equation for the layer.

To conclude, we have presented an analytical expression for the exciton binding energies in 2D semiconductors that only depends on the static 2D polarizability and the effective mass and produces quantitative agreement with the solution of the full screened 2D Schrödinger equation. It has also been shown that for large polarizabilities, the result becomes independent of mass and yields a linear relation between exciton binding energies and band gaps. It has previously been anticipated that the non-hydrogenic Rydberg series could be attributed to an n-dependent value of the effective screening [22]. Here

	2H-MoS_2	2H-MoS_2@2H-MoS_2	2H-MoS_2@1T-MoS_2
E_B^{BSE} [eV]	0.50	0.37	0.10
$E_B^{Schr.}$ [eV]	0.54	0.40	0.17
E_B^{Model} [eV]	0.48	0.30	0.10
$\tilde{\alpha}$ [Å]	5.83	10.0	30.1

TABLE I. Exciton binding energies for 2H-MoS_2 in different environments calculated from the Bethe-Salpeter equation (BSE), the 2D Schrödinger equation, and the generalized screened hydrogen model. We also display the values of $\tilde{\alpha}$, which is the polarizability of the single 2H-MoS_2 layer used in the calculations. For all calculations we used an effective exciton mass of 0.276, which was obtained from the ab initio band structure.

we have obtained an explicit expression for ϵ_n that provides an accurate account of the full exciton spectrum. It has also been shown, that the model can be generalized to incorporate the effect of a simple screening environment. We do not claim that the presented expression for the effective screening (3) in the linear model is unique. In fact, it is based on an unweighted average of a linear model for the non-local 2D screening over the extend of the exciton and it is easy to imagine more elaborate averaging schemes. However, we believe that the simplicity is the main merit of this procedure and the resulting analytical expressions are very easy to apply to a given 2D material. In particular, for complicated structures it may not be possible to treat the electron-hole interaction by a first principles approach and our model results could be a crucial ingredient in understanding the excitonic structure in such materials.

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 (CNG) is sponsored by the Danish National Research Foundation, Project DNRF58.

* tolsen@fysik.dtu.dk

[1] L. Wirtz, A. Marini, and A. Rubio, Phys. Rev. Lett. **96**, 1 (2006)
[2] P. Cudazzo, C. Attaccalite, I. V. Tokatly, and A. Rubio, Phys. Rev. Lett. **104**, 1 (2010)
[3] F. Hüser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **88**, 245309 (2013)
[4] H. P. Komsa and A. V. Krasheninnikov, Phys. Rev. B - Condens. Matter Mater. Phys. **86**, 1 (2012
[5] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **111**, 216805 (nov 2013)
[6] Z. Ye, T. Cao, K. OâBrien, H. Zhu, X. Yin, Y. Wang, S. G. Louie, and X. Zhang, Nature **513**, 214 (aug 2014)
[7] M. M. Ugeda, A. J. Bradley, S.-F. Shi, F. H. da Jornada, Y. Zhang, D. Y. Qiu, W. Ruan, S.-K. Mo, Z. Hussain, Z.-X. Shen, F. Wang, S. G. Louie, and M. F. Crommie, Nat. Mater. **13**, 1091 (2014)
[8] A. Hanbicki, M. Currie, G. Kioseoglou, A. Friedman, and B. Jonker, Solid State Commun. **203**, 16 (2015)
[9] X. L. Yang, S. H. Guo, F. T. Chan, K. W. Wong, and W. Y. Ching, Phys. Rev. A **43**, 1186 (1991)
[10] P. Cudazzo, I. V. Tokatly, and A. Rubio, Phys. Rev. B - Condens. Matter Mater. Phys. **84**, 1 (2011)
[11] S. Albrecht, L. Reining, R. Del Sole, and G. Onida, Phys. Rev. Lett. **80**, 4510 (1998)
[12] M. Rohlfing and S. Louie, Phys. Rev. Lett. **81**, 2312 (1998)
[13] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002)
[14] K. Andersen, S. Latini, and K. S. Thygesen, Nano Lett., 4616(2015)
[15] G. H. Wannier, Phys. Rev. **52**, 191 (aug 1937)
[16] T. C. Berkelbach, M. S. Hybertsen, and D. R. Reichman, Phys. Rev. B - Condens. Matter Mater. Phys. **88**, 1 (2013)
[17] J. Enkovaara, C. Rostgaard, J. J. Mortensen, J. Chen, M. Dulak, L. Ferrighi, J. Gavnholt, C. Glinsvad, V. Haikola, H. A. Hansen, H. H. Kristoffersen, M. Kuisma, A. H. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsaris, M. Vanin, M. Walter, B. Hammer, H. Häkkinen, G. K. H. Madsen, R. M. Nieminen, J. K. Nørskov, M. Puska, T. T. Rantala, J. Schiøtz, K. S. Thygesen, and K. W. Jacobsen, J. Phys. Condens. Matter **22**, 253202 (2010)
[18] J. Yan, J. J. Mortensen, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **83**, 245122 (2011)
[19] F. A. Rasmussen and K. S. Thygesen, J. Phys. Chem. C **119**, 13169 (2015)
[20] J.-H. Choi, P. Cui, H. Lan, and Z. Zhang, Phys. Rev. Lett. **115**, 066403 (2015)
[21] S. Huang, Y. Liang, and L. Yang, Phys. Rev. B - Condens. Matter Mater. Phys. **88**, 1 (2013)
[22] A. Chernikov, T. C. Berkelbach, H. M. Hill, A. Rigosi, Y. Li, O. B. Aslan, D. R. Reichman, M. S. Hybertsen, and T. F. Heinz, Phys. Rev. Lett. **113**, 076802 (aug 2014)
[23] H. Shi, H. Pan, Y.-W. Zhang, and B. I. Yakobson, Phys. Rev. B **87**, 155304 (2013)
[24] S. Latini, T. Olsen, and K. S. Thygesen(2015), arXiv:1509.07972, `http://arxiv.org/abs/1509.07972`
[25] C. A. Rozzi, D. Varsano, A. Marini, E. K. U. Gross, and A. Rubio, Phys. Rev. B **73**, 205119 (2006)
[26] F. Hüser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **87**, 235132 (2013)
[27] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Phys. Rev. Lett. **105**, 2 (2010)
[28] A. Molina-Sánchez, D. Sangalli, K. Hummer, A. Marini, and L. Wirtz, Phys. Rev. B **88**, 045412 (2013)
[29] X. Qian, J. Liu, L. Fu, and J. Li, Science (80-.). **346**, 1344 (2014)
[30]
[31] 08(1)

Paper V

Mohnish Pandey et al. “Defect Tolerance in Monolayer Transition Metal Dichalcogenides”. Submitted to Nano Letters. December 2015

Defect Tolerance in Monolayer Transition Metal Dichalcogenides

Mohnish Pandey,[†] Filip A. Rasmussen,[†] Korina Kuhar,[†] Thomas Olsen,[†] Karsten W. Jacobsen,[†] and Kristian S. Thygesen[*,†,‡]

Center for Atomic-scale Materials Design, Department of Physics, Technical University of Denmark, DK - 2800 Kongens Lyngby, Denmark, and Center for Nanostructured Graphene (CNG), Department of Physics, Technical University of Denmark, DK - 2800 Kongens Lyngby, Denmark

E-mail: thygesen@fysik.dtu.dk

[*]To whom correspondence should be addressed

[†]Center for Atomic-scale Materials Design, Department of Physics, Technical University of Denmark, DK - 2800 Kongens Lyngby, Denmark

[‡]Center for Nanostructured Graphene (CNG), Department of Physics, Technical University of Denmark, DK - 2800 Kongens Lyngby, Denmark

Abstract

Localized electronic states formed inside the band gap of a semiconductor due to crystal defects can be detrimental to its opto-electronic properties. Defect tolerant semiconductors have in general lower tendency to form such defect-induced deep gap states. Here we provide a systematic first-principles investigation of defect tolerance in more than 40 monolayer transition metal dichalcogenides and oxides (TMDs) of interest for nano-scale optoelectronics. We show that upon formation of a chalcogen/oxygen (S, Se, Te, O) vacancy, all TMDs with similar orbital character of their valence and conduction bands form deep gap states while the remaining materials do not. The analysis is made quantitative by introducing a descriptor that measures the difference in average orbital character of two groups of bands. The study is generalized to the nanoribbons of the TMDs where cleavage of the monolayer to form nanoribbons results to deep gap states only in the defect sensitive TMDs thus opening new avenues for one-dimensional TMDs as well. Our result thus allows for quantitative measures of the degree of defect tolerance in semiconductors and should be useful for computational materials design.

Single layers of semi-conducting transition metal dichalcogenides (TMDs) are attracting much attention due to their unique opto-electronic properties. The Mo and W based TMDs have been found to undergo a transition from indirect to direct band gap materials when thinned down from the bulk to a single layer.[1–4] Additionally, interesting properties like valley sensitive spin-orbit interactions,[5,6] decent charge carrier mobilities,[7–9] and strong interaction with light resulting from large oscillator strengths and tightly bound excitons,[10,11] has led to a urge in research on these novel 2D materials. We note in passing that similar properties are expected for the Cr-based TMDs[12] but to our knowledge this has still not been experimentally verified. The electronic properties of 2D crystals can be further tuned e.g. by applying strain,[13] electrostatic gating[14,15] or by varying the dielectric environment.[16] Most recently, the idea of stacking different 2D materials into van der Waals heterostructures has opened new avenues in 2D materials research.[17,18] In parallel with this

development new 2D materials are continuously being discovered, for example, monolayers and multilayers of $MoTe_2$, $NbSe_2$, $NiTe_2$, TaS_2, $TaSe_2$, TiS_2, WS_2, WSe_2, ZrS_2 have recently been synthesized.[19]

One of the main performance limiting factors of semiconductors for optoelectronics is the presence of defects such as vacancies, impurities or crystal disorder. Such defects can act as local scattering centers which reduce the mobility of charge carriers and enhance recombination of photo-excited electron-hole pairs. The effectiveness of a defect to scatter charge carriers, trap excitons and induce recombination between electrons and holes depends crucially on the way the defect affects the electronic structure around the band edges; in particular whether or not it introduces localized states inside the band gap (deep gap states). Semiconductors which tend not to form deep gap states due to defects are termed defect tolerant.[20] The microscopic origin of defect tolerance was recently theoretically investigated in the context of photovoltaic materials by Zakutayev *et al.* who found that the orbital character of the valence band maximum (VBM) and conduction band minimum (CBM) is crucial for the tendency of the material to form deep gap states.[20] Specifically, they proposed that materials with anti-bonding VBM and bonding CBM are less likely to form localized states in the gap than materials whose VBM has bonding character and/or CBM has anti-bonding character. The reason is that the dangling bonds created in the vicinity of the defect tend to fall outside the gap in the former case and inside it in the latter case, see Figure 1. We stress that this rule should apply to the case of vacancies, crystal distortions, or other perturbations whose effect is to distort the intrinsic bonding. On the other hand, in the case of impurity atoms, the presence of deep gap states depends also on the energy of the atomic orbitals of the impurity atom relative to the band edges.

There have been several experimental[21,22] and theoretical[23] studies of defects and their influence on the electronic properties of few-layer MoS_2 - the most well studied of the TMDs. These studies indicate that S vacancies are the most common type of defects and that they lead to the formation of localized states inside the band gap. These sulphur vacancies are

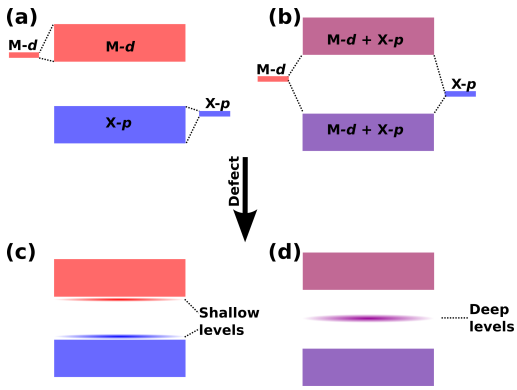


Figure 1: (a) and (b) show the nature of the band structures near the band edges for the defect tolerant and defect sensitive cases respectively. In all the TMDs studied here, the states near the band edges primarily have contributions from the metal d states and chalcogen/oxygen p states. In the defect tolerant case the nature of the bands near the band edges are significantly different whereas in the defect sensitive case they are of mixed nature. (c) and (d) show the shallow and deep levels introduced after the creation of defects.

likely to play an important role for the relatively low mobility of back gated MoS_2 which is usually one order of magnitude lower than the theoretical limit set by phonon scattering.[7,8] We mention that mobilities very close to the theoretical limit was recently obtained in van der Waals heterostructure devices where the MoS_2 layer was fully encapsulated into hexagonal boron nitride.[9]

In this Letter, we systematically explore the tolerance of monolayer TMDs to chalcogen vacancies. Using ab-initio methods we calculate band structures of 41 representative monolayer semiconducting TMDs with and without chalcogen vacancies. The compounds have been selected from a 2D materials database which contains different electronic properties of TMDs calculated with the first-principle methods.[12,24] To quantify the orbital character of the conduction/valence band manifolds we introduce a measure based on the projected density of states. We find that localized states in the band gap are introduced for all the materials whose valence/conduction bands have similar orbital characters at the conduction and valence bands (normalized orbital distance close to 0) while no states or only shallow

states are introduced for the other materials (normalized orbital distance close to 1). This provides further evidence for using the orbital character of the conduction/valence bands as a simple descriptor when searching for new opto-electronic materials. Using this descriptor we predict 19 out of the 41 TMDs to be defect tolerant. Additionally, we also explore nanoribbons of all the TMDs and find that cleaving the monolayer along a non-polar direction of the defect tolerant TMDs induces only shallow states in the band gap as opposed to the nanoribbons of defect sensitive TMDs which have deep gap states.

All calculations were performed with the GPAW electronic structure code[25] which is based on the projector augmented wave method for describing the ion cores. The wavefunctions were expanded on a real space grid with a grid spacing of 0.18 Å and we use the PBE xc-functional.[26] All pristine structures have been relaxed until the forces on each atom converged down to 0.05 eV/Å. The structures have not been relaxed after creating the chalcogen vacancy because of the neglible difference in the band structure of the relaxed and unrelaxed structures of the represtative compounds i.e. MoS_2 and HfS_2. We note that the PBE is known to underestimate the band gap of semiconductors. However, since the present study is concerned with qualitative features of the band structure, i.e. whether or not the vacancies introduced localized states in the gap, rather than the absolute values of band edges and gap states, the PBE description can be justified. A prototypical structure used for the calculation of defects is shown in the Figure 2.

Figure 3 (a) shows the PBE band structures, the total density of states (DOS) and the projected density of states (PDOS) projected onto the chalcogen/oxygen p-orbitals and metal d-orbitals of 2H-MoS_2 in its pristine form (left) and and with an S vacancy (right). The narrow band inside the band gap is due to deep gap states localized around the S vacancy. From the projected density of states (PDOS) it is seen that the VBM and CBM have very similar orbital character indicating that they consist of bonding and anti-bonding combinations of sulphur p and metal d states, respectively. We note that the finite size of the supercell (3×3) is responsive for the small dispersion of the deep gap state as well as

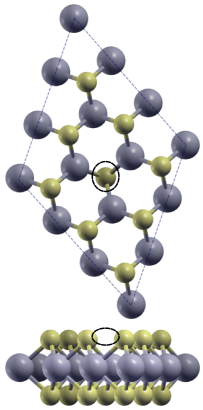


Figure 2: Top and sideview of a prototypical structure of a metal dichalcogenide in the 2H structure used for the defect calculations. The unit cell is shown with the dotted lines. The defect structure is created by making a chalcogen vacancy as shown with a black ellipse. Similar chalcogen vacancies have been created in the 1T structure as well.

the general spectral smearing of the band structure. The smearing significantly reduces in 4×4 supercell, however, due to qualitative similar band structures of 3×3 and 4×4 supercells all the calculations have been performed for 3×3 supercell to save computational resources without affecting the conclusion.

Figure 3 (b) shows a similar plot as Figure 3 (a) for 1T-HfS_2. In contrast to MoS_2 this compound largely conserves its electronic structure around the band edges and no defect state is introduced. Additionally, the DOS plot shows that the states near the VBM are mostly dominated by the chalcogen p states whereas the CBM edge states mainly consist of the metal d states.

The above examples indicate that the orbital character of the valence/conduction bands are crucial for the tendency to form deep gap states. To quantify the orbital character of the electronic states in a given energy window from E_1 to E_2 we introduce the orbital fingerprint

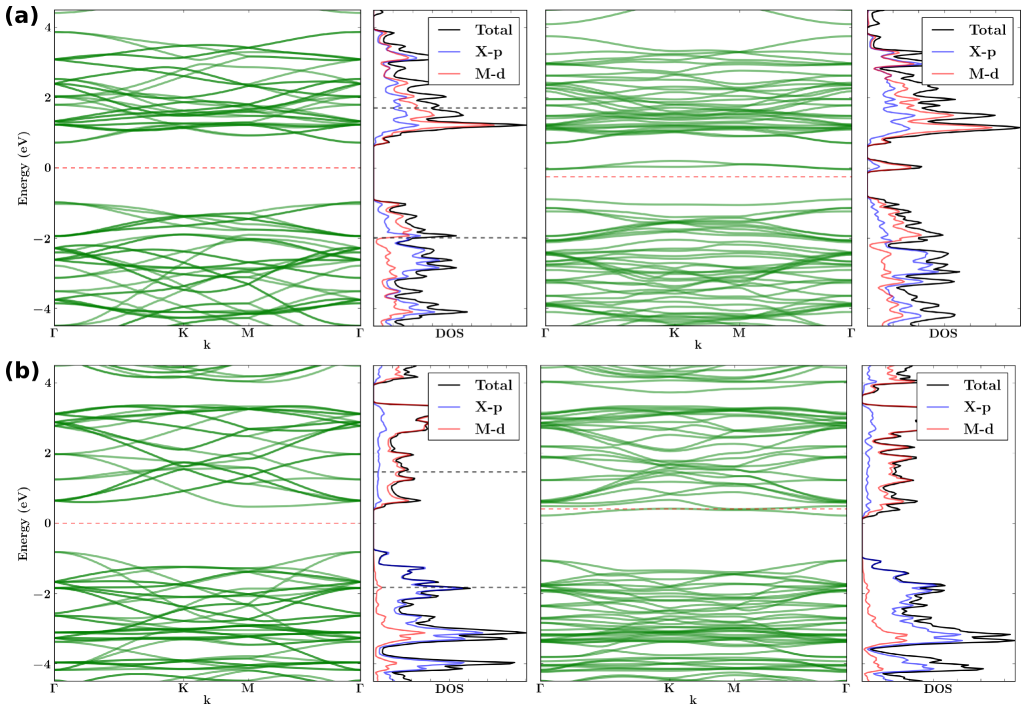


Figure 3: PBE band structures, the total density of states (DOS) and the projected density of states (PDOS) projected onto the chalcogen/oxygen *p*-orbitals and metal *d*-orbitals of 2H-MoS_2 in its pristine form (left) and and with an S vacancy (right). The energy levels have been aligned to the Fermi level of the pristine monolayer.

vector,

$$|\alpha\rangle = \frac{1}{\sqrt{c}} \begin{bmatrix} \rho_{\nu_1} \\ \rho_{\nu_2} \\ \vdots \\ \rho_{\nu_N} \end{bmatrix} \tag{1}$$

where c is a normalization constant, ν_i's are d and p states and ρ_{ν_i} is the projected density of states onto the atomic orbital ϕ_{ν_i} integrated over the energy window,

$$\rho_{\nu_i} = \sum_n \int_{E_1}^{E_2} |\langle \psi_n | \phi_{\nu_i} \rangle|^2 \delta(E - E_n) dE \tag{2}$$

Using the orbital character vector we can define the normalized distance between two manifolds of bands located in the energy windows E_1^v to E_2^v and E_1^c to E_2^c as

$$D = ||\,|\alpha\rangle - |\beta\rangle\,|| \tag{3}$$

where α and β correspond to the valence and conduction band manifolds. By taking the two energy windows to lie around the VBM and CBM, respectively, we have a measure of the difference in average orbital character around the valence and conduction band edges. We note that $D = 1$ for materials with completely different character of the valence and conduction bands while $D = 0$ for materials with identical orbital character at the valence and conduction bands.

We have computed D (with an energy window of 1 eV above/below are conduction/valence band extrema) for a set of 41 monolayer TMDs, see Figure 4. The defect sensitive materials are identified with $D < 0.5$ whereas the defect tolerant materials with $D > 0.5$. We have carried out the band structure calculations like those shown in Figure 3 for all 41 TMDs (see supplementary information). For all the materials with $D < 0.5$ we find localized states inside the band gap while for the materials with $D > 0.5$ we find no deep gap states. The

compounds exhibiting deep defect states after the removal of a chalcogen atom are shown with red circles in the Figure 4 whereas the compounds where no deep defect states are introduced after a chalcogen atom is removed are marked with green' circles. As expected, the figure clearly shows that the band structure of the compounds with low D values is sensitive towards defects whereas high D values represent a class of defect tolerant compounds. This clearly shows that the normalized orbital distance between the valence and conduction band manifolds represent a reliable and quantitative descriptor for the degree of defect tolerance of the material.

Additionally, we also calculated the band structure of nanoribbons cleaved from the monolayers of TMDs. The monolayer cleaved along a polar direction will have metallic edge states induced due to the presence of dipole across the edges.[27] However, cleaving the monolayer along a non-polar direction will introduce edge states purely due to the dangling bonds thus having a close resemblance to the case of a monolayer with a vacancy where the shallow/deep levels arise due to the presence of dangling bonds. Therefore, we expect that the arguments for the monolayers with vacancies will also be applicable for the edges in the nanoribbons.

Figure 5 (a) and (b) show the structure of the non-polar nanoribbons of the 2H and 1T structures respectively used for the analysis of the edge states. The edges of the nanoribbons lie along the x-axis with a finite width along the y-axis. In Figure 4 the nanoribbons manifesting edge states lying deep in the band gap are shown with hollow squares. The figure clearly shows that the only defect sensitive TMDs introduce edge states deep in the band gap whereas the defect tolerant TMDs introduce only shallow or no edges in deep in the band gap (see supplementary information). The correspondence between the band structure of the monolayer with vacancies and the edge states in nanoribbons indicates that the analysis based on the orbital character of the states near the valence and conduction band edge is completely general and can be applied to other system with imperfections involving dangling bonds.

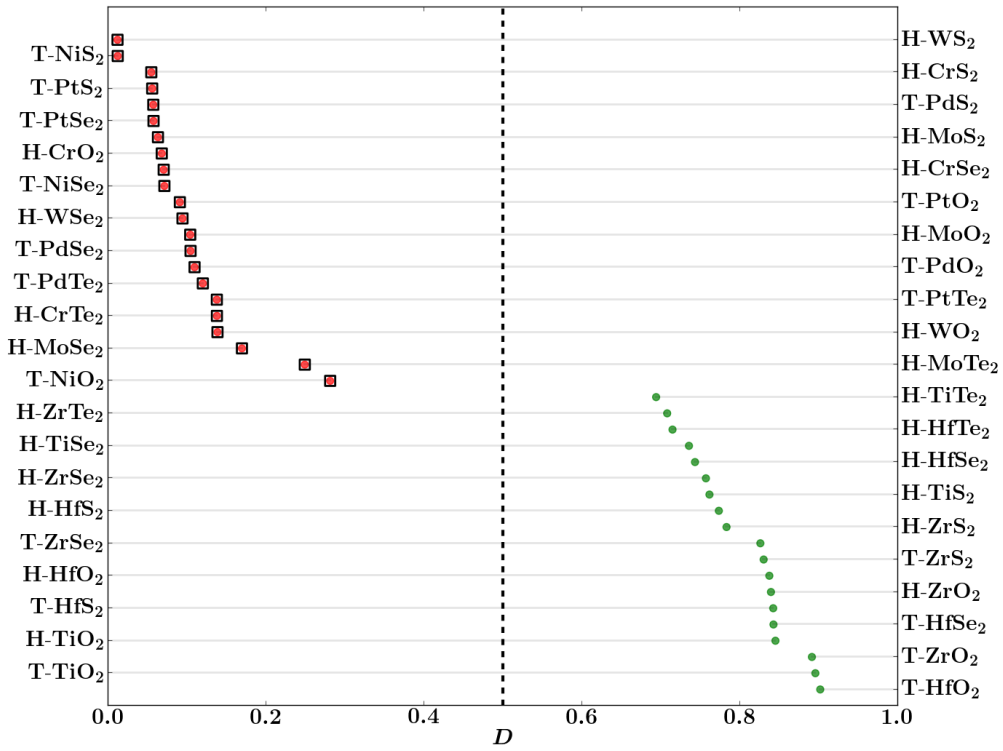


Figure 4: The plot showing the compounds (y-axis) and the corresponding D values (x-axis). Red circles indicate the compounds manifesting deep defect states and the green circles indicating the compounds showing no deep defect states after the removal of a chalcogen atom. The black squares represent compounds showing states appearing deep in the band gap after cleaving the monolayer to form nanoribbons.

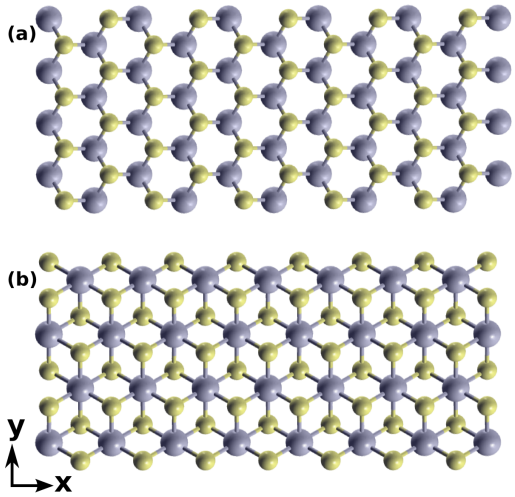


Figure 5: (a) shows the structure of a nanoribbon of the 2H structure cleaved along the non-polar direction. The edges of the nanoribbons lie along the x-axis with a finite width along the y-axis; (b) shows a nanoribbon cleaved from the 1T structure.

In summary, we have explored the sensitivity of the band structure of 2D TMDs towards chalcogen vacancies. Our analysis shows that the tendency of the materials to form localized states within the band gap levels strongly depends on the similarity of the orbital character of the states near the conduction and valence band. The results are made quantitative by introducing a descriptor to measure the difference in the orbital character of the valence and conduction bands. A strong predictablity for the nature of the defect states and the edge states in non-polar nanoribbons using the descriptor provides an elegant way to search for defect tolerant materials. Since the approach used is completely general e.g. independent of the dimensionality of the material, therefore, it can be employed to study defect tolerance in other semiconductors.

Supplementary Information

Band structures of 41 TMDs monolayers in the pristine and defect forms and the band structure of the non-polar nanoribbobs are provided. The plot of the descriptor for different energy windows is also shown. The 2H and 1T structures used for the analysis can be accessed at http://cmr.fysik.dtu.dk.

Acknowledgments

The authors acknowledge support from the Danish Council for Independent Research's Sapere Aude Program, Grant No. 11-1051390. The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation, Project DNRF58.

References

(1) Mak, K. F.; Lee, C.; Hone, J.; Shan, J.; Heinz, T. F. Atomically Thin MoS_2: A New Direct-Gap Semiconductor. *Phys. Rev. Lett.* **2010**, *105*, 136805.

(2) Splendiani, A.; Sun, L.; Zhang, Y.; Li, T.; Kim, J.; Chim, C.-Y.; Galli, G.; Wang, F. Emerging Photoluminescence in Monolayer MoS_2. *Nano Lett.* **2010**, *10*, 1271–1275.

(3) Zeng, H.; Liu, G.-B.; Dai, J.; Yan, Y.; Zhu, B.; He, R.; Xie, L.; Xu, S.; Chen, X.; Yao, W.; Cui, X. Optical Signature of Symmetry Variations and Spin-valley Coupling in Atomically Thin Tungsten Dichalcogenides. *Sci. Rep.* **2013**, *3*.

(4) Zhang, C.; Johnson, A.; Hsu, C.-L.; Li, L.-J.; Shih, C.-K. Direct Imaging of Band Profile in Single Layer MoS_2 on Graphite: Quasiparticle Energy Gap, Metallic Edge States, and Edge Band Bending. *Nano Lett.* **2014**, *14*, 2443–2447.

(5) Cao, T.; Wang, G.; Han, W.; Ye, H.; Zhu, C.; Shi, J.; Niu, Q.; Tan, P.; Wang, E.; Liu, B.;

Feng, J. Valley-selective Circular Dichroism of Monolayer Molybdenum Disulphide. *Nat. Commun.* **2012**, *3*, 887.

(6) Mak, K. F.; He, K.; Shan, J.; Heinz, T. F. Control of Valley Polarization in Monolayer MoS_2 by Optical Helicity. *Nat. Nanotechnol.* **2012**, *7*, 494–498.

(7) Kaasbjerg, K.; Thygesen, K. S.; Jacobsen, K. W. Phonon-limited Mobility in N-type Single-layer MoS_2 from First Principles. *Phys. Rev. B* **2012**, *85*, 115317.

(8) Kaasbjerg, K.; Thygesen, K. S.; Jauho, A.-P. Acoustic Phonon Limited Mobility in Two-dimensional Semiconductors: Deformation Potential and Piezoelectric Scattering in Monolayer MoS_2 from First Principles. *Phys. Rev. B* **2013**, *87*, 235312.

(9) Cui, X. et al. Multi-Terminal Transport Measurements of MoS_2 Using a van der Waals Heterostructure Device Platform. *Nat. Nanotechnol.* **2015**, *10*, 534–540.

(10) Britnell, L.; Ribeiro, R. M.; Eckmann, A.; Jalil, R.; Belle, B. D.; Mishchenko, A.; Kim, Y.-J.; Gorbachev, R. V.; Georgiou, T.; Morozov, S. V.; Grigorenko, A. N.; Geim, A. K.; Casiraghi, C.; Neto, A. H. C.; Novoselov, K. S. Strong Light-Matter Interactions in Heterostructures of Atomically Thin Films. *Science* **2013**, *340*, 1311–1314.

(11) Bernardi, M.; Palummo, M.; Grossman, J. C. Extraordinary Sunlight Absorption and One Nanometer Thick Photovoltaics Using Two-Dimensional Monolayer Materials. *Nano Lett.* **2013**, *13*, 3664–3670.

(12) Rasmussen, F. A.; Thygesen, K. S. Computational 2D Materials Database: Electronic Structure of Transition-Metal Dichalcogenides and Oxides. *J. Phys. Chem. C* **2015**, *119*, 13169–13183.

(13) Conley, H. J.; Wang, B.; Ziegler, J. I.; Haglund, R. F.; Pantelides, S. T.; Bolotin, K. I.

Bandgap Engineering of Strained Monolayer and Bilayer MoS_2. *Nano Lett.* **2013**, *13*, 3626–3630.

(14) Liu, Q.; Li, L.; Li, Y.; Gao, Z.; Chen, Z.; Lu, J. Tuning Electronic Structure of Bilayer MoS_2 by Vertical Electric Field: A First-Principles Investigation. *J. Phys. Chem. C* **2012**, *116*, 21556–21562.

(15) Rostami, H.; Moghaddam, A. G.; Asgari, R. Effective Lattice Hamiltonian for Monolayer MoS_2: Tailoring Electronic Structure with Perpendicular Electric and Magnetic Fields. *Phys. Rev. B* **2013**, *88*, 085440.

(16) Andersen, K.; Latini, S.; Thygesen, K. S. Dielectric Genome of van der Waals Heterostructures. *Nano Lett.* **2015**, *15*, 4616–4621.

(17) Gao, G.; Gao, W.; Cannuccia, E.; Taha-Tijerina, J.; Balicas, L.; Mathkar, A.; Narayanan, T. N.; Liu, Z.; Gupta, B. K.; Peng, J.; Yin, Y.; Rubio, A.; Ajayan, P. M. Artificially Stacked Atomic Layers: Toward New Van der Waals Solids. *Nano Lett.* **2012**, *12*, 3518–3525.

(18) Geim, A. K.; Grigorieva, I. V. Van der Waals Heterostructures. *Nature* **2013**, *499*, 419–425.

(19) Chhowalla, M.; Shin, H. S.; Eda, G.; Li, L.-J.; Loh, K. P.; Zhang, H. The Chemistry of Two-Dimensional Layered Transition Metal Dichalcogenide Nanosheets. *Nat. Chem.* **2013**, *5*, 263–275.

(20) Zakutayev, A.; Caskey, C. M.; Fioretti, A. N.; Ginley, D. S.; Vidal, J.; Stevanovic, V.; Tea, E.; Lany, S. Defect Tolerant Semiconductors for Solar Energy Conversion. *J. Phys. Chem. Lett.* **2014**, *5*, 1117–1125.

(21) McDonnell, S.; Addou, R.; Buie, C.; Wallace, R. M.; Hinkle, C. L. Defect-Dominated Doping and Contact Resistance in MoS_2. *ACS Nano* **2014**, *8*, 2880–2888.

(22) Qiu, H.; Xu, T.; Wang, Z.; Ren, W.; Nan, H.; Ni, Z.; Chen, Q.; Yuan, S.; Miao, F.; Song, F.; Long, G.; Shi, Y.; Sun, L.; Wang, J.; Wang, X. Hopping Transport Through Defect-Induced Localized States in Molybdenum Disulphide. *Nat. Commun.* **2013**, *4*, 2642.

(23) Komsa, H.-P.; Krasheninnikov, A. V. Native Defects in Bulk and Monolayer MoS_2 From First Principles. *Phys. Rev. B* **2015**, *91*, 125304.

(24) CAMd, Computational Materials Repository. 2015; `https://cmr.fysik.dtu.dk/`.

(25) Enkovaara, J. et al. Electronic Structure Calculations with GPAW: a Real-space Implementation of the Projector Augmented-wave Method. *J. Phys.: Condens. Matter* **2010**, *22*, 253202.

(26) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(27) Gibertini, M.; Marzari, N. Emergence of One-Dimensional Wires of Free Carriers in Transition-Metal-Dichalcogenide Nanostructures. *Nano Lett.* **2015**, *15*, 6229–6238.

{2.7182818284}ϑφερτυθιοπσδφγηξκλ