



New Non-Local XC Functionals and High-Throughput Studies of 2D Materials

PhD Thesis

Asbjørn Rasmussen, July 2021

New Non-local XC Functionals and High-Throughput Studies of 2D Materials

PhD Thesis
August, 2021

By
Asbjørn Rasmussen

Copyright: Reproduction of this publication in whole or in part must include the customary bibliographic citation, including author attribution, report title, etc.

Published by: DTU, Department of Physics, Fysikvej, Building 309, 2800 Kgs. Lyngby Denmark
www.fysik.dtu.dk

Abstract

This thesis introduces a new type of exchange-correlation functional called the *Weighted Local Density Approximation* (WLDA), describes the development of a variety of tools for high-throughput studies, and finally concludes with a high-throughput study of bilayer materials.

It is increasingly important to develop new approximation methods to handle highly correlated materials as well as other systems that are beyond the reach of most commonly used electronic structure methods. WLDA attempts to improve on correlation effects by matching both the energy and the correlation hole of the homogeneous electron gas. This yields a fully non-local functional. Different variants are proposed and several perform comparably to the PBE functional on atomic energies, atomization energies, and lattice constants. Suggestions for future development and improvement of WLDA are also discussed.

When Density Functional Theory (DFT) methods are not sufficient for the task at hand, methods such as many-body perturbation theory must be used. Typically such methods are very computationally expensive which often prohibits the use of these methods in high-throughput projects. In this thesis, one of the most commonly used many-body perturbation methods, namely G_0W_0 , in a high-throughput context by analysing the results of more than 60.000 individual self-energy calculations. Problematic materials that require careful handling in a potential high-throughput study are identified. Errors in the quasiparticle energies stemming from the linear approximation to the quasiparticle equation are investigated and several error-reduction schemes are proposed. The validity of the extrapolation of the plane-wave cutoff to an infinite cutoff and the scissor-operator approximation are also discussed.

As science advances, increasingly complex calculations need to be performed in order to extract the interesting information from the material under investigation. Much of this burden is alleviated by the various electronic structure codes that are available. These typically implement DFT, Quantum Monte Carlo, or Dynamical Mean Field Theory, to name a few. However, many interesting properties have to be calculated by combining or processing the results of an electronic structure calculation, often in non-trivial ways. This thesis discusses the *Atomic Simulation Recipes* (ASR) Python package and contributions made to ASR. ASR standardizes common electronic structure computational tasks by implementing *recipes* to perform such tasks.

Finally, this thesis describes a high-throughput screening project for bilayer materials. Monolayers are extracted from the Computational 2D Materials Database and bilayer structures are generated from a heuristic stacking algorithm. A workflow is implemented that calculates bilayer binding energies and various electronic structure properties of the most interesting candidate bilayers. The candidates are then analysed for *emergent* and for *switchable* properties, i.e. properties that change when going from monolayer to bilayer, and properties that change between different stacking orders, respectively.

Resume

Denne afhandling introducerer en ny type exchange-correlation funktionale ved navn *Weighted Local Density Approximation* (WLDA). Ydermere beskrives udviklingen af en mængde værktøjer til high-throughput studier, og til slut konkluderes med et high-throughput studie af atomart tynde to-lags materialer.

Det er bliver fortsat vigtigere at udvikle nye approksimationsmetoder der kan beskrive stærkt korrelerede materials så vel som andre systemer der er uden for rækkevidde for de mest almindelige metoder indenfor udregninger af elektronisk struktur. WLDA forsøger at forbedre beskrivelse af korrelations effektor ved at matche både energien og korrelationshullet for den homogene gas. Dette resulterer i et fuldstændigt ikke-lokalt funktionale. Forskellige varianter af WLDA foreslås og flere af disse har samme nøjagtighed som PBE for atomare energier, atomiserings energier for molekyler, og for gitterkonstanter. Foreslag til fremtidig udvikling og forbedring af WLDA diskuteres også.

Når Density Functional Theory (DFT) ikke er tilstrækkeligt for det givne problem, bliver det nødvendigt at bruge metoder såsom mange-legeme perturbationsteori. Sådanne metoder er typisk meget udregningsmæssigt dyre, og dette gør det ofte umuligt at anvende metoderne i high-throughput studier. I denne afhandling analyseres en af de mest alment anvendte mange-legeme metoder, navnlig " G_0W_0 ". Ved at analysere mere en 60.000 selv-energi-udregninger identificeres problematiske materiale-typer samt typiske fejlkilder i diverse udregningsmetoder der bruges i forbindelse med G_0W_0 . Resultaterne fra disse analyser udgør et skridt imod fuldt-automatiserede G_0W_0 -udregninger, hvilket er essentielt hvis G_0W_0 skal bruges i et high-throughput studie.

I takt med at videnskaben udvikler sig, er det nødvendigt at foretage udregninger med stigende kompleksitet for at ekstrahere den interessante information for det givne materiale man studerer. Den ekstra kompleksitet udgør en stor byrde når programmet skal udvikles og anvendes, men meget af denne byrde løftes af de forskellige elektronstruktur-koder som er tilgængelige. Typisk implementerer disse koder enkeltstående metoder, men ofte skal resultaterne kombineres på ikke-trivielle måder for at udregne de ønskede egenskaber. Denne afhandling diskuterer Python-pakken Atomic Simulation Recipes (ASR) og diverse bidrag foretaget til ASR. ASR standardiserer en række alment anvendt analyser indenfor elektronstruktur og gør det let at kombinere og udvide sådanne analyser.

Til slut beskrives et high-throughput projekt der omhandler atomart tynde to-lags materialer. Enkelt-lags materialer ekstraheres fra Computational 2D Materials Database og to-lags strukturerne genereres vha. en heuristisk algoritme. Et workflow implementeres som udregner to-lags bindingsenergier og diverse elektronstruktur-egenskaber for de mest interessante to-lags materialer. Disse materialer analyseres derefter for hvilke ændringer der forekommer når et ekstra lag tilføjes, samt hvilke ændringer der forekommer når man sammenligner forskellige måder at stable to lag.

Preface

This thesis is submitted in candidacy for the Ph.D. degree in physics from the Technical University of Denmark. The work was carried out at the section for Computational Atomic-scale Materials Design (CAMD) at the Department of Physics, in the period from August 2018 to August 2021, and supervised by professors Kristian S. Thygesen and Thomas Olsen. This Ph.D. project was funded by the Center for Nanostructure Graphene, Project DNRF103.

Kongens Lyngby, 31st July 2021



Asbjørn Rasmussen

Acknowledgements

I would like to thank Kristian and Thomas for supervising my projects, for suggesting new directions of research, and for teaching me how to do research and about physics.

Thank you, also, to Jens Jørgen. I have learned innumerable things about scientific software development and numerical methods from you. Thank you for your patience and continual willingness to help and advise.

Morten, thank you for working with me, for interesting discussions, and for encouragement when I thought things were going badly. It was a pleasure to have you as a colleague.

Thank you to Ole for keeping things running smoothly, for helping me teach the *erhvervspraktikanter* about Niflheim and computational research, and for letting me use Niflheim for my PhD photo.

Thank you, Bettina, for taking care of all the administrative stuff. You were always positive and happy to help. Having you on the CAMD team was a joy.

Mom, you were always there to support and encourage me, especially when my motivation was low. Thank you also for all the dinners with Signe, the walks, and the talks. Thank you for all the rest - there is too much to list. It means a lot. Thank you, Henrik, for taking an interest in my work and for always being fun to be around. Dad, thank you for both challenging my points of view and supporting my course of action. I always have something to think about after our discussions. Thank you for encouraging me to find the story of my work. Kia, thank you for lively conversations that make me reconsider my standpoints and thank you for always being so warm-hearted. Thank you, Signe, for being the coolest sister and for all the board games, kubb, and music we have played. Thank you, M and M, for always being supportive. Thank you, Carsten and Stine, for helping me believe it would work out even when things were not going well, and for always showing an interest. Thank you to Asger, Malene, Olivia, and Albert. Summers and Christmases at your place is always fun, and was a great way to escape the PhD at times.

Thank you to all the students and postdocs who were at CAMD during my PhD: Alireza, Anders, Ask, Casper, Cuauhtémoc, Daniele, Estefanía, Fabian, Hadeel, Jiban, Joachim, Kåre, Luca, Mads, Mark, Matthew, Mikkel, Morten, Nikolaj, Sahar, Sajid, Sami, Simone, Stefano, Sten, Søren, Thorbjørn, Thorsten, and Urko. If I forgot anyone, I am sorry. Thank you all for the incredibly warm, fun, and funny environment you help create. Thank you for the debates on bird-based supercomputers, spaghetti physics, anti-supervisor defensive constructions, on whether tigers make good PhD students, the merits of breakfast buffet pancakes vs. physics talks, and on the much-needed puppy room. A special thanks to Thorbjørn for working with me on WLDA. Your enthusiasm is unrivaled.

Thank you to the boys for daily shenanigans and for bringing me down to earth. Thank you to Morten Munk for reminding me of physics outside of CAMD. Thank you to *Fysikklubben* for fascinating, if increasingly rare, meetings. Thank you, Estefanía, for

all your support, for extensive proof reading, for making my work sound much cooler,
and for life outside of work.

Thank you to all of you.

Contents

Abstract	iii
Resume	v
Preface	vii
Acknowledgements	ix
1 Introduction	1
2 Theoretical Background	3
2.1 Density Functional Theory	3
2.2 Projector Augmented Wave Method	5
3 The Weighted Local Density Approximation	9
3.1 Background for WLDA	10
3.2 WLDA Theory	13
3.3 HEG Optimizations	29
3.4 WLDA for Spin-Polarized Systems	32
3.5 Potential Regularization	34
3.6 Use of the All-Electron Density	34
3.7 Benchmarks	35
3.8 Perspective: Comparison to other Functionals	41
3.9 Open Questions	44
3.10 Conclusion	45
4 High-Throughput GW	47
4.1 Quasiparticle Theory	47
4.2 Quasiparticle Weight	49
4.3 Hedin's Equations and the GW Approximation	52
4.4 Results of GW Paper	53
5 Automation of Material Simulations	65
5.1 Introduction to ASR	65
5.2 Enhanced Stability Analysis	68
5.3 Effective Masses	78
5.4 Other Contributions to ASR	83
6 High-throughput Screening of Bilayer Materials	87
6.1 Introduction and Motivation	87
6.2 Generating the Bilayers	88
6.3 Bilayer Structure Optimization	89
6.4 Stacking and Relaxation Verification	90
6.5 Workflow	93
6.6 Emergent and Switchable Properties	96

6.7 Conclusion	105
7 Conclusions	107
Bibliography	109
Paper I	117
Paper II	139
Paper III	149
Paper IV	181
A WLDA	203
A.1 Optimizations	203
A.2 All-electron plane-wave convergence	212
A.3 Atomization Results - All Variants	213
B Bilayer Screening	215
B.1 Indirect to Direct Materials	215
B.2 High Effective Mass Changes	216
B.3 Large Switchable Dipole Moment	216
B.4 Switchable Magnetic Order	217

1 Introduction

In the twenty-first century devices that rely on quantum effects are ubiquitous. Such devices can be found in technologies ranging from the transistors in your smartphone to solar panels powering your house to the LEDs in your television to the catalysts in your car engine. Our current scientific understanding of such materials is highly developed and we can calculate catalytic activity or solar panel power efficiency with high accuracy if the appropriate methods are used. Often computational methods are employed to understand or supplement experimental methods when the experiments are too expensive or time-consuming, or to access information that is not directly available in an experimental setup. The most commonly used computational technique for describing solid state devices is *Density Functional Theory* (DFT). DFT is remarkably accurate while still being computationally feasible and which is why it is the most popular method in computational materials science. However, DFT has its shortcomings, in that calculated properties can differ in critical ways from the actual, physical values. For example, the guitar amplifiers that enabled Jimi Hendrix's riffs relied on Germanium transistors [1] and the fact that Germanium is a semiconductor [2]. However, according to DFT, Germanium is a metal [3] and not a semiconductor. So, if DFT were to be believed, Hendrix's concerts would have been a much less interesting affair.

Science is continually improving and adding to the capabilities of state-of-the-art devices and this calls for theoretical and computational techniques that can satisfactorily deal with the ever-increasing complexity of new materials. One option is to develop new approximations for the central quantity in DFT, the *exchange-correlation functional* (XC functional), which contains information about the non-trivial ways in which electrons interact. Current commonly employed approximations to the XC functional suffer from a number of deficits: band gaps are often underestimated, excitations are poorly described, large systems can be hard to tackle, calculations for strongly correlated materials often fail in various ways, and there is no systematic way to improve approximations to the XC functional [4], to name a few. It seems likely that not all the problems of DFT can be solved at the same time; in fact, if it were possible to calculate the exact XC functional efficiently, this would imply efficient solutions to the quantum version of problems in the Non-deterministic Polynomial-time (NP) complexity class which is generally believed to be impossible [5]. In practice we must choose the functional or method appropriate for the task at hand. In chapter 3 of this thesis we describe a new type of non-local XC functional approximation that attempts to solve some of these issues, most notably attempting a better description of how electrons are displaced due to the mutual electrostatic interactions. This is done by matching the so-called *correlation hole*, which is one part of the charge displacement. Hopefully, these new functionals can also inspire new ideas for functional development.

Once accurate tools have been developed, the remaining problem becomes how to find the structure and composition of the next generation of materials. The problem

of designing a material when given a set of desired properties is very hard; it is much easier to calculate the properties of a material rather than designing a material with certain properties, not least because there may be several materials with the same properties. A widely used technique is that of *high-throughput screening* where the properties of a very large amount, sometimes on the order of thousands, of candidate materials are calculated and evaluated. Afterwards the most promising candidates can be extracted and trends in the data can be sought. In the past, high throughput screening has been used, for instance, to discover new atomically thin materials [6, 7] and other two-dimensional materials [8], new solar cell materials [9], materials for water-splitting [10], and even one-dimensional materials [11]. The technique is also widely used outside of materials science, for example in the discovery of new medical compounds [12].

In chapter 4 we discuss many-body perturbation theory calculations, in a particular approximation called the G_0W_0 approximation, in the context of high-throughput studies. We do this by analysing data from the Computational 2D Materials Database (C2DB), which is a database developed at CAMD that currently contains about 4000 atomically thin materials [6]. In chapter 5 we discuss the Atomic Simulation Recipes framework (ASR), a new Python package for performing standard electronic structure tasks on arbitrary materials. ASR is intended to facilitate expert-level calculations across a wide selection of properties without the user necessarily being an expert in all the relevant fields, thus extending the reach of any given researcher substantially. In chapter 6 we detail a high-throughput screening of new bilayer materials that have potential application to atomically thin non-volatile memory devices, controllable magnetic or conductive properties, among other relevant properties.

Before we get to any of that, we start with a general introduction to DFT in chapter 2.

2 Theoretical Background

In this chapter we give a brief introduction to the theoretical topics that are relevant for all the work done in this thesis. This pertains to the method used to describe interacting many-body quantum system, namely Density Functional Theory (DFT), and the numerical methods used to simulate DFT. Methods beyond DFT and more details about exchange-correlation energy approximations will be described in later chapters where it is relevant.

2.1 Density Functional Theory

Density Functional Theory (DFT) was developed as an attempt to provide a tractable solution to the many-body quantum problem. It relies on a crucial theorem developed by Hohenberg and Kohn which states that the Hamiltonian is uniquely determined by the ground state density, up to a constant energy shift [13]. Since the density is easily determined by the Hamiltonian, it follows that there is a one-to-one correspondence between the density and the Hamiltonian and hence every quantity associated that can be determined from the Hamiltonian, such as the wavefunction. This is quite surprising, since the wavefunction seems to contain so much more data than the density, as the wavefunction is a function of N coordinate-vectors but the density is only a function of 1 coordinate-vector. Proofs of the Hohenberg-Kohn theorem are given in many standard textbooks such as [14, 15], so the proof will not be repeated here.

In particular, the ground state energy is a functional of the density, and it can be shown that the ground state density is the minimizer of the energy functional, subject to the constraint that the total number of electrons is unchanged upon variation of the density. Furthermore, the Hohenberg-Kohn theorem can be applied equally well to non-interacting and interacting systems, and it is the "interpretation" of the interacting energy-functional *as the energy-functional of a non-interacting system* that affords a tractable method for describing interacting systems in DFT. This approach is often called Kohn-Sham theory [16].

For a non-interacting system the energy functional may be written as

$$E_s[n] = T_s[n] + \int v_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r}, \quad (2.1)$$

where T_s is the kinetic energy functional, the subscript s denotes "single-particle", and v_{ext} is the external potential. The interacting energy functional can be written as

$$E[n] = T_s[n] + \frac{1}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \int v_{\text{ext}}(\mathbf{r})n(\mathbf{r}) + E_{\text{xc}}[n]. \quad (2.2)$$

The first term is again the kinetic energy; the second term is the so-called *Hartree energy* which is the classical electrostatic energy of an electron gas with the given density; the third term is the energy of the electrons in the external potential; and the

last term is the *exchange-correlation energy*. The exchange-correlation energy contains "all the rest", i.e. energy contributions from the interaction that is not accounted for by the Hartree term. Crucially, it is a universal functional that do not depend on the particular system (here parametrized by v_{ext})[13].

To understand the construction of the Kohn-Sham equations, we consider the variation of the non-interacting energy-functional:

$$\frac{\delta E_s[n]}{\delta n(\mathbf{r})} = \frac{\delta T_s[n]}{\delta n(\mathbf{r})} + v_{\text{ext}}(\mathbf{r}). \quad (2.3)$$

Setting this quantity to zero and constraining the density properly determines the ground state density uniquely

$$\frac{\delta E_s[n]}{\delta n(\mathbf{r})} = 0, \quad (2.4)$$

$$\int n(\mathbf{r}) d\mathbf{r} = N, \quad (2.5)$$

$$\Rightarrow n \text{ is the ground state density.} \quad (2.6)$$

Here N is the total number of electrons. However, we know from elementary quantum mechanics that the ground state density is also uniquely determined by solving the Schrödinger equation:

$$-\frac{1}{2}\nabla^2\psi_i + v_{\text{ext}}\psi_i = E_i\psi_i, \quad (2.7)$$

$$\Rightarrow n(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2. \quad (2.8)$$

In other words, solving the Schrödinger equation and forming the density as above is equivalent to minimizing the energy functional.

Turning back to the interacting energy functional we again take the variation

$$\frac{\delta E[n]}{\delta n(\mathbf{r})} = \frac{\delta T_s[n]}{\delta n(\mathbf{r})} + \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{\text{ext}}(\mathbf{r}) + \frac{\delta E_{\text{xc}}[n]}{\delta n(\mathbf{r})}. \quad (2.9)$$

The ground state density is again uniquely given by $\frac{\delta E[n]}{\delta n(\mathbf{r})} = 0$, but as we also saw, this is equivalent to solving a Schrödinger equation, which for the interacting energy functional is

$$-\frac{1}{2}\nabla^2\psi_i(\mathbf{r}) + (v_{\text{ext}}(\mathbf{r}) + v_H[n](\mathbf{r}) + v_{\text{xc}}[n](\mathbf{r}))\psi_i(\mathbf{r}) = E_i\psi_i(\mathbf{r}), \quad (2.10)$$

which gives a density of

$$n(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2. \quad (2.11)$$

We have introduced two new quantities, the *Hartree potential* and the *exchange-correlation potential*:

$$v_H[n](\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}', \quad (2.12)$$

$$v_{xc}[n](\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}. \quad (2.13)$$

Because the Hamiltonian now depends on the density, we have to solve the equation self-consistently, i.e. start with an initial guess for the density; solve the Schrödinger equation; calculate the new density; and repeat until the density doesn't change.

This approach is called Kohn-Sham theory and it allows us to calculate the ground state density of an interacting system via an auxiliary non-interacting system, which gives a monumental decrease in the computational complexity of the problem.

However, there are a few caveats.

The Kohn-Sham wavefunctions are not physical, i.e. this construction in principle does not give us any information about anything other than the density. In practice, it usually works well to interpret the Kohn-Sham wavefunctions as physical wavefunctions of some quasiparticles.

Another caveat is that E_{xc} is not known analytically, and one could speculate that computing the true expression is probably as computationally complex as solving the full interacting Schrödinger equation, unless there is some hidden redundancies in the current formulation of quantum mechanics. That is to say, the information contained in the "true" quantum wavefunction seems to be so much larger than what can be encoded in a density, so the exact energy functional must somehow extract all this extra information from the density by performing very complex computations that not only depend on the density at each point but also the relationship between the densities at different points. In any case, the development of better approximations to E_{xc} is an important field of research, and we give a brief description of various types of functionals below, because it will be relevant for the development of the Weighted Local Density Approximation, to be described in chapter 3.

2.2 Projector Augmented Wave Method

Implementing density functional theory in a numerical program requires sophisticated methods to ensure that the computations are numerically stable, accurate and not too time consuming.

A major issue is that close to the nuclei wavefunctions high in energy are forced to have very sharp features due to the orthogonality requirement with lower energy wavefunctions. Numerically this poses an issues as it essentially requires a very fine grid, or many fourier-components, to describe such oscillations accurately.

One method to solve this is the *Projector Augmented Wave* method (PAW). The discussion below follows [17] to a large extent.

PAW relies on the following idea: since the true wavefunctions, $\{\psi_i\}$, are highly oscillatory, let us try to construct an alternate basis, $\{\tilde{\psi}_i\}$ that is better behaved numerically. The $\tilde{\psi}$ are called the *smooth wavefunctions*. We assume there exists a linear operator, \mathcal{T} , that maps between the two bases as

$$\psi_i = \mathcal{T}\tilde{\psi}_i. \quad (2.14)$$

The transformed states satisfy as transformed Schrodinger equation

$$H\psi_i = E_i\psi_i \Rightarrow \mathcal{T}^\dagger H\mathcal{T}\tilde{\psi}_i = E_i\mathcal{T}^\dagger\mathcal{T}\tilde{\psi}_i. \quad (2.15)$$

We cannot yet assume anything about the operator \mathcal{T} , so in general it is non-unitary, which means that the Schrodinger equation for $\tilde{\psi}_i$ is a generalized eigenvalue problem.

Next we attempt to construct \mathcal{T} . A core assumption is that \mathcal{T} only corrects the wavefunctions close to the nuclei. Here "close to" means inside a so-called *augmentation sphere*. An important assumption is that the augmentation spheres do not overlap which dramatically simplifies the resulting equations: it turns out that everything can be written as a part operating on the smooth wavefunction plus local, atom-centered corrections. So we write

$$\mathcal{T} = 1 + \sum_a \mathcal{T}_a, \quad (2.16)$$

where the sum over a runs over all atoms and the operators \mathcal{T}_a only have support inside the augmentation sphere for atom a .

To define \mathcal{T}_a we choose a basis, $\{\phi_i^a\}$, so-called *partial waves*, with which to expand the true wavefunctions close to atom a . For physical reasons, a good basis to choose is the atomic eigenfunctions. We also construct a set of smooth partial waves $\{\tilde{\phi}_i^a\}$ and define \mathcal{T}_a implicitly through

$$\phi_i^a = (1 + \mathcal{T}_a)\tilde{\phi}_i^a \quad (2.17)$$

\Rightarrow

$$\mathcal{T}_a\tilde{\phi}_i^a = \phi_i^a - \tilde{\phi}_i^a. \quad (2.18)$$

where we've used the locality of \mathcal{T}_a explicitly, i.e. $\mathcal{T}_{a'}\tilde{\phi}_i^a(\mathbf{r}) = 0$ when \mathbf{r} is inside the augmentation sphere of atom a and $a \neq a'$. Locality of \mathcal{T}_a also implies that

$$\text{Outside of augmentation sphere: } \phi_i^a(\mathbf{r}) = \tilde{\phi}_i^a(\mathbf{r}). \quad (2.19)$$

Inside the augmentation sphere for atom a we can expand $\tilde{\psi}_i$ only in terms of the smooth partial waves for atom a , $\tilde{\phi}_i^a$:

$$\tilde{\psi}_i = \sum_j P_{ij}^a \tilde{\phi}_j^a. \quad (2.20)$$

To find the expansion coefficients we need a basis, $\{\tilde{p}_i^a\}$, that is dual to $\{\tilde{\phi}_i^a\}$ and this dual basis is called the *projectors* which give the name to PAW:

$$\langle p_i^a | \tilde{\phi}_j^b \rangle = \delta_{ij} \delta_{ab}. \quad (2.21)$$

With these we get

$$P_{ij}^a = \langle p_j^a | \tilde{\psi}_i \rangle. \quad (2.22)$$

For this and eq. (2.20) to hold for any $\tilde{\psi}_i$ we must have that the bases $\{\tilde{\phi}_i^a\}$ and $\{\tilde{p}_i^a\}$ are complete, i.e. that

$$\sum_i |\tilde{\phi}_i^a\rangle \langle \tilde{p}_i^a| = 1. \quad (2.23)$$

This, along with eq. (2.18), allows the final definition of \mathcal{T}_a :

$$\mathcal{T}_a = \mathcal{T}_a \sum_i |\tilde{\phi}_i^a\rangle \langle \tilde{p}_i^a| = \sum_i \left(|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle \right) \langle \tilde{p}_i^a|. \quad (2.24)$$

As we noted above, we could now proceed to derive expressions for e.g. operator expectation values in terms of the smooth wavefunctions using eq. (2.24), and one finds that everything reduces to a contribution from the smooth wavefunctions and atom-centered corrections, which by virtue of the above construction *only depend on the smooth wavefunctions*. The PAW constructions in principle maintains the full all-electron description but does in a numerically stable fashion since only smooth wavefunctions are used. This is in contrast to pseudopotential methods which are not all-electron methods. In practice, however, the PAW is not perfect, for example there is a limit to how many projectors can be constructed before the construction procedure becomes unstable. This means the completeness relation, eq. (2.23), is sometimes not fully satisfied which can cause problems for higher-lying states. Sometimes it is also not possible for the partial waves to be fully normalized.

The PAW method is implemented in GPAW [18, 19] which is used for all calculations in this thesis.

3 The Weighted Local Density Approximation

As discussed in chapter 2 the exchange-correlation functional is the central quantity in density functional theory. It contains the description of all the non-trivial interaction effects and the exact exchange-correlation functional is unknown and so must be approximated.

Approximations to the exchange-correlation functional is sometimes described as falling on a "Jacob's ladder" of approximations [20]. In this metaphor "Earth" is Hartree-Fock theory while "Heaven" refers to chemical accuracy. Each rung contains approximations of increasing sophistication and complication.

The first rung consists of the *local density approximation* (LDA) which uses the exchange-correlation functional of a homogeneous electron gas (HEG). The HEG is a model system that can be solved exactly and so one can find an expression for the exchange-correlation functional. This is then used in inhomogeneous systems by replacing the homogeneous density with a spatially varying one $n \rightarrow n(\mathbf{r})$. Clearly this will not be a perfect approximation but one expects that it works well enough for systems where the density is slowly varying. The second rung is the *generalized gradient approximation* (GGA) which is the class of approximations that also include information about the gradient of the density. The most well-known and widely used is the so-called PBE functional [21]. The next rung of *meta generalized gradient approximation* (meta-GGA) contains also the kinetic energy density, τ :

$$\tau_{\sigma} = \sum_{\alpha} \Theta(\mu - \epsilon_{\alpha\sigma}) \frac{1}{2} |\nabla \psi_{\alpha\sigma}(\mathbf{r})|^2 \quad (3.1)$$

The following rungs include additional wavefunction information, including exact exchange, more sophisticated correlation energy expressions, or the random phase approximation plus correlation corrections.

These "rungs" are the standard categories of exchange-correlation functionals and indicate the usual way in which additional information is incorporated into more elaborate approximations.

In this chapter we describe the development of a new functional that uses a weighted density in the LDA functional. This weighted density is in principle fully non-local, so in the context of the Jacob's Ladder metaphor we have jumped from the lowest rung all the way to the top rung of fully non-local exchange-correlation functionals. We call this functional the Weighted Local Density Approximation (WLDA).

We find that by constructing a functional that both matches the energy of the HEG and the correlation hole we can achieve accuracies comparable to PBE on atomic energies, atomization energies, and lattice constants. There are still unresolved issues with

WLDA but these first steps are promising. The work was done together with Thorbjørn Skovhus and is still unpublished.

The motivation and the principle of the design of this functional are described in section 3.1. WLDA is inspired by the work on the renormalized adiabatic LDA kernel in ref. [22]. The theory behind WLDA is explained in section 3.2. We discuss several variants of WLDA and their resulting adiabatic kernels and find that particular variation leads to an especially simple kernel. Section 3.3 deals with optimizing the remaining free parameters of WLDA by matching the correlation hole in the homogeneous electron gas as closely as possible. Section 3.4 deals with the extension to spin-polarized systems where we propose separate ways to deal with the exchange and correlation parts. A potential issue with the potential is highlighted in section 3.5 and we propose a solution. WLDA is defined using the all-electron density but this is not always numerically easy to describe, even though it is in principle accessible in PAW. Section 3.6 investigates convergence properties of atomization energies as a function of plane-wave cutoff for a set of molecular systems. Section 3.7 describes benchmark calculations for WLDA on atomic energies, atomization energies, and lattice constants. We find a performance comparable to PBE for the optimal variant of WLDA. Section 3.8 discusses the methodological differences and similarities between WLDA and other functionals in an attempt to highlight possible future development options. Section 3.9 summarizes the remaining open questions in WLDA and finally section 3.10 concludes the chapter.

3.1 Background for WLDA

Using the adiabatic connection and fluctuation-dissipation theorem it is possible to connect the energy of the fully interacting system with that of a non-interacting system. If we denote the electron-electron interaction by v , then the correlation energy can be written in terms of the response function, χ^λ , for a system where v is replaced by λv [22]:

$$E_c[n] = - \int_0^1 d\lambda \int_0^\infty \frac{d\omega}{2\pi} \int d\mathbf{r} \int d\mathbf{r}' v(\mathbf{r}, \mathbf{r}') \left[\chi^\lambda(\mathbf{r}, \mathbf{r}', i\omega) - \chi^{\text{KS}}(\mathbf{r}, \mathbf{r}', i\omega) \right]. \quad (3.2)$$

This expression connects the non-interacting $\lambda = 0$ system to the fully interacting $\lambda = 1$ where the external potential for all $\lambda < 1$ is chosen to reproduce the fully interacting density.

The response function χ^λ can be expressed in terms of the Kohn-Sham response function, which corresponds to $\chi^{\lambda=0}$, and the *Hartree-exchange-correlation kernel* $f_{\text{Hxc}}^\lambda = \lambda v + f_{\text{xc}}^\lambda$:

$$\chi^\lambda = \chi^{\text{KS}} + \chi^{\text{KS}} f_{\text{Hxc}}^\lambda \chi^\lambda. \quad (3.3)$$

Exchange-correlation kernels are in principle time-dependent quantities and so should be derived from time-dependent DFT but often the adiabatic approximation is used

where the kernel from time-independent DFT is used with trivial time-dependence:

$$f_{a-xc}(\mathbf{r}, \mathbf{r}', t - t') = \frac{\delta^2 E_{xc}[n]}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \delta(t - t'). \quad (3.4)$$

In ref. [22] and later in ref. [23] corrections to the adiabatic LDA kernel (ALDA kernel) that improve upon the description of correlations were studied. This modification of the ALDA kernel was termed the *renormalized ALDA* (rALDA) kernel and was based upon the HEG where accurate parametrizations of the correlation energy is known. The WLDA is based upon this modification: we have attempted to find an energy functional which closely matches the exact kernel in the HEG limit, essentially attempting to solve the functional differential equation eq. (3.4) for E_{xc} , a highly non-trivial task.

To fully motivate the functional form of WLDA we first need to understand rALDA and to that end we want to consider a quantity related E_c namely the coupling constant averaged *correlation hole*. In the following all quantities are those belonging to the HEG. The correlation hole is defined by re-writing the expression for E_c . First we re-express χ^λ by solving eq. (3.3):

$$\chi^\lambda(q, \omega) = \frac{\chi^{\text{KS}}(q, \omega)}{1 - \chi^{\text{KS}}(q, \omega) f_{\text{Hxc}}^\lambda(q, \omega)}. \quad (3.5)$$

Inserting $v(\mathbf{r}, \mathbf{r}') = \frac{e^2}{|\mathbf{r} - \mathbf{r}'|}$ and using the convolution theorem to express the spatial integrals in eq. (3.2) in reciprocal space one finds that for the HEG the correlation energy per electron is

$$\epsilon_c = \frac{E_c}{N} = -\frac{e^2}{\pi^2 n} \int_0^\infty \mathbf{d}q \int_0^1 \mathbf{d}\lambda \int_0^\infty \mathbf{d}\omega \frac{[\chi^{\text{KS}}(q, i\omega)]^2 f_{\text{Hxc}}^\lambda(q, i\omega)}{1 - \chi^{\text{KS}}(q, i\omega) f_{\text{Hxc}}^\lambda(q, i\omega)}, \quad (3.6)$$

where N is the number of electrons and n is the electron density. The coupling-constant average correlation hole, $\bar{g}_c(q)$, is then finally defined via

$$\epsilon_c = \int_0^\infty \frac{\epsilon_c(q)}{2k_F} \mathbf{d}q, \bar{g}_c(q) = \epsilon_c(q) \frac{\pi}{2e^2 k_F n}. \quad (3.7)$$

Using eq. (3.6) we find

$$\bar{g}_c(q) = \frac{1}{2\pi k_F n^2} \int_0^1 \mathbf{d}\lambda \int_0^\infty \mathbf{d}\omega \frac{[\chi^{\text{KS}}(q, i\omega)]^2 f_{\text{Hxc}}^\lambda(q, i\omega)}{1 - \chi^{\text{KS}}(q, i\omega) f_{\text{Hxc}}^\lambda(q, i\omega)}. \quad (3.8)$$

Using the expression for χ^{KS} (also known as the Lindhard function) [24]

$$\chi^{\text{KS}}(q, i\omega) = \frac{mk_F}{2\pi^2} \left(\frac{Q^2 - \tilde{\omega}^2 - 1}{4Q} \ln \left(\frac{\tilde{\omega}^2 + (Q+1)^2}{\tilde{\omega}^2 + (Q-1)^2} \right) - 1 + \tilde{\omega} \arctan \left(\frac{1+Q}{\tilde{\omega}} \right) + \tilde{\omega} \arctan \left(\frac{1-Q}{\tilde{\omega}} \right) \right), \quad (3.9)$$

where

$$Q = \frac{q}{2k_F}, \quad (3.10)$$

$$\tilde{\omega} = \frac{m\omega}{qk_F}, \quad (3.11)$$

we can calculate various how various approximations to the kernel affects the correlation hole.

The exact, RPA ($f_{xc} = 0$), ALDAx ($f_c = 0$), and ALDA correlation holes are shown in figure 3.1 for various values of the density n expressed in terms of the average separation

$$r_s = \left(\frac{3}{4\pi n} \right)^{1/3}. \quad (3.12)$$

The exact hole uses the parametrization from ref. [25]. The main deficit of the ALDA kernels is the long range tail for $q > 2k_F$ where the exact kernel is close to zero. RPA, on the other hand, matches the exact kernel poorly at intermediate q , $k_F \lesssim q \lesssim 3k_F$.

In [22] it was proposed to modify the ALDA Hartree-exchange-correlation kernel so that the match with the exact kernel was better. Specifically, they considered the ALDA Hartree-exchange kernel and applied a hard cutoff of the kernel at $q = 2k_F$ where the ALDAx kernel has an exact zero:

$$f_{\text{Hxc}}^{\text{rALDA}}(q) = \Theta(2k_F - q) f_{\text{Hx}}^{\text{ALDA}}(q). \quad (3.13)$$

From eq. (3.6) it is clear that this will simply set the correlation hole to zero after $q = 2k_F$. In ref. [22] it is shown how this improves the total correlation energy over RPA, and the PGG functional [26] for all densities and the functional of ref. [27], and [28] for all but high densities. In physical terms, because this renormalization brings the kernel close to the exact kernel in the high q limit we expect that short-range correlation effects are better described with rALDA than with ALDA. Indeed, ref. [22] finds that atomization energies for a small set of molecules is better described with rALDA than in ALDA with the mean absolute error relative to experimental atomization energies going from 19 kcal/mol with ALDA to 3.7 kcal/mol with rALDA.

In ref. [29] it was noted in the supplementary material that if we use a weighted density

$$n^*(\mathbf{r}) = \int d\mathbf{r}' \phi(\mathbf{r} - \mathbf{r}') n(\mathbf{r}') \quad (3.14)$$

in an energy functional, the kernel transforms according to

$$f^*[n] = \frac{\delta^2 E[n^*]}{\delta n(\mathbf{r}) n(\mathbf{r}')} = \phi * f[n^*] * \phi, \quad (3.15)$$

where $*$ denotes convolution. The hope was then that if one chooses ϕ judiciously and applies it to the LDA functional in the above manner, one could get a kernel that

approximates rALDA. One benefit would be that the Kohn-Sham response function used in the correlation energy would be defined consistently with the kernel, instead of using a LDA response function with the rALDA kernel, say.

This is the central idea behind the WLDA: can we use a weighted density, n^* , defined similarly to eq. (3.14), such that we get a kernel that closely approximates the exact kernel in HEG limit?

The form of eq. (3.14) is probably a bit too simple as the weight function ϕ should depend on the density, so that we can capture screening effects accurately. For example, for high densities the LDA becomes more accurate so we would want the weight function to be narrow for high densities. Thus we use the more general (and very unspecified) form

$$n^*(\mathbf{r}) = \int d\mathbf{r}' \phi[n](\mathbf{r} - \mathbf{r}')n(\mathbf{r}'). \quad (3.16)$$

When ϕ has a non-trivial dependence on n the transformation of the kernel will no longer be as simple as eq. (3.15) and a large part of the theoretical work on WLDA is extracting form that produces a tractable and useable transformation of the kernel.

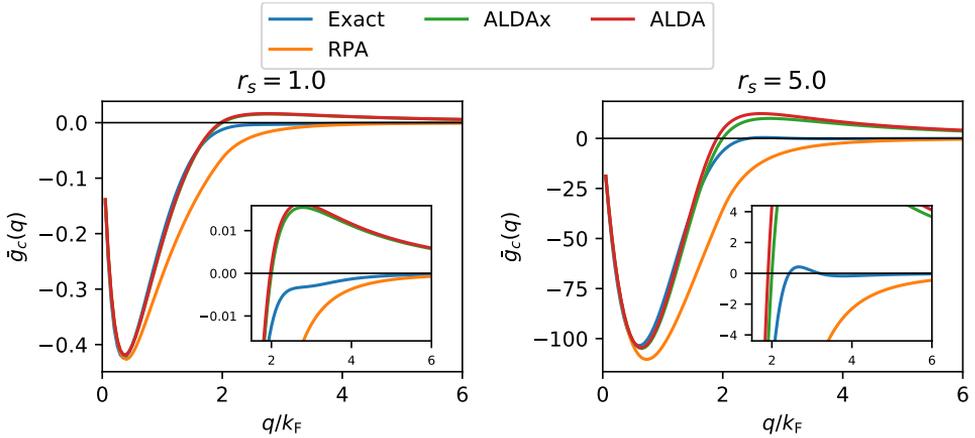


Figure 3.1: Comparisons of the exact, ALDA with exchange only (ALDAx), ALDA, and RPA correlation holes. The exact correlation hole uses the parametrization from [25]. r_s is defined via the density as $r_s = \left(\frac{3}{4\pi n}\right)^{1/3}$.

3.2 WLDA Theory

3.2.1 Narrowing down the functional form of the weighted density

The general form of eq. (3.16) allows for an infinite number of different, potentially non-local ways that ϕ can depend on the density. To narrow it down and maintain

computational tractability we restrict ourselves to forms where ϕ depends on the density at one point. This leads us to consider two forms

$$n^*(\mathbf{r}) = \int d\mathbf{r}' \phi(\mathbf{r} - \mathbf{r}', n(\mathbf{r}))n(\mathbf{r}'), \quad (3.17)$$

and

$$n^*(\mathbf{r}) = \int d\mathbf{r}' \phi(\mathbf{r} - \mathbf{r}', n(\mathbf{r}'))n(\mathbf{r}'). \quad (3.18)$$

The form (3.17) has the benefit that it is a convolution which allows efficient evaluation using the convolution theorem:

$$\mathcal{F} \left[\int f(x-y)g(y)dy \right] (k) = \mathcal{F}[f](k)\mathcal{F}[g](k), \quad (3.19)$$

where \mathcal{F} denotes the Fourier transform. Using this, convolutions can be quickly evaluated by fourier transforming integrands, multiplying the result, and performing the inverse fourier transform to get back to the spatial grid. This has time complexity of $\mathcal{O}(N \log N)$, where N is the number of points in the spatial grid, so it is vastly preferable to direct integration which has time complexity $\mathcal{O}(N^2)$. However, eq. (3.17) has the deficit that the norm of the weighted density is uncontrolled. The idea of WLDA is to use the weighted density in a standard energy functional (specifically LDA) so if we use a density that essentially contains an incorrect number of electrons this may lead to problems.

If we define the weighted density via eq. (3.18) we can fix the normalization of n^* by required that ϕ is normalized to 1:

$$\int n^*(\mathbf{r})d\mathbf{r} = \int d\mathbf{r} \int d\mathbf{r}' \phi(\mathbf{r} - \mathbf{r}', n(\mathbf{r}'))n(\mathbf{r}') = \int d\mathbf{r}' n(\mathbf{r}') \quad (3.20)$$

Eq. (3.18) is not in the form of a convolution so efficient evaluation is not possible in this form. To circumvent this we use a method developed in ref. [30] for van-der-Waals functionals. The method uses what we here call *indicator functions*, f_α . An indicator function, f_α , is defined to be close to 1 whenever its argument is close to α . This is very similar to how indicator functions, also know as *characteristic functions*, are defined in mathematics:

$$f_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases} \quad (3.21)$$

where $x \in X$ for some set X and $A \subset X$ is some subset of X .

The indicators used here are, however, *smooth* versions that go smoothly to zero. For the set of indicators we also require that if we sum over all the indicators we should get 1: $\sum_\alpha f_\alpha(x) = 1, \forall x$. We call the set of α values used to define the indicators the *anchors*. There are many ways to satisfy these requirements but the simplest is

linear interpolators. The easiest way to understand how they are defined is to see an illustration, so consider fig. 3.2. Mathematically, suppose that α_i is the closest anchor to x . Then:

$$f_{\alpha_i}(x) = \begin{cases} \frac{x - \alpha_{i-1}}{\alpha_i - \alpha_{i-1}}, & x < \alpha_i, \\ \frac{\alpha_{i+1} - x}{\alpha_{i+1} - \alpha_i}, & x \geq \alpha_i \end{cases}, \quad (3.22)$$

$$f_{\alpha_{i-1}}(x) = \begin{cases} \frac{\alpha_i - x}{\alpha_i - \alpha_{i-1}}, & x < \alpha_i, \\ 0, & \text{otherwise} \end{cases}, \quad (3.23)$$

$$f_{\alpha_{i+1}}(x) = \begin{cases} \frac{x - \alpha_i}{\alpha_{i+1} - \alpha_i}, & x > \alpha_i, \\ 0, & \text{otherwise} \end{cases}, \quad (3.24)$$

and the rest of the indicators are zero. It is easy to see that this fulfills the two requirements we imposed on a set of indicator functions.

With a given set of indicator functions, we can approximate eq. (3.18) as

$$n^*(\mathbf{r}) \approx \sum_{\alpha} \int d\mathbf{r}' \phi(\mathbf{r} - \mathbf{r}', \alpha) f_{\alpha}(n(\mathbf{r}')) n(\mathbf{r}'). \quad (3.25)$$

Now the weighted density has the form of a convolution between ϕ and $f_{\alpha}(n(\mathbf{r}'))n(\mathbf{r}')$, and we expect it to converge to eq. (3.18) in the limit of an infinitely dense grid of indicators. In practice a few tens of indicators are sufficient.

In summary, the form of the weighted density in eq. (3.18) is both numerically tractable and satisfy the physically relevant criterion of correct normalization. We also note that by construction the weighted density is equal to the physical density in the HEG.

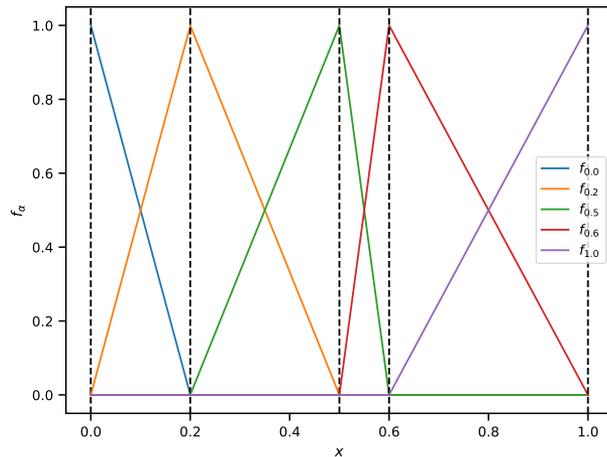


Figure 3.2: An illustration of indicator functions defined via linear interpolation. We use anchors 0.0, 0.2, 0.5, 0.8, and 1.0 and their location is indicated by the vertical dashed lines.

3.2.2 Selecting the form of the energy functional

The first idea is to define the WLDA functional via the LDA functional. Since ref. [22] showed that it is beneficial to truncate the Hartree-exchange-correlation kernel, we seek to define a new Hartree-exchange-correlation energy functional:

$$E_{\text{Hxc}}^{\text{WLDA}}[n] = E_{\text{Hxc}}^{\text{LDA}}[n^*] = \int d\mathbf{r} \int d\mathbf{r}' \epsilon_{\text{Hxc}}(\mathbf{r}, \mathbf{r}', n^*(\mathbf{r})) n^*(\mathbf{r}), \quad (3.26)$$

where

$$\epsilon_{\text{Hxc}}(\mathbf{r}, \mathbf{r}', n(\mathbf{r})) = \frac{n(\mathbf{r}')}{2|\mathbf{r} - \mathbf{r}'|} + \epsilon_{\text{xc}}(n(\mathbf{r}')) \delta(\mathbf{r} - \mathbf{r}'). \quad (3.27)$$

Here ϵ_{xc} refers to the exchange-correlation energy density per electron for LDA.

Since $n^* = n$ in the HEG $E_{\text{WLDA}} = E_{\text{LDA}}$ in the HEG, which is good since LDA is correct in this limit. The question remains whether the WLDA kernel improves on ALDA.

We start by considering the exchange-correlation part. It is straightforward to derive an expression for the kernel for the above functional form:

$$\begin{aligned} f_{\text{xc}}^{\text{WLDA}}[n](\mathbf{r}, \mathbf{r}') &= \frac{\delta^2 E_{\text{LDA}}[n^*]}{\delta n(\mathbf{r}) n(\mathbf{r}')} \\ &= \int d\mathbf{r}'' \left[\frac{\partial^2 \epsilon_{\text{xc}}}{\partial n^2} \Big|_{n=n^*(\mathbf{r}'')} n^*(\mathbf{r}'') \frac{\delta n^*(\mathbf{r}'')}{\delta n(\mathbf{r})} \frac{\delta n^*(\mathbf{r}'')}{\delta n(\mathbf{r}')} \right. \\ &\quad + \frac{\partial \epsilon_{\text{xc}}}{\partial n} \Big|_{n=n^*(\mathbf{r}'')} \left(2 \frac{\delta n^*(\mathbf{r}'')}{\delta n(\mathbf{r})} \frac{\delta n^*(\mathbf{r}'')}{\delta n(\mathbf{r}')} + n^*(\mathbf{r}'') \frac{\delta^2 n^*(\mathbf{r}'')}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \right) \\ &\quad \left. + \epsilon_{\text{xc}}(n^*(\mathbf{r}'')) \frac{\delta^2 n^*(\mathbf{r}'')}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \right] \end{aligned} \quad (3.28)$$

To compare with LDA it turns out to be useful to introduce a measure of the non-trivial change in the weighted density:

$$\begin{aligned} \gamma(\mathbf{r}, \mathbf{r}') &\equiv \frac{\delta n^*(\mathbf{r})}{\delta n(\mathbf{r}')} - \delta(\mathbf{r} - \mathbf{r}') \\ &= \frac{\partial \phi}{\partial n} \Big|_{|\mathbf{r}-\mathbf{r}'|, n(\mathbf{r}')} n(\mathbf{r}') + \phi(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r}')) - \delta(\mathbf{r} - \mathbf{r}'), \end{aligned} \quad (3.29)$$

If $n^* = n$ then $\gamma = 0$ everywhere. With this definition the kernel becomes

$$\begin{aligned}
f_{xc}^{W LDA}[n](\mathbf{r}, \mathbf{r}') &= f_{xc}^{LDA}(|\mathbf{r} - \mathbf{r}'|, n^*(\mathbf{r})) \\
&+ \left(\frac{\partial^2 \epsilon_{xc}}{\partial n^2} \Big|_{n^*(\mathbf{r})} n^*(\mathbf{r}) + 2 \frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n^*(\mathbf{r})} \right) \gamma(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r})) \\
&+ \left(\frac{\partial^2 \epsilon_{xc}}{\partial n^2} \Big|_{n^*(\mathbf{r}')} n^*(\mathbf{r}') + 2 \frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n^*(\mathbf{r}')} \right) \gamma(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r}')) \\
&+ \int d\mathbf{r}'' \left(\frac{\partial^2 \epsilon_{xc}}{\partial n^2} \Big|_{n^*(\mathbf{r}'')} n^*(\mathbf{r}'') + 2 \frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n^*(\mathbf{r}'')} \right) \\
&\quad \times \gamma(|\mathbf{r}'' - \mathbf{r}|, n(\mathbf{r})) \gamma(|\mathbf{r}'' - \mathbf{r}'|, n(\mathbf{r}')) \\
&+ \int d\mathbf{r}'' \left(\frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n^*(\mathbf{r}'')} n^*(\mathbf{r}'') + \epsilon_{xc}(n^*(\mathbf{r}'')) \right) \\
&\quad \times \frac{\partial \gamma}{\partial n} \Big|_{|\mathbf{r}'' - \mathbf{r}|, n(\mathbf{r})} \delta(\mathbf{r} - \mathbf{r}') \quad (3.30)
\end{aligned}$$

We now turn to the Hartree part of the kernel which, with

$$E_H[n] = \int d\mathbf{r} \int d\mathbf{r}' \frac{n(\mathbf{r}')n(\mathbf{r})}{2|\mathbf{r} - \mathbf{r}'|}, \quad (3.31)$$

becomes

$$\begin{aligned}
f_H^{W LDA}[n](\mathbf{r}, \mathbf{r}') &= \iint d\mathbf{r}_1 d\mathbf{r}_2 \left[\frac{\delta n^*(\mathbf{r}_1)}{\delta n(\mathbf{r})} \frac{\delta n^*(\mathbf{r}_2)}{\delta n(\mathbf{r}')} \right. \\
&\quad \left. + n^*(\mathbf{r}_1) \frac{\delta^2 n^*(\mathbf{r}_2)}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \right] \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \\
&= f_H(\mathbf{r} - \mathbf{r}') \\
&+ \iint d\mathbf{r}_1 d\mathbf{r}_2 \left[\gamma(\mathbf{r}_1, \mathbf{r}) \delta(\mathbf{r}_2 - \mathbf{r}') + \delta(\mathbf{r}_1 - \mathbf{r}) \gamma(\mathbf{r}_2, \mathbf{r}') \right. \\
&+ \gamma(\mathbf{r}_1, \mathbf{r}) \gamma(\mathbf{r}_2, \mathbf{r}') \\
&\quad \left. + n^*(\mathbf{r}_1) \frac{\partial \gamma}{\partial n} \Big|_{|\mathbf{r}_2 - \mathbf{r}'|, n(\mathbf{r}')} \delta(\mathbf{r} - \mathbf{r}') \right] \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}. \quad (3.32)
\end{aligned}$$

The question now is, can we choose ϕ or γ such that $f_{HXC}^{W LDA}$ closely approximates the exact kernel in the HEG limit? If we knew other limits exactly, we could attempt to match those as well. In the present case we attempt to match the exact correlation hole of the HEG rather than the kernel as the kernel itself is not known. If we refer back to the formula for the correlation hole, eq. (3.8), one question that we have not considered is whether there is any "gauge" freedom for the kernel, i.e. are there any modifications of the kernel that do not change the correlation hole? Such degrees of freedom, if they exist, could be used to match other exact limits.

When we consider the form of the Hartree and XC kernels it does not seem clear at all how to choose ϕ such that we recover something resembling a truncated version of the ALDA kernel. The rALDA kernel has a very simple form as we saw above:

$$f_{\text{Hx}}^{\text{rALDA}}[n](q) = \Theta(2k_F - q) f_{\text{Hx}}^{\text{ALDA}}(q), \quad (3.33)$$

but the expression in eq. (3.30) contains many terms that are not directly related to the ALDA kernel. For this reason it is hard to see how to get something like the rALDA kernel from eq. (3.30). In an attempt to circumvent this issue we also considered two alternative forms of the energy functional:

$$E_{\text{Hxc}}^{\text{WLDA-1}}[n] = \int d\mathbf{r} \int d\mathbf{r}' \epsilon_{\text{Hxc}}(\mathbf{r}, \mathbf{r}', n(\mathbf{r})) n^*(\mathbf{r}), \quad (3.34)$$

$$E_{\text{Hxc}}^{\text{WLDA-2}}[n] = \int d\mathbf{r} \int d\mathbf{r}' \epsilon_{\text{Hxc}}(\mathbf{r}, \mathbf{r}', n^*(\mathbf{r})) n(\mathbf{r}), \quad (3.35)$$

$$E_{\text{Hxc}}^{\text{WLDA-3}}[n] = \int d\mathbf{r} \int d\mathbf{r}' \epsilon_{\text{Hxc}}(\mathbf{r}, \mathbf{r}', n^*(\mathbf{r})) n^*(\mathbf{r}), \quad (3.36)$$

where we have repeated the original form, here named "WLDA-3".

The kernels for WLDA-1 and WLDA-2 turn out to be as complicated as the WLDA-3 kernel.

We start by noting that the Hartree part of the kernel is identical for WLDA-1 and WLDA-2:

$$\begin{aligned} f_H^{\text{WLDA}_1}[n](\mathbf{r}, \mathbf{r}') &= f_H^{\text{WLDA}_2}[n](\mathbf{r}, \mathbf{r}') \\ &+ n(\mathbf{r}_1) \frac{\delta^2 n^*(\mathbf{r}_2)}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \Big] \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \\ &= f_H(\mathbf{r} - \mathbf{r}') \\ &+ \frac{1}{2} \iint d\mathbf{r}_1 d\mathbf{r}_2 \left[\gamma(\mathbf{r}_1, \mathbf{r}) \delta(\mathbf{r}_2 - \mathbf{r}') + \delta(\mathbf{r}_1 - \mathbf{r}) \gamma(\mathbf{r}_2, \mathbf{r}') \right. \\ &+ n(\mathbf{r}_1) \left. \frac{\partial \gamma}{\partial n} \Big|_{|\mathbf{r}_2 - \mathbf{r}'|, n(\mathbf{r}')} \delta(\mathbf{r} - \mathbf{r}') \right] \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \end{aligned} \quad (3.37)$$

The exchange-correlation kernel for WLDA-1 is given by

$$\begin{aligned} f_{xc}^{\text{WLDA}_1}[n](\mathbf{r}, \mathbf{r}') &= f_{xc}^{\text{LDA}}(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r})) \\ &+ \frac{\partial^2 \epsilon_{xc}}{\partial n^2} \Big|_{n(\mathbf{r})} (n^*(\mathbf{r}) - n(\mathbf{r})) \delta(\mathbf{r} - \mathbf{r}') \\ &+ \frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n(\mathbf{r})} \gamma(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r}')) + \frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n(\mathbf{r}')} \gamma(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r})) \\ &+ \int d\mathbf{r}'' \epsilon_{xc}(n(\mathbf{r}'')) \frac{\partial \gamma}{\partial n} \Big|_{|\mathbf{r}'' - \mathbf{r}|, n(\mathbf{r})} \delta(\mathbf{r} - \mathbf{r}'). \end{aligned} \quad (3.38)$$

For WLDA-2 the exchange-correlation kernel is

$$\begin{aligned}
f_{xc}^{WLDA_2}[n](\mathbf{r}, \mathbf{r}') &= f_{xc}^{LDA}(|\mathbf{r} - \mathbf{r}'|, n^*(\mathbf{r})) \\
&+ \left. \frac{\partial^2 \epsilon_{xc}}{\partial n^2} \right|_{n^*(\mathbf{r})} (n(\mathbf{r}) - n^*(\mathbf{r})) \delta(\mathbf{r} - \mathbf{r}') \\
&+ \left(\left. \frac{\partial^2 \epsilon_{xc}}{\partial n^2} \right|_{n^*(\mathbf{r})} n(\mathbf{r}) + \left. \frac{\partial \epsilon_{xc}}{\partial n} \right|_{n^*(\mathbf{r})} \right) \gamma(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r}')) \\
&+ \left(\left. \frac{\partial^2 \epsilon_{xc}}{\partial n^2} \right|_{n^*(\mathbf{r}')} n(\mathbf{r}') + \left. \frac{\partial \epsilon_{xc}}{\partial n} \right|_{n^*(\mathbf{r}')} \right) \gamma(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r})) \\
&+ \int d\mathbf{r}'' \left. \frac{\partial^2 \epsilon_{xc}}{\partial n^2} \right|_{n^*(\mathbf{r}'')} n(\mathbf{r}'') \gamma(|\mathbf{r}'' - \mathbf{r}|, n(\mathbf{r})) \gamma(|\mathbf{r}'' - \mathbf{r}'|, n(\mathbf{r}')) \\
&+ \int d\mathbf{r}'' \left. \frac{\partial \epsilon_{xc}}{\partial n} \right|_{n^*(\mathbf{r}'')} n(\mathbf{r}'') \left. \frac{\partial \gamma}{\partial n} \right|_{|\mathbf{r}'' - \mathbf{r}|, n(\mathbf{r})} \delta(\mathbf{r} - \mathbf{r}'). \quad (3.39)
\end{aligned}$$

By staring at this long enough one finds that a particular combination of the different functional forms leads to a dramatic simplification. It is most easy to see for the Hartree-parts where

$$\begin{aligned}
&f_{\text{H}}^{\text{WLDA-1}}[n](\mathbf{r}, \mathbf{r}') + f_{\text{H}}^{\text{WLDA-2}}[n](\mathbf{r}, \mathbf{r}') - f_{\text{H}}^{\text{WLDA-3}}[n](\mathbf{r}, \mathbf{r}') \\
&= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \left[\delta(\mathbf{r}_1 - \mathbf{r}) \delta(\mathbf{r}_2 - \mathbf{r}') \right. \\
&\quad - (n^*(\mathbf{r}_1) - n(\mathbf{r}_1)) \left. \frac{\partial \gamma}{\partial n} \right|_{|\mathbf{r}_2 - \mathbf{r}'|, n(\mathbf{r}')} \delta(\mathbf{r} - \mathbf{r}') \\
&\quad \left. - \gamma(\mathbf{r}_1, \mathbf{r}) \gamma(\mathbf{r}_2, \mathbf{r}') \right] f_{\text{H}}(\mathbf{r}_1, \mathbf{r}_2). \quad (3.40)
\end{aligned}$$

In the HEG this expression becomes even simpler since $n^* = n$:

$$f_{\text{H}}^{\text{WLDA}}[n](\mathbf{r}, \mathbf{r}') = \int d\mathbf{r}_1 \int d\mathbf{r}_2 \left[\delta(\mathbf{r}_1 - \mathbf{r}) \delta(\mathbf{r}_2 - \mathbf{r}') - \gamma(\mathbf{r}_1, \mathbf{r}) \gamma(\mathbf{r}_2, \mathbf{r}') \right] f_{\text{H}}(\mathbf{r}_1, \mathbf{r}_2). \quad (3.41)$$

With this expression we hope to have a much simpler time trying to match the exact correlation hole. It still remains to be seen if the exchange-correlation kernel also simplifies with this particular combination:

$$\begin{aligned}
& f_{xc}^{WLDA-1}[n](\mathbf{r}, \mathbf{r}') + f_{xc}^{WLDA-2}[n](\mathbf{r}, \mathbf{r}') - f_{xc}^{WLDA-3}[n](\mathbf{r}, \mathbf{r}') \\
&= f_{xc}^{LDA}(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r})) \\
&- \left(\frac{\partial^2 \epsilon_{xc}}{\partial n^2} \Big|_{n(\mathbf{r})} - \frac{\partial^2 \epsilon_{xc}}{\partial n^2} \Big|_{n^*(\mathbf{r})} \right) (n(\mathbf{r}) - n^*(\mathbf{r})) \delta(\mathbf{r} - \mathbf{r}') \\
&+ \left(\frac{\partial^2 \epsilon_{xc}}{\partial n^2} \Big|_{n^*(\mathbf{r})} (n(\mathbf{r}) - n^*(\mathbf{r})) \right. \\
&+ \left. \frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n(\mathbf{r})} - \frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n^*(\mathbf{r})} \right) \gamma(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r})) \\
&+ \left(\frac{\partial^2 \epsilon_{xc}}{\partial n^2} \Big|_{n^*(\mathbf{r}')} (n(\mathbf{r}') - n^*(\mathbf{r}')) \right. \\
&+ \left. \frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n(\mathbf{r}')} - \frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n^*(\mathbf{r}')} \right) \gamma(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r})) \\
&- \int d\mathbf{r}'' \left(f_{xc}^{LDA}(n^*(\mathbf{r}'')) - \frac{\partial^2 \epsilon_{xc}}{\partial n^2} \Big|_{n^*(\mathbf{r}'')} n(\mathbf{r}'') \right) \\
&\times \gamma(|\mathbf{r}'' - \mathbf{r}|, n(\mathbf{r})) \gamma(|\mathbf{r}'' - \mathbf{r}'|, n(\mathbf{r}')) \\
&+ \int d\mathbf{r}'' \left(\frac{\partial \epsilon_{xc}}{\partial n} \Big|_{n^*(\mathbf{r}'')} (n(\mathbf{r}'') - n^*(\mathbf{r}'')) \right. \\
&+ \left. \epsilon_{xc}(n(\mathbf{r}'')) - \epsilon_{xc}(n^*(\mathbf{r}'')) \right) \times \frac{\partial \gamma}{\partial n} \Big|_{|\mathbf{r}'' - \mathbf{r}|, n(\mathbf{r})} \delta(\mathbf{r} - \mathbf{r}'). \quad (3.42)
\end{aligned}$$

This expression is quite complicated but it simplifies significantly in the HEG limit where, again, $n^* = n$

$$\begin{aligned}
& f_{xc}^{WLDA-1}[n](\mathbf{r} - \mathbf{r}') + f_{xc}^{WLDA-2}[n](\mathbf{r} - \mathbf{r}') - f_{xc}^{WLDA-3}[n](\mathbf{r} - \mathbf{r}') \\
&= f_{xc}^{LDA}(n) \delta(\mathbf{r} - \mathbf{r}') \\
&- \left(f_{xc}^{LDA}(n) - \frac{\partial^2 \epsilon_{xc}}{\partial n^2} n \right) \int d\mathbf{r}'' \gamma(|\mathbf{r}'' - \mathbf{r}|, n), \gamma(|\mathbf{r}'' - \mathbf{r}'|, n) \\
& \quad (3.43)
\end{aligned}$$

If we Fourier transform eqs. (3.41) and (3.43) we get the Hartree-exchange-correlation kernel in the HEG as

$$\begin{aligned}
& f_{Hxc}^{WLDA-1}(n, q) + f_{Hxc}^{WLDA-2}(n, q) - f_{Hxc}^{WLDA-3}(n, q) \\
&= [1 - \gamma^2(q, n)] [f_H(q) + f_{xc}^{LDA}(n)] + \gamma^2(q, n) \frac{\partial^2 \epsilon_{xc}}{\partial n^2} n, \quad (3.44)
\end{aligned}$$

The fact that this is relatively simple and thus tractable leads us to consider the following form of the WLDA energy functional:

$$E^{\text{WLDA}}[n] = E^{\text{WLDA-1}}[n] + E^{\text{WLDA-2}}[n] - E^{\text{WLDA-3}}[n]. \quad (3.45)$$

It remains to be seen whether we can choose a weight function ϕ such that the WLDA correlation hole approximates the exact one. From the form of eq. (3.44) we see that we should probably have something like $1 - \gamma^2(2k_F, n) \approx 0$ such that the first part, which is the ALDA kernel, approximates the rALDA kernel. The influence of the last term remains to be seen.

3.2.3 Choice of weight function

We start by reiterating the required properties of the weight function and listing a few new properties.

1. Normalization The weight function should ensure the normalization of the weighted density is the same as the physical density. This means that the weighting procedure represents a kind of moving around of the electrons.

2. Weight function is positive definite If the weight function is not positive definite we cannot guarantee that the weighted density is positive everywhere. Negative density values would lead to undefined behaviour, at least for the LDA functional where square roots are sometimes used.

3. Match exact correlation hole in long-range limit In the long range, or $q \approx 0$ limit we should match the exact kernel. Since ALDA is exact in this limit, we see from eq. (3.44) that we want

$$\gamma(q, n) \xrightarrow{q \rightarrow 0} 0. \quad (3.46)$$

4. Match exact correlation hole in the short-range limit We also want to match the exact kernel in the short-range or $q \rightarrow \infty$ limit. This means we want

$$1 - \gamma(q, n)^2 \approx 0, \quad q > 2k_F. \quad (3.47)$$

5. Match exact correlation hole everywhere We actually want to match the exact hole *everywhere*. We will select reasonable guesses for the functional form of ϕ using requirements 1-4 and then optimize these guesses by minimizing the difference between the WLDA correlation hole and the exact hole for all q .

We will consider three different weight functions. There is no reason to suppose that this selection is exhaustive but it provides a decent starting point. For all the functional forms we assume that the weight functions are localized in space but there is no *a priori* reason to do limit oneself in this way, although it may make intuitive sense.

Exponential weight function

The first type of function we consider has exponential dependence on the radius

$$\phi(r, n) = c_0 n \exp\left(-c_1 r n^{1/3}\right). \quad (3.48)$$

The prefactor of n and the factor of $n^{1/3}$ in the exponential function are chosen so as to make the parameter c_0 and c_1 dimensionless.

c_0 is determined by the normalization requirement:

$$4\pi \int_0^\infty \mathbf{d}r r^2 \phi(r, n) = \frac{8\pi}{c_1^3} c_0 = 1. \quad (3.49)$$

To investigate the limit $q \rightarrow 1$ where we require $\gamma(q, n) \rightarrow 0$ we first remind ourselves that

$$\begin{aligned} \gamma(\mathbf{r}, \mathbf{r}') &= \frac{\delta n^*(\mathbf{r})}{\delta n(\mathbf{r}')} - \delta(\mathbf{r} - \mathbf{r}') \\ &= \frac{\partial \phi}{\partial n}(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r}')) n(\mathbf{r}') + \phi(|\mathbf{r} - \mathbf{r}'|, n(\mathbf{r}')) - \delta(\mathbf{r} - \mathbf{r}'). \end{aligned} \quad (3.50)$$

For $q = 0$ in the HEG we have

$$\gamma(q = 0, n) = n \int \frac{\partial \phi}{\partial n}(|\mathbf{r}|, n) \mathbf{d}\mathbf{r} + \phi(q = 0, n) - 1. \quad (3.51)$$

Since $\phi(q = 0, n) = 1$ by the normalization requirement we should have that

$$\int \frac{\partial \phi}{\partial n}(|\mathbf{r}|, n) \mathbf{d}\mathbf{r} = 0. \quad (3.52)$$

Inserting the functional form (3.48) we find

$$\begin{aligned} 4\pi \int_0^\infty \mathbf{d}r r^2 \frac{\partial \phi(r, n)}{\partial n} &= \frac{c_1^3}{2} \int_0^\infty \mathbf{d}r r^2 \left(1 - \frac{c_1 r n^{1/3}}{3}\right) e^{-c_1 r n^{1/3}} \\ &= \frac{c_1^3}{2} \left(\frac{2}{c_1^3 n} - \frac{c_1 n^{1/3}}{3} \frac{6}{c_1^4 n^{4/3}}\right) = 0. \end{aligned} \quad (3.53)$$

Thus the $q = 0$ requirement is satisfied as long as the weight function is normalized (for this functional form, at least).

Next we want to consider the truncation of the kernel in reciprocal space in the HEG limit. That is, we want to calculate the prefactor $1 - \gamma^2$ in eq. (3.44). To that end we consider

$$\Phi(r, n) := \gamma(r) + \delta(r) = \frac{\partial \phi(r, n)}{\partial n} n + \phi(r, n) = \left(2 - \frac{c_1 r n^{1/3}}{3}\right) \frac{(c_1 n^{1/3})^3}{8\pi} e^{-c_1 r n^{1/3}}. \quad (3.54)$$

By Fourier transforming we find

$$\begin{aligned}
\Phi(q, n) &= 2\pi \int_0^\pi \sin(\theta) d\theta \int_0^\infty r^2 dr \Phi(r, n) e^{-i \cos(\theta) qr} \\
&= 4\pi \int_0^\infty r^2 dr \Phi(r, n) \frac{\sin(qr)}{qr} \\
&= \left(2 - \frac{1 - \frac{1}{3} [q/(c_1 n^{1/3})]^2}{1 + [q/(c_1 n^{1/3})]^2} \right) \frac{1}{(1 + [q/(c_1 n^{1/3})]^2)^2}. \tag{3.55}
\end{aligned}$$

This results in the following truncation

$$1 - \gamma^2(\tilde{q}) = \Phi(\tilde{q})(2 - \Phi(\tilde{q})), \quad \tilde{q} = \frac{q}{c_1 n^{1/3}}. \tag{3.56}$$

This function is shown in figure 3.3a. By choosing c_1 we can adjust where on the curve $2k_F$ falls, thus determining how strongly the kernel is truncated at $q = 2k_F$. Rather than choosing "sensible" values we will later optimize c_1 by comparing the WLDA correlation hole with the exact hole.

Gaussian weight function

We also consider a Gaussian weight function given by

$$\phi(r, n) = c_0 n e^{-(c_1 r n^{1/3})^2}. \tag{3.57}$$

c_0 is again determined by normalization and it turns out the $q = 0$ limit is satisfied as long as the weight function is normalized (as was also the case for the exponential weight function).

For the truncation function we find

$$\begin{aligned}
\Phi(q, n) &= 4\pi \int_0^\infty r^2 dr \Phi(r, n) \frac{\sin(qr)}{qr} \\
&= \left(1 + \frac{[q/(c_1 n^{1/3})]^2}{6} \right) e^{-\frac{1}{4}[q/(c_1 n^{1/3})]^2} = \left(1 + \frac{\tilde{q}^2}{6} \right) e^{-\tilde{q}^2/4}. \tag{3.58}
\end{aligned}$$

Again we illustrate the resulting truncation function $\Phi(\tilde{q})(2 - \Phi(\tilde{q}))$, in fig. 3.3b.

Lorentzian weight function

The final form for the weight function we consider is the Lorentzian:

$$\phi(r, n) = \frac{c_0 n}{(1 + (c_1 r n^{1/3})^2)^2}. \tag{3.59}$$

c_0 and the $q = 0$ limit are again fixed by normalization.

For the truncation we find

$$\begin{aligned}\Phi(q, n) &= 4\pi \int_0^\infty r^2 \mathbf{d}r \Phi(r, n) \frac{\sin(qr)}{qr} \\ &= \left(1 + \frac{1}{3} \frac{q}{(c_1 n^{1/3})}\right) e^{-q/(c_1 n^{1/3})} = \left(1 + \frac{\tilde{q}}{3}\right) e^{-\tilde{q}}.\end{aligned}\quad (3.60)$$

The resulting truncation function is shown in figure 3.3c.

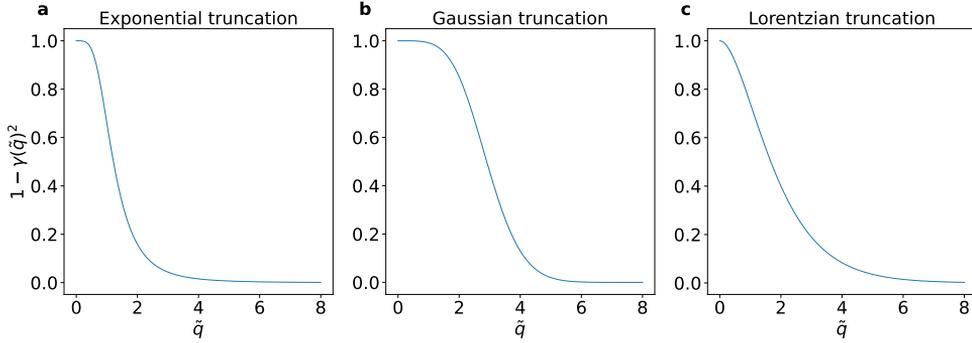


Figure 3.3: The kernel truncation function for the a) exponential weight function, b) Gaussian weight function, and c) Lorentzian weight function.

Choice of weight function - Summary

The weight function is a kind of "free parameter" in WLDA and we have seen above that it is possible to choose various kinds of weight functions that fulfill the necessary requirements but have quite different truncations of the kernel. It remains to be seen which choice is optimal and it is possible that better choices exist but we will not investigate that here.

3.2.4 WLDA as a correction

It is possible to reformulate WLDA slightly. Recall that we can write the WLDA energy as

$$\begin{aligned}E^{\text{WLDA}}[n] &= E^{\text{WLDA-1}}[n] + E^{\text{WLDA-2}}[n] - E^{\text{WLDA-3}}[n] \\ &= \iint \mathbf{d}\mathbf{r} \mathbf{d}\mathbf{r}' [\epsilon_{\text{Hxc}}(\mathbf{r} - \mathbf{r}', n(\mathbf{r}')) n^*(\mathbf{r}) + \epsilon_{\text{Hxc}}(\mathbf{r} - \mathbf{r}', n^*(\mathbf{r}')) (n(\mathbf{r}) - n^*(\mathbf{r}))].\end{aligned}\quad (3.61)$$

By adding and subtracting the LDA energy we can rewrite this as

$$\begin{aligned}E^{\text{WLDA}}[n] &= E^{\text{LDA}}[n] + \Delta E^{\text{WLDA}}[n] \\ &= E^{\text{LDA}}[n] + \iint \mathbf{d}\mathbf{r} \mathbf{d}\mathbf{r}' [\epsilon_{\text{Hxc}}(\mathbf{r} - \mathbf{r}', n^*(\mathbf{r}')) - \epsilon_{\text{Hxc}}(\mathbf{r} - \mathbf{r}', n(\mathbf{r}'))] (n(\mathbf{r}) - n^*(\mathbf{r})).\end{aligned}\quad (3.62)$$

From a numerical standpoint this is convenient for the Hartree-term because it is now written as

$$\Delta E_{\text{H}}^{\text{WLDA}}[n] = -\frac{1}{2} \int \int \mathbf{dr} \mathbf{dr}' \frac{((n(\mathbf{r}) - n^*(\mathbf{r})) (n(\mathbf{r}') - n^*(\mathbf{r}'))}{|\mathbf{r} - \mathbf{r}'|}. \quad (3.63)$$

Since n^* has the same normalization as n , the density different $n - n^*$ is *charge neutral*: $\int \mathbf{dr} (n(\mathbf{r}) - n^*(\mathbf{r})) = 0$. This allows us to evaluate $\Delta E_{\text{H}}^{\text{WLDA}}[n]$ using Fourier transforms by setting the $q = 0$ contribution equal to zero where the Fourier transform of the Coulomb potential has a divergence. The "tricky" part of the Coulomb energy is then included in $E_{\text{Hxc}}^{\text{LDA}}$ which is already implemented in GPAW.

Another reason this is convenient is that it suggests a way to interpret the WLDA energy functional: WLDA is a *correction* to the LDA energy. To understand how the correction works we divide space into regions where $n > n^*$ and regions where $n < n^*$. We then define two quantities that we call *excess localization charge*,

$$\delta n_{\text{excess}}(\mathbf{r}) = (n(\mathbf{r}) - n^*(\mathbf{r}))\Theta(n(\mathbf{r}) - n^*(\mathbf{r})) \quad (3.64)$$

and *localization charge deficit*

$$\delta n_{\text{deficit}}(\mathbf{r}) = (n^*(\mathbf{r}) - n(\mathbf{r}))\Theta(n^*(\mathbf{r}) - n(\mathbf{r})). \quad (3.65)$$

Here $\Theta(x)$ is the Heaviside step-function.

Since $\int \mathbf{dr} (n(\mathbf{r}) - n^*(\mathbf{r})) = \int \mathbf{dr} (\delta n_{\text{excess}}(\mathbf{r}) - \delta n_{\text{deficit}}(\mathbf{r})) = 0$ we have that the amount of excess and deficit charge is equal:

$$\int \mathbf{dr} \delta n_{\text{excess}}(\mathbf{r}) = \int \mathbf{dr} \delta n_{\text{deficit}}(\mathbf{r}). \quad (3.66)$$

In terms of the excess and deficit charges we have

$$\begin{aligned} \Delta E_{\text{Hxc}}^{\text{WLDA}}[n] &= \int \int \mathbf{dr} \mathbf{dr}' [\epsilon_{\text{Hxc}}(\mathbf{r} - \mathbf{r}', n^*(\mathbf{r}')) - \epsilon_{\text{Hxc}}(\mathbf{r} - \mathbf{r}', n(\mathbf{r}'))] \delta n_{\text{excess}}(\mathbf{r}) \\ &+ \int \int \mathbf{dr} \mathbf{dr}' [\epsilon_{\text{Hxc}}(\mathbf{r} - \mathbf{r}', n(\mathbf{r}')) - \epsilon_{\text{Hxc}}(\mathbf{r} - \mathbf{r}', n^*(\mathbf{r}'))] \delta n_{\text{deficit}}(\mathbf{r}). \end{aligned} \quad (3.67)$$

The interpretation of WLDA is then that in regions of excess charge we take a fraction of the physical electron density and change the amount of interaction energy it experiences. We do this by replacing the Hartree-exchange-correlation energy density by that of a less dense gas. In regions of deficit charge we add extra interaction energy. One could hope that these changes reduce the delocalization error of LDA. Because of condition (3.66) we can think of the excess charges as being moved to the deficit regions which lets us think about WLDA in the following way: for the purposes of the energy we effectively use an electron density that is more delocalized than the physical density. Thus, intuitively speaking, a localized density in WLDA will have the same energy as a delocalized density in LDA and one can (again) hope that this means delocalization is less of a problem in WLDA. However, we can't be sure the minimum of

the energy occurs at this localized density, so it is not a given that WLDA will fix the delocalization error.

When formulated as a correction we also see the possible for new functional forms: to remain exact in the HEG limit we need to include $E_{\text{Hxc}}^{\text{LDA}}$, i.e. we need the LDA energy for both the Hartree, the exchange, and the correlation part, but the *corrections* - which go to zero in the HEG - can be applied or not at will. In the following we will consider forms where both Hartree, exchange, and correlation energies are corrected but we also define a new variation where the correlation correction is neglected - mirroring rALDAx - but where both the Hartree and exchange corrections are included. We call this form WLDAx:

$$E_{\text{Hxc}}^{\text{WLDAx}}[n] = E_{\text{Hxc}}^{\text{LDA}}[n] + \Delta E_{\text{Hx}}^{\text{WLDA}}[n]. \quad (3.68)$$

Renormalized WLDA

We now introduce a variant of the WLDA functional based on improving the match to the rALDA kernel. Recall that the WLDA kernel in the HEG is given by

$$f_{\text{Hxc}}^{\text{WLDA}}(n, q) = [1 - \gamma^2(q, n)] [f_{\text{H}}(q) + f_{\text{xc}}^{\text{LDA}}(n)] + \gamma^2(q, n) \frac{\partial^2 \epsilon_{\text{xc}}}{\partial n^2} n. \quad (3.69)$$

It is clear that because we required $\gamma \rightarrow 1$ as $q \rightarrow \infty$ we will have a leftover term for large q :

$$f_{\text{Hxc}}^{\text{WLDA}}(n, q) \xrightarrow{q \rightarrow \infty} \frac{\partial^2 \epsilon_{\text{xc}}}{\partial n^2} n. \quad (3.70)$$

If we for a moment focus on the exchange part, throughout using that $E_{\text{Hxc}}^{\text{WLDA}} = E_{\text{H}}^{\text{WLDA}} + E_{\text{x}}^{\text{WLDA}} + E_{\text{c}}^{\text{WLDA}}$, we see first that with $\epsilon_{\text{x}} = -C_{\text{x}} n^{1/3}$

$$\frac{\partial^2 \epsilon_{\text{x}}}{\partial n^2} n = \frac{2}{9} C_{\text{x}} n^{-2/3}. \quad (3.71)$$

The exchange part of the WLDA kernel includes a term that is similar to this term and it is given by eq. (3.43), which we restate here for convenience:

$$f_{\text{x}}^{\text{WLDA}}(n, q) = f_{\text{x}}^{\text{LDA}}(n) - \left(f_{\text{x}}^{\text{LDA}}(n) - \frac{\partial^2 \epsilon_{\text{x}}}{\partial n^2} n \right) \gamma(q, n)^2. \quad (3.72)$$

Now,

$$f_{\text{x}}^{\text{LDA}}(n) = \frac{\partial^2 \epsilon_{\text{x}}}{\partial n^2} n + 2 \frac{\partial \epsilon_{\text{x}}}{\partial n} = -\frac{4C_{\text{x}}}{9} n^{-2/3}, \quad (3.73)$$

$$f_{\text{x}}^{\text{LDA}}(n) - \frac{\partial^2 \epsilon_{\text{x}}}{\partial n^2} n = 2 \frac{\partial \epsilon_{\text{x}}}{\partial n} = -\frac{2C_{\text{x}}}{3} n^{-2/3} = \frac{3}{2} f_{\text{x}}^{\text{LDA}}(n). \quad (3.74)$$

Thus

$$f_{\text{x}}^{\text{WLDA}}(n, q) = f_{\text{x}}^{\text{LDA}}(n) - \frac{3}{2} f_{\text{x}}^{\text{LDA}}(n) \gamma(q, n)^2. \quad (3.75)$$

One way to get rid of the extra term in the $q \rightarrow \infty$ limit is then by using the following functional:

$$E_{\text{Hxc}}[n] = E_{\text{Hxc}}^{\text{LDA}}[n] + \Delta E_{\text{H}}^{\text{WLDA}}[n] + \frac{2}{3} \Delta E_{\text{x}}^{\text{WLDA}}[n] + E_{\text{c}}^{\text{WLDA}}[n]. \quad (3.76)$$

Here and below, quantities without a superscript such as "WLDA" are to be understood as *preliminary* quantities, i.e. quantities that we use to guide the discussion but which are not final.

The above form leads to the kernel correction

$$\Delta f_{\text{x}}(n, q) = -f_{\text{x}}^{\text{LDA}}(n) \gamma(q, n)^2, \quad (3.77)$$

so that the full kernel can be written

$$\begin{aligned} f_{\text{Hxc}}(n, q) &= f_{\text{Hxc}}^{\text{LDA}}(n, q) + \Delta f_{\text{H}}^{\text{WLDA}}(n, q) + \Delta f_{\text{x}}^{\text{WLDA}}(n, q) \\ &= [1 - \gamma(q, n)^2] [f_{\text{H}}(q) + f_{\text{x}}^{\text{LDA}}(n)] + f_{\text{c}}^{\text{LDA}}(n). \end{aligned} \quad (3.78)$$

In this form the WLDA functional leads to a kernel that is very similar to rALDA. We call this functional the *renormalized WLDA with exchange only* functional or rWLDAx:

$$E_{\text{Hxc}}^{\text{rWLDAx}}[n] = E_{\text{Hxc}}^{\text{LDA}}[n] + \Delta E_{\text{H}}^{\text{WLDA}}[n] + \frac{2}{3} \Delta E_{\text{x}}^{\text{WLDA}}[n]. \quad (3.79)$$

We could also include the correlation correction, e.g. as

$$E_{\text{Hxc}}[n] = E_{\text{Hxc}}^{\text{LDA}}[n] + \Delta E_{\text{H}}^{\text{WLDA}}[n] + \frac{2}{3} \Delta E_{\text{x}}^{\text{WLDA}}[n] + \Delta E_{\text{c}}^{\text{WLDA}}[n], \quad (3.80)$$

which yields the kernel

$$f_{\text{Hxc}}(n, q) = [1 - \gamma(q, n)^2] [f_{\text{H}}(q) + f_{\text{x}}^{\text{LDA}}(n) + f_{\text{c}}^{\text{LDA}}(n)] + \gamma(q, n)^2 \frac{\partial^2 \epsilon_{\text{c}}}{\partial n^2} n. \quad (3.81)$$

Clearly, we can ask the same question as before: Is it possible to multiply the correlation correction by a prefactor such that the last term disappears? Consider the correction to the part of the kernel coming from the correlation energy:

$$\Delta f_{\text{c}}^{\text{WLDA}}(n, q) = - \left(f_{\text{c}}^{\text{LDA}}(n) - \frac{\partial^2 \epsilon_{\text{c}}}{\partial n^2} n \right) \gamma(q, n)^2. \quad (3.82)$$

As before we would seek a prefactor, ρ , such that

$$\rho \left(f_{\text{c}}^{\text{LDA}}(n) - \frac{\partial^2 \epsilon_{\text{c}}}{\partial n^2} n \right) = f_{\text{c}}^{\text{LDA}}(n). \quad (3.83)$$

However, this is not possible as we see in figure 3.4 where we plot ρ as a function of density for various spin-polarizations ζ ; ρ depends on the density n .

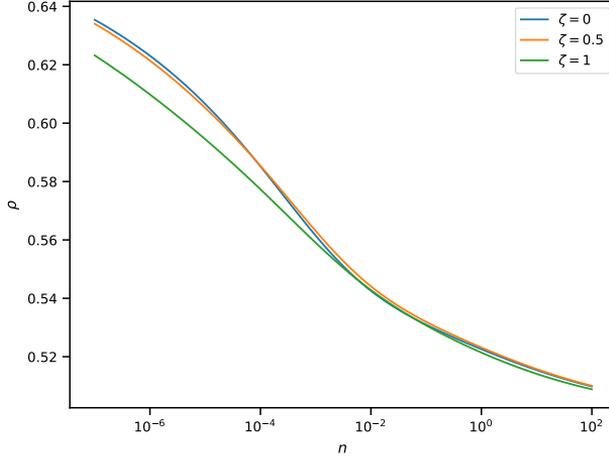


Figure 3.4: The factor ρ (see main text) as a function of density n for various spin-polarizations.

It may be possible to introduce some additional modification of the correlation energy to get a n -dependent ρ so as to exactly cancel the extra term in the kernel but we have not pursued that. To keep things simply we define a renormalized WLDA with correlation energy as

$$E_{\text{Hxc}}^{\text{rWLDA}}[n] = E_{\text{Hxc}}^{\text{LDA}}[n] + \Delta E_{\text{H}}^{\text{WLDA}}[n] + \frac{2}{3}\Delta E_{\text{xc}}^{\text{WLDA}}[n]. \quad (3.84)$$

The factor of $2/3$ is included in front of both the exchange part and correlation part. This gives full cancellation for the extra exchange part but only partial for the correlation part.

Fractionalized WLDA

The final variant of WLDA is based on the observation that the ALDAx correlation hole has a zero at $q = 2k_F$, contrary to the exact correlation hole as we see from figure 3.1. Because the rWLDA kernel is approximately equal to a truncated version of the ALDA kernel (up to a small correction from the correlation energy), we expect the rWLDA kernel to also have a zero around $q = 2k_F$. This will be a problem because the correlation hole will be overestimated for large q . A way to fix this is to subtract something from the kernel. In an admittedly *ad hoc* fashion we define the *fractionalized WLDA* (fWLDA) functional:

$$E_{\text{Hxc}}^{\text{fWLDA}}[n] = E_{\text{Hxc}}^{\text{LDA}}[n] + \frac{2}{3}\Delta E_{\text{H}}^{\text{WLDA}}[n] + \frac{2}{3}\Delta E_{\text{xc}}^{\text{WLDA}}[n]. \quad (3.85)$$

The factor of $2/3$ in front of the Hartree energy is arbitrary and simply chosen to keep things simple. We also define the exchange-only variant (fWLDAx):

$$E_{\text{Hxc}}^{\text{fWLDAx}}[n] = E_{\text{Hxc}}^{\text{LDA}}[n] + \frac{2}{3}\Delta E_{\text{H}}^{\text{WLDA}}[n] + \frac{2}{3}\Delta E_{\text{x}}^{\text{WLDA}}[n]. \quad (3.86)$$

Flavors of WLDA - Summary

In summary we have defined multiple variants of the WLDA functional, which are shown in table 3.1.

Variant	Energy functional
WLDA	$E_{xc}^{LDA}[n] + \Delta E_{Hxc}^{WLDA}[n]$
WLDAX	$E_{xc}^{LDA}[n] + \Delta E_{Hx}^{WLDA}[n]$
rWLDA	$E_{xc}^{LDA}[n] + \Delta E_H^{WLDA}[n] + \frac{2}{3} \Delta E_{xc}^{WLDA}[n]$
rWLDAX	$E_{xc}^{LDA}[n] + \Delta E_H^{WLDA}[n] + \frac{2}{3} \Delta E_x^{WLDA}[n]$
fWLDA	$E_{xc}^{LDA}[n] + \frac{2}{3} \Delta E_{Hxc}^{WLDA}[n]$
fWLDAX	$E_{xc}^{LDA}[n] + \frac{2}{3} \Delta E_{Hx}^{WLDA}[n]$

Table 3.1: The different variants of the WLDA functional with corresponding energy functional.

In the next section we will calculate correlation holes for these variants and compare to the exact correlation hole. This will let us optimize three things: 1) The type of WLDA functional, 2) The choice of weight function (recall that we considered a Gaussian, exponential, or Lorentzian weight function), and 3) The c_1 parameter of the weight function (which is related to the width of the weight function).

3.3 HEG Optimizations

In this section we will calculate the correlation holes for the different flavors of WLDA and optimize the c_1 parameter for the different weight function types by minimizing the integrated difference between the WLDA and exact correlation holes.

3.3.1 WLDA and WLDAX

The first flavors we consider is WLDA and WLDAX. We start by calculating the correlation hole, using eq. (3.8), for the exponential weight function for various different values of c_1 and r_s . To determine the optimal choice of parameters we calculate the normalized mean absolute error (NMAE):

$$\text{NMAE}(c_1, r_s) = \frac{\int dq |g_c(q, c_1, r_s, \text{WLDA}) - g_c(q, r_s, \text{exact})|}{\int dq |g_c(q, r_s, \text{exact})|}, \quad (3.87)$$

where $g_c(q, r_s, c_1, \text{WLDA})$ denotes the correlation hole as a function of wavevector q , density $n = \frac{3}{4\pi r_s^3}$, and c_1 value.

The NMAE as a function of $1/c_1$ are shown in figure 3.5. The small figures show the NMAEs for various r_s values. We plot the error as a function of $1/c_1$ so we can see the convergence to ALDA which happens for $1/c_1 = 0$ (where the weight function is a Dirac delta). The dashed lines show the location of the c_1 value that gives the lowest

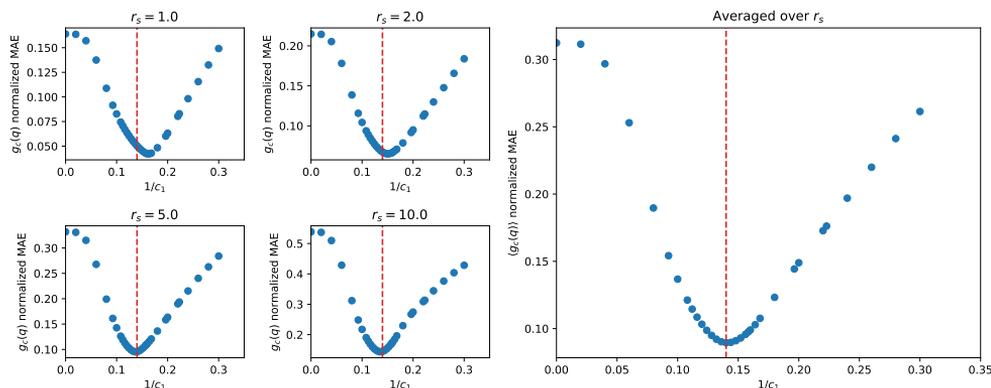


Figure 3.5: The normalized mean absolute error (NMAE, defined in main text) between the WLDA functional using the exponential weight function for various c_1 values. The right plot shows the error averaged over all r_s . The vertical dashed, red lines shown the location of the minimum of the averaged errors in all cases. We see that the optimal value of the average error is close to the minima in all cases. The optimal c_1 value is 7.14286.

average NMAE. For WLDA with exponential weight function this value is close to the optimal c_1 value for all densities with the biggest deviations for high densities.

The next steps are to optimize the Gaussian and Lorentzian WLDA kernels as well. Finally, we repeat the whole procedure using the exchange-only variant (WLDax). The results are shown in appendix A.

For the WLDA functional the best performing weight function is the Lorentzian with an average NMAE of 0.058. For WLDax it is again the Lorentzian weight function with a NMAE of 0.070. The correlation holes for the Lorentzian for these values of c_1 are shown in figure 3.6. As we can see the optimized functionals match the exact correlation hole quite well.

3.3.2 rWLDA and rWLDax

We now repeat this procedure for the rWLDA(x) functional. The optimization results are shown in the appendix. The optimal weight function turns out to be the Gaussian for both rWLDA and rWLDax with a c_1 of 2.08333 and NMAE 0.081950, and $c_1 = 2.18023$ and with NMAE 0.131995 respectively. The correlation holes are shown in figure 3.7.

3.3.3 fWLDA and fWLDax

Finally, we consider fractionalized WLDA(x). The optimal for fWLDA turns out to be the exponential weight function with $c_1 = 5.09393$ and NMAE 0.035367. For fWLDax it is the Gaussian weight function, with $c_1 = 1.83824$ and NMAE 0.0439024. The correlation holes are shown in figure 3.8.

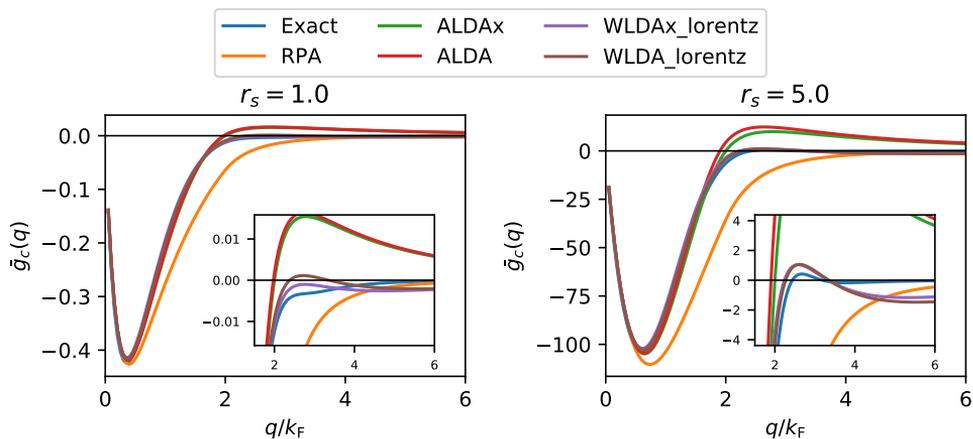


Figure 3.6: The correlation holes for the optimized version of the Lorentzian weight function for WLDA and WLDAx.

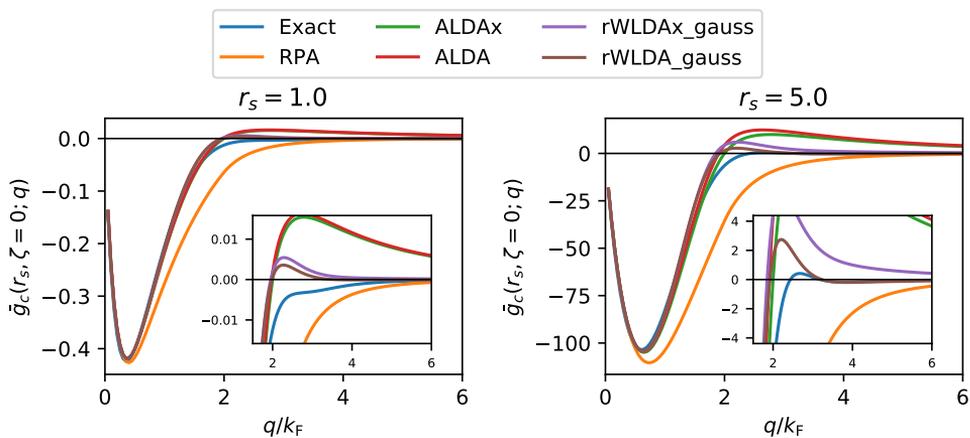


Figure 3.7: The correlation holes for the optimized version of the Gaussian weight function for rWLDA and rWLDAx.

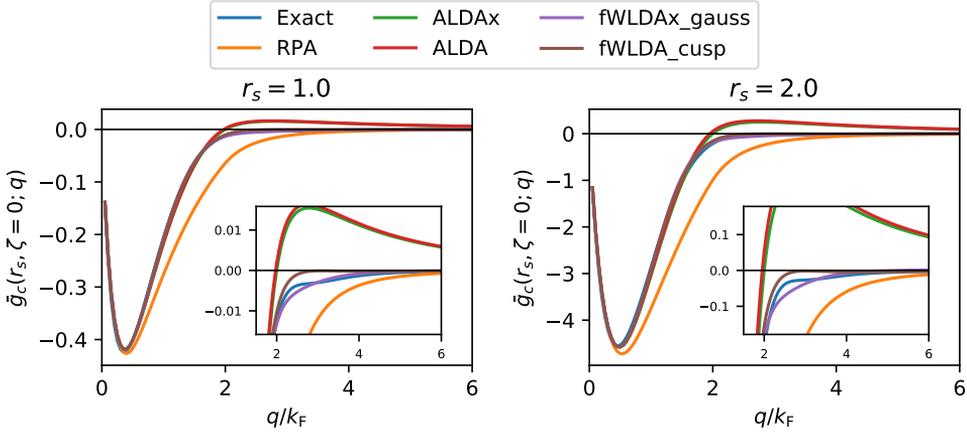


Figure 3.8: The correlation holes for the optimized version of the exponential (for fWLDA) and Gaussian (for fWLDax) weight functionx.

3.3.4 Summary: c_1 Optimization

We have performed optimizations of the c_1 parameter for all possible combinations of the functional types WLDA, WLDax, rWLDA, rWLDax, fWLDA, fWLDax with the weight functions of exponential, Gaussian, and Lorentzian types. In table 3.2 we tabulate the results. It turns out that for both exchange-only and exchange+correlation fractionalized WLDA is best. Overall fWLDA with the exponential weight function performs best of functionals with exchange+correlation corrections and fWLDA with the Gaussian weight function performs best of the exchange-only functionals.

	Exp. c_1	Exp. NMAE	Gau. c_1	Gau. NMAE	Lor. c_1	Lor. NMAE
WLDA	7.14286	0.089473	3.02419	0.128982	5.5556	0.058134
WLDax	6.09756	0.0725738	2.60417	0.10648	4.6875	0.070153
rWLDA	5.0	0.104596	2.08333	0.081950	4.28571	0.157958
rWLDax	5.43478	0.1560827	2.18023	0.131995	5.17241	0.202066
fWLDA	5.09393	0.035367	2.08333	0.081950	3.48837	0.046904
fWLDax	4.09836	0.0520419	1.83824	0.0439024	2.77778	0.0867764

Table 3.2: Optimized values of the c_1 parameter and the corresponding normalized mean absolute error (NMAE) for the various functionals and weight functions. To optimal values across those functionals using both exchange and correlation, and across those using just exchange are highlighted in bold.

3.4 WLDA for Spin-Polarized Systems

For magnetic systems it is essential to perform a spin-polarized calculation to obtain accurate results. The above optimizations and derivations were all done in the spin-paired HEG so the question remains how to treat spin-polarized systems.

For the exchange part of the energy we can use the relation

$$E_x[n_\uparrow, n_\downarrow] = \frac{E_x[2n_\uparrow] + E_x[2n_\downarrow]}{2}, \quad (3.88)$$

which holds for the exact exchange. We use this to define the WLDA exchange correction as

$$\Delta E_x^{\text{WLDA}}[n_\uparrow, n_\downarrow] = \frac{E_x^{\text{WLDA}}[2n_\uparrow] + E_x^{\text{WLDA}}[2n_\downarrow]}{2}. \quad (3.89)$$

It is not so clear whether the Hartree correction can be interpreted as an exchange or a correlation energy, or even a mix. We will consider two different ways of calculating the Hartree correction in spin-polarized systems: Hartree-As-eXchange (HAX) or spin-neutral Hartree correction. The HAX correction is defined similar to the exchange correction in eq. (3.89). The spin-neutral Hartree correction is calculated using the total electron density:

$$\Delta E_H^{\text{WLDA}}[n_\uparrow, n_\downarrow] = \Delta E_H^{\text{WLDA}}[n_\uparrow + n_\downarrow] \quad (3.90)$$

where the right-hand-side is defined as in eq. (3.63), i.e. exactly as in a spin-paired system.

The correlation correction cannot be treated as simply. We need to consider how the weighted density should be defined and there are probably many ways to do this. One requirement is that the results reduce to the spin-paired case when $n_\uparrow = n_\downarrow$. To make some relatively reasonable choice we fix the total weighted density to be

$$n_\uparrow^*(\mathbf{r}) + n_\downarrow^*(\mathbf{r}) = n^*(\mathbf{r}) = \int d\mathbf{r}' \phi(\mathbf{r} - \mathbf{r}', n(\mathbf{r}'))n(\mathbf{r}'), \quad (3.91)$$

and we require that the unit-cell averaged magnetization is the same for the weighted density and the physical density:

$$\int_{\Omega} d\mathbf{r} (n_\uparrow^*(\mathbf{r}) - n_\downarrow^*(\mathbf{r})) = \int_{\Omega} d\mathbf{r} (n_\uparrow(\mathbf{r}) - n_\downarrow(\mathbf{r})). \quad (3.92)$$

These requirements can be fulfilled by defining

$$n_\sigma^*(\mathbf{r}) = \int d\mathbf{r}' \phi(\mathbf{r} - \mathbf{r}', n(\mathbf{r}'))n_\sigma(\mathbf{r}'). \quad (3.93)$$

With this the correlation correction becomes

$$\begin{aligned} \Delta E_c^{\text{WLDA}}[n_\uparrow, n_\downarrow] = \int d\mathbf{r} [& \epsilon_c(n_\uparrow^*(\mathbf{r}), n_\downarrow^*(\mathbf{r}))(n(\mathbf{r}) - n^*(\mathbf{r})) \\ & - \epsilon_c(n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r}))(n(\mathbf{r}) - n^*(\mathbf{r}))]. \end{aligned} \quad (3.94)$$

3.5 Potential Regularization

We discovered during our studies that there is a term in the potential that is not completely regular, i.e. it seems to grow without bound as one goes into vacuum. This phenomenon is most easily seen if we consider the part of the exchange correction to the potential that comes from "WLDA-1":

$$\begin{aligned} v_x^{\text{WLDA-1}}[n](\mathbf{r}) &= \frac{\delta E_x^{\text{WLDA-1}}}{\delta n(\mathbf{r})} = \frac{\delta}{\delta n(\mathbf{r})} \int d\mathbf{r}' \epsilon_x(n(\mathbf{r}')) n^*(\mathbf{r}') \\ &= \int d\mathbf{r}' \epsilon_x(n(\mathbf{r}')) \frac{\delta n^*(\mathbf{r}')}{\delta n(\mathbf{r})} + \partial_n \epsilon_x(n(\mathbf{r})) n^*(\mathbf{r}) \end{aligned} \quad (3.95)$$

Of interest here is the term

$$\partial_n \epsilon_x(n(\mathbf{r}')) n^*(\mathbf{r}') \sim \partial_n (n)^{1/3} n^* \sim \frac{n^*}{n^{2/3}}. \quad (3.96)$$

The question is then how $n^*/n^{2/3}$ behaves in low density regions. Intuitively, the weighted density at a point \mathbf{r} is given by how much density is received from other points \mathbf{r}' . The range of how far density is "sent" from a point is given by the density at that point. Thus in a low density region the weighted density will receive contributions from a large region around it. We should therefore worry that $n^*/n^{2/3}$ will grow without bound as we go into vacuum regions. This does indeed seem to be the case from studies of the weighted density on a Hydrogen 1s state and from self-consistent calculations on various atomic systems.

This divergence is a fundamental problem with WLDA and strictly speaking probably requires a redefinition of the functional form that avoids the vacuum divergence. However, due to time constraints it was not possible to do this.

Instead we use a more or less *ad hoc* regularization. We simply cutoff the divergent term when the density is sufficiently small. On tests on small atoms and molecules the exact value for the cutoff was not important. We also attempted another regularization scheme which produced similar results. Generally, the regularization does not seem to be necessary always but it does greatly improve convergence rate and in some cases it is impossible to converge without regularization. Altogether this seems to imply that the exact form of the regularization is not important, as one would hope, but, in my opinion, the divergent term remains a fundamental problem for this version of WLDA.

A solution to the divergence could be to redefine the weight function such that n^*/n does not diverge. Alternatively one could try to use the WLDA-2 or WLDA-3 forms exclusively as these forms do not suffer from the divergence. How well these would perform remain a question for future work.

3.6 Use of the All-Electron Density

WLDA is defined using the all-electron density. In most electronic structure codes this is not directly available or it is hard to converge with respect to basis set size. GPAW is a PAW code which means most quantities are calculated as "smooth part + atom-centered corrections" (see sec. 2.2). Using PAW corrections is equivalent to working

with the all-electron quantities and it renders the numerics stable. However, it is only possible to define PAW corrections for semi-local quantities; non-local quantities will in general involve terms that couple different atoms. There is therefore no way to write PAW corrections for WLDA, at least in the standard formulation. It may be possible to include the extra coupling terms in some expanded PAW formalism but whether it can be made tractable remains to be seen. Not all is lost as we can use the PAW scheme for the LDA part of WLDA and do something else for the correction part.

For the corrections we can either use the PAW pseudo-density or we can reconstruct the all-electron density from the pseudo density + suitable PAW corrections. As mentioned above, it can be hard to converge the all-electron density but whether or not it will be an issue for WLDA is not certain *a priori*. To determine converge with respect to plane-wave cutoff we calculate atomization energies for several small molecules with the fWLDAX variant using the all-electron density for various plane-wave cutoffs. The results are shown in fig. A.19 in sec. A.2. With the exception of H₂ the atomization energies are not converged even at ~ 1500 eV plane-wave cutoff so at this stage we conclude that it is not practical to use the all-electron density. The following results are therefore calculated using the pseudo density for the corrections.

3.7 Benchmarks

3.7.1 Atomic Energies

We have implemented the various WLDA variants in the atomic module of GPAW. This module can do DFT calculations for atoms using a spherically averaged density. Utilizing the spherical symmetry allows for fast evaluation of the various quantities involved. In this section we describe the performance of the WLDA variants on atomic systems. We compare atomic ground state energies for WLDA, LDA, and PBE with the best-estimate reference energies from refs. [31, 32]. The references are calculated from experimental energies plus a relativistic "deperturbation" scheme. The calculations were done using a spherically average density, on a radial grid with 2000 grid points and a cutoff radius of 50 Bohr radii, and using 50 Gaussians to form a final basis set of approx. 20-25 linear combinations of Gaussians.

The performance results for the optimal version of the exchange-only functionals, i.e. using the Lorentzian weight function for WLDA, and the Gaussian for rWLDAX and fWLDAX are shown in figure 3.9 and the mean absolute error (MAE) and mean absolute relative error is shown in table 3.3. With the exception of WLDA@Lorentzian the results all improve on LDA and are even comparable to PBE.

Next we will consider the variants that include the correlation energy correction. The results for these are shown in figure 3.10 and the errors are tabulated in table 3.4. Again, with the exception of WLDA with the Lorentzian weight function, WLDA performs comparably to PBE.

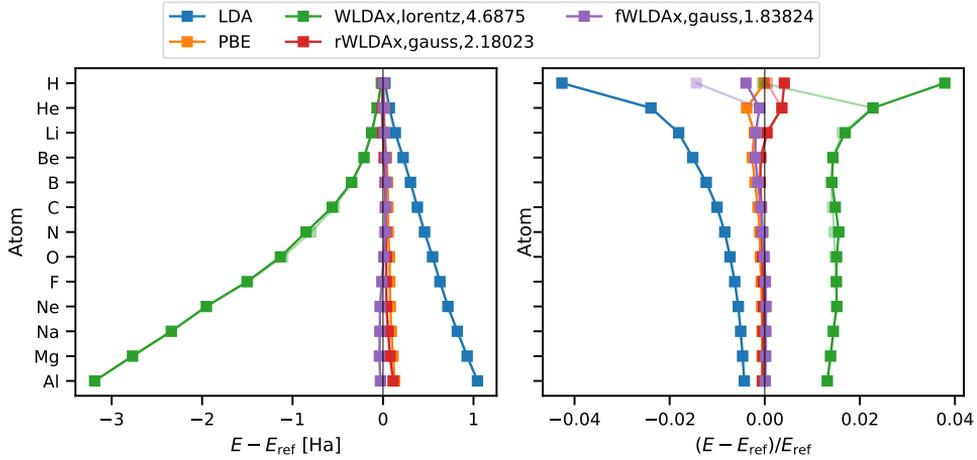


Figure 3.9: Errors in atomic energies for the exchange-only variants of WLDA. The opaque points show the results using Hartree-as-exchange and the transparent points show the results when using spin-neutral Hartree (see sec. 3.4).

	MAE [Ha]	MARE
LDA	0.482995	0.012596
PBE	0.060484	0.001326
WLDax	1.159472(1.148340)	0.017170(0.014074)
rWLDax	0.035884(0.034815)	0.001050(0.000765)
fWLDax	0.023108(0.021874)	0.000981(0.001750)

Table 3.3: Atomic performance benchmarks for the exchange-only variants of WLDA. With the exception of WLDax with the Lorentzian weight function, the WLDA variants performs comparably to PBE. The results using spin-neutral Hartree correction are shown in parentheses.

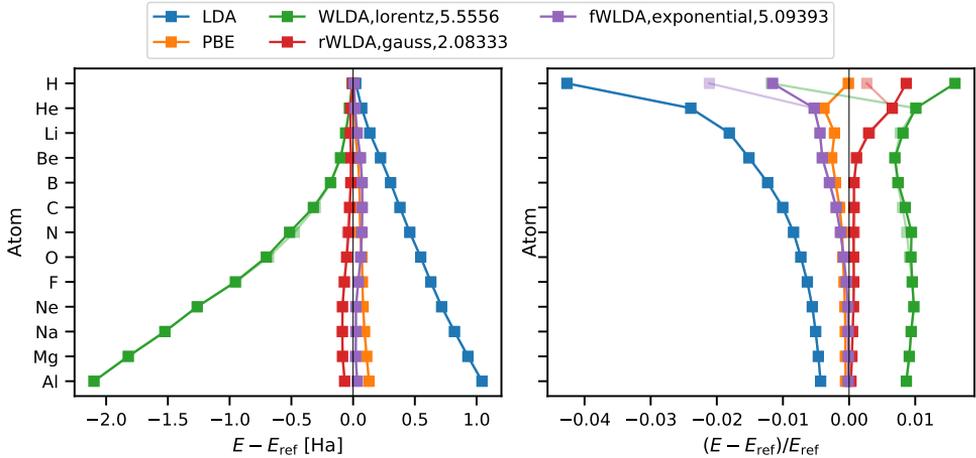


Figure 3.10: Errors in atomic energies for the variants of WLDA the include a correlation correction. The opaque points show the results using Hartree-as-exchange and the transparent points show the results when using spin-neutral Hartree (see sec. 3.4).

	MAE [Ha]	MARE
LDA	0.482995	0.012596
PBE	0.060484	0.001326
WLDA	0.737114(0.730207)	0.009418(0.008936)
rWLDA	0.045518(0.046966)	0.001898(0.001482)
fWLDA	0.041321(0.040525)	0.002566(0.003283)

Table 3.4: Atomic performance benchmarks for the variants of WLDA that include the correlation correction. The results using spin-neutral Hartree correction are shown in parentheses.

These results are quite promising as PBE is in some sense tuned to give good atomic energies [33]. However, atomic energies in themselves are not so useful and tests have shown that the atomic energies can be somewhat sensitive to the choice of weight function and c_1 parameter. The most striking effect is whether or not one uses the Lorentzian weight function which seems to perform worse than other weight functions for WLDA(x) even though it is optimal in the HEG. This can be seen in figure 3.11 where atomic energy calculations for different weight functions for WLDA are shown. It is not known why this is; perhaps it is related to the fact that the truncation function that follows from the Lorentzian weight has a much less sharp cutoff and a non-zero derivative at $q = 0$, see fig. 3.3. It remains to be seen if the results translate to physical properties of interest such as molecular atomization energies and lattice constants of solids. This is the topic of the next two sections.

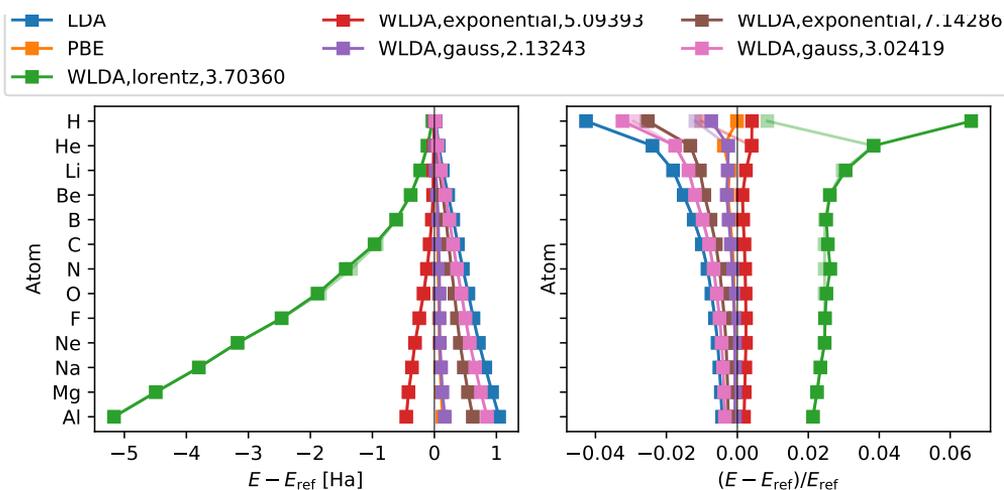


Figure 3.11: Error in ground state atomic energies calculated with the LDA, PBE and WLDA xc-functionals. The green, red and violet points indicate c_1 -parameters chosen such that the truncation function Φ satisfies $\Phi(q = 2k_F) = 1/2$. Brown and pink indicate the optimized c_1 values. Opaque and translucent markers indicate the *Hartree correction as exchange* and *Spin-neutral Hartree corrections* respectively.

3.7.2 Molecular Atomization Energies

In this section we test the WLDA variants on molecular atomizations energies for a subset of the G2 set of molecules from [34].

The results are shown in table 3.5. As with the atomic energies, WLDA(x) performs very poorly compared to rWLDA and fWLDA as well as compared to LDA and PBE. Again, this poor performance seems to stem from the Lorentzian weight function. In table A.1 results on the G2 dataset are shown for all weight functions for all variants. WLDA@exp and WLDA@gauss both have much lower error than WLDA@lorentz. (r/f)WLDA is however capable of reaching an accuracy on the level of PBE.

As to the effects of exchange-only vs. exchange + correlation corrections it is hard to draw any definitive conclusions. For rWLDA it looks preferable to include correlation whereas for fWLDA it looks counterproductive. To investigate this further we have calculated the atomization energies on a smaller subset of molecules for WLDA@exp with exchange-only and exchange + correlation but for the same c_1 value to isolate the effects of the correlation contribution. This is shown in table 3.6. From this table we see that adding the correlation correction only changes atomization energies by about 1% for a "well-behaved" weight function. In table 3.5 we see much larger difference between, say fWLDAx and fWLDA which is therefore probably mainly caused by the difference in c_1 value and the choice of weight function but not the inclusion of the correlation correction.

	ME	MAE	MARE
LDA	84.18	84.19	0.1984
PBE	26.23	26.87	0.08238
WLDax@Lor.	261 (329.7)	261 (329.7)	0.5688 (0.7163)
WLDA@Lor.	207.9 (257.7)	207.9 (257.7)	0.4551 (0.5613)
rWLDax@Gau.	26.8 (24.32)	28.04 (25.71)	0.07489 (0.06813)
rWLDA@Gau.	21.9 (20.74)	24.16 (23.2)	0.06555 (0.06098)
fWLDax@Gau.	35.87 (38.77)	37.3 (40.33)	0.09261 (0.0956)
fWLDA@Exp.	50.98 (55.79)	51.35 (56.2)	0.1205 (0.1277)

Table 3.5: Subset of G2/97 atomization energies (142 molecules) in kcal / mol. The values are calculated using the optimal c_1 parameters for each functional variant, see table 3.2. Results for spin-neutral Hartree is shown in parenthesis and other results are calculated with Hartree as exchange.

	LDA	LDAx	PBE	WLDax@Exp.	WLDA@Exp.	Expt.
CO	301	267	280	287 (284)	289 (285)	259
F ₂	78.4	67.8	66.1	45.6 (45.4)	46.3 (46.1)	38.4
H ₂	113	80.9	104	112 (121)	111 (120)	110
H ₂ O	268	216	243	268 (275)	269 (276)	233
HF	162	137	148	163 (167)	163 (168)	141
N ₂	266	210	243	253 (242)	255 (244)	228
O ₂	179	156	161	148 (143)	148 (144)	120
MAE	34	20	18	21 (21)	22 (22)	
MARE	0.31	0.22	0.19	0.14 (0.14)	0.14 (0.15)	

Table 3.6: Atomization energies in kcal / mol calculated with WLDA using the exponential weight functions. The c_1 value is chosen to be $c_1 = 5.09393$ such as to fulfill $\Phi(\bar{q}) = 1/2$. These results compared with table 3.5 suggest that the correlation contribution is negligible compared to choice of weight function and c_1 value.

3.7.3 Lattice Constant Benchmarks

To test how well WLDA describes structures we test the WLDA variants on a set of 27 cubic solids. The experimental lattice constants are adapted from [35] where they were used to test a Bayesian DFT functional. The calculations were run with a plane-wave cutoff of 800 eV and a 8x8x8 Monkhorst-Pack k-point grid.

The results for the exchange-only variants are shown in table 3.7 and the results for exchange + correlation correction is shown in table 3.8. Remarkably, it is possible to achieve an improvement over PBE. The WLDA functional with the Lorentzian weight function seems to perform best, in contrast to its performance on atomic and atomization energies, however it also often fails to converge. For lattice constants it may therefore be best to use the rWLDA variant with the Gaussian weight function.

Mater.	LDA	PBE	WLDax	rWLDax	fWLDax	Exp.
Li	3.360	3.425	3.348	3.354	3.344	3.451
Na	4.052	4.196	4.146	4.133	4.095	4.209
K	5.035	5.275	none	5.726	5.813	5.212
Rb	5.363	5.660	none	6.206	6.374	5.577
Ca	5.351	5.534	none	5.615	5.667	5.556
Sr	5.809	6.045	none	6.183	6.274	6.040
Ba	4.767	5.028	none	5.282	5.436	5.002
V	2.937	3.009	2.968	2.956	2.955	3.024
Nb	3.236	3.303	3.372	3.275	3.281	3.294
Ta	3.260	3.331	3.343	3.287	3.291	3.299
Mo	3.108	3.160	3.205	3.138	3.142	3.141
W	3.132	3.184	3.183	3.151	3.152	3.160
Fe	2.756	2.849	2.800(2.790)	2.784(2.780)	2.777(2.774)	2.853
Rh	3.761	3.839	3.793	3.782	3.773	3.793
Ir	3.805	3.868	3.844	3.827	3.820	3.831
Ni	3.436	3.529	3.442(3.442)	3.446(3.446)	3.436(3.436)	3.508
Pd	3.841	3.944	3.914	3.885	3.871	3.876
Pt	3.889	3.969	3.941	3.919	3.909	3.913
Cu	3.528	3.642	3.581	3.572	3.561	3.596
Ag	4.000	4.147	none	4.114	4.107	4.062
Au	4.072	4.174	4.266	4.144	4.134	4.062
Pb	4.868	5.023	none	5.133	5.224	4.912
Al	3.982	4.038	4.002	4.005	3.997	4.019
C	3.533	3.573	3.583	3.556	3.559	3.544
Si	5.407	5.476	5.532	5.459	5.462	5.415
Ge	5.632	5.767	none	5.826	5.872	5.639
Sn	6.458	6.640	none	6.670	6.707	6.474
MAE	0.078	0.049	0.014(0.014)	0.040(0.040)	0.051(0.051)	

Table 3.7: Lattice constant results for the exchange-only variants of WLDA. Entries that are "none" indicate that the calculation did not converge. Numbers in parenthesis are the results with spin-neutral Hartree correction. Most entries do not have a spin-neutral Hartree result because those calculations were spin-paired and spin-neutral Hartree and Hartree-As-Exchange are equivalent in that case.

Mater.	LDA	PBE	WLDA	rWLDA	fWLDA	Exp.
Li	3.367	3.433	3.354	3.357	3.352	3.451
Na	4.052	4.198	4.108	4.134	4.082	4.209
K	5.038	5.277	none	5.898	5.712	5.212
Rb	5.369	5.664	none	6.397	6.219	5.577
Ca	5.351	5.534	none	5.669	5.625	5.556
Sr	5.809	6.045	none	6.264	6.221	6.040
Ba	4.771	5.028	none	5.385	5.371	5.002
V	2.926	3.001	2.946	2.948	2.940	3.024
Nb	3.243	3.304	3.329	3.288	3.283	3.294
Ta	3.247	3.316	3.304	3.280	3.276	3.299
Mo	3.106	3.157	3.170	3.142	3.136	3.141
W	3.130	3.181	3.164	3.152	3.148	3.160
Fe	2.756	2.849	2.784(2.781)	2.788(2.784)	2.774(2.769)	2.853
Rh	3.760	3.839	3.779	3.783	3.769	3.793
Ir	3.803	3.866	3.828	3.827	3.816	3.831
Ni	3.436	3.529	3.438(3.438)	3.446(3.446)	3.435(3.434)	3.508
Pd	3.840	3.942	3.882	3.888	3.863	3.876
Pt	3.889	3.969	3.920	3.922	3.905	3.913
Cu	3.530	3.645	3.563	3.579	3.551	3.596
Ag	4.002	4.146	4.185	4.134	4.086	4.062
Au	4.072	4.176	4.173	4.156	4.124	4.062
Pb	4.872	5.033	none	5.203	5.192	4.912
Al	3.984	4.041	3.996	4.010	3.994	4.019
C	3.533	3.573	3.567	3.562	3.553	3.544
Si	5.407	5.476	5.483	5.466	5.450	5.415
Ge	5.632	5.767	6.318	5.871	5.832	5.639
Sn	6.458	6.640	7.239	6.715	6.665	6.474
MAE	0.078	0.049	0.031(0.031)	0.049(0.049)	0.043(0.043)	

Table 3.8: Lattice constant results for the exchange and correlation variants of WLDA. Entries that are "none" indicate that the calculation did not converge. Numbers in parenthesis are the results with spin-neutral Hartree correction. Most entries do not have a spin-neutral Hartree result because those calculations were spin-paired and spin-neutral Hartree and Hartree-As-Exchange are equivalent in that case.

3.8 Perspective: Comparison to other Functionals

It may be instructive to consider the differences between WLDA and other functional types as this may provide some insights into possible improvements of the shortcomings of WLDA.

Local Functionals To my knowledge, essentially the only used local density functional is the Local Density Approximation. LDA simply gives the exchange-correlation energy density at a point as the XC energy of a HEG with the local density. Non-local

correlations seem impossible to describe in LDA but it can actually be expected to work well for a large range of systems [36].

WLDA, being defined from LDA, shares the feature of producing exact energies in the HEG. Beyond that, WLDA also tried to match other features of the HEG, namely the correlation hole. Thus we can expect WLDA to outperform LDA.

Meta-LDA Another intriguing possibility is to combine a WLDA procedure with a *meta-LDA* functional [37]. Meta-LDAs use the kinetic energy density to calculate an effective HEG density, $\tilde{n}(\mathbf{r})$. They therefore depend on the wavefunctions and not just the density. The effective density is combined with the local density in the form $n^{\text{avg}}(\mathbf{r}) = n^{1-x}(\mathbf{r})\tilde{n}^x(\mathbf{r})$ for $x \in [0, 1]$, and n^{avg} can then be used in the LDA exchange-correlation function where the energy density per electron is evaluated using n^{avg} . It is unknown whether a "meta-WLDA" could improve on WLDA; the two types would behave identically in the HEG since $\tilde{n} = n$ there. The optimal x , as well as the optimal variant of WLDA, would have to be determined some other way.

Semi-local Functionals Functionals that employ density gradient information fall in the class of Generalized Gradient Approximations (GGA) and this class of functionals generally improve upon LDA for properties in atoms, molecules, and solids [38]. They have almost the same computational cost as LDA but offer significant improvements. However, GGAs suffer from a fundamental limitation. GGA exchange can be written as

$$E_x^{\text{GGA}}[n] = \int d\mathbf{r} \epsilon_x^{\text{HEG}}(n(\mathbf{r}))F_x(s(\mathbf{r})), \quad (3.97)$$

where the *enhancement factor* F_x depends on the density gradient through

$$s = \frac{|\nabla n|}{2k_F n}. \quad (3.98)$$

If the GGA is to match the HEG in the limit of a slowly varying density, $s \rightarrow 0$, then F_x must obey

$$F_x(s) \xrightarrow{s \rightarrow 0} 1 + \mu s^2, \quad (3.99)$$

with $\mu = 10/81$ [33]. However, for atoms it is found that using instead $F_x(s) \approx 1 + 2\mu s^2$ is necessary to get good values for atomic energies and by extension for atomization energies[33]. Matching the exact limit is relevant for solids where the density is more slowly varying than in atoms. Thus GGA exchange can either describe atomization energies or solids well, but not both and a similar set of opposing constraints exist for the correlation energy [33].

As good as GGA is in practice, it seems necessary to move beyond it to find universally applicable functionals.

The strength of GGA is the many well-known and well understood limits which can be used to fix the parameters in the functional. Because of this it may be easier to interpret what GGA is doing or predict when it will work than it seems to be for WLDA.

Beyond GGA Introducing the kinetic energy density $\tau = \frac{1}{2} \sum_i^{\text{occ}} |\nabla \phi_i|^2$ as a fundamental variable leads to the class of *meta*-GGAs that are no longer explicit functionals of the density only. Another class is *hybrid*-GGAs that mix in exact exchange (a fully non-local quantity and again a functional of the wavefunctions rather than the density). Hybrid functionals are of much higher computational complexity than meta-GGAs and generally achieve greater accuracy. However, in 2016 the meta-GGA "SCAN" constructed by matching a list of known exact limits was shown to have comparable accuracy to hybrid-GGAs at meta-GGA cost [39].

Beyond GGA functionals have considerable complexity but the use of the kinetic energy density allows the functional to distinguish between covalent and metallic bonds [39] and one can apply these limits to the functional form to reason about the behaviour of the exchange-correlation energy. Again WLDA has the short-coming that it is not easy to see *a priori* how it will behave for different types of physical systems or even that it can distinguish them.

Non-local Functionals Fully non-local functionals have also been tried. One example is the van der Waals functional [30, 40] which includes a description of the van der Waals interaction with a fully non-local energy contribution of the form

$$E_{\text{vdW}}[n] = \frac{1}{2} \int \int \mathbf{drdr}' n(\mathbf{r})n(\mathbf{r}')\phi(n, \nabla n, |\mathbf{r} - \mathbf{r}'|). \quad (3.100)$$

This functional captures van der Waals dominated physics such as the double wall carbon nanotubes investigated in [30] or conceivably van der Waals bonded multilayer structures.

In the late 1970s the *Weighted Density Approximation* (WDA) was proposed [41–43] and it was later developed further by David Singh [44]. Despite the similarity in name, it is quite different from WLDA. In WDA the exchange-correlation energy is written in terms of the exchange-correlation hole:

$$\begin{aligned} E_{\text{xc}}[n] &= \int \int \mathbf{drdr}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} [g[n](\mathbf{r}, \mathbf{r}') - 1], \\ &= \int \int \mathbf{drdr}' \frac{n(\mathbf{r})n_{\text{xc}}(\mathbf{r}, \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \end{aligned} \quad (3.101)$$

where g is the pair-distribution function and n_{xc} is the xc hole. This expression is exact but g is unknown in general, and whereas LDA corresponds to the approximation

$$n(\mathbf{r}')g[n](\mathbf{r}, \mathbf{r}') \rightarrow n(\mathbf{r})g^h(n(\mathbf{r}), |\mathbf{r} - \mathbf{r}'|), \quad (3.102)$$

where g^h is the pair-distribution function in the HEG, WDA uses the approximation

$$n(\mathbf{r}')g[n](\mathbf{r}, \mathbf{r}') \rightarrow n(\mathbf{r}')g^w(\bar{n}(\mathbf{r}), |\mathbf{r} - \mathbf{r}'|), \quad (3.103)$$

where g^w is a model pair-distribution function (not necessarily equal to g^h), and \bar{n} is the eponymous weighted density. \bar{n} is determined such that the xc hole sum rule holds:

$$\int \mathbf{dr}' n(\mathbf{r}') [g(\bar{n}(\mathbf{r}), |\mathbf{r} - \mathbf{r}'|) - 1] = -1. \quad (3.104)$$

However, the "weighted" density is therefore not *a priori* related to the physical density. WDA improves upon LDA for lattice constants and bulk moduli [44, 45] and in some cases can even get bandstructures that agree well with GW results [46].

Non-local functionals as a class has the potential to be exact since it trivially includes all possible functionals, including the exact xc functional. The question remains how to construct useful approximations. In the two examples above there were two philosophies: In the van der Waals functional a specific non-local physical effect (van der Waals bonding) was identified and described, and in WDA one attempts to construct a functional that satisfies exact constraints, similar to LDA. The philosophy of WLDA is similar to WDA's but while WDA tries to satisfy a kind of "coarse-grained" constraint (the integral of a function is correct) for all systems, WLDA tries to satisfy a "fine-grained" constraint (the function is correct) but just for one system (the HEG). Perhaps future versions of WLDA can attempt to satisfy the sum-rule and match the HEG correlation hole simultaneously.

3.9 Open Questions

There remain a few open questions that we summarize here.

Performance on Other Properties It remains to be seen how well WLDA performs on other electronic structure properties such as the band gap, determination of magnetic state.

Reproduction of the exact density Another relevant question is how well WLDA reproduces the exact electron density. This question is separate from the question of the accuracy of the predicted energies [47].

Vacuum Regularization As we saw in sec. 3.5 there is a divergent term in the potential that seems to indicate a flaw in the definition of WLDA. Future versions of WLDA should be redefined to eliminate this divergence.

Spin-polarized Correlation Treatment To define WLDA for spin-polarized systems we guessed a reasonable form of the correlation correction, see sec. 3.4. The correlation correction form could instead be chosen by matching different alternatives to the spin-polarized HEG or by attempting to find theoretical arguments for the form of the correlation.

HAX or no Spin-neutral Hartree It is not fully resolved whether it is best to use Hartree-as-exchange or the spin-neutral version of the Hartree correction. So far it seems that there is not a huge difference in accuracy, and, barring any theoretical justifications one way or the other, it may therefore be preferable to use the spin-neutral Hartree correction because the implementation should be slightly faster.

WLDA Setups We have implemented WLDA in GPAW which is a PAW code and as such uses so-called "PAW setups" (the PAW partial waves and projectors). These setups are typically optimized for the given exchange-correlation functional. In this work we have simply used the LDA setups which should be reasonable given that we correct directly on LDA but it is possible that better convergence can be achieved with WLDA specific setups.

Additional Constraints In this version of WLDA we have attempted to match the HEG energy and correlation hole. It may be possible to satisfy further constraints such as the sum-rule for the exchange-correlation hole. Matching exact constraints should hopefully improve the performance of the functional and it is at least a systematic way to develop a functional.

3.10 Conclusion

WLDA achieves systematic improvements over LDA and thus seems to be a promising functional. However, it is clearly not the last word: WLDA is fully non-local yet "only" performs at the level of PBE. More work is needed to improve the functional and test its performance on more systems and properties. The first step is to design a form that is free of the potential divergence perhaps by simply using the WLDA-3 form. However, because WLDA is designed to have a correctly behaving kernel it may be suitable for TDDFT calculations. Of course, at this stage, only the adiabatic kernel is defined so properties where the time-dependence is crucial will probably not be well described by WLDA. In the future, it may also be interesting to investigate how WLDA performs when used as a basis for many-body perturbation methods, such as GW. If nothing else, hopefully WLDA can inspire new avenues of exchange-correlation functional development.

4 High-Throughput GW

In this chapter we give an introduction to quasiparticle theory and the GW approximation and detail the results of paper II which pertains to the accuracy of various approximations employed within typical GW calculations, and suggestions for improving the accuracy within the context of high throughput GW calculations.

4.1 Quasiparticle Theory

In a physical systems, electrons will interact with each other. A given electron will repel other electrons due to the Coulomb interaction, and a picture that is often used is that the local electron density around the given electron is depleted. Thus the electron is surrounded by a "cloud" of positive charge that moves along with it. The true physical picture probably involves more complicated correlations. This picture is called the quasiparticle picture because the electron + positive charge cloud behaves as if it were a particle. Quasiparticles are only approximate eigenstates of the true Hamiltonian both because we must use approximations to describe the quasiparticles but also because interacting systems may not have eigenstates that are single-particle states.

In more mathematical terms, quasiparticles arise because the Green's function for the electron in the interacting system is not equal to that of a particle in a non-interacting system. The Green's function is defined as[48, 49]

$$G(\mathbf{r}, t, \mathbf{r}', t') = -i\langle \mathcal{T} \Psi(\mathbf{r}, t) \Psi^\dagger(\mathbf{r}', t') \rangle_T, \quad (4.1)$$

where \mathcal{T} is the time-ordering operator, $\Psi(\mathbf{r}, t)$ is the annihilation operator for an electron at \mathbf{r}, t , and $\langle \cdot \rangle_T$ is a thermal average taken at temperature T . For systems that do not change in time the Green's function only depends on $t - t'$ and the Fourier transform in time of G will then only depend on one frequency. We recall that the Green's function is equivalent to the *spectral function* through

$$A(\mathbf{r}, \mathbf{r}', E) = \pi^{-1} |\text{Im}(G(\mathbf{r}, \mathbf{r}', E))|, \quad (4.2)$$

$$G(\mathbf{r}, \mathbf{r}', E) = \int_{-\infty}^{\infty} \frac{A(\mathbf{r}, \mathbf{r}', E')}{E - E'} dE'. \quad (4.3)$$

The spectral function and the Green's function are important because these are the quantities involved when we perform single-particle experiments, such as photo-emission and inverse photo-emission experiments which, respectively, use light to emit an electron from a material and inject electrons into a material and measure the emitted light[48]. Spectral functions also arise in tunneling spectroscopy experiments when measuring currents through a system.

The *meaning* of the spectral function can be understood by considering a non-interacting system. Then one can show [49] that the spectral function for the Green's function $-i\langle \mathcal{T} \Psi(\nu, t) \Psi^\dagger(\nu', t') \rangle$, where ν denotes eigenstates of the system, diagonal in ν, ν'

and the diagonal elements are Dirac-delta functions in energy with the peak located at the energy of the eigenstate. In other words

$$A(\nu, \nu', E) = \delta_{\nu, \nu'} \delta(E_\nu - E). \quad (4.4)$$

When we consider the effect of single-particle perturbations to the Hamiltonian the only effect on $A(\nu, \nu', E)$ will be to shift the location of the peaks. Multi-particle perturbations, such as the Coulomb interaction, will however both shift the peaks and broaden them [49]. The interpretation of the spectral function is therefore that the peaks occur at single-particle energies, or more generally speaking, the peaks occur at the energies of the quasi-particles. One can also show that the width of the peak is proportional to the inverse lifetime of the quasi-particles[49]. More precisely, provided that the quasiparticle picture is physical, the poles of the single-particle Green's function occur at certain complex values of the energy, ϵ^{QP} , where the real part of ϵ^{QP} is the quasiparticle energy and the imaginary part is the lifetime of the quasiparticle[50]. The problem of determining the poles of the interacting single-particle Green's function is then central to calculating the quasiparticle energies.

To calculate the Green's function one defines a reference system, re-expresses eq. 4.1 in terms of a thermal average in the reference system, takes the time-derivative of G , and finally Fourier transforms, see [49]. The result is the *Dyson equation* [48, 49]:

$$G(\mathbf{r}, \mathbf{r}', E) = G^0(\mathbf{r}, \mathbf{r}', E) + \int \int G^0(\mathbf{r}, \mathbf{r}_1, E) \Sigma(\mathbf{r}_1, \mathbf{r}_2, E) G(\mathbf{r}_2, \mathbf{r}', E) d\mathbf{r}_1 d\mathbf{r}_2, \quad (4.5)$$

where G^0 is the Green's function of the non-interacting system, and Σ is the *self-energy* operator. The self-energy contains all the information about interactions in the system.

Finding an equation for the poles of G is most easily done by considering the Lehmann representation of G [51]

$$G(z) = \sum_i \frac{|\psi_i^{QP}\rangle \langle \psi_i^{QP}|}{z - \epsilon_i^{QP}}. \quad (4.6)$$

Here $z \in \mathbb{C}$ and ψ_i^{QP} are as of yet undetermined states, which we identify with the quasiparticle wavefunctions. Writing the solution to the Dyson equation in the form $G(z) = [z - H_0 - \Sigma(z)]^{-1}$ (neglecting spatial coordinates for brevity) and integrating $1 = [z - H_0 - \Sigma(z)]G(z)$ in a small circular contour, C around ϵ_i^{QP} we find

$$\int_C 1 = 0 = \int_C [z - H_0 - \Sigma(z)]G(z) = (\epsilon_i^{QP} - H_0 - \Sigma(\epsilon_i^{QP})) |\psi_i^{QP}\rangle \langle \psi_i^{QP}|, \quad (4.7)$$

which follows from the residue theorem [51]. Since the LHS is the zero-operator, we must also have

$$\left[\epsilon_i^{QP} - H_0 - \Sigma(\epsilon_i^{QP}) \right] |\psi_i^{QP}\rangle = 0. \quad (4.8)$$

This determines the quasiparticle energies ϵ_i^{QP} .

To make further progress one usually starts from a Kohn-Sham wavefunction, treating $\Sigma - V_{\text{xc}}$ perturbatively and writing $\epsilon_i^{\text{QP}} = \epsilon_i^{\text{KS}} + \Delta\epsilon_i$ which yields

$$\Delta\epsilon_i = \langle \psi_i^{\text{KS}} | \Sigma(\epsilon_i^{\text{QP}}) - V_{\text{xc}} | \psi_i^{\text{KS}} \rangle. \quad (4.9)$$

Already one could attempt to solve this equation. It would however require evaluating the self-energy at multiple frequencies.

Instead one usually uses the linear approximation to this equation:

$$\Delta\epsilon_i \approx \langle \psi_i^{\text{KS}} | \Sigma(\epsilon_i^{\text{KS}}) + \Delta\epsilon \Sigma'(\epsilon_i^{\text{KS}}) - V_{\text{xc}} | \psi_i^{\text{KS}} \rangle \quad (4.10)$$

\Rightarrow

$$\epsilon_i^{\text{QP}} = \epsilon_i^{\text{KS}} + Z_i \langle \psi_i^{\text{KS}} | \Sigma(\epsilon_i^{\text{KS}}) - V_{\text{xc}} | \psi_i^{\text{KS}} \rangle, \quad (4.11)$$

where

$$Z_i = \langle \psi_i^{\text{KS}} | 1 - \Sigma'(\epsilon_i^{\text{KS}}) | \psi_i^{\text{KS}} \rangle. \quad (4.12)$$

Below, we will be studying the solutions to this equation, including the effects of doing the linear approximation.

4.2 Quasiparticle Weight

The quantity Z has a useful physical meaning: it is the *quasiparticle weight*. The quasiparticle weight is the spectral weight of the quasiparticle state. This means that Z must be $0 \leq Z \leq 1$ and for a single-particle state that is an eigenstate of the system the quasiparticle weight will be 1. Smaller values of Z indicates that the quasiparticle is further from being a true eigenstate as we will see below in a rigorous manner. This derivation is based on [51].

The first step is to consider G at zero temperature assuming $t > t'$ (it is easily extended to finite temperature and $t' < t$):

$$G(\mathbf{r}, t, \mathbf{r}', t') = \langle N, 0 | e^{iHt} \Psi(\mathbf{r}) e^{-iH(t-t')} \Psi^\dagger(\mathbf{r}') e^{-iHt'} | N, 0 \rangle. \quad (4.13)$$

Here $|N, 0\rangle$ is the N -particle ground state. If we denote the excited state i for $N + 1$ particles by $|N + 1, i\rangle$ we get via the Lehmann representation the expression

$$\begin{aligned} G(\mathbf{r}, t, \mathbf{r}', t') &= \langle N, 0 | e^{iHt} \Psi(\mathbf{r}) \sum_i |N + 1, i\rangle \langle N + 1, i| e^{-iH(t-t')} \Psi^\dagger(\mathbf{r}') e^{-iHt'} | N, 0 \rangle \\ &= \sum_i \langle N, 0 | e^{iE_0^N t} \Psi(\mathbf{r}) | N + 1, i\rangle \langle N + 1, i| e^{-iE_i^{N+1}(t-t')} \Psi^\dagger(\mathbf{r}') e^{-iE_0^N t'} | N, 0 \rangle \\ &= \sum_i e^{-i\epsilon_{i+}^{\text{QP}}(t-t')} \langle N, 0 | \Psi(\mathbf{r}) | N + 1, i\rangle \langle N + 1, i| \Psi^\dagger(\mathbf{r}') | N, 0 \rangle. \end{aligned} \quad (4.14)$$

Here $\epsilon_{i+}^{\text{QP}} = E_i^{N+1} - E_0^N$. We also define the quasiparticle wavefunctions

$$\psi_{i+}^{\text{QP}}(\mathbf{r}) = \langle N + 1, i | \Psi^\dagger(\mathbf{r}) | N, 0 \rangle, \quad (4.15)$$

$$\psi_{i-}^{\text{QP}}(\mathbf{r}) = \langle N - 1, i | \Psi(\mathbf{r}) | N, 0 \rangle, \quad (4.16)$$

where the second quasiparticle wavefunction would arise from considering $t' < t$ and it has the associated energy $\epsilon_{i-}^{\text{QP}} = E_0^N - E_i^{N-1}$. From the definitions we see that $\epsilon_{i\pm}^{\text{QP}}$ are related to electron addition (+) and removal (-) energies and with μ the chemical potential we have

$$\epsilon_{i+}^{\text{QP}} > \mu \quad (4.17)$$

and

$$\epsilon_{i-}^{\text{QP}} \leq \mu. \quad (4.18)$$

If we now take G and Fourier transform with respect to $t - t'$ and switch to basis-free notation we arrive again at eq. (4.6):

$$G(z) = \sum_i \frac{|\psi_i^{\text{QP}}\rangle \langle \psi_i^{\text{QP}}|}{z - \epsilon_i^{\text{QP}}}, \quad (4.19)$$

where it is now understood that $\epsilon_i^{\text{QP}} = \epsilon_{i+}^{\text{QP}}$ if $\epsilon_{i+}^{\text{QP}} > \mu$ and $\epsilon_i^{\text{QP}} = \epsilon_{i-}^{\text{QP}}$ otherwise. We have both rederived eq. (4.6) and derived an expression for the quasiparticle wavefunctions, which we will use below.

We now turn to the meaning of the quasiparticle wavefunctions. First we consider the overlap of the quasiparticle wavefunction with an arbitrary single-particle orbital ϕ :

$$\langle \phi | \psi_{i+}^{\text{QP}} \rangle = \int d\mathbf{r} \phi(\mathbf{r})^* \langle N+1, i | \Psi^\dagger(\mathbf{r}) | N, 0 \rangle = \langle N+1, i | c_\phi^\dagger | N, 0 \rangle, \quad (4.20)$$

where c_ϕ^\dagger creates a particle in the state ϕ , $c_\phi^\dagger = \int d\mathbf{r} \phi^*(\mathbf{r}) \Psi^\dagger(\mathbf{r})$.

We next use the fact that the norm, defined as $\|f\| = \sqrt{\langle f | f \rangle}$, is also given by

$$\|f\| = \max_\varphi \{ \langle \varphi | f \rangle, \|\varphi\| = 1 \}. \quad (4.21)$$

Thus the norm of the quasiparticle wavefunction can be expressed as

$$\|\psi_{i+}^{\text{QP}}\| = \max_\varphi \{ \langle N+1, i | c_\varphi^\dagger | N, 0 \rangle, \|\varphi\| = 1 \}. \quad (4.22)$$

That is, the norm of the quasiparticle wavefunction is the *maximal overlap between a single-particle state added to the ground state and the true excited state*. The interpretation is that it measures how close the quasiparticle state is to a true eigenstate.

We now seek an expression for the quasiparticle norm in terms of the self-energy.

The first step is to consider the defining equation for the Green's function at a general complex energy z :

$$[z - H_0 - \Sigma(z)] G = 1. \quad (4.23)$$

We can form a representation of G using this equation by considering the eigenstates of the, generally non-hermitian, operator $H_0 + \Sigma(z)$ at some complex energy z :

$$[H_0 + \Sigma(z)] |\psi_n(z)\rangle = \epsilon_n(z) |\psi_n(z)\rangle. \quad (4.24)$$

In terms of these states, G has the spectral representation

$$G(z) = \sum_n \frac{|\psi_n(z)\rangle \langle \psi^n(z)|}{z - \epsilon_n(z)}. \quad (4.25)$$

Here $\psi^n(z)$ is the dual element of $\psi_n(z)$ satisfying $\langle \psi^m(z) | \psi_n(z) \rangle = \delta_{mn}$ (we don't necessarily have $\langle \psi_n | = \langle \psi^n |$ because the operator is non-Hermitian). The correctness of this expression can be seen by multiplying $G(z)$ by $z - H_0 - \Sigma(z)$ and using the (assumed) completeness of the states $\psi_n(z)$.

The two spectral representations of G must of course coincide for all z and in general this means there is no relation between ψ_i^{QP} and $\psi_n(z)$.

However, at $z = \epsilon_i^{\text{QP}}$ we must have

$$|\psi_i^{\text{QP}}\rangle \propto |\psi_i(\epsilon_i^{\text{QP}})\rangle, \quad (4.26)$$

as can be readily seen from the defining equations. If we choose $\psi_i(z)$ to be normalized we get

$$|\psi_i(\epsilon_i^{\text{QP}})\rangle = \frac{|\psi_i^{\text{QP}}\rangle}{\| |\psi_i^{\text{QP}}\rangle \|}. \quad (4.27)$$

If we now consider the matrix element $\langle \psi^i(z) | G(z) | \psi_i(z) \rangle$ in the two different spectral representations we get

$$\langle \psi^i(z) | G(z) | \psi_i(z) \rangle = \sum_n \frac{\langle \psi^i(z) | \psi_n^{\text{QP}} \rangle \langle \psi_n^{\text{QP}} | \psi_i(z) \rangle}{z - \epsilon_n^{\text{QP}}} = \sum_n \frac{\langle \psi^i(z) | \psi_n(z) \rangle \langle \psi_n(z) | \psi_i(z) \rangle}{z - \epsilon_n(z)}. \quad (4.28)$$

Integrating in a contour around ϵ_i^{QP} and using the orthogonality relations and the residue theorem we get

$$\langle \psi^i(\epsilon_i^{\text{QP}}) | \psi_i^{\text{QP}} \rangle \langle \psi_i^{\text{QP}} | \psi_i(\epsilon_i^{\text{QP}}) \rangle = \frac{1}{1 - \epsilon_i'(z)}, \quad (4.29)$$

with $\epsilon_i'(z) = \frac{d}{dz} \epsilon_i(z)$.

From eq. (4.27) we then find an expression for the norm of the quasiparticle wavefunction

$$\| |\psi_i^{\text{QP}}\rangle \|^2 = \frac{1}{1 - \epsilon_i'(z)}. \quad (4.30)$$

Expressed in terms of the self-energy we have

$$\|\psi_i^{\text{QP}}\|^2 = \left[\langle \psi_i(\epsilon_i^{\text{QP}}) | 1 - \Sigma'(\epsilon_i^{\text{QP}}) | \psi_i(\epsilon_i^{\text{QP}}) \rangle \right]^{-1} \quad (4.31)$$

Comparing with eq. (4.12) we see that

$$Z \approx \|\psi_i^{\text{QP}}\|^2, \quad (4.32)$$

if $\epsilon_i^{\text{KS}} \approx \epsilon_i^{\text{QP}}$ and $\psi_i(\epsilon_i^{\text{QP}}) \approx \psi_i^{\text{KS}}$. We will generally refer to Z as the quasiparticle weight, even though, as we saw here, it is only an approximation.

4.2.1 Restrictions on Z

Eq. (4.32) imposes some consistency requirements on calculated values of Z . Since Z is (approximately) a state norm we must have $Z \geq 0$. Also, as noted above, the quasiparticle weight is the integral of the quasiparticle peak of the spectral function. Because the spectral function is normalized we also should have $Z \leq 1$. Other than that, there are no hard requirements. Because of the assumptions made, Z should only be used for relatively weakly correlated systems, i.e. systems that are well-described by Kohn-Sham theory.

4.3 Hedin's Equations and the GW Approximation

The remaining problem of many-body perturbation theory is then to find approximations to Σ that are accurate and tractable. One choice is simply to expand Σ diagrammatically in terms of the bare Coulomb interaction, v , however, it is well known that this expansion is divergent in metals and generally the convergence is poor.

An alternate approach was developed by Hedin [52], which expands instead in the *screened* Coulomb interaction, W . The screened interaction is not only expected to be smaller, thus leading generally to faster convergence, but one can also use various approximation schemes for the screened interaction, such as model dielectric approaches or plasmon-pole models [48]. Additionally, one can also calculate the screened interaction using density functional perturbation theory which effectively includes the most relevant parts of the dielectric function without calculating it explicitly [53–55].

Hedin's equations will not be derived here but they may be written in the form [48, 52]

$$\Gamma(1, 2; 3) = \delta(1, 2)\delta(1, 3) + \int \frac{\delta\Sigma(1, 2)}{\delta G(4, 5)} G(4, 6)G(7, 5)\Gamma(6, 7; 3)d(4, 5, 6, 7), \quad (4.33)$$

$$P(1, 2) = -i \int G(2, 3)G(4, 2)\Gamma(3, 4; 1)d(3, 4), \quad (4.34)$$

$$W(1, 2) = v(1, 2) + \int W(1, 3)P(3, 4)v(4, 2)d(3, 4), \quad (4.35)$$

$$\Sigma(1, 2) = i \int G(1, 4)W(1^+, 3)\Gamma(4, 2; 3)d(3, 4), \quad (4.36)$$

$$G(1, 2) = G^0(1, 2) + \int G^0(1, 3)\Sigma(3, 4)G(4, 2)d(3, 4). \quad (4.37)$$

Here the notation $\mathbf{1}$ stands for all the coordinates of particle 1 (e.g. space, spin, time), and we have introduced the vertex correction Γ and the polarizability P . These equations are to be solved self-consistently, so that we start with a guess for G , typically one starts with G^0 , then calculates Γ , then P , W , Σ and finally G again.

A common approximation is to fix $\Gamma = 1$. This is called the GW Approximation because the self-energy reduces to $\Sigma_{\text{GW}} = GW$. Furthermore, one can also neglect to iterate to self-consistency and instead just use the first iteration of the self-consistency cycle. This leads to the so-called G_0W_0 approximation so named because G and W from the non-interacting system are used. This is approximation used in paper II.

The GW approximation is generally much more accurate than LDA or GGA methods. Compared to LDA it usually significantly improves the bandgap [48]. The GW approximation is also qualitatively correct in certain limiting cases and satisfies important conservation criteria [48, 56]. Furthermore, GGA is generally designed to improve total energies but not the description of quasiparticles, so for properties derived from quasiparticle physics, something like GW is probably needed [56].

4.4 Results of GW Paper

A few years ago, a database of computational 2D materials was developed by the CAMD group and compiled into the Computational 2D materials Database (C2DB) [6]. A wide variety of properties were calculated, ranging from bandstructure to effective masses and polarizabilities. For 370 of these materials, the G_0W_0 eigenvalues and Z values were also calculated and these calculations made up a trove of G_0W_0 data much larger than most other published databases. Even though G_0W_0 is empirically better than LDA/GGA the desirability of doing G_0W_0 calculations is offset by the large computational cost and the unknown factor of how often such calculations produce unphysical or inaccurate results. For the purposes of high-throughput studies, which are becoming increasingly more important as computational power increases, it is beneficial to study these questions. High-throughput studies have a tendency to bring out all sorts of bugs and edge cases you did not expect. The G_0W_0 data in the C2DB had not been studied in a statistical fashion and because the size of the dataset one would expect that useful information was waiting to be found. In paper II we studied this G_0W_0 data and in this section we will go through the main results.

Calculation Details

A few details about the implementation of G_0W_0 in GPAW is relevant insofar as it informs us of possible error sources.

The G_0W_0 data in the C2DB is based on DFT calculations with the PBE functional, using a 800 eV plane-wave cutoff and including spin-orbit coupling by diagonalizing the spin-orbit Hamiltonian for each k -point using the states from PBE.

The implementation follows that described in [57]. Three different energy cutoffs for the screened interaction were used and the results were extrapolated to infinite cutoff. Furthermore, full frequency integration of the screened interaction is used, and truncations for the exchange and correlation parts of the self-energy are used to avoid spurious effects from artificially repeated layers [58]. These truncations lead to slower

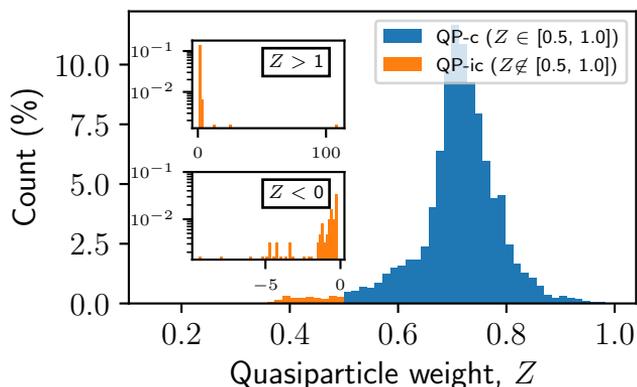


Figure 4.2: The observed distribution of quasiparticle weights, Z , in the C2DB G_0W_0 dataset. Source: paper II.

We also see that some fraction of the Z values are strictly *unphysical*, in the sense that they cannot correspond to a quasiparticle norm. $Z < 0$ implies a wavefunction norm less than zero, which is mathematically impossible and $Z > 1$ implies that more than all the spectral weight is in quasiparticle peak. Hence we denote a Z with $Z > 1$ or $Z < 0$ as unphysical. It is a relatively small fraction of points that are unphysical: less than 0.3% are outside the physical range.

In summary, most Z values are "good" in the sense of being physical and QP-c. The fraction of such states will most likely depend on the materials studied; strongly correlated systems should generally have lower quasiparticle weights. Knowing the typical distribution of Z values is still relevant for future high-throughput studies, and we can already suggest some important considerations for screening workflows. We saw that QP-ic states do form a not-insignificant fraction (2.5 %) of all states. From the physical reasoning described in sec. 4.1 alone, G_0W_0 might not be adequate for such states and perhaps some form of self-consistent is needed. Furthermore, we show below that the largest errors in the solution of the Quasiparticle equation are associated with QP-ic states. Both physical reasoning and empirical data suggests that QP-ic eigenvalues are probably not accurate.

The next question is then what causes QP-ic or unphysical Z values. In general, we found no correlation between the Z values and obvious quantities such as the PBE energies or the G_0W_0 correction. Furthermore, while we generally see a higher percentage of magnetic materials having QP-ic states see fig. 4.3, there are no other obvious patterns. A previous study had found that unphysical Z values could stem from a poor PAW basis set [59–61], but we found no such correlation either.

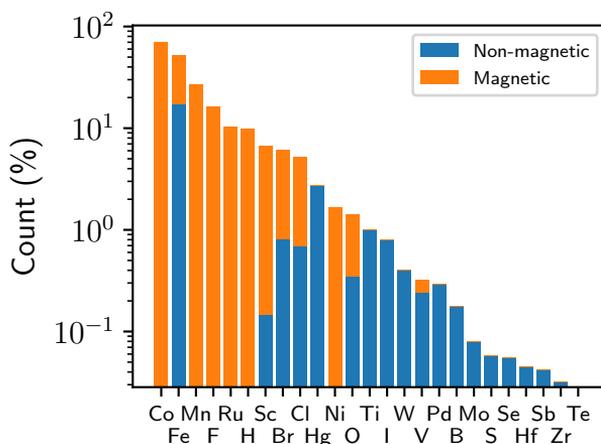


Figure 4.3: The percentages of QP-ic Z values for the indicated elements. Source: paper II.

In general then, we cannot say *when* QP-ic Z will occur, except that it is more likely for magnetic materials and in general occurs in about 3% of calculations.

The final thing we note is that while a QP-ic Z probably implies that the G_0W_0 eigenvalue is less trustworthy, a QP-c Z does not necessarily imply that the G_0W_0 eigenvalue is correct as we will see below.

4.4.3 Frequency-dependent Self-Energies

To try to investigate when G_0W_0 with the linear approximation to the quasiparticle equation might fail, we have calculated full frequency dependent self-energies for 12 of the 370 materials. This was done with a simple adaption of the current G_0W_0 implementation in GPAW which follows [57]. Two of these materials were chosen because they had unphysical Z ($Z \notin [0, 1]$) and the remaining 10 were chosen because all Z were physical ($Z \in [0, 1]$). In this way, we hope to get a somewhat representative sample. More data is always nice, but practical considerations restricted us to calculate the frequency dependent self-energy for only a few materials. The chosen materials are summarized in table 4.1.

In figure 4.4 and 4.5, a few self-energies are plotted.

Material	Prototype	Mag. state	PBE gap [eV]	G_0W_0 gap [eV]
HfBrI	MoSSe	NM	0.71	1.61
HfClI	MoSSe	NM	0.81	1.78
ZrBrCl	MoSSe	NM	0.91	1.88
ZrClI	MoSSe	NM	0.88	1.74
FeCl ₂	MoS ₂	FM	0.35	0.00
MnBr ₂	MoS ₂	FM	1.59	2.02
MoS ₂	MoS ₂	NM	1.58	2.53
PdSe ₂	CdI ₂	NM	0.56	1.61
Al ₂ Se ₂	Ga ₂ S ₂	NM	1.99	3.54
Ga ₂ S ₂	Ga ₂ S ₂	NM	2.32	4.08
Ga ₂ Se ₂	Ga ₂ S ₂	NM	1.76	3.44
In ₂ S ₂	Ga ₂ S ₂	NM	1.67	3.15

Table 4.1: A summary of the materials for which frequency-dependent self-energies have been calculated. Adapted from paper II.

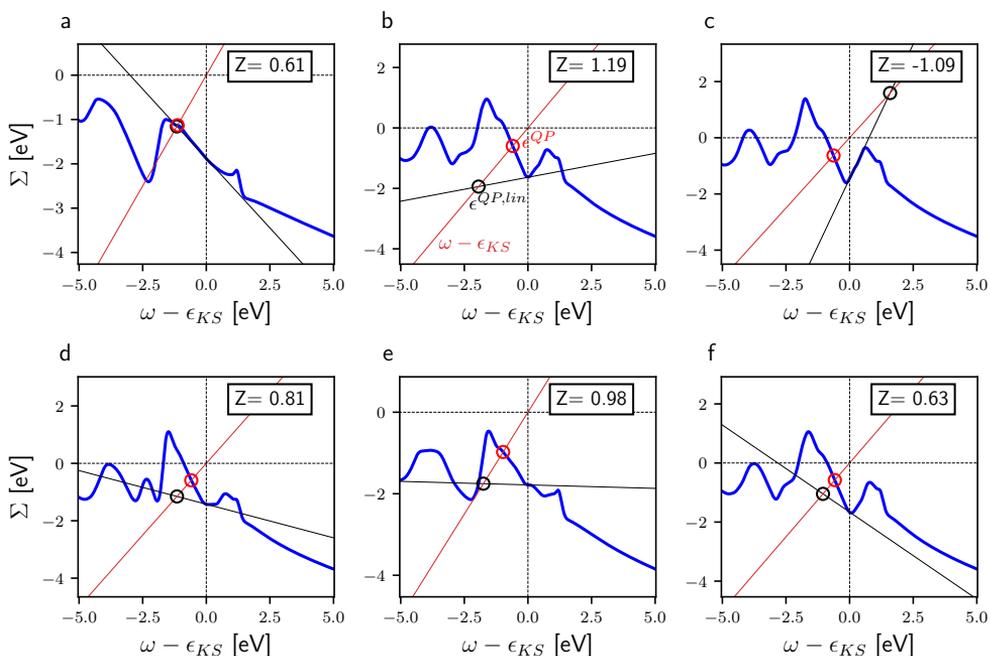


Figure 4.4: Spline-interpolated calculations of the frequency dependent self-energies for a few example states. The Z for each are indicated in the inset box.

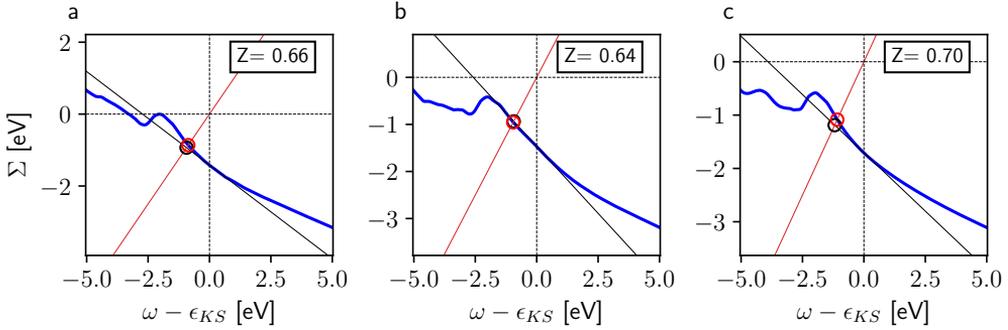


Figure 4.5: A set of frequency dependent self-energies that exemplify cases where the linear approximation is good.

To understand the figures, first recall the equation for the quasiparticle energy, eq. (4.9), in the G_0W_0 approximation

$$\epsilon_i^{\text{QP}} - \epsilon_i^{\text{KS}} = \langle \psi_i^{\text{KS}} | \Sigma_{G_0W_0}(\epsilon_i^{\text{QP}}) - V_{\text{xc}} | \psi_i^{\text{KS}} \rangle. \quad (4.38)$$

The solutions to this equation are given by the intersection a slope 1 line

$$f(\omega) = \omega - \epsilon_i^{\text{KS}}, \quad (4.39)$$

and the function

$$\Sigma(\omega) := \langle \psi_i^{\text{KS}} | \Sigma_{G_0W_0}(\omega) - V_{\text{xc}} | \psi_i^{\text{KS}} \rangle. \quad (4.40)$$

The slope 1 line is indicated by a red line in the plots and the blue curve denotes $\Sigma(\omega)$.

The true solution to eq. (4.9) is given by the intersection of the red line and the blue curve, which is highlighted by the red circles in the various plots. As is evident from some of the plots, sometimes multiple solutions exist. In these cases we have selected the solution closest to the Kohn-Sham energy. Since (4.9) rests on the assumption that the Kohn-Sham states are close to the quasiparticle states, this seems like the "best guess" and corresponds to the solution with the smallest G_0W_0 correction.

The solution afforded by the linear approximation, eq. (4.11), can be thought as the intersection between the red line and another linear function, as we can see from eq. (4.11):

$$\begin{aligned} \epsilon_i^{\text{QP}} - \epsilon_i^{\text{KS}} &= \langle \psi_i^{\text{KS}} | \Sigma_{G_0W_0}(\epsilon_i^{\text{KS}}) + (\epsilon_i^{\text{QP}} - \epsilon_i^{\text{KS}}) \Sigma'(\epsilon_i^{\text{KS}}) - V_{\text{xc}} | \psi_i^{\text{KS}} \rangle \\ &= \Sigma(\epsilon_i^{\text{KS}}) + \Sigma'(\epsilon_i^{\text{KS}}) (\epsilon_i^{\text{QP}} - \epsilon_i^{\text{KS}}). \end{aligned} \quad (4.41)$$

Looking at the last expression we see that it is simply the linear approximation at the Kohn-Sham energy to the full self-energy. This line is shown in black in the figures, and the corresponding solutions (red-black intersections) are shown with black circles. Finally, the Z values are also shown in the plots.

A couple of things are worth noting. First of all, sometimes, as in fig. 4.4a, the linear approximation does very well. In other cases, the linear approximation performs poorly, and these examples also highlight the point we made above, that a QP-c Z does not necessarily entail a good G_0W_0 eigenvalue (in the linear approximation). We also see that large errors seem to occur whenever the Kohn-Sham energy is close to a cusp in the self-energy.

However, these examples are hand-picked both to illustrate the different cases that can occur and to make nice-looking plots. In general, the linear solutions are quite good as we illustrate in figure 4.5. Below we will quantify this statement and look at the errors due to the linear approximation across the set of 12 materials.

4.4.4 Analysis of errors due to linear approximation

Let us now quantify the errors of the linear approximation. The error distribution for QP-c and QP-ic states are shown in figure 4.6. The distributions quantify the statement we made above, that QP-c states generally have a lower error. The mean absolute error (MAE) for QP-c and QP-ic is 0.04 eV and 0.27 eV respectively.

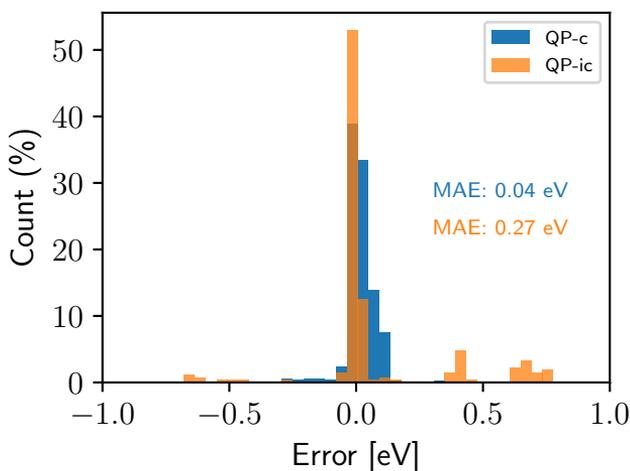


Figure 4.6: Errors arising from the linear approximation to the QPE for QP-c and QP-ic states. Source: paper II.

The next question we try to answer is whether there are any simple and efficient schemes for reducing this error.

First of all, the linear approximation can also be viewed as one iteration of the Newton-Raphson (NR) method. To find the true solution one could simply iterate NR multiple times until the energy is converged, however it would be preferably not to do this as it would reduce computational efficiency. We will, however, investigate the effects of one additional iteration of NR, as this might not be too expensive.

The second method we have investigated is motivated the narrow distribution of Z

values, see again figure 4.2. As we noted above, large errors can occur when the Kohn-Sham energy is at a cusp in the self-energy. Since the Z values are narrowly distributed, one could hope that replacing the calculated Z value with the empirical mean of Z , namely $Z_0 = 0.75$. A slightly more refined method is to only replace the calculated Z for QP-ic states, as these are the ones we expect to have large errors. We have dubbed this method "empirical Z " or "empZ" since it uses the empirical distribution of Z values. When only applied to QP-ic states we call it "empZ@QP-ic".

The final method we show here is the " ΣdE ", which gives the correction to the linear solution as

$$\delta = \Sigma(\epsilon^{\text{QP, lin}}) - \left(\Sigma(\epsilon^{\text{KS}}) + \Sigma'(\epsilon^{\text{KS}})(\epsilon^{\text{QP, lin}} - \epsilon^{\text{KS}}) \right). \quad (4.42)$$

Here $\epsilon^{\text{QP, lin}}$ is the quasiparticle energy in the linear approximation, note that this means we have to evaluate the self-energy at one extra point in addition to the Kohn-Sham energy. The motivation for this expression is that it approximates the true solution by essentially approximating Σ with a second order polynomial using points at $\omega = \epsilon^{\text{KS}}$ and $\omega = \epsilon^{\text{QP, lin}}$. We have also tested polynomial approximations using points near ϵ^{KS} but none showed any systematic improvement in the solutions.

The errors after applying these methods are shown in figures 4.7 and 4.8. The plots only show the main range, and some outliers are omitted. The single iteration Newton-Raphson (NR1), which is equivalent to the linear approximation, is shown in each plot to facilitate comparison with the standard linear approximation.

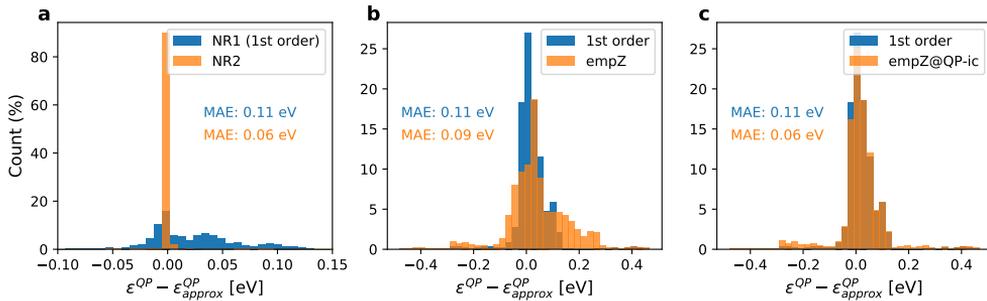


Figure 4.7: Error distributions after applying error correcting schemes. NR-X stands for X iterations of Newton-Raphson where NR-1 corresponds to the linear approximation to the QPE. The empirical Z (empZ) method is described in the main text. Source: paper II.

NR2, figure 4.7a, improves the error distribution and reduces the MAE significantly, as one would expect.

The empirical Z method, shown in figure 4.7b, despite being so simple and despite adjusting *all* Z , actually works well if you consider the overall MAE. However, inside the main range, where most points fall, it seems to produce larger errors. When we

instead apply empZ only to the QP-ic states, as shown in figure 4.7c, the errors in the main range are virtually unchanged but the overall MAE is reduced, showing that the outliers are reduced by the empZ@QP-ic method. This is quite promising because empZ requires no additional calculation and can be applied as a post-processing step to a regular G_0W_0 calculation.

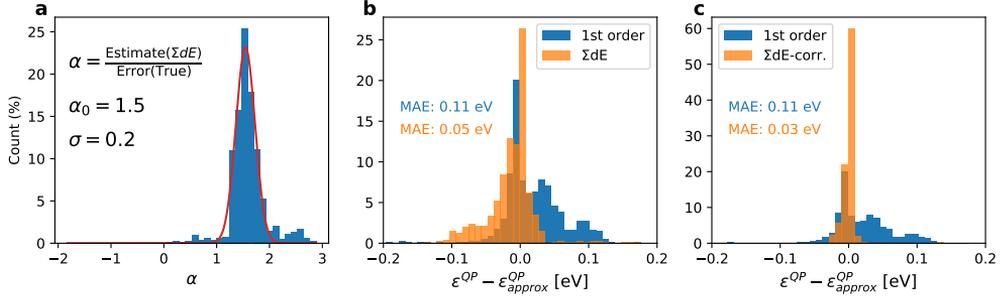


Figure 4.8: (a) Shows the distribution of the "overshooting" factor for ΣdE , α , defined in the main text. (b) and (c) show the error distributions of ΣdE and the corrected ΣdE . Also shown for comparison, in blue, is the errors for the linear approximation to the QPE. Source: paper II.

The results of the ΣdE correction scheme is shown in figure 4.8b. Errors are generally reduced but there seem to be some systematic overcorrection, since the peak in the main range is shifted. This motivated correcting the ΣdE by considering the empirical overcorrection values $\alpha = \text{Error}(\Sigma dE)/\text{True error}$. Note that we have one α for each state. The distribution of these errors is shown in figure 4.8a. We fitted a gaussian, shown in red, to the distribution and extracted the average value, α_0 . Defining a corrected ΣdE as

$$\delta_{\text{corrected}} = \delta / \alpha_0, \quad (4.43)$$

we get the errors shown in figure 4.8c, showing a significant improvement. To verify that there were no systematic biases introduced by calculating α_0 from the same dataset it was applied to, we split the dataset randomly into two sets of equal size, the "training" set and the "test" set. We then calculated α_0 from the training set and applied it to the test set. The entire procedure was repeated multiple times and the MAE was always in the range 0.02 - 0.03 eV. We also tested on dataset of varying size, whence the error was < 0.03 eV as long as more than $\sim 5\%$ of the points were used in the training set. In the future it may be desirable to do a more statistically rigorous test.

A summary of the results are shown in table 4.2. The number of required self-energy evaluations plus the number of required quasiparticle weight evaluations for each method is also shown in the table. Since Z can usually be acquired at negligible computational cost if one calculates Σ and the empZ@QP-ic just requires Σ and Z , this method seems the most promising. One can use it to correct energies that are likely to be large at essentially zero additional cost compared to a standard G_0W_0 calculation.

Method	MAE [eV]	# Σ/Z evals
1st order	0.11	2
empZ	0.09	1
empZ@QP-ic	0.06	2
Σ dE	0.05	3
Σ dE-corr.	0.03	3

Table 4.2: Mean absolute errors (MAE) and effective number of self-energy (Σ) evaluations for the various methods proposed in the main text. Adapted from paper II.

The empZ@QP-ic method has been implemented as part of a recipe in ASR (see paper IV). A prominent example where G_0W_0 fails is MoS₂ and standard G_0W_0 bandstructure along the the empZ@QP-ic bandstructure is shown in figure 4.9.

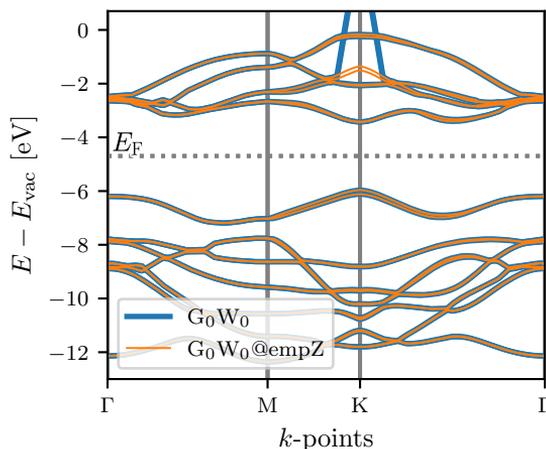


Figure 4.9: Standard G_0W_0 bandstructure for MoS₂ is shown in blue. The empZ-corrected bandstructure is shown in the orange. Adapted from paper III.

4.4.5 The plane-wave extrapolation and scissor-operator approximation

In paper II we also studied the effects of the plane-wave extrapolation and the scissor-operator approximation. We will not repeat the results here, but merely mention that the plane-wave extrapolation scheme is generally well-founded, at least judging from the linearity measures of the self-energy and the derivative of the self-energy as a function of plane-waves.

The scissor-operator approximation, where the G_0W_0 correction is calculated for a single k -point and this correction is applied to all other k -points, is generally not valid, unless relatively large errors on the order of an electronvolt, are acceptable.

4.4.6 Summary of Paper II

In summary, we discovered no general trends in when quasiparticle energy calculations fail, except that magnetic materials seem harder to describe. We also proposed a post-processing error correction method for the eigenvalues which was based on whether the calculated Z value was *quasiparticle-consistent* or not. We also found that plane-wave extrapolation is a valid approach but the scissor-operator approximation generally produces large errors.

For the purposes of high-throughput studies, the workflow could include a post-processing step that applied the empirical Z method to get the most out of the calculations and circumventing the need for laborious convergence studies which may not be desirable in screening studies. It is generally not advisable to use the scissor-operator approximation to cut down on calculation time, but plane-wave extrapolation should be used and it might even be possible to use lower cutoffs than that used in the C2DB. In the future it might be interesting to study whether this statement can be made rigorous. It would also be interesting to see how the results presented here change when separating materials into classes that are physically distinct. For example, it would be interesting to look at a large set of magnetic materials compared to a large set of non-magnetic materials, or materials that are known to be strongly correlated vs. less strongly correlated materials.

5 Automation of Material Simulations

Studies in material design are beginning to involve a wide range of physical properties spanning from Raman spectra to polarizability to stiffness calculations. The large span of properties touch a variety of areas of physics and it becomes prohibitively difficult for any single researcher to be an expert in all these domains. One solution is to reduce the number of properties that one investigates, but often it is desirable to get as much out of the data as possible. Another option is to work in large collaborations so as to include experts on every domain of interest.

A different approach, which is taken by the *Atomic Simulation Recipes* (ASR) python package described in paper IV, is to formulate different property calculations as relatively small, isolated calculation steps called *recipes* which are implemented as python modules. A recipe might, for example, perform a ground state calculation or a structure optimization. The recipes are implemented by appropriate domain experts and, if the recipes are sufficiently robust and require little user-input, end-users can leverage the expert knowledge independently of the experts. This can greatly extend the reach of any single researcher but also increases the utilization of the expert's knowledge beyond what the expert can achieve on their own with only 24 hours in day.

ASR was conceptualized, largely single-handedly, by Morten Gjerding who also implemented the core functionality.

The chapter starts with an overview of the ASR infrastructure and a brief comparison to other, similar tools. The remainder of the chapter deals with my principal contributions to ASR, namely recipes for enhanced stability analysis and effective mass recipes. Other contributions are also briefly mentioned but the bulk of the details are left to other chapters to avoid needless repetition or confusing cross-referencing. The ASR package is the subject of paper IV. Finally, the results of papers I and III, both of which used ASR, are discussed.

5.1 Introduction to ASR

ASR is essentially a framework for *organizing* electronic structure calculations and therefore it does not implement the electronic structure methods themselves but rather interfaces with external codes that do the computations. Currently only GPAW [18] is supported but other codes such as ABINIT [62] or Octopus [63] could in principle also be used. The interface with external codes happens either directly with the codes or through ASE [64]. Recipes in ASR handle setup, management, analysis, pre- and post-processing in a close to fully-automated fashion. The user accesses the recipes and the results through various user interfaces: the command line, via python scripting, or by using the app module of ASR. Various tools are also implemented to perform data migration and work with databases.

This infrastructure is shown in figure 5.1.

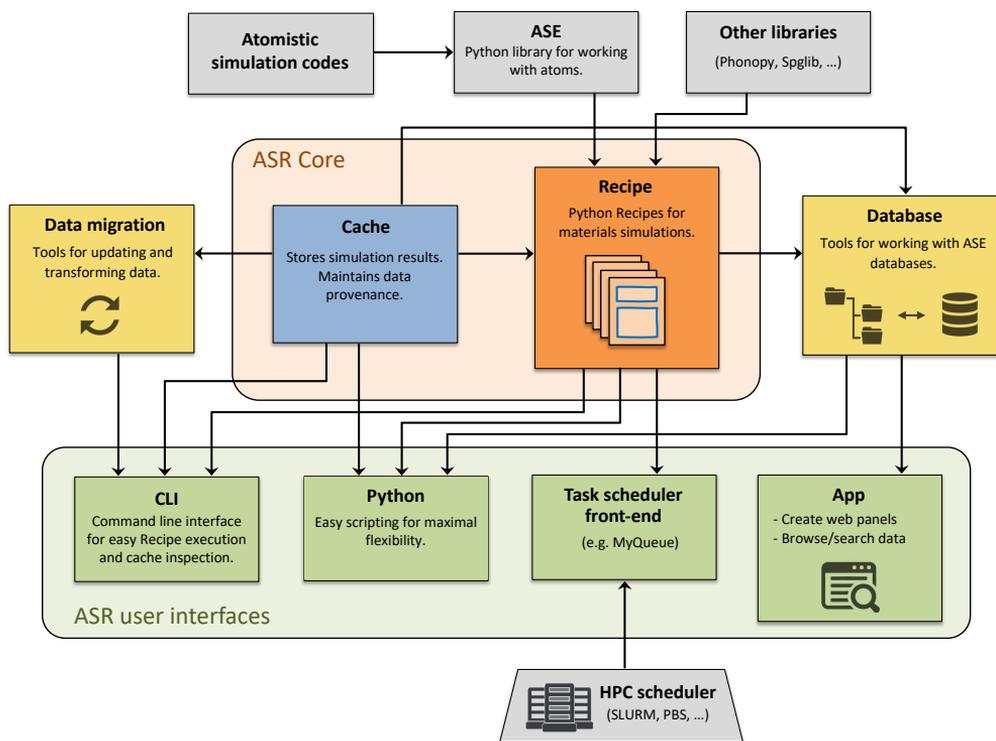


Figure 5.1: The architecture of ASR. Source: paper IV.

As mentioned above, isolated property simulations are defined in so-called *recipes*. Recipes are simply python modules that perform some well-specified, generally useful task. An example of a simplified recipe that calculates the ground state energy by directly interfacing with GPAW is shown in the following code listing.

```

1 from asr.core import command, option, AtomsFile
2 from ase import Atoms
3
4 @command('asr.gs')
5 @option('-a', '--atoms', type=AtomsFile, default='structure.json')
6 def main(atoms: Atoms):
7     from gpaw import GPAW
8     calculator = GPAW(mode=PW(400), xc='LDA', txt='asr.gs.out')
9     atoms.calc = calculator
10    e = atoms.get_potential_energy()
11    return e

```

Such a recipe can be run on the command line with the command

```
$ asr run gs
```

A recipe can equally well be submitted to a job scheduler such as slurm. One can for example use myqueue [65]:

```
$ mq submit asr.gs -R 40:1h
```

which submits the `asr.gs` recipe to 40 cores with a maximum time limit of 1 hour.

ASR automatically intercepts the return-result and serializes it to a json-file which can subsequently be used to inspect the data. Generally, the author of a recipe is encouraged to document the return result of a recipe by defining an `ASRResult` class and returning an instance of this. This also facilitates reading and using the results in a script.

Obviously, this is not an exhaustive explanation of the various interfaces to ASR nor is a full description of what a recipe can do. An important thing we have neglected in this discussion is how to write a webpanel to display the results in a web browser. Such a webpanel can include plots, tables, and text descriptions of various elements. Examples abound on the C2DB website [66] and we show one material in figure 5.2.

The preceding discussion is attempt to explain the strength of ASR: with the correct formalization of a property calculation and the powerful ASR infrastructure, running and analysing complex simulations becomes significantly easier, allowing the researcher to focus on the science rather than the implementation. The main point of ASR is then a strict and robust specification of how to calculate a given property. This greatly aids reproducibility and tools that allow for data provenance tracking are also implemented in ASR; ASR adheres to the FAIR principles [67].

Other research groups have developed tools that address similar problems.

`atomate` [68] has a large overlap with ASR. Like ASR, `atomate` defines specific calculation tasks (recipes in ASR, workflows in `atomate`). Unlike ASR, `atomate` is also a job manager that handles submission of jobs to a compute cluster. ASR is agnostic about job management and CAMD has been using `myqueue` [65]. A downside of `atomate` is that it relies on the licensed software VASP [68–70] whereas ASR is fully open-source.

`AiiDA` [71, 72] is another software package in the same space. `AiiDA` seems to be more code agnostic and is more focused on workflow management, data provenance, and data management. A weaker coupling to specific implementation (code agnosticism) can be a huge strength when testing and benchmarking different versions of a method as also noted in [71]. On the other hand, if one is doing standard methods code agnosticism may lead to additional up-front work.

In summary, there seems to be large overlap between ASR and other solutions but scopes and approaches differ. ASR is focused only on doing, storing, and displaying calculations whereas other packages also include a HPC job management component. ASR is generally for doing specific, well-defined, well-tested tasks in an efficient, easy

way. This makes it less suited for method-development but ideal for almost all other materials science tasks.

O₈Te₄

Download raw data

Browse raw data

C2DB
The Computational 2D Materials Database

Summary

Thermodynamic stability

Stiffness tensor

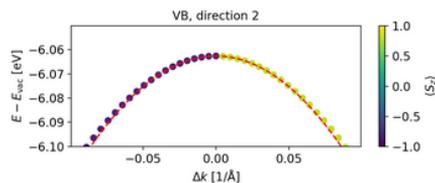
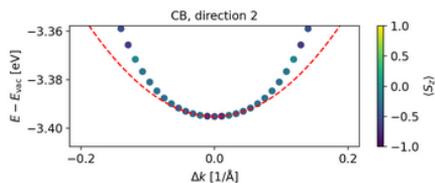
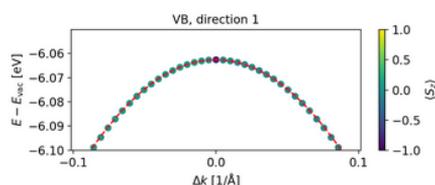
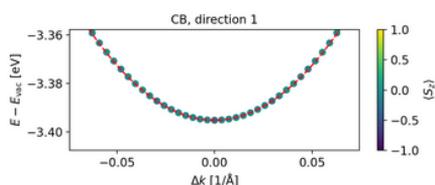
Phonons

Basic electronic properties (PBE)

Electronic band structure (PBE)

Projected band structure and DOS (PBE)

Effective masses (PBE)



Electron effective mass	Value	MARE (25 meV)
CB, direction 1	0.42 m_0	0.006 %
CB, direction 2	3.58 m_0	0.139 %
CB + 1, direction 1	0.42 m_0	0.006 %
CB + 1, direction 2	3.57 m_0	0.138 %

Hole effective mass	Value	MARE (25 meV)
VB, direction 1	0.79 m_0	0.002 %
VB, direction 2	0.75 m_0	0.004 %
VB - 1, direction 1	0.78 m_0	0.002 %
VB - 1, direction 2	0.75 m_0	0.004 %

Figure 5.2: An example of an ASR webpanel. Source: paper IV.

5.2 Enhanced Stability Analysis

In computational design of new materials, robust theoretical assessments of material stability is obviously important: if the proposed material is unstable it can be hard or impossible to manufacture in the lab. Two approaches commonly employed are *thermodynamic* and *dynamic* stability measures which describe whether the material is

energetically and structurally stable, respectively. In this section we describe a set of recipes in ASR which allow for an additional stability assessment to be made, namely whether the material is stable in the presence of a reservoir of other materials. The most relevant example is an Oxygen reservoir which often occurs in Earth's atmosphere wherein most labs are located.

We start with a brief review of the standard stability measures to set the stage and introduce relevant definitions.

5.2.1 Thermodynamic Stability

One thing that is often important, is to determine whether it is energetically favorable to form the material in question. The answer to this is given by the *heat of formation* which, in C2DB and for our purposes, is the change in energy per atom when producing the material from the reactants. The reactants are always the *standard phases* of the given reactant. If we take MoS₂ as an example, we would compare the energy of the MoS₂ crystal with those of bulk crystals of Mo and S. These energies are calculated at zero temperature with DFT which is usually good enough. This discussion leads us to define the heat of formation, E_{hof} , as

$$E_{\text{hof}} = \frac{1}{N} \left(E_{\text{product}} - \sum_{r \in \text{reactants}} E_{\text{standard } r} \right). \quad (5.1)$$

We see that a negative heat of formation implies that is energetically favorable to have the product rather than the reactants. However, and this is a big however, in the presence of other possible products stability is not implied by a negative heat of formation. Instead we have to consider the *convex hull* of possible products.

A convex hull of a set of points is the smallest convex set that contains those points. Intuitively it corresponds to taking an elastic sheet (as from a trampoline) and stretching it as much as possible over the points, see figure 5.3. Mathematically, a convex set, C , is a set

$$\forall p_1, p_2 \in C : tp_1 + (1 - t)p_2 \in C \forall t \in [0, 1], \quad (5.2)$$

i.e. given two points p_1, p_2 from C all points between them are in C as well. The hull for a set of points, as mentioned, is the smallest such set. As illustrated in the figure it contains the "outermost" points.

To explain the relation of the convex hull to stability considerations, let us take as an example Ir₂S₂ (<https://cmrdb.fysik.dtu.dk/c2db/row/Ir2S2-bf9a5b4e8e88>). We want to compare the energy of Ir₂S₂ to that of Ir and S, so we define an x-axis that indicates the content of S, say. Various points on the x-axis will correspond to different materials of the form Ir_{1-x}S_x. On the y-axis we plot the energy of the given material. By definition, the reference phases Ir and S have zero heat of formation. Because, as one can calculate, Ir₂S₂ has negative heat of formation it will be below the two points corresponding to Ir and S, see figure 5.4. However, multiple other Ir-S compounds exist. If we for simplicity consider only Ir₄S₆ located at $x = 6/10 = 0.6$ we see that *the lower edge of the convex hull lies below Ir₂S₂*. This is emphasized because if we have

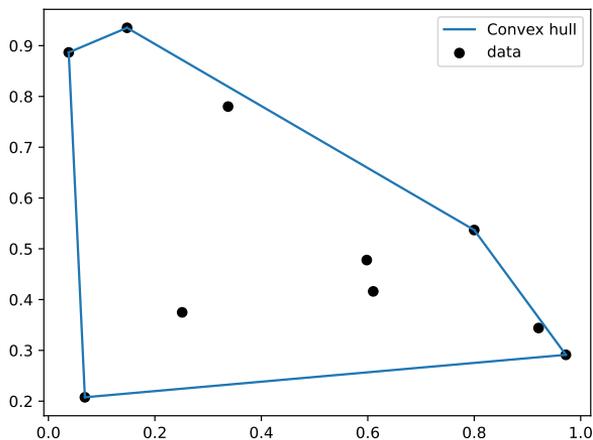


Figure 5.3: An illustration of the convex hull concept. The convex hull is the smallest convex set enclosing the data.

any material corresponding to $x \in [0.0, 0.6]$ it will always be energetically favorable to form Ir (in the reference) and Ir_4S_6 rather than, say, Ir_2S_2 . This is because the lower edge of the convex hull is the energy of a particular linear combination of Ir and Ir_4S_6 , refer to the definition of the convex hull. The relative amounts of Ir and Ir_4S_6 that is formed is, of course, determined by x .

We see that for a 2-component system, the hull is 1-dimensional. In general for an N -dimensional system the hull is $N - 1$ -dimensional. In higher dimensions we are still interested in the lower edge of the hull, where the "lower" edge is in general defined as the surfaces with outward normals that have a negative component in the energy-direction. In the 1D example the energy direction was along the y -axis.

Strictly speaking a material is stable when it lies on the hull. However, practically speaking it is possible to fabricate materials that are not on the hull, i.e. *metastable* materials. In C2DB a classification scheme for thermodynamic stability was defined based on experimentally synthesizable monolayers [6]. If the material has a heat of formation larger than 0.2 eV/atom the material is considered to have "low" stability. A heat of formation *relative to the convex hull* of less than 0.2 eV/atom is considered to have "high" stability. Materials that fall into neither of these classes have "medium" thermodynamic stability. There are several reasons to not impose the strict stability requirements mentioned above. First of all DFT has limited accuracy so materials that are predicted to be a little bit above the hull may in fact be on the hull. Secondly, materials are often fabricated on substrates which can dramatically change the preferred structure. Additionally, reaction barriers may be large such that a given material is stable for practical purposes [6].

As a final note, the stability assessment discussed above is limited by the number of materials one compares to. It is possible that the convex hull can be lowered by adding more references, but it can never be raised. The stability assessments are always best guesses or, to put it another way, upper bounds.

5.2.2 Dynamical Stability

The dynamical stability is not strictly relevant for the work in this thesis but we include a brief explanation for completeness and context.

Dynamical, or structural, stability is based on the phonon modes of the material. A full description of phonon physics is outside the scope of this thesis but essentially one considers various displacements of the atoms in the structure. If the forces tend to restore the structure to its original configuration, the structure is dynamically stable. However, if there exists a displacement mode where the forces are not restoring, the structure is unstable. The relation to phonons is of course that phonons are oscillatory atomic displacement modes. DFT calculations at these displacements allow one to calculate the *dynamical matrix* which essentially the Hessian of the energy with respect atomic positions.

Even though the structures in C2DB are structurally optimized, the optimal structure may not be a minimum energy structure because imposed symmetry and limited unit cell size may constrain the optimization such that a saddle point rather than a minimum is found [6].

As with thermodynamic stability three different stability levels are defined. The eigenvalues of dynamical matrix correspond to the squared frequencies of the phonon modes. A negative eigenvalues will then correspond to an imaginary frequency, which indicates an unstable mode. However, as with the heat of formation, some experimentally known materials would be predicted to be unstable if we just use the sign of the eigenvalues. Therefore the definition of the three dynamical stability levels in C2DB are are guided by calculations done on experimentally synthesized monolayers. Thus "low" stability corresponds to a minimum eigenvalue of the dynamical matrix of less than -2 eV/\AA^2 and "high" corresponds to a minimum eigenvalue greater than $-10^{-5} \text{ eV/\AA}^2$. "Medium" are those materials that are not "low" and not "high" [6].

5.2.3 FERE

Accurate prediction of a materials stability is obviously of fundamental importance when doing computational materials design but DFT is known to have certain shortcomings. There are errors both when calculating the energies of the compound phase (e.g. Ir_2S_2) and when calculating the energies of the reference phases (e.g. Ir, S). When the compound and the references are of the same type (e.g. both are metals) these errors tend to cancel. However, if the compound and references have different character large errors in the heat of formation can occur [73, 74]. For example, metal oxides have quite different bonding than in pure oxygen and in the pure metal [74].

In [73] a method was proposed to alleviate this problem: *Fitted Elemental Reference Energies* (FERE). The idea is to adjust the energies of some or all of the reference states by fitting certain heats of formation to experimental heats of formation. This is

done by finding a least-squares solution to the set of equations

$$\begin{aligned}
 \text{Reaction 1: } E_{\text{exp-hof}}^1 &= \frac{1}{N_1} \left(E_{\text{product}}^1 - \sum_{r \in \text{references}} E_r^1 \right) \\
 \text{Reaction 2: } E_{\text{exp-hof}}^2 &= \frac{1}{N_2} \left(E_{\text{product}}^2 - \sum_{r \in \text{references}} E_r^2 \right) \\
 &\dots \\
 \text{Reaction n: } E_{\text{exp-hof}}^n &= \frac{1}{N_n} \left(E_{\text{product}}^n - \sum_{r \in \text{references}} E_r^n \right) \tag{5.3}
 \end{aligned}$$

The left-hand sides are the experimentally measured heats of formation. These are typically done under ambient temperature and pressure which is hard to simulate in DFT, but the difference in energy from zero temperature and pressure is usually negligible [73]. The right-hand sides consist of calculated energies of the products, E_{product}^j and the reference energies, E_r^j . The reference energies are then adjusted to match the left-hand side. Because the same reference usually occurs in multiple reactions it is in general not possible to exactly satisfy all equalities which is why a least-squares solution is sought.

The reference energies thus found are not physical, i.e. they cannot be used as accurate energies for the reference phases. Rather they contain a bit of the error present in the product so as to cancel this error [73]. This procedure was shown to greatly decrease the mean absolute error (MAE) of the heat of formation from about 250 meV/atom to 54 meV/atom [73].

As part of the work done for this thesis on enhanced stability analysis FERE was implemented in ASR as the recipe `asr.fere`. The user supplies a list of elemental references to adjust, a list of target heats of formation, and finally a ASE database containing materials. The output is a database with FERE heats of formation. Since the core algorithm is just least-squares it is not of great interest to discuss further here but the recipe was used in the work described below that deals with the presence of reservoir.

5.2.4 Stability in the Presence of a Reservoir

A question that is not usually addressed in stability assessments is how stability is affected when the material in question is placed in contact with a reservoir of some other chemical. The most commonly encountered reservoir is the atmosphere which besides relatively inert gasses, contains oxygen which is highly reactive. A very relevant question for experimental synthesis is therefore whether the material will oxidize when exposed to the atmosphere. The question can be generalized to whether a material will, when in contact with a reservoir containing some type of particles, react with those particles and form new compounds. This process essentially destroys the original material, so it is generally not desirable. Besides oxygen, it could be useful to know whether one's material will react with other gasses if one for example is doing chemical vapor deposition on top of an existing material.

So far we have been dealing with fixed stoichiometries, that is, fixed amounts of the various components (e.g. Ir, S). When dealing with a reservoir, there is an essentially infinite amount of the reservoir-compound available but the tendency of the reservoir-compound to interact with the experimental-compound is not infinite and can be described by the *chemical potential*. The chemical potential essentially describes the energy of the reservoir-compound when it is inside the reservoir. A very low chemical potential thus qualitatively means that there is a very low reactivity between the reservoir and the experimental system and similarly for a high chemical potential. In general, the governing principle is that in equilibrium the total free energy, $\Phi = E - \mu_r N_r$, is minimized. Here μ_r stands for the chemical potential of the reservoir and N_r for the amount of reservoir-compound. The internal energy E can be minimized by reacting or disintegrating different parts in the system ($\text{Ir}_2\text{S}_2 \leftrightarrow 2\text{Ir} + 2\text{S}$). However, the reaction paths that are available will depend on the amount of reservoir-compound N_r . This makes the problem slightly more complicated to think about, but thankfully the convex hull construction described above can make the minimizing solution easy to understand.

We consider again figure 5.4 but now we imagine that the amount of Ir is fixed but it is in contact with a reservoir of sulphur gas. We now interpret the height of the sulphur gas circle (S_{48} in the figure) as the value of the chemical potential and thus it is no longer fixed to be zero but for now we assume it is still zero.

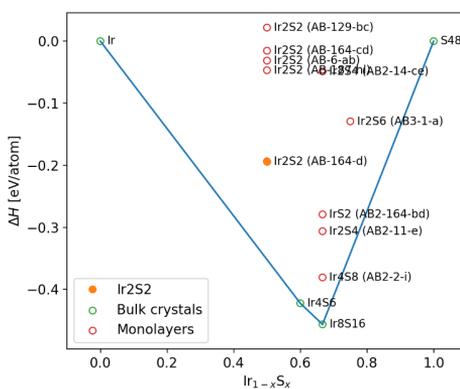


Figure 5.4: The convex hull for Ir-S. Adapted from [66].

For any given sulphur concentration x the minimum energy configuration is given by the convex hull as we discussed before. However, the equilibrium state is now found by additionally minimizing over the convex hull. To realize this we note that adding additional sulphur atoms is essentially "free" (because we temporarily assumed for simplicity that the chemical potential is zero), so if we can decrease the height of the hull by changing the sulphur concentration this will be favorable with respect to the free energy. Thus the equilibrium state of Iridium in contact with a sulphur reservoir is Ir_8S_{16} as we can see from the figure.

When the chemical potential is non-zero there is a "cost" in free energy incurred by changing the sulphur concentration; the cost is just the negative chemical potential, $-\mu_S$. However, we can incorporate this cost into the convex hull by simply adjusting the height of each point on the convex hull by $-\mu_S x$ where x again is the sulphur concentration. In this adjusted convex hull the free energy minimizer is again the minimum on the hull.

The above construction obviously generalizes to any reservoir-compound and the generalization to higher dimensions is also straightforward. As an example, let's go one dimension up and consider MoTe_2 in the presence of O_2 . The Mo to Te content is fixed to the ratio 1:2 but the oxygen content is variable since it is sourced by a reservoir. As we now have 3 different elements, the convex hull is two-dimensional as discussed above. One normally plots the convex hull view from above, where the edges joining adjacent faces of the convex hull are shown as lines, see figure 5.5a. Obviously this destroys any information about the height of the hull, i.e. the heats of formation. The equilibrium state is then found by minimizing along the one-dimensional subspace on the convex hull that goes from pure MoTe_2 to pure O_2 , see the dashed lines in figure 5.5b. To determine whether the material is stable in oxygen we need to determine the height of the convex hull at the intersections between the dashed line and the hull edges, these are shown as red crosses in figure 5.5b. We don't need to calculate the intermediate energies because of the way the convex hull is constructed: it is the minimal convex set containing the points thus the surfaces that join the edges are flat planes. If we then plot the height of the convex hull along the dashed line we get figure 5.5c. Along this line we minimize the free energy $E - \mu N_{\text{O}_2}$ and we get the point at O content ≈ 0.75 . From figure 5.5b we can see that this is a combination of Mo_8O_{24} and O_{10}Te_4 , and we conclude that this is the equilibrium state (for this chemical potential of O_2). In other words, we predict that if you expose MoTe_2 to air it will oxidize.

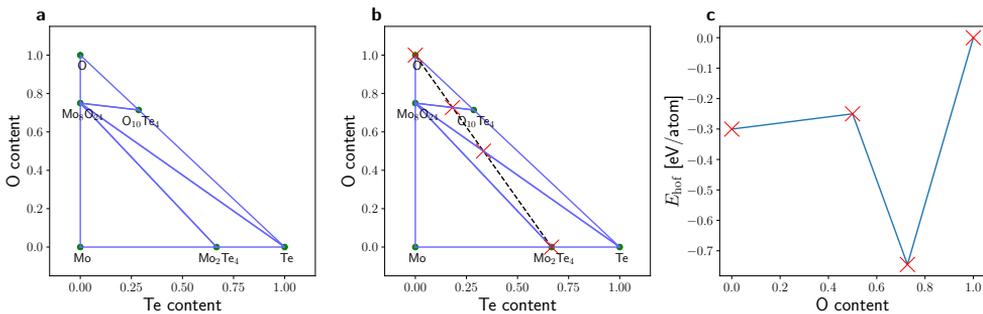


Figure 5.5: (a) shows a convex hull for Te-O. (b) shows a line indicating the relevant material compositions when Mo_2Te_4 is in contact with an oxygen reservoir. (c) shows the height of the convex hull when following the dashed line in (b).

It is possible to adjust the chemical potential, and the most accessible way to change the chemical potential of a gas is to change the pressure. An obvious question is then

how much we would have to reduce the chemical potential of O_2 to make $MoTe_2$ stable. The $Mo_8O_{24} - O_{10}Te_4$ mixture that makes up the equilibrium state has an oxygen percentage of ≈ 0.75 , which means that the free energy of this state changes linearly with the oxygen chemical potential with a slope of -0.75 . Thus we can easily calculate the change in chemical potential needed to stabilize $MoTe_2$:

$$\Delta\mu = \frac{-0.3 \text{ eV/atom} - (-0.7 \text{ eV/atom})}{-0.75 \text{ atom}^{-1}} \approx -0.5 \text{ eV}. \quad (5.4)$$

We call this the μ -*adjustment* measure which is one way to quantify how difficult it is to make a material stable or how likely it is oxidize. Below we discuss some evidence that this measure correlates with actual stability in experiments.

The procedure described above is implemented in ASR as the recipe `asr.chc`, where "chc" stands for Convex Hull Cut which refers to the one-dimensional cut along which we minimize to find the equilibrium state.

5.2.5 Summary of results from Enhanced Stability Analysis

The work discussed above was done in collaboration with Morten Gjerding. However, we did not find any application that we felt warranted its own publication so the results we discuss here are not unpublished.

The first question we tried to address was whether the μ -adjustment actually correlates with real-world stability. One would certainly expect that the measure as defined is relevant for stability, since it is simply related to the free energy gain a mixture can get by reacting with Oxygen or whatever the reservoir material is. In this way it is on a similar basis as the other, standard stability measures as there is, as far as I am aware, no systematic comparison of these vs. experimental stability.

One reason for this could be that materials are never fabricated in an "ideal" way, meaning, first of all, that materials may have defects or consist of an uncontrolled mixture of competing states due to technical limitations. Secondly, materials are often fabricated on a substrate which can greatly influence stability. Thirdly, that a material has been synthesized and measurements have been performed does not necessarily mean that it is stable: it may rapidly degenerate or it may necessitate special protection and information like this is not always reported in literature (because it is "common" knowledge for experimentalists). Finally, even if a material is "unstable" it may degenerate only over many hours, days or months and in such cases it is not clear what demarcates stable materials from unstable ones.

Nevertheless, we attempted to go through the literature and determine if there was a correlation between the μ -adjustment and the reported stability. The results are reported in table 5.1. Note that it is not always known which structure has been synthesized. In those cases and where we have multiple theoretical structures we report a high/low for the μ -adjustment. Recall that higher is better (more stable) and a negative μ -adjustment means the material is strictly speaking unstable. The table thus shows best and worst case $\Delta\mu$, predicted stability based on $\Delta\mu$, experimentally reported stability, and references.

Before discussing the results we will go through each item and list the reasons for the reported experimental stability as it is not always completely clear how one should interpret the experimental report.

- PtSe₂: Stable

[75]: does not mention anything about oxidation or special synthesis requirements such as ultra-high vacuum. This is not in itself conclusive.

[76]: reports "air-stable" PtSe₂.

- hBN: Stable.

[77]: reports "strong oxidation resistance" of BN nanosheets.

- GaN: Stable

[78]: "evident upon ... exposure of 2D GaN to air, where no detectable oxidation ... is observed".

[79]: A study of oxidation of GaN where they have to use high temperatures (700-1000° C) to oxidize the material which suggest stability at room temperature.

- Ge: Unstable

[80]: First principles study reporting a low barrier and exothermic reaction with O₂.

[81]: Uses synthesis methods similar to silicene, i.e. ultra-high vacuum which suggests instability in air.

- MoS₂: Stable

[80]: "MoS₂ shows good stability in air at room temperature...".

[82]: Oxidation of MoS₂ edges occur rapidly but long exposure times are required for oxidation beyond that.

- P₄: Unstable

[83]: Black phosphorous "tends to be oxidized in the atmosphere".

[84]: Study of oxidation mechanisms of P₄.

- Si₂: Unstable

[81]: reports that Si₂ requires ultra-high vacuum to synthesize.

[85]: "A dramatic change in the silicene composition can be deduced already after 3 min exposure to air." However exposure to pure oxygen showed little reactivity so perhaps the picture is unclear, e.g. water could play a significant role.

- VS₂: Stable

[86]: Synthesization of VS₂ with no explicit prevention of oxidation.

Mater.	$\Delta\mu$	Pred. Stab.	Exp. Stab.
PtSe ₂	-0.3/-1.0	Unstable	Stable
hBN	-0.7	Unstable	Stable
GaN	-2.15	Unstable	Stable
MoS ₂	-0.55	Unstable	Stable
VS ₂	-0.79/-1.47	Unstable	Stable
Ge	-3.19	Unstable	Unstable
P ₄	-2.71	Unstable	Unstable
Si ₂	-5.09	Unstable	Unstable

Table 5.1: Reference materials with predicted stability from the convex hull cut and experimentally found stability. See main text for experimental references.

[87]: Another study reporting synthesization of VS₂ with no explicit prevention of oxidation.

As we can see it is not always clear from an experimental report whether the material is stable or not. Additionally, from table 5.1 we see that basically all materials are predicted to be unstable, a point we will see repeated later. The reason is probably that oxygen is highly reactive and most things will react with oxygen given the chance. However, reaction barrier heights are also relevant for determining whether a material will react on short timescales and our measure does not provide any information on the reaction barrier height.

We do see a rough, qualitative correspondence between low $|\Delta\mu|$ and stable materials, and high $|\Delta\mu|$ and unstable materials. For example one could choose $|\Delta\mu| < 2.5$ or $\Delta\mu > 0$ as likely indicating stability. Of course the dataset is quite small so it is hard to form a strong conclusion, but these results combined with the strong theoretical basis for the $\Delta\mu$ -measure (it is just basic thermodynamics) suggests that it could be useful, at least to separate candidate materials into groups that are likely/unlikely to oxidize.

Next we will consider the $\Delta\mu$ for oxygen calculated for the materials in C2DB. A histogram is shown in figure 5.6 which shows the distributions using the uncorrected references (blue) and the FERE corrected references (orange). The picture is largely the same: most materials are strictly speaking unstable which is probably to be expected given oxygens high reactivity. If we use the cutoff of -2.5 instead we see that, using FERE, around half of materials are stable.

In general the differences between the uncorrected and FERE distributions show how important it can be to use accurate reference energies. The work presented here, implemented as ASR recipes, can be used as supplementary stability information or in cases where information about reactivity with external materials is critical.

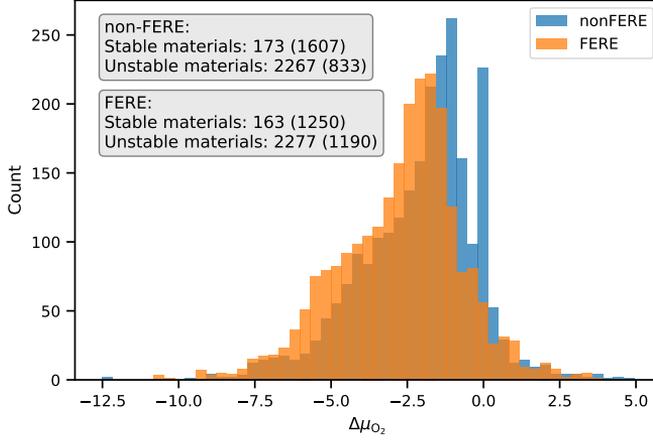


Figure 5.6: Distribution of required chemical potential adjustments for O₂ in order for the materials in C2DB to be stable. The blue distribution shows the uncorrected references while the orange distribution shows the results after applying FERE.

5.3 Effective Masses

A free electron, that is, an electron that is not subject to any forces, has an energy given by

$$E(\mathbf{p}) = \frac{\mathbf{p}^2}{2m_0}, \quad (5.5)$$

where m_0 is the electron mass and \mathbf{p} is the momentum quantum number. In periodic systems the momentum is no longer a good quantum number but following Bloch's theorem one can define a *crystal momentum* \mathbf{k} [14]. The crystal momentum is essentially defined through the eigenvalues of the translation operator

$$\hat{T}(\mathbf{a}) = \exp(i\hat{\mathbf{p}} \cdot \mathbf{a}), \quad (5.6)$$

where \mathbf{a} is the translation vector. If the system is periodic with respect to translations of the form $\mathbf{R}_n = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, then we can simultaneously diagonalize \hat{T} and the Hamiltonian. These simultaneous eigenstates are typically written as $\psi_{\mathbf{k}n}$ where \mathbf{k} is the aforementioned crystal momentum which defines the eigenvalue of the translation operator through

$$\hat{T}(\mathbf{R}_n)\psi_{\mathbf{k}n} = \exp(i\mathbf{k} \cdot \mathbf{R}_n)\psi_{\mathbf{k}n}. \quad (5.7)$$

The energy, $\langle \psi_{\mathbf{k}n} | H | \psi_{\mathbf{k}n} \rangle = E_{\mathbf{k}n}$, can be thought of as a function of crystal momentum and this is what defines the *bands* or *bandstructure* in periodic systems:

$$E_n(\mathbf{k}) = \langle \psi_{\mathbf{k}n} | H | \psi_{\mathbf{k}n} \rangle. \quad (5.8)$$

There is a superficial similarity between $E(\mathbf{p})$ and $E(\mathbf{k})$ but in general the bandstructure for a periodic system is a very complicated function of \mathbf{k} whereas the free electron "bandstructure" is very simple. However, for semiconductors at low temperatures or low doping there is a useful analogy to draw. When the system is gapped and the temperature is zero, the lowest conduction band is completely unfilled. However, if we introduce a finite but small temperature or a little bit of electron doping, the lowest conduction band will become partially filled. Since we are assuming that the temperature or doping is small only the bandstructure around the conduction band minimum will be probed because other regions are energetically prohibited. This suggests that low doping physics can be effectively described not by the full bandstructure but just by the bandstructures behaviour to low orders in \mathbf{k} around the conduction band minimum. Thus we seek a Taylor expansion (assuming for simplicity that the band minimum is at $\mathbf{k} = 0$)

$$E(\mathbf{k}) \approx E_0 + \mathbf{d} \cdot \mathbf{k} + \frac{1}{2} \mathbf{k} \cdot \mathbf{A} \cdot \mathbf{k}. \quad (5.9)$$

By assumption $\mathbf{d} = 0$, since we consider an expansion around the conduction band minimum. E_0 is typically also irrelevant. We are left with the Hessian of the bandstructure \mathbf{A} as the lowest order relevant term of the bandstructure:

$$E(\mathbf{k}) \approx \frac{1}{2} \mathbf{k} \cdot \mathbf{A} \cdot \mathbf{k}. \quad (5.10)$$

If we write this equation in terms of the eigenvalues, λ_i , and eigenvectors, \mathbf{w}_i , of the Hessian we get

$$\mathbf{k} = u_1 \mathbf{w}_1 + u_2 \mathbf{w}_2 + u_3 \mathbf{w}_3, \quad (5.11)$$

$$E(\mathbf{k}) \approx \frac{1}{2} (\lambda_1 u_1^2 \mathbf{w}_1^2 + \lambda_2 u_2^2 \mathbf{w}_2^2 + \lambda_3 u_3^2 \mathbf{w}_3^2). \quad (5.12)$$

The bandstructure now looks very similar to the free electron bandstructure and this analogy is completed by defining the *effective masses*, m_i^* :

$$m_i^* = \frac{1}{\lambda_i}. \quad (5.13)$$

If we absorb the expansion coefficients, u_i , into \mathbf{w}_i the bandstructure is then takes the form

$$E(\mathbf{k}) \approx \frac{\mathbf{w}_1^2}{2m_1^*} + \frac{\mathbf{w}_2^2}{2m_2^*} + \frac{\mathbf{w}_3^2}{2m_3^*}. \quad (5.14)$$

The discussion above was done for 3D materials but specialization to lower dimensionalities is straightforward.

The basic algorithm used to calculate effective masses in C2DB and improved upon in this work is to perform a second order fit to the DFT bandstructure and from this extract the effective masses from the Hessian as described above. In this definition we attempt to find the Hessian exactly at the band extrema, as opposed to finding a

mean curvature over a larger range (e.g. corresponding to room temperature). There are other possible definitions of the effective mass such as using $\mathbf{k} \cdot \mathbf{p}$ -theory [88] but the approach taken here is simple and robust and can be applied to any material without modification whereas $\mathbf{k} \cdot \mathbf{p}$ -theory relies on the symmetry of the lattice.

Since the effective masses describe low-energy excitations they are useful for a number of things. For example, it can be shown that the carrier mobility in 2D, μ_{2D} depends on the effective mass as [89]

$$\mu_{2D} \propto \frac{1}{(m^*)^2}, \quad (5.15)$$

and, as another example, the Seebeck (thermopower) coefficient, S , depends on effective mass as [89]

$$S \propto m^*. \quad (5.16)$$

5.3.1 Effective Mass Recipe

The effective mass algorithm was originally designed and implemented in connection with the first version of C2DB. In my thesis work I adapted the implementation to ASR and implemented it as the recipe `asr.emasses`, generalized the algorithm to work for materials in any dimension, and modified the algorithm to improve robustness. I have also implemented *parabolicity estimates*, that compare the quadratic fit over a large energy range thus providing a different measure of how parabolic the band is and allowing users of the data to find interesting outliers.

Algorithm The core algorithm is shown here and we go through a detailed explanation of the steps below.

1. Perform a ground state calculation
2. Determine preliminary locations of the VBM and CBM.
3. Non-self-consistently calculate the KS eigenvalues on a denser k -grid around the VBM and CBM.
4. Repeat step 2 based on the new grid and calculate a final k -grid.
5. Use the calculated energies on the final k -grid to fit a 2nd order polynomial. From the fit we extract the Hessian and then the effective masses.
6. Calculate bandstructures along the Hessian eigenvectors.
7. Calculate parabolicity over a range of 25 meV for the parabolicity estimates.

In step 1 we use `asr.gs` to calculate the ground state density. This is used as a starting point for the rest of the algorithm.

To fit the effective masses we need to construct a denser k -grid around the VBM/CBM to get an accurate enough fit. The philosophy employed in `asr.emasses` is that the effective masses should be defined from the best approximation to the mathematical

curvature at VBM/CBM, as mentioned in the introduction to effective masses. This means we need a much denser grid than would be computationally feasible in the ground state calculation. This is why we have steps 2 and 3: we first find the VBM/CBM based on the ground state calculation and then determine the eigenvalues non-self-consistently on a much denser grid that is localized around the VBM/CBM. We refer to this a *grid-refinement* or *k-grid-refinement*.

In step 4 we repeat this procedure, starting from the grid from step 3 rather than the `asr.gs-grid`, to define a final, highest-density *k*-grid. This is then used to find the effective masses. The final region spans approximately 1 meV. Such a small region is chosen to approximate the mathematical curvature very closely without suffering from numerical noise. The double grid-refinement improves effective masses in cases where the band extremum is hard to locate, as in e.g. bands with Rashba splitting.

Step 5 simply performs a least-squares fit to the energies. If we have N *k*-points \mathbf{k}_i the least-squares problem can be written as the matrix equation

$$\mathbf{A}\mathbf{x}=\mathbf{b}, \quad (5.17)$$

$$\mathbf{A} = \begin{pmatrix} 1 & k_{1x} & k_{1x}^2 & k_{1y} & k_{1y}^2 & k_{1z} & k_{1z}^2 & k_{1x}k_{1y} & k_{1x}k_{1z} & k_{1y}k_{1z} \\ 1 & k_{2x} & k_{2x}^2 & k_{2y} & k_{2y}^2 & k_{2z} & k_{2z}^2 & k_{2x}k_{2y} & k_{2x}k_{2z} & k_{2y}k_{2z} \\ \vdots & \vdots \\ 1 & k_{Nx} & k_{Nx}^2 & k_{Ny} & k_{Ny}^2 & k_{Nz} & k_{Nz}^2 & k_{Nx}k_{Ny} & k_{Nx}k_{Nz} & k_{Ny}k_{Nz} \end{pmatrix}, \quad (5.18)$$

$$\mathbf{x} = \begin{pmatrix} E_0 \\ a_x \\ b_x \\ a_y \\ b_y \\ a_z \\ b_z \\ c_{xy} \\ c_{xz} \\ c_{yz} \end{pmatrix}, \quad (5.19)$$

$$\mathbf{b} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix}, \quad (5.20)$$

where k_{jx} is the *x*-component of the *j*th *k*-vector and so on, ϵ_j is the energy at the *j*th *k*-point, and the components of \mathbf{x} define a quadratic fit given by

$$E(\mathbf{k}) = E_0 + a_x k_x + a_y k_y + a_z k_z + b_x k_x^2 + b_y k_y^2 + b_z k_z^2 + c_{xy} k_x k_y + c_{xz} k_x k_z + c_{yz} k_y k_z. \quad (5.21)$$

Eq. (5.17) will be overdetermined and there will be no exact solution (unless the band is exactly quadratic). It can only be solved in a least-squares manner which is easily

done using standard libraries. With these definitions the Hessian, \mathbf{A} is easily seen to be

$$\mathbf{A} = \begin{pmatrix} 2b_x & c_{xy} & c_{xz} \\ c_{xy} & 2b_y & c_{yz} \\ c_{xz} & c_{yx} & 2b_z \end{pmatrix}, \quad (5.22)$$

and from this the effective masses are extracted. Fits performed with this method are almost always correct, i.e. the residual is small, however if the energies are degenerate, which happens when the band is very flat, the 1 meV fitting range can be too small to perform a numerically stable fit. The result in those cases can be a curvature with the wrong sign, e.g. implying that the valence band maximum is really a minimum or saddle point. Another case where the algorithm can be faulty is when the band has a Rashba spin-orbit splitting that is in a critical region. The critical region is when the Rashba splitting is not very small (so as to be unnoticeable) but not very large either. In those cases the energies in the fitting range effectively have curvature with the wrong sign which produces incorrect fits. We do not have quantitative estimates for the size of this region. Faulty cases are the minority, as we will see below.

In step 6 GPAW is used to calculate bandstructures along the Hessian eigenvectors.

In step 7 we calculate a parabolicity measure. This is defined by first finding a energy scale, E_s , by averaging the energy variation of the band over a 25 meV window, e.g. $E_s = \langle E(k) - E_{\text{vbm}} \rangle_{25 \text{ meV}}$. This defines an energy scale which is independent of the absolute position of the band. Then the mean absolute relative error (MARE) is calculated where the error is relative to E_s , $\text{MARE} = \langle (E(k) - E_{\text{vbm}})/E_s \rangle$.

Effective mass overview An illustration of the calculated quantities are shown in figure 5.7. The plots on the left show the bandstructures (dots) for Rh_2Br_6 and MoS_2 with the effective mass plots (red dashed line). The plots just show a single Hessian-eigenvector direction. MoS_2 is an example of a bandstructure that is very parabolic, with a MARE of 2.13% whereas Rh_2Br_6 showcases an asymmetric, non-parabolic band with corresponding higher MARE. On the right side of figure 5.7 we see the distribution of MAREs for the C2DB. 77.32% of bands have a MARE of less than 10%. A small percentage have extremely large MAREs, i.e. $>1000\%$. These materials most likely have very flat or asymmetric bands or are in one of the faulty classes discussed above in step 5.

Outliers To get a sense for the types of errors that can occur we will investigate a couple of outliers that have a very high MARE, i.e. higher than 100%.

The first case is BaI_2 shown in figure 5.8. The conduction band in the first direction has a very flat bottom. The resulting fit has the wrong curvature which results in a high MARE in excess of 500% (the plotting code also breaks down).

The second case is Cl_2Ti_2 which is shown in figure 5.9. The band has an initially quite flat region, as evidenced by the small energy differences seen in the plots, which then flattens even further away from the fitting region. The result is that, since the

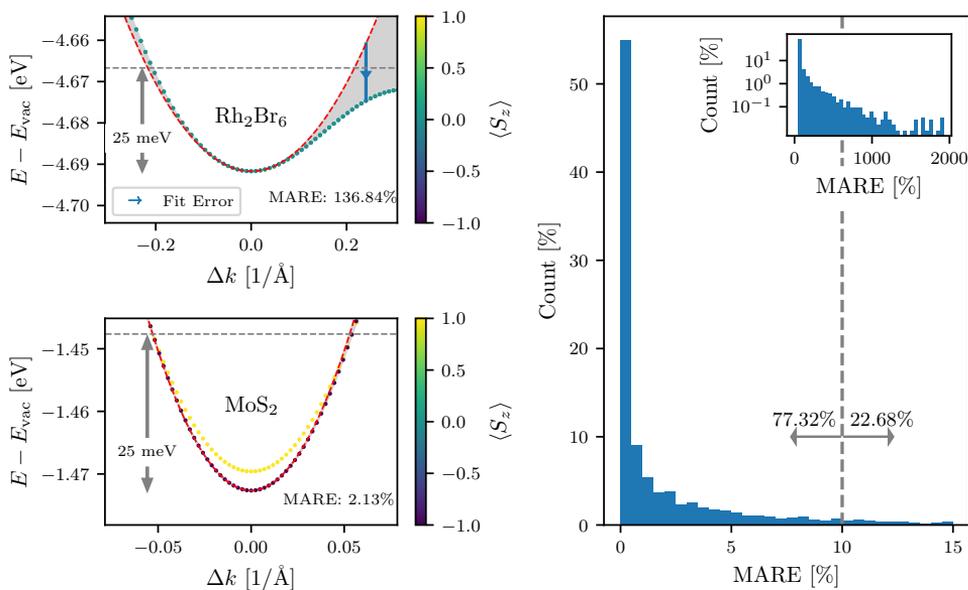


Figure 5.7: Illustration of effective mass fits and parabolicity error measures, and the distribution of these errors. Source: paper III.

initial region was flat, the 25 meV range will correspond to a large k -range over which it might be unreasonable to expect that the fit is correct.

5.4 Other Contributions to ASR

In this thesis I have also implemented other recipes. The $G_0W_0@empirical-Z$ method, described in chapter 4, is now a standard part of the `asr.gw` recipe.

Recipes were also developed to construct bilayers by stacking a monolayer, and optimize the resulting bilayer structures. Several utility recipes for constructing a bilayer database were also developed. This work is further described in chapter 6.

5.4.1 Overview of Papers IV, III, and I

In this section we will briefly go through the results of the papers that are related to the work described in this chapter.

In paper IV the implementation and design of ASR is described. It contains a detailed description of the architecture, exact definitions of recipes and other terminology, technicalities of the caching and dependency-management system, information about the various interfaces to ASR (CLI, python scripting, web browser), and a use case example, namely the updated version of C2DB. Most of the central points of this paper were already touched upon in the introduction to this chapter, so we will not repeat them here.



Figure 5.8: Effective mass webpanel for BaI_2 . Adapted from [66].

In paper III we describe the updates that have been made to C2DB since its inception. This includes data on more materials, improvements to various property calculations, new properties, as well as an outlook describing projects that are currently in development that use C2DB data. There is too much to go through all of it here, and in any case it can be found in the paper, so we will restrict ourselves to the work done in this thesis. The effective mass recipe described here and the G_0W_0 empirical Z method described in chapter 4 is part of the improved properties described in III. In chapter 6 we describe a bilayer screening project undertaken in this thesis. This consisted of forming bilayers by stacking monolayers from the C2DB and calculating various properties. One useful property that we can extract is the exfoliation energy which is slated to be one of the new properties in C2DB. The enhanced stability analysis is not part of the paper as it was deemed to have lower priority than other projects and so computational time and manhours were not dedicated to calculating e.g. oxygen reactivity for all the materials in the C2DB. As part of the outlook is the aforementioned bilayer screening project which we will return to in chapter 6.

In paper I we undertook a statistical study of the already calculated properties in the C2DB, searching for materials with interesting and potentially useful anisotropies. Anisotropic properties are interesting because they can act as a kind of filter or transformer of incoming signals. One example is polarization filters which block light polarized along a certain axis. Anisotropies can happen for optical response properties, mechanical properties such as strain-response, mobility can be different along different

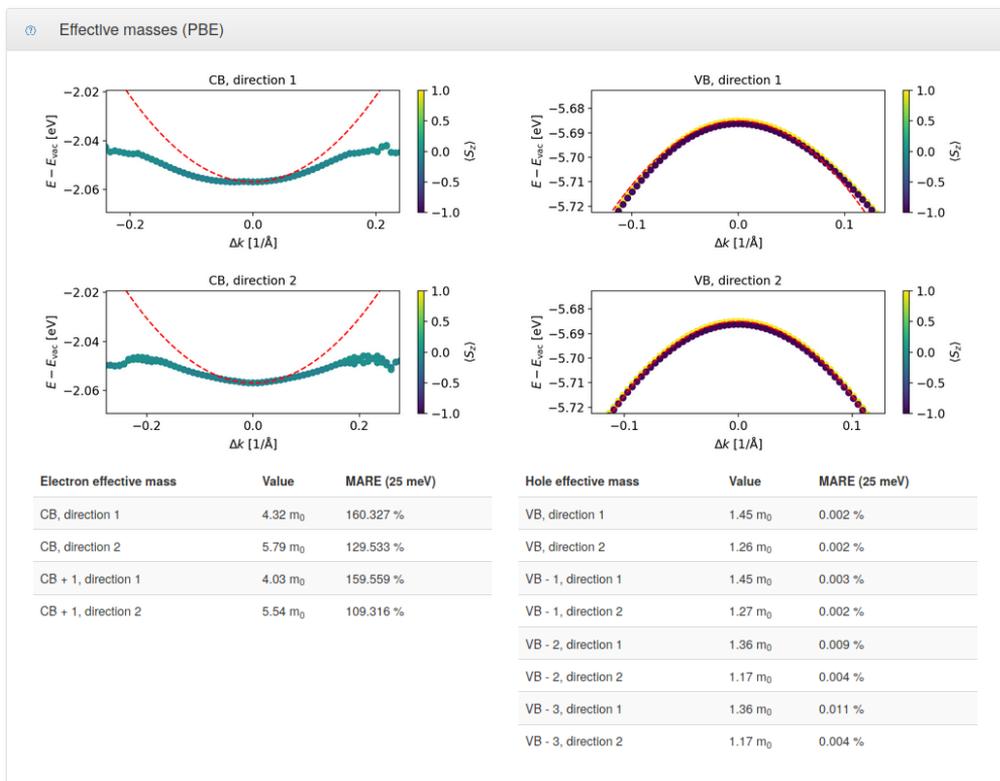


Figure 5.9: Effective mass webpanel for Cl_2Ti_2 . Adapted from [66].

directions, and electromagnetic response can also be anisotropic. Such effects can have various exotic applications, see e.g. [90–92]. In the paper we studied anisotropy measures for 4 different properties: the magnetic easy axis, elastic response, effective masses, and polarizability.

To illustrate our results, we describe here the magnetic and effective mass anisotropies. The magnetic easy axis is defined as the direction along which it takes the least amount of energy to magnetize a material. For 2D materials the Mermin-Wagner theorem [93] prohibits spontaneous symmetry breaking, so some kind of asymmetry is required for a 2D material to be magnetic. We considered the magnetic anisotropy energy defined as

$$\Delta_{xy} = |E(\mathbf{M}|y) - E(\mathbf{M}|x)|, \quad (5.23)$$

i.e. the difference in ground state energies per unit cell when the magnetization is fixed to be along the x vs. y axis. Here the x -axis is along one of the lattice vectors and the y is perpendicular to this. It was found that those magnetic materials with a large magnetic anisotropy energy had a markedly different distribution of point group symmetries. The point groups mmm and $2/m$ comprise together more than 50% of all materials with a magnetic anisotropy energy higher than 0.005 meV / unit cell and more than 80% of materials with a magnetic anisotropy energy higher than 0.7 meV / unit cell. We found one particularly interesting candidate material: V_2I_4 which is predicted to be stable and has a magnetic anisotropy energy of 1.09 meV / unit cell which rivals the highest known anisotropies [94].

For the effective masses we considered the anisotropy measures

$$\delta_{me} = \frac{|m_1^e - m_2^e|}{m_1^e + m_2^e}, \quad (5.24)$$

$$\delta_{mh} = \frac{|m_1^h - m_2^h|}{m_1^h + m_2^h}. \quad (5.25)$$

Here $m_i^{e(h)}$ is the effective electron (hole) mass along the i th direction at the conduction (valence) band extremum.

To remove as many numerical outliers as possible from the analysis we removed materials that had an effective masses greater than $m_i^{e/h} > 20m_e$ as well as materials with a high ratio $m_i^{e/h}/m_j^{e/h}$ (for any values of i, j). At the time of writing, the C2DB contained 574 materials that were both dynamically and thermodynamically stable and that were gapped. 106 were removed according to the above criteria. Quite a large number of materials have an anisotropy factor, $\delta_{me/mh}$, of greater than 0.7: we found 101 monolayers. These are reported in the paper in figure 6. We also found 3 materials that were previously experimentally fabricated (TiS₃, SnSe, GaTe, see [66] for references) and 1 material (AuSe) that was evaluated to be easily exfoliable in [7].

6 High-throughput Screening of Bilayer Materials

6.1 Introduction and Motivation

Two-dimensional materials are interesting from the point of view of both fundamental science and technological applications. A two-dimensional material has first and foremost significantly reduced electrostatic screening compared to three-dimensional materials which combined with the increased electron confinement leads to stronger quantum effects. Furthermore, two-dimensional materials are basically "all surface" which means the properties of the quantum states in a two-dimensional material are much more tunable by external gating, are more susceptible to chemical functionalization, and can interact more effectively with adjacent materials. This opens the way for many potential technological uses that are not available in three-dimensional materials.

With two-dimensional materials as a starting point it is possible to construct layered materials by stacking two or more monolayer materials. By combining different types of monolayers, exotic hybridisation effects can occur but even layered materials formed from the same type of monolayer can show exotic behaviour. For example, bilayer graphene at a special twist angle becomes superconducting [95]. It can also happen that the monolayer has one magnetic state but the bilayer has another. For example, CrI_3 is ferromagnetic in the monolayer [96] but is anti-ferromagnetic or ferromagnetic in the bilayer depending on how the two monolayers are stacked [97]. Other examples include WTe_2 which has a non-zero Berry curvature dipole in odd layers that can be switched by changing the stacking configuration [98]; hexagonal Boron Nitride which has a reversible out-of-plane electric dipole moment in the bilayer [99]; and famously MoS_2 which has a direct gap in the monolayer but indirect in multilayer structures [100].

It is then clear that the interlayer hybridisation that occurs in multilayer structures can give rise to very interesting physics both from a fundamental and an applied point of view. Given the combinatorially large space of possible multilayer structures there is plenty to investigate.

In this thesis we look at homo-bilayers, i.e. bilayers formed from stacking two copies of the same monolayer. There is still a large amount of possible stackings because the two layers may be translated and rotated with respect to each other. For most rotations the unit cell of the bilayer will not be the same as the monolayer and one gets so-called Moiré effects. Here we consider the subset of rotations which ensure that the bilayer has the same unit cell as the monolayer. This reduces the space of possible rotations from an infinite number to a finite number and by avoiding Moiré effects which necessitate large unit cells we can deal with more materials. We start the chapter by describing the algorithm that generates the bilayers. Then we detail the workflow that forms the basis for a highthroughput study of bilayer materials based on

C2DB monolayers [6]. Finally, we analyse the data looking for interesting candidate materials and propose further work that can be done in this field of research.

The manuscript for this project is currently in preparation so there is no associated paper in the "Papers" section.

6.2 Generating the Bilayers

To generate the bilayers we start from a given monolayer. The question is how to find relevant *stackings* where we define a *stacking* as a combination of two identical layers with some relative rotation and translation.

The first step of generating the stackings is to find all the space group operations that preserve the unit cell of the monolayer. As mentioned above, this is to avoid overly large unit cells and to maintain a reasonable scope for the project. Each operation is applied to the monolayer and duplicate structures are filtered out: some operations will result in identical atomic positions. This gives us a set of transformed monolayers with the same unit cell.

Next, for each monolayer in this set we combine it with the untransformed seed monolayer. The layers are placed with some arbitrary initial interlayer distance; we use a default of 5 Å but this will be optimized later during the workflow so it is essentially irrelevant. One layer is then given an additional in-plane translation according to the two following schemes:

1. Position matching
2. Cell-specific translations

For position matching we take each pair formed from atoms in the bottom layer and atoms in top layer and form the set of translations, \mathcal{T}

$$\mathcal{T} = \{\mathbf{p}_i - \mathbf{p}_k | i \in \text{bottom layer}, k \in \text{top layer}\}. \quad (6.1)$$

Here $\mathbf{p}_{i/k}$ is to be understood as the *in-plane* position. If the monolayers are oriented such that they span the $x - y$ plane then e.g. \mathbf{p}_i will just be the x, y components of the i th atom in the bottom layer.

Applying the translation $\mathbf{t}_{ik} = \mathbf{p}_i - \mathbf{p}_k$ to the top layer will ensure that atom i in the bottom layer and atom k in the top layer have the same x, y coordinates. Thus, we consider all the translations that place one atom in the top layer on top of an atom in the bottom layer.

For monolayers with only one atom in the unit cell (or all atoms at same x, y coordinate) this will only give a simple AA stacking. Additionally, for some materials this will not produce experimentally known stackings such as the Bernal stacking in bilayer graphene [101]. To alleviate this we add cell-specific translations. For hexagonal cells we perform a translation that is equivalent to the Bernal stackings in graphene: we translate $2/3$ along one lattice vector and $1/3$ along the other. For other cell types we translate the top layer by half of each lattice vector and by half of one lattice vector plus half of the other. A few examples of these stackings are shown in figure 6.1. Fig.

6.1a shows a honeycomb lattice (i.e. an example of a hexagonal lattice) and the rest show rectangular lattices.

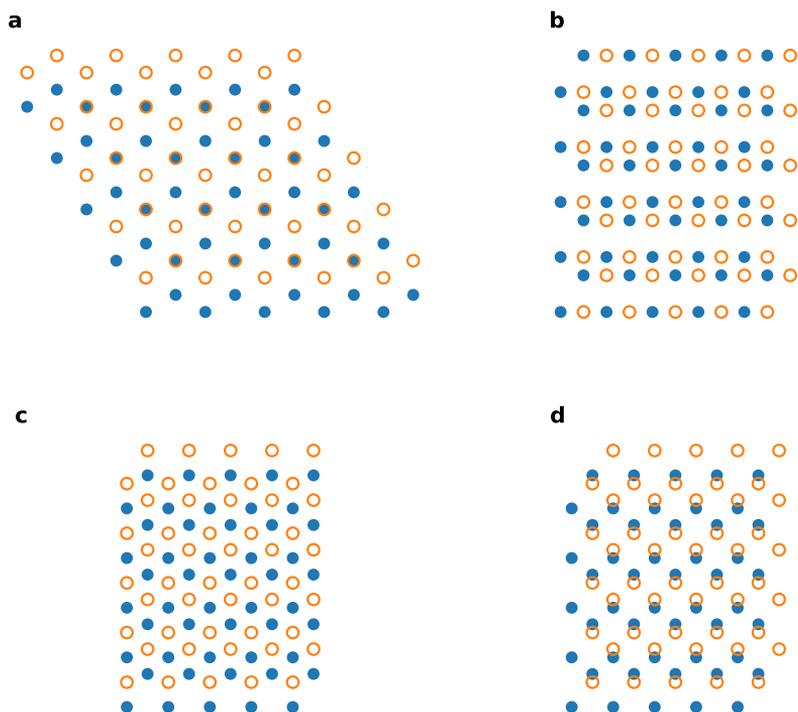


Figure 6.1: Examples of the cell specific stackings. The bottom layer is shown as filled blue circles and the top layer is shown as orange rings. a) shows a honeycomb lattice while the rest show a rectangular lattice. These examples do not correspond to any real material but have just been constructed for illustrative purposes.

This heuristic method generates a set of candidate stackings for the bilayers. The next step is to optimize the structures, i.e. find the minimum energy geometry for the bilayer.

6.3 Bilayer Structure Optimization

Once the bilayers are generated it is necessary to find the minimum energy structure. In principle this could be done via a standard structure relaxation algorithm, e.g. one of the algorithms implemented in ASE [64]. However, since the main layer-layer bonding mechanism is van der Waals bonding it is possible to devise a simpler scheme. Instead of fully relaxing the structure, it may be sufficient to only vary the interlayer distance while keeping in-plane positions fixed to find the optimal structure. If non-van der Waals effects are small enough, this will not affect the positions of the atoms within the layers and this procedure will be sufficient and faster than a full relaxation.

To capture the van der Waals physics it is not sufficient to use semi-local functionals as they are unable to capture non-local effects. It would be possible to use e.g. the van der Waals functional of [30] but since this has a relatively high computational complexity we decided to use, instead, the DFT-D3 correction [102]. This is a van der Waals correction to the energies (and forces) of a DFT calculation based on atomic positions, i.e. it consists of a correction, E_{disp} ,

$$E_{\text{disp}} = \sum_{a,b} \sum_{n=6,8} s_n \frac{C_n^{ab}}{r_{ab}^n} f_{d,n}(r_{ab}), \quad (6.2)$$

where a, b range over the atoms in the system, r_{ab} is the distance between atoms a and b , C_n^{ab} are *ab initio* parameters, and $f_{d,n}$ is an empirical damping function (hence the subscript d). Being based only on atomic positions this method is fast but it still achieves an accuracy of within 10% of coupled cluster methods [102]. In the bilayer structure optimization we use it with the PBE functional via the DFT-D3 interface in ASE.

For future reference we denote the interlayer distance optimization with PBE-D3 functional as the *standard relaxation* (standard within the context of this project).

After we have calculated the optimal bilayer structure we can find the *binding energy* of the bilayer, E_b . This is defined as

$$E_b = 2E_{\text{ML}} - E_{\text{BL}}. \quad (6.3)$$

Here E_{ML} is the energy of the monolayer and E_{BL} is the bilayer energy. It is of course important that both are calculated with the D3 correction and the same underlying functional. With this definition binding energies are positive and a higher binding energy corresponds to a more strongly bound bilayer.

6.4 Stacking and Relaxation Verification

Since we only change the interlayer separation the translations and rotations we started with will remain fixed throughout the optimization. This makes it essential that the rotations and translations are chosen in such a way that the most stable or relevant stackings are generated. To verify our structure generation and optimization we first consider the effects of performing a full relaxation and comparing it with a standard relaxation. The comparison is done by calculating the Root-Mean-Square-Distance (RMSD) measure implemented in pymatgen [103]. We will not go over the RMSD algorithm in detail here as it is defined in their documentation but it involves attempting to match all possible lattices and all possible positions of one structure with the other and calculating the RMSD between the atoms. In ASR (see chapter 5), two structures are considered equal if the RMSD is less than 0.3 Å and we adopt the same criterion here. Figure 6.2 shows the RMSD vs. binding energy of various randomly selected bilayers on the left side. 87% of bilayers are essentially unaffected by the full relaxation and have a RMSD of less than 0.3 Å. Some materials have higher RMSD and are indicated in red, green, orange and purple. Half of the outliers (RMSD > 0.3 Å), with the exception of Sb_2 and $\text{Cl}_2\text{Cu}_2\text{Se}_4$, have that all bilayers for a given monolayer either fall above

or below the 0.3 Å criterion. There also seems to be no tendency that a large RMSD correlates with a large binding energy. Outliers can be caused by two things: either the monolayer structure changes significantly when relaxed with PBE+D3, or there are interlayer effects not accounted for by the simple interlayer-distance-only relaxation. We observe that the bilayer RMSD is pseudo-lower-bounded by the monolayer RMSD; it cannot be a strict bound because interlayer effects, although weak, can in principle change things. Therefore, we may tentatively conclude that since materials with significant change in the monolayer will correspond to a high minimum bilayer RMSD. That type of outlier can then be explained by the monolayer relaxing significantly. The other class, where the minimum bilayer RMSD is very low but the maximum bilayer RMSD is high, e.g. Sb_2 , probably have significant interlayer effects not accounted for by the interlayer-distance-only relaxation, at least in some stackings. On the right of figure 6.2 the RMSDs for the monolayers after PBE+D3 relaxation vs. the exfoliation energy is shown. Note that monolayer $\text{Au}_2\text{Br}_2\text{Te}_4$ experiences a large change after PBE+D3 relaxation and the bilayers have a correspondingly high RMSD. The same goes for FeBr_2 but interestingly the monolayer RMSD for $\text{Ag}_2\text{Br}_2\text{S}_4$ is larger than for FeBr_2 whereas all $\text{Ag}_2\text{Br}_2\text{S}_4$ bilayers have a low RMSD. Thus we see that the ordering of the monolayers is not necessarily the same as the bilayers and complicated things can happen in general.

In summary, there will be materials where the monolayer structure changes significantly when van der Waals effects are taken into account, and there will also be materials where the optimal structure is not found when doing interlayer-distance-only relaxation. However, both of these classes are in the minority.

We have also compared calculated exfoliation energies for a select set of materials with those reported in [7]. The results are shown in table 6.1. The last two columns report the results from [7] while the "PBE+D3" are the results from this project. We see that [7] gets a spread of the order of 20-30% in the calculated exfoliation energy of a material between different functionals. Our results match reasonably well in most cases.

These results show the extent to which the relaxations can be relied on. For most cases the structures are well described. Some cases, however, are not in the minimum energy structure, likely because the source monolayer was not adequately described with just PBE. As always in high-throughput studies, one must choose a reasonable trade-off between accuracy and speed so the study is both feasible and useful.

One further verification we have performed is to check whether our stacking procedure reproduces the stackings in known, layered bulk materials. Layered bulk materials are understood as effectively consisting of 2D layers that are bonded by van-der-Waals forces. To identify such structures we use the method of [104], which can identify subcomponents in a unit cell and construct a geometric "2D score" for each component. 3D materials that contain components with a high 2D score are classified as layered. Starting from the lowdim database [104] we identify layered materials and the layers of which they consist.

To begin with we remove those materials where there is a large difference between

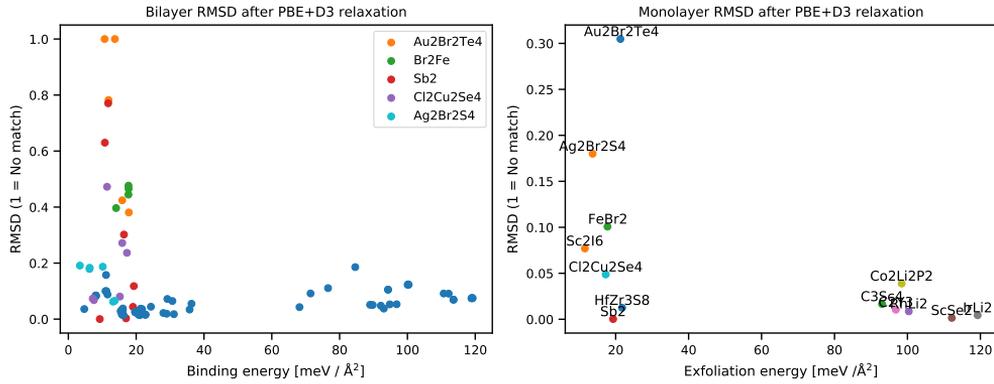


Figure 6.2: (left) RMSD between the interlayer-distance-only relaxed bilayer and the fully relaxed bilayer vs. binding energy. (right) RMSD between PBE-relaxed monolayer and PBE+D3-relaxed monolayer vs. exfoliation energy.

Material	PBE+D3	DF2	rVV10
MoS ₂	28.9	21.6	28.8
MoTe ₂	30.3	25.2	30.4
ZrNBr	18.5	10.5	18.5
C	18.9	20.3	25.5
P	21.9	38.4	30.7
BN	18.9	19.4	24.4
WTe ₂	32.0	24.7	30.0
PbTe	23.2	27.5	33.0

Table 6.1: Exfoliation energies for selected materials calculated with PBE+D3 are compared with the DF2 and rVV10 results from [7]. All numbers are in units of meV/Å².

the bulk monolayer and the freestanding monolayer. This can happen due to relaxation effects and may be due e.g. strong buckling effects. For these cases it is not meaningful to ask whether we find the correct stackings. From the remaining layered materials we extract the bulk stackings, i.e. the possible pairs of layers within the bulk material. In general, a freestanding monolayer cannot be directly compared to its bulk counterpart because of deformations that are likely to occur, such as lattice mismatches due to strains, or space groups that change (which could also be a computational artefact). Therefore, we only compare the *relative* positions of the layers within a given bulk stacking with the *relative* positions of the stackings generated by our algorithm. To do this we determine what point group operation and what translation is necessary to construct the second layer from the first layer. This can be directly compared with the output of our stacking algorithm.

In figure 6.3 we show a scatter plot where each point corresponds to one of our bilayer stackings, and we show the difference between the stacking's binding energy and the binding energy of the stacking that we predict to be most stable, $E_b^{\max} - E_b$, on the vertical axis. On the horizontal axis we show the binding energy, E_b , we predict for that stacking. Green points indicate materials where the bulk stackings are constructed by our stacking algorithm and one of the most stable stackings (that is, materials where $E_b^{\max} - E_b < 2$ meV) is one of the bulk stackings. Red points indicate other cases where none of the most stable stackings are in bulk, which could be because they are predicted to be less stable or because they are not generated by the stacking algorithm. We find that for 77.2% of the monolayers extracted from bulk the stacking algorithm finds all of the stackings for those monolayers.

These results indicate that we are able to both construct relevant stackings and also correctly predict them to be stable in the majority of cases.

6.5 Workflow

With these preliminaries in place we can describe the workflow used to screen for interesting bilayer materials.

An overview is shown in figure 6.4. Initially highly stable monolayers from C2DB are selected. For more on the stability criteria see chapter 5. Suffice to say that there are two stability categories: thermodynamic and dynamic. We select monolayers that have the highest stability rating in both categories. The stability criterion is to avoid as many unrealistic materials as possible. Furthermore, we only select monolayers with less than or equal to 8 atoms per unit cell to keep the computational cost manageable. In the future this could be expanded to include larger systems. After applying these criteria we find 960 monolayers¹. After selecting the monolayers we generate bilayer stackings using the algorithm in sec. 6.2. This produces 8270 bilayers. Once the stackings have been produced we relax the interlayer distance using the method described in sec. 6.3.

We next apply two filters designed to select the most promising bilayer candidates. We

¹in practice it was more, but some systems were very hard to converge for one or more properties so we discarded them in the end.

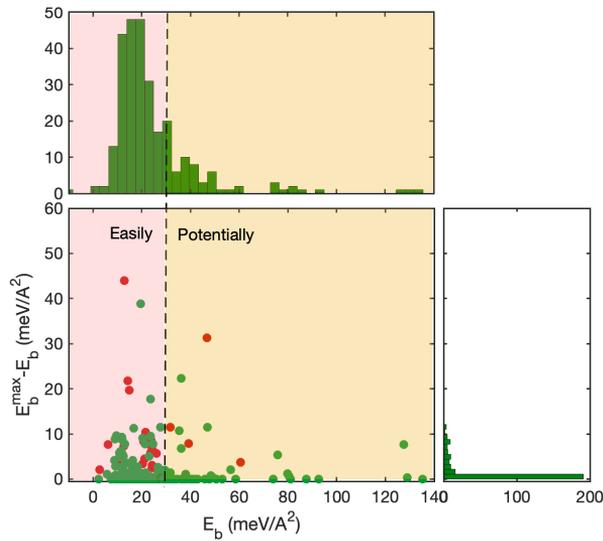


Figure 6.3: Scatter plot of binding energies for materials derived from bulk structures. The green points indicate when the bulk stacking has been correctly identified and the red points indicate when the bulk stacking was not correctly identified. The histograms show the distribution of the green points.

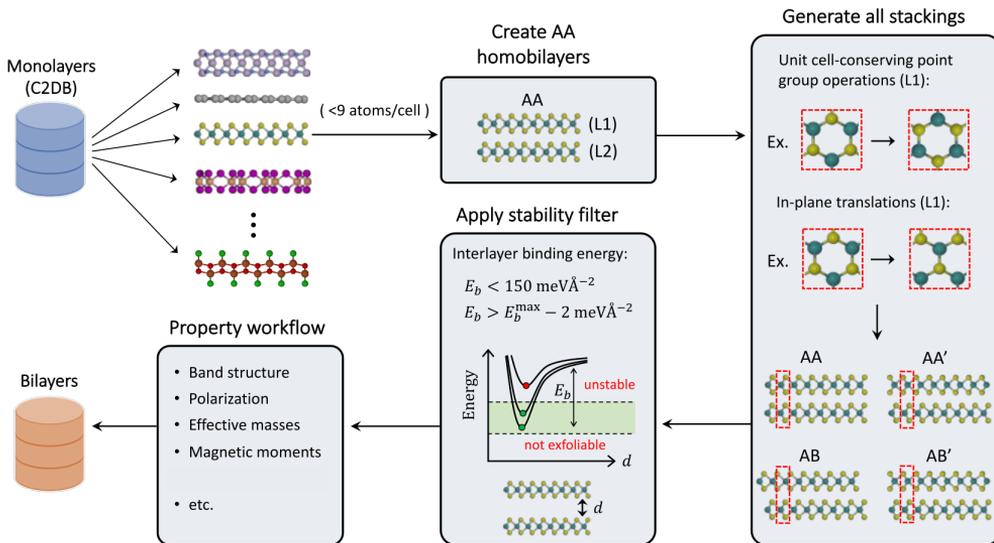


Figure 6.4: An illustration of the bilayer workflow. Highly stable monolayers are extracted from the C2DB. Different unique bilayer structures are then created and the structures are optimized. Various properties are calculated for the most stable bilayer materials.

are mainly interested in experimentally realizable materials so the interlayer binding should not be too large otherwise it is probably difficult to synthesize. To select a cutoff we use the definition of *potentially exfoliable* materials of [7]; if the most stable bilayer has a binding energy larger than $150 \text{ meV}/\text{\AA}^2$ we discard the material. Secondly, for any given monolayer there may be several stackings with a spread in energy. If this spread is large, it may be unlikely that the weakly bound bilayer stackings will be synthesizable. To determine what constitutes a "large" energy difference we refer to MoS_2 where it is known that 3 stackings are significantly more stable than the others [105]. We use our calculated spread of $1.4 \text{ meV}/\text{\AA}^2$ between these three stackings to determine a suitable cutoff: we only include bilayers within $2 \text{ meV}/\text{\AA}^2$ of the most stable stacking. For an illustration of this criterion see figure 6.5. After applying these criteria we are left with 2376 bilayers.

For these bilayers we then run through a property workflow using recipes from ASR (see paper IV). Here we calculate the bandstructure, out-of-plane dipole moment, effective masses, magnetic moments, and a measure of the interlayer magnetic coupling to be detailed later.

The final results are collected into an ASE database.

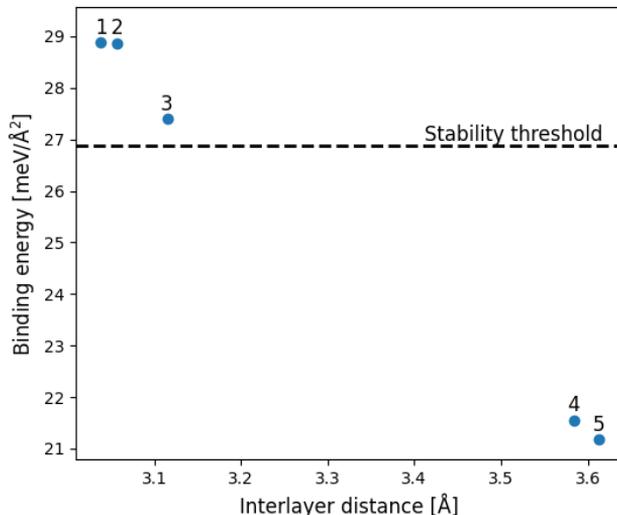


Figure 6.5: A scatterplot of MoS_2 binding energies and interlayer distances. The three most stable stackings are separated from the remaining two by a relatively large gap. The spread in energy of the three most stable stackings is $1.4 \text{ meV}/\text{\AA}^2$, and this leads us to only consider materials within $2 \text{ meV}/\text{\AA}^2$ of the most stable stacking. This is indicated by the dashed line; materials below this line are not included in further calculations.

6.6 Emergent and Switchable Properties

With the results of the workflow in hand we are ready to look for interesting materials and physics. To this end, we frame the discussion by considering two classes of phenomena: *emergent* properties and *switchable* properties.

Emergent properties are here defined to be changes that occur in going from the monolayer to the bilayer. Usually, emergent effects are taken to be, roughly speaking, non-anticipated or non-predictable behaviours in complex systems, i.e. phenomena that involve a complicated interplay of many different parts and forces in such a way as to be very computationally complex and therefore essentially un-derivable from the microscopic laws. When stacking two 2D materials the very least we can expect is some hybridisation between the states in each layer; every band will be doubled and there will be some splitting. We will take the liberty of calling everything beyond that *emergent*. For example, if the band gap character changes from indirect to direct or the material becomes magnetic in the bilayer, we call that an emergent effect.

By switchable properties we mean quantities that can be changed by considering different stackings of the same bilayer. For example, if the band gap character can be changed when going from an AA stacking to an AB stacking, we say that the band gap character is *switchable*. As defined here, this includes changes across all pairs of stackings of a given monolayer. However, it may not actually be experimentally feasible to "switch" between to arbitrary stackings, e.g. if the "switch" would involve a rotation or an inversion of one plane. Rotations and inversions are probably not possible to do on any single device and we may wish to exclude such pairs from the analysis. The actual mechanism for switching will probably be an external field coupling to a dipole moment in the bilayer as in the experiments on hBN in ref. [99], so it will be interesting to consider only those pairs of stackings that have a "large" difference in dipole moment which would suggest they could be switched via an external field.

Studying emergent properties can tell us what kind of changes are possible in going from monolayer to bilayer. It may suggest bilayers for synthesis, if for example a monolayer is easily synthesizable but does not have good optical properties because of an indirect bandgap while a bilayer with a direct gap exists.

Switchable properties offer opportunities for such things as topological memory devices, controllable magnetism, and switchable transport properties, and such bilayers may become a part of quantum materials devices in the future.

6.6.1 Bandstructure Effects

To start with, we consider emergent bandstructure changes by comparing each bilayer to the corresponding monolayer and classifying the change between monolayer and bilayer. We ensure that negligible changes do not affect the statistics by applying three cutoffs. 1) If the gap is smaller than 30 meV, then the material is classified as metallic (or gapless). 2) If the minimum gap is within 30 meV of the direct gap, then the material is classified as having a direct gap. 3) If the minimum gap is not within 30 meV of the direct gap but the valence band maximum is closer than 0.05 \AA^{-1} of the conduction band minimum then the material is also classified as having a direct

gap. With these criteria we avoid spuriously classifying metals as having a gap and we classify materials that "almost" have a direct gap as having a direct gap.

With these cutoffs the distribution of changes is shown in the left side of figure 6.6. Most bilayers are similar to their parent monolayer but some show a change. The most common changes are when the bilayer changes from/to a direct gap to/from an indirect gap and when a gapped material becomes metallic. Since we expect the gap to narrow due to interlayer hybridisation, as mentioned above, the latter category can perhaps be explained by this effect.

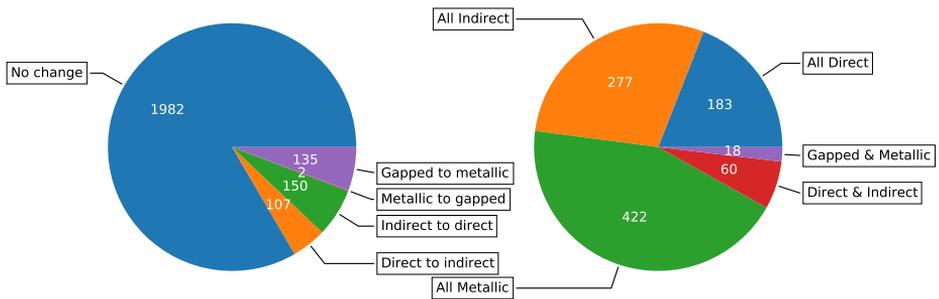


Figure 6.6: (left) The distribution of gap changes for each bilayer. (right) The distribution of gap variation for each monolayer.

We then consider the amount variation that exists between different stackings for a given monolayer. That is, for a given monolayer how much and in what way will the bandstructure vary? This is shown in the right side of figure 6.6. For example, 277 monolayers possess an indirect gap for all possible stackings. Since most bilayer do not change with respect to the monolayer, we have that for most monolayers there is no variation between bilayers. However, for 8.125% of monolayers multiple types of gaps can be found.

Materials with a direct gap are more useful for optical applications since photons with an energy corresponding to typical band gaps essentially have zero momentum. Therefore it is interesting to look for candidates from the class of monolayers that change from indirect to direct. To find the most promising candidates we look for monolayers with an indirect gap and with a energy above the convex hull of less than 50 meV (in chapter 5 we saw that the usual criterion for "high" stability was a hull energy less than 200 meV) where *all* stackings have a direct gap. It may not always be possible to control which stacking order is achieved in an experiment so these criteria should let us select materials that are highly stable and where it is feasible to construct a directly gapped bilayer. We also apply the cutoffs for the bandgap and the VBM/CBM locations described above. This leaves us with 24 monolayers. The full list is shown in the appendix. We highlight here the materials with the clearest indirect to direct

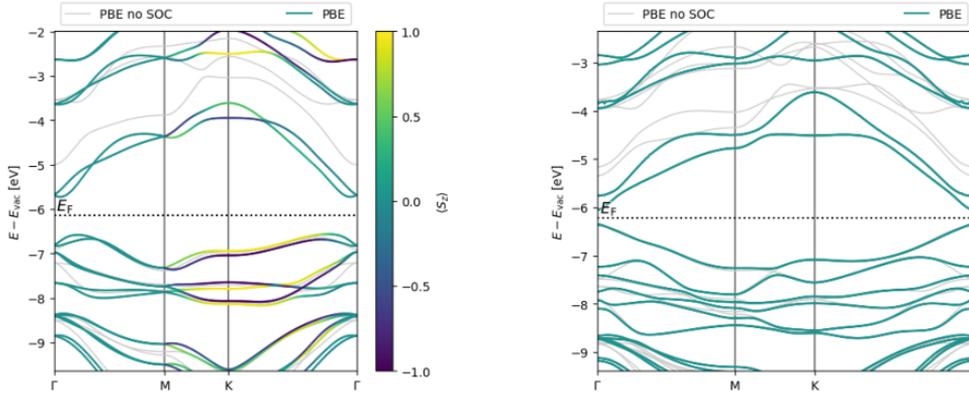


Figure 6.7: The left side shows the monolayer bandstructure for BiClSe and the right side shows the bilayer bandstructure for the most stable stacking of BiClSe. The monolayer has an indirect gap whereas the bilayer has a direct gap.

transition: PbI_2 , BiBrSe , BiClSe , BiISe , BrSbTe , ClSbSe , and ClSbTe . Besides PbI_2 these are all Janus structures. In figure 6.7 we show the bandstructures of BiClSe and its most stable stacking which is representative of the Janus structures, and in figure 6.8 we show the bandstructures of PbI_2 . Materials such as these would be interesting to synthesize both to verify our calculations and to open the door to new direct gap materials.

For the purpose of identifying interesting switchable materials we look for monolayers that have pairs of stacking orders that are related by a translation since pure translations probably have the lowest energy barrier. In ref. [99] they found that bilayer hBN was switchable via external gating and they estimate a dipole moment of 2.25×10^{-12} C/m. Assuming the dipole switching mechanism is related to the interaction of the gates with the dipole moment of the material, this value can provide a guideline for when a dipole moment is large enough to enable switching. We thus apply the pure-translation and a dipole moment change of larger than 10^{-12} C/m as criteria for when two stackings are switchable. We then look for bilayer-pairs that are switchable and where one member of the pair has a direct gap and the other has an indirect gap. Such materials could be interesting in materials where controllable optical response is desired.

Unfortunately, with these criteria we do not find any promising candidates. Even lowering the dipole moment threshold to 10^{-14} C/m does not produce any promising candidates. Perhaps a more targeted search for such materials is necessary; at present the most obvious way forward would be to increase the scope of materials considered.

6.6.2 Dipole Moments

In this section we investigate the presence of dipole moments in bilayer materials. We again split the discussion into emergent dipole moments and switchable dipole

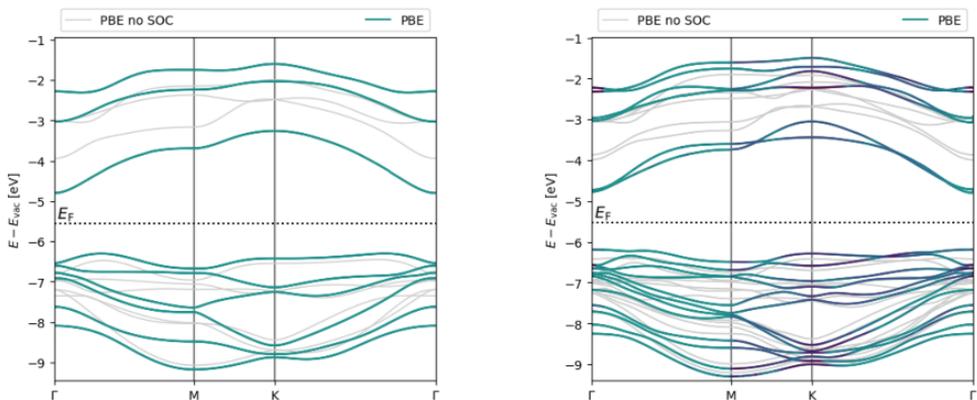


Figure 6.8: The left side shows the monolayer bandstructure of Pbl_2 and the right side shows the bilayer bandstructure of the most stable stacking of Pbl_2 . The monolayer has an indirect gap whereas the bilayer has a direct gap.

moments. Emergent dipole moments are perhaps interesting only from a fundamental point of view but switchable dipole moments can have applications in e.g. memory devices or electro-mechanical actuators, see for example ref. [99]. We only consider out-of-plane dipole moments.

First, we calculate the absolute value of the change in dipole between monolayer and bilayer for each monolayer, Δp_z . We express it in units of $\log_{10}(\text{C/m})$. We also consider the minimum change for each monolayer, $\min|\Delta p_z|$, i.e. we find the bilayer with minimal change in dipole moment. The distribution of these quantities is shown in figure 6.9. 27% of bilayers have an increase in dipole moment greater than 10^{-13} C/m and 12% of monolayers have that all bilayers increase the dipole moment by at least 10^{-13} C/m. Perhaps if an intrinsic dipole moment is required it is worthwhile to consider a stacked structure.

We then, as in sec. 6.6.1, look for bilayer pairs that are related by a rigid translation of one layer.

However, we also need to take special care to include stackings where the structure can be effectively inverted by translating the top layer. That is, by moving the top layer we can end up in a configuration that is equivalent to the initial configuration mirrored in the x-y plane. Such a mirroring produces two structures that are equivalent, and hence only one structure would be included in the workflow. Such structures are relevant for the dipole moments, because mirroring the structure in the x-y plane inverts the out-of-plane dipole moment. Other properties, such as the bandstructure or effective masses are, of course, unchanged.

For the pairs we find in this way we calculate the change in dipole moment, Δp_z , and show the distribution in figure 6.10. We may also consider restricting to bilayer pairs that have the same bandstructure character (direct gap, metallic, etc.) such that trans-

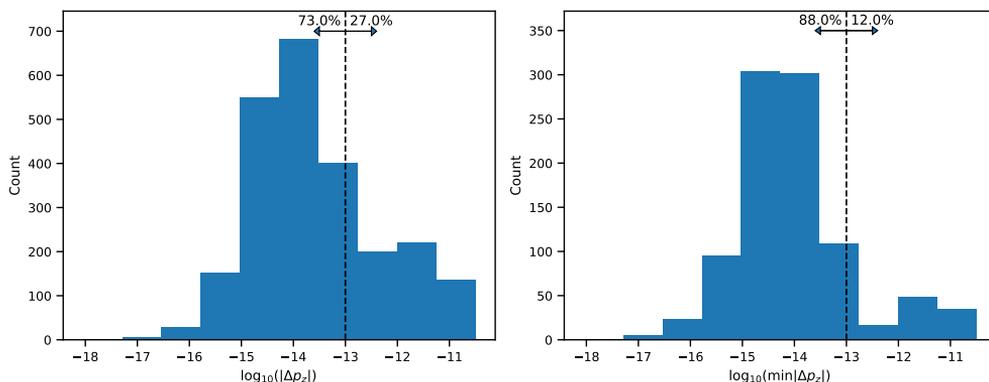


Figure 6.9: Logarithmic changes in dipole moments in going from monolayer to bilayer. Left side shows all changes, the right show shows the minimum change.

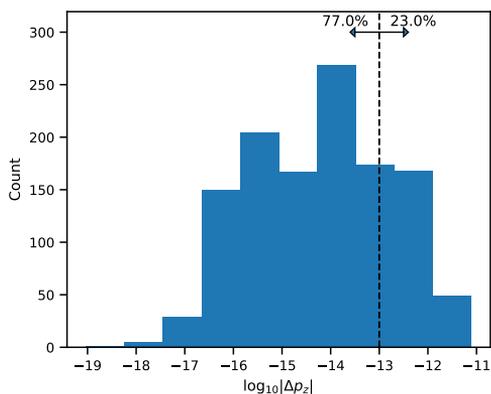


Figure 6.10: Dipole changes across switchable bilayer pairs.

port and optical properties are not changed when switching the dipole moment. However, the distribution is essentially the same (but the absolute counts are of course lower). 23% of switchable bilayer pairs show a dipole change larger than 10^{-13} C/m which seems quite promising for possible applications. If we restrict to those monolayers that are highly stable, i.e. with an energy above the convex hull of less than 50 meV, then we find 125 monolayers, 27 of which are already known in bulk and 19 which have been synthesized in the monolayer which indicates that there is plenty of possibilities left to explore in the lab. If we consider dipole changes larger than 10^{-12} C/m instead we find 28 monolayers, 5 that are known in bulk and 9 which have been synthesized in the monolayer. We list these 28 monolayers in appendix B section B.3.

A few materials that are worth highlighting is hexagonal Boron Nitride which is known to have switchable dipole moment [99], and the well-studied WTe_2 [106, 107] and MoS_2 [100].

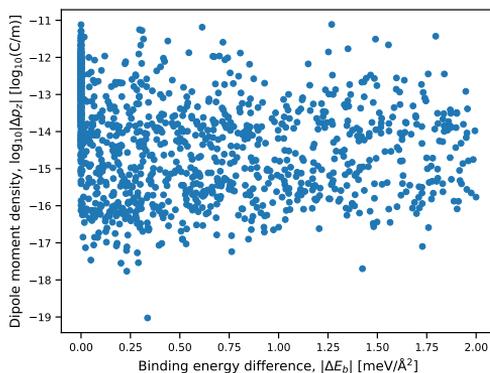


Figure 6.11: Change in dipole density vs. bilayer binding energy between switchable pairs of bilayers.

To get a different view on the same data, we show in figure 6.11 a scatter plot of dipole density changes vs. the different in binding energy between switchable pairs of bilayers. The points at $\Delta E_b = 0$ correspond to bilayer stackings that can effectively be inverted by rigid translation of a layer. The most interesting candidates are those where the dipole change is large but the change in energy, or equivalent binding energy, is small. These are location in the upper left corner of the plot.

6.6.3 Effective Masses

The effective masses are defined as described in chapter 5 but this makes it non-trivial how to compare effective masses for two materials: the masses are defined as essentially the inverse eigenvalues of Hessian at the band extrema which means that it is not guaranteed that the masses of one material correspond to the same *directions* as the masses in another material. To make things simple we compare the lightest masses of one material with the lightest of the other and the heaviest with the heaviest (there are only 2 masses since the materials are 2D). We perform this comparison for the valence bands effective masses, henceforth termed *hole masses*, and for the conduction bands masses, henceforth *electron masses*.

For this analysis we neglect effective masses that are larger than $100 m_e$. Very large masses occur for very flat bands where the effective mass is perhaps not well-defined. We also eliminate cases where the effective mass has the wrong sign; such cases indicate pathological fits and the results have no meaning.

The resulting changes in effective masses when going from the monolayer to the bilayer are shown in figure 6.12. The correlation coefficient between the light mass change and the heavy mass change, C , is also shown. In most cases the masses are essentially unchanged but there is still a significant amount of cases where the change is several tens of m_e . There is little correlation between the light mass change and the heavy mass change but they do have a slight tendency to have the same sign as evidenced by $C \approx 0.23$.

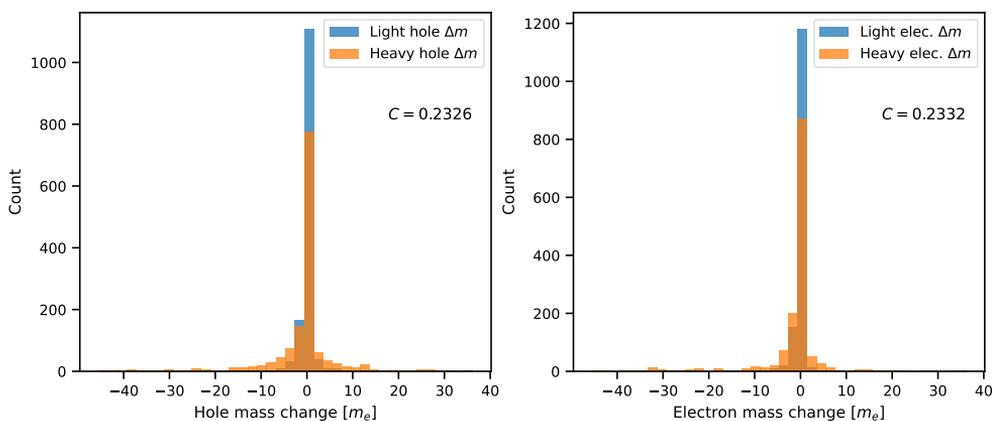


Figure 6.12: Changes in effective masses when going from monolayer to bilayer. The correlation coefficient between the light mass and heavy mass changes, C , is also shown.

These results indicate that it may be possible to both increase and decrease transport coefficients by constructing stacked structures.

Next, we look at switchable bilayer pairs with no restriction on dipole moment change. We just consider the absolute values of the possible changes since there is no intrinsic sense of directionality for two bilayer pairs. The distributions are shown in figure 6.13.

Compared to the monolayer to bilayer changes, the absolute size of the mass changes when switching between bilayers is about an order of magnitude smaller. This corresponds to the intuition that the largest changes in bandstructures will occur between monolayer and bilayer and not between different stacking orders.

If we instead consider the relative changes in effective masses with the additional criterion that the dipole moment must change by more than 10^{-13} C/m we get the distribution shown in figure 6.14. Easily switchable pairs where the mass changes by a factor of 2-3 are clearly possible but they do not seem to be extremely common. Future studies might include a more directed search for high-dipole materials.

Switchable effective masses allow for devices where for example transport can be turned on and off, offering a new type of building block for quantum devices.

Identifying switchable bilayer pairs with a dipole moment change greater than 10^{-13} C/m, having at least one mass that changes by more than 100%, while also having an energy above the convex hull of less than 50 meV, we find only 9 monolayers which are shown in the appendix. Future analyses can try to identify more promising candidates where, for example, all the valence band masses changes by a significant amount or similarly, depending on the application. We find relatively few materials so if such materials turn out to be especially interesting the database could be expanded with the hope of finding more switchable effective masses.

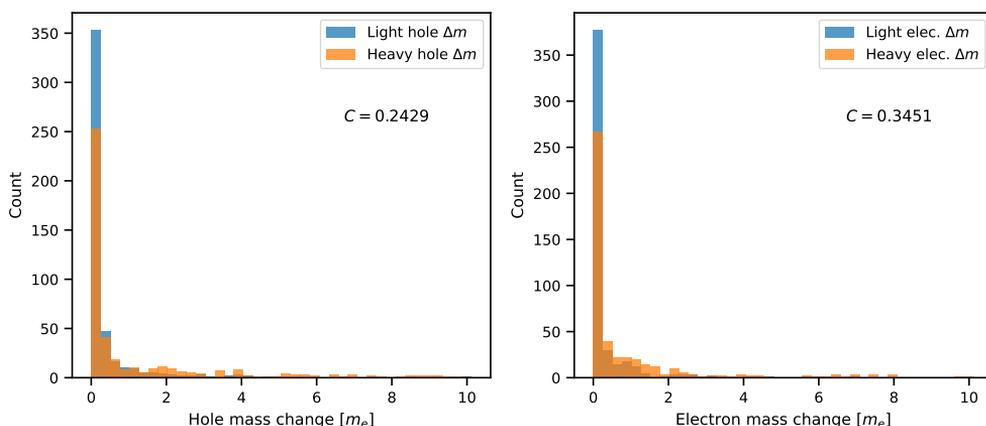


Figure 6.13: Changes in effective masses when going between switchable bilayer pairs. The correlation coefficient between the light mass and heavy mass changes, C , is also shown.

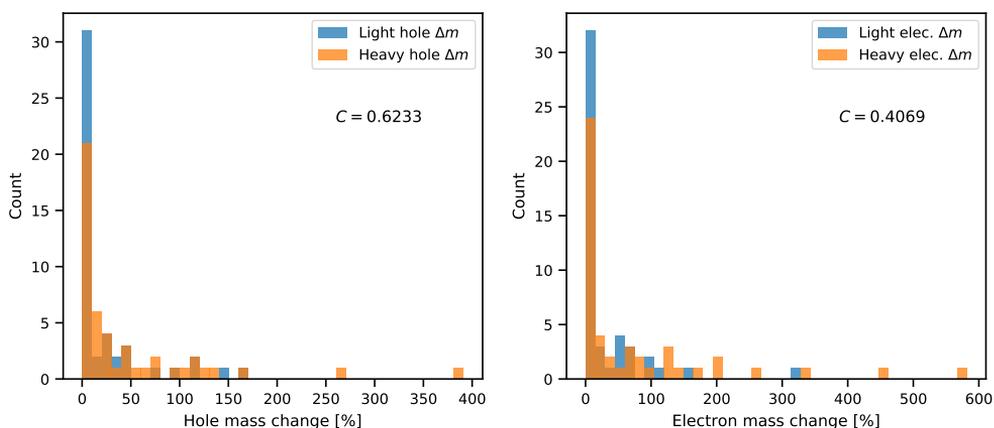


Figure 6.14: Relative changes in effective masses when going between switchable bilayer pairs. The correlation coefficient between the light mass and heavy mass changes, C , is also shown. Only pairs where the out-of-plane dipole moment changes by more than 10^{-13} C/m are considered.

6.6.4 Magnetism

Some materials become magnetically ordered when stacked or the magnetic order can depend on the number of layers and the stacking order. One example is CrI_3 where the bulk structure is ferromagnetic, the few-layer structures display stacking dependent magnetic ordering, and the monolayer is ferromagnetic [96, 97].

In this section we investigate emergent and switchable magnetic effects. We select materials that have either a non-zero total magnetic moment (deemed ferromagnetic)

or a non-zero maximum magnetic moment but a zero total magnetic moment (deemed anti-ferromagnetic). For these materials we calculate the energies of two configurations: 1) interlayer ferromagnetic (ILFM), and 2) interlayer antiferromagnetic (ILAFM). ILFM (ILAFM) is the configuration where the initial magnetic moments are parallel (anti-parallel) in the two layers. For the bottom layer the initial magnetic moments are chosen from the monolayer structure. These calculations are done using PBE+D3+U. We use PBE+D3 to be consistent with the method used to optimize the structure and we need the Hubbard correction to match the reported energetic orderings of the FM and AFM states for various stackings in CrI₃ and a few other materials. We use a Hubbard U of 3 eV for the TM atoms with 3d electrons: Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn. Finally the magnetic order is determined by comparing the energies of the FM and AFM states. We also check that the AFM state has zero total magnetic moment; if not, we say the AFM state does not exist. For the FM state we check that the magnetic moments have not changed direction during the SC cycle; if so, we say that the FM state does not exist. If neither the AFM or the FM state exist the material is classified as non-magnetic (NM). It may be more accurate to say that the magnetic state is not known but in this first pass over the data this analysis should suffice.

We start by looking at changes when going from monolayer to bilayer. These are shown in figure 6.15. Similar to the other properties we have looked at, most materials do not change their magnetic order when stacked. However, a sizable percentage does experience some change.

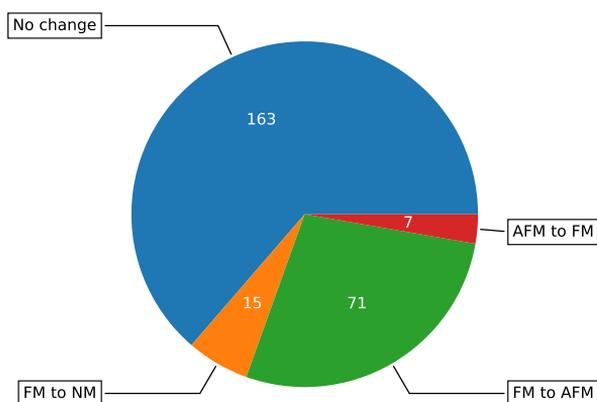


Figure 6.15: Changes in magnetic ordering when going from monolayer to bilayer.

Moving to switchable materials without any restriction on dipole moment change we find 50 monolayers where the magnetic state can be switched between FM and AFM. Of these, 29 have a hull energy of less than 50 meV. If we restrict ourselves to materials where the dipole moment changes by more than 10^{-13} C/m, such that the material might be electrically switchable, we find only 3 monolayers. 2 of these are highly stable and these are listed in the appendix. In this project we have a relatively small sample

of only 101 magnetic monolayers and we are able to find 2-30% of promising materials. In the future a search more directed towards magnetic materials seems promising given these high percentages. However, the energy differences that determine the magnetic state are of the order of meV and there is a risk of under-converged calculations. Therefore any candidates one finds should probably be double-checked if one intends to use them for applications.

6.7 Conclusion

In this project we have developed an algorithm for generating bilayer materials as well as a workflow to perform high throughput screening of bilayer materials. We find several interesting materials with potential applications to optical devices, nonvolatile memory, controllable transport properties, and switchable magnetic devices.

There are many avenues for future studies. Here, we have only included monolayers with 8 or less atoms in the unit cell, and this is the first obvious restriction to lift. The stacking algorithm could also be expanded to include even more realistic stackings. Another option in a similar direction is to change the relaxation algorithm to optimize not only the interlayer distance. This would reduce dependence on the heuristic stacking algorithm but increase computational cost. A further step would then also be needed in the workflow to remove stackings that relaxed into the same local minimum. If specific applications were of interest, more targeted selection or generation of source monolayer as well as generation of potential stackings may be desirable. Finally, the fact that many properties can change when going from one layer to two layers begs the question of what can happen in multilayer structures. One major hurdle that has to be overcome is the combinatorial explosion of the number of possible multilayers that happens if one simply tries all possible multilayer stackings.

7 Conclusions

In chapter 3 we described work on the development of a new type of exchange-correlation functional, called the *Weight Local Density Approximation* (WLDA). The guiding principle was to match both the energy of the HEG, as in LDA, but also the correlation hole. The functionals that we developed are fully non-local and can be thought of as a correction to LDA. They improve over LDA and it was possible to perform as well as PBE on atomic energies, atomization energies, and lattice constants. However, WLDA in the current formulation has a short-coming, namely that the potential is not strictly regular. Luckily, the divergence of the potential happens at low densities, so it is expected that it does not affect the results too much. Nonetheless, future versions of WLDA should be defined so they have a regular potential. It may also be possible to define a "weighted density" functional with a different starting point than LDA.

Chapter 4 dealt with sources of error in G_0W_0 calculations in the context of high-throughput studies. We found several ways to reduce the expected error, including the empirical Z method which only requires the quasiparticle weight, Z , and the self-energy at the Kohn-Sham energy. It can therefore be used as a post-processing step in most G_0W_0 implementations. One can imagine doing similar work on e.g. the BSE method. An ongoing CAMD project is to use machine learning to predict G_0W_0 energies, so, in the future, high-throughput G_0W_0 may be performed very efficiently this way. A crucial step will be to have a way to determine the confidence of the machine learning model, so actual G_0W_0 calculations can be performed where necessary.

The Python package ASR was described in chapter 5. ASR consists of several *recipes* that each perform a well-defined electronic structure task. ASR also has code infrastructure to facilitate recipe versioning, database creation, and data analysis. In this thesis several recipes were contributed and ASR was used to do studies on anisotropy, and bilayer materials, as well as updating the Computational 2D Materials Database. ASR can be extended with many more recipes, and work is ongoing to generalize the interface with the underlying electronic structure code so as to make ASR usable with almost any code the user wished to use. It may also be advisable to extend ASR to better support method development; such tasks are inevitably part of modern materials research. On the other hand, ASR cannot, and should not, do everything.

Finally, in chapter 6 we described a high-throughput screening project which studied homo-bilayers constructed from highly stable monolayers in the C2DB. We found that is sometimes possible to engineer better materials properties by stacking layers, for example, bandstructures and magnetic states may change. We also found a set of candidate *switchable* materials. These are materials where two closely related stackings have different properties, e.g. one has a direct gap and the other has an indirect gap. At present we have restricted the workflow to run on monolayers with at most 8 atoms in the unit cell, and this should be expanded in the future. Additionally, the most interesting candidate materials should be investigated in depth. Finding the can-

didates that are most likely to be synthesizable and collaborating with experimentalists to actually make these materials is another, highly interesting avenue of research.

Bibliography

- [1] James J Fricke. “Jimi Hendrix’ use of distortion to extend the performance vocabulary of the electric guitar”. In: (1998).
- [2] *Los Alamos National Laboratory*. <https://periodic.lanl.gov/32.shtml>. Accessed: 2021-06-28.
- [3] G Ceder and K Persson. *The materials project: A materials genome approach*. 2010.
- [4] Kieron Burke. “Perspective on density functional theory”. In: *The Journal of chemical physics* 136.15 (2012), p. 150901.
- [5] Norbert Schuch and Frank Verstraete. “Computational complexity of interacting electrons and fundamental limitations of density functional theory”. In: *Nature Physics* 5.10 (2009), pp. 732–735.
- [6] Sten Hastrup et al. “The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals”. In: *2D Materials* 5.4 (2018), p. 042002.
- [7] Nicolas Mounet et al. “Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds”. In: *Nature nanotechnology* 13.3 (2018), pp. 246–252.
- [8] Xu Zhang, An Chen, and Zhen Zhou. “High-throughput computational screening of layered and two-dimensional materials”. In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 9.1 (2019), e1385.
- [9] Korina Kuhar et al. “High-throughput computational assessment of previously synthesized semiconductors for photovoltaic and photoelectrochemical devices”. In: *ACS Energy Letters* 3.2 (2018), pp. 436–446.
- [10] Ivano E Castelli et al. “New cubic perovskites for one-and two-photon water splitting using the computational materials repository”. In: *Energy & Environmental Science* 5.10 (2012), pp. 9034–9043.
- [11] Davide Campi et al. “High-throughput searches for novel 2D and 1D materials”. In: *Bulletin of the American Physical Society* (2021).
- [12] Sandra Fox et al. “High-throughput screening: update on practices and success”. In: *Journal of biomolecular screening* 11.7 (2006), pp. 864–869.
- [13] Pierre Hohenberg and Walter Kohn. “Inhomogeneous electron gas”. In: *Physical review* 136.3B (1964), B864.
- [14] Neil W Ashcroft, N David Mermin, et al. *Solid state physics*. Vol. 2005. holt, rinehart and winston, new york London, 1976.
- [15] Jorge Kohanoff. *Electronic structure calculations for solids and molecules: theory and computational methods*. Cambridge University Press, 2006.
- [16] Walter Kohn and Lu Jeu Sham. “Self-consistent equations including exchange and correlation effects”. In: *Physical review* 140.4A (1965), A1133.
- [17] Carsten Rostgaard. *The Projector Augmented-wave Method*. 2009. arXiv: 0910.1921 [cond-mat.mtrl-sci].

- [18] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen. “Real-space grid implementation of the projector augmented wave method”. In: *Phys. Rev. B* 71 (3 Jan. 2005), p. 035109. DOI: 10.1103/PhysRevB.71.035109. URL: <https://link.aps.org/doi/10.1103/PhysRevB.71.035109>.
- [19] J. Enkovaara et al. “Electronic structure calculations with GPAW: A real-space implementation of the projector augmented-wave method”. In: *Journal of Physics Condensed Matter* 22.25 (2010). DOI: 10.1088/0953-8984/22/25/253202.
- [20] John P. Perdew and Karla Schmidt. “Jacob’s ladder of density functional approximations for the exchange-correlation energy”. In: *AIP Conference Proceedings* 577.1 (2001), pp. 1–20. DOI: 10.1063/1.1390175. eprint: <https://aip.scitation.org/doi/pdf/10.1063/1.1390175>. URL: <https://aip.scitation.org/doi/abs/10.1063/1.1390175>.
- [21] John P Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized gradient approximation made simple”. In: *Physical review letters* 77.18 (1996), p. 3865.
- [22] Thomas Olsen and Kristian S Thygesen. “Extending the random-phase approximation for electronic correlation energies: The renormalized adiabatic local density approximation”. In: *Physical Review B* 86.8 (2012), p. 081103.
- [23] Thomas Olsen and Kristian S Thygesen. “Beyond the random phase approximation: Improved description of short-range correlation by a renormalized adiabatic local density approximation”. In: *Physical Review B* 88.11 (2013), p. 115131.
- [24] Manfred Lein, EKV Gross, and John P Perdew. “Electron correlation energies from scaled exchange-correlation kernels: Importance of spatial versus temporal nonlocality”. In: *Physical Review B* 61.20 (2000), p. 13431.
- [25] John P Perdew and Yue Wang. “Accurate and simple analytic representation of the electron-gas correlation energy”. In: *Physical review B* 45.23 (1992), p. 13244.
- [26] MGUJ Petersilka, UJ Gossmann, and EKV Gross. “Excitation energies from time-dependent density-functional theory”. In: *Physical review letters* 76.8 (1996), p. 1212.
- [27] Massimiliano Corradini et al. “Analytical expressions for the local-field factor $G(q)$ and the exchange-correlation kernel $K_{xc}(r)$ of the homogeneous electron gas”. In: *Physical Review B* 57.23 (1998), p. 14569.
- [28] CF Richardson and NW Ashcroft. “Dynamical local-field factors and effective interactions in the three-dimensional electron liquid”. In: *Physical Review B* 50.12 (1994), p. 8170.
- [29] Thomas Olsen and Kristian S Thygesen. “Accurate ground-state energies of solids and molecules from time-dependent density-functional theory”. In: *Physical Review Letters* 112.20 (2014), p. 203001.
- [30] Guillermo Román-Pérez and José M Soler. “Efficient implementation of a van der Waals density functional: application to double-wall carbon nanotubes”. In: *Physical review letters* 103.9 (2009), p. 096102.
- [31] Ernest R Davidson et al. “Ground-state correlation energies for two-to ten-electron atomic ions”. In: *Physical Review A* 44.11 (1991), p. 7071.
- [32] Subhas J Chakravorty et al. “Ground-state correlation energies for atomic ions with 3 to 18 electrons”. In: *Physical Review A* 47.5 (1993), p. 3649.

- [33] John P Perdew et al. “Restoring the density-gradient expansion for exchange in solids and surfaces”. In: *Physical review letters* 100.13 (2008), p. 136406.
- [34] Larry A Curtiss et al. “Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation”. In: *The Journal of Chemical Physics* 106.3 (1997), pp. 1063–1079.
- [35] Jess Wellendorff et al. “Density functionals for surface science: Exchange-correlation model development with Bayesian error estimation”. In: *Physical Review B* 85.23 (2012), p. 235149.
- [36] David C Langreth and MJ Mehl. “Beyond the local-density approximation in calculations of ground-state electronic properties”. In: *Physical Review B* 28.4 (1983), p. 1809.
- [37] Susi Lehtola and Miguel AL Marques. “Meta-Local Density Functionals: A New Rung on Jacob’s Ladder”. In: *Journal of chemical theory and computation* 17.2 (2021), pp. 943–948.
- [38] 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 (1992), p. 6671.
- [39] Jianwei Sun et al. “Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional”. In: *Nature chemistry* 8.9 (2016), p. 831.
- [40] Max Dion et al. “Van der Waals density functional for general geometries”. In: *Physical review letters* 92.24 (2004), p. 246401.
- [41] JA Alonso and LA Girifalco. “Nonlocal approximation to the exchange potential and kinetic energy of an inhomogeneous electron gas”. In: *Physical Review B* 17.10 (1978), p. 3735.
- [42] O Gunnarsson, M Jonson, and BI Lundqvist. “Descriptions of exchange and correlation effects in inhomogeneous electron systems”. In: *Physical Review B* 20.8 (1979), p. 3136.
- [43] O Gunnarsson and RO Jones. “Density functional calculations for atoms, molecules and clusters”. In: *Physica Scripta* 21.3-4 (1980), p. 394.
- [44] David J Singh. “Weighted-density-approximation ground-state studies of solids”. In: *Physical Review B* 48.19 (1993), p. 14099.
- [45] Zhigang Wu, Ronald E Cohen, and David J Singh. “First-Principles WDA Calculations for Ferroelectric Materials”. In: *AIP Conference Proceedings*. Vol. 677. 1. American Institute of Physics. 2003, pp. 276–286.
- [46] Zhigang Wu et al. “Weighted-density-approximation description of rare-earth trihydrides”. In: *Physical Review B* 69.8 (2004), p. 085104.
- [47] Michael G Medvedev et al. “Density functional theory is straying from the path toward the exact functional”. In: *Science* 355.6320 (2017), pp. 49–52.
- [48] Wilfried G Aulbur, Lars Jönsson, and John W Wilkins. “Quasiparticle calculations in solids”. In: *Solid state physics (New York. 1955)* 54 (2000), pp. 1–218.
- [49] Henrik Bruus and Karsten Flensberg. *Many-body quantum theory in condensed matter physics: an introduction*. Oxford university press, 2004.
- [50] Viktor Mikhailovich Galitskii and Arkadii Beinusovich Migdal. “Application of quantum field theory methods to the many body problem”. In: *Sov. Phys. JETP* 7.96 (1958), p. 18.

- [51] Falco Hüser, Thomas Olsen, and Kristian S. Thygesen. “Quasiparticle GW calculations for solids, molecules, and two-dimensional materials”. In: *Phys. Rev. B* 87 (23 June 2013), p. 235132. DOI: 10.1103/PhysRevB.87.235132. URL: <https://link.aps.org/doi/10.1103/PhysRevB.87.235132>.
- [52] Lars Hedin. “New Method for Calculating the One-Particle Green’s Function with Application to the Electron-Gas Problem”. In: *Phys. Rev.* 139 (3A Aug. 1965), A796–A823. DOI: 10.1103/PhysRev.139.A796. URL: <https://link.aps.org/doi/10.1103/PhysRev.139.A796>.
- [53] Feliciano Giustino, Marvin L Cohen, and Steven G Louie. “GW method with the self-consistent Sternheimer equation”. In: *Physical Review B* 81.11 (2010), p. 115105.
- [54] Henry Lambert and Feliciano Giustino. “Ab initio Sternheimer-GW method for quasiparticle calculations using plane waves”. In: *Physical Review B* 88.7 (2013), p. 075117.
- [55] Honghui Shang. “The Sternheimer approach to all-electron real-space density-functional perturbation theory with atomic basis set”. In: *AIP Advances* 11.1 (2021), p. 015224.
- [56] Ferdi Aryasetiawan and Olle Gunnarsson. “The GW method”. In: *Reports on Progress in Physics* 61.3 (1998), p. 237.
- [57] M. Shishkin and G. Kresse. “Implementation and performance of the frequency-dependent *GW* method within the PAW framework”. In: *Phys. Rev. B* 74 (3 July 2006), p. 035101. DOI: 10.1103/PhysRevB.74.035101. URL: <https://link.aps.org/doi/10.1103/PhysRevB.74.035101>.
- [58] Ravishankar Sundararaman and TA 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 (2013), p. 165122.
- [59] Jiří Klimeš, Merzuk Kaltak, and Georg Kresse. “Predictive *G W* calculations using plane waves and pseudopotentials”. In: *Physical Review B* 90.7 (2014), p. 075125.
- [60] Hong Jiang and Peter Blaha. “*G W* with linearized augmented plane waves extended by high-energy local orbitals”. In: *Physical Review B* 93.11 (2016), p. 115203.
- [61] Hong Jiang. “Revisiting the *G W* approach to d-and f-electron oxides”. In: *Physical Review B* 97.24 (2018), p. 245132.
- [62] Xavier Gonze et al. “The ABINIT project: Impact, environment and recent developments”. In: *Computer Physics Communications* 248 (2020), p. 107042.
- [63] Miguel AL Marques et al. “octopus: a first-principles tool for excited electron–ion dynamics”. In: *Computer Physics Communications* 151.1 (2003), pp. 60–78.
- [64] Ask Hjorth Larsen, Jens Jørgen Mortensen, et al. “The atomic simulation environment - a Python library for working with atoms”. In: *Journal of Physics: Condensed Matter* 29.27 (2017), p. 273002.
- [65] Jens Jørgen Mortensen, Morten Gjerding, and Kristian Sommer Thygesen. “MyQueue: Task and workflow scheduling system”. In: *Journal of Open Source Software*

- 5.45 (2020), p. 1844. DOI: 10.21105/joss.01844. URL: <https://doi.org/10.21105/joss.01844>.
- [66] *Computational 2D Materials Database (C2DB)*. <https://cmr.fysik.dtu.dk/c2db/c2db.html>. Accessed: 2021-05-04.
- [67] Mark D Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific data* 3.1 (2016), pp. 1–9.
- [68] Kiran Mathew et al. “Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows”. In: *Computational Materials Science* 139 (2017), pp. 140–152. ISSN: 0927-0256. DOI: <https://doi.org/10.1016/j.commatsci.2017.07.030>. URL: <https://www.sciencedirect.com/science/article/pii/S0927025617303919>.
- [69] Georg Kresse and Jürgen Furthmüller. “Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set”. In: *Physical review B* 54.16 (1996), p. 11169.
- [70] Georg Kresse and D. Joubert. “From ultrasoft pseudopotentials to the projector augmented-wave method”. In: *Physical Review B* 59.3 (1999), p. 1758.
- [71] Giovanni Pizzi et al. “AiiDA: automated interactive infrastructure and database for computational science”. In: *Computational Materials Science* 111 (2016), pp. 218–230. ISSN: 0927-0256. DOI: <https://doi.org/10.1016/j.commatsci.2015.09.013>. URL: <https://www.sciencedirect.com/science/article/pii/S0927025615005820>.
- [72] Sebastiaan P Huber et al. “AiiDA 1.0, a scalable computational infrastructure for automated reproducible workflows and data provenance”. In: *Scientific data* 7.1 (2020), pp. 1–18.
- [73] Vladan Stevanović et al. “Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies”. In: *Phys. Rev. B* 85 (11 Mar. 2012), p. 115104. DOI: 10.1103/PhysRevB.85.115104. URL: <https://link.aps.org/doi/10.1103/PhysRevB.85.115104>.
- [74] Mohnish Pandey and Karsten W. Jacobsen. “Heats of formation of solids with error estimation: The mBEEF functional with and without fitted reference energies”. In: *Phys. Rev. B* 91 (23 June 2015), p. 235201. DOI: 10.1103/PhysRevB.91.235201. URL: <https://link.aps.org/doi/10.1103/PhysRevB.91.235201>.
- [75] Yeliang Wang et al. “Monolayer PtSe₂, a new semiconducting transition-metal-dichalcogenide, epitaxially grown by direct selenization of Pt”. In: *Nano letters* 15.6 (2015), pp. 4013–4018.
- [76] Yuda Zhao et al. “High-electron-mobility and air-stable 2D layered PtSe₂ FETs”. In: *Advanced Materials* 29.5 (2017), p. 1604230.
- [77] Lu Hua Li et al. “Strong oxidation resistance of atomically thin boron nitride nanosheets”. In: *ACS nano* 8.2 (2014), pp. 1457–1462.
- [78] Zakaria Y Al Balushi et al. “Two-dimensional gallium nitride realized via graphene encapsulation”. In: *Nature materials* 15.11 (2016), pp. 1166–1171.
- [79] Takahiro Yamada et al. “Comprehensive study on initial thermal oxidation of GaN (0001) surface and subsequent oxide growth in dry oxygen ambient”. In: *Journal of Applied Physics* 121.3 (2017), p. 035303.

- [80] Gaoxue Wang, Ravindra Pandey, and Shashi P Karna. "Physics and chemistry of oxidation of two-dimensional nanomaterials by molecular oxygen". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 7.1 (2017), e1280.
- [81] ME Dávila et al. "Germanene: a novel two-dimensional germanium allotrope akin to graphene and silicene". In: *New Journal of Physics* 16.9 (2014), p. 095002.
- [82] Pavel Afanasiev and Chantal Lorentz. "Oxidation of nanodispersed MoS₂ in ambient air: the products and the mechanistic steps". In: *The Journal of Physical Chemistry C* 123.12 (2019), pp. 7486–7494.
- [83] Hong Ni, Xiaochen Liu, and Qunfeng Cheng. "A new strategy for air-stable black phosphorus reinforced PVA nanocomposites". In: *Journal of Materials Chemistry A* 6.16 (2018), pp. 7142–7147.
- [84] Yuan Huang et al. "Interaction of black phosphorus with oxygen and water". In: *Chemistry of Materials* 28.22 (2016), pp. 8330–8339.
- [85] Alessandro Molle et al. "Hindering the oxidation of silicene with non-reactive encapsulation". In: *Advanced Functional Materials* 23.35 (2013), pp. 4340–4344.
- [86] Jun Feng et al. "Metallic few-layered VS₂ ultrathin nanosheets: high two-dimensional conductivity for in-plane supercapacitors". In: *Journal of the American Chemical Society* 133.44 (2011), pp. 17832–17838.
- [87] Lu Li et al. "Vanadium disulfide flakes with nanolayered titanium disulfide coating as cathode materials in lithium-ion batteries". In: *Nature communications* 10.1 (2019), pp. 1–10.
- [88] Andor Kormányos et al. "k·p theory for two-dimensional transition metal dichalcogenide semiconductors". In: *2D Materials* 2.2 (2015), p. 022001.
- [89] Bo Peng et al. "Chemical intuition for high thermoelectric performance in monolayer black phosphorus, α -arsenene and aW-antimonene". In: *J. Mater. Chem. A* 6 (5 2018), pp. 2018–2033. DOI: 10.1039/C7TA09480A. URL: <http://dx.doi.org/10.1039/C7TA09480A>.
- [90] Morten Niklas Gjerding et al. "Layered van der Waals crystals with hyperbolic light dispersion". In: *Nature communications* 8.1 (2017), pp. 1–8.
- [91] Zubin Jacob, Leonid V Alekseyev, and Evgenii Narimanov. "Optical hyperlens: far-field imaging beyond the diffraction limit". In: *Optics express* 14.18 (2006), pp. 8247–8256.
- [92] Dylan Lu et al. "Enhancing spontaneous emission rates of molecules using nanopatterned multilayer hyperbolic metamaterials". In: *Nature nanotechnology* 9.1 (2014), pp. 48–53.
- [93] N. D. Mermin and H. Wagner. "Absence of Ferromagnetism or Antiferromagnetism in One- or Two-Dimensional Isotropic Heisenberg Models". In: *Phys. Rev. Lett.* 17 (22 Nov. 1966), pp. 1133–1136. DOI: 10.1103/PhysRevLett.17.1133. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.17.1133>.
- [94] Daniele Torelli and Thomas Olsen. "Calculating critical temperatures for ferromagnetic order in two-dimensional materials". In: *2D Materials* 6.1 (2018), p. 015028.
- [95] Yuan Cao et al. "Unconventional superconductivity in magic-angle graphene superlattices". In: *Nature* 556.7699 (2018), pp. 43–50.
- [96] Bevin Huang et al. "Layer-dependent ferromagnetism in a van der Waals crystal down to the monolayer limit". In: *Nature* 546.7657 (2017), pp. 270–273.

- [97] Nikhil Sivadas et al. "Stacking-dependent magnetism in bilayer CrI₃". In: *Nano letters* 18.12 (2018), pp. 7658–7664.
- [98] Jun Xiao et al. "Berry curvature memory through electrically driven stacking transitions". In: *Nature Physics* 16.10 (2020), pp. 1028–1034.
- [99] Kenji Yasuda et al. "Stacking-engineered ferroelectricity in bilayer boron nitride". In: *Science* (2021).
- [100] Kin Fai Mak et al. "Atomically thin MoS₂: a new direct-gap semiconductor". In: *Physical review letters* 105.13 (2010), p. 136805.
- [101] Kai Yan et al. "Formation of bilayer bernal graphene: layer-by-layer epitaxy via chemical vapor deposition". In: *Nano letters* 11.3 (2011), pp. 1106–1110.
- [102] Stefan Grimme et al. "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu". In: *The Journal of chemical physics* 132.15 (2010), p. 154104.
- [103] Shyue Ping Ong et al. "Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis". In: *Computational Materials Science* 68 (2013), pp. 314–319.
- [104] Peter Mahler Larsen et al. "Definition of a scoring parameter to identify low-dimensional materials components". In: *Physical Review Materials* 3.3 (2019), p. 034003.
- [105] Jiangang He, Kerstin Hummer, and Cesare Franchini. "Stacking effects on the electronic and optical properties of bilayer transition metal dichalcogenides MoS₂, MoSe₂, WS₂, and WSe₂". In: *Physical Review B* 89.7 (2014), p. 075409.
- [106] Mazhar N Ali et al. "Large, non-saturating magnetoresistance in WTe₂". In: *Nature* 514.7521 (2014), pp. 205–208.
- [107] Zaiyao Fei et al. "Edge conduction in monolayer WTe₂". In: *Nature Physics* 13.7 (2017), pp. 677–682.

Paper I

Anisotropic properties of monolayer 2D materials: An overview from the C2DB database

Luca Vannucci, Urko Petralanda, **Asbjørn Rasmussen**, Thomas Olsen, Kristian S. Thygesen

J. Appl. Phys. 128, 105101 (2020) - Published 08 September 2020

Anisotropic properties of monolayer 2D materials: An overview from the C2DB database

Cite as: J. Appl. Phys. **128**, 105101 (2020); <https://doi.org/10.1063/5.0021237>

Submitted: 08 July 2020 . Accepted: 21 August 2020 . Published Online: 08 September 2020

 Luca Vannucci,  Urko Petralanda, Asbjørn Rasmussen,  Thomas Olsen, and  Kristian S. Thygesen

COLLECTIONS

Paper published as part of the special topic on [Beyond Graphene: Low Symmetry and Anisotropic 2D Materials](#)

 This paper was selected as an Editor's Pick



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Beyond Graphene: Low-Symmetry and Anisotropic 2D Materials](#)

Journal of Applied Physics **128**, 140401 (2020); <https://doi.org/10.1063/5.0030751>

[Point defects in two-dimensional hexagonal boron nitride: A perspective](#)

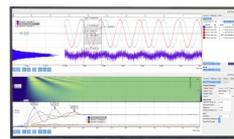
Journal of Applied Physics **128**, 100902 (2020); <https://doi.org/10.1063/5.0021093>

[A practical guide for crystal growth of van der Waals layered materials](#)

Journal of Applied Physics **128**, 051101 (2020); <https://doi.org/10.1063/5.0015971>

Challenge us.

What are your needs for
periodic signal detection?



Zurich
Instruments

Anisotropic properties of monolayer 2D materials: An overview from the C2DB database



Cite as: J. Appl. Phys. **128**, 105101 (2020); doi: [10.1063/5.0021237](https://doi.org/10.1063/5.0021237)

Submitted: 8 July 2020 · Accepted: 21 August 2020 ·

Published Online: 8 September 2020



Luca Vannucci,^{1,a)} Urko Petralanda,^{1,b)} Asbjørn Rasmussen,^{1,2} Thomas Olsen,¹ and Kristian S. Thygesen^{1,2,c)}

AFFILIATIONS

¹CAMD, Department of Physics, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

²Center for Nanostructured Graphene (CNG), Technical University of Denmark, 2800 Kongens Lyngby, Denmark

Note: This paper is part of the Special Topic on: Beyond Graphene: Low Symmetry and Anisotropic 2D Materials.

a) Author to whom correspondence should be addressed: lucav@dtu.dk

b) Electronic mail: upeho@dtu.dk

c) Electronic mail: thygesen@fysik.dtu.dk

ABSTRACT

We analyze the occurrence of in-plane anisotropy in the electronic, magnetic, elastic, and transport properties of more than 1000 2D materials from the C2DB database. We identify hundreds of anisotropic materials and classify them according to their point group symmetry and degree of anisotropy. A statistical analysis reveals that a lower point group symmetry and a larger amount of different elements in the structure favor all types of anisotropies, which could be relevant for future material design approaches. Besides, we identify novel compounds, predicted to be easily exfoliable from a parent bulk compound, with anisotropies that largely outscore those of already known 2D materials. Our findings provide a comprehensive reference for future studies of anisotropic response in atomically thin crystals and point to new previously unexplored materials for the next generation of anisotropic 2D devices.

Published under license by AIP Publishing. <https://doi.org/10.1063/5.0021237>

I. INTRODUCTION

Anisotropy is the characteristic of a material whereby it displays different physical properties along different directions. It is intrinsic to the atomic structure and can, therefore, influence the electric, magnetic, optical, or mechanical response of a material to an external perturbation. In fact, anisotropic materials have become increasingly present in modern devices, finding applications in diverse fields. One paradigmatic example of anisotropic material is an optical polarizer, which is transparent to electromagnetic radiation polarized along a well defined axis, while blocking or deviating light that is polarized along a different direction.

Layered van der Waals (vdW) materials represent an interesting class of naturally anisotropic materials. In vdW materials, the anisotropy derives from the dispersive nature of the bonds between the two-dimensional (2D) atomic layers, which is much weaker than the covalent bonds existing between atoms within the 2D layers. This intrinsic anisotropy can be exploited for various applications. For example, in certain layered materials, the interplay between the structural and electronic properties is so strong that it

changes the iso-frequency surfaces of light from elliptic to hyperbolic¹ with fascinating perspectives for sub-wavelength imaging and radiative emission control.^{2,3}

Individual 2D atomic layers are obviously anisotropic due to the missing third dimension, but they can exhibit in-plane anisotropy as well. However, the most widely studied 2D materials—graphene,⁴ hexagonal boron nitride (hBN),⁵ and the family of transition metal dichalcogenides (TMDCs)⁶—have in-plane isotropic properties due their highly symmetric crystal structure. Graphene, for instance, has a sixfold rotational symmetry and three mirror planes, while hBN and TMDCs such as MoS₂ have a sixfold roto-inversion symmetry with two mirror planes. Such large sets of crystal symmetries turn out to inhibit any form of anisotropic response.

The prototypical example of in-plane anisotropic 2D material is phosphorene.⁷ Phosphorene is obtained by mechanical exfoliation of black phosphorus down to the monolayer limit and exhibits a highly anisotropic puckered structure, which differs along the zigzag and armchair direction (as shown in Fig. 1). This strong anisotropy has motivated a large number of theoretical and

experimental studies of phosphorene, which have revealed the effect of the structural anisotropy on its electronic, optoelectronic, electro-mechanical, thermal, and excitonic properties.³⁻¹⁷

Other notable examples of in-plane anisotropic 2D materials are TMDCs in the distorted 1T'-phase (such as WTe₂^{1,18-23}), titanium trichalcogenides (most notably TiS₃²³⁻²⁸), ReSe₂ and ReSe₂,^{16,29-32} GaTe,³³ and pentagonal structures such as PdSe₂.^{34,35} Such materials exhibit anisotropy in their mechanical, electrical, optical, and magnetic properties, with intriguing applications for optical devices (such as birefringent wave plate or hyperbolic plasmonic surfaces), high mobility transistors, ultra-thin memory devices, and controllable magnetic devices among others.

As of today, more than 50 different 2D materials have been identified and synthesized or exfoliated in monolayer form,³⁶ but they represent only a small fraction of all the possible stable 2D

materials that have been predicted by computations.³⁶⁻⁴¹ It is, therefore, reasonable to expect that the above mentioned examples of anisotropic 2D materials will be soon complemented by additional atomically thin layers with highly direction-dependent properties.

Here, we take a first step in this direction by presenting an extensive analysis of the occurrence of in-plane anisotropic features in the magnetic, elastic, transport, and optical properties of more than 1000 predicted stable 2D materials from the Computational 2D Materials Database (C2DB).³⁶ We discuss trends and similarities in the atomic and electronic structure of anisotropic monolayer materials by classifying them according to their point symmetry group and highlight the most interesting candidates for different applications.

After introducing the C2DB database and the criteria used to assess the stability of the materials in Sec. II, we analyze the occurrence

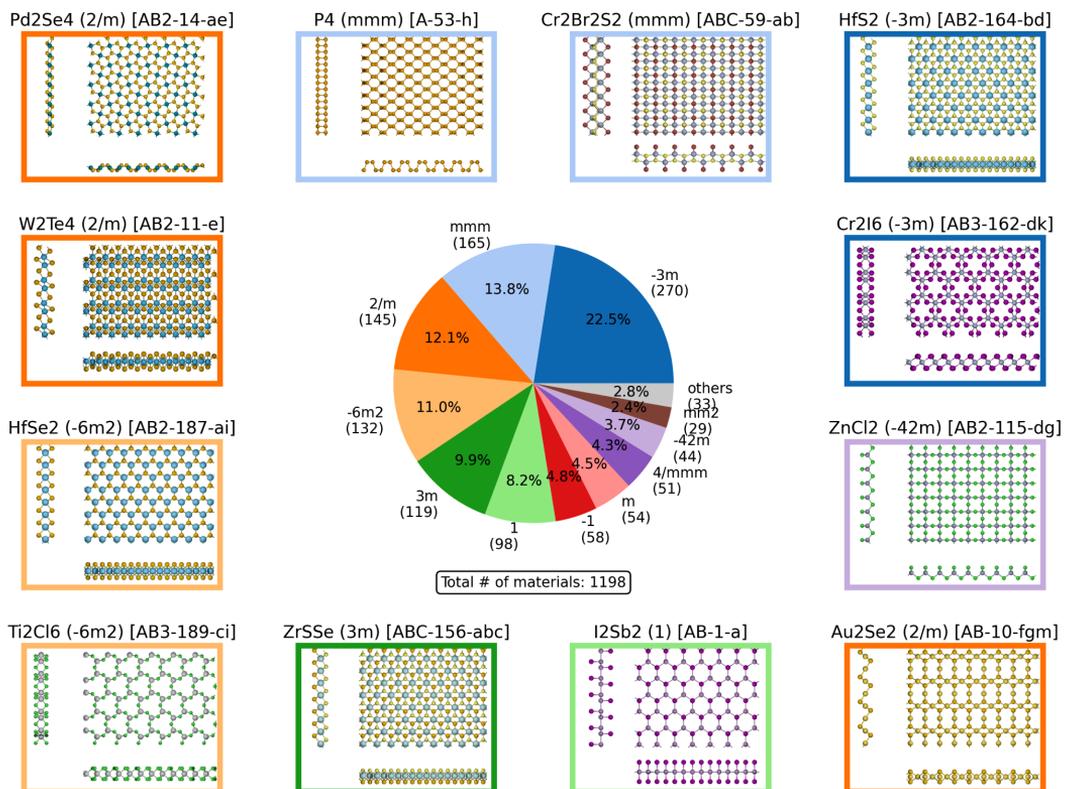


FIG. 1. Overview of the point group distribution of highly stable materials in C2DB, with some examples of corresponding crystal structures. Each example is accompanied by the crystal structure prototype, which is a label of the form $S-n-p$ with S being the stoichiometry, n being the space group number, and p being the set of occupied Wyckoff positions. The following point groups representing less than 2% of stable materials are omitted in the figure: 2 (17%, 20 materials); -3 (0.8%, 8 materials); 3 (0.2%, 2 materials); 32 (0.2%, 2 materials); $6/mmm$ (0.1%, 1 material).

of anisotropy in the magnetic easy axis direction, elastic response, effective masses, and polarizability in four separate sub-sections of Sec. III. We have tried to make these sections as self-contained as possible so they can be read independently, with a separate introduction to the formalism used and the relevant literature for each of them. We conclude by summarizing the main results in Sec. IV, where we highlight the most interesting anisotropic and potentially exfoliable 2D materials identified in the study.

II. OVERVIEW OF THE C2DB DATABASE

The Computational 2D Materials Database (C2DB) is an open database containing thermodynamic, elastic, electronic, magnetic, and optical properties of 2D monolayer materials.³⁶ All properties were calculated with the electronic structure code GPAW⁴² using additional software packages for atomic simulation and workflows handling such as ASE,⁴³ ASR,⁴⁴ and MyQueue.⁴⁵ Unless explicitly stated, all properties reported in this work were calculated with the PBE exchange-correlation functional.⁴⁶

The latest development version of C2DB contains 4262 fully relaxed structures at the time of writing, which are categorized in terms of their dynamic and thermodynamic stability. The dynamic stability determines whether a material is stable with respect to distortions of the atomic positions or the unit cell and is established from phonon frequencies (at the Γ -point and high-symmetry points of the Brillouin zone boundary) and the stiffness tensor. A material is dynamically stable only if all phonon frequencies are real and the stiffness tensor eigenvalues are positive. On the other hand, the thermodynamic stability of a given 2D material is assessed in terms of its heat of formation and total energy with respect to other competing phases (taken as the most stable elemental and binary compounds from the OQMD database⁴⁷)—also known as *energy above the convex hull*.³⁶ A material's thermodynamic stability is classified as *high* if the heat of formation is negative and the energy above the convex hull is below 0.2 eV/atom.

Materials with high thermodynamic and dynamic stability are the most likely to be exfoliated or synthesized in the lab. Although these criteria are not sufficient to ensure experimental realization, we note that they have been determined from a detailed analysis of more than 50 already synthesized monolayers.³⁶ We will, therefore, focus on the subset of highly stable materials (according to the criteria used in the C2DB) in the remainder of this work. For further details on the stability assessment and a complete overview of the C2DB, we refer the reader to Ref. 36.

As shown by Neumann more than one century ago, the symmetries of any physical property of a material must include the symmetry operations of the point group to which the crystalline lattice belongs.⁴⁸ Figure 1 provides an overview of the 1198 thermodynamically and dynamically stable materials in C2DB grouped according to their point group symmetry, with the latter obtained from the Spglib library.⁴⁹ Some specific examples of materials are shown with their point group indicated by the color of the frame. The selected materials represent some of the most interesting anisotropic 2D materials identified in this work and discussed in the following.

From Fig. 1, we make the following general observations:

- Materials with trigonal symmetry, that is, materials belonging to the point groups $-3m$, $3m$, -3 , 3 , 32 in the international

- notation,⁵⁰ are the most frequently occurring (33% of the total). These include, among others, TMDCs in the 1T phase such as HfS₂,⁵¹ group IV monolayers,⁵² hydrogenated graphene (i.e., graphene⁵³), MXY Janus structures^{54–56} such as ZrSSe, and monolayer magnetic materials such as CrI₃.⁵⁷
- Monoclinic materials (groups 2, m , $2/m$) account for 18% of the total. They include TMDCs in the 1T' phase, such as WTe₂,⁵⁸ TiS₃,^{23,24} and the pentagonal PdSe₂.³⁴ Of the 219 monoclinic materials investigated in this work, 140 of them bear orthogonal structure, while the remaining 79 have an inclined crystal system.
- The orthorhombic structures comprise 16% of the total (groups mmm , $mm2$). Notable examples are the highly anisotropic puckered phosphorene (that is, monolayer black phosphorus⁷) and puckered arsenene.⁵⁹ We also point out the magnetic ternary compound CrSBr, which has been recently exfoliated from the layered bulk structure^{60,61} and whose crystal prototype is largely recurrent in C2DB among orthorhombic structures.
- Triclinic materials (groups 1 , -1) account for 13% of the total. This group include materials with low symmetry, such as TMDC alloys,^{62,63} the topological insulator SbI₃,^{64,65} and other potentially exfoliable materials such as AuSe.
- 11% of structures have hexagonal point group symmetry (groups $-6m2$, $6/mmm$). They include TMDCs in the H phase such as HfSe₂, hexagonal boron nitride (hBN), graphene (which is the only stable representative of the point group $6/mmm$), and other less common structures possessing sixfold rotation symmetry such as TiCl₃.
- The remaining 8% correspond to tetragonal structures (groups $-42m$, $4/mmm$) such as ZnCl₂, which is predicted to be easily exfoliable.³⁷

This set of 1198 known or potentially exfoliable/synthesizable materials forms the basis for the anisotropy analysis presented in this work.

III. RESULTS AND DISCUSSION

A. Magnetic easy axis

Magnetic anisotropy is defined in a material as the dependence of its properties on the direction of its magnetization. The main manifestation of magnetic anisotropy is the existence of an easy axis, along which it takes the least energy to magnetize the crystal, and a hard axis, where it takes the most. In order to quantify the degree of anisotropy, the magnetic anisotropy energy (MAE) is defined, which accounts for the energy necessary to deflect the magnetization from the easy to the hard direction. In general, the MAE may have contributions from different features of a crystal such as strain or defects. In this work, we will consider perfect crystals, wherein only the so-called magnetocrystalline anisotropy, given by the coupling of the lattice to the spin magnetic moment, contributes to the MAE.

In 2D materials, magnetic anisotropy gains a special importance due to the Mermin-Wagner theorem,⁶⁶ which prohibits a broken symmetry phase at finite temperatures. This means that for a magnetic order to emerge, the spin rotational symmetry has to be broken explicitly by magnetic anisotropy. This has attracted a wide

interest on the topic in the recent years, both in the light of new fundamental questions^{57,67-71} and applications.⁷²⁻⁷⁶

In this work, we will focus on the in-plane MAE and define the x and y axes to span the atomic plane of the material. We define the in-plane MAE, Δ_{xy} , as

$$\Delta_{xy} = |E(\vec{M} \parallel y) - E(\vec{M} \parallel x)|, \quad (1)$$

where $E(\vec{M} \parallel x)$ and $E(\vec{M} \parallel y)$ are the electronic energies including spin-orbit coupling (SOC) with magnetization parallel to the x and y axes, respectively. In the electronic structure code GPAW, SOC is added non-self-consistently on top of a converged PBE calculation as described in Ref. 77. This method yields very accurate results, as evidenced, for instance, by the excellent agreement of MoS₂ SOC-induced band splitting with other first-principles approaches and with experiments,⁷⁷ or by the correct description of topological physics governed by SOC.^{65,78} The MAE obtained with this method has been used in the calculation of critical temperature of magnetic 2D materials, providing good agreement with experiments.^{70,71}

Let us note that the definition of x and y axes is of course arbitrary. In C2DB, the x axis is systematically chosen to be parallel to one of the lattice vectors, with the y axis orthogonal to it. All properties are then calculated with respect to this reference frame. With this choice, the anisotropy is assessed with respect to the high symmetry axes for most of the materials in C2DB, except only for low symmetric oblique structures such as the ones in point group 1 and -1 (which represent only a 13% of the total). For such cases, the degree of anisotropy for some properties might be slightly underestimated.

In Fig. 2, we show the distribution of the magnetic materials in the C2DB database, sorted by their point group symmetry and the value of their in-plane magnetic anisotropy Δ_{xy} . In comparison with Fig. 1, we find a similar landscape once we filter for ferromagnetic (FM) or anti-ferromagnetic (AFM) materials, as shown in Fig. 2(a). This indicates that being magnetic or not is not strongly correlated to the point group but to the presence of magnetic atoms in the structure. However, once anisotropy comes into play, we do observe, in Fig. 2(b), important structural features that condition it. In fact, one can expect a magnetic easy axis to appear in the direction where magnetic atoms are packed more loosely, creating an anisotropy in the magnetic properties. For example, as we set a very low (0.005 meV/unit cell) threshold for Δ_{xy} , all hexagonal and tetragonal point groups vanish and only point groups not restricting the in-plane perpendicular directions by symmetry hold an in-plane magnetic anisotropy. Another feature we observe from Fig. 2(c) is the prevalence of the orthorhombic (mmm) and, to a lower degree monoclinic ($2/m$), systems with crystals of mmm point group symmetry representing over 35% of the materials with a low Δ_{xy} threshold. When we increase the Δ_{xy} threshold to 0.7 meV/unit cell, we see that the trend is enhanced: mmm dominates with half of the materials and $2/m$ comprises a third of the materials [Fig. 2(d)]. The materials above this threshold are classified and sorted according to their anisotropy in Fig. 2(e). We see that both FM and AFM magnetic orders are equally represented, indicating little influence of the type of magnetic order on the anisotropy. In Fig. 2(e), we also show the direction of the magnetic

easy axis, indicated by a full orange marker if it lies within the plane and an empty blue one if it is oriented out-of-plane. It is clear that most of the selected anisotropic materials indeed present an in-plane easy axis.

Among the 113 materials with $\Delta_{xy} > 0.005$ meV/unit cell, we find that the ternary compound structure prototype of orthorhombic symmetry $ABC-59-ab$ ³⁶ is the most frequently occurring with 47 entries (see CrSBr in Fig. 1 for an example of this structure). The main reason for this might be the mentioned lack of symmetry between the x and y directions in the plane, along with the fact that it is more likely to contain a magnetic atom due to its ternary nature (most other crystals in C2DB are binary). To the best of our knowledge, the only 2D material from this group that has been successfully synthesized and exfoliated is CrSBr, whose FM order down to the monolayer limit has been very recently confirmed in experiments.^{60,61} We note that several materials bearing the very same structure prototype are listed as easily exfoliable in Ref. 37, e.g., CrOBr, CrOCl, FeOCl, VOBr, and VOCl. Finally, the monoclinic T' phase of transition metal dichalcogenides occurs 15 times, followed by the trigonal MoS₂⁷⁹ type with 10 occurrences.

We also cross-checked the rest of our selected anisotropic materials against the list of exfoliable 2D materials in Ref. 37. We found, out of the 113 materials with $\Delta_{xy} > 0.005$ meV/unit cell, over 20 materials whose stoichiometry match entries in the list of exfoliable materials. Among these, perhaps the most promising material with regard to a potential experimental realization is the AFM T' di-halide V₂I₄, which lies at the convex hull according to the C2DB database.³⁶ V₂I₄ shows an in-plane magnetic easy axis and $\Delta_{xy} = 1.09$ meV/unit cell, that competes with the highest out-of-plane anisotropies known to date.⁷⁰ In addition, we find several materials that are only a few meV above the convex hull and show remarkably high anisotropies. Among these, the AFM Ni₂I₄ compound stands out with an exceptional in-plane anisotropy of over 20 meV/unit cell and an in-plane easy axis. Other materials in the same stability category, such as Ni₂Br₄, Co₂O₄, and CrBr₂, also show large Δ_{xy} values and are listed in Table I.

B. Elastic response and auxetic effect

The elastic response of 2D materials to strains and deformations is usually expressed in terms of the Young modulus E and the Poisson ratio ν .^{80,81} The former measures the response along a direction that is parallel to the applied strain, while the latter describes how the material reacts along orthogonal directions. For anisotropic materials, both the Young modulus and the Poisson ratio depend on the directions of stresses and strains. Assuming that the 2D material lies in the xy plane, and neglecting the elastic response along the out-of-plane axis z , we will denote the axis-dependent Young's modulus with E_i , $i = \{x, y\}$. Similarly, the coefficient relating the stress along the i axis to an applied strain in the perpendicular j direction will be quantified by the Poisson ratio ν_{ij} , with $i \neq j$.

More generally, the elastic response of a continuous 2D medium is quantified in terms of the 2D stiffness tensor C , which is a linear map between the strain tensor ϵ and the stress tensor σ ,⁸²

$$\sigma_{ij} = \sum_{kl} C_{ijkl} \epsilon_{kl}. \quad (2)$$

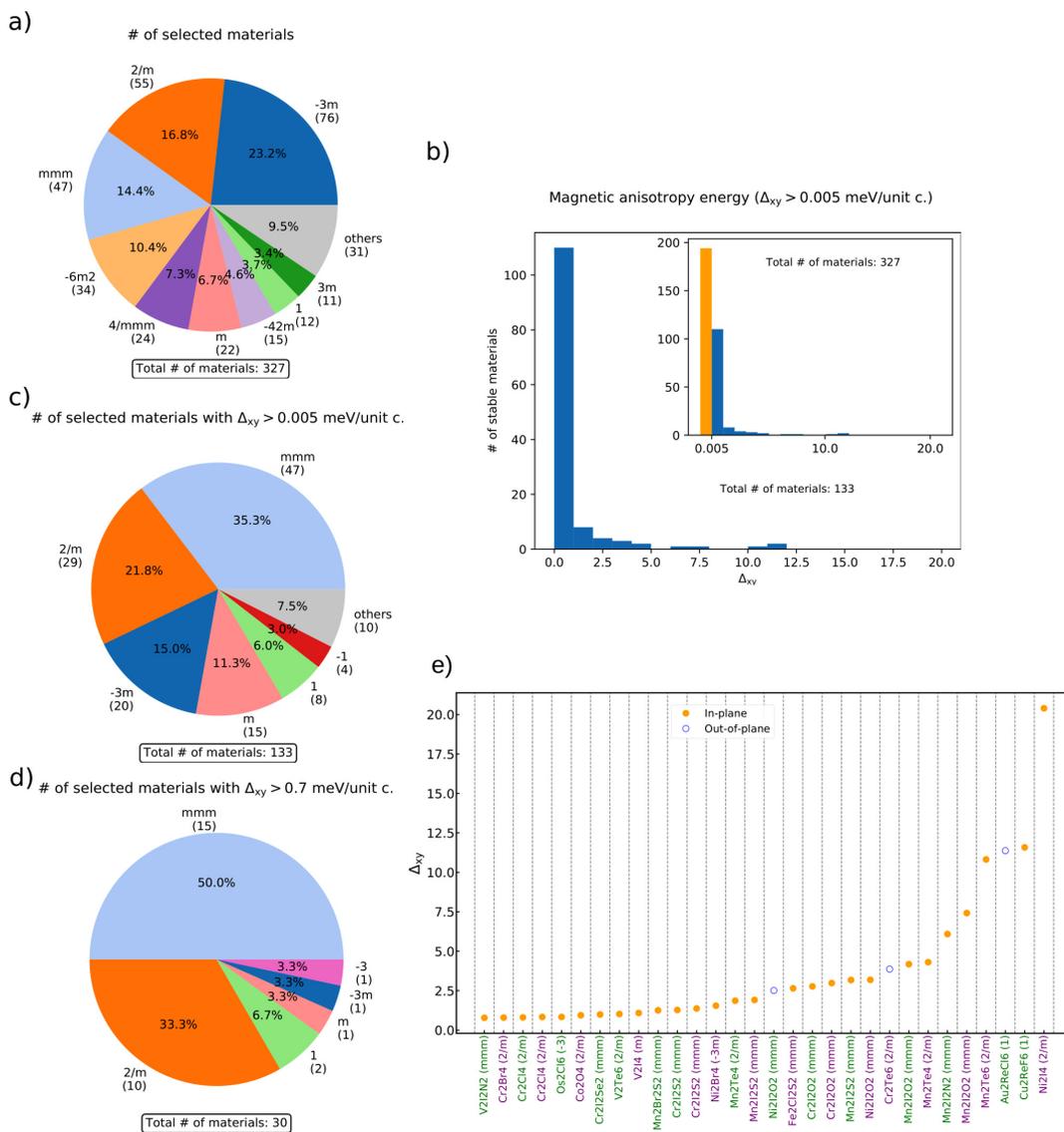


FIG. 2. (a) Distribution of predicted stable magnetic materials (FM or AFM) from the C2DB database grouped according to their point group symmetry. The percentage of materials represented by each point group and their number of occurrences in the database is explicitly shown, except for those point groups representing less than 3% of materials. Panels (b) and (c) show the distribution of magnetic materials in the database with a low (0.005 meV/unit cell) in-plane magnetic anisotropy (MA) threshold. The main chart in (b) shows the distribution of the materials above the threshold, while the orange bin in the inset represents all magnetic materials below the threshold for comparison. Panels (d) and (e) show the distribution of the materials with a high MA (>0.7 meV/unit cell). In (e), the materials are sorted by the size of their MA and their magnetic state is indicated by the font color (green for FM, purple for AFM). The marker color corresponds to the signs of Δ_{xx} and Δ_{yy} : empty marker with blue edge for $\Delta_{xx} < 0$ and $\Delta_{yy} < 0$ (out-of-plane easy axis) and full orange if any of the two is above 0 (in-plane easy axis).

TABLE I. Monolayers predicted stable and with the highest in-plane magnetic anisotropy in the C2DB database and in-plane magnetic easy axis, whose stoichiometry matches that of entries in the list of easily exfoliable 2D materials in Ref. 37. The table shows the chemical formula, the space group symmetry, magnetic state, energy above the convex hull, and in-plane magnetic anisotropy.

	Sym.	Mag.	E_{hull} (meV)	Lowest E_{hull} monolayer?	Δ_{xy} (meV/unit c.)
Cr ₂ Br ₄	P2 ₁ /m	FM	54.4	No	0.79
Co ₂ O ₄	C2/m	FM	7.1	No	0.94
V ₂ I ₄	Pm	AFM	0.0	Yes	1.09
Ni ₂ Br ₄	P $\bar{3}$ m1	AFM	8.8	No	1.55
Ni ₂ I ₄	C2/m	AFM	10.3	No	20.40

Here, we have $i = \{x, y\}$, since we restrict to in-plane stresses and strains. A generic matrix element σ_{ij} represents the i component of the stress acting on a plane perpendicular to the j direction, while the strain components ε_{ij} are given by $\varepsilon_{ij} = (\partial_i u_j + \partial_j u_i)/2$ in terms of the infinitesimal deformations u_i .

Being a linear map between two 2nd rank tensors, the stiffness tensor is naturally a 4th rank tensor. However, one can make use of the symmetric properties of both σ and ε at equilibrium to write both of them as one-dimensional vectors, namely,

$$\vec{\sigma} = (\sigma_{xx}, \sigma_{yy}, \sigma_{xy})^T := (\sigma_1, \sigma_2, \sigma_3)^T, \quad (3)$$

$$\vec{\varepsilon} = (\varepsilon_{xx}, \varepsilon_{yy}, 2\varepsilon_{xy})^T := (\varepsilon_1, \varepsilon_2, \varepsilon_3)^T. \quad (4)$$

Such a notation is often called *Voigt* notation. Then, the stiffness tensor becomes a 2nd rank symmetric tensor with only six independent components,

$$\vec{\sigma} = \begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{12} & C_{22} & C_{23} \\ C_{13} & C_{23} & C_{33} \end{pmatrix} \vec{\varepsilon}. \quad (5)$$

We will restrict the following analysis to the class of *orthotropic materials*, that is, materials having three mutually orthogonal planes of reflection symmetry. In such a case, the stiffness tensor takes the form

$$C = \begin{pmatrix} C_{11} & C_{12} & 0 \\ C_{12} & C_{22} & 0 \\ 0 & 0 & C_{33} \end{pmatrix}. \quad (6)$$

In practice, this means that we restrict attention to materials where the shear deformations ε_{xy} are decoupled from xx and yy stresses. This allows us to straightforwardly relate the components C_{ij} of the stiffness tensor to the in-plane Young modulus E_i and the in-plane Poisson ratio ν_{ij} via the following relations:

$$E_x = \frac{C_{11}C_{22} - C_{12}^2}{C_{22}}, \quad (7a)$$

$$E_y = \frac{C_{11}C_{22} - C_{12}^2}{C_{11}}, \quad (7b)$$

$$\nu_{xy} = \frac{C_{12}}{C_{11}}, \quad (7c)$$

$$\nu_{yx} = \frac{C_{12}}{C_{22}}. \quad (7d)$$

In C2DB, each component of the 2D stiffness tensor is calculated by straining a material along a given direction (x or y) and calculating the forces acting on the unit cell after relaxing the position of the atoms within the fixed unit cell.³⁶ To restrict to orthotropic materials only, we have discarded all materials whose stiffness tensor components C_{13} or C_{23} exceed a certain tolerance C_{max} , which we set to $C_{\text{max}} = 0.01$ N/m. With this method, we have obtained a subset of 555 materials (roughly 50% of all the stable materials) that we analyze in the following.

In Fig. 3(a), we show an overview of the direction-dependent Young modulus for all orthotropic and stable materials in C2DB. We plot $\max(E_x, E_y)$ against $\min(E_x, E_y)$, which means that all data point lying outside the diagonal represent a material with anisotropic elastic properties. Well known anisotropic structures such as WTe₂, PdSe₂, TiS₃, P₄, and As₄ are all identified by this method, while hundreds of unexplored anisotropic materials are predicted as well. In Fig. 3(b), we use a similar method to show the anisotropy of the Poisson ratio. While this does not add much information with respect to panel (a)—since $E_x/E_y = \nu_{yx}/\nu_{xy}$, as one can easily infer from Eqs. (7)—we notice that Poisson ratios can also take negative values, differently from the Young modulus. In such a case, a material stretched (or compressed) along the x direction will also expand (shrink) along the perpendicular y direction, a quite counterintuitive property called *auxetic* behavior.^{80,83} We will investigate such cases in more detail in the following.

To describe elastic anisotropy in a more quantitative manner, we define an elastic anisotropy degree (or anisotropy parameter) for each material as

$$\delta_E = \frac{|E_x - E_y|}{E_x + E_y}. \quad (8)$$

Such a parameter will be always bounded between 0 and 1, with $\delta_E = 0$ signifying a perfectly isotropic materials, while $\delta \approx 1$ for an extremely anisotropic medium.

In Fig. 3(c), we show the distribution of the elastic anisotropy degree for all materials having $\delta_E \geq 0.05$ (corresponding to a difference of at least 10% between x and y Young's modulus), with the

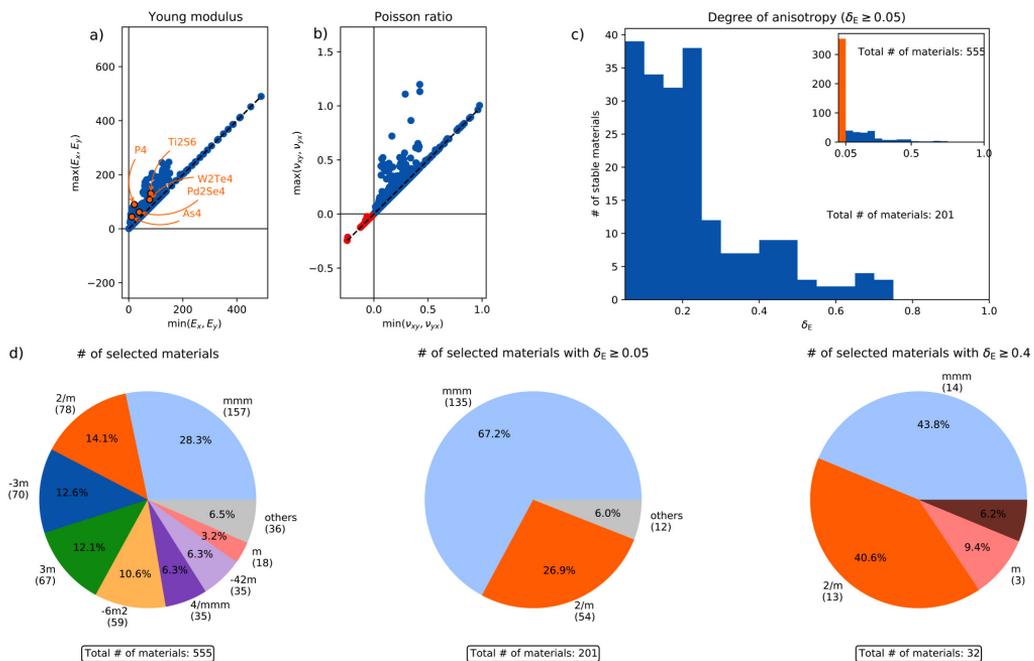


FIG. 3. Elastic properties of the 555 orthotropic materials in C2DB. Panels (a) and (b) show a comparison between the Young modulus and Poisson ratios along opposite directions, respectively. In panel (a), some known anisotropic materials are highlighted in orange, while in panel (b), we use red markers for auxetic materials. Panel (c) shows the distribution of the anisotropic degree δ_E for the 201 materials having $\delta_E \geq 0.05$, with the inset showing the full distribution of all 555 materials for comparison. Panel (d) shows the point group distributions of orthotropic materials for three different threshold values of δ_E (namely, 0, 0.05, and 0.4 from left to right).

inset showing the full distribution including materials with $\delta_E < 0.05$. We notice that more than one third of the selected materials (201 out of 555) show an elastic anisotropy exceeding this threshold value, while 162 of them exceed the value $\delta_E = 0.1$ (signifying a difference of roughly 20% or more between E_x and E_y) and 32 of them show a highly anisotropic elastic behavior with $\delta_E \geq 0.4$.

The distribution of point groups corresponding to different threshold values for δ_E is shown as a series of pie charts in Fig. 3(d). On the left, we plot the distribution of point groups for all 555 selected materials. A comparison with Fig. 1 shows that our choice of selecting only orthotropic materials tends to favor orthorhombic structures (especially, group mmm) with respect to trigonal ones, while the proportions between remaining point groups remain basically unaffected. However, when selecting all materials with at least 10% of difference between E_x and E_y ($\delta_E \geq 0.05$, in the middle), the proportions change drastically, with all trigonal and hexagonal groups suppressed in favor of orthorhombic and monoclinic structures. This shows that symmetric crystal structures such as the ones of TMDCs in the H and T phase, graphene, and hBN are generally isotropic, with little difference in the elastic properties along x and y directions. On the other hand, TMDCs in the distorted T' phase

(such as WTe_2), pentagonal structures ($PdSe_2$), and puckered layers (phosphorene) stand out for their markedly anisotropic elastic properties due to their asymmetric crystal lattice.

When restricting to highly anisotropic materials having $\delta_E \geq 0.4$, monoclinic and orthorhombic structures share exactly 50% of the total each. The Young moduli of these 32 structures are plotted in the top panel of 4, sorted from the lowest to highest value of δ_E . Besides known structures such as phosphorene (P_4) and puckered arsenene (As_4), we find many new stable structures with exceptionally high elastic anisotropy. Four out of the first six materials are compounds of the form CrX_2 (with X a halogen element) in both the AFM and FM magnetic state, which also stand out for their markedly anisotropic magnetic behavior as described previously. These are, however, not the most stable structures with the same constituent elements, since they all have a competing phase of the form CrX_3 with a more favorable formation energy (one of them is shown in Fig. 1). However, this is not the case for the monoclinic structures AuSe and AuTe, which represent the most stable phase of their respective elements. One of them (AuSe, which is shown in Fig. 1 as well) has also been identified as an easily exfoliable materials by independent work of Mounet

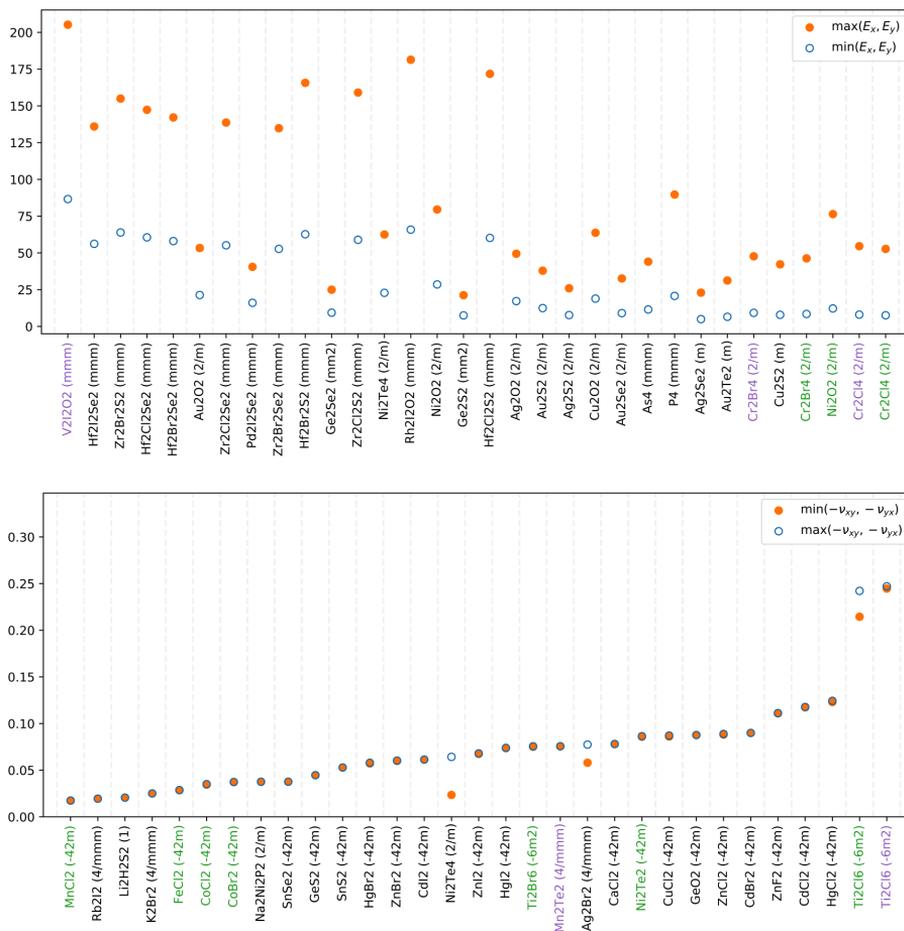


FIG. 4. Top: Materials with highly anisotropic Young modulus $\delta_E \geq 0.4$, sorted from the lowest to highest δ_E . Bottom: Materials with negative Poisson ratio, sorted from the lowest to highest value of $\max(-\nu_{xy}, -\nu_{yx})$. Green (purple) labels indicates ferromagnetic (antiferromagnetic) materials.

et al.,³⁷ making this one of the most appealing material for anisotropic elastic applications found in this work. We also note that puckered compounds GeS and GeSe seem also to be easily exfoliable from their respective three-dimensional parent structures, which is again confirmed in the literature.³⁷

Finally, it is worth mentioning the presence of several entries in the structural prototype *ABC-59-ab* (especially, Hf- and Zr-based compounds), whose relevance has already been discussed Sec. III A. We note that Ref. 37 lists HfNBr, ZrNBr, and ZrNI as easily exfoliable layered materials. We find a relatively low elastic anisotropy degree $\delta_E = 0.07$ for ZrNI, but we suggest that materials

with much higher elastic anisotropy such as HfBrX and ZrBrX (with X = S, Se) should, in principle, be available by substitution of nitrogen with an element from the halogen group.

Let us now come back to the subset of materials showing a negative in-plane Poisson ratio, that is, the ones highlighted with red markers in Fig. 3(b). The auxetic effect is not necessarily associated to anisotropy as both Poisson ratios ν_{xy} and ν_{yx} can take negative values without necessarily being different from each other. Indeed, such an effect does not originate from the material having a different elastic response along orthogonal axes but rather from the presence of special re-entrant structures or rigid blocks linked

by flexible hinges in the crystalline structure that can compress or extend in counter-intuitive fashions. Nevertheless, our framework allows for the systematic search of novel 2D materials with a negative Poisson ratio, which itself is an active field of research.⁸³ Moreover, Poisson ratios of anisotropic materials can take arbitrarily large values (positive or negative), differently from ordinary isotropic media.⁸⁴

In the bottom panel of Fig. 4, we plot the Poisson ratios of the 31 stable materials in C2DB showing auxetic behavior, sorting from the lowest to highest value of $\max(-\nu_{xy}, -\nu_{yx})$. The largest negative Poisson ratio is found for TiCl_3 in the hexagonal crystal structure (shown in Fig. 1), with both AFM and FM magnetic configurations. Once again, this is not the most stable phase for such a compound, which reaches the lowest energy configuration when arranged in a trigonal phase, in the same crystal prototype as the ferromagnetic insulator CrI_3 .³⁷

There are several interesting candidates among the materials with the tetragonal structure. In particular, materials with stoichiometry AB_2 in point group $-42m$, such as the case of ZnCl_2 shown in Fig. 1, represent a large majority of stable auxetic materials in C2DB. Notable examples are metal di-halides involving Co, Mn, or Fe as the metallic element. Such materials are all exfoliable from a 3D parent compound with a trigonal point group,³⁷ but their tetragonal phases generally have total energies that are comparable or even lower than the trigonal monolayer phase (which is also present in C2DB). A second notable example is given by group 12 di-halides involving Zn, Cd, and Hg, for which the tetragonal auxetic structure turns out to be the most stable phase. Interestingly, both HgI_2 and ZnCl_2 are reported as easily exfoliable materials by Mounet *et al.*,³⁷ making these two materials very appealing candidates for novel auxetic 2D materials. We also note that MnTe, AgBr, and GeO_2 all seem to have total energies very close to the convex hull and thus also belong to the set of predicted stable auxetic monolayers.

Let us note that a significant majority of known auxetic 2D structures display negative Poisson ratio in the out-of-plane direction,^{85–89} while only very few materials were previously predicted to exhibit in-plane auxetic response.^{90,91} Reference 90 reports a negative Poisson ratio for monolayers of groups 6–7 transition-metal dichalcogenides (MX_2 with $M = \text{Mo}, \text{W}, \text{Tc}, \text{Re}$, and $X = \text{S}, \text{Se}, \text{Te}$) in the 1T-phase. We do find a negative ν_{xy} in C2DB for all of them, but they have low dynamical and thermodynamical stability and, therefore, are not identified by our analysis.

C. Effective masses

Monolayer 2D materials with a finite bandgap can display large anisotropies in the effective masses along two orthogonal directions. This makes them appealing for highly directional-dependent transport, with applications in anisotropic field-effect transistors, polarization-sensitive detectors, and non-volatile memory devices among others.^{12,26,29,30,32,33}

In C2DB, effective masses for conduction and valence bands are calculated for all materials having a finite bandgap greater than 0.01 eV at the PBE level. We define the effective mass, m , from the curvature, a , at the band maximum (minimum) for valence

(conduction) bands as $m = 1/2a$. To determine the curvature we start from a self-consistent ground state calculation performed at a k -point density of 12 \AA^{-1} . From these k -points a preliminary band extremum is found and a second, non-self-consistent calculation is performed with higher density of k -points centered around the preliminary extremum. Then, from these values a final extremum is determined and the energies for a number of k -points spaced very closely around the extremum are calculated non-self-consistently. The k -points used for the first refinement step are by default chosen to lie in a sphere around the extremum with a radius of 250 meV (for a mass of 1) and the same number of k -points as the original calculation (but at least 19). The last refinement uses a 1 meV sphere and 9 points. The points calculated in the final refinement step are used to determine the curvature. We first do a fit to a second order polynomial to determine a preliminary extremum. Then, we perform a fit to a third order polynomial and find the new extremum, unless the optimization algorithm diverges (as may happen for third order polynomials) in which case, we revert to the original fit. We have found that the third order polynomial fit does provide an improvement to the description of the band extremum and in some cases is necessary, e.g., in the presence of parabolic bands crossing as in Rashba splitting. From the fit, we find the curvature a at the extremum and the mass is calculated as $m = 1/2a$.

To measure the presence of anisotropic effects in the effective masses, we define the parameters

$$\delta_{me} = \frac{|m_x^{(e)} - m_y^{(e)}|}{m_x^{(e)} + m_y^{(e)}}, \quad (9a)$$

$$\delta_{mh} = \frac{|m_x^{(h)} - m_y^{(h)}|}{m_x^{(h)} + m_y^{(h)}}, \quad (9b)$$

where

- $m_i^{(e)}$ is the effective electron mass calculated along the i direction around the conduction band minimum and
- $m_i^{(h)}$ is the effective hole mass calculated along the i direction around the valence band maximum.

Unfortunately, getting a very accurate value for the effective masses in a fully automated fashion turns out to be a quite challenging task, with some fits being not accurate enough, or picking a wrong sign for the electron or hole mass in the case of a particularly heavy effective mass. We, therefore, remove all materials having $m_i^{(e/h)} \geq 20m_e$, with m_e the free electron mass, and materials with extremely high ratio $m_i^{(e/h)}/m_j^{(e/h)} \geq 20$. We stress that these threshold values are arbitrary. They have been primarily chosen so that we discard all wrong results, while also keeping the highly anisotropic materials with accurate results into the analysis as much as possible.

The C2DB database contains 574 dynamically and thermodynamically stable materials with a PBE bandgap greater than 0.01 eV, of which 106 fall outside the range of validity described above. This leaves us with a total of 468 materials, whose effective electron and hole masses are shown as a scatterplot in Figs. 5(a)

and 5(b). We find a rather large set of materials with anisotropic effective masses, as one can immediately notice from the large number of points falling outside the main diagonal. Indeed, as shown in Fig. 5(c), 60% of the selected materials (282 out of 468) show a difference of at least 10% between electron effective masses, while 54% of them (253 out of 468) have at least 10% of difference between hole effective masses. Moreover, a quite large fraction of

materials have extremely high values of δ_{me} or δ_{mh} as compared with the case of elastic anisotropy in Fig. 3(c).

When considering the distribution of point groups for materials with effective masses anisotropy, a quite different behavior with respect to previous cases emerges. First, as shown in the left column of Fig. 6, let us notice that the restriction to semiconductors with a bandgap of at least 0.01 eV removes many structures in the

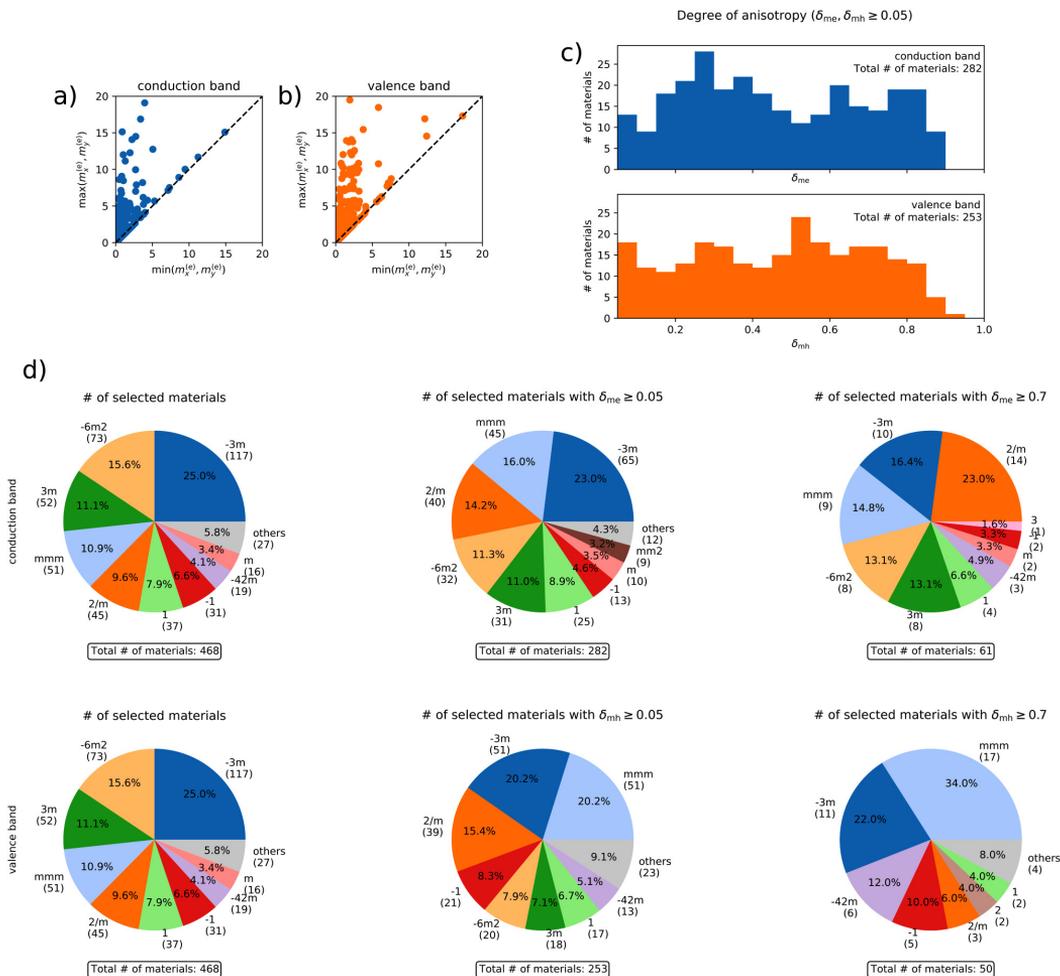


FIG. 5. Distribution of the effective masses in C2DB. In panels (a) and (b), we show the effective mass along the heaviest direction against the effective mass along the lightest direction for conduction and valence bands respectively. Effective masses are plotted in units of the free electron mass m_e . Panel (c) shows a distribution of the anisotropic degrees δ_{me} (δ_{mh}) for the such cases where δ_{me} (δ_{mh}) is greater than 0.05. Panel (d) shows, from left to right, the distribution of δ_{me} (δ_{mh}) in terms of point group symmetry for three different threshold values of 0, 0.05, and 0.7 respectively.

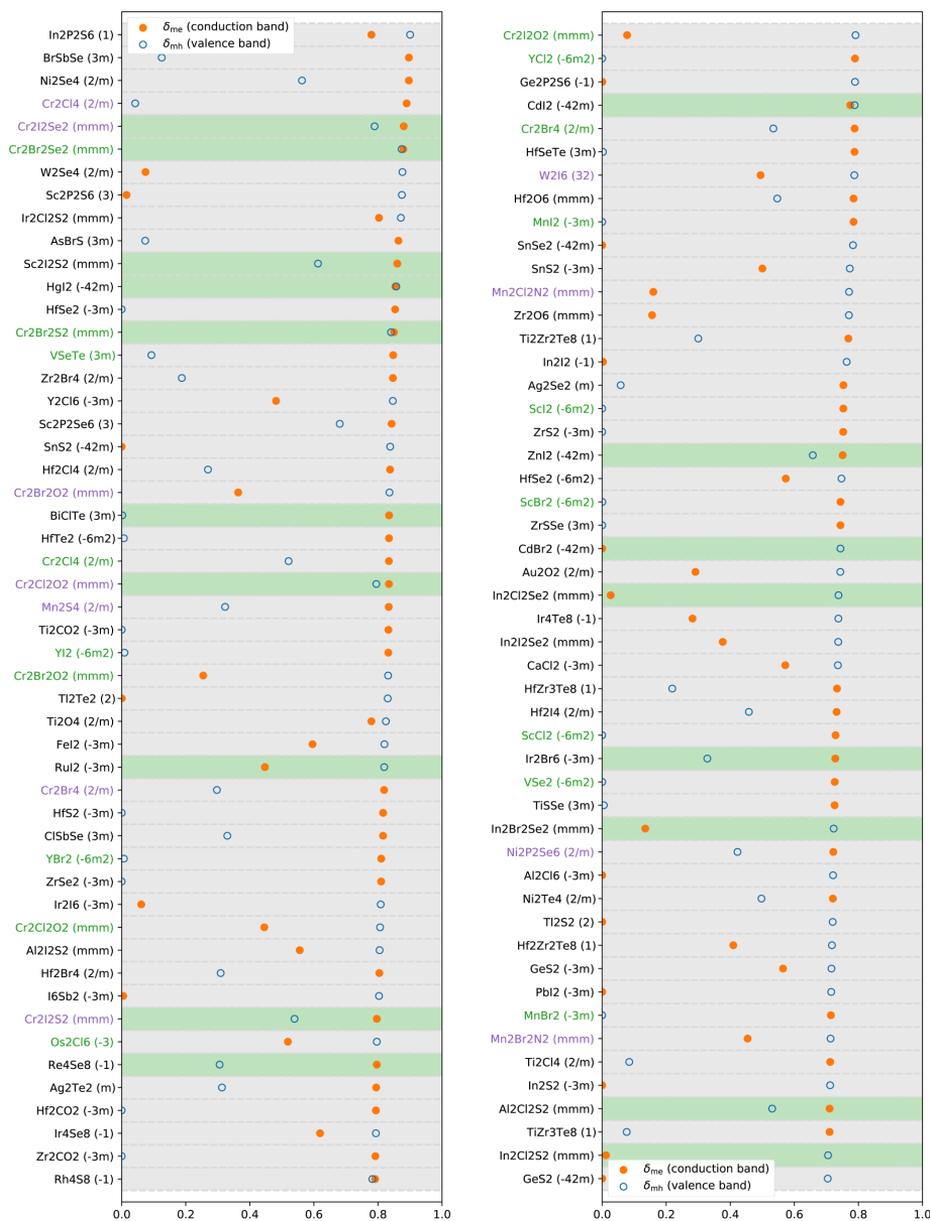


FIG. 6. Materials with highly anisotropic effective masses $\delta_{me} \geq 0.7$ or $\delta_{mh} \geq 0.7$, sorted from the highest (top left) to lowest (bottom right) value of $\max(\delta_{me}, \delta_{mh})$. Green (purple) labels indicates ferromagnetic (antiferromagnetic) materials. A green (light gray) background indicates the presence of a direct (indirect) bandgap at the PBE level.

orthorhombic (mmm) and monoclinic ($2/m$) groups, while favoring structures with trigonal ($-3m, 3m$) and hexagonal ($-62m$) symmetry with respect to the general case shown in Fig. 1. More importantly, we notice that these structures are not filtered out even when we select materials with increasingly high effective masses anisotropy ($\delta_{me(mh)} \geq 0.05$ in the middle, $\delta_{me(mh)} \geq 0.7$ on the right). This means that, despite their structural symmetry, materials such as TMDCs in the T and H phase and Janus structures display a quite strong anisotropy in the effective masses. One should bear in mind that we only calculate the curvatures of valence and conduction bands in one particular valley. While these are not bound to symmetries of the crystal, the overall transport properties (such as, for instance, the mobility) are determined by adding up contributions from all valleys, which in the end cancels out any anisotropic effect and restores the Neumann principle. However, it is worth noting that transport properties of a single anisotropic valley should in principle be accessible in experiments with valley-selection techniques such as circularly polarized optical excitation and gating.^{92–94}

For the case of electron effective masses, we find that 61 stable materials have a rather high anisotropy degree $\delta_{me} \geq 0.7$. While this group is dominated by monoclinic structures in the $2/m$ point group (mostly TMCD in the distorted T' phase), we find a rather large set of triclinic structures which sum up to roughly one third of the total, and a significant 13% of share for hexagonal structures. The situation is different for the case of hole effective masses, where orthorhombic structures represent a 34% of the 50 materials with $\delta_{mh} \geq 0.7$. However, we still find that 22% of structures are in a trigonal point group as well.

We have reported all the 101 structures with δ_{me} or δ_{mh} greater than 0.7 in Fig. 6, with 10 of them having both parameters above the threshold value. Among this group, we notice a recurrent presence of hafnium, bismuth, and antimony-based Janus structures, and chromium based compounds (especially, chromium halides, which also bear elastic and magnetic anisotropy). Most importantly, we find few materials that have already been exfoliated down to monolayer thickness in experiments, including four hafnium and zirconium-based TMDC in the stable T phase and the already mentioned FM compound CrBrS. We also point out the presence of monolayers SbI₃, BiClTe, SbTe, and other ternary compounds in the crystal prototype $ABC-59-ab$ (point group mmm) such as CrBrO and CrClO, which all seem to be easily exfoliable from the layered 3D parent bulk structure.³⁷ Finally, we note that the direct or indirect character of the bandgap does not seem to be relevant for attaining high anisotropy in the effective masses. Indeed, we find 18 out of 101 highly anisotropic materials with direct gap (that is, 18%), which is not dramatically different from the total fraction of semiconductors and insulators with direct gap among the 468 materials considered in this analysis (29%).

A list of experimentally available or easily exfoliable materials with highly anisotropic effective masses is presented in Table II, where we also report three experimentally available materials and one easily exfoliable material having $0.5 \leq \delta_{me/mh} \leq 0.7$ (namely, TiS₃, SnSe₂, GaTe, and AuSe). We also include additional structures that can be obtained from this subset by replacing a constituent element with atoms from the same group. This a relevant case

for Janus monolayer, which can be obtained from already available structures by stripping off an outer layer of chalcogen atoms and substituting them with an element from the same family.⁵⁴ A similar method is likely to be applicable to halogen and chalcogen atoms in ternary orthorhombic compounds. For each entry of Table II, we have double-checked the accuracy of the parabolic fit.

D. Polarizability

The polarizability of a material relates the induced electric dipole moment density to an applied electric field to linear order.¹⁰⁴ For 2D materials, this relation takes the form

$$P_i^{2D}(\vec{q}, \omega) = \sum_j \alpha_{ij}^{2D}(\vec{q}, \omega) E_j(\vec{q}, \omega), \quad (10)$$

where P^{2D} is the induced polarization in the material averaged over the area of the unit cell, $E(\vec{q}, \omega)$ is the applied electric field, and α^{2D} is the polarizability.³⁶

In general, the polarizability can be split into a contribution from the electrons, $\alpha_{ij}^e(\vec{q}, \omega)$, and a contribution from the lattice, $\alpha_{ij}^{lat}(\vec{q}, \omega)$. Since the characteristic response time of the electrons is much faster than that of the lattice, the relevance of the two contributions depends on the time scale of the considered problem. For optical processes involving electromagnetic waves with frequency well above the characteristic phonon frequency of the lattice, only the electronic polarizability is relevant and we can write $\alpha_{ij}(\vec{q}, \omega) \approx \alpha_{ij}^e(\vec{q}, \omega)$. On the other hand, for processes involving infrared light, the lattice response must be considered as well and can in some case even dominate the electronic response.

The polarizability determines the degree of dielectric screening in a material and as such it sets the strength of Coulomb interaction between charged particles.^{105,106} It thereby governs several of the unique properties that made 2D materials famous over the last decade^{4,6,107} including excitons, plasmons, and bandgap renormalization effects.¹⁰⁸ In this context, the in-plane anisotropy of 2D materials has attracted significant interest since the synthesis of few-layer black phosphorus (P₄) in 2014.^{7,9,109} For example, the anisotropic optical absorption (essentially the imaginary part of the electronic polarizability) makes the material act as a linear polarizer,¹¹⁰ which finds applications in diverse fields such as liquid-crystal displays, medical applications, or optical quantum computers.^{111,112} In addition, other fundamental properties, such as the electron–phonon coupling and electron–hole interactions, are influenced by an anisotropic polarizability resulting in formation of quasiparticles, e.g., polarons, excitons, trions, with unconventional shapes and dispersion relations.^{110,113–118}

In the C2DB, the electronic polarizability is calculated within the random phase approximation (RPA)^{119,120} using PBE wave functions and eigenvalues, see Ref. 36 for further details. To keep the discussion general, we focus here on the polarizability in the static ($\omega = 0$) and long wavelength ($q = 0$) limits. As a measure of the degree of anisotropy we adopt the δ parameter defined above and define

TABLE II. List of experimentally available or easily exfoliable materials with high anisotropy in the effective electron or hole masses. The list includes materials that can be obtained from available or exfoliable compounds by chemical substitution of one or more atomic species.

Material	Pointgroup	δ_{me}	δ_{mh}	Ref. (Exp.)	Exfol. from bulk
CrBrS (AFM)	mmm	0.89	0.94	61	Yes
W ₂ Se ₄	2/m	0.07	0.88	95	
CrBrS (FM)	mmm	0.85	0.84	61	Yes
HfSe ₂	-3m	0.85	0.00	96	Yes
Ti ₂ CO ₂	-3m	0.83	0.00	97	
HfS ₂	-3m	0.81	0.00	51	Yes
ZrSe ₂	-3m	0.81	0.00	98	Yes
Re ₄ Se ₈	-1	0.80	0.30	30	Yes
SnS ₂	-3m	0.50	0.78	99	Yes
ZrS ₂	-3m	0.75	0.00	100	Yes
PbI ₂	-3m	0.00	0.72	101	Yes
Ti ₂ S ₆	2/m	0.60	0.52	23	Yes
SnSe ₂	-3m	0.55	0.00	102	Yes
Ga ₂ Te ₂	-6m2	0.51	0.37	103	Yes
CrBrO (AFM)	mmm	0.37	0.84		Yes ^{36,37}
CrClO (AFM)	mmm	0.83	0.79		Yes ^{36,37}
CrBrO (FM)	mmm	0.25	0.83		Yes ^{36,37}
CrClO (FM)	mmm	0.45	0.81		Yes ^{36,37}
I ₆ Sb ₂	-3m	0.01	0.80		Yes ³⁶
Au ₂ Se ₂	2/m	0.65	0.26		Yes ^{36,37}
Material	Pointgroup	δ_{me}	δ_{mh}	Obtainable via substitution from	
CrBrSe (AFM)	mmm	0.94	0.92	CrBrS ³⁷	
CrIS (FM)	mmm	0.93	0.73	CrBrS ³⁷	
BrSbSe	3m	0.90	0.13	ISbTe ³⁷	
CrISe (AFM)	mmm	0.89	0.80	CrBrS ³⁷	
CrBrSe (FM)	mmm	0.88	0.88	CrBrS ³⁷	
CrIS (AFM)	mmm	0.79	0.53	CrBrS ³⁷	
CrIO (FM)	mmm	0.08	0.79	CrBrO, ³⁷ CrClO ³⁷	
HfSeTe	3m	0.79	0.01	HfS ₂ , ⁵¹ HfSe ₂ ⁹⁶	
ZrSSe	3m	0.75	0.00	ZrS ₂ , ¹⁰⁰ ZrSe ₂ ⁹⁸	
CrIO (AFM)	mmm	0.15	0.58	CrBrO, ³⁷ CrClO ³⁷	

$$\delta_{ae} = \frac{|\alpha_x^p - \alpha_y^p|}{|\alpha_x^p| + |\alpha_y^p|}, \quad (11)$$

with $p = \{e, lat\}$ for the electronic and lattice polarizability, respectively.

In Fig. 7 panels (c) and (d), we show the statistical distribution of the materials with electronic polarizability anisotropy above 0.005 and 0.4, respectively. For reference, the distribution of all the materials for which the polarizability has been calculated is shown in panel (a), and the materials are classified according to point group symmetry as usual. As the threshold is increased we see the same trend as for the magnetic, elastic, and effective masses anisotropies, namely, the orthorhombic *mmm* point group followed by the monoclinic *2/m* becomes increasingly dominant. Both the trigonal and tetragonal phases disappear from the distribution already for $\delta_{ae} > 0.005$. The orthorhombic *mmm* group is particularly ubiquitous

among the materials with high δ_{ae} , surpassing 70% of the remaining materials already at a moderate threshold of $\delta_{ae} > 0.4$. In Fig. 7(e), we show the materials with the largest anisotropies found in the range $\delta_{ae} > 0.7$. We note that our analysis correctly identifies the known in-plane anisotropic compounds such as P₄⁷ (phosphorene), As₄⁵⁹ (arsenene), MoS₂ (in the T' phase) and WTe₂⁵⁸ among others.

The materials with the highest δ_{ae} are ternary compounds and, therefore, more challenging to realize experimentally than the more common binary 2D materials, a notable exception being the recently synthesized Cs₂Br₂S₂,⁶¹ mentioned above, which has, in addition, a high electronic polarizability anisotropy. Regarding binary compounds, several of the materials with large δ_{ae} have been predicted to be exfoliable from known parent bulk materials.³⁷ Among our anisotropic materials that match the stoichiometry of the materials listed in Ref. 37 as easily exfoliable, the most promising of these are listed in Table III.

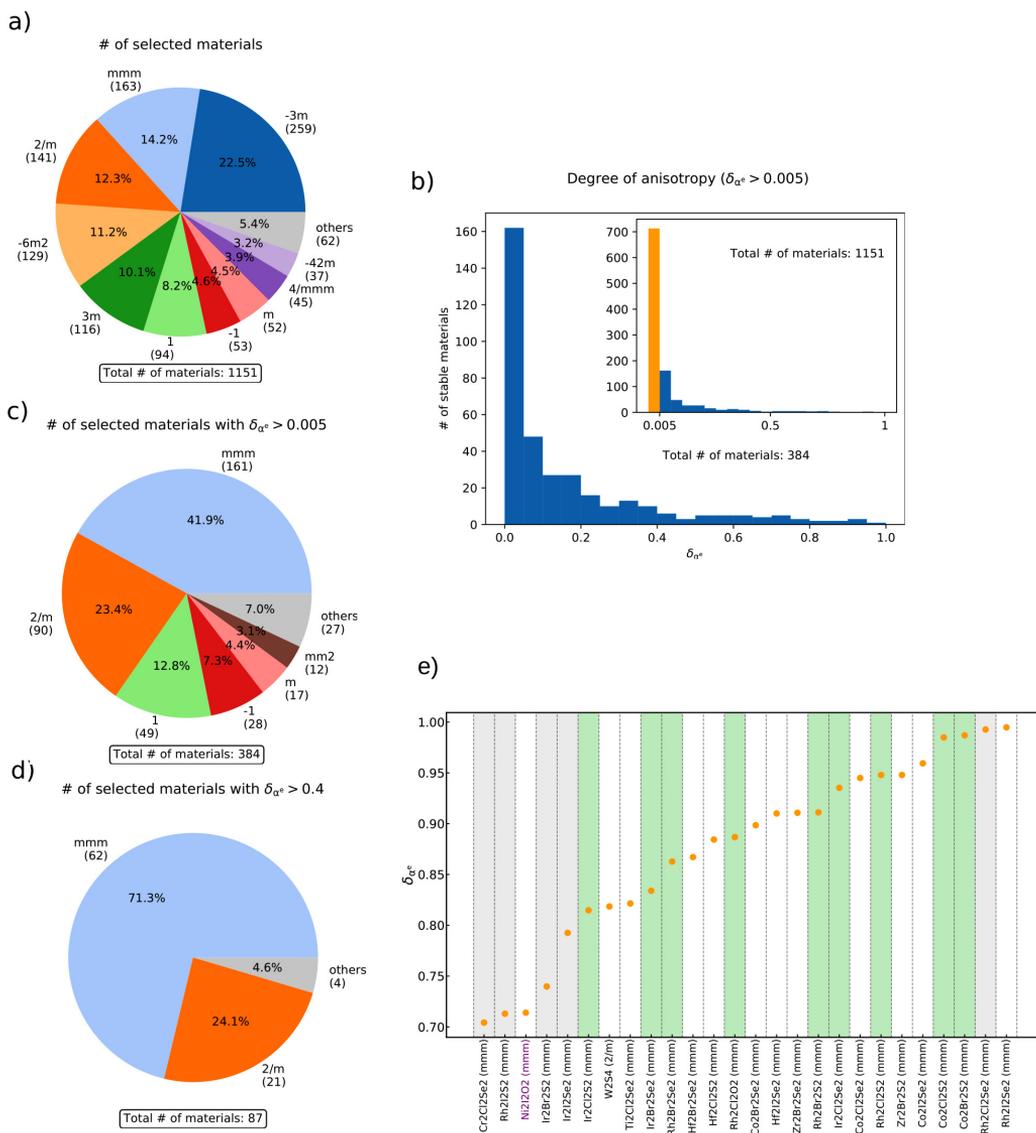


FIG. 7. (a) Distribution of materials predicted to be stable in the C2DB database, according to their point group and their in-plane optical polarizability anisotropy. The percentage of materials represented by each point group and their number of occurrences in the database is explicitly shown, except for those point groups representing less than 3% of materials. In (b) and (c), classification of materials in the database with a low ($\delta_{\alpha^e} > 0.005$) in-plane optical polarizability anisotropy threshold; in a histogram according to their degree of anisotropy in (b), and in a pie chart according to their point group in (c). The main chart in (b) shows the distribution of the materials above the threshold, while in the inset an orange bin represents all materials below the threshold. In (d) and (e), the materials with a high anisotropy are classified. In (d), a threshold of $\delta_{\alpha^e} > 0.4$ is set. In (e), the materials over a threshold of $\delta_{\alpha^e} > 0.7$ are sorted by their anisotropy, and their magnetic state is indicated by the font color (purple for AFM). Green, gray, and white background colors indicate the presence of a direct and indirect bandgap and zero bandgap according to HSE calculations.

TABLE III. Monolayers with the highest in-plane electronic polarizability anisotropy (δ_{ac}) in the C2DB database whose stoichiometry matches that of the entries predicted to be easily exfoliable from a known layered bulk material in Ref. 37. Their space group symmetry, energy above convex hull, magnetic state, and in-plane electronic polarizability anisotropies are listed.

	Sym.	E_{hull} (meV)	Lowest E_{hull} monolayer?	δ_{ac}
$\text{Cs}_2\text{Br}_2\text{S}_2$ ⁶¹	Pm3m	0.0	Yes	0.41
W_2Se_4 ⁹⁵	$P2_1/m$	91.7	No	0.47
Zr_2I_4	$P2_1/m$	0.0	Yes	0.51
Mo_2Se_4	$P2_1/m$	109.4	No	0.54
Zr_2Cl_4	$P2_1/m$	31.9	No	0.60
Ti_2Cl_4	$P2_1/m$	0.0	Yes	0.65
W_2S_4	$P2_1/m$	177.9	No	0.82

We highlight the experimentally known ternary compound $\text{Cs}_2\text{Br}_2\text{S}_2$ among the easily exfoliable materials from Table III, presenting a high optical polarizability anisotropy of $\delta_{ac} > 0.41$. Among the compounds not known yet experimentally, we highlight two in the table, that have $\delta_{ac} > 0.65$ and $\delta_{ac} > 0.5$, respectively: Ti_2Cl_4 and Zr_2I_4 . We stress that $\delta_{ac} = 0.65$, for instance, implies that the polarizability in one direction of the plane is 4.7 times larger than in the other direction. Consequently, these materials are very promising candidates for anisotropic optical applications such as light polarizers.

The lattice or infrared polarizability α^{lat} is calculated in the C2DB database only for materials meeting the requirements of high stability and bandgap > 10 meV, which represent about 15% of the materials in the database. Hence we will limit our

TABLE IV. Stable materials with the highest in-plane infrared polarizability anisotropy in the C2DB database whose stoichiometry matches that of entries in the easily exfoliable 2D materials list in Ref. 37. Their space group symmetry, energy over the hull minimum, magnetic state and in-plane infrared polarizability anisotropies are given.

	Sym.	E_{hull} (meV)	Lowest E_{hull} monolayer?	$\delta_{\alpha^{lat}}$
Sn_2S_2	Pmn2 ₁	42.6	No	0.26
Sn_2Se_2	Pmn2 ₁	42.9	No	0.38
Sn_2Te_2	Pmn2 ₁	62.9	Yes	0.45
ZrI_2	P6m2	27.0	No	0.46
Ge_2Se_2	Pmn2 ₁	24.9	No	0.50

analysis to extracting the most promising individual materials, since a statistical analysis would not be representative of the real distribution of the materials in the database. In Fig. 8, we show all 16 materials with $\delta_{\alpha^{lat}} > 0.2$. As it is the case with the rest of properties, we find several materials with a significant anisotropy. For instance, the monolayer Sn_2Te_2 has a $\delta_{\alpha^{lat}}$ value 0.45, is at the bottom of the convex hull combining Sn and Te among monolayers and is considered to be easily exfoliable.³⁷ This makes it a very interesting material for further experimental and theoretical exploration. We find other materials whose stoichiometry matches easily exfoliable entries in Ref. 37 among those with $\delta_{\alpha^{lat}} > 0.25$ and are listed in Table IV. Taking into account that these promising materials are selected among only a little fraction of the entire C2DB database, we anticipate that there are a large amount of promising infrared anisotropic materials in the database yet to be discovered.

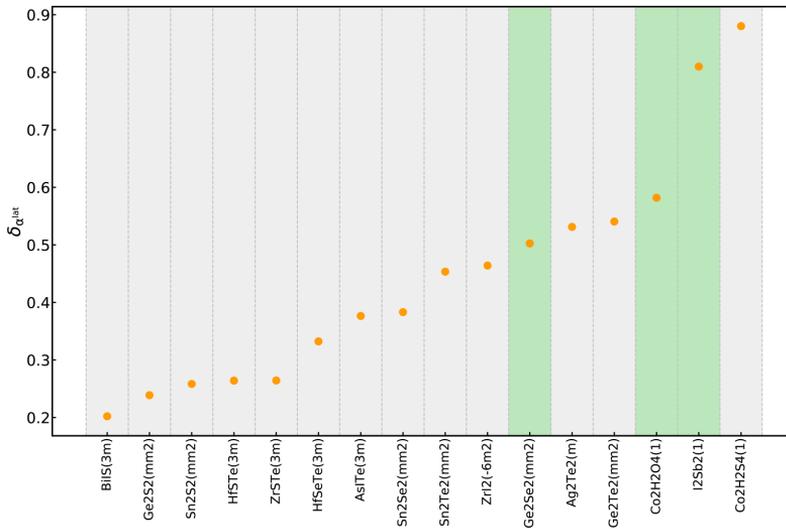


FIG. 8. Materials in the C2DB database with high infrared polarizability anisotropy ($\delta_{\alpha^{lat}} > 0.2$) predicted to be stable. Green, gray, and white background colors indicate the presence of a direct and indirect bandgap and zero bandgap according to HSE calculations.

IV. CONCLUSIONS

In this work, we have analyzed the presence of anisotropic behavior among more than 1000 2D materials predicted to be stable in the C2DB database.³⁶ Specifically, we have identified materials with in-plane magnetic anisotropy, anisotropic Young's modulus and/or negative Poisson ratio, anisotropic effective masses, and anisotropic polarizabilities.

Consistent with the Neumann principle, we have found that there are two main features in the C2DB database that favor anisotropy, namely, (i) a lower symmetry and (ii) a larger number of constituent elements. In addition, our analysis satisfactorily captures the specific symmetry requirements of each anisotropy type: elastic and polarizability anisotropies, derived from second order tensors, are forbidden for trigonal, tetragonal, or hexagonal compounds; the magnetic anisotropic materials do not include hexagonal and tetragonal groups, and the effective mass anisotropy is allowed for all symmetry groups in the database. Several of the materials identified in this study outperform the known 2D materials in terms of anisotropic figures of merit and are predicted to be stable and/or exfoliable from known parent bulk crystals,³⁷ providing useful guidelines for future experimental investigations.

The most prominent material class resulting from our analysis is the ternary orthorhombic compound prototype *ABC-59-ab*.³⁶ This material class combines three different atomic species in a low symmetry structure, often resulting in strongly anisotropic properties. To the best of our knowledge, one of such materials (namely, CrSBr) has been isolated in monolayer form very recently,⁶¹ and further experimental efforts in this direction could hopefully be motivated by our work.

We find several binary monolayers with interesting anisotropic behaviors that are predicted to be stable and in some cases even predicted as easily exfoliable. A transition metal (in particular, Ni, V, Cr, and Os) combined with a halide in a low symmetry structure appears to be the best recipe for obtaining in-plane magnetic and elastic anisotropy. For instance, VI_2 is an exfoliable and very stable compound with an in-plane magnetic anisotropy that competes with the highest out-of-plane anisotropies known to date. Moreover, there are multiple compounds with large predicted anisotropies, that match the stoichiometry of exfoliable materials and with similar total energies. Such materials could be stabilized under the right experimental conditions. Among them, we highlight anti-ferromagnetic Ni_2I_4 , which has an exceptional in-plane anisotropy exceeding 20 meV/unit cell, that makes it a candidate for realization of high temperature in-plane 2D antiferromagnetism. Likewise, chromium di-halides stand out for their markedly anisotropic behavior in the elastic response and effective electron and hole masses. Among the non-magnetic materials, we identify AuX ($X = \text{S}, \text{Se}, \text{Te}$) as a new class of potentially stable 2D materials with high anisotropy in several physical properties, with AuSe being reported as easily exfoliable in the literature.

On the other hand, transport properties of a single valley deserve to be mentioned separately, as they do not seem to be bound to the symmetries of the crystal lattice and could be experimentally accessed by means of circularly polarized light. Several TMDCs and Janus structures have highly symmetric crystal structure with trigonal symmetry but strong effective mass anisotropy,

and we identify HfX_2 , ZrX_2 , and SnX_2 ($X = \text{S}, \text{Se}$) as the most interesting monolayers for anisotropic transport applications already available in experiments, together with the above mentioned CrSBr. Moreover, new undiscovered structures with very low inter-layer binding energy such as SbI₃, AuSe, and ternary magnetic compounds CrOBr and CrOCl display strong effective masses anisotropies.

Regarding the electronic polarizability, we also find a large amount of anisotropic materials, nearly all being ternary compounds of orthorhombic symmetry. In addition, some binary compounds, mostly involving a transition metal and a halide or a chalcogen, that are predicted to be easily exfoliable are identified and listed in the text. Finally, we also identified materials with interesting infrared polarizability anisotropy values among a smaller set of candidates in the C2DB. The most promising prospects for experimental realization are listed in the text.

Among the materials with negative Poisson ratios (so-called auxetic materials) identified in our study, we highlight HgI_2 and ZnCl_2 , which are both predicted as easily exfoliable,⁵⁷ and MnTe, AgBr, and GeO_2 , which are predicted to be stable in their monolayer form.

AUTHORS' CONTRIBUTIONS

L.V. and U.P. contributed equally.

ACKNOWLEDGMENTS

We thank Alireza Taghizadeh for useful discussions and suggestions. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 754462 (EuroTechPostdoc). K.S.T. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 773122, LIMA). The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation (Project No. DNR103).

DATA AVAILABILITY

The data that support the findings of this study are openly available in C2DB website, Ref. 121.

REFERENCES

- ¹M. N. Gjerding, R. Petersen, T. G. Pedersen, N. A. Mortensen, and K. S. Thygesen, "Layered van der Waals crystals with hyperbolic light dispersion," *Nat. Commun.* **8**, 948 (2017).
- ²Z. Jacob, L. V. Alekseyev, and E. Narimanov, "Optical hyperlens: Far-field imaging beyond the diffraction limit," *Opt. Express* **14**, 8247–8256 (2006).
- ³D. Lu, J. J. Kan, E. E. Fullerton, and Z. Liu, "Enhancing spontaneous emission rates of molecules using nanopatterned multilayer hyperbolic metamaterials," *Nat. Nanotechnol.* **9**, 48–53 (2014).
- ⁴K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, "Electric field effect in atomically thin carbon films," *Science* **306**, 666–669 (2004).
- ⁵L. Ci, L. Song, C. Jin, D. Jariwala, D. Wu, Y. Li, A. Srivastava, Z. Wang, K. Storr, L. Balicas *et al.*, "Atomic layers of hybridized boron nitride and graphene domains," *Nat. Mater.* **9**, 430–435 (2010).

- ⁶K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, "Atomically thin MoS_2 : A new direct-gap semiconductor," *Phys. Rev. Lett.* **105**, 136805 (2010).
- ⁷L. Li, Y. Yu, G. J. Ye, Q. Ge, X. Ou, H. Wu, D. Feng, X. H. Chen, and Y. Zhang, "Black phosphorus field-effect transistors," *Nat. Nanotechnol.* **9**, 372 (2014).
- ⁸R. Fei and L. Yang, "Strain-engineering the anisotropic electrical conductance of few-layer black phosphorus," *Nano Lett.* **14**, 2884–2889 (2014).
- ⁹F. Xia, H. Wang, and Y. Jia, "Rediscovering black phosphorus as an anisotropic layered material for optoelectronics and electronics," *Nat. Commun.* **5**, 4458 (2014).
- ¹⁰Q. Wei and X. Peng, "Superior mechanical flexibility of phosphorene and few-layer black phosphorus," *Appl. Phys. Lett.* **104**, 251915 (2014).
- ¹¹T. Low, R. Roldán, H. Wang, F. Xia, P. Avouris, L. M. Moreno, and F. Guinea, "Plasmons and screening in monolayer and multilayer black phosphorus," *Phys. Rev. Lett.* **113**, 106802 (2014).
- ¹²L. Wang, A. Kutana, X. Zou, and B. I. Yakobson, "Electro-mechanical anisotropy of phosphorene," *Nanoscale* **7**, 9746–9751 (2015).
- ¹³A. Jain and A. McGaughey, "Strongly anisotropic in-plane thermal transport in single-layer black phosphorene," *Sci. Rep.* **5**, 8501 (2015).
- ¹⁴X. Wang, A. M. Jones, K. L. Seyler, V. Tran, Y. Jia, H. Zhao, H. Wang, L. Yang, X. Xu, and F. Xia, "Highly anisotropic and robust excitons in monolayer black phosphorus," *Nat. Nanotechnol.* **10**, 517–521 (2015).
- ¹⁵H. Tian, Q. Guo, Y. Xie, H. Zhao, C. Li, J. J. Cha, F. Xia, and H. Wang, "Anisotropic black phosphorus synaptic device for neuromorphic applications," *Adv. Mater.* **28**, 4991–4997 (2016).
- ¹⁶H. Yang, H. Jussila, A. Autere, H.-P. Komsa, G. Ye, X. Chen, T. Hasan, and Z. Sun, "Optical waveplates based on birefringence of anisotropic two-dimensional layered materials," *ACS Photonics* **4**, 3023–3030 (2017).
- ¹⁷C. Wang, G. Zhang, S. Huang, Y. Xie, and H. Yan, "The optical properties and plasmonics of anisotropic 2D materials," *Adv. Opt. Mater.* **8**, 1900996 (2020).
- ¹⁸J. Ma, Y. Chen, Z. Han, and W. Li, "Strong anisotropic thermal conductivity of monolayer WTe_2 ," *2D Mater.* **3**, 045010 (2016).
- ¹⁹E. Torun, H. Sahin, S. Cahangirov, A. Rubio, and F. M. Peeters, "Anisotropic electrical, mechanical, and optical properties of monolayer WTe_2 ," *J. Appl. Phys.* **119**, 074307 (2016).
- ²⁰Y.-J. Zhang, R.-N. Wang, G.-Y. Dong, S.-F. Wang, G.-S. Fu, and J.-L. Wang, "Mechanical properties of 1T-, 1T', and 1H-MX₂ monolayers and their 1H/1T'-MX₂ ($M = \text{Mo}, \text{W}$ and $X = \text{S}, \text{Se}, \text{Te}$) heterostructures," *AIP Adv.* **9**, 125208 (2019).
- ²¹Q. Zhang, R. Zhang, J. Chen, W. Shen, C. An, X. Hu, M. Dong, J. Liu, and L. Zhu, "Remarkable electronic and optical anisotropy of layered 1T'- WTe_2 2D materials," *Beilstein J. Nanotechnol.* **10**, 1745–1753 (2019).
- ²²C. Wang, S. Huang, Q. Xing, Y. Xie, C. Song, F. Wang, and H. Yan, "Van der Waals thin films of WTe_2 for natural hyperbolic plasmonic surfaces," *Nat. Commun.* **11**, 192 (2020).
- ²³J. O. Island, M. Buscema, M. Barawi, J. M. Clamagirand, J. R. Ares, C. Sánchez, I. J. Ferrer, G. A. Steele, H. S. J. van der Zant, and A. Castellanos-Gomez, "Ultrahigh photoreponse of few-layer TiS_3 nanoribbon transistors," *Adv. Opt. Mater.* **2**, 641–645 (2014).
- ²⁴J. O. Island, M. Barawi, R. Biele, A. Almazán, J. M. Clamagirand, J. R. Ares, C. Sánchez, H. S. J. van der Zant, J. V. Álvarez, R. D'Agosta, I. J. Ferrer, and A. Castellanos-Gomez, " TiS_3 transistors with tailored morphology and electrical properties," *Adv. Mater.* **27**, 2595–2601 (2015).
- ²⁵J. Kang, H. Sahin, and F. M. Peeters, "Mechanical properties of monolayer sulphides: A comparative study between MoS_2 , HfS_2 and TiS_3 ," *Phys. Chem. Chem. Phys.* **17**, 27742–27749 (2015).
- ²⁶Y. Jin, X. Li, and J. Yang, "Single layer of MX_3 ($M = \text{Ti}, \text{Zr}$; $X = \text{S}, \text{Se}, \text{Te}$): A new platform for nano-electronics and optics," *Phys. Chem. Chem. Phys.* **17**, 18665–18669 (2015).
- ²⁷J. A. Silva-Guillén, E. Canadell, P. Ordejón, F. Guinea, and R. Roldán, "Anisotropic features in the electronic structure of the two-dimensional transition metal trichalcogenide TiS_3 : Electron doping and plasmons," *2D Mater.* **4**, 025085 (2017).
- ²⁸A. Khatibi, R. H. Godiksen, S. B. Basuvalingam, D. Pellegrino, A. A. Bol, B. Shokri, and A. G. Curto, "Anisotropic infrared light emission from quasi-1D layered TiS_3 ," *2D Mater.* **7**, 015022 (2019).
- ²⁹E. Liu, Y. Fu, Y. Wang, Y. Feng, H. Liu, X. Wan, W. Zhou, B. Wang, L. Shao, C.-H. Ho, Y.-S. Huang, Z. Cao, L. Wang, A. Li, J. Zeng, F. Song, X. Wang, Y. Shi, H. Yuan, H. Y. Hwang, Y. Cui, F. Miao, and D. Xing, "Integrated digital inverters based on two-dimensional anisotropic ReS_2 field-effect transistors," *Nat. Commun.* **6**, 6991 (2015).
- ³⁰E. Zhang, P. Wang, Z. Li, H. Wang, C. Song, C. Huang, Z.-G. Chen, L. Yang, K. Zhang, S. Lu, W. Wang, S. Liu, H. Fang, X. Zhou, H. Yan, J. Zou, X. Wan, P. Zhou, W. Hu, and F. Xiu, "Tunable ambipolar polarization-sensitive photodetectors based on high-anisotropy ReSe_2 nanosheets," *ACS Nano* **10**, 8067–8077 (2016).
- ³¹J. P. Echeverry and I. C. Gerber, "Theoretical investigations of the anisotropic optical properties of distorted 1T ReS_2 and ReSe_2 monolayers, bilayers, and in the bulk limit," *Phys. Rev. B* **97**, 075123 (2018).
- ³²X. Liu, Y. Yuan, Z. Wang, R. S. Deacon, W. J. Yoo, J. Sun, and K. Ishibashi, "Directly probing effective-mass anisotropy of two-dimensional ReSe_2 in schottky tunnel transistors," *Phys. Rev. Appl.* **13**, 044056 (2020).
- ³³H. Wang, M.-L. Chen, M. Zhu, Y. Wang, B. Dong, X. Sun, X. Zhang, S. Cao, X. Li, J. Huang, L. Zhang, W. Liu, D. Sun, Y. Ye, K. Song, J. Wang, Y. Han, T. Yang, H. Guo, C. Qin, L. Xiao, J. Zhang, J. Chen, Z. Han, and Z. Zhang, "Gate tunable giant anisotropic resistance in ultra-thin GaTe ," *Nat. Commun.* **10**, 2302 (2019).
- ³⁴A. D. Oyedele, S. Yang, L. Liang, A. A. Puzetky, K. Wang, J. Zhang, P. Yu, P. R. Pudasaini, A. W. Ghosh, Z. Liu, C. M. Rouleau, B. G. Sumpster, M. F. Chisholm, W. Zhou, P. D. Rack, D. B. Geoghegan, and K. Xiao, "PdSe₂: Pentagonal two-dimensional layers with high air stability for electronics," *J. Am. Chem. Soc.* **139**, 14090–14097 (2017).
- ³⁵L.-S. Lu, G.-H. Chen, H.-Y. Cheng, C.-P. Chuu, K.-C. Lu, C.-H. Chen, M.-Y. Lu, T.-H. Chuang, D.-H. Wei, W.-C. Chueh, W.-B. Jian, M.-Y. Li, Y.-M. Chang, L.-J. Li, and W.-H. Chang, "Layer-dependent and in-plane anisotropic properties of low-temperature synthesized few-layer PdSe_2 single crystals," *ACS Nano* **14**, 4963–4972 (2020).
- ³⁶S. Hastrup, M. Strange, M. Pandey, T. Deilmann, P. S. Schmidt, N. F. Hinsche, M. N. Gjerding, D. Torelli, P. M. Larsen, A. C. Riis-Jensen, J. Gath, K. W. Jacobsen, J. J. Mortensen, T. Olsen, and K. S. Thygesen, "The computational 2D materials database: High-throughput modeling and discovery of atomically thin crystals," *2D Mater.* **5**, 042002 (2018).
- ³⁷N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohier, I. E. Castelli, A. Cepellotti, G. Pizzi, and N. Marzari, "Two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds," *Nat. Nanotechnol.* **13**, 246–252 (2018).
- ³⁸M. Ashton, J. Paul, S. B. Sinnott, and R. G. Hennig, "Topology-scaling identification of layered solids and stable exfoliated 2D materials," *Phys. Rev. Lett.* **118**, 106101 (2017).
- ³⁹K. Choudhary, I. Kalish, R. Beams, and F. Tavazza, "High-throughput identification and characterization of two-dimensional materials using density functional theory," *Sci. Rep.* **7**, 197 (2017).
- ⁴⁰G. Cheon, K.-A. N. Duerloo, A. D. Sendek, C. Porter, Y. Chen, and E. J. Reed, "Data mining for new two- and one-dimensional weakly bonded solids and lattice-commensurate heterostructures," *Nano Lett.* **17**, 1915–1923 (2017).
- ⁴¹J. Zhou, L. Shen, M. D. Costa, K. A. Persson, S. P. Ong, P. Huck, Y. Lu, X. Ma, Y. Chen, H. Tang, and Y. P. Feng, "2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches," *Sci. Data* **6**, 201 (2019).
- ⁴²J. 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. Lehtovaa, M. Ljungberg, O. Lopez-Acevedo, P. G. Moses, J. Ojanen, T. Olsen, V. Petzold, N. A. Romero, J. Stausholm-Møller, M. Strange, G. A. Tritsarolis, 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, "Electronic structure calculations

- with GPAW: A real-space implementation of the projector augmented-wave method," *J. Phys. Condens. Matter* **22**, 253202 (2010).
- ⁴³A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schott, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, "The atomic simulation environment—A Python library for working with atoms," *J. Phys. Condens. Matter* **29**, 273002 (2017).
- ⁴⁴See <https://asr.readthedocs.io/en/latest/> for "Atomic Simulation Recipes (ASR)".
- ⁴⁵J. J. Mortensen, M. Gjerding, and K. S. Thygesen, "MyQueue: Task and workflow scheduling system," *J. Open Source Softw.* **5**, 1844 (2020).
- ⁴⁶J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- ⁴⁷J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, "Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD)," *JOM* **65**, 1501–1509 (2013).
- ⁴⁸F. Neumann and O. Meyer, *Vorlesungen über die Theorie der Elasticität der festen Körper und des Lichtäthers*, Vorlesungen über Mathematische Physik (Teubner, 1885).
- ⁴⁹A. Togo and I. Tanaka, "Spglib: A software library for crystal symmetry search," arXiv e-prints, [arXiv:1808.01590\[cond-mat.mtrl-sci\]](https://arxiv.org/abs/1808.01590) (2018).
- ⁵⁰Note that we will substitute the common overline notation with a dash (that is, we will use $-n$ instead of \bar{n}).
- ⁵¹K. Xu, Z. Wang, F. Wang, Y. Huang, F. Wang, L. Yin, C. Jiang, and J. He, "Ultrasensitive phototransistors based on few-layered HfS_2 ," *Adv. Mater.* **27**, 7881–7887 (2015).
- ⁵²F.-F. Zhu, W.-J. Chen, Y. Xu, C.-L. Gao, D.-D. Guan, C.-H. Liu, D. Qian, S.-C. Zhang, and J.-F. Jia, "Epitaxial growth of two-dimensional stanene," *Nat. Mater.* **14**, 1020–1025 (2015).
- ⁵³D. C. Elias, R. R. Nair, T. M. G. Mohiuddin, S. V. Morozov, P. Blake, M. P. Halsall, A. C. Ferrari, D. W. Boukhvalov, M. I. Katsnelson, A. K. Geim, and K. S. Novoselov, "Control of graphene's properties by reversible hydrogenation: Evidence for graphane," *Science* **323**, 610 (2009).
- ⁵⁴A.-Y. Lu, H. Zhu, J. Xiao, C.-P. Chuu, Y. Han, M.-H. Chiu, C.-C. Cheng, C.-W. Yang, K.-H. Wei, Y. Yang *et al.*, "Janus monolayers of transition metal dichalcogenides," *Nat. Nanotech.* **12**, 744–749 (2017).
- ⁵⁵B. Fülöp, Z. Tajkov, J. Pető, P. Kun, J. Koltai, L. Oroszlány, E. Tóvári, H. Murakawa, Y. Tokura, S. Bordács, L. Tapasztó, and S. Consona, "Exfoliation of single layer BiTeI flakes," *2D Mater.* **5**, 031013 (2018).
- ⁵⁶A. C. Riis-Jensen, S. Manti, and K. S. Thygesen, "Engineering atomically sharp potential steps and band alignment at solid interfaces using 2D Janus layers," *J. Phys. Chem. C* **124**, 9572–9580 (2020).
- ⁵⁷B. Huang, G. Clark, E. Navarro-Moratalla, D. R. Klein, R. Cheng, K. L. Seyler, D. Zhong, E. Schmidgall, M. A. McGuire, D. H. Cobden *et al.*, "Layer-dependent ferromagnetism in a van der Waals crystal down to the monolayer limit," *Nature* **546**, 270–273 (2017).
- ⁵⁸S. Tang, C. Zhang, D. Wong, Z. Pedramrazi, H.-Z. Tsai, C. Jia, B. Moritz, M. Claassen, H. Ryu, S. Kahn, J. Jiang, H. Yan, M. Hashimoto, D. Lu, R. G. Moore, C.-C. Hwang, C. Hwang, Z. Hussain, Y. Chen, M. M. Ugeda, Z. Liu, X. Xie, T. P. Devereaux, M. F. Crommie, S.-K. Mo, and Z.-X. Shen, "Quantum spin hall state in monolayer $1T'$ - WTe_2 ," *Nat. Phys.* **13**, 683–687 (2017).
- ⁵⁹C. Kamal and M. Ezawa, "Arsenene: Two-dimensional buckled and puckered honeycomb arsenic systems," *Phys. Rev. B* **91**, 085423 (2015).
- ⁶⁰E. J. Telford, A. H. Dismukes, K. Lee, M. Cheng, A. Wieteska, A. K. Bartholomew, Y.-S. Chen, X. Xu, A. N. Pasupathy, X. Zhu, C. R. Dean, and X. Roy, "Layered antiferromagnetism induces large negative magnetoresistance in the van der Waals semiconductor CrSBr ," arXiv e-prints, [arXiv:2005.06110](https://arxiv.org/abs/2005.06110) (2020).
- ⁶¹K. Lee, A. H. Dismukes, E. J. Telford, R. A. Wiscoson, X. Xu, C. Nuckolls, C. R. Dean, X. Roy, and X. Zhu, "Magnetic order and symmetry in the 2D semiconductor CrSBr ," arXiv e-prints, [arXiv:2007.10715](https://arxiv.org/abs/2007.10715) (2020).
- ⁶²H.-P. Komsa and A. V. Krasheninnikov, "Two-dimensional transition metal dichalcogenide alloys: Stability and electronic properties," *J. Phys. Chem. Lett.* **3**, 3652–3656 (2012).
- ⁶³L. M. Xie, "Two-dimensional transition metal dichalcogenide alloys: Preparation, characterization and applications," *Nanoscale* **7**, 18392–18401 (2015).
- ⁶⁴Z. Song, C.-C. Liu, J. Yang, J. Han, M. Ye, B. Fu, Y. Yang, Q. Niu, J. Lu, and Y. Yao, "Quantum spin hall insulators and quantum valley hall insulators of BiX/SbX ($X = \text{H, F, Cl}$ and Br) monolayers with a record bulk band gap," *NPG Asia Mater.* **6**, e147 (2014).
- ⁶⁵L. Vannucci, T. Olsen, and K. S. Thygesen, "Conductance of quantum spin hall edge states from first principles: The critical role of magnetic impurities and inter-edge scattering," *Phys. Rev. B* **101**, 155404 (2020).
- ⁶⁶N. D. Mermin and H. Wagner, "Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models," *Phys. Rev. Lett.* **17**, 1133 (1966).
- ⁶⁷Z. Fei, B. Huang, P. Malinowski, W. Wang, T. Song, J. Sanchez, W. Yao, D. Xiao, X. Zhu, A. F. May, W. Wu, D. H. Cobden, J.-H. Chu, and X. Xu, "Two-dimensional itinerant ferromagnetism in atomically thin Fe_3GeTe_2 ," *Nat. Mater.* **17**, 778–782 (2018).
- ⁶⁸W.-B. Zhang, Q. Qu, P. Zhu, and C.-H. Lam, "Robust intrinsic ferromagnetism and half semiconductivity in stable two-dimensional single-layer chromium trihalides," *J. Mater. Chem. C* **3**, 12457–12468 (2015).
- ⁶⁹J. L. Lado and J. Fernández-Rossier, "On the origin of magnetic anisotropy in two dimensional CrI_3 ," *2D Mater.* **4**, 035002 (2017).
- ⁷⁰D. Torelli and T. Olsen, "Calculating critical temperatures for ferromagnetic order in two-dimensional materials," *2D Mater.* **6**, 015028 (2018).
- ⁷¹D. Torelli, K. S. Thygesen, and T. Olsen, "High throughput computational screening for 2D ferromagnetic materials: The critical role of anisotropy and local correlations," *2D Mater.* **6**, 045018 (2019).
- ⁷²C. Gong and X. Zhang, "Two-dimensional magnetic crystals and emergent heterostructure devices," *Science* **363**, eaav4450 (2019).
- ⁷³C. Cardoso, D. Soriano, N. A. García-Martínez, and J. Fernández-Rossier, "Van der Waals spin valves," *Phys. Rev. Lett.* **121**, 067701 (2018).
- ⁷⁴D. Zhong, K. L. Seyler, X. Linpeng, R. Cheng, N. Sivadans, B. Huang, E. Schmidgall, T. Taniguchi, K. Watanabe, M. A. McGuire, W. Yao, D. Xiao, K.-M. C. Fu, and X. Xu, "Van der Waals engineering of ferromagnetic semiconductor heterostructures for spin and valleytronics," *Sci. Adv.* **3**, e1603113 (2017).
- ⁷⁵K. Burch, D. Mandrus, and J. Park, "Magnetism in two-dimensional van der Waals materials," *Nature* **563**, 47–52 (2018).
- ⁷⁶Z. Wang, I. Gutiérrez-Lezama, N. Ubrig, M. Kroner, M. Gibertini, T. Taniguchi, K. Watanabe, A. Imamoğlu, E. Giannini, and A. F. Morpurgo, "Very large tunneling magnetoresistance in layered magnetic semiconductor CrI_3 ," *Nat. Commun.* **9**, 2516 (2018).
- ⁷⁷T. Olsen, "Designing in-plane heterostructures of quantum spin hall insulators from first principles: $1T' - \text{MoS}_2$ with adsorbates," *Phys. Rev. B* **94**, 235106 (2016).
- ⁷⁸T. Olsen, E. Andersen, T. Okugawa, D. Torelli, T. Deilmann, and K. S. Thygesen, "Discovering two-dimensional topological insulators from high-throughput computations," *Phys. Rev. Mater.* **3**, 024005 (2019).
- ⁷⁹R. Kappera, D. Voiry, S. E. Yalcin, B. Branch, G. Gupta, A. D. Mohite, and M. Chhowalla, "Phase-engineered low-resistance contacts for ultrathin MoS_2 transistors," *Nat. Mater.* **13**, 1128–1134 (2014).
- ⁸⁰D. Akinwande, C. J. Brennan, J. S. Bunch, P. Egberts, J. R. Felts, H. Gao, R. Huang, J.-S. Kim, T. Li, Y. Li, K. M. Liechti, N. Lu, H. S. Park, E. J. Reed, P. Wang, B. I. Yakobson, T. Zhang, Y.-W. Zhang, Y. Zhou, and Y. Zhu, "A review on mechanics and mechanical properties of 2D materials—Graphene and beyond," *Extreme Mech. Lett.* **13**, 42–77 (2017).
- ⁸¹C. Androulidakis, K. Zhang, M. Robertson, and S. Tawfik, "Tailoring the mechanical properties of 2D materials and heterostructures," *2D Mater.* **5**, 032005 (2018).
- ⁸²L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 2nd ed., Course of Theoretical Physics Vol. 7 (Pergamon Press, Oxford, 1970).

- ⁸⁵J.-W. Jiang, S. Y. Kim, and H. S. Park, "Auxetic nanomaterials: Recent progress and future development," *Appl. Phys. Rev.* **3**, 041101 (2016).
- ⁸⁴T. C. T. Ting and T. Chen, "Poisson's ratio for anisotropic elastic materials can have no bounds," *Q. J. Mech. Appl. Math.* **58**, 73–82 (2005).
- ⁸⁵B. Liu, M. Niu, J. Fu, Z. Xi, M. Lei, and R. Quhe, "Negative Poisson's ratio in puckered two-dimensional materials," *Phys. Rev. Mater.* **3**, 054002 (2019).
- ⁸⁶X. Kong, J. Deng, L. Li, Y. Liu, X. Ding, J. Sun, and J. Z. Liu, "Tunable auxetic properties in group-IV monochalcogenide monolayers," *Phys. Rev. B* **98**, 184104 (2018).
- ⁸⁷L. C. Gomes, A. Carvalho, and A. H. Castro Neto, "Enhanced piezoelectricity and modified dielectric screening of two-dimensional group-IV monochalcogenides," *Phys. Rev. B* **92**, 214103 (2015).
- ⁸⁸J.-W. Jiang and H. S. Park, "Negative Poisson's ratio in single-layer black phosphorus," *Nat. Commun.* **5**, 4727 (2014).
- ⁸⁹Y. Du, J. Maassen, W. Wu, Z. Luo, X. Xu, and P. D. Ye, "Auxetic black phosphorus: A 2D material with negative Poisson's ratio," *Nano Lett.* **16**, 6701–6708 (2016).
- ⁹⁰L. Yu, Q. Yan, and A. Ruzsinszky, "Negative Poisson's ratio in 1T-type crystalline two-dimensional transition metal dichalcogenides," *Nat. Commun.* **8**, 15224 (2017).
- ⁹¹G. Qin and Z. Qin, "Negative Poisson's ratio in two-dimensional honeycomb structures," *NPJ Comput. Mater.* **6**, 51 (2020).
- ⁹²T. Cao, G. Wang, W. Han, H. Ye, C. Zhu, J. Shi, Q. Niu, P. Tan, E. Wang, B. Liu, and J. Feng, "Valley-selective circular dichroism of monolayer molybdenum disulfide," *Nat. Commun.* **3**, 887 (2012).
- ⁹³K. F. Mak, K. L. McGill, J. Park, and P. L. McEuen, "The valley hall effect in MoS₂ transistors," *Science* **344**, 1489–1492 (2014).
- ⁹⁴J. Lee, K. F. Mak, and J. Shan, "Electrical control of the valley hall effect in bilayer MoS₂ transistors," *Nat. Nanotechnol.* **11**, 421–425 (2016).
- ⁹⁵P. Chen, W. W. Pai, Y. H. Chan, W. L. Sun, C. Z. Xu, D. S. Lin, M. Y. Chou, A. V. Fedorov, and T. C. Chiang, "Large quantum-spin-hall gap in single-layer 1T' WSe₂," *Nat. Commun.* **9**, 2003 (2018).
- ⁹⁶K. E. Aretouli, P. Tsipas, D. Tsoutsou, J. Marquez-Velasco, E. Xenogiannopoulou, S. A. Giannini, E. Vassalou, N. Kelaidis, and A. Dimoulas, "Two-dimensional semiconductor HfSe₂ and MoSe₂/HfSe₂ van der Waals heterostructures by molecular beam epitaxy," *Appl. Phys. Lett.* **106**, 143105 (2015).
- ⁹⁷S. A. Melchior, K. Raju, I. S. Ike, R. M. Erasmus, G. Kabongo, I. Sigalas, S. E. Iyuke, and K. I. Ozoemena, "High-voltage symmetric supercapacitor based on 2D titanium carbide (MXene, Ti₂C₂X)/carbon nanosphere composites in a neutral aqueous electrolyte," *J. Electrochem. Soc.* **165**, A501–A511 (2018).
- ⁹⁸S. Mañas-Valero, V. García-López, A. Cantarero, and M. Galbiati, "Raman spectra of ZrS₂ and ZrSe₂ from bulk to atomically thin layers," *Appl. Sci.* **6**, 264 (2016).
- ⁹⁹Y. Sun, H. Cheng, S. Gao, Z. Sun, Q. Liu, Q. Liu, F. Lei, T. Yao, J. He, S. Wei, and Y. Xie, "Freestanding tin disulfide single-layers realizing efficient visible-light water splitting," *Angew. Chem. Int. Ed.* **51**, 8727–8731 (2012).
- ¹⁰⁰M. Zhang, Y. Zhu, X. Wang, Q. Feng, S. Qiao, W. Wen, Y. Chen, M. Cui, J. Zhang, C. Cai, and L. Xie, "Controlled synthesis of ZrS₂ monolayer and few layers on hexagonal boron nitride," *J. Am. Chem. Soc.* **137**, 7051–7054 (2015).
- ¹⁰¹W. Zheng, B. Zheng, C. Yan, Y. Liu, X. Sun, Z. Qi, T. Yang, Y. Jiang, W. Huang, P. Fan, F. Jiang, W. Ji, X. Wang, and A. Pan, "Direct vapor growth of 2D vertical heterostructures with tunable band alignments and interfacial charge transfer behaviors," *Adv. Sci.* **6**, 1802204 (2019).
- ¹⁰²Y. W. Park, S.-K. Jerng, J. H. Jeon, S. B. Roy, K. Akbar, J. Kim, Y. Sim, M.-J. Seong, J. H. Kim, Z. Lee, M. Kim, Y. Yi, J. Kim, D. Y. Noh, and S.-H. Chun, "Molecular beam epitaxy of large-area SnSe₂ with monolayer thickness fluctuation," *2D Mater.* **4**, 014006 (2016).
- ¹⁰³O. D. Pozo-Zamudio, S. Schwarz, M. Sich, I. A. Akimov, M. Bayer, R. C. Schofield, E. A. Chekhovich, B. J. Robinson, N. D. Kay, O. V. Kolosov, A. I. Dmitriev, G. V. Lashkarev, D. N. Borisenko, N. N. Kolesnikov, and A. I. Tartakovskii, "Photoluminescence of two-dimensional GaTe and GaSe films," *2D Mater.* **2**, 035010 (2015).
- ¹⁰⁴E. C. Le Ru and P. G. Etchegoin, *Principles of Surface-Enhanced Raman Spectroscopy and Related Plasmonic Effects*, 1st ed. (Elsevier, 2008), Vol. 7.
- ¹⁰⁵F. Hüser, T. Olsen, and K. S. Thygesen, "How dielectric screening in two-dimensional crystals affects the convergence of excited-state calculations: Monolayer MoS₂," *Phys. Rev. B* **88**, 245309 (2013).
- ¹⁰⁶T. Tian, D. Scullion, D. Hughes, L. H. Li, C.-J. Shih, J. Coleman, M. Chhowalla, and E. J. G. Santos, "Electronic polarizability as the fundamental variable in the dielectric properties of two-dimensional materials," *Nano Lett.* **20**, 841–851 (2020).
- ¹⁰⁷A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, "The electronic properties of graphene," *Rev. Mod. Phys.* **81**, 109 (2009).
- ¹⁰⁸K. S. Thygesen, "Calculating excitons, plasmons, and quasiparticles in 2D materials and van der Waals heterostructures," *2D Mater.* **4**, 022004 (2017).
- ¹⁰⁹H. Liu, A. T. Neal, Z. Zhu, Z. Luo, X. Xu, D. Tománek, and Y. P. D., "Phosphorene: An unexplored 2D semiconductor with a high hole mobility," *ACS Nano* **8**, 4033–4041 (2014).
- ¹¹⁰T. Wy, R. Soklaski, Y. Liang, and L. Yang, "Layer-controlled band gap and anisotropic excitons in few-layer black phosphorus," *Phys. Rev. B* **89**, 235319 (2014).
- ¹¹¹E. Knill, R. Laflamme, and G. A. Milburn, "A scheme for efficient quantum computation with linear optics," *Nature* **409**, 46–52 (2001).
- ¹¹²N. Zeng, X. Jiang, Q. Gao, Y. He, and H. Ma, "Linear polarization difference imaging and its potential applications," *Appl. Opt.* **48**, 6734–6739 (2009).
- ¹¹³X. Ling, S. Huang, E. H. Hasdeo, L. Liang, W. M. Parkin, Y. Tatsumi, A. R. T. Nugraha, A. A. Puzetzyk, P. M. Das, B. G. Sumpter, D. B. Geohegan, J. Kong, R. Saito, M. Drndic, V. Meunier, and M. S. Dresselhaus, "Anisotropic electron-photon and electron-phonon interactions in black phosphorus," *Nano Lett.* **16**, 2260–2267 (2016).
- ¹¹⁴L. H. Li, E. J. G. Santos, T. Xing, E. Cappelluti, R. Roldán, Y. Chen, K. Watanabe, and T. Taniguchi, "Dielectric screening in atomically thin boron nitride nanosheets," *Nano Lett.* **15**, 218–223 (2015).
- ¹¹⁵R. Xu, S. Zhang, F. Wang, J. Yang, Z. Wang, J. Pei, Y. W. Myint, B. Xing, Z. Yu, L. Fu *et al.*, "Extraordinarily bound quasi-one-dimensional trions in two-dimensional phosphorene atomic semiconductors," *ACS Nano* **10**, 2046–2053 (2016).
- ¹¹⁶J. Yang, R. Xu, J. Pei, Y. W. Myint, F. Wang, Z. Wang, S. Zhang, Z. Yu, and Y. Lu, "Optical tuning of exciton and trion emissions in monolayer phosphorene," *Light Sci. Appl.* **4**, e312–e312 (2015).
- ¹¹⁷T. Deilmann and K. S. Thygesen, "Unraveling the not-so-large trion binding energy in monolayer black phosphorus," *2D Mater.* **5**, 041007 (2018).
- ¹¹⁸M. N. Gjerding, L. S. Cavalcante, A. Chaves, and K. S. Thygesen, "Efficient ab-initio based modeling of dielectric screening in 2D van der Waals materials: Including phonons, substrates, and doping," *J. Phys. Chem. C* **124**, 11609–11616 (2020).
- ¹¹⁹D. Langreth and J. Perdew, "The exchange-correlation energy of a metallic surface," *Solid State Commun.* **17**, 1425–1429 (1975).
- ¹²⁰T. Olsen and K. S. Thygesen, "Random phase approximation applied to solids, molecules, and graphene-metal interfaces: From van der Waals to covalent bonding," *Phys. Rev. B* **87**, 075111 (2013).
- ¹²¹"Computational 2D materials database (C2DB)," <https://cmr.fysik.dtu.dk/c2db/c2db.html>

Paper II

Towards fully automatized GW band structure calculations: What we can learn from 60.000 self-energy evaluations

Asbjørn Rasmussen, Thorsten Deilmann and Kristian S. Thygesen

npj Computational Materials 7(1), 1-0 (2021) - Published 29 January 2021

ARTICLE OPEN



Towards fully automated GW band structure calculations: What we can learn from 60.000 self-energy evaluations

Asbjørn Rasmussen^{1,2}✉, Thorsten Deilmann³ and Kristian S. Thygesen^{1,2}

We analyze a data set comprising 370 GW band structures of two-dimensional (2D) materials covering 14 different crystal structures and 52 chemical elements. The band structures contain a total of 61716 quasiparticle (QP) energies obtained from plane-wave-based one-shot G_0W_0 @PBE calculations with full frequency integration. We investigate the distribution of key quantities, like the QP self-energy corrections and QP weights, and explore their dependence on chemical composition and magnetic state. The linear QP approximation is identified as a significant error source and we propose schemes for controlling and drastically reducing this error at low computational cost. We analyze the reliability of the $1/N$ basis set extrapolation and find that is well-founded with a narrow distribution of coefficients of determination (r^2) peaked very close to 1. Finally, we explore the accuracy of the scissors operator approximation and conclude that its validity is very limited. Our work represents a step towards the development of automatized workflows for high-throughput G_0W_0 band structure calculations for solids.

npj Computational Materials (2021)7:22; <https://doi.org/10.1038/s41524-020-00480-7>

INTRODUCTION

In computational materials science, the high-throughput mode of operation is becoming increasingly popular¹. The development of automatized workflow engines capable of submitting, controlling, and receiving thousands of interlinked calculations^{2–4} with minimal human intervention has greatly expanded the range of materials, and properties, that can be investigated by a single researcher. Several high-throughput studies have been conducted over the past decade mostly with the aim of identifying new prospect materials for various applications including catalysis⁵, batteries^{6,7}, thermoelectrics^{8,9}, photocatalysts¹⁰, transparent conductors¹¹, and photovoltaics^{12,13}, just to mention some. The vast amounts of data generated by such screening studies have been stored in open databases^{14–17} making them available for further processing, testing, and comparison of methods and codes, training of machine learning algorithms, etc. With very few exceptions, the high-throughput screening studies and the generation of materials databases, have been based on density functional theory (DFT) at the level of the generalized gradient approximation (GGA).

While DFT is fairly accurate for structural parameters and other properties related to the electronic ground state, it is well known that electronic band structures, in particular the size of band gaps, are not well reproduced by most xc-functionals¹⁸. This holds in particular for the LDA and GGA functionals, which hugely underestimate band gaps, often by about a factor of 2 or more^{19,20}. Hybrid functionals and certain metaGGAs perform significantly better²¹, but are not fully ab-initio and miss fundamental physics such as nonlocal screening effects²². Instead, the gold standard for quasiparticle band structure calculations of solids is the many-body GW method^{23–26}, which explicitly accounts for exchange and dynamical screening. In its simplest non-self-consistent form, i.e., G_0W_0 , this approximation reproduces experimental band gaps to within 0.3 eV (mean absolute error) or 10% (mean relative error)^{19,20,27}. We note in passing that for partially self-consistent GW_0 ²⁰ or when vertex corrections are

included^{28,29}, the deviation from experiments falls below 0.2 eV, which is comparable to the uncertainty of the experimental reference data. The improved accuracy of the GW method(s) comes at the price of a significantly more involved methodology both conceptually and numerically as compared to DFT. While DFT calculations can be routinely performed by non-experts using codes that despite very different numerical implementations produce identical results³⁰, GW calculations remain an art for the expert.

The high complexity of GW calculations is due to several factors including (i) The basic quantities of the theory, i.e., the Greens function (G) and screened Coulomb interaction (W) are dynamical quantities that depend on time/frequency. Several possibilities for handling the frequency dependence exists including the formally exact direct integration¹⁹ and contour deformation techniques³¹ as well as the controlled approximate analytic continuation methods³² and the rather uncontrolled but inexpensive plasmon-pole approximations²⁴. (ii) The formalism involves infinite sums over the unoccupied bands. While most implementations perform the sum explicitly up to a certain cutoff, schemes to avoid the sum over empty states have been developed^{33,34}. (iii) The basic quantities are two-point functions in real space (or reciprocal space) that couple states at different k -points. This leads to large memory requirements and makes it unfeasible to fully converge GW calculations with respect to the basis set. Consequently, strategies for extrapolation to the infinite basis set limit must be employed^{35,36}. (vi) Unless the GW equations are solved fully self-consistently, which is rarely done and does not improve accuracy^{29,37}, there is always a starting point dependence. This has been systematically explored for molecules where it was found that LDA/GGA often comprise a poor starting point whereas hybrids perform better in the sense that they lead to better agreement with experimental ionization potentials and produce more well-defined spectral peaks with higher quasiparticle weights^{38,39}. These and other factors imply that GW calculations not only become significantly more demanding than DFT in terms

¹CAMD, Department of Physics, Technical University of Denmark, 2800 Kongens Lyngby, Denmark. ²Center for Nanostructured Graphene (CNG), Technical University of Denmark, 2800 Kongens Lyngby, Denmark. ³Institut für Festkörperteorie, Westfälische Wilhelms-Universität Münster, 48149 Münster, Germany. ✉email: asbrae@fysik.dtu.dk

of computer resources, but they also involve more parameters making it difficult to assess whether the obtained results are properly converged or perhaps even erroneous.

Successful application of the high-throughput approach to problems involving excited electronic states, e.g., light absorption/emission, calls for the development of automatized and robust algorithms for setting the parameters of many-body calculations such as GW (according to available computational resources and required accuracy level), extrapolating the basis set, and assessing the reliability of the obtained results. The first step towards this goal is to analyze and systematize the data from large-scale GW studies. With a similar goal in mind, van Setten et al. compared G_0W_0 @PBE band gaps, obtained with the plasmon-pole approximation, to the experimental band gaps. They analyzed the correlations between different quantities and concluded that G_0W_0 (with plasmon-pole approximation) is more accurate than using an empirical correction of the PBE gap, but that, for accurate predictive results for a broad class of materials, an improved starting point or some type of self-consistency is necessary.

In this work we perform a detailed analysis of an extensive GW data set consisting of G_0W_0 @PBE band structures of 370 two-dimensional semiconductors comprising a total of 61,716 QP energies. Our focus is not on the ability of the G_0W_0 to reproduce experiments, i.e., its accuracy, which is well established by numerous previous studies, but rather on the numerical robustness and reliability of the method and the basis set extrapolation procedure. The calculations employ a plane-wave basis set and direct frequency integration; thus the use of projector augmented wave (PAW) potentials represent the only significant numerical approximation. We investigate the distribution of self-energy corrections and quasiparticle weights, Z , and explore their dependence on the materials composition and magnetic state. By investigating the full frequency-dependent self-energy for selected materials we analyze the error caused by the linear approximation to the QP equation and propose methods to estimate and correct this error. We assess the reliability of a plane-wave basis set extrapolation scheme finding it to be very accurate with a coefficient of determination, r^2 , values above 0.95 in more than 90% of the cases when extrapolation is performed from 200 eV. Finally, we assess the accuracy of the scissors operator approach and conclude that it should only be used when the average (maximal) band energy errors of 0.2 eV (2 eV) are acceptable.

RESULTS AND DISCUSSION

The G_0W_0 data set

The 370 G_0W_0 calculations were performed as part of the Computational 2D Materials Database (C2DB) project⁴⁰. Below we briefly recapitulate the computational details behind the G_0W_0 calculations and refer to Ref. ⁴⁰ for more details. All calculations were performed with the projector augmented wavefunction code GPAW⁴¹.

The C2DB database contains about 4000 monolayers comprising both known and hypothetical 2D materials constructed by decorating experimentally known crystal prototypes with a subset of elements from the periodic table⁴⁰. Currently, G_0W_0 calculations have been performed for 370 materials covering 14 different crystal structures and 52 different chemical elements. Figure 1a illustrates the distribution of elements. The number of materials containing a given element is shown below the element symbol. The number of magnetic materials containing the elements is shown in parenthesis next to the total number.

To give an overview of some of the data analyzed in this work, the distribution of the 61716 G_0W_0 corrections for the six bands around the bandgap is shown in Fig. 1b. The distribution for the valence bands is shown in blue and for the conduction in orange.

It is usually the case in GW studies that the DFT valence bands are shifted down and the conduction bands are shifted up. Similar behavior is found for the main part of our data, but we also observe a small subset of states for which the correction has the opposite sign. It is difficult to provide a clear physical explanation for why some occupied states are shifted up and some empty states are shifted down. We stress, however, that the GW corrections are measured relative to the PBE band energies, which is a somewhat arbitrary reference. For example, G_0W_0 @LDA and G_0W_0 @HSE would give different results—not so much for the resulting QP energies, which are relatively independent of the starting point—but for the size and sign of the GW corrections, which would now be measured relative to the LDA and HSE energies, respectively.

Figure 1c shows a scatter plot of the PBE energies versus the G_0W_0 energies. We only show energies from -10 to 10 eV for clarity. The color of a point shows the Z value. The latter has been truncated to the region $[0.5, 1.0]$ to show the variation of the main part of the distribution. The main observation we can make from this figure is that there is no obvious correlation between the energies and the Z values. This is also verified by the calculated correlation coefficient, C , between E_{PBE} and Z ($C = 0.27$), $E_{G_0W_0}$ and Z ($C = 0.23$) and between the G_0W_0 correction, $E_{G_0W_0} - E_{\text{PBE}}$, and Z ($C = 0.10$). We conclude that there is no significant correlation between the energies and Z , meaning that low Z values (which signals a break down of the QP approximation) may occur in any energy range.

Quasiparticle weight Z

The quasiparticle weight, Z , gives a rough measure of the validity of the quasiparticle picture, i.e., how well the charged excitations of the interacting electron system can be described by single-particle excitations from the ground state. In the “Methods” section, we prove a physical interpretation of the quasiparticle weight.

In the following, we analyze the 61,716 calculated QP weights, Z , contained in the C2DB database. As discussed in the Methods section, for the QP approximation to be well-founded Z should be close to 1. We split the Z values into two classes: quasiparticle-consistent (QP-c) for $Z \in [0.5, 1.0]$ and quasiparticle-inconsistent (QP-ic) for $Z \notin [0.5, 1.0]$. With this definition, QP-c states will have at least half of their spectral weight in the quasiparticle peak, but there is no deeper principle behind the threshold value of 0.5. We can expect that the QP approximation is more accurate for QP-c states than for QP-ic states.

Figure 2 shows a histogram of the Z -values (all extrapolated to the infinite plane-wave limit) corresponding to the 3 highest valence bands and 3 lowest conduction band of 370 semiconductors. The vast majority of the values are distributed around ≈ 0.75 with only 0.28% lying outside the physical range from 0 to 1 (0.16% are larger than one and 0.12% are negative). We find that 97.5% of the states are QP-c.

It is of interest to investigate if there are specific types of materials/elements that are particularly challenging to describe by G_0W_0 . Figure 3 shows a barplot of the percentage of QP-ic states in materials containing a given element (note the logarithmic scale). The result of this analysis performed on the non-magnetic (ferromagnetic) materials is shown in blue (orange). For example, a large percentage (about 65%) of the states in Co-containing materials are QP-ic. It is clear that magnetic materials contribute a large fraction of the QP-ic states. In fact, 0.36% of the non-magnetic states are QP-ic while 22% of the magnetic states are QP-ic. In general, it thus seems that the QP approximation is generally worse for magnetic materials.

We note that the employed PAW potentials are not strictly norm-conserving. It has previously been found that the use of norm-conserving pseudopotentials can be crucial for the

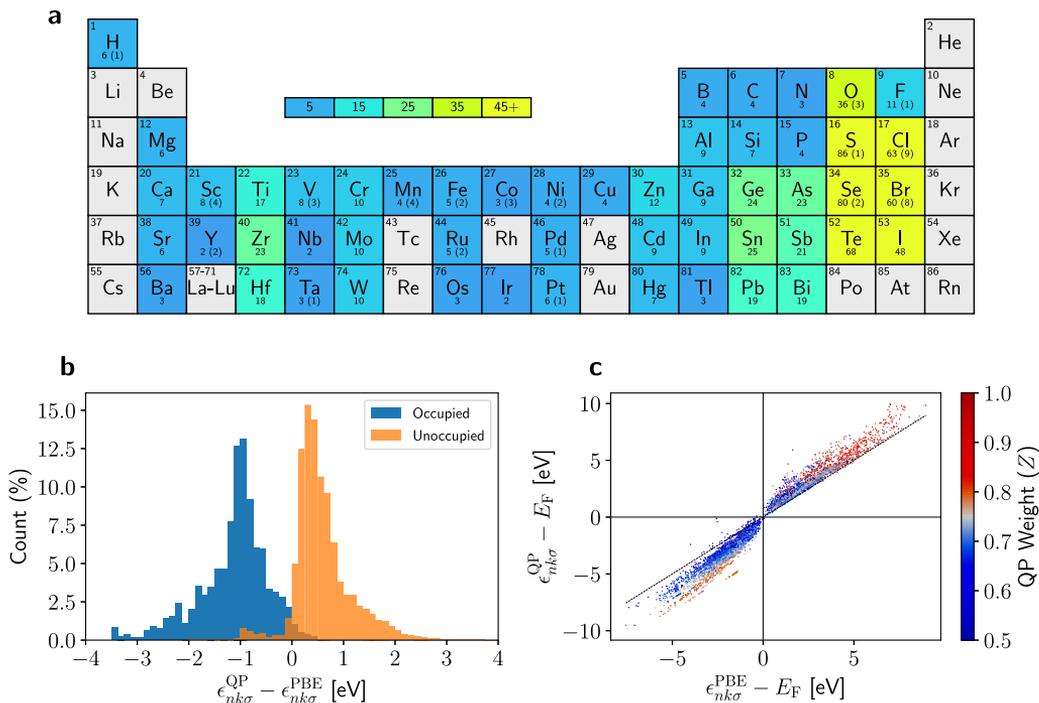


Fig. 1 The G_0W_0 data set. **a** The representation of individual elements in the G_0W_0 data set. The number of materials containing a given element is shown under the element's symbol. The number of magnetic materials, if any, is shown in the parenthesis next to the total number. **b** Histograms of quasiparticle energy corrections calculated from G_0W_0 . The blue histogram shows the three topmost occupied valence bands, while the orange shows the three lowest unoccupied conduction bands. **c** A scatter plot of the PBE energy vs. the G_0W_0 energy. The colors show the Z value truncated to the interval $[0.5, 1.0]$. The points are plotted so that a point with smaller Z is plotted on top of a point with larger Z if the two points overlap.

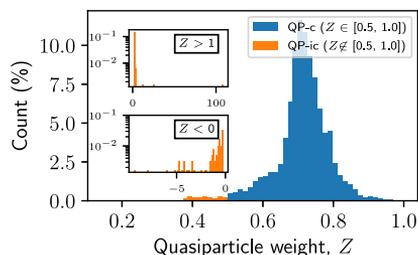


Fig. 2 Quasiparticle weights. Histogram of QP weights, Z , for the 61716 QP states in the C2DB⁴⁰. The Z values have been extrapolated to the infinite plane-wave limit (see next section). The main panel shows the distribution of Z values within the range, $Z \in [0, 1]$, while the upper and lower insets show the distribution outside the physical range, i.e., $Z > 1$ and $Z < 0$, respectively. 0.16% of points lie in the $Z > 1$ range, while 0.12% lie in the $Z < 0$ range.

quantitative accuracy of G_0W_0 results for materials with localized d or f states^{35,42,43}. To investigate this potential issue, we checked the distributions of G_0W_0 corrections and QP weights for materials containing at least one element with a pseudo partial wave of norm < 0.5 , i.e., materials where the norm-conservation could potentially be strongly violated for certain states. Out of the 370 materials, there were 279 materials in this category. The resulting distributions were not found to deviate qualitatively from those of all the materials (shown in Figs. 1b and 2, respectively), and the

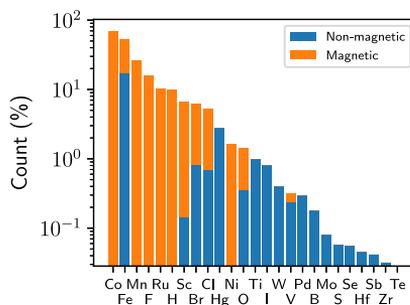


Fig. 3 QP-inconsistent solutions by element. Barplot showing the percentage of QP-ic Z values ($Z \notin [0.5, 1.0]$) for the given element. Non-magnetic materials are shown in blue and magnetic materials are shown in orange.

strongest indicator of unphysical Z values or opposite-sign G_0W_0 corrections remained the magnetic state of the material. On basis of this analysis, we conclude that the use of non-norm-conserving PAW potentials does not affect the conclusions of our study.

Based on the distribution of QP weights in Fig. 2, it appears that the QP approximation is valid for essentially all the states in the non-magnetic materials and most of the states in the magnetic materials. However, while a QP- c Z value is likely a necessary condition for predicting an accurate QP energy from the linearized QP equation [Eq. (6) in the "Method" section], it is not sufficient.

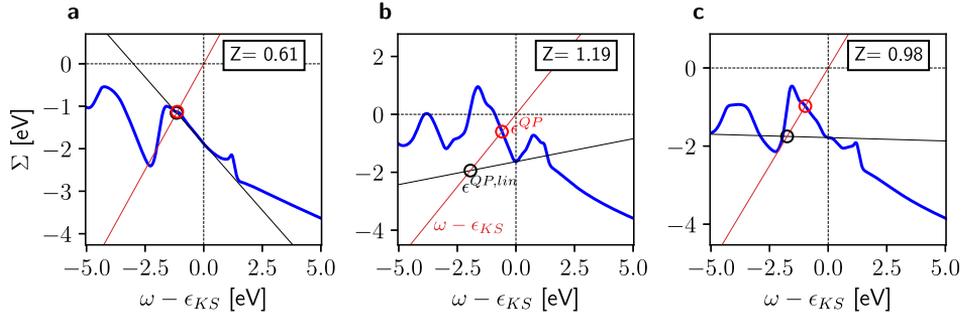


Fig. 4 Self-energies and the linear approximation. Frequency-dependent self-energy (blue) for three electronic states with different quasiparticle weights, Z . The red line indicates $\omega - \epsilon_{KS}$ while the black line is the linear approximation of the self-energy. The intersection of the blue and red lines indicate the solution to the quasiparticle equation, while the intersection between the red and black lines indicate the solution given by the linear approximation to the self-energy.

This is because the assumption behind Eq. (6), i.e., that $\Sigma(\epsilon)$ varies linearly with ϵ in the range between the KS energy and the QP energy, is not guaranteed for QP-c states. This is illustrated in Fig. 4, which shows the full frequency-dependent self-energy for three states in the ferromagnetic FeCl_2 . Case (a) is a typical example where the self-energy of a QP-c state ($Z=0.61$) varies linearly around ϵ_{KS} and the 1st order approximation works well. The second case (b) shows an example where the 1st order approximation breaks down for a QP-ic state ($Z=1.19$). The final case (c) illustrates that the 1st order approximation can break down even in cases where Z is very close to 1. Unfortunately, there is no simple way to diagnose such cases from the information available in a standard G_0W_0 calculation ($\Sigma(\epsilon_{KS})$ and Z). We stress that the example in Fig. 4c is a special case and that in general, the linear approximation is significantly more likely to hold for QP-c states than for QP-ic states (see discussion below).

Beyond the linear QP approximation

Under the assumption that the KS wave functions constitute a good approximation to the QP wave functions so that off-diagonal elements can be neglected, the solution to the QP equation reduces to solving an equation of the form

$$\omega - \epsilon_{KS} = \Sigma(\omega), \quad (1)$$

where $\Sigma(\omega) = \Sigma_{GW}(\omega) - v_{xc}$ is the frequency-dependent self-energy (see “Methods” section).

In this section, we investigate different root-finding schemes to estimate the size of the error introduced by the linear approximation and obtain an improved QP energy. With high-throughput computations in mind, a good algorithm provides a reasonable balance between computation time (number of Σ/Z evaluations) and accuracy. To benchmark the different schemes we computed the full frequency-dependent self-energy for 3192 states, corresponding to the 3 highest valence bands and 3 lowest conduction bands, for 12 of the 370 2D materials (including two ferromagnetic materials). The two ferromagnetic materials were chosen at random from materials that had some $Z \notin [0, 1]$. The remaining 10 materials were chosen at random from materials with all $Z \in [0, 1]$ and typical Z distributions. An overview of the materials is shown in Table 1. The self-energy is evaluated on a uniform frequency grid and interpolated using cubic splines. The “true” solution of the QP equation is then determined and used to evaluate the errors of the approximate schemes. In cases where there are multiple solutions, the smallest correction is selected.

We first determine the errors introduced by the linear approximation. Histograms of the errors for QP-c and QP-ic states are shown in Fig. 5. This shows that QP-ic generally has larger error and thus warrant particular attention.

Table 1. Properties of test materials summary of the 12 materials used to study the frequency-dependent self-energy.

Material	Prototype	Mag. state	PBE gap [eV]	G_0W_0 gap [eV]
HfBrI	MoSSe	NM	0.71	1.61
HfClI	MoSSe	NM	0.81	1.78
ZrBrCl	MoSSe	NM	0.91	1.88
ZrClI	MoSSe	NM	0.88	1.74
FeCl_2	MoS_2	FM	0.35	0.00
MnBr_2	MoS_2	FM	1.59	2.02
MoS_2	MoS_2	NM	1.58	2.53
PdSe_2	CdI_2	NM	0.56	1.61
Al_2Se_2	Ga_2S_2	NM	1.99	3.54
Ga_2S_2	Ga_2S_2	NM	2.32	4.08
Ga_2Se_2	Ga_2S_2	NM	1.76	3.44
In_2S_2	Ga_2S_2	NM	1.67	3.15

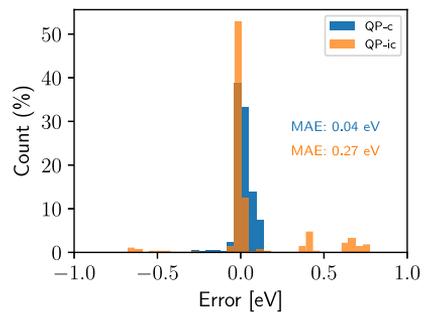


Fig. 5 Errors of quasiparticle-consistent and -inconsistent solutions. The distributions of the error incurred by the linear approximation as estimated from 3192 states in 12 different materials for which we have calculated the full frequency-dependent self-energy and determined the exact QP energy (see main text). The distribution for QP-c states is shown in blue, while the distribution for QP-ic states is shown in orange. The inset shows the full distribution for QP-ic states.

We first consider the iterative Newton–Raphson (NR) method where we limit ourselves to 1 and 2 iterations to keep the number of self-energy evaluations and thus the computational cost low. We note that 1 iteration (NR1) is equivalent to the linear

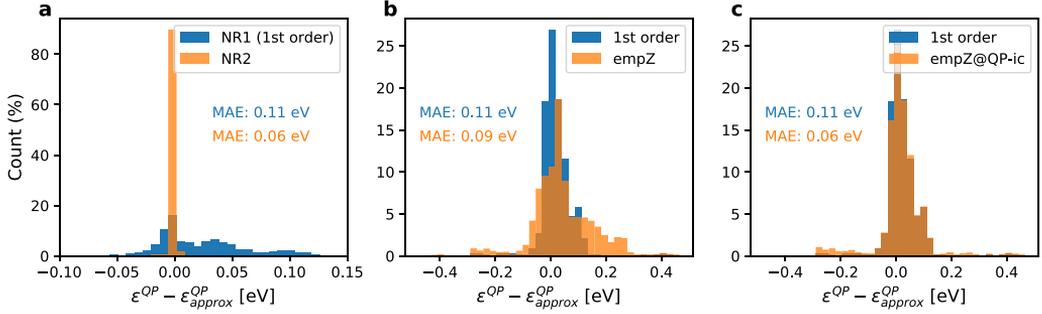


Fig. 6 Newton–Raphson and the empirical Z method. **a** The error distributions for first-order Newton–Raphson (NR1) (blue) and second-order Newton–Raphson (NR2) (orange). NR1 is equivalent to solving the linearized QP equation. **b** The NR1 distribution from **a** is again shown in blue for comparison. The orange distribution shows the error for the empirical empZ scheme. **c** The NR1 distribution is again shown in blue. The orange distribution is the error when the empZ scheme is applied only to the QP-ic states.

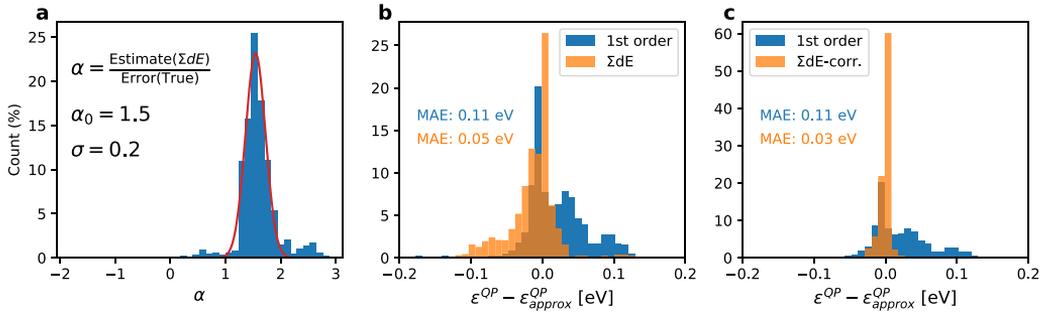


Fig. 7 Estimated errors and the ΣdE method. **a** The distribution of the ratio of the estimated error and the true error. Also shown in red is a gaussian fit to the distribution. The text annotations are shown the definition of α (top), the mean of the fitted gaussian, α_0 (middle), and the standard deviation of the fitted gaussian, σ , (bottom). **b** Distribution of the error of the linear approximation (blue) and the error of solution derived from the estimated error (orange). **c** Correcting for the mean of α yields improved solution estimates (orange).

approximation. The distribution of the errors is shown in Fig. 6a. Although 87% of the errors from NR1 are below 0.1 eV, the mean absolute error (MAE) is 0.11 eV due to outliers. Most of these errors are significantly reduced by performing one more iteration of Newton–Raphson (NR2), but again outliers increase the MAE. If we evaluate the MAE without the outliers (those lying outside the displayed error range), the MAE reduces to only 0.006 eV.

Motivated by the relatively narrow distribution of Z values in Fig. 2, we consider an empirical solution estimate consisting of replacing the actual Z value with the mean value of the distribution, i.e., we simply set $Z = 0.75$. This has the advantage of being simple, computationally cheap, and robust in the sense of avoiding outlier Z-values arising from local irregularities in Σ at the KS energy (Fig. 4b). The resulting error distribution is shown in Fig. 6b. While the central part of the distribution is slightly broadened compared to the 1st order approximation, the MAE is reduced due to a reduction of outliers (enhanced robustness). As shown in Fig. 6c, the central part of the distribution can be narrowed by applying the empirical approach only for QP-ic states, i.e., when $Z \notin [0.5, 1]$. In fact, this approach (empZ@QP-ic) has a MAE equal to that of NR2 but with half the computational cost (two Σ/Z evaluations compared to four).

Next, we examine the polynomial fitting of the self-energy. We construct second and fourth-order polynomials, $P_n(\omega)$, from the self-energy at energies in a range of ± 1 eV around the KS energy. The cost of the second and fourth-order fits is equivalent to three and five self-energy evaluations, respectively. In general, the polynomial fits have rather low correlation coefficients of $C < 0.9$ and are sensitive to the choice of frequency points and self-energy

data used for the fit. As a consequence, the resulting errors are large (not shown) and the approach is not suitable. We attribute this to our observation that self-energies are often irregular (on the relevant scale of 1 eV) and not well-described by low-order polynomials.

Finally, we consider a scheme that we refer to as ΣdE , which estimates the error as

$$\delta = \Sigma(\epsilon^{\text{QP,lin}}) - \left(\Sigma(\epsilon^{\text{KS}}) + \frac{d\Sigma}{d\omega} \Big|_{\omega=\epsilon^{\text{KS}}} (\epsilon^{\text{QP,lin}} - \epsilon^{\text{KS}}) \right). \quad (2)$$

The motivation for this expression is the following. If the linear approximation is exact, then δ vanishes as it should. Moreover, if the self-energy has a non-zero curvature it can be shown that δ equals the true error to leading order in the curvature. In that sense, it is similar to the second-order polynomial fit, but with the important difference that whereas the polynomial fit was based on uniformly distributed points, ΣdE uses the value and slope at E^{KS} and the value at $E^{\text{QP,lin}}$.

In Fig. 7a, the distribution of the ratios of the estimated error and true error is shown and the errors resulting from Eq. (2) are shown in Fig. 7b. Compared to the linear approximation, the ΣdE reduces the MAE from 0.11 to 0.05 eV, at the cost of one additional self-energy evaluation. Interestingly, Eq. (2) systematically overestimates the error as shown in Fig. 7a. A Gaussian fit to the distribution (red curve) has a mean value of $\alpha_0 = 1.5$ and a standard deviation of 0.2. Since the distribution of α is fairly narrow, it is tempting to correct for the systematic error using $\alpha = \alpha_0$, i.e., replacing $\delta \rightarrow \delta/\alpha_0$. We denote this estimate as ΣdE -corrected. To verify this procedure we randomly bisect the data into a “training” and a “test” set of equal size. α_0 is determined

from the training set and the MAE is calculated on the test set. The MAEs thus found were always 0.02–0.03 eV. We performed the same analysis using different sizes of the training set and found that an MAE of 0.03 eV is robust with a training set down to $\geq 5\%$ of data points. This indicates the approach is insensitive to data used to determine a_0 . In Fig. 7c, the Σ dE-corrected values are shown, where a_0 was determined from the full distribution for simplicity. The Σ dE-corrected scheme shows excellent performance with an almost four-fold reduction of the MAE from 0.11 eV for the linear approximation to only 0.03 eV at a computational overhead of just one additional self-energy evaluation.

The performance of the different correction schemes is summarized in Table 2.

Plane-wave extrapolation

The self-energy and the derivative of the self-energy (both evaluated at the KS energy) are calculated at three cutoff energies: 170, 185, and 200 eV. These values are then extrapolated to infinite cutoff, or an infinite number of plane waves, $N_{PW} \rightarrow \infty$, by assuming a linear dependence on the inverse number of plane waves⁴⁴. An example of this fitting procedure is shown in Fig. 8a. The extrapolation procedure saves computational time while improving the accuracy of the results—provided the extrapolation is sufficiently accurate. Extrapolation can fail if convergence as a function of the plane-wave cutoff for the given quantity does not follow the expected $1/N_{PW}$ behavior in the considered cutoff range.

To validate this approach, we investigate the distribution of the r^2 values for all 61716 extrapolations in C2DB. We split them into two cases: extrapolation of the self-energy and extrapolation of

Method	MAE [eV]	# Σ/Z evals
1st order	0.11	2
empZ	0.09	1
empZ@QP-ic	0.06	2
Σ dE	0.05	3
Σ dE-corr.	0.03	3

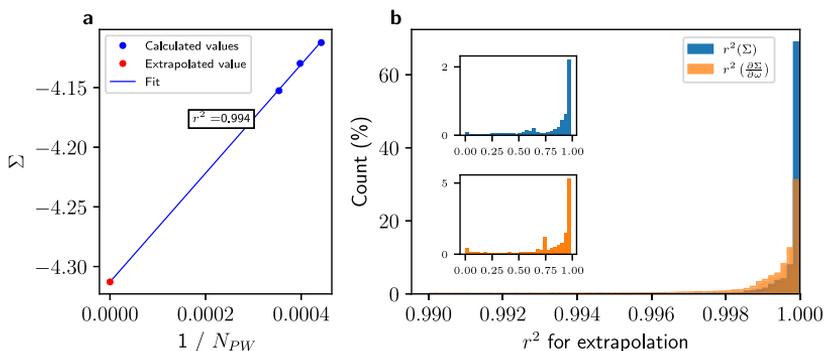


Fig. 8 Plane-wave extrapolation. **a** Example of the plane-wave extrapolation procedure for the G_0W_0 self-energy and its derivative. The quantity of interest, e.g., the self-energy, is calculated for three different cutoff energies, here 170, 185, and 200 eV, and the assumed linear dependence on $1/N_{PW}$ (N_{PW} is the number of plane waves) is extrapolated to the infinite basis set limit. The coefficient of determination for the fit, r^2 , is shown in the box. **b** Histogram of the coefficient of determination, r^2 , for the 61,716 plane-wave extrapolations of self-energies (blue) and the derivatives of the self-energy (orange). The plot shows the distribution for the coefficient of determination $r^2 \geq 0.99$, while the insets show values outside this range. A total of 5.5% and 14.1% of the values are <0.99 for the self-energy and its derivative, respectively.

the derivative of the self-energy. The distributions are shown as histograms in Fig. 8b. The distributions are clearly peaked very close to 1, and in general, it seems that the extrapolation is very good. The distribution for the derivatives is somewhat broader, and the extrapolation is generally less accurate than for the self-energies, which indicates a slower convergence with plane waves than for the self-energies. If we choose $r^2 = 0.8$ as an acceptable threshold, we find that 1.7% of the r^2 values of the self-energy extrapolation fall below this criterion while 5.0% are below for the derivative extrapolation. While these numbers might seem large, the problem is readily diagnosed (by the r^2 value) and can be alleviated by using higher plane-wave cutoffs.

Scissors operator approximation

Within the so-called scissors operator approximation (SOA) it is assumed that the G_0W_0 correction is independent of band- and k -index. Consequently, the G_0W_0 correction calculated at, e.g., the Γ point is applied to all the eigenvalues thus saving computational time as only one G_0W_0 correction is required. In Fig. 9a, the idea is illustrated for a generic band. With the notation from the figure, the SOA consists of setting $\Delta(k) = \Delta$ (or $\Delta_{n\sigma}(k) = \Delta_{n\sigma}$ when more than one band and spin is involved).

To test the accuracy of the SOA, we evaluate the mean absolute error ($\langle |\epsilon| \rangle$) and maximum absolute error ($\max(|\epsilon|)$) of the band energies obtained with the SOA for each of the 370 materials:

$$\langle |\delta| \rangle = \frac{1}{N_\sigma N_k N_n} \sum_{n,k,\sigma} |\Delta_{n\sigma}(k) - \Delta_{n\sigma}| \quad (3)$$

and

$$\max(|\delta|) = \max_{n,k,\sigma} \{ |\Delta_{n\sigma}(k) - \Delta_{n\sigma}| \}. \quad (4)$$

The distribution of these errors is shown in Fig. 9b, c. From Fig. 9b, we see that the mean error exceeds 100 meV for about half of all materials—a rather large error, comparable to the target accuracy of the G_0W_0 method itself. Furthermore, it follows from Fig. 9c that the maximum absolute error is often 0.5–1.0 eV. We conclude that while the average error of the SOA might be acceptable, it can produce significant errors for specific bands and should be used with care.

Summary and conclusions

As high-throughput computations are gaining popularity in the electronic structure community, it becomes important to establish

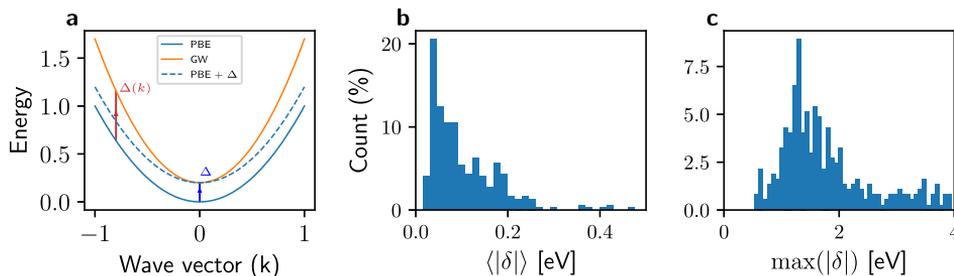


Fig. 9 Scissor operator approximation. **a** Illustration of the scissors operator approximation for a generic band. The G_0W_0 correction (Δ) is calculated at, e.g., the Γ -point and is used to correct the energies at all every k -point. This yields the scissors shifted band structure, here labeled “PBE + Δ ”. The actual G_0W_0 correction at the point k is labeled $\Delta(k)$. **b** Histogram showing the mean absolute error. **c** Maximum absolute error (**b**) of the scissors operator approximation. In both **b** and **c**, the average (maximum) is taken over the 3 highest valence bands and 3 lowest conduction bands in each of the 370 2D materials considered in this work.

protocols for performing various types of calculations in an automated, robust, and error-controlled manner. In this work, we have taken steps towards the development of automated workflows for G_0W_0 band structure calculations of solids. With G_0W_0 representing the state-of-the-art for predicting QP energies in condensed matter systems, such workflows are essential for continued progress in the field of computational materials design.

Based on our detailed analysis of 61,716 G_0W_0 self-energy evaluations for the eigenstates of 370 two-dimensional semiconductors we were able to draw several conclusions relevant to large-scale GW studies. First of all, we found it useful to divide the states into two categories, namely quasiparticle-consistent (QP-c) and quasiparticle-inconsistent (QP-ic) states defined by $Z \in [0.5, 1.0]$ and $Z \notin [0.5, 1.0]$, respectively. Importantly, we found that the QP energies obtained from the standard linearized QP equation are significantly more accurate for QP-c states than for QP-ic state. Moreover, we found the fraction of QP-ic states to be much larger in magnetic materials (22%) than in non-magnetic materials (0.36%). Thus, extra care must be taken when performing G_0W_0 calculations for magnetic materials; in particular, such materials might require special treatment in high-throughput workflows.

The mean absolute error (MAE) on the QP energies resulting from the linearized QP equation was found to be 0.11 eV. The MAE evaluated separately for QP-c and QP-ic states were 0.04 and 0.27 eV, respectively. In comparison, the accuracy of the GW approximation itself (compared to experiments) is on the order of 0.2 eV. It is therefore of interest to reduce or at least estimate the numerical error bar on the QP energies obtained from G_0W_0 calculations. We found that an empirical scheme, where we set $Z = 0.75$ (corresponding to the mean of the Z -distribution) for QP-ic states, reduces the MAE from 0.11 to 0.06 eV with no computational overhead. Similarly, the method dubbed the corrected Σ DE scheme reduces the MAE to 0.03 eV, at the cost of one additional self-energy evaluation. From these studies, it seems natural to accompany the QP energies obtained from G_0W_0 with estimated error bars derived from one of these correction schemes. In fact, we have used the empZ@QP-ic method to correct all the GW band structures in the C2DB database.

Our analysis of the well known and widely used scissors operator approximation shows that the errors introduced on the individual QP energies when averaged over all bands (specifically the 3 highest valence and 3 lowest conduction bands) typically is on the order of 0.1 eV while the maximum error typically exceeds 1 eV. We stress that our scissors operator fits each of the six bands separately using the G_0W_0 corrections at the Γ -point. Thus the errors introduced by the more standard scissors approximation that fits only the bandgap, are expected to be even larger. We conclude that the scissors operator should be used with care and

only in cases where errors on specific band energies of 1–3 eV are acceptable.

Finally, the plane-wave extrapolation scheme was found to be highly reliable for our PAW calculations when applied to cutoff energies in the range 180–200 eV. In fact, only 1.7% (5.0%) of the self-energy (a derivative of self-energy) extrapolations had an r^2 below 0.8. However, for the purpose of high-throughput studies, it may be prudent to store and make available information on the r^2 for the extrapolation so that the quality of the extrapolation can always be examined and improved calculations with higher cutoff can be performed if deemed necessary.

METHODS

G_0W_0 calculations

For the materials considered here, DFT calculations using PBE⁴⁵ were performed using an 800 eV plane-wave cutoff. Spin-orbit coupling is included by diagonalizing the spin-orbit Hamiltonian in the k -subspace of the Bloch states found from PBE.

Those materials that have a finite gap and up to 5 atoms in the unit cell are selected for G_0W_0 calculations. The QP energies in C2DB are calculated for the 8 highest occupied and the 4 lowest unoccupied bands, however, in this study we only use the 6 bands closest to the Fermi level (3 valence and 3 conduction bands). Furthermore, we only include materials with a PBE gap greater than 0.2 eV as the accuracy of G_0W_0 for materials with very small PBE gaps is questionable. Three energy cutoffs are used: 170, 185, and 200 eV. The results are then extrapolated to infinite energy, i.e., to an infinite number of plane waves. This extrapolation is done by expressing the self-energies in terms of the inverse number of plane waves, $1/N_{PW}$, performing a linear fit, and determining the value of the fit at $1/N_{PW} = 0$ ^{35,46}.

The screened Coulomb interaction entering in the self-energy is calculated using full frequency integration in real frequency space. To avoid effects from the (artificially) repeated layers. A Wigner-Seitz truncation scheme is used for the exchange part of the self-energy⁴⁷ and a 2D truncation of the Coulomb interaction is used for the correlation part^{44,48}. A truncated Coulomb interaction leads to significantly slower k -point convergence because the dielectric function strongly depends on q around $q = 0$; this is remedied by handling the integral around $q = 0$ analytically^{49,50}. A k -point density of $5.0/\text{\AA}^{-3}$ was used.

The statistical analyses performed here use the data from all spins, k -points, and the three highest occupied bands, and the three lowest unoccupied bands. In section IV B we consider several examples of the full frequency-dependent self-energies for a randomly selected spin, k -point, and band combination, subject to some requirements on the quasiparticle weight, Z , which are described below.

Quasiparticle theory

The G_0W_0 quasiparticle energies are found by solving the quasiparticle equation (QPE)³⁷:

$$E_{n\kappa}^{\text{QP}} = \text{Re}(\langle \psi_{n\kappa} | H_{KS} + \Sigma(E_{n\kappa}^{\text{QP}}) | \psi_{n\kappa} \rangle) \quad (5)$$

Here $\psi_{n\mathbf{k}\sigma}$ is the Kohn–Sham wavefunction for band n , crystal momentum \mathbf{k} , and spin σ , H_{KS} is the single-particle Kohn–Sham Hamiltonian, $\Sigma(\omega) = \Sigma_{\text{GW}}(\omega) - v_{\text{xc}}$ is the self-energy, and v_{xc} is the exchange–correlation potential.

Typically, and in C2DB, the QPE is solved via one iteration of the Newton–Raphson method starting from the KS energy, $\epsilon_{n\mathbf{k}\sigma}$, which is equivalent to making a linear approximation of the self-energy. This yields the solution

$$E_{n\mathbf{k}\sigma}^{\text{QP}} \approx \epsilon_{n\mathbf{k}\sigma} + Z \text{Re}[(\psi_{n\mathbf{k}\sigma} | \Sigma(\epsilon_{n\mathbf{k}\sigma}) | \psi_{n\mathbf{k}\sigma})], \quad (6)$$

$$Z = \left(1 - \frac{\partial \Sigma}{\partial \omega} \Big|_{\omega=\epsilon_{n\mathbf{k}\sigma}} \right)^{-1}. \quad (7)$$

Z is known as the quasiparticle weight. The G_0W_0 correction is defined as the difference between the G_0W_0 energy and KS energy, $\Delta E_{n\mathbf{k}\sigma} = E_{n\mathbf{k}\sigma}^{\text{QP}} - \epsilon_{n\mathbf{k}\sigma}$.

Following ref. 49, we provide here a physical interpretation of Z . We denote the many-body eigenstates for the N particle system by $|\Psi_i^N\rangle$, where i is the excitation index. An interesting question is how well the state $|\Psi_i^{N+1}\rangle$ can be described as the addition of a single electron to the ground state $|\Psi_0^N\rangle$. In other words, can we find a state ϕ such that $|\Psi_i^{N+1}\rangle \approx c_\phi^\dagger |\Psi_0^N\rangle$? The optimal ϕ is determined from maximizing the overlap, i.e.,

$$\phi = \arg \max_{\phi} (|\langle \Psi_i^{N+1} | c_\phi^\dagger | \Psi_0^N \rangle|, \|\phi\| = 1) \quad (8)$$

If the maximal overlap is close to 1 the excited many-body state is well approximated by a single-particle excitation.

It turns out that the square of this maximal overlap is exactly equal to the QP weight Z defined by Eq. (6) if it is evaluated at the true QP energy and with the true QP wavefunction rather than at the KS energy and with the KS wavefunction. Furthermore, Z can be shown to be equal to the squared norm of the QP wavefunction, which is defined as

$$\psi_i^{\text{QP}}(\mathbf{r}) = \langle \Psi_i^{N+1} | \hat{\psi}^\dagger(\mathbf{r}) | \Psi_0^N \rangle. \quad (9)$$

For proof of these results, we refer to ref. 49. In standard G_0W_0 calculations, the self-energy is evaluated at the KS energy using KS eigenstates. In this case, Z is no longer equal to the exact QP weight but only approximates it. If Z deviates significantly from 1, we can only conclude that either (1) the system is strongly correlated so that the QP approximation fails, or (2) the Kohn–Sham energy and/or wavefunction are a bad approximation to the true QP energy and/or wavefunction. In either case, we would expect that the G_0W_0 calculation is problematic and requires special attention.

DATA AVAILABILITY

Data are available as an ASE⁵¹ database at <https://cmr.fysik.dtu.dk/htgw/htgw.html>.

Received: 20 July 2020; Accepted: 12 December 2020;

Published online: 29 January 2021

REFERENCES

1. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
2. Jain, A. et al. Fireworks: a dynamic workflow system designed for high-throughput applications. *Concurr. Comput.* **27**, 5037–5059 (2015).
3. Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N. & Kozinsky, B. Aiida: automated interactive infrastructure and database for computational science. *Comput. Mater. Sci.* **111**, 218–230 (2016).
4. Mortensen, J., Gjerding, M. & Thygesen, K. Myqueue: Task and workflow scheduling system. *J. Open Source Softw.* **5**, 1844 (2020).
5. Greeley, J., Jaramillo, T. F., Bonde, J., Chorkendorff, I. & Nørskov, J. K. Computational high-throughput screening of electrocatalytic materials for hydrogen evolution. *Nat. Mater.* **5**, 909–913 (2006).
6. Kirklin, S., Meredig, B. & Wolverton, C. High-throughput computational screening of new li-ion battery anode materials. *Adv. Energy Mater.* **3**, 252–262 (2013).
7. Zhang, Z. et al. Computational screening of layered materials for multivalent ion batteries. *ACS Omega* **4**, 7822–7828 (2019).
8. Chen, W. et al. Understanding thermoelectric properties from high-throughput calculations: trends, insights, and comparisons with experiment. *J. Mater. Chem. C* **4**, 4414–4426 (2016).

9. Bhattacharya, S. & Madsen, G. K. High-throughput exploration of alloying as design strategy for thermoelectrics. *Phys. Rev. B Condens. Matter* **92**, 085205 (2015).
10. Castellì, I. E. et al. Computational screening of perovskite metal oxides for optimal solar light capture. *Energy Environ. Sci.* **5**, 5814–5819 (2012).
11. Hautier, G., Miglio, A., Ceder, G., Rignanese, G.-M. & Gonze, X. Identification and design principles of low hole effective mass p-type transparent conducting oxides. *Nat. Commun.* **4**, 1–7 (2013).
12. Yu, L. & Zunger, A. Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials. *Phys. Rev. Lett.* **108**, 068701 (2012).
13. Kuhar, K., Pandey, M., Thygesen, K. S. & Jacobsen, K. W. High-throughput computational assessment of previously synthesized semiconductors for photovoltaic and photoelectrochemical devices. *ACS Energy Lett.* **3**, 436–446 (2018).
14. Thygesen, K. S. & Jacobsen, K. W. Making the most of materials computations. *Science* **354**, 180–181 (2016).
15. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (oqmd). *JOM* **65**, 1501–1509 (2013).
16. Jain, A. et al. Commentary: The materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
17. Curtarolo, S. et al. Aflow: an automatic framework for high-throughput materials discovery. *Comput. Mater. Sci.* **58**, 218–226 (2012).
18. Godby, R., Schlüter, M. & Sham, L. Accurate exchange–correlation potential for silicon and its discontinuity on addition of an electron. *Phys. Rev. Lett.* **56**, 2415 (1986).
19. Hüser, F., Olsen, T. & Thygesen, K. S. Quasiparticle GW calculations for solids, molecules, and two-dimensional materials. *Phys. Rev. B Condens. Matter* **87**, 235132 (2013).
20. Shishkin, M. & Kresse, G. Self-consistent GW calculations for semiconductors and insulators. *Phys. Rev. B Condens. Matter* **75**, 235102 (2007).
21. Borlido, P. et al. Exchange–correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning. *Npj Comput. Mater.* **6**, 1–17 (2020).
22. 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* **80**, 245427 (2009).
23. Hedin, L. New method for calculating the one-particle green’s function with application to the electron–gas problem. *Phys. Rev.* **139**, A796 (1965).
24. Hybertsen, M. S. & Louie, S. G. Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies. *Phys. Rev. B Condens. Matter* **34**, 5390 (1986).
25. Aryasetiawan, F. & Gunnarsson, O. The GW method. *Rep. Prog. Phys.* **61**, 237 (1998).
26. Golze, D., Dvorak, M. & Rinke, P. The GW compendium: a practical guide to theoretical photoemission spectroscopy. *Front. Chem.* **7**, 377 (2019).
27. Nabok, D., Gulans, A. & Draxl, C. Accurate all-electron G₀W₀ quasiparticle energies employing the full-potential augmented plane-wave method. *Phys. Rev. B Condens. Matter* **94**, 035118 (2016).
28. Shishkin, M., Marsman, M. & Kresse, G. Accurate quasiparticle spectra from self-consistent GW calculations with vertex corrections. *Phys. Rev. Lett.* **99**, 246403 (2007).
29. Schmidt, P. S., Patrick, C. E. & Thygesen, K. S. Simple vertex correction improves GW band energies of bulk and two-dimensional crystals. *Phys. Rev. B Condens. Matter* **96**, 205206 (2017).
30. Lejaeghere, K. et al. Reproducibility in density functional theory calculations of solids. *Science* **351**, aad3000 (2016).
31. Faber, C., Attaccalite, C., Olevano, V., Runge, E. & Blase, X. First-principles GW calculations for dna and rna nucleobases. *Phys. Rev. B Condens. Matter* **83**, 115123 (2011).
32. Caruso, F., Rinke, P., Ren, X., Scheffler, M. & Rubio, A. Unified description of ground and excited states of finite systems: The self-consistent GW approach. *Phys. Rev. B Condens. Matter* **86**, 081102 (2012).
33. Umari, P., Stenuit, G. & Baroni, S. GW quasiparticle spectra from occupied states only. *Phys. Rev. B Condens. Matter* **81**, 115104 (2010).
34. Govoni, M. & Galli, G. Large scale GW calculations. *J. Chem. Theory Comput.* **11**, 2680–2696 (2015).
35. Klimeš, J., Kaltak, M. & Kresse, G. Predictive GW calculations using plane waves and pseudopotentials. *Phys. Rev. B Condens. Matter* **90**, 075125 (2014).
36. Rasmussen, F. A. & Thygesen, K. S. Computational 2d materials database: electronic structure of transition-metal dichalcogenides and oxides. *J. Phys. Chem. C* **119**, 13169–13183 (2015).
37. Shishkin, M. & Kresse, G. Implementation and performance of the frequency-dependent GW method within the paw framework. *Phys. Rev. B Condens. Matter* **74**, 035101 (2006).

38. Rostgaard, C., Jacobsen, K. W. & Thygesen, K. S. Fully self-consistent gw calculations for molecules. *Phys. Rev. B Condens. Matter* **81**, 085103 (2010).
39. Bruneval, F. & Marques, M. A. Benchmarking the starting points of the gw approximation for molecules. *J. Chem. Theory Comput.* **9**, 324–329 (2013).
40. Hastrup, S. et al. The computational 2d materials database: high-throughput modeling and discovery of atomically thin crystals. *2D Mater.* **5**, 042002 (2018).
41. Enkovaara, J. E. et al. Electronic structure calculations with gpaw: a real-space implementation of the projector augmented-wave method. *J. Phys. Condens. Matter* **22**, 253202 (2010).
42. Jiang, H. & Blaha, P. GW with linearized augmented plane waves extended by high-energy local orbitals. *Phys. Rev. B Condens. Matter* **93**, 115203 (2016).
43. Jiang, H. Revisiting the GW approach to d- and f-electron oxides. *Phys. Rev. B Condens. Matter* **97**, 245132 (2018).
44. Rozzi, C. A., Varsano, D., Marini, A., Gross, E. K. & Rubio, A. Exact coulomb cutoff technique for supercell calculations. *Phys. Rev. B Condens. Matter* **73**, 205119 (2006).
45. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
46. Tiago, M. L., Ismail-Beigi, S. & Louie, S. G. Effect of semicore orbitals on the electronic band gaps of Si, Ge, and GaAs within the GW approximation. *Physical Rev. B* **69**, 125212 (2004).
47. Sundararaman, R. & Arias, T. Regularization of the coulomb singularity in exact exchange by wigner-seitz truncated interactions: Towards chemical accuracy in nontrivial systems. *Phys. Rev. B Condens. Matter* **87**, 165122 (2013).
48. Ismail-Beigi, S. Truncation of periodic image interactions for confined systems. *Phys. Rev. B Condens. Matter* **73**, 233103 (2006).
49. Hüser, F., Olsen, T. & Thygesen, K. S. Quasiparticle gw calculations for solids, molecules, and two-dimensional materials. *Phys. Rev. B Condens. Matter* **87**, 235132 (2013).
50. Rasmussen, F. A., Schmidt, P. S., Winther, K. T. & Thygesen, K. S. Efficient many-body calculations for two-dimensional materials using exact limits for the screened potential: Band gaps of mos 2, h-bn, and phosphorene. *Phys. Rev. B Condens. Matter* **94**, 155406 (2016).
51. Larsen, A. H. et al. The atomic simulation environment—a python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).

ACKNOWLEDGEMENTS

We acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant No. 773122, LIMA). The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation, Project DNRF103. This project has received funding in the

European Union's Horizon 2020 research and innovation program under the European Union's Grant Agreement No. 951786 (NOMAD CoE). T.D. acknowledges financial support from the German Research Foundation (DFG Project No. DE 2749/2-1).

AUTHOR CONTRIBUTIONS

A.R. performed the statistical analyses and full, frequency-dependent self-energy calculations. T.D. performed the G_0W_0 calculations. K.S.T. conceptualized the project. All authors interpreted the analyses and wrote the article.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to A.R.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Paper III

Recent Progress of the Computational 2D Materials Database (C2DB)

Morten Niklas Gjerding, Alireza Taghizadeh, **Asbjørn Rasmussen**, Sajid Ali, Fabian Bertoldo, Thorsten Deilmann, Urko Petralanda Holguin, Nikolaj RørbækKnøsgaard, Mads Kruse, Simone Manti, Thomas Garm Pedersen, Thorbjørn Skovhus, Mark Kamper Svendsen, Jens Jørgen Mortensen, Thomas Olsen and Kristian Sommer Thygesen

Submitted

Recent Progress of the Computational 2D Materials Database (C2DB)

Morten Niklas Gjerding¹, Alireza Taghizadeh^{2,1}, Asbjørn Rasmussen¹, Sajid Ali¹, Fabian Bertoldo¹, Thorsten Deilmann³, Nikolaj Rørbæk Knøsgaard¹, Mads Kruse¹, Simone Manti¹, Thomas Garm Pedersen², Urko Petralanda¹, Thorbjørn Skovhus¹, Mark Kamper Svendsen¹, Jens Jørgen Mortensen¹, Thomas Olsen¹ and Kristian Sommer Thygesen¹

¹ Computational Atomic-scale Materials Design (CAMD), Department of Physics, Technical University of Denmark, 2800 Kgs. Lyngby Denmark

² Department of Materials and Production, Aalborg University, 9220 Aalborg Øst, Denmark

³ Institut für Festkörpertheorie, Westfälische Wilhelms-Universität Münster, 48149 Münster, Germany

E-mail: thygesen@fysik.dtu.dk

Abstract. The C2DB is a highly curated open database organizing a wealth of computed properties for more than 4000 atomically thin two-dimensional (2D) materials. Here we report on new materials and properties that were added to the database since its first release in 2018. The set of new materials comprise several hundred monolayers exfoliated from experimentally known layered bulk materials, (homo)bilayers in various stacking configurations, native point defects in semiconducting monolayers, and chalcogen/halogen Janus monolayers. The new properties include exfoliation energies, Bader charges, spontaneous polarisations, Born charges, infrared polarisabilities, piezoelectric tensors, band topology invariants, exchange couplings, Raman- and second harmonic generation spectra. We also describe refinements of the employed material classification schemes, upgrades of the computational methodologies used for property evaluations, as well as significant enhancements of the data documentation and provenance. Finally, we explore the performance of Gaussian process-based regression for efficient prediction of mechanical and electronic materials properties. The combination of open access, detailed documentation, and extremely rich materials property data sets make the C2DB a unique resource that will advance the science of atomically thin materials.

Submitted to: *2D Mater.*

1. Introduction

The discovery of new materials, or new properties of known materials, to meet a specific industrial or scientific requirement, is an exciting intellectual challenge of the utmost importance for our environment and economy. For example, the successful transition to a society based on sustainable energy sources and the realisation of quantum technologies (e.g. quantum computers and quantum communication) depend critically on new materials with novel functionalities. First-principles quantum mechanical calculations, e.g. based on density functional theory (DFT)[1], can predict the properties of materials with high accuracy even before they are made in the lab. They provide insight into mechanisms at the most fundamental (atomic and electronic) level and can pinpoint and calculate key properties that determine the performance of the material at the macroscopic level. Powered by high-performance computers, atomistic quantum calculations in combination with data science approaches, have the potential to revolutionize the way we discover and develop new materials.

Atomically thin, two-dimensional (2D) crystals represent a fascinating class of materials with exciting perspectives for both fundamental science and technology[2, 3, 4, 5]. The family of 2D materials has been growing steadily over the past decade and counts about a hundred materials that have been realised in single- or few-layer form[6, 7, 8, 9, 10]. While some of these materials, including graphene, hexagonal boron-nitride (hBN), and transition metal dichalcogenides (TMDs), have been extensively studied, the majority have only been scarcely characterized and remain poorly understood. Computational studies indicate that around 1000 already known layered crystals have sufficiently weak interlayer bonding to allow the individual layers to be mechanically exfoliated[11, 12]. Supposedly, even more 2D materials could be realized beyond this set of already known crystals. Adding to this the possibility of stacking individual 2D layers (of the same or different kinds) into ultrathin van der Waals (vdW) crystals[13], and tuning the properties of such structures by varying the relative twist angle between adjacent layers[14, 15] or intercalating atoms into the vdW gap[16, 17], it is clear that the prospects of tailor made 2D materials are simply immense. To support experimental efforts and navigate the vast 2D materials space, first-principles calculations play a piv-

otal role. In particular, FAIR‡[18] databases populated by high-throughput calculations can provide a convenient overview of known materials and point to new promising materials with desired (predicted) properties. Such databases are also a fundamental requirement for the successful introduction and deployment of artificial intelligence in materials science.

Many of the unique properties exhibited by 2D materials have their origin in quantum confinement and reduced dielectric screening. These effects tend to enhance many-body interactions and lead to profoundly new phenomena such as strongly bound excitons[19, 20, 21] with nonhydrogenic Rydberg series[22, 23, 24], phonons and plasmons with anomalous dispersion relations[25, 26], large dielectric band structure renormalizations[27, 28], unconventional Mott insulating and superconducting phases[14, 15], and high-temperature exciton condensates[29]. Recently, it has become clear that long range magnetic order can persist[30, 31] and (in-plane) ferroelectricity even be enhanced[32], in the single layer limit. In addition, first-principles studies of 2D crystals have revealed rich and abundant topological phases[33, 34]. The peculiar physics ruling the world of 2D materials entails that many of the conventional theories and concepts developed for bulk crystals break down or require special treatments when applied to 2D materials[35, 26, 36]. This means that computational studies must be performed with extra care, which in turn calls for well-organized and well-documented 2D property data sets that can form the basis for the development, benchmarking, and consolidation of physical theories and numerical implementations.

The Computational 2D Materials Database (C2DB)[6, 37] is a highly curated and fully open database containing elementary physical properties of around 4000 two-dimensional (2D) monolayer crystals. The data has been generated by automatic high-throughput calculations at the level of density functional theory (DFT) and many-body perturbation theory as implemented in the GPAW electronic structure code. The computational workflow is constructed using the Atomic Simulation Recipes (ASR) – a recently developed Python framework for high-throughput materials modeling building on the Atomic Simulation Environment (ASE) – and managed/executed using the

‡ FAIR data are data which meet principles of findability, accessibility, interoperability, and reusability

MyQueue task scheduler[38].

The C2DB differentiates itself from existing computational databases of bulk[39, 40, 41] and low-dimensional[12, 11, 42] materials, by the large number of physical properties available, see Table 1. The use of beyond-DFT theories for excited state properties (GW band structures and BSE absorption for selected materials) and Berry-phase techniques for band topology and polarization quantities (spontaneous polarization, Born charges, piezoelectric tensors), are other unique features of the database.

The C2DB can be downloaded in its entirety or browsed and searched online. As a new feature, all data entries presented on the website are accompanied by a clickable help icon that presents a scientific documentation ("what does this piece of data describe?") and technical documentation ("how was this piece of data computed?"). This development enhances the usability of the database and improves the reproducibility and provenance of the data contained in C2DB. As another novelty it is possible to download all property data pertaining to a specific material or a specific type of property, e.g. the band gap, for all materials thus significantly improving data accessibility.

In this paper, we report on the significant C2DB developments that have taken place during the past two years. These developments can be roughly divided into four categories: (i) General updates of the workflow used to select, classify, and stability assess the materials. (ii) Computational improvements for properties already described in the 2018 paper. (iii) New properties. (vi) New materials. The developments, described in four separate sections, cover both original work and review of previously published work. In addition, we have included some outlook discussions of ongoing work. In the last section we illustrate an application of statistical learning to predict properties directly from the atomic structure.

2. Selection, classification, and stability

Figure 1 illustrates the workflow behind the C2DB. In this section we describe the first part of the workflow until the property calculations (red box), focusing on aspects related to selection criteria, classification, and stability assessment, that have been changed or updated since the 2018 paper.

2.1. Structure relaxation

Given a prospect 2D material, the first step is to carry out a structure optimization. This calculation is performed with spin polarization and with the symmetries of the original structure enforced. The latter is done to keep the highest level of control

over the resulting structure by avoiding "uncontrolled" symmetry breaking distortions. The prize to pay is a higher risk of generating dynamically unstable structures.

2.2. Selection: Dimensionality analysis

A dimensionality analysis[43] is performed to identify and filter out materials that have disintegrated into non-2D structures during relaxation. Covalently bonded clusters are identified through an analysis of the connectivity of the structures where two atoms are considered to belong to the same cluster if their distance is less than some scaling of the sum of their covalent radii, i.e. $d < k(r_i^{\text{cov}} + r_j^{\text{cov}})$, where i and j are atomic indices. A scaling factor of $k = 1.35$ was determined empirically. Only structures that consist of a single 2D cluster after relaxation are further processed. Figure 2 shows three examples (graphene, Ge_2Se_2 , and Pb_2O_6) of structures and their cluster dimensionalities before and after relaxation. All structures initially consist of a single 2D cluster, but upon relaxation Ge_2Se_2 and Pb_2O_6 disintegrate into two 2D clusters as well as one 2D and two 0D clusters, respectively. On the other hand, the relaxation of graphene decreases the in-plane lattice constant but does not affect the dimensionality. According to the criterion defined above only graphene will enter the database.

2.3. Selection: Ranking similar structures

Maintaining a high-throughput database inevitably requires a strategy for comparing similar structures and ranking them according to their relevance. In particular, this is necessary in order to identify different representatives of the same material e.g. resulting from independent relaxations, and thereby avoid duplicate entries and redundant computations. The C2DB strategy to this end involves a combination of structure clustering and Pareto analysis.

First, a single-linkage clustering algorithm is used to group materials with identical reduced chemical formula and "similar" atomic configurations. To quantify configuration similarity a slightly modified version of PyMatGen's[44] distance metric is employed where the cell volume normalization is removed to make it applicable to 2D materials surrounded by vacuum. Roughly speaking, the metric measures the maximum distance an atom must be moved (in units of Å) in order to match the two atomic configurations. Two atomic configurations belong to the same cluster if their distance is below an empirically determined threshold of 0.3 Å.

At this point, the simplest strategy would be to remove all but the most stable compound within

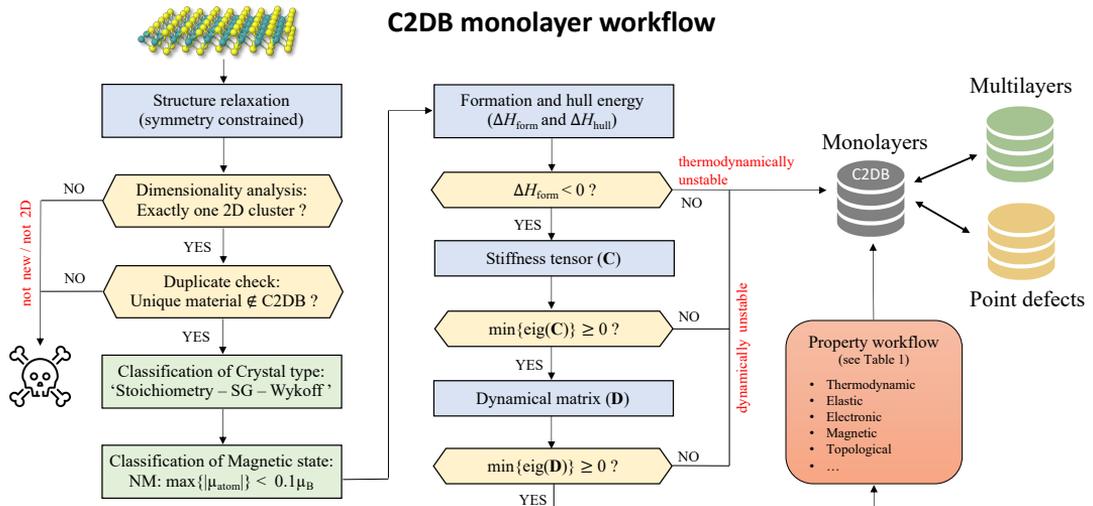


Figure 1. The workflow behind the C2DB. After the structural relaxation, the dimensionality of the material is checked and it is verified that the material is not already present in the database. Next, the material is classified according to its chemical composition, crystal structure, and magnetic state. Finally, the thermodynamic- and dynamic stability is assessed from the energy above the convex hull and the sign of the minimum eigenvalues of the dynamical matrix and stiffness tensor. Unstable materials are stored in the database; stable materials are subject to the property workflow. The C2DB monolayer database is interlinked with databases containing structures and properties of multilayer stacks and point defects in monolayers from the C2DB.

a cluster. However, this procedure would remove many high symmetry crystals for which a more stable distorted version exists. For example, the well known T-phase of MoS_2 would be removed in favor of the more stable T'-phase. This is undesired as high-symmetry structures, even if dynamically unstable at $T = 0$, may provide useful information and might in fact become stabilized at higher temperatures[45]. Therefore, the general strategy adopted for the C2DB, is to keep a material that is less stable than another material of the same cluster if it has fewer atoms in its primitive unit cell (and thus typically higher symmetry). Precisely, materials within a given cluster are kept only if they represent a defining point of the $(N, \Delta H)$ -Pareto front, where N is the number of atoms in the unit cell and ΔH is the heat of formation. A graphical illustration of the Pareto analysis is shown in Figure 3 for the case of ReS_2 .

2.4. Classification: Crystal structure

The original C2DB employed a *crystal prototype* classification scheme where specific materials were promoted to prototypes and used to label groups of materials with the same or very similar crystal structure. This approach was found to be difficult to maintain (as well as being non-transparent). Instead, materials are now classified according to their *crystal*

type defined by the reduced stoichiometry, space group number, and the alphabetically sorted labels of the occupied Wyckoff positions. As an example, MoS_2 in the H-phase has the crystal type: AB2-187-bi.

2.5. Classification: Magnetic state

In the new version of the C2DB, materials are classified according to their magnetic state as either *non-magnetic* or *magnetic*. A material is considered magnetic if any atom has a local magnetic moments greater than $0.1 \mu_B$.

In the original C2DB, the *magnetic* category was further subdivided into ferromagnetic (FM) and anti-ferromagnetic (AFM). But since the simplest anti-ferromagnetically ordered state typically does not represent the true ground state, all material entries with an AFM state have been removed from the C2DB and replaced by the material in its FM state. Although the latter is less stable, it represents a more well defined state of the material. Crucially, the nearest neighbor exchange couplings for all magnetic materials have been included in the C2DB (see Sec. 5.8). This enables a more detailed and realistic description of the magnetic order via the Heisenberg model. In particular, the FM state of a material is not expected to represent the true magnetic ground if the exchange coupling $J < 0$.

Property	Method	Criteria	Count
Bader charges	PBE	None	3809
Energy above convex hull	PBE	None	4044
Heat of formation	PBE	None	4044
Orbital projected band structure	PBE	None	2487
Out-of-plane dipole	PBE	None	4044
Phonons (Γ and BZ corners)	PBE	None	3865
Projected density of states	PBE	None	3332
Stiffness tensor	PBE	None	3968
Exchange couplings	PBE	Magnetic	538
Infrared polarisability	PBE	$E_{\text{gap}}^{\text{PBE}} > 0$	784
Second harmonic generation	PBE	$E_{\text{gap}}^{\text{PBE}} > 0$, non-magnetic, non-centrosymmetric	375
Electronic band structure PBE	PBE*	None	3496
Magnetic anisotropies	PBE*	Magnetic	823
Deformation potentials	PBE*	$E_{\text{gap}}^{\text{PBE}} > 0$	830
Effective masses	PBE*	$E_{\text{gap}}^{\text{PBE}} > 0$	1272
Fermi surface	PBE*	$E_{\text{gap}}^{\text{PBE}} = 0$	2505
Plasma frequency	PBE*	$E_{\text{gap}}^{\text{PBE}} = 0$	3144
Work function	PBE*	$E_{\text{gap}}^{\text{PBE}} = 0$	4044
Optical polarisability	RPA@PBE	None	3127
Electronic band structure	HSE06@PBE*	None	3155
Electronic band structure	G ₀ W ₀ @PBE*	$E_{\text{gap}}^{\text{PBE}} > 0$, $N_{\text{atoms}} < 5$	357
Born charges	PBE, Berry phase	$E_{\text{gap}}^{\text{PBE}} > 0$	639
Raman spectrum	PBE, LCAO basis set	Non-magnetic, dyn. stable	708
Piezoelectric tensor	PBE, Berry phase	$E_{\text{gap}}^{\text{PBE}}$, non-centrosym.	353
Optical absorbance	BSE@G ₀ W ₀ *	$E_{\text{gap}}^{\text{PBE}} > 0$, $N_{\text{atoms}} < 5$	378
Spontaneous polarisation	PBE, Berry phase	$E_{\text{gap}}^{\text{PBE}} > 0$, nearly centrosym. polar space group	151
Topological invariants	PBE*, Berry phase	$0 < E_{\text{gap}}^{\text{PBE}} < 0.3$ eV	242

Table 1. Properties calculated by the C2DB monolayer workflow. The computational method and the criteria used to decide whether the property should be evaluation for a given material is also shown. A '*' indicates that spin-orbit coupling (SOC) is included. All calculations are performed with the GPAW code using a plane wave basis except for the Raman calculations, which employ a double-zeta polarized (DZP) basis of numerical atomic orbitals.

2.6. Stability: Thermodynamic

The heat of formation, ΔH , of a compound is defined as its energy per atom relative to its constituent elements in their standard states.[46] The

thermodynamic stability of a compound is evaluated in terms of its energy above the *convex hull*, ΔH_{hull} , which gives the energy of the material relative to other competing phases of the same chemical composition, including mixed phases[6], see Fig. 4 for an example.

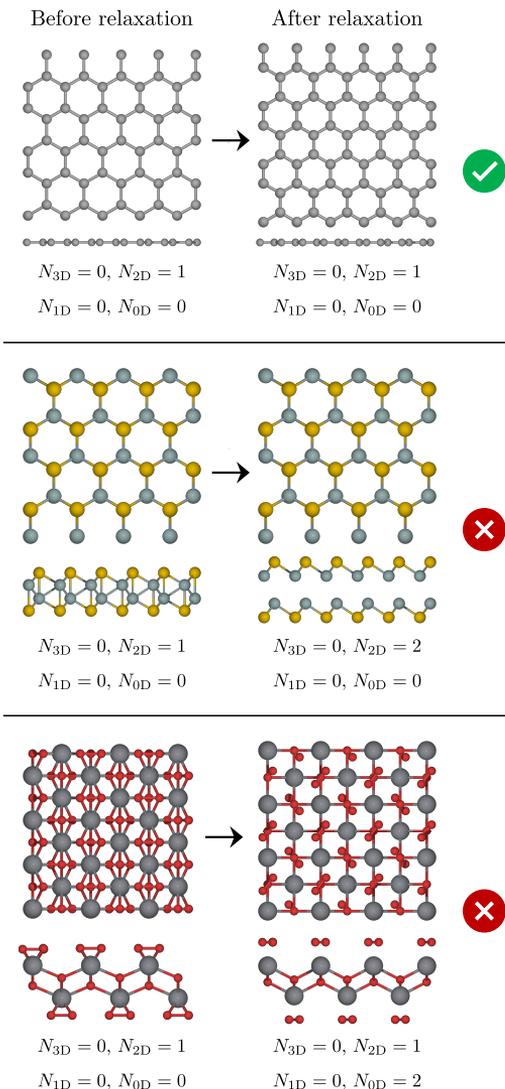


Figure 2. Three example structures from C2DB (top: graphene, middle: Ge_2Se_2 , bottom: Pb_2O_6) with their respective cluster dimensionalities cluster before (left) and after (right) relaxation. The number N_{xD} denotes the number of clusters of dimensionality x .

Clearly, ΔH_{hull} , depends on the pool of reference phases, which in turn defines the convex hull. The original C2DB employed a pool of reference phases comprised by 2807 elemental and binary bulk crystals from the convex hull of the Open Quantum Materials Database (OQMD)[46]. In the new version, this set

Thermodynamic stability indicator	Criterion (eV/atom)
LOW	$\Delta H > 0.2$
MEDIUM	$\Delta H < 0.2$ and $\Delta H_{\text{hull}} > 0.2$
HIGH	$\Delta H < 0.2$ and $\Delta H_{\text{hull}} < 0.2$

Table 2. Thermodynamic stability indicator assigned to all materials in the C2DB.

has been extended by approximately 6783 ternary bulk compounds from the convex hull of OQMD, making a total of 9590 stable bulk reference compounds.

As a simple indicator for the thermodynamic stability of a material, the C2DB employs three labels (low, medium, high) as defined in Table 2.

It should be emphasized that the energies of both monolayers and bulk reference crystals are calculated with the PBE xc-functional. This implies that some inaccuracies must be expected, in particular for materials with strongly localized d -electrons, e.g. certain transition metal oxides, and materials for which dispersive interactions are important, e.g. layered van der Waals crystals. The latter implies that the energy of a monolayer and its layered bulk parent (if such exists in the pool of references) will have the same energy. For further details and discussions see Ref. [6].

2.7. Stability: Dynamical

Dynamically stable materials are situated at a local minimum of the potential energy surface and are thus stable to small structural perturbations. Structures resulting from DFT relaxations can end up in saddle point configurations because of imposed symmetry constraints or an insufficient number of atoms in the unit cell.

In C2DB, the dynamical stability is assessed from the signs of the minimum eigenvalues of (i) the stiffness tensor (see Sec. 3.1) and (ii) the Γ -point Hessian matrix for a supercell containing 2×2 repetitions of the unit cell (the structure is not relaxed in the 2×2 supercell). If one of these minimal eigenvalues is negative the material is classified as dynamically unstable. This indicates that the energy can be reduced by displacing an atom and/or deforming the unit cell, respectively. The use of two categories for dynamical stability, i.e. stable/unstable, differs from the original version of the C2DB where an intermediate category was used for materials with negative but numerically small minimal eigenvalue of either the Hessian or stiffness tensors.

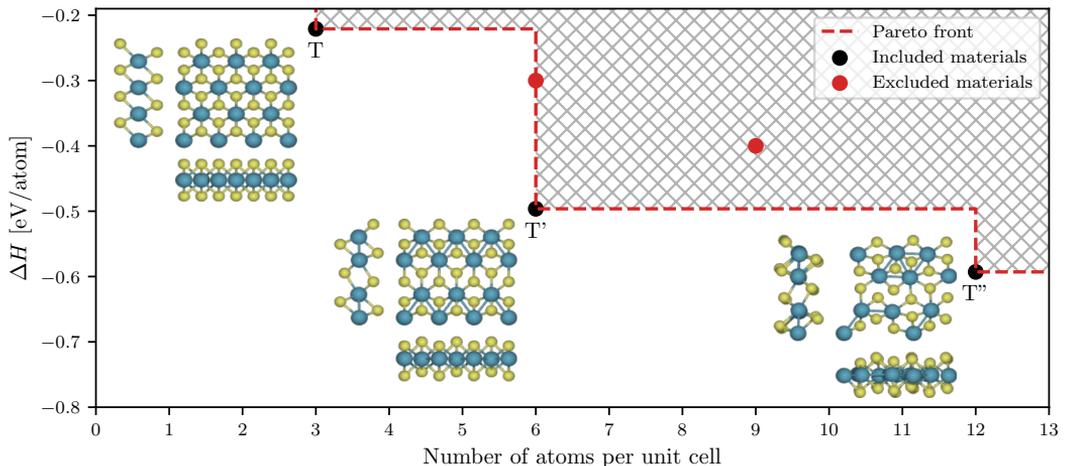


Figure 3. Illustration of the Pareto analysis used to filter out duplicates or irrelevant structures from the C2DB. All points represent materials with the same reduced chemical formula (in this case ReS_2) that belong to the same cluster defined by the structure metric. Only structures lying on the $(N, \Delta H)$ -Pareto front are retained (black circles) while other materials are excluded (red circles). The philosophy behind the algorithm is to keep less stable materials if they contain fewer atoms per unit cell than more stable materials and thus represent structures of higher symmetry.

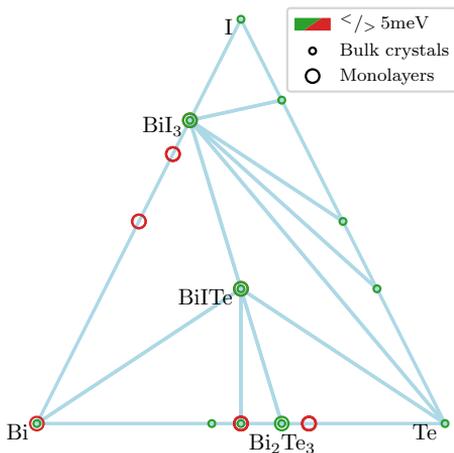


Figure 4. Convex hull diagram for $(\text{Bi}, \text{I}, \text{Te})$ -compounds. Green (red) coloring indicate materials that have a convex hull energy of less than (greater than) 5 meV. The monolayers BiI_3 , Bi_2Te_3 and BiTe lie on the convex hull. The monolayers are degenerate with their layered bulk parent because the vdW interactions are not captured by the PBE xc-functional.

3. Improved property methodology

The new version of the C2DB has been generated using a significantly extended and improved workflow for property evaluations. This section focuses on improvements relating to properties that were already present in the original version of the C2DB while new properties are discussed in the next section.

3.1. Stiffness tensor

The stiffness tensor, C , is a rank-4 tensor that relates the stress of a material to the applied strain. In Mandel notation (a variant of Voigt notation) C is expressed as an $N \times N$ matrix relating the N independent components of the stress and strain tensors. For a 2D material $N = 3$ and the tensor takes the form

$$\mathbf{C} = \begin{bmatrix} C_{xxxx} & C_{xyyy} & \sqrt{2}C_{xxxy} \\ C_{xxyy} & C_{yyyy} & \sqrt{2}C_{yyxy} \\ \sqrt{2}C_{xxxy} & \sqrt{2}C_{yyxy} & 2C_{xyxy} \end{bmatrix} \quad (1)$$

where the indices on the matrix elements refer to the rank-4 tensor. The factors multiplying the tensor elements account for their multiplicities in the full rank-4 tensor. In the C2DB workflow, C is calculated as a finite difference of the stress under an applied strain with full relaxation of atomic coordinates. A negative eigenvalue of C signals a dynamical instability, see Sec. 2.7

In the first version of the C2DB only the diagonal elements of the stiffness tensor were calculated. The new version also determines the shear components such that the full 3×3 stiffness tensor is now available. This improvement also leads to a more accurate assessment of dynamical stability[47].

3.2. Effective masses with parabolicity estimates

For all materials with a finite band gap the effective masses of electrons and holes are calculated for bands within 100 meV of the conduction band minimum (CBM) and valence band maximum (VBM), respectively. The Hessian matrices at the band extrema (BE) are determined by fitting a second order polynomial to the PBE band structure including SOC, and the effective masses are obtained by subsequent diagonalization of the Hessian. The main fitting-procedure is unaltered from the first version of C2DB, but two important improvements have been made.

The first improvement consists in an additional k -mesh refinement step for better localization of the BE in the Brillouin zone. After the location of the BE have been estimated based on a uniformly sampled band structure with k -point density of 12 Å, another one-shot calculation is performed with a denser k -mesh around the estimated BE positions. This ensures a more accurate and robust determination of the location of the BE, which can be important in cases with a small but still significant spin-orbit splitting or when the band is very flat or non-quadratic around the BE. The second refinement step is the same as in the first version of C2DB, i.e. the band energies are calculated on a highly dense k -mesh in a small disc around the BE, and the Hessian is obtained by fitting the band energies in the range up to 1 meV from the band minimum/maximum.

The second improvement is the calculation of the mean absolute relative error (MARE) of the polynomial fit in a 25 meV range from the band minimum/maximum. The value of 25 meV corresponds to the thermal energy at room temperature and is thus the relevant energy scale for many applications. The MARE provides a useful measure of the parabolicity of the energy bands and thus the validity of the effective mass approximation over this energy scale.

Figure 5 shows two examples of band structures with the effective mass fits and corresponding fit errors indicated. Additionally, the distribution of MARE for all the effective mass fits in the C2DB are presented. Most materials have an insignificant MARE, but a few materials have very large errors. Materials with a MARE above a few tens of percentages fall into two classes. For some materials the algorithm does not correctly find the position of the BE. An example is

Ti₂S₂ in the space group C2/m. For others, the fit and BE location are both correct, but the band flattens away from the BE which leads to a large MARE. An example of this latter class is Cl₂Tl₂ in the space group P-1. In general a small MARE indicates a parabolic band while materials with large MARE should be handled on a case-by-case basis.

3.3. Orbital projected band structure

To facilitate a state-specific analysis of the PBE Kohn-Sham wave functions, an orbital projected band structure (PBS) is provided to complement the projected density of states (PDOS). In the PAW methodology, the all-electron wave functions are projected onto atomic orbitals inside the augmentation spheres centered at the position of each atom. The PBS resolves these atomic orbital contributions to the wave functions as a function of band and k -point whereas the PDOS resolves the atomic orbital character of the total density of states as a function of energy. The spin-orbit coupling is not included in the PBS or PDOS, as its effect is separately visualized by the spin-projected band structure also available in the C2DB.

As an example, Figure 6 shows the PBS (left) and PDOS (right) of monolayer MoS₂ calculated with PBE. The relative orbital contribution to a given Bloch state is indicated by a pie chart symbol. In the present example, one can deduce from the PBS that even though Mo- p orbitals and S- p orbitals contribute roughly equally to the DOS in the valence band, the Mo- p orbital contributions are localized to a region in the BZ around the M -point, whereas the S- p orbitals contribute throughout the entire BZ.

3.4. Corrected G_0W_0 band structures

The C2DB contains G_0W_0 quasiparticle (QP) band structures of 370 monolayers covering 14 different crystal structures and 52 chemical elements. The details of these calculations can be found in the original C2DB paper[6]. A recent in-depth analysis of the 61.716 G_0W_0 data points making up the QP band structures led to several important conclusions relevant for high-throughput G_0W_0 calculations. In particular, it identified the linear QP approximation as a significant error source in standard G_0W_0 calculations and proposed an extremely simple correction scheme (the *empirical Z* (empZ) scheme), that reduces this error by a factor of two on average.

The empZ scheme divides the electronic states into two classes according to the size of the QP weight, Z . States with $Z \in [0.5, 1.0]$ are classified as QP consistent (QP-c) while states with $Z \notin [0.5, 1.0]$ are classified as QP inconsistent (QP-ic). With this definition, QP-c states will have at least half of their

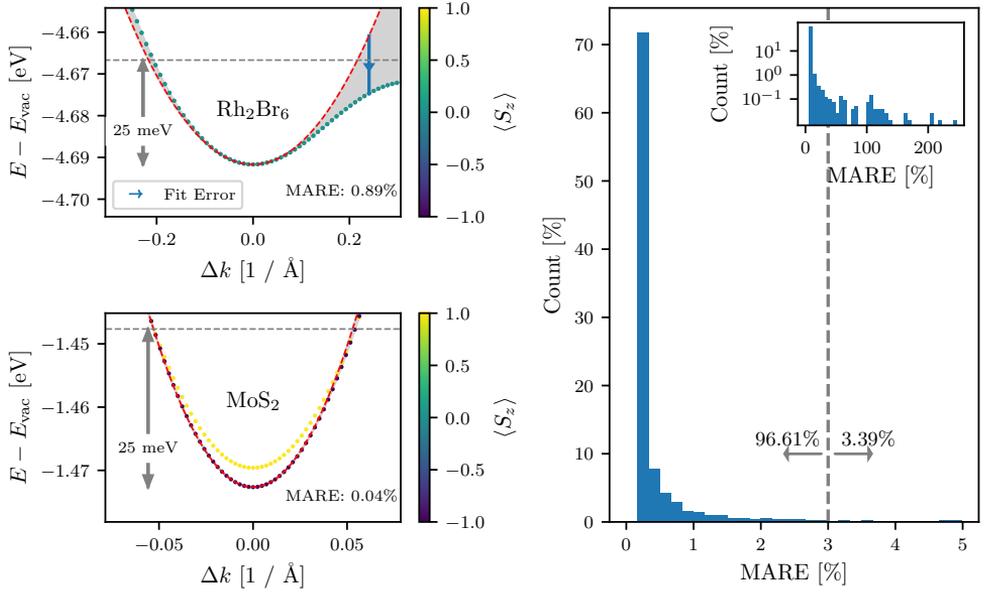


Figure 5. Left: The PBE band structures of Rh₂Br₆ and MoS₂ (colored dots) in regions around the conduction band minimum. The dashed red line shows the fit made to estimate the effective masses of the lowest conduction band. The shaded grey region highlights the error between the fit and the true band structure. The mean absolute relative error (MARE) is calculated for energies within 25 meV of the band minimum. For MoS₂ the fit is essentially on top of the band energies. Right: The distribution of the MARE of all effective mass fits in the C2DB. The inset shows the full distribution on a log scale. As mentioned in the main text, very large MAREs indicate that the band minimum/maximum was incorrectly identified by the algorithm and/or that the band is very flat.

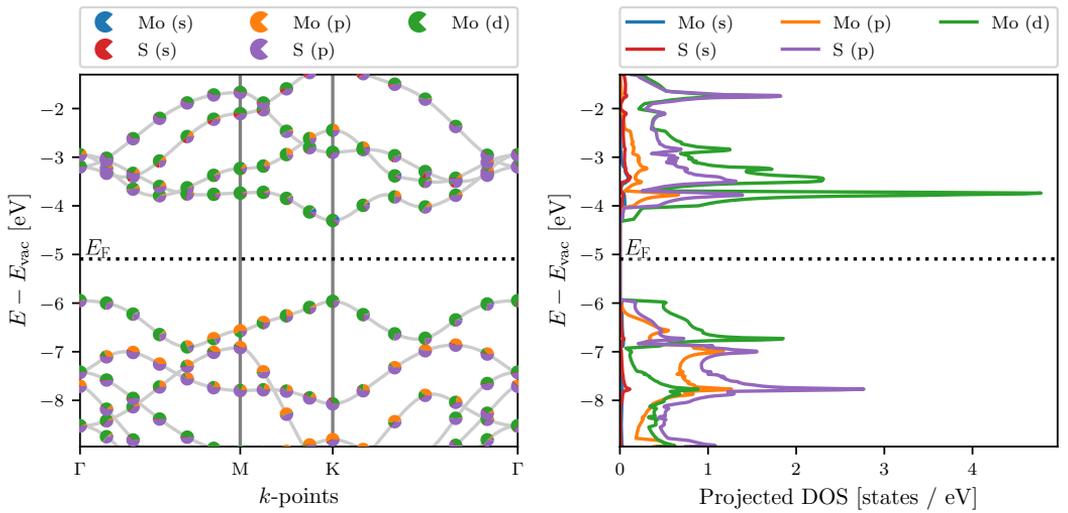


Figure 6. Orbital projected band structure and orbital projected density of states of MoS₂ in the H-phase. The pie chart symbols indicate the fractional atomic orbital character of the Kohn-Sham wave functions.

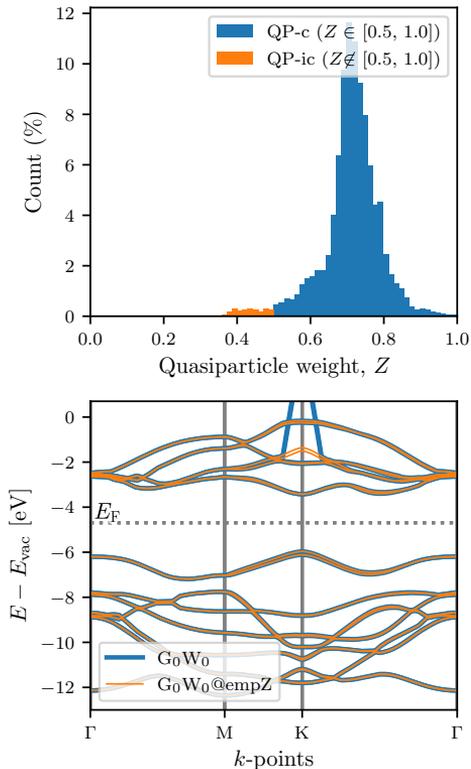


Figure 7. Top: Distribution of the 61716 QP weights (Z) contained in the C2DB. The blue part of the distribution shows QP-consistent (QP-c) Z -values while the orange part shows QP-inconsistent (QP-ic) Z values. In general, the linear expansion of the self-energy performed when solving the QP equation works better for Z closer to 1. About 0.3% of the Z -values lie outside the interval from 0 to 1 and are not included in the distribution. Bottom: G_0W_0 band structure before (blue) and after (orange) applying the empZ correction, which replaces Z by the mean of the distribution for QP-ic states. In the case of MoS_2 only one state at K is QP-ic.

spectral weight in the QP peak. The distribution of the 60.000+ Z -values is shown in Figure 7. It turns out that the linear approximation to the self-energy, which is the gist of the QP approximation, introduces significantly larger errors for QP-ic states than for QP-c states. Consequently, the empZ method replaces the calculated Z of QP-ic states with the mean of the Z -distribution, $Z_0 \approx 0.75$. This simple replacement reduces the average error of the linear approximation from 0.11 eV to 0.06 eV.

An illustration of the method applied to MoS_2 is shown in Figure 7. The original uncorrected G_0W_0 band structure is shown in blue while the empZ corrected band structure is shown in orange. MoS_2

has only one QP-ic state in the third conduction band at the K -point. Due to a break-down of the QP approximation for this state, the G_0W_0 correction is greatly overestimated leading to a local discontinuity in the band structure. The replacement of Z by Z_0 for this particular state resolves the problem. All G_0W_0 band structures in the C2DB are now empZ corrected.

3.5. Optical absorbance

In the first version of the C2DB, the optical absorbance was obtained from the simple expression [6]

$$A(\omega) \approx \frac{\omega \text{Im}\{\alpha^{2D}(\omega)\}}{\epsilon_0 c}, \quad (2)$$

where α^{2D} is the long wavelength limit of the in-plane sheet polarisability density (Note that the equation is written here in SI units). The sheet polarisability is related to the sheet conductivity via $\sigma^{2D}(\omega) = -i\omega\alpha^{2D}(\omega)$. The expression (2) assumes that the electric field inside the layer equals the incoming field (i.e. reflection is ignored), and hence, it may overestimate the absorbance.

In the new version, the absorbance is evaluated from $A = 1 - R - T$, where R and T are the reflected and transmitted powers of a plane wave at normal incidence, respectively. These can be obtained from the conventional transfer matrix method applied to a monolayer suspended in vacuum. The 2D material is here modelled as an infinitely thin layer with a sheet conductivity. Alternatively, it can be modelled as quasi-2D material of thickness d with a “bulk” conductivity of $\sigma = \sigma^{2D}/d$ [48], but the two approaches yield very similar results, since the optical thickness of a 2D material is much smaller than the optical wavelength. Within this model, the expression for the absorbance of a suspended monolayer with the sheet conductivity σ^{2D} reads

$$A(\omega) = \text{Re}\{\sigma^{2D}(\omega)\eta_0\} \left| \frac{2}{2 + \sigma^{2D}(\omega)\eta_0} \right|^2, \quad (3)$$

where $\eta_0 = 1/(\epsilon_0 c) \approx 377 \Omega$ is the vacuum impedance.

If the light-matter interaction is weak, i.e. $|\sigma^{2D}\eta_0| \ll 1$, Eq. (3) reduces to Eq. (2). Nonetheless, due the strong light-matter interaction in some 2D materials, this approximation is not reliable in general. In fact, it can be shown that the maximum possible absorption from Eq. (3) is 50%, which is known as the upper limit of light absorption in thin films [49]. This limit is not guaranteed by Eq. (2), which can even yield an absorbance above 100%.

As an example, Fig. 8 shows the absorption spectrum of monolayer MoS_2 for in- and out-of-plane polarized light as calculated with the exact Eq. (3) and the approximate Eq. (2), respectively. In all cases the

sheet polarisability is obtained from the Bethe-Salpeter Equation (BSE) to account for excitonic effects [6]. For weak light-matter interactions, e.g. for the z -polarized light, the two approaches agree quite well, but noticeable differences are observed in regions with stronger light-matter interaction.

4. New materials in the C2DB

In this section we discuss the most significant extensions of the C2DB in terms of new materials. The set of materials presented here is not complete, but represents the most important and/or well defined classes. The materials discussed in Secs. 4.1 and 4.2 (MXY Janus monolayers and monolayers extracted from experimental crystal structure databases) are already included in the C2DB. The materials described in Secs. 4.3 and 4.4 (homo-bilayers and monolayer point defect systems) will soon become available as separate C2DB-interlinked databases.

4.1. MXY Janus monolayers

The class of transition metal dichalcogenide (TMDC) monolayers of the type MX_2 (where M is the transition metal and X is a chalcogen) exhibits a large variety of interesting and unique properties and has been widely discussed in the literature [50]. Recent experiments have shown that it is not only possible to synthesize different materials by changing the metal M or the chalcogen X, but also by exchanging the X on one side of the layer by another chalcogen (or halogen) [51, 52, 53]. This results in a class of 2D materials known as MXY Janus monolayers with broken mirror symmetry and finite out-of-plane dipole moments. The prototypical MXY crystal structures are shown in Fig. 9 for the case of MoSSe and BiTeI, which have both been experimentally realized [51, 52, 53]. Adopting the nomenclature from the TMDCs, the crystal structures are denoted as H- or T-phase, depending on whether X and Y atoms are vertically aligned or displaced, respectively.

In a recent work [54], the C2DB workflow was employed to scrutinize and classify the basic electronic and optical properties of 224 different MXY Janus monolayers. All data from the study is available in the C2DB. Here we provide a brief discussion of the Rashba physics in these materials and refer the interested reader to Ref. [54] for more details and analysis of other properties.

A key issue when considering hypothetical materials, i.e. materials not previously synthesized, is their stability. The experimentally synthesized MoSSe and BiTeI are both found to be dynamically stable and lie within 10 meV of the convex hull confirming their thermodynamic stability. Out of the 224 initial monolayers

93 are classified as stable according to the C2DB criteria (dynamically stable and $\Delta H_{\text{hull}} < 0.2 \text{ eV/atom}$). Out of the 93 stable materials, 70 exhibits a finite band gap when computed with the PBE xc-functional.

The Rashba effect is a momentum dependent splitting of the band energies of a 2D semiconductor in the vicinity of a band extremum arising due to the combined effect of spin-orbit interactions and a broken crystal symmetry in the direction perpendicular to the 2D plane. The simplest model used to describe the Rashba effect is a 2D electron gas in a perpendicular electric field (along the z -axis). Close to the band extremum, the energy of the two spin bands is described by the Rashba Hamiltonian [55, 56]

$$H = \alpha_R(\boldsymbol{\sigma} \times \mathbf{k}) \cdot \hat{\mathbf{e}}_z, \quad (4)$$

where $\boldsymbol{\sigma}$ is the vector of Pauli matrices, $\mathbf{k} = \mathbf{p}/\hbar$ is the wave number, and the Rashba parameter is proportional to the electric field strength, $\alpha_R \propto E_0$,

Although the Rashba Hamiltonian is only meant as a qualitative model, it is of interest to test its validity on the Janus monolayers. The electric field of the Rashba model is approximately given by $E_0 = \Delta V_{\text{vac}}/d$, where ΔV_{vac} is the shift in vacuum potential on the two sides of the layer (see left inset of Fig. 10) and d is the layer thickness. Assuming a similar thickness for all monolayers, the electric field is proportional to the potential shift. Not unexpected, the latter is found to correlate strongly with the difference in electronegativity of the X and Y atoms, see left panel of Fig. 10.

The Rashba energy, E_R , can be found by fitting $E(k) = \hbar^2 k^2/2m^* \pm \alpha_R k$ to the band structure (see right inset of Fig. 10) and should scale with the electric field strength. However, as seen from the right panel of Fig. 10, there is no correlation between the two quantities. Hence we conclude that the simple Rashba model is completely inadequate and that the strength of the perpendicular electric field cannot be used to quantify the effect of spin-orbit interactions on band energies.

4.2. Monolayers from known layered bulk crystals

The C2DB has been extended with a number of monolayers that are likely exfoliable from experimentally known layered bulk compounds. Specifically, the Inorganic Crystal Structure Database (ICSD)[57] and Crystallography Open Database (COD)[58] have first been filtered for corrupted, duplicate and theoretical compounds, which reduce the initial set of 585.485 database entries to 167.767 unique materials. All of these have subsequently been assigned a "dimensionality score" based on a purely geometrical descriptor. If the 2D score is larger than the sum of 0D, 1D and 3D

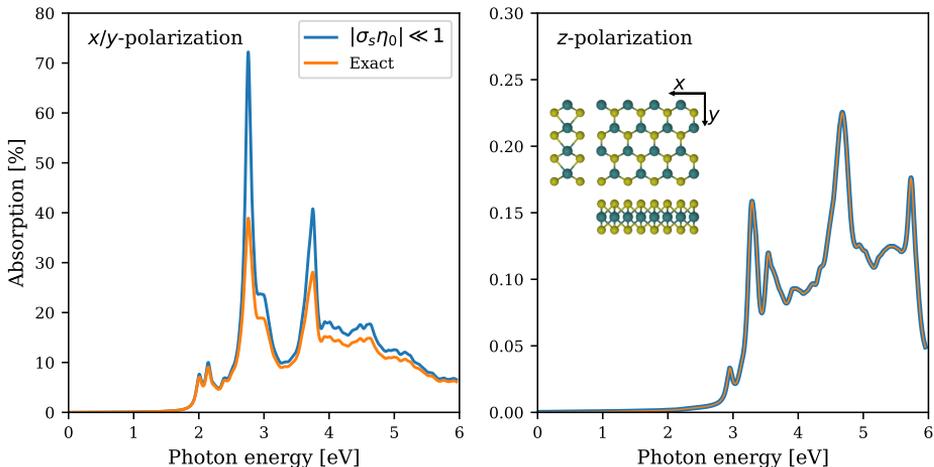


Figure 8. Optical absorption of standalone monolayer MoS₂ for x/y -polarization (left) and z -polarization (right) at normal incident in the BSE framework, obtained using Eq. (2) (blue) or Eq. (3) (orange). The crystal structure cross-sectional views are shown in the inset with the definition of directions.

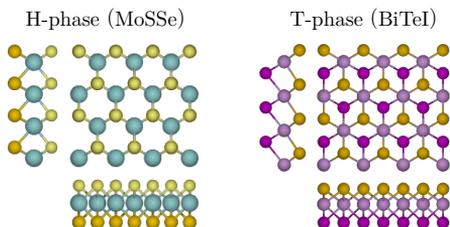


Figure 9. Atomic structure of the MXY Janus monolayers in the H-phase (left) and T-phase (right). The two prototype materials MoSSe and BiTeI are examples of experimentally realized monolayers adopting these crystal structures (not to scale).

scores we regard the material as being exfoliable and we extract the individual 2D components that comprise the material (see also Sec. 2.2). We refer to the original work on the method for details [43] and note that similar approaches were applied in Refs. [12, 11] to identify potentially exfoliable monolayers from the ICSD and COD.

The search has been limited to bulk compounds containing less than 6 different elements and no rare earth elements. This reduces the set of relevant bulk materials to 2991. For all of these we extracted the 2D components containing less than 21 atoms in the unit cell, which were then relaxed and sorted for duplicates following the general C2DB workflow steps described in Secs. 2.1-2.3. At this point 781 materials remain. This set includes most known 2D materials and 207 of

the 781 were already present in the C2DB prior to this addition. All the materials (including those that were already in C2DB) have been assigned an ICSD/COD identifier that refers to the parent bulk compound from which the 2D material was computationally exfoliated. We emphasize that we have not considered exfoliation energies in the analysis and a subset of these materials may thus be rather strongly bound and challenging to exfoliate even if the geometries indicate van der Waals bonded structures of the parent bulk compounds.

Fig. 11 shows the distribution of energies above the convex hull for materials derived from parent structures in ICSD or COD as well as for the entire C2DB, which includes materials obtained from combinatorial lattice decoration as well. As expected, the materials derived from experimental bulk materials are situated rather close to the convex hull whereas those obtained from lattice decoration extend to energies far above the convex hull. It is also observed that a larger fraction of the experimentally derived materials are dynamically stable. There are, however, well known examples of van der Waals bonded structures where the monolayer undergoes a significant lattice distortion, which will manifest itself as a dynamical instability in the present context. For example, bulk MoS₂ exists in van der Waals bonded structures composed of either 2H-MoS₂ or 1T-MoS₂ layers, but a monolayer of the 1T phase undergoes a structural deformation involving a doubling of the unit cell [59] and is thus categorized as dynamically unstable by the C2DB workflow. The dynamically stable materials derived from parent bulk structures

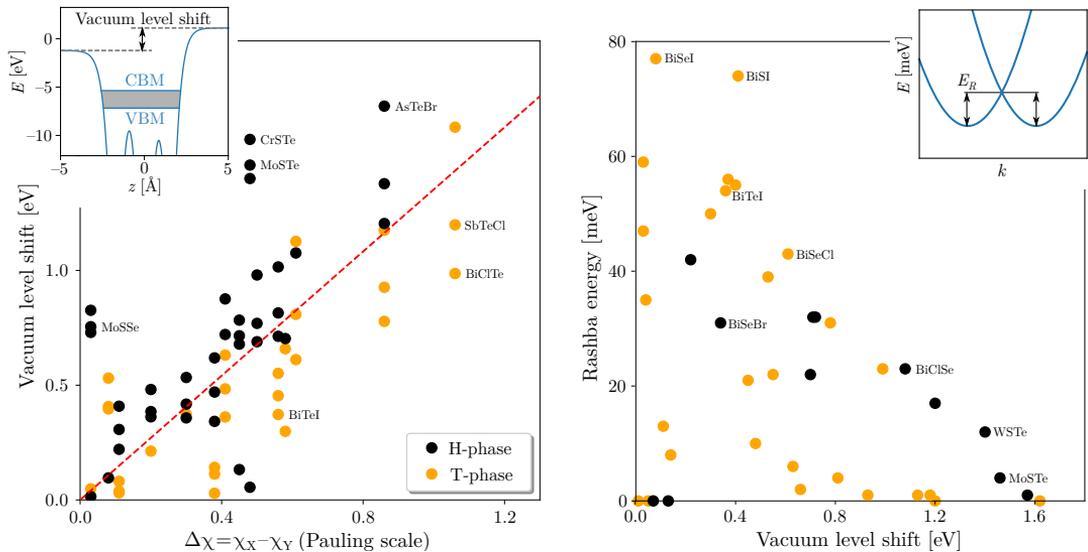


Figure 10. Left: Correlation between the electronegativity difference of X and Y in MX_Y Janus monolayers and the vacuum level shift across the layer. Right: Correlation between the Rashba energy and the vacuum level shift. Structures in the H-phase (e.g. MoSSe) are shown in black while structures in the T-phase (e.g. BiTeI) are shown in orange. The linear fit has a slope 1.35 eV/ $\Delta\chi$ (Pauling scale). The insets show the definition of the vacuum level shift and the Rashba energy, respectively. Modified from [54].

in the ICSD and COD may serve as a useful subset of the C2DB that are likely to be exfoliable from known compounds and thus facilitate experimental verification. As a first application the subset has been used to search for magnetic 2D materials, which resulted in a total of 85 ferromagnets and 61 anti-ferromagnets [60].

4.3. Outlook: Multilayers

The C2DB is concerned with the properties of covalently bonded monolayers (see discussion of dimensionality filtering in Sec. 2.2). However, multilayer structures composed of two or more identical monolayers are equally interesting and often have properties that deviate from those of the monolayer. In fact, the synthesis of layered vdW structures with a controllable number of layers represents an interesting avenue for atomic-scale materials design. Several examples of novel phenomena emerging in layered vdW structures have been demonstrated including direct-indirect band gap transitions in MoS₂[61, 62], layer-parity selective Berry curvatures in few-layer WTe₂[63], thickness-dependent magnetic order in CrI₃[64, 65], and emergent ferroelectricity in bilayer hBN[66].

As a first step towards a systematic exploration of multilayer 2D structures, the C2DB has been used as basis for generating homobilayers in various stacking configurations and subsequently computing

their properties following a modified version of the C2DB monolayer workflow. Specifically, the most stable monolayers (around 1000) are combined into bilayers by applying all possible transformations (unit cell preserving point group operations and translations) of one layer while keeping the other fixed. The candidate bilayers generated in this way are subject to a stability analysis, which evaluates the binding energy and optimal interlayer distance based on PBE-D3 total energy calculations keeping the atoms of the monolayers fixed in their PBE relaxed geometry, see Fig. 12 and Table 3.

The calculated interlayer binding energies are generally in the range from a few to a hundred meV/Å² and interlayer distances range from 1.5Å to 3.8Å. A scatter plot of preliminary binding energies and interlayer distances is shown in Fig. 13. The analysis of homobilayers provides an estimate of the energy required to peel a monolayer off a bulk structure. In particular, the binding energy for the most stable bilayer configuration provides a measure of the *exfoliation energy* of the monolayer. This key quantity is now available for all monolayers in the C2DB, see Sec. 5.1.

4.4. Outlook: Point defects

The C2DB is concerned with the properties of 2D materials in their pristine crystalline form. However, as is well known the perfect crystal is an idealized

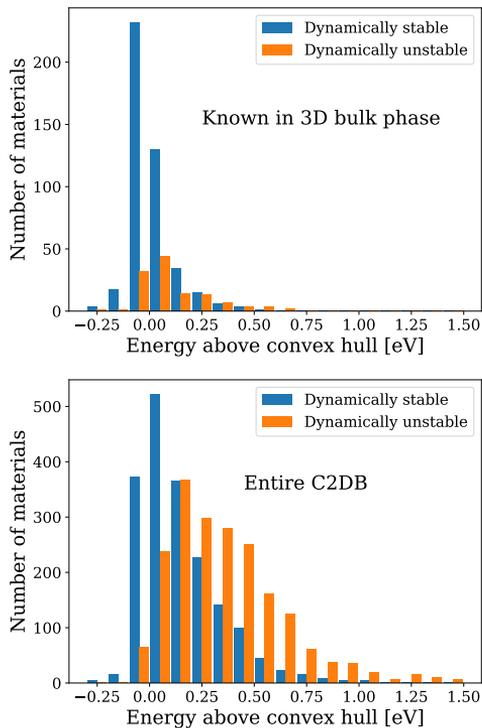


Figure 11. Distribution of energies above the convex hull for the 2D materials extracted from bulk compounds in ICSD and COD (top) and for the entire C2DB including those constructed from combinatorial lattice decoration (bottom). Dynamically stable materials are indicated in blue.

model of real materials, which always contain defects in smaller or larger amounts depending on the intrinsic materials properties and growth conditions. Crystal defects often have a negative impact on physical properties, e.g. they lead to scattering and life time-reduction of charge carriers in semiconductors. However, there are also important situations where defects play a positive enabling role, e.g. in doping of semiconductors, as color centers for photon emission[67, 68] or as active sites in catalysis.

To reduce the gap between the pristine model material and real experimentally accessible samples, a systematic evaluation of the basic properties of the simplest native point defects in a selected subset of monolayers from the C2DB has been initiated. The monolayers are selected based on the stability of the pristine crystal. Moreover, only non-magnetic semiconductors with a PBE band gap satisfying $E_{\text{gap}} > 1$ eV, are currently considered as such materials are candidates for quantum technology applications like single-photon sources and spin

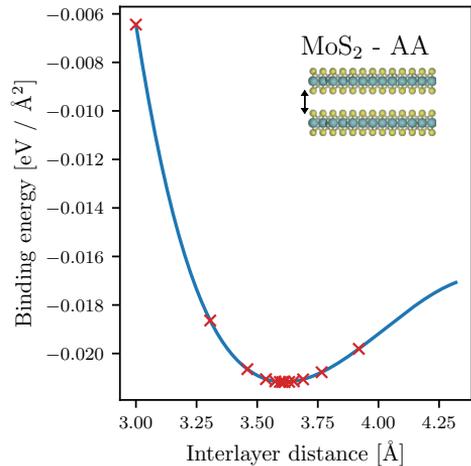


Figure 12. An illustration of the optimization of the interlayer (IL) distance for MoS₂ in the AA stacking. The black crosses are the points sampled by the optimization algorithm while the blue curve is a spline interpolation of the black crosses. The inset shows the MoS₂ AA stacking and the definition of the IL distance is indicated with a black double-sided arrow.

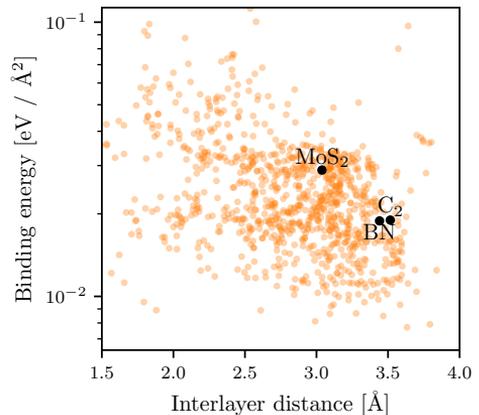


Figure 13. Scatter plot of the calculated interlayer distance and binding energies of (homo)bilayers of selected materials from C2DB. A few well known materials are highlighted: MoS₂, graphene (C₂), and hexagonal boron-nitride (BN). The bilayer binding energies provide an estimate of the monolayer exfoliation energies, see Sec. 5.1

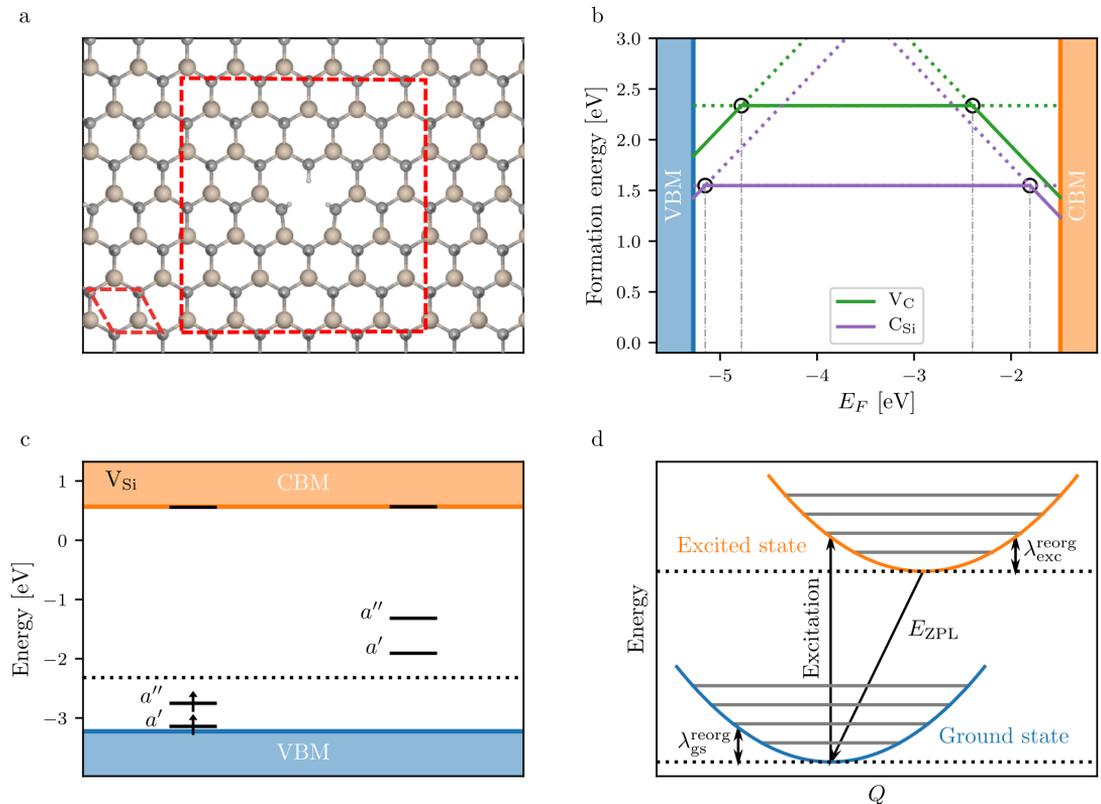


Figure 14. Overview of some of the properties included in the 2D defect database project for the example host material CH_2Si . (a) The supercell used to represent the defects (here a Si vacancy). The supercell is deliberately chosen to break the symmetry of the host crystal lattice. (b) Formation energies of a C vacancy (green) and C-Si substitutional defect (purple). (c) Energy and orbital symmetry of the localized single-particle states of the V_{Si} defect for both spin channels (left and right). The Fermi level is shown by the dotted line. (d) Schematic excited state configuration energy diagram. The transitions corresponding to the vertical absorption and the zero-phonon emission are indicated.

qubits. Following these selection criteria around 300 monolayers are identified and their vacancies and intrinsic substitutional defects are considered, yielding a total of about 1500 defect systems.

Each defect system is subject to the same workflow, which is briefly outlined below. To enable point defects to relax into their lowest energy configuration, the symmetry of the pristine host crystal is intentionally broken by the chosen supercell, see Fig. 14 (a). In order to minimize defect-defect interaction, supercells are furthermore chosen such that the minimum distance between periodic images of defects is larger than 15 Å. Unique point defects are created based on the analysis of equivalent Wyckoff positions for the host material. To illustrate some of the properties that will feature in the upcoming point defect database, we consider the specific example of monolayer CH_2Si .

First, the formation energy[69, 70] of a given defect is calculated from PBE total energies. Next, Slater-Janak transition state theory is used to obtain the charge transition levels[71, 72]. By combining these results, one obtains the formation energy of the defect in all possible charge states as a function of the Fermi level. An example of such a diagram is shown in Fig. 14 (b) for the case of the V_C and C_{Si} defects in monolayer CH_2Si . For each defect and each charge state, the PBE single-particle energy level diagram is calculated to provide a qualitative overview of the electronic structure. A symmetry analysis[73] is performed for the defect structure and the individual defect states lying inside the band gap. The energy level diagram of the neutral V_{Si} defect in CH_2Si is shown in Fig. 14 (c), where the defect states are labeled according to the irreducible representations of the C_s point group.

In general, excited electronic states can be modelled by solving the Kohn-Sham equations with non-Aufbau occupations. The excited-state solutions are saddle points of the Kohn-Sham energy functional, but common self-consistent field (SCF) approaches often struggle to find such solutions, especially when nearly degenerate states are involved. The calculation of excited states corresponding to transitions between localized states inside the band gap is therefore performed using an alternative method based on the direct optimization (DO) of orbital rotations in combination with the maximum overlap method (MOM) [74]. This method ensures fast and robust convergence of the excited states, as compared to SCF. In Fig. 14 (d), the reorganization energies for the ground- and excited state, as well as the zero-phonon line (ZPL) energy are sketched. For the specific case of the Si vacancy in CH_2Si , the DO-MOM method yields $E_{\text{ZPL}} = 3.84$ eV, $\lambda_{\text{gs}}^{\text{reorg}} = 0.11$ eV and $\lambda_{\text{exc}}^{\text{reorg}} = 0.16$ eV. For systems with large electron-phonon coupling (i.e. Huang-Rhys factor > 1) a one-dimensional approximation for displacements along the main phonon mode is used to produce the configuration coordinate diagram (see Fig. 14 (d)). In addition to the ZPL energies and reorganization energies, the Huang-Rhys factors, photoluminescence spectrum from the 1D phonon model, hyperfine coupling and zero field splitting are calculated.

5. New properties in the C2DB

This section reports on new properties that have become available in the C2DB since the first release. The employed computational methodology is described in some detail and results are compared to the literature where relevant. In addition, some interesting property correlations are considered along with general discussions of the general significance and potential application of the available data.

5.1. Exfoliation energy

The exfoliation energy of a monolayer is estimated as the binding energy of its bilayer in the most stable stacking configuration (see also Sec. 4.3). The binding energy is calculated using the PBE+D3 xc-functional [75] with the atoms of both monolayers fixed in the PBE relaxed geometry. Table 3 compares exfoliation energies obtained in this way to values from Mounet *et al.* [11] for a representative set of monolayers.

5.2. Bader charges

For all monolayers we calculate the net charge on the individual atoms using the Bader partitioning scheme [76]. The analysis is based purely on the

Material	SG	PBE+D3	DF2	rVV10
MoS ₂	P-6m2	28.9	21.6	28.8
MoTe ₂	P-6m2	30.3	25.2	30.4
ZrNBr	Pmmm	18.5	10.5	18.5
C	P6/mmm	18.9	20.3	25.5
P	Pmna	21.9	38.4	30.7
BN	P-6m2	18.9	19.4	24.4
WTe ₂	P-6m2	32.0	24.7	30.0
PbTe	P3m1	23.2	27.5	33.0

Table 3. Exfoliation energies for selected materials calculated with the PBE+D3 xc-functional as described in Sec. 4.3 and compared with the DF2 and rVV10 results from Ref. [11]. The spacegroups are indicated in the column "SG". All numbers are in units of meV/Å².

electron density, which we calculate from the PAW pseudo density plus compensation charges using the PBE xc-functional. Details of the method and its implementation can be found in Tang *et al.* [77]. In Sec. 5.4 we compare and discuss the relation between Bader charges and Born charges.

5.3. Spontaneous polarization

The spontaneous polarization (\mathbf{P}_s) of a bulk material is defined as the charge displacement with respect to that of a reference centrosymmetric structure [78, 79]. Ferroelectric materials exhibit a finite value of \mathbf{P}_s that may be switched by an applied external field and have attracted a large interest for a wide range of applications [80, 81, 82].

The spontaneous polarization in bulk materials can be regarded as electric dipole moment per unit volume, but in contrast to the case of finite systems this quantity is ill-defined for periodic crystals [78]. Nevertheless, one can define the formal polarization density

$$\mathbf{P} = \frac{1}{2\pi} \frac{e}{V} \sum_l \phi_l \mathbf{a}_l \quad (5)$$

where \mathbf{a}_l (with $l \in \{1, 2, 3\}$) are the lattice vectors spanning the unit cell, V is the cell volume and e is the elementary charge. ϕ_l is the polarization phase along the lattice vector defined by

$$\phi_l = \sum_i Z_i \mathbf{b}_l \cdot \mathbf{u}_i - \phi_l^{\text{elec}} \quad (6)$$

where \mathbf{b}_l is the reciprocal lattice vector satisfying $\mathbf{b}_l \cdot \mathbf{R}_l = 2\pi$ and \mathbf{u}_i is the position of nucleus i

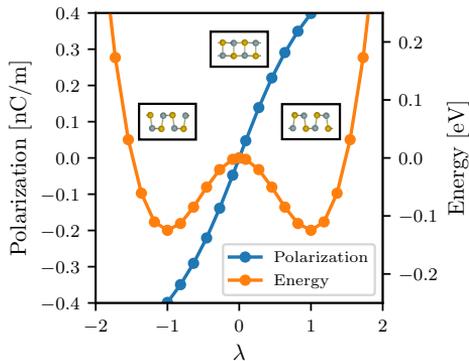


Figure 15. Depicted in the blue plot is the formal polarization calculated along the adiabatic path for GeSe, using the methods described in the main text. The orange plot shows the energy potential along the path as well as outside. Figure inset: The structure of GeSe in the 2 non-centrosymmetric configurations corresponding to $-\mathbf{P}_s$ and \mathbf{P}_s and the centrosymmetric configuration.

with charge eZ_i . The electronic contribution to the polarization phase is defined as

$$\phi_l^{\text{elec}} = \frac{1}{N_{k \perp \mathbf{b}_l}} \text{Im} \sum_{\mathbf{k} \in \text{BZ}_{\perp \mathbf{b}_l}} \times \ln \prod_{j=0}^{N_{k \parallel \mathbf{b}_l} - 1} \det_{\text{occ}} \left[\left\langle u_{n\mathbf{k}+j\delta\mathbf{k}} \left| u_{m\mathbf{k}+(j+1)\delta\mathbf{k}} \right\rangle \right], \quad (7)$$

where $\text{BZ}_{\perp \mathbf{b}_l} = \{\mathbf{k} | \mathbf{k} \cdot \mathbf{b}_l = 0\}$ is a plane of \mathbf{k} -points orthogonal to \mathbf{b}_l , $\delta\mathbf{k}$ is the distance between neighbouring \mathbf{k} -points in the \mathbf{b}_l direction and $N_{k \parallel \mathbf{b}_l}$ ($N_{k \perp \mathbf{b}_l}$) is the number of \mathbf{k} -points along (perpendicular to) the \mathbf{b}_l direction. These expression generalize straightforwardly to 2D.

The formal polarization is only well-defined modulo $e\mathbf{R}_n/V$ where \mathbf{R}_n is any lattice vector. However, changes in polarization are well defined and the spontaneous polarization may thus be obtained by

$$\mathbf{P}_s = \int_0^1 \frac{d\mathbf{P}(\lambda)}{d\lambda} d\lambda, \quad (8)$$

where λ is a dimensionless parameter that defines an adiabatic structural path connecting the polar phase ($\lambda = 1$) with a non-polar phase ($\lambda = 0$).

The methodology has been implemented in GPAW and used to calculate the spontaneous polarization of all stable materials in the C2DB with a PBE band gap above 0.01 eV and a polar space group symmetry. For each material, the centrosymmetric phase with smallest atomic displacement from the

polar phase is constructed and relaxed under the constraint of inversion symmetry. The adiabatic path connecting the two phases is then used to calculate the spontaneous polarization using Eqs. (5)-(8). An example of a calculation for GeSe is shown in Fig. 15 where the polarization along the path connecting two equivalent polar phases via the centrosymmetric phase is shown together with the total energy. The spontaneous polarization obtained from the path is 39.8 nC/m in good agreement with previous calculations [83].

5.4. Born charges

The Born charge of an atom a at position \mathbf{u}_a in a solid is defined as

$$Z_{ij}^a = \frac{V}{e} \frac{\partial P_i}{\partial u_{aj}} \Big|_{E=0}. \quad (9)$$

It can be understood as an effective charge assigned to the atom to match the change in polarization in direction i when its position is perturbed in direction j . Since the polarization density and the atomic position are both vectors, the Born charge of an atom is a rank-2 tensor. The Born charge is calculated as a finite difference and relies on the Modern theory of polarization [84] for the calculation of polarization densities, see Ref. [85] for more details. The Born charge has been calculated for all stable materials in C2DB with a finite PBE band gap.

It is of interest to examine the relation between the Born charge and the Bader charge (see Sec. 5.2). In materials with strong ionic bonds one would expect the charges to follow the atoms. On the other hand, in covalently bonded materials the hybridization pattern and thus the charge distribution, depends on the atom positions in a complex way, and the idea of charges following the atom is expected to break down. In agreement with this idea, the (in-plane) Born charges in the strongly ionic hexagonal boron-nitride ($\pm 2.71e$ for B and N, respectively) are in good agreement with the calculated Bader charges ($\pm 3.0e$). In contrast, (the in-plane) Born charges in MoS₂ ($-1.08e$ and $0.54e$ for Mo and S, respectively) deviate significantly from the Bader charges ($1.22e$ and $-0.61e$ for Mo and S, respectively). In fact, the values disagree even on the sign of the charges underlining the non-intuitive nature of the Born charges in covalently bonded materials.

Note that the out-of-plane Born charges never match the Bader charges, even for strongly ionic insulators, and are consistently smaller in value than the in-plane components. The smaller out-of-plane values are consistent with the generally smaller out-of-plane polarisability of 2D materials (for both electronic and phonon contributions) and agrees with

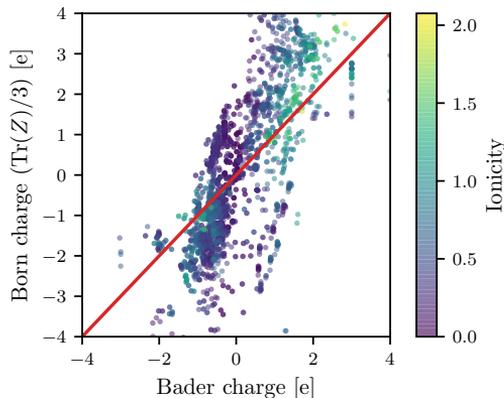


Figure 16. Born charges, $\text{Tr}(Z)/3$, vs. Bader charges for 3025 atoms in the 585 materials for which the Born charges are calculated. The colors indicate the ionicity of the atoms (see main text).

the intuitive expectation that it is more difficult to polarize a 2D material in the out-of-plane direction as compared to the in-plane direction.

Fig. 16 shows the average of the diagonal of the Born charge tensor, $\text{Tr}(Z^a)/3$, plotted against the Bader charges for all 585 materials in the C2DB for which the Born charges have been computed. The data points have been colored according to the ionicity of the atom a defined as $I(a) = |\chi_a - \langle\chi\rangle|$, where χ_a and $\langle\chi\rangle$ are the Pauling electronegativity of atom a and the average electronegativity of all atoms in the unit cell, respectively. The ionicity is thus a measure of the tendency of an atom to donate/accept charge relative to the average tendency of atoms in the material. It is clear from Fig. 16 that there is a larger propensity for the Born and Bader charges to match in materials with higher ionicity.

Fig. 17 plots the average (in-plane) Born charge and the Bader charge versus the band gap. It is clear that large band gap materials typically exhibit integer Bader charges, whereas there is no clear correlation between the Born charge and the band gap.

5.5. Infrared polarizability

The original C2DB provided the frequency dependent polarisability computed in the random phase approximation (RPA) with inclusion of electronic interband and intraband (for metals) transitions[6]. However, phonons carrying a dipole moment (so-called infrared (IR) active phonons) also contribute to the polarizability at frequencies comparable to the frequency of optical phonons. This response is described by the IR polarizability,

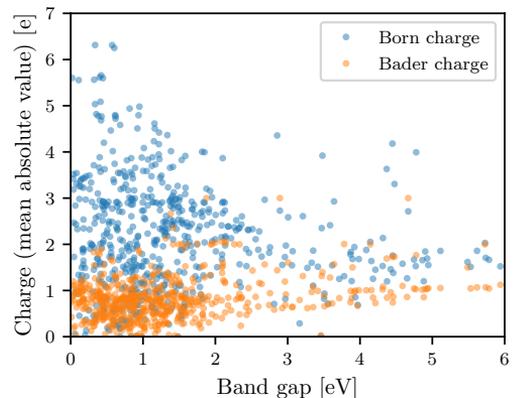


Figure 17. Bader and in-plane Born charges vs. band gap.

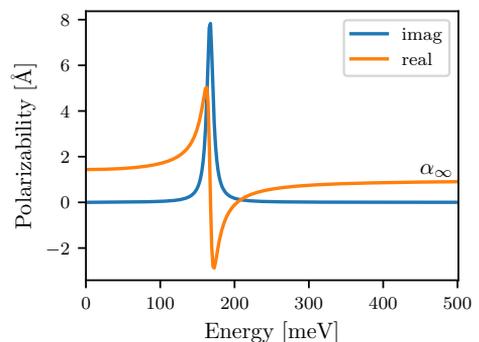


Figure 18. Total polarizability, including both electrons and phonons, of monolayer hBN in the infrared frequency regime. The resonance at around 180 meV is due to the Γ -point longitudinal optical phonon. At energies above all phonon frequencies (but below the band gap) the polarizability is approximately constant and equal to the static limit of the electronic polarizability, α_∞ .

$$\alpha^{\text{IR}}(\omega) = \frac{e^2}{A} \mathbf{Z}^T \mathbf{M}^{-1/2} \left(\sum_i \frac{\mathbf{d}_i \mathbf{d}_i^T}{\omega_i^2 - \omega^2 - i\gamma\omega} \right) \mathbf{M}^{-1/2} \mathbf{Z}, \quad (10)$$

where \mathbf{Z} and \mathbf{M} are matrix representations of the Born charges and atomic masses, ω_i^2 and \mathbf{d}_i are eigenvectors and eigenvalues of the dynamical matrix, A is the in-plane cell area and γ is a broadening parameter representing the phonon lifetime and is set to 10 meV. The total polarizability is then the sum of the electronic polarizability and the IR polarizability.

The new C2DB includes the IR polarisability of all monolayers for which the Born charges have been

calculated (stable materials with a finite band gap), see Sec. (5.4). As an example, Fig. 18 shows the total polarizability of monolayer hexagonal boron nitride. For details on the calculation of the IR polarizability see Ref. [85].

5.6. Piezoelectric tensor

The piezoelectric effect is the accumulation of charges, or equivalently the formation of an electric polarisation, in a material in response to an applied mechanical stress or strain. It is an important material characteristic with numerous scientific and technological applications in sonar, microphones, accelerometers, ultrasonic transducers, energy conversion etc [86, 87]. The change in polarization originates from the movement of positive and negative charge centers as the material is deformed.

Piezoelectricity can be described by the (proper) piezoelectric tensor c_{ijk} with $i, j, k \in \{x, y, z\}$, given by [88]

$$c_{ijk} = \frac{e}{2\pi V} \sum_l \frac{\partial \phi_l}{\partial \epsilon_{jk}} a_{li}. \quad (11)$$

which differs from Eq. (5) only by a derivative of the polarization phase with respect to the strain tensor ϵ_{jk} . Note that c_{ijk} does not depend on the chosen branch cut.

The piezoelectric tensor is a symmetric tensor with at most 18 independent components. Furthermore, the point group symmetry restricts the number of independent tensor elements and their relationships due to the well-known Neumann's principle [89]. For example, monolayer MoS₂ with point group D_{3h} , has only one non-vanishing independent element of c_{ijk} . Note that c_{ijk} vanishes identically for centrosymmetric materials. Using a finite-difference technique with a finite but small strain (1% in our case), Eq. (11) has been used to compute the proper piezoelectric tensor for all non-centrosymmetric materials in the C2DB with a finite band gap. Table 4 shows a comparison of the piezoelectric tensors in the C2DB with literature for a selected set of monolayer materials. Good agreement is obtained for all these materials.

5.7. Topological invariants

For all materials in the C2DB exhibiting a direct band gap below 1 eV, the k -space Berry phase spectrum of the occupied bands has been calculated from the PBE wave functions. Specifically, a particular k -point is written as $k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2$ and the Berry phases $\gamma_n(k_2)$ of the occupied states on the path $k_1 = 0 \rightarrow k_1 = 1$ is calculated for each value of k_2 . The connectivity of the Berry phase spectrum determines the topological properties of the 2D Bloch Hamiltonian [92, 93].

Material	Exp.	Theory [90]	C2DB
BN	-	0.14	0.13
MoS ₂	0.3	0.36	0.35
MoSe ₂	-	0.39	0.38
MoTe ₂	-	0.54	0.48
WS ₂	-	0.25	0.24
WSe ₂	-	0.27	0.26
WTe ₂	-	0.34	0.34

Table 4. Comparison of computed piezoelectric tensor versus experimental values and previous calculations for hexagonal BN and a selected set of TMDs (space group 187). All number are in units of nC/m. Experimental data for MoS₂ is obtained from Ref. [91].

The calculated Berry phase spectra of the relevant materials are available for visual inspection on the C2DB webpage. Three different topological invariants have been extracted from these spectra and are reported in the C2DB: 1) The Chern number, C , takes an integer value and is well defined for any gapped 2D material. It determines the number of chiral edge states on any edge of the material. For any non-magnetic material the Chern number vanishes due to time-reversal symmetry. It is determined from the Berry phase spectrum as the number of crossings at any horizontal line in the spectrum. 2) The mirror Chern number, C_M , defined for gapped materials with a mirror plane in the atomic layer[94]. For such materials, all states may be chosen as mirror eigenstates with eigenvalues $\pm i$ and the Chern numbers C_{\pm} can be defined for each mirror sector separately. For a material with vanishing Chern number, the mirror Chern number is defined as $C_M = (C_+ - C_-)/2$ and takes an integer value corresponding to the number of edge states on any mirror symmetry preserving edge. It is obtained from the Berry phase spectrum as the number of chiral crossings in each of the mirror sectors. 3) The Z_2 invariant, ν , which can take the values 0 and 1, is defined for materials with time-reversal symmetry. Materials with $\nu = 1$ are referred to as quantum spin Hall insulators and exhibit helical edge states at any time-reversal conserving edge. It is determined from the Berry phase spectrum as the number of crossing points modulus 2 at any horizontal line in the interval $k_2 \in [0, 1/2]$.

Figure 19 shows four representative Berry phase spectra corresponding to the three cases of non-vanishing C , C_M and ν as well as a trivial insulator. The four materials are: OsCl₃ (space group 147) - a Chern insulator with $C = 1$, OsTe₂ (space group 14) - a mirror crystalline insulator with $C_M = 2$, SbI

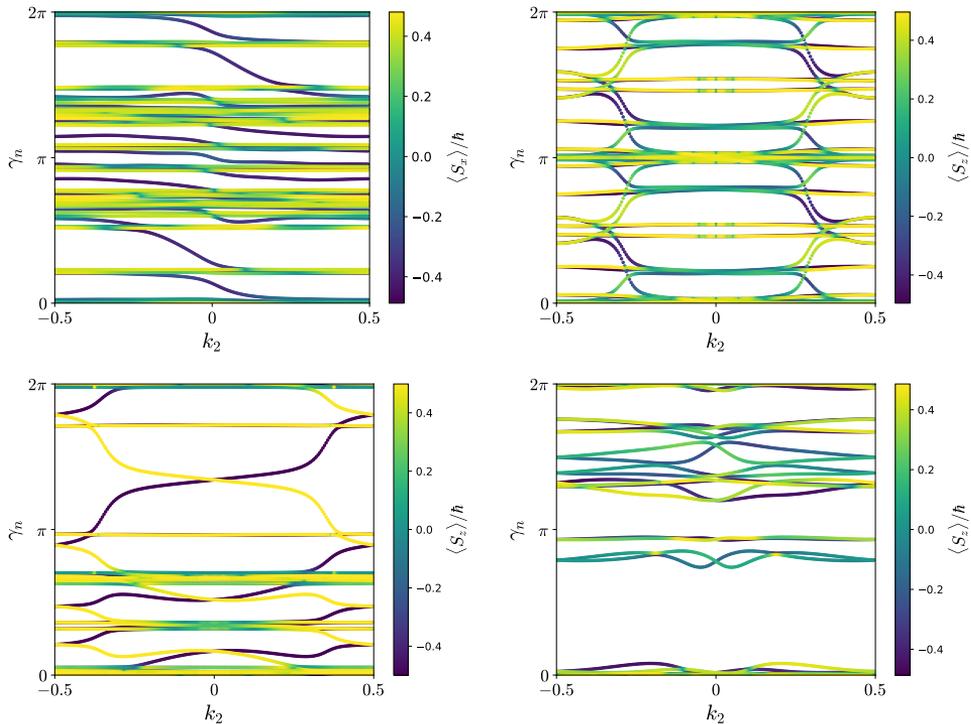


Figure 19. Berry phase spectra of the Chern insulator OsCl₃ (top left), the crystalline topological insulator OsTe₂ (top right), the quantum spin Hall insulator Sbi (lower left) and the trivial insulator BiTe (lower right).

(spacegroup 1) - a quantum spin Hall insulator with $\nu = 1$ and BiTe (spacegroup 156) - a trivial insulator. Note that a gap in the Berry phase spectrum always implies a trivial insulator.

In Ref. [95] the C2DB was screened for materials with non-trivial topology. At that point it was found that the database contained 7 Chern insulators, 21 mirror crystalline topological insulators and 48 quantum spin Hall insulators. However, that does not completely exhaust the the topological properties of materials in the C2DB. In particular, there may be materials that can be topologically classified based on crystalline symmetries other than the mirror plane of the layer. In addition, second order topological effects may be present in certain materials, which imply that flakes will exhibit topologically protected corner states. Again, the Berry phase spectra may be used to unravel the second order topology by means of nested Wilson loops [96].

5.8. Exchange coupling constants

The general C2DB workflow described in Secs. 2.1-2.3 will identify the ferromagnetic ground state of a

material and apply it as starting point for subsequent property calculations, whenever it is more stable than the spin-paired ground state. In reality, however, the ferromagnetic state is not guaranteed to comprise the magnetic ground state. In fact, anti-ferromagnetic states often have lower energy than the ferromagnetic one, but in general it is non-trivial to obtain the true magnetic ground state. We have chosen to focus on the ferromagnetic state due to its simplicity and because its atomic structure and stability are often very similar to those of other magnetic states. Whether or not the ferromagnetic state is the true magnetic ground state is indicated by the nearest neighbor exchange coupling constant as described below.

When investigating magnetic materials the thermodynamical properties (for example the critical temperatures for ordering) are of crucial interest. In two dimensions the Mermin-Wagner theorem[97] comprises an extreme example of the importance of thermal effects since it implies that magnetic order is only possible at $T = 0$ unless the spin-rotational symmetry is explicitly broken. The thermodynamic properties cannot be accessed directly by DFT. Consequently, magnetic models that capture the crucial features of mag-

netic interactions must be employed. For insulators, the Heisenberg model has proven highly successful in describing magnetic properties of solids in 3D as well as 2D[98]. It represents the magnetic degrees of freedom as a lattice of localized spins that interact through a set of exchange coupling constants. If the model is restricted to include only nearest neighbor exchange and assume magnetic isotropy in the plane, it reads

$$H = -\frac{J}{2} \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j - \frac{\lambda}{2} \sum_{\langle ij \rangle} S_i^z S_j^z - A \sum_i (S_i^z)^2 \quad (12)$$

where J is the nearest neighbor exchange constant, λ is the nearest neighbor anisotropic exchange constant and A measures the strength of single-ion anisotropy. We also neglect off-diagonal exchange coupling constants that give rise to terms proportional to $S_i^x S_j^y$, $S_i^y S_j^x$ and $S_i^z S_j^x$. The out-of-plane direction has been chosen as z and $\langle ij \rangle$ implies that for each site i we sum over all nearest neighbor sites j . The parameters J , λ and A may be obtained from an energy mapping analysis involving four DFT calculations with different spin configurations[99, 60, 100]. The thermodynamic properties of the resulting "first principles Heisenberg model" may subsequently be analysed with classical Monte Carlo simulations or renormalized spin wave theory [101, 36].

The C2DB provides the values of J , λ , and A as well as the number of nearest neighbors N_{nn} and the maximum eigenvalue of S_z (S), which is obtained from the total magnetic moment per atom in the ferromagnetic ground state (rounded to nearest half-integer for metals). These key parameters facilitate easy post-processing analysis of thermal effects on the magnetic structure. In Ref. [102] such an analysis was applied to estimate the critical temperature of all ferromagnetic materials in the C2DB based on a model expression for T_C and the parameters from Eq. (12).

For metals, the Heisenberg parameters available in C2DB should be used with care because the Heisenberg model is not expected to provide an accurate description of magnetic interactions in this case. Nevertheless, even for metals the sign and magnitude of the parameters provide an important qualitative measure of the magnetic interactions that may be used to screen and select materials for more detailed investigations of magnetic properties.

A negative value of J implies the existence of an anti-ferromagnetic state with lower energy than the ferromagnetic state used in C2DB. This parameter is thus crucial to consider when judging the stability and relevance of a material classified as magnetic in C2DB (see Sec. 2.5). Fig. 20 shows the distribution of exchange coupling constants (weighted by S^2) of the magnetic materials in the C2DB. The distribution is slightly skewed to the positive side

indicating that ferromagnetic order is more common than anti-ferromagnetic order.

The origin of magnetic anisotropy may stem from either single-ion anisotropy or anisotropic exchange and it is in general difficult *a priori* to determine, which mechanism is most important. There is, however, a tendency in the literature to neglect anisotropic exchange terms in a Heisenberg model description of magnetism and focus solely on the single-ion anisotropy. In Fig. 20 we show a scatter plot of the anisotropy parameters A and λ for the ferromagnetic materials ($J > 0$). The spread of the parameters indicate that the magnetic anisotropy is in general equally likely to originate from both mechanisms and neglecting anisotropic exchange is not advisable. For ferromagnets, the model (Eq. (12)) only exhibits magnetic order at finite temperatures if $A(2S - 1) + \lambda N_{nn} > 0$ [102]. Neglecting anisotropic exchange thus excludes materials with $A < 0$ that satisfies $A(2S - 1) + \lambda N_{nn} > 0$. This is in fact the case for 11 ferromagnetic insulators and 31 ferromagnetic metals in the C2DB.

5.9. Raman spectrum

Raman spectroscopy is an important technique used to probe the vibrational modes of a solid (or molecule) by means of inelastic scattering of light [103]. In fact, Raman spectroscopy is the dominant method for characterising 2D materials and can yield detailed information about chemical composition, crystal structure and layer thickness. There exist several different types of Raman spectroscopies that differ mainly by the number of photons and phonons involved in the scattering process [103]. The first-order Raman process, in which only a single phonon is involved, is the dominant scattering process in samples with low defect concentrations.

In a recent work, the first-order Raman spectra of 733 monolayer materials from the C2DB were calculated, and used as the basis for an automatic procedure for identifying a 2D material entirely from its experimental Raman spectrum[104]. The Raman spectrum is calculated using third-order perturbation theory to obtain the rate of scattering processes involving creation/annihilation of one phonon and two photons, see Ref. [104] for details. The light field is written as $\mathcal{F}(t) = \mathcal{F}_{in} \mathbf{u}_{in} \exp(-i\omega_{in}t) + \mathcal{F}_{out} \mathbf{u}_{out} \exp(-i\omega_{out}t) + c.c.$, where $\mathcal{F}_{in/out}$ and $\omega_{in/out}$ denote the amplitudes and frequencies of the input/output electromagnetic fields, respectively. In addition, $\mathbf{u}_{in/out} = \sum_i u_{in/out}^i \mathbf{e}_i$ are the corresponding polarization vectors, where \mathbf{e}_i denotes the unit vector along the i -direction with $i \in \{x, y, z\}$. Using this light field, the final expression for the Stokes Raman intensity involving scattering events by only one phonon

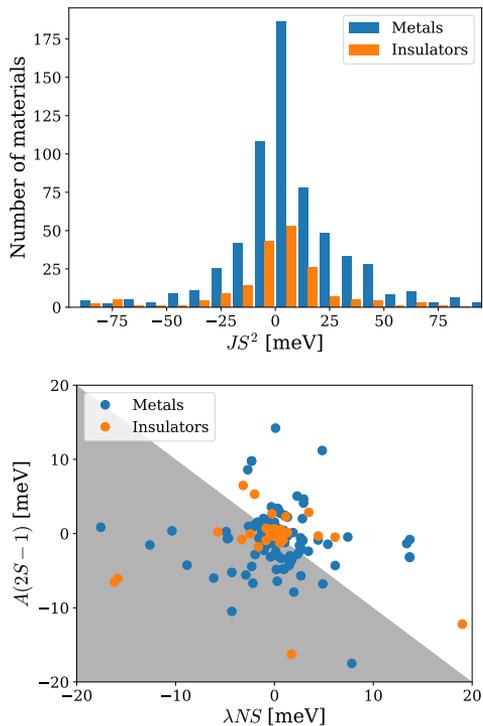


Figure 20. Top: Distribution of exchange coupling constants in C2DB. Bottom: Single-ion anisotropy A vs anisotropic exchange λ for ferromagnetic materials with $S > 1/2$. The shaded area indicates the part of parameter space where the model (Eq. (12)) does not yield an ordered state at finite temperatures.

reads [104]

$$I(\omega) = I_0 \sum_{\nu} \frac{n_{\nu} + 1}{\omega_{\nu}} \left| \sum_{ij} u_{in}^i R_{ij}^{\nu} u_{out}^j \right|^2 \delta(\omega - \omega_{\nu}). \quad (13)$$

Here, I_0 is an unimportant constant (since Raman spectra are always reported normalized), and n_{ν} is obtained from the Bose-Einstein distribution, i.e. $n_{\nu} \equiv (\exp[\hbar\omega_{\nu}/k_B T] - 1)^{-1}$ at temperature T for a Raman mode with energy $\hbar\omega_{\nu}$. Note that only phonons at the Brillouin zone center (with zero momentum) contribute to the one-phonon Raman processes due to momentum conservation. In Eq. (13), R_{ij}^{ν} is the Raman tensor for phonon mode ν , which involves electron-phonon and dipole matrix elements as well as the electronic transition energies and the incident excitation frequency. Eq. (13) has been used to compute the Raman spectra of the 733 most stable, non-magnetic monolayers in C2DB for a range of excitation frequencies and polarization configurations. Note that the Raman shift $\hbar\omega$ is typically expressed

in cm^{-1} with 1 meV equivalent to 8.0655 cm^{-1} . In addition, for generating the Raman spectra, we have used a Gaussian $[G(\omega) = (\sigma\sqrt{2\pi})^{-1} \exp\{(-\omega^2/2\sigma^2)\}]$ with a variance $\sigma = 3 \text{ cm}^{-1}$ to replace the Dirac delta function, which accounts for the inhomogeneous broadening of phonon modes.

As an example, Fig. 21 shows the calculated Raman spectrum of monolayer MoS₂ and the Janus monolayer MoSSe (see Sec. 4.1). Experimental Raman spectra extracted from Ref. [52] are shown for comparison. For both materials, good agreement between theory and experiment is observed for the peak positions and relative amplitudes of the main peaks. The small deviations can presumably be attributed to substrate interactions and defects in the experimental samples as well as the neglect of excitonic effects in the calculations. The qualitative differences between the Raman spectra can be explained by the different point groups of the materials (C_{3v} and D_{3h} , respectively), see Ref. [104]. In particular, the lower symmetry of MoSSe results in a lower degeneracy of its vibrational modes leading to more peaks in the Raman spectrum.

5.10. Second harmonics generation

Nonlinear optical (NLO) phenomena such as harmonic generation, Kerr, and Pockels effects are of great technological importance for lasers, frequency converters, modulators, etc. In addition, NLO spectroscopy has been extensively employed to obtain insight into materials properties [105] that are not accessible by e.g. linear optical spectroscopy. Among numerous nonlinear processes, second-harmonic generation (SHG) has been widely used for generating new frequencies in lasers as well as identifying crystal orientations and symmetries.

Recently, the SHG spectrum was calculated for 375 non-magnetic, non-centrosymmetric semiconducting monolayers of the C2DB, and multiple 2D materials with giant optical nonlinearities were identified [106]. In the SHG process, two incident photons at frequency ω generate an emitted photon at frequency of 2ω . Assume that a mono-harmonic electric field written $\mathcal{F}(t) = \sum_i \mathcal{F}_i \mathbf{e}_i e^{-i\omega t} + \text{c.c.}$ is incident on the material, where \mathbf{e}_i denotes the unit vector along direction $i \in \{x, y, z\}$. The electric field induces a SHG polarization density $\mathbf{P}^{(2)}$, which can be obtained from the quadratic susceptibility tensor $\chi_{ijk}^{(2)}$,

$$P_i^{(2)}(t) = \epsilon_0 \sum_{jk} \chi_{ijk}^{(2)}(\omega, \omega) \mathcal{F}_j \mathcal{F}_k e^{-2i\omega t} + \text{c.c.}, \quad (14)$$

where ϵ_0 denotes the vacuum permittivity. $\chi_{ijk}^{(2)}$ is a symmetric (due to intrinsic permutation symmetry i.e. $\chi_{ijk}^{(2)} = \chi_{jik}^{(2)}$) rank-3 tensor with at most 18 independent elements. Furthermore, similar to the piezoelectric

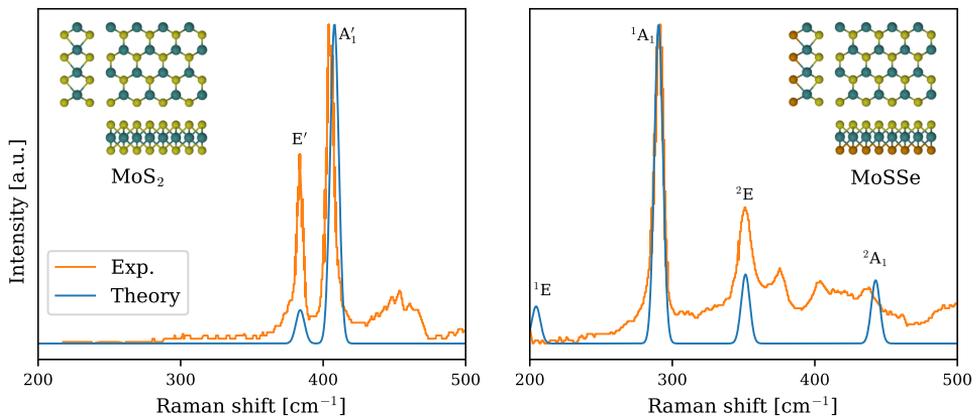


Figure 21. Comparison of the calculated and experimental (extracted from Ref. [52]) Raman spectrum of MoS₂ (left) and MoSSe (right). The excitation wavelength is 532 nm, and both the polarization of both the incoming and outgoing photons are along the y -direction. The Raman peaks are labeled according to the irreducible representations of the corresponding vibrational modes. Adapted from Ref. [104].

tensor, the point group symmetry reduces the number of independent tensor elements.

In the C2DB, the quadratic susceptibility is calculated using density matrices and perturbation theory [107, 108] with the involved transition dipole matrix elements and band energies obtained from DFT. The use of DFT single-particle orbitals implies that excitonic effects are not accounted for. The number of empty bands included in the sum over bands was set to three times the number of occupied bands. The width of the Fermi-Dirac occupation factor was set to $k_B T = 50$ meV, and a line-shape broadening of $\eta = 50$ meV was used in all spectra. Furthermore, time-reversal symmetry was imposed in order to reduce the \mathbf{k} -integrals to half the BZ. For various 2D crystal classes, it was verified by explicit calculation that the quadratic tensor elements fulfill the expected symmetries, e.g. that they all vanish identically for centrosymmetric crystals.

As an example, the calculated SHG spectra for monolayer Ge₂Se₂ is shown in Fig. 22 (left panel). Monolayer Ge₂Se₂ has 5 independent tensor elements, $\chi_{xxx}^{(2)}$, $\chi_{xyy}^{(2)}$, $\chi_{xzz}^{(2)}$, $\chi_{yyx}^{(2)} = \chi_{yxy}^{(2)}$, and $\chi_{zzx}^{(2)} = \chi_{zxx}^{(2)}$, since it is a group-IV dichalcogenide with an orthorhombic crystal structure (space group 31 and point group C_{2v}). Note that, similar to the linear susceptibility, the bulk quadratic susceptibility (with SI units of m/V) is ill-defined for 2D materials (since the volume is ambiguous) [106]. Instead, the unambiguous *sheet* quadratic susceptibility (with SI units of m²/V) is evaluated. In addition to the frequency-dependent SHG spectrum, the angular dependence of the static ($\omega = 0$) SHG intensity at normal incidence for parallel and perpendicular polarizations (relative to

the incident electric field) is calculated, see Fig. 22 (right panel). Such angular resolved SHG spectroscopy has been widely used for determining the crystal orientation of 2D materials. The calculated SHG spectra for all non-vanishing inequivalent polarization configurations and their angular dependence, are available in the C2DB.

Since C2DB has already gathered various material properties of numerous 2D materials, it provides a unique opportunity to investigate interrelations between different material properties. For example, the strong dependence of the quadratic optical response on the electronic band gap was demonstrated on basis of the C2DB data [106]. As another example of a useful correlation, the static quadratic susceptibility is plotted versus the static linear susceptibility for 67 TMDCs (with formula MX₂, space group 187) in Fig. 23. Note that for materials with several independent tensor elements, only the largest is shown. There is a very clear correlation between the two quantities. This is not unexpected as both the linear and quadratic optical responses are functions of the transition dipole moments and transition energies. More interestingly, the strength of the quadratic response seems to a very good approximation to be given by a universal constant times the linear susceptibility to the power of three (ignoring polarisation indices), i.e.

$$\chi^{(2)}(0, 0) \approx A\chi^{(1)}(0)^3, \quad (15)$$

where A is only weakly material dependent. Note that this scaling law is also known in classical optics as semi-empirical Miller's rule for non-resonant quadratic responses [109], which states that the second order

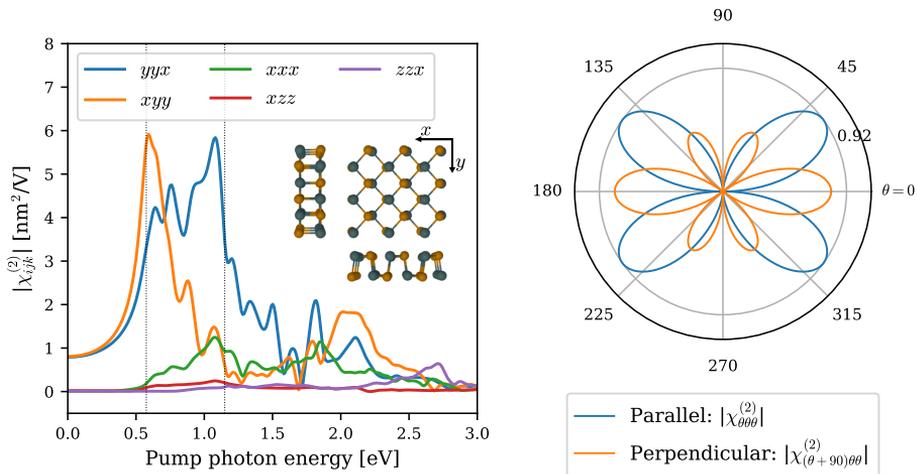


Figure 22. (Left panel) SHG spectra of monolayer Ge_2Se_2 , where only non-vanishing independent tensor elements are shown. The vertical dashed lines mark $h\omega = E_g/2$ and $h\omega = E_g$, respectively. The crystal structure of Ge_2Se_2 structure is shown in the inset. (Right panel) The rotational anisotropy of the static ($\omega = 0$) SHG signal for parallel (blue) and perpendicular (red) polarization configurations with θ defined with respect to the crystal x -axis.

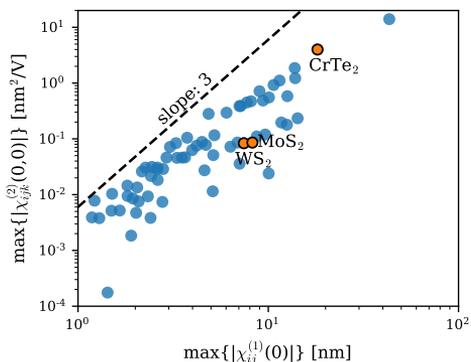


Figure 23. Scatter plot (double log scale) of the static sheet quadratic susceptibility $|\chi_{ijk}^{(2)}|$ versus the static sheet linear susceptibility $|\chi_{ij}^{(1)}|$ for 67 TMDCs (with chemical formula MX_2 and space group 187). A few well known materials are highlighted.

electric susceptibility is proportional to the product of the first-order susceptibilities at the three frequencies involved.

6. Machine learning properties

In recent years, material scientists have shown great interest in exploiting the use of machine learning (ML) techniques for predicting materials properties and guiding the search for new materials. ML is the

scientific study of algorithms and statistical models that computer systems can use to perform a specific task without using explicit instructions but instead relying on patterns and inference. Within the domain of materials science, one of the most frequent problems is the mapping from atomic configuration to material property, which can be used e.g. to screen large material spaces in search of optimal candidates for specific applications. [110, 111]

In the ML literature, the mathematical representation of the input observations is often referred to as a fingerprint. Any fingerprint must satisfy a number of general requirements. [112] In particular, a fingerprint must be

Complete: The fingerprint should incorporate all the relevant input for the underlying problem, i.e. materials with different properties should have different fingerprints.

Compact: The fingerprint should contain no or a minimal number of features redundant to the underlying problem. This includes being invariant to rotations, translations and other transformations that leave the properties of the system invariant.

Descriptive: Materials with similar target values should have similar fingerprints.

Simple: The fingerprint should be efficient to evaluate. In the present context, this means that calculating the fingerprint should be significantly faster than calculating the target property.

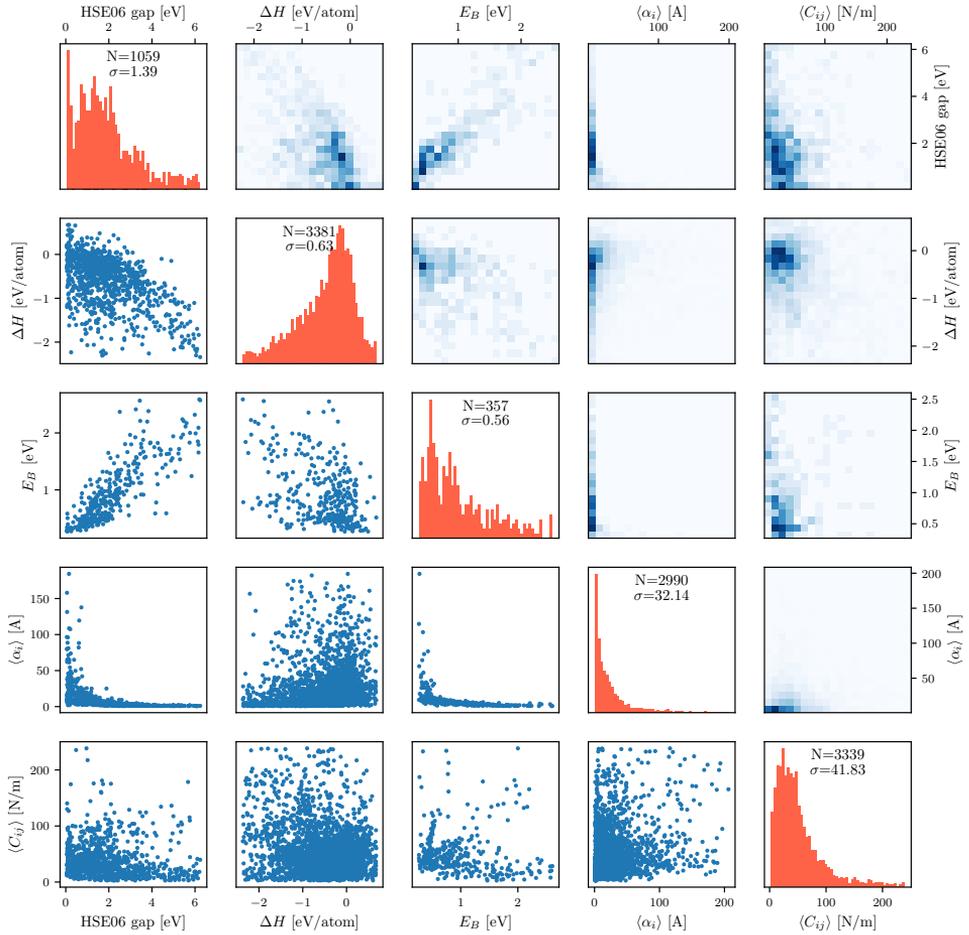


Figure 24. Pair-plot of selected properties from C2DB. The diagonal contains the single property histograms. Below the diagonal are two-property scatter plots showing the correlation between properties and above the diagonal are two-property histograms. properties include the HSE06 band gap, the PBE heat of formation (ΔH), the exciton binding energy (E_B) calculated from the Bethe-Salpeter equation (BSE), the in-plane static polarisability calculated in the random phase approximation (RPA) and averaged over the x and y polarisation directions ($\langle\alpha_i\rangle$), and the in-plane Voigt modulus ($\langle C_{ij}\rangle$) defined as $\frac{1}{4}(C_{11} + C_{22} + 2C_{12})$, where C_{ij} is a component of the elastic stiffness tensor.

Several types of atomic-level materials fingerprints have been proposed in the literature, including general purpose fingerprints based on atomistic properties[113, 114] possibly encoding information about the atomic structure, i.e. atomic positions[115, 112, 116], and specialized fingerprints tailored for specific applications (materials/properties)[117, 118].

The aim of this section is to demonstrate how the C2DB may be utilized for ML-based prediction of general materials properties. Moreover, the study serves to illustrate the important role of the fingerprint for such problems. The 2D materials are represented using three different fingerprints: two popular structural fingerprints and a more advanced

fingerprint that encodes information about the the electronic structure via the projected density of states (PDOS). The target properties include the HSE06 band gap, the PBE heat of formation (ΔH), the exciton binding energy (E_B) obtained from the many-body Bethe-Salpeter equation (BSE), the in-plane static polarisability calculated in the random phase approximation (RPA) averaged over the x and y polarisation directions ($\langle\alpha_i\rangle$), and the in-plane Voigt modulus ($\langle C_{ii}\rangle$) defined as $\frac{1}{4}(C_{11} + C_{22} + 2C_{12})$, where C_{ij} is a component of the elastic stiffness tensor in Mandel notation.

To introduce the data, Figure 24 shows pair-plots of the dual-property relations of these properties. The plots in the diagonal show the single-property histograms, whereas the off-diagonals show dual-property scatter plots below the diagonal and histograms above the diagonal. Clearly, there are only weak correlations between most of the properties, with the largest degree of correlation observed between the HSE06 gap and exciton binding energy. The lack of strong correlations motivates the use of machine learning for predicting the properties.

The prediction models are build using the Ewald sum matrix and many-body tensor representation (MBTR) as structural fingerprints. The Ewald fingerprint is a version of the simple Coulomb matrix fingerprint[115] modified to periodic systems [112]. The MBTR encodes first, second and third order terms like atomic numbers, distances and angles between atoms in the system [116]. As an alternative to the structural fingerprints, a representation based on the PBE projected density of states (PDOS) is also tested. This fingerprint (to be published) encodes the coupling between the PDOS at different atomic orbitals and the distance between atoms. Since this fingerprint requires a DFT-PBE calculations to be performed, additional features derivable from the DFT calculation can be added to the fingerprint. In this study, the PDOS fingerprint is amended by the PBE band gap. The latter can in principle be extracted from the PDOS, but its explicit inclusion improves performance (see below).

A Gaussian process regression using a simple Gaussian kernel with a noise component is used as learning algorithm. The models are trained using 5-fold cross validation on a training set consisting of 80% of the materials with the remaining 20% held aside as test data. Prior to training the model, the input space is reduced to 50 features using principal component analysis (PCA). This step is necessary to reduce the huge number of features in the MBTR fingerprint to a manageable size. Although this is not required for the Ewald and PDOS fingerprints, we perform the same feature reduction in all cases. The optimal number of features depends on the choice of fingerprint, target

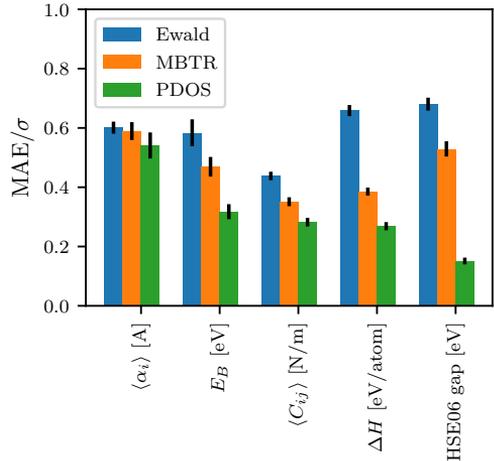


Figure 25. Prediction scores (MAE normalized to standard deviation of property values) for the test sets of selected properties using a Gaussian process regression.

property and learning algorithm, but for consistency 50 PCA components are used for all fingerprints and properties in this study.

Figure 25 shows the prediction scores obtained for the 5 properties using the three different fingerprints. The employed prediction score is the mean absolute error of the test set normalized by the standard deviation of the property values (standard deviations are annotated in the diagonal plots in Fig. 24). In general, the PDOS fingerprint outperforms the structural fingerprints. The difference between prediction scores is smallest for the static polarisability $\langle\alpha_i\rangle$ and largest for the HSE06 gap. It should be stressed that although the evaluation of the PBE-PDOS fingerprint is significantly more time consuming than the evaluation of the structural fingerprints, it is still much faster than the evaluation of all the target properties. Moreover, structural fingerprints require the atomic structure, which in turns requires a DFT structure optimization (unless the structure is available by other means).

The HSE06 band gap shows the largest sensitivity to the employed fingerprint. To elaborate on the HSE06 results, Fig. 26 shows the band gap predicted using each of the three different fingerprints plotted against the true band gap. The mean absolute errors on the test set is 0.95 eV and 0.74 eV for Ewald and MBTR fingerprints, respectively, while the PDOS significantly outperforms the other fingerprints with a test MAE of only 0.21 eV. This improvement in prediction accuracy is partly due to the presence of

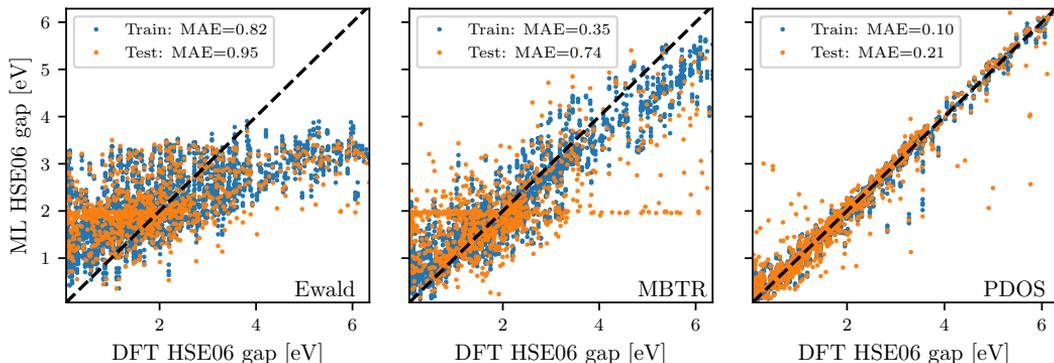


Figure 26. ML predicted HSE06 gap values vs. true values for Ewald, MBTR and PDOS fingerprints with MAE’s for train and test set included. The PDOS is found to perform significantly better for the prediction of HSE06 gap.

the PBE gap in the PDOS fingerprint. However, our analysis shows that the pure PDOS fingerprint without the PBE gap still outperforms the structural fingerprints. Using only the PBE gap as feature results in a test MAE of 0.28 eV.

The current results show that the precision of ML-based predictions are highly dependent on the type of target property and the chosen material representation. For some properties, the mapping between atomic structure and property is easier to learn while others might require more/deeper information, e.g. in terms of electronic structure fingerprints. Our results clearly demonstrate the potential of encoding electronic structure information into the material fingerprint, and we anticipate more work on this relevant and exciting topic in the future.

7. Summary and outlook

We have documented a number of extensions and improvements of the Computational 2D Materials Database (C2DB) made in the period 2018-2020. The new developments include: (i) A refined and more stringent workflow for filtering prospect 2D materials and classifying them according to their crystal structure, magnetic state and stability. (ii) Improvements of the methodology used to compute certain challenging properties such as the full stiffness tensor, effective masses, G_0W_0 band structures, and optical absorption spectra. (iii) New materials including 216 MXY Janus monolayers and 574 monolayers exfoliated from experimentally known bulk crystals. In addition, ongoing efforts to systematically obtain and characterize bilayers in all possible stacking configurations as well as point defects in the semiconducting monolayers, have been

described. (iv) New properties including exfoliation energies, spontaneous polarisations, Bader charges, piezoelectric tensors, infrared (IR) polarisabilities, topological invariants, magnetic exchange couplings, Raman spectra, and second harmonic generation spectra. It should be stressed that the C2DB will continue to grow as new structures and properties are being added, and thus the present paper should not be seen as a final report on the C2DB but rather a snapshot of its current state.

In addition to the above mentioned improvements relating to data quantity and quality, the C2DB has been endowed with a comprehensive documentation layer. In particular, all data presented on the C2DB website are now accompanied by an information field that explains the meaning and representation (if applicable) of the data and details how it was calculated thus making the data easier to understand, reproduce, and deploy.

The C2DB has been produced using the Atomic Simulation Recipes (ASR) in combination with the GPAW electronic structure code and the MyQueue task and workflow scheduling system. The ASR is a newly developed Python-based framework designed for high-throughput materials computations. The highly flexible and modular nature of the ASR and its strong coupling to the well established community-driven ASE project, makes it a versatile framework for both high- and low-throughput materials simulation projects. The ASR and the C2DB-ASR workflow are distributed as open source code. A detailed documentation of the ASR will be published elsewhere.

While the C2DB itself is solely concerned with the properties of perfect monolayer crystals, ongoing efforts focus on the systematic characterisation of homo-bilayer structures as well as point defects

in monolayers. The data resulting from these and other similar projects will be published as separate, independent databases, but will be directly interlinked with the C2DB making it possible to switch between them in a completely seamless fashion. These developments will significantly broaden the scope and usability of the C2DB+ (+ stands for associated databases) that will help theoreticians and experimentalists to navigate one of the most vibrant and rapidly expanding research fields at the crossroads of condensed matter physics, photonics, nanotechnology, and chemistry.

8. Acknowledgments

The Center for Nanostructured Graphene (CNG) is sponsored by the Danish National Research Foundation, Project DNRFF103. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement No 773122 (LIMA). T.D. acknowledges financial support from the German Research Foundation (DFG Projects No. DE 2749/2-1).

References

- [1] Kohn W and Sham L J 1965 *Physical review* **140** A1133
- [2] Schwierz F 2010 *Nature nanotechnology* **5** 487
- [3] Novoselov K, Mishchenko A, Carvalho A and Neto A C 2016 *Science* **353**
- [4] Ferrari A C, Bonaccorso F, Fal'Ko V, Novoselov K S, Roche S, Bøggild P, Borini S, Koppens F H, Palermo V, Pugno N *et al.* 2015 *Nanoscale* **7** 4598–4810
- [5] Bhimanapati G R, Lin Z, Meunier V, Jung Y, Cha J, Das S, Xiao D, Son Y, Strano M S, Cooper V R *et al.* 2015 *ACS nano* **9** 11509–11539
- [6] Hastrup S, Strange M, Pandey M, Deilmann T, Schmidt P S, Hinsche N F, Gjerding M N, Torelli D, Larsen P M, Riis-Jensen A C *et al.* 2018 *2D Materials* **5** 042002
- [7] Shivayogimath A, Thomsen J D, Mackenzie D M, Geisler M, Stan R M, Holt A J, Bianchi M, Crovetto A, Whelan P R, Carvalho A *et al.* 2019 *Nature communications* **10** 1–7
- [8] Zhou J, Lin J, Huang X, Zhou Y, Chen Y, Xia J, Wang H, Xie Y, Yu H, Lei J *et al.* 2018 *Nature* **556** 355–359
- [9] Anasori B, Lukatskaya M R and Gogotsi Y 2017 *Nature Reviews Materials* **2** 1–17
- [10] Dou L, Wong A B, Yu Y, Lai M, Kornienko N, Eaton S W, Fu A, Bischak C G, Ma J, Ding T *et al.* 2015 *Science* **349** 1518–1521
- [11] Mounet N, Gibertini M, Schwaller P, Campi D, Merkys A, Marrazzo A, Sohler T, Castellì E, Cepellotti A, Pizzi G *et al.* 2018 *Nature nanotechnology* **13** 246–252
- [12] Ashton M, Paul J, Sinnott S B and Hennig R G 2017 *Physical review letters* **118** 106101
- [13] Geim A K and Grigorieva I V 2013 *Nature* **499** 419–425
- [14] Cao Y, Fatemi V, Fang S, Watanabe K, Taniguchi T, Kaxiras E and Jarillo-Herrero P 2018 *Nature* **556** 43–50
- [15] Bistritzer R and MacDonald A H 2011 *Proceedings of the National Academy of Sciences* **108** 12233–12237
- [16] Zhao X, Song P, Wang C, Riis-Jensen A C, Fu W, Deng Y, Wan D, Kang L, Ning S, Dan J *et al.* 2020 *Nature* **581** 171–177
- [17] Wan J, Lacey S D, Dai J, Bao W, Fuhrer M S and Hu L 2016 *Chemical Society Reviews* **45** 6742–6765
- [18] Wilkinson M D, Dumontier M, Aalbersberg I J, Appleton G, Axton M, Baak A, Blomberg N, Boiten J W, da Silva Santos L B, Bourne P E *et al.* 2016 *Scientific data* **3** 1–9
- [19] Wirtz L, Marini A and Rubio A 2006 *Physical review letters* **96** 126104
- [20] Cudazzo P, Tokatly I V and Rubio A 2011 *Physical Review B* **84** 085406
- [21] Klots A, Newaz A, Wang B, Prasai D, Krzyzanowska H, Lin J, Caudel D, Ghimire N, Yan J, Ivanov B *et al.* 2014 *Scientific reports* **4** 6608
- [22] Chernikov A, Berkelbach T C, Hill H M, Rigosi A, Li Y, Aslan O B, Reichman D R, Hybertsen M S and Heinz T F 2014 *Physical review letters* **113** 076802
- [23] Olsen T, Latini S, Rasmussen F and Thygesen K S 2016 *Physical review letters* **116** 056401
- [24] Riis-Jensen A C, Gjerding M N, Russo S and Thygesen K S 2020 *arXiv preprint arXiv:2009.12317*
- [25] Felipe H, Xian L, Rubio A and Louie S G 2020 *Nature communications* **11** 1–10
- [26] Sohler T, Gibertini M, Calandra M, Mauri F and Marzari N 2017 *Nano letters* **17** 3758–3763
- [27] Ugeda M M, Bradley A J, Shi S F, Felipe H, Zhang Y, Qiu D Y, Ruan W, Mo S K, Hussain Z, Shen Z X *et al.* 2014 *Nature materials* **13** 1091–1095
- [28] Winther K T and Thygesen K S 2017 *2D Materials* **4** 025059
- [29] Wang Z, Rhodes D A, Watanabe K, Taniguchi T, Hone J C, Shan J and Mak K F 2019 *Nature* **574** 76–80
- [30] Gong C, Li L, Li Z, Ji H, Stern A, Xia Y, Cao T, Bao W, Wang C, Wang Y *et al.* 2017 *Nature* **546** 265–269
- [31] Huang B, Clark G, Navarro-Moratalla E, Klein D R, Cheng R, Seyler K L, Zhong D, Schmidgall E, McGuire M A, Cobden D H *et al.* 2017 *Nature* **546** 270–273
- [32] Chang K, Liu J, Lin H, Wang N, Zhao K, Zhang A, Jin F, Zhong Y, Hu X, Duan W *et al.* 2016 *Science* **353** 274–278
- [33] Olsen T, Andersen E, Okugawa T, Torelli D, Deilmann T and Thygesen K S 2019 *Physical Review Materials* **3** 024005
- [34] Marrazzo A, Gibertini M, Campi D, Mounet N and Marzari N 2019 *Nano Letters* **19** 8431–8440
- [35] Thygesen K S 2017 *2D Materials* **4** 022004
- [36] Torelli D and Olsen T 2018 *2D Mater.* **6** 015028
- [37] Rasmussen F A and Thygesen K S 2015 *The Journal of Physical Chemistry C* **119** 13169–13183
- [38] Mortensen J J, Gjerding M and Thygesen K S 2020 *Journal of Open Source Software* **5** 1844 URL <https://doi.org/10.21105/joss.01844>
- [39] Saal J E, Kirklin S, Aykol M, Meredig B and Wolverton C 2013 *Jom* **65** 1501–1509
- [40] Jain A, Ong S P, Hautier G, Chen W, Richards W D, Dacek S, Cholia S, Gunter D, Skinner D, Ceder G *et al.* 2013 *Apl Materials* **1** 011002
- [41] Curtarolo S, Setyawan W, Hart G L, Jahnatek M, Chepulskii R V, Taylor R H, Wang S, Xue J, Yang K, Levy O *et al.* 2012 *Computational Materials Science* **58** 218–226
- [42] Zhou J, Shen L, Costa M D, Persson K A, Ong S P, Huck P, Lu Y, Ma X, Chen Y, Tang H *et al.* 2019 *Scientific data* **6** 1–10
- [43] Larsen P M, Pandey M, Strange M and Jacobsen K W 2019 *Phys. Rev. Mater.* 034003
- [44] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2013 *Computational Materials Science* **68**

- 314–319
- [45] Patrick C E, Jacobsen K W and Thygesen K S 2015 *Physical Review B* **92** 201205
- [46] Kirklin S, Saal J E, Meredig B, Thompson A, Doak J W, Aykol M, Rühl S and Wolverton C 2015 *npj Computational Materials* **1** 1–15
- [47] Maździarz M 2019 *2D Materials* **6** 048001 URL <https://doi.org/10.1088/2053-1583/ab2ef3>
- [48] Li Y and Heinz T F 2018 *2D Mater.* **5** 025021 URL <https://doi.org/10.1088/2053-1583/aab0cf>
- [49] Hadley L N and Dennison D 1947 *JOSA* **37** 451–465
- [50] Wang G, Chernikov A, Glazov M M, Heinz T F, Marie X, Amand T and Urbaszek B 2018 *Reviews of Modern Physics* **90** 021001
- [51] Lu A Y, Zhu H, Xiao J, Chuu C P, Han Y, Chiu M H, Cheng C C, Yang C W, Wei K H, Wang Y, Wang Y, Sokaras D, Nordlund D, Yang P, Muller D A, Chou M Y, Zhang X and Li L J 2017 *Nature Nanotechnology* **12** 744–749 ISSN 1748-3387
- [52] Zhang J, Jia S, Kholmanov I, Dong L, Er D, Chen W, Guo H, Jin Z, Shenoy V B, Shi L and Lou J 2017 *ACS Nano* **11** 8192–8198 URL <https://doi.org/10.1021/acsnano.7b03186>
- [53] Fülöp B, Tajkov Z, Pető J, Kun P, Koltai J, Oroszlány L, Tóvári E, Murakawa H, Tokura Y, Bordács S *et al.* 2018 *2D Materials* **5** 031013
- [54] Riis-Jensen A C, Deilmann T, Olsen T and Thygesen K S 2019 *ACS Nano* **13** 13354 ISSN 1936-086X
- [55] Bychkov Y A and Rashba E I 1984 *Journal of physics C: Solid state physics* **17** 6039
- [56] Petersen L and Hedegård P 2000 *Surface science* **459** 49–56
- [57] Allmann R and Hinek R 2007 *Acta Crystallogr. Sect. A Found. Crystallogr.* **63** 412–417
- [58] Gražulis S, Dašević A, Merkys A, Chateigner D, Lutterotti L, Quirós M, Serebryanaya N R, Moeck P, Downs R T and Le Bail A 2012 *Nucleic Acids Res.* **40** D420–D427
- [59] Qian X, Liu J, Fu L and Li J 2014 *Science* **346** 1344–1347
- [60] Torelli D, Moustafa H, Jacobsen K W and Olsen T 2020 *npj Computational Materials* **6** 158
- [61] Mak K F, Lee C, Hone J, Shan J and Heinz T F 2010 *Physical review letters* **105** 136805
- [62] Splendiani A, Sun L, Zhang Y, Li T, Kim J, Chim C Y, Galli G and Wang F 2010 *Nano letters* **10** 1271–1275
- [63] Xiao J, Wang Y, Wang H, Pemmaraju C, Wang S, Muscher P, Sie E J, Nyby C M, Devereaux T P, Qian X *et al.* 2020 *Nature Physics* **16** 1028–1034
- [64] Sivadas N, Okamoto S, Xu X, Fennie C J and Xiao D 2018 *Nano letters* **18** 7658–7664
- [65] Liu Y, Wu L, Tong X, Li J, Tao J, Zhu Y and Petrovic C 2019 *Scientific reports* **9** 1–8
- [66] Yasuda K, Wang X, Watanabe K, Taniguchi T and Jarillo-Herrero P 2020 *arXiv preprint arXiv:2010.06600*
- [67] Northrup T and Blatt R 2014 *Nature photonics* **8** 356–363
- [68] O’Brien J, Furusawa A and Vuckovic J 2009 Photonic quantum technologies nat *Photonics* vol 3 p 687
- [69] Zhang S and Northrup J E 1991 *Physical review letters* **67** 2339
- [70] Van de Walle C G, Laks D, Neumark G and Pantelides S 1993 *Physical Review B* **47** 9425
- [71] Janak J F 1978 *Physical Review B* **18** 7165
- [72] Pandey M, Rasmussen F A, Kuhar K, Olsen T, Jacobsen K W and Thygesen K S 2016 *Nano letters* **16** 2234–2239
- [73] Kaappa S, Malola S and Häkkinen H 2018 *The Journal of Physical Chemistry A* **122** 8576–8584
- [74] Levi G, Ivanov A V and Jonsson H 2020 *Faraday Discussions*
- [75] Grimme S, Antony J, Ehrlich S and Krieg H 2010 *The Journal of chemical physics* **132** 154104
- [76] Bader R F W 1990 *Atoms in molecules : a quantum theory* The International series of monographs on chemistry ; 22 (Oxford: Clarendon Press) ISBN 0198551681
- [77] Tang W, Sanville E and Henkelman G 2009 *Journal of Physics: Condensed Matter* **21** 084204 URL <https://doi.org/10.1088/0953-8984/21/8/084204>
- [78] Resta R 1992 *Ferroelectrics* **136** 51–55 URL <https://doi.org/10.1080/00150199208016065>
- [79] King-Smith R D and Vanderbilt D 1993 *Physical Review B* **47**(3) URL <https://doi.org/10.1103/PhysRevB.47.1651>
- [80] Zhang S and Yu F 2011 *Journal of the American Ceramic Society* **94** 3153–3170
- [81] Maeder M D, Damjanovic D and Setter N 2004 *Journal of Electroceramics* **13** 385–392
- [82] Scott J F 2000 *Ferroelectric memories* vol 3 (Springer Science & Business Media)
- [83] Rangel T, Fregoso B M, Mendoza B S, Morimoto T, Moore J E and Neaton J B 2017 *Phys. Rev. Lett.* **119**(6-11) URL <https://doi.org/10.1103/PhysRevLett.119.067402>
- [84] Resta R and Vanderbilt D 2007 Theory of Polarization: A Modern Approach *Physics of Ferroelectrics* vol 105 (Springer Berlin Heidelberg) pp 31–68 ISBN 3540345906
- [85] Gjerding M N, Cavalcante L S R, Chaves A and Thygesen K S 2020 *The Journal of Physical Chemistry C* **124** 11609–11616 ISSN 1932-7447 URL <https://doi.org/10.1021/acs.jpcc.0c01635>
- [86] Ye Z G 2008 *Handbook of advanced dielectric, piezoelectric and ferroelectric materials: Synthesis, properties and applications* (Elsevier)
- [87] Ogawa T 2016 *Piezoelectric Materials* (BoD-Books on Demand)
- [88] Vanderbilt D 1999 *Journal of Physics and Chemistry of Solids* **61**(2) URL [https://doi.org/10.1016/S0022-3697\(99\)00273-5](https://doi.org/10.1016/S0022-3697(99)00273-5)
- [89] Authier A 2003 *International tables for crystallography: Volume D: Physical properties of crystals* (Wiley Online Library)
- [90] Duerloo K A N, Ong M T and Reed E J 2012 *The Journal of Physical Chemistry Letters* **3** 2871–2876 URL <https://doi.org/10.1021/jz3012436>
- [91] Zhu H, Wang Y, Xiao J, Liu M, Xiong S, Wong Z J, Ye Z, Ye Y, Yin X and Zhang X 2015 *Nature Nanotechnology* **10** 151–155 ISSN 1748-3395 URL <https://doi.org/10.1038/nnano.2014.309>
- [92] Taherinejad M, Garrity K F and Vanderbilt D 2014 *Physical Review B* **89** 115102
- [93] Olsen T 2016 *Phys. Rev. B* **94** 235106
- [94] Fu L 2011 *Physical Review Letters* **106** 106802
- [95] Olsen T, Andersen E, Okugawa T, Torelli D, Deilmann T and Thygesen K S 2019 *Physical Review Materials* **3** 024005
- [96] Benalcazar W A, Bernevig B A and Hughes T L 2017 *Physical Review B* **96** 245115
- [97] Mermin N D and Wagner H 1966 *Physical Review Letters* **17** 1133–1136
- [98] Olsen T 2019 *MRS Communications* **9** 1142–1150
- [99] Olsen T 2017 *Physical Review B* **96** 125143
- [100] Torelli D and Olsen T 2020 *Journal of Physics: Condensed Matter* 335802
- [101] Lado J L and Fernández-Rossier J 2017 *2D Mater.* **4** 035002
- [102] Torelli D, Thygesen K S and Olsen T 2019 *2D Materials* **6** 045018
- [103] Long D A 2002 *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules* (John Wiley & Sons Ltd, Chichester) ISBN 0471490288 URL <https://onlinelibrary.wiley.com/doi/book/10.1002/0470845767>

- [104] Taghizadeh A, Leffers U, Pedersen T G and Thygesen K S 2020 *Nat. Commun.* **11** 3011 URL <https://doi.org/10.1038/s41467-020-16529-6>
- [105] Prylepa A, Reitböck C, Cobet M, Jesacher A, Jin X, Adeling R, Schatzl-Linder M, Luckeneder G, Stellnberger K H, Steck T, Faderl J, Stehrer T and Stifter D 2018 *J. Phys. D: Appl. Phys.* **51** 043001 URL <https://doi.org/10.1088/1361-6463/aa9df4>
- [106] Taghizadeh A, Thygesen K S and Pedersen T G 2020 *arXiv:2010.11596 [cond-mat, physics:physics]* ArXiv: 2010.11596 URL <http://arxiv.org/abs/2010.11596>
- [107] Aversa C and Sipe J E 1995 *Phys. Rev. B* **52**(20) 14636–14645 URL <https://link.aps.org/doi/10.1103/PhysRevB.52.14636>
- [108] Taghizadeh A, Hipolito F and Pedersen T G 2017 *Phys. Rev. B* **96**(19) 195413 URL <https://link.aps.org/doi/10.1103/PhysRevB.96.195413>
- [109] Miller R C 1964 *Appl. Phys. Lett.* **5** 17–19 URL <https://doi.org/10.1063/1.1754022>
- [110] Schmidt J, Marques M R G, Botti S and Marques M A L 2019 *npj Computational Materials* **5** 83 ISSN 2057-3960 URL <https://doi.org/10.1038/s41524-019-0221-0>
- [111] Zhuo Y, Mansouri Tehrani A and Brgoch J 2018 *The Journal of Physical Chemistry Letters* **9** 1668–1673 pMID: 29532658 (Preprint <https://doi.org/10.1021/acs.jpcllett.8b00124>) URL <https://doi.org/10.1021/acs.jpcllett.8b00124>
- [112] Faber F, Lindmaa A, von Lilienfeld O A and Armiento R 2015 *International Journal of Quantum Chemistry* **115** 1094–1101 (Preprint <https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.24917>) URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.24917>
- [113] Ward L, Agrawal A, Choudhary A and Wolverton C 2016 *npj Computational Materials* **2** 1–7
- [114] Ghiringhelli L M, Vybiral J, Levchenko S V, Draxl C and Scheffler M 2015 *Physical review letters* **114** 105503
- [115] Rupp M, Tkatchenko A, Müller K R and Von Lilienfeld O A 2012 *Physical review letters* **108** 058301
- [116] Huo H and Rupp M 2018 Unified representation of molecules and crystals for machine learning (Preprint 1704.06439)
- [117] Jørgensen P B, Mesta M, Shil S, García Lastra J M, Jacobsen K W, Thygesen K S and Schmidt M N 2018 *The Journal of chemical physics* **148** 241735
- [118] Rajan A C, Mishra A, Satsangi S, Vaish R, Mizuseki H, Lee K R and Singh A K 2018 *Chemistry of Materials* **30** 4031–4038

Paper IV

Atomic Simulation Recipes – a Python framework and library for automated workflows

Morten Niklas Gjerding, Thorbjørn Skovhus, **Ashbjørn Rasmussen**, Fabian Bertoldo, Ask Hjorth Larsen, Jens Jørgen Mortensen, Kristian Sommer Thygesen

Submitted

Morten Gjerding¹, Thorbjørn Skovhus¹, Asbjørn Rasmussen¹, Fabian Bertoldo¹, Ask Hjorth Larsen¹, Jens Jørgen Mortensen¹ and Kristian Sommer Thygesen¹

¹Computational Atomic-scale Materials Design (CAMD), Department of Physics, Technical University of Denmark, 2800 Kgs. Lyngby Denmark

ARTICLE INFO

Keywords:

High-throughput, Database, Data provenance, Workflow, Python, Materials computation, Density functional theory,

ABSTRACT

The Atomic Simulation Recipes (ASR) is an open source Python framework for working with atomistic materials simulations in an efficient and sustainable way that is ideally suited for high-throughput projects. Central to ASR is the concept of a Recipe: a high-level Python script that performs a well defined simulation task robustly and accurately while keeping track of the data provenance. The ASR leverages the functionality of the Atomic Simulation Environment (ASE) to interface with external simulation codes and attain a high abstraction level. We provide a library of Recipes for common simulation tasks employing density functional theory and many-body perturbation schemes. These Recipes utilize the GPAW electronic structure code, but may be adapted to other simulation codes with an ASE interface. Being independent objects with automatic data provenance control, Recipes can be freely combined through Python scripting giving maximal freedom for users to build advanced workflows. ASR also implements a command line interface that can be used to run Recipes and inspect results. The ASR Migration module helps users maintain their data while the Database and App modules makes it possible to create local databases and present them as customized web pages.

ORCID(s):

¹email: mortengjerding@gmail.com

1. Introduction

As computing power continues to increase and the era of exascale approaches, the development of software solutions capable of exploiting the immense computational resources becomes a key challenge for the scientific community. In the field of materials science, ab initio electronic structure (aiES) calculations are increasingly being conducted in a high-throughput fashion to screen thousands of materials for various applications[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16] and to generate large reference data sets for training machine learning algorithms to predict fundamental materials properties[17, 18, 19, 20, 21, 22] or design interatomic potentials[23, 24, 25, 26]. The results from such aiES high-throughput calculations are often stored in open databases allowing the data to be efficiently shared and deployed beyond the original purpose[27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37].

While a few thousands of calculations can be managed manually, a paradigm in which data drives scientific discovery calls for dedicated workflow solutions that automatically submit and retrieve the calculations, store the results in organized data structures, and keep track of the origin, history and dependencies of all data, i.e. the data provenance. Ideally, the workflow should also attach explanatory descriptions to the data that allows them to be easily accessed, understood, and deployed – also by users with limited domain knowledge.

Materials scientists from the aiES community are employing a large and heterogeneous set of simulation codes based mainly on density functional theory (DFT)[38]. These codes differ substantially in the way they implement and solve the fundamental physical equations. This is due to the fact that different types of problems require different numerical approaches, e.g. high accuracy vs. large system sizes, periodic vs. finite vs. open boundary conditions, or ground state vs. excited state properties. In principle, the large pool of available aiES codes provides users with a great deal of flexibility and freedom to pick the code that best suits the problem at hand. In practice, however, the varying numerical implementations and the diverse and often rudimentary user interfaces make it challenging for users to switch between the different aiES codes leading to a significant “code barrier”.

To some extent, a similar situation exists with respect to materials properties. Although aiES codes provide access to a rich variety of physical and chemical properties, individual researchers often focus on properties within a specific scientific domain. While this may be sufficient in many cases, several important contemporary problems addressed by the aiES community are multi-physical in nature and require properties and insights from several domains. For example, evaluating the potential of a material as a photocatalyst involves an assessment of solar light absorption, charge transport, and chemical reactions at a solid–liquid interface. Calculating new types of properties for the first time is often a time-consuming process involving trial and error and the acquisition of technical, implementation-specific knowledge of no direct benefit for the user or the overall project aim. This situation may result in a “property barrier” that hampers researchers’ exploitation of the full capacity of aiES codes.

In this paper, we introduce The Atomic Simulation Recipes (ASR) – a highly flexible Python framework for developing and working with computational materials workflows. The ASR reduces code and property barriers and makes it easy to perform high-throughput computations with advanced workflows while adhering to the FAIR Data Principles[39]. There are already some workflow solutions available in the field, some of the most prominent being AFlow[30], Fireworks[40], AiiDA[41], and Atomate[42]. However, these are either designed for one specific simulation code and/or constitute rather colossal integrated entities, the complexity of which could represent an entry barrier to some users. The ASR differs from the existing solutions in several important ways, and we expect it to appeal to a large crowd of computational researchers, e.g. those with Python experience who like to develop their own personalized (workflow) scripts and databases, less experienced users who prefer plug-and-play solutions, and those who wish to apply non-standard methodologies, e.g. compute GW band structures or Raman spectra, but feel they lack the expertise required for using standard low-level codes.

The basic philosophy of ASR is to prioritize usability and simplicity over system perfection. More specifically, ASR is characterized by the following qualities:

- **Flexibility:** The Python scripting interface and high degree of modularity provide users with almost unlimited freedom for developing and deploying workflows.
- **Modularity:** The key components of ASR, namely the workflow development framework (ASR core), the Database and App modules, the task scheduler (MyQueue), and the simulation codes, are separate independent entities. Moreover, the Recipe library concept supports modular workflow designs and reuse of code.
- **Data locality:** Generated data is stored in a special folder named `.asr` where it can be accessed transparently via command line tools (similar to Git).

- **Compatibility:** For compatibility with external simulation codes, the ASR core is fully simulation code-independent while specific Recipe implementations communicate with simulation codes via the abstract ASE Calculator interface.
- **Minimalism and pragmatism:** ASR is based on simple solutions that work efficiently in practice. This makes ASR fast to learn, easy to use, and relatively uncomplicated to adapt to future demands.

At the core of ASR is the concept of a *Recipe*. In essence, a Recipe is a piece of code that can perform a certain simulation task (e.g. relax an atomic structure, calculate a Raman spectrum, or identify covalently bonded components of a material) while recording all relevant results and metadata. The use of Recipes makes it simple to run simulations from either Python or the command line. For example,

```
$ asr run "asr.bandstructure --atoms structure.json"
```

will calculate the electronic band structure of the material `structure.json`. Subsequently, the command

```
$ asr results asr.bandstructure
```

will produce a plot of the band structure. With two additional commands, the ASR results can be inspected in a web browser, see example in Fig. 6.

In practice, Recipes are implemented as Python modules building on the Atomic Simulation Environment (ASE)[43]. Recipes conform to certain naming and structured programming conventions, making them largely self-documenting and easy to read. To keep track of data provenance, Recipes utilize a caching mechanism that automatically logs all exchange of data with the user and other Recipes in a uniquely identifiable Record object. Not only does this guarantee the documentation and reproducibility of the results, it also allows ASR to determine whether a given Recipe task has already been performed (such that its result can be directly loaded and returned) and to detect if a Recipe task needs to be rerun because another piece of data in its dependency chain has changed. In addition, Recipes implement presentation and explanatory descriptions of their outputs and may also define a web panel for online presentation.

The Recipes of the current ASR library cover a variety of computational tasks and properties (see Table 1). Most of the 40+ available Recipes utilize DFT. However, some Recipes do not involve calls to a simulation code (e.g. symmetry analysis or construction of phase diagrams) while others employ beyond-DFT methodology (e.g. the GW method or the Bethe–Salpeter equation). These library Recipes can be used “out of the box” or modified to fit the user’s need. New Recipes may be developed straightforwardly following the documentation and large body of available examples. Recipes can be combined into complex workflows using Python scripting for maximal flexibility and compatibility with ASE and other relevant Python libraries like PymatGen[44], Spglib[45] and Phonopy[46]. The Python workflows may be executed on supercomputers using the MyQueue[47] task scheduler front-end or other similar systems.

The ASR contains a number of tools for working with the ASE database module, which makes it easy to generate and maintain local materials databases. Relying on the Recipes’ web panel implementations, these databases may be straightforwardly presented in a browser allowing for easy inspection, querying, and sharing of results on a local or public network. As an example of an ASR-driven database project we refer to the Computational 2D Materials Database (C2DB)[32, 48]¹.

While the core of ASR, i.e. the Recipe concept and caching system, is fully simulation code-independent, most Recipe implementations of the current library contain calls to the specific aiES code GPAW[49]. We are currently working on a generalization of the ASE Calculator interface which will decouple Recipe implementations from simulation codes. In the future, many Recipes will therefore work with multiple simulation codes.

Another on-going effort is to generalize the organization of calculated results. For example results are currently presented mainly by material. This is practical for a database which primarily associates a number of properties with each material, but not for presenting sets of results parametrized over other variables than the material. These limitations will be removed over the next releases.

The rest of this paper is organised as follows: In Section 2 we provide a general overview of the main components of ASR. Section 3 zooms in on the central Recipe concept and its caching system while Section 4 gives an overview of the currently available Recipes. In Section 5, the Database and App modules are described. Section 6 gives a brief presentation of the Computational 2D Materials Database as an example of an ASR-driven high-throughput database project and provides a few concrete examples of Recipe implementations. Section 7 describes the different

¹<http://c2db.fysik.dtu.dk>

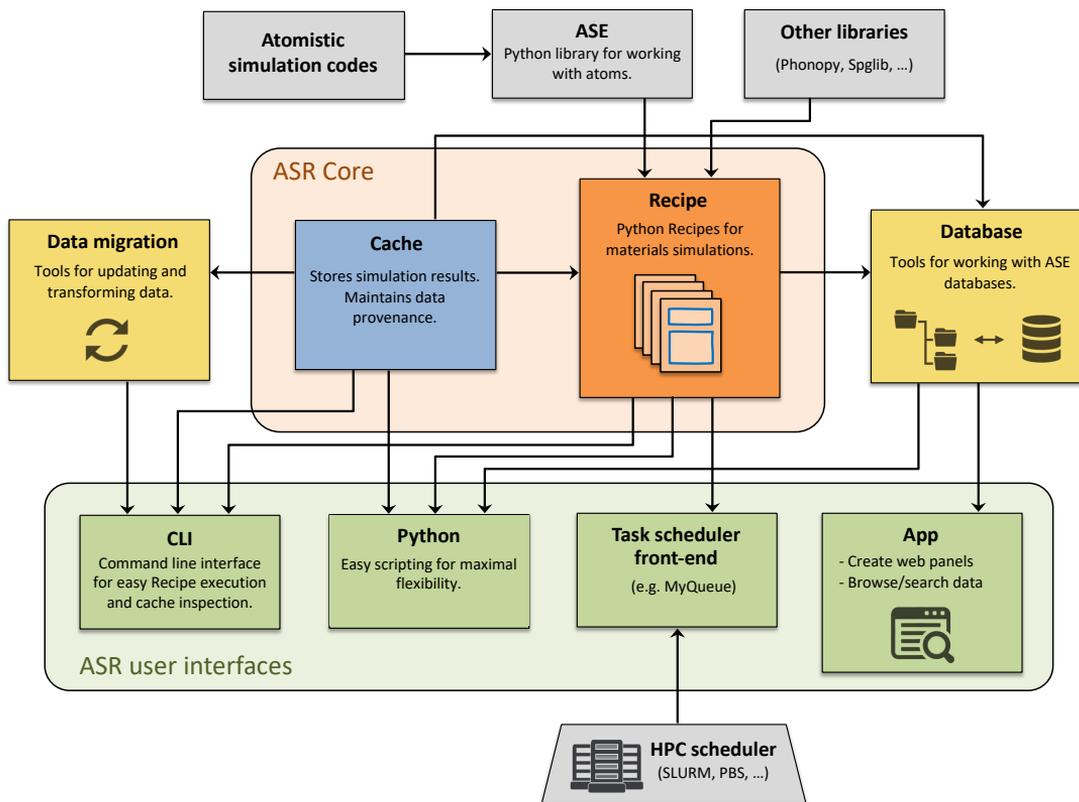


Figure 1: Schematic overview of the main modules of the ASR and their interrelations. ASR consists of a Python library of Recipes for materials simulations and a caching system for recording of results and metadata. Recipes are envisioned to communicate with simulation codes via ASE interfaces, although most current Recipe implementations contain parts that are specific to the GPAW code. An arrow from X to Y means that Y calls X. The blue frames on the Instructions Recipe box symbolise a caching layer that records all data flow to/from the Recipes.

user interfaces supported by ASR while Sections 8 and 9 explain how ASR manages data migration and provenance, respectively. Sections 10 and 11 cover documentation and technical specifications. Finally, Section 12 summarises the paper and presents our future perspectives for ASR.

2. Overview of ASR

Fig. 1 shows a schematic overview of the main components of the ASR and their mutual dependencies. An arrow from X to Y indicates a direct dependence of Y on X, e.g. via function calls (Y calls X). The ASR modules have been divided into the ASR core modules (Cache and Recipe) and the ASR user interfaces (command-line interface, Python, Task scheduler front-end, and Apps). In addition, the ASR Database and Data migration modules contain tools for working with databases and maintaining data, respectively.

Recipes implement specific, well defined materials simulation tasks as Python modules building on the ASE[43] and other Python libraries. A Recipe integrates with a Cache module that keeps track of performed tasks and manages all relevant metadata. The Cache also allows the user to inspect the data generated by a Recipe via the ASR command line interface (CLI) or using Python. Likewise, the Recipes may be executed directly from the CLI or called via Python scripts, the latter giving maximal flexibility and compatibility with existing Python libraries. For the purpose of high-

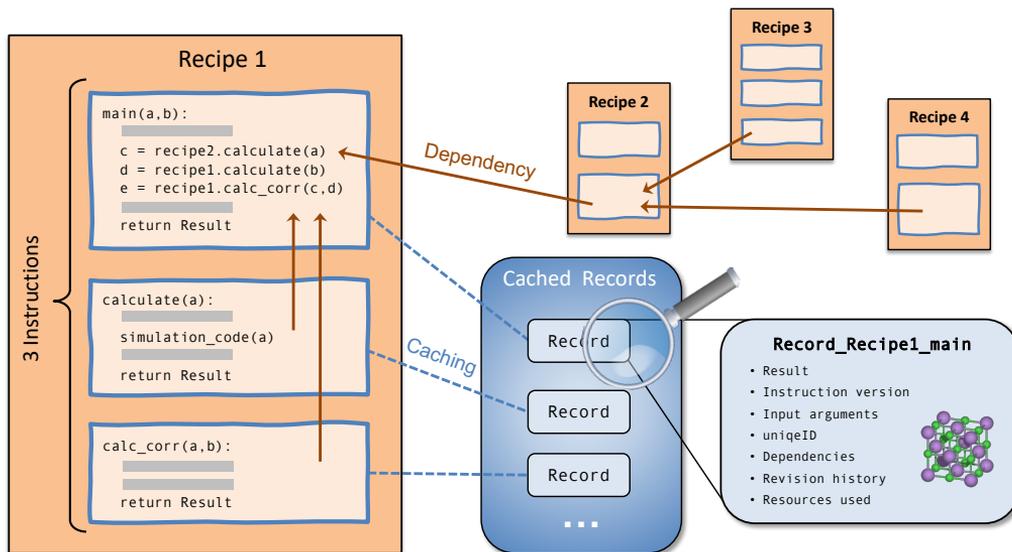


Figure 2: A Recipe consists of a set of Instructions (see Fig. 3) implementing the computational steps needed to obtain a desired result. An Instruction may call other Instructions of the same, or separate, Recipes. An Instruction always returns a Record holding its result, normally represented as a Result data structure, together with the dependencies on other Instructions and all additional metadata required to trace back and reproduce the result.

throughput computations, advanced Python workflows combining several Recipes may be constructed and executed remotely using task scheduling systems like MyQueue[47].

The ASR Cache and Recipe modules work on a folder/file basis. This locality of data makes the ASR highly transparent for the user. The ASR Database module contains functions for converting the ASR data stored in a tree of folders into an ASE database and vice versa. The ASR App module generates web pages for online presentation, browsing and searching of the databases generated by the ASR Database module. Finally, the Data migration module provides tools for transforming data (results or metadata) to ensure backward compatibility when Recipes are updated.

3. What is a Recipe?

A Recipe is a Python module implementing the Instructions needed to obtain a particular result, for example to relax an atomic structure, calculate an electronic band structure or a piezoelectric tensor. This section describes the structure and main components of a Recipe. A schematic overview of the Recipe concept is shown in Fig. 2.

3.1. Instruction

An Instruction is to be understood as a Python function wrapped in a caching layer provided by ASR, see Fig. 3. Whenever an Instruction is called, the caching layer intercepts the input arguments and asks the cache whether the result of the particular Instruction call already exists (cache hit) or whether there are no matching results (cache miss). If a matching result exists (because it was calculated previously), the caching layer skips the actual evaluation of the Instruction and simply reads and returns the previously calculated result. In the case of a cache miss, the Instruction is evaluated, after which the result is intercepted by the caching layer and stored together with the relevant metadata in a Record object. The precise content of the Record object and the conditions for a cache hit/miss are described in Section 3.3.

One of the great benefits of this design is its simplicity. Because the Instruction/caching layer is implemented as a simple wrapper around a Python function, usage of the caching functionality requires minimal additional knowledge.

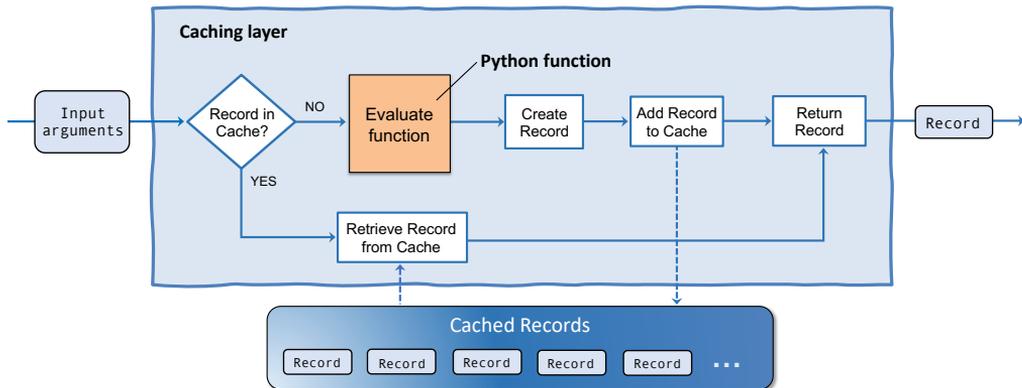


Figure 3: An Instruction is a Python function (orange) wrapped in a caching layer (light blue). When the function is called with a set of input arguments, the caching layer consults the cache to check if a Record for that exact function call already exists. In case of a cache hit, the Record is read and returned. In case of a cache miss, the function is evaluated and the Record is stored before it is returned.

In practice, this means that working with ASR and implementing new ASR Recipes becomes really simple.

The caching system works on a per-folder basis (similar to Git): a cache is initialized by the user in a folder and any instruction evaluated within this folder or sub-folders will utilize this cache. This mimics the behaviour of the MyQueue task scheduler so as to maximize the synergy between these tools. In practice, the “one-cache-per-folder” system works well together with a “one-material-per-folder” structure. The latter is currently still a requirement for utilizing the Database functionalities described in Section 5. However, the caching system can work with several atomic structures in the same folder as the cache can distinguish ASR tasks performed on different atomic structures. Data written by ASR is encoded as JSON.

Any Instruction can be called directly by the user (from Python or the CLI), but special importance is given to the “main Instruction”. The main Instruction usually provides the primary interface for the user to the Recipe and returns the final result of the Recipe. Other Instructions are called by the main Instruction and evaluated as needed. These may be Instructions implemented in the Recipe itself but may also be Instructions of separate Recipes. The main Instruction takes all input arguments required by the Recipe and uses them to call other Instructions.

Having multiple Instructions in a Recipe is usually motivated by code reusability or reduction of resources. The former is relevant when another Recipe needs to perform an identical Instruction (see Section on *Dependencies*). The latter is relevant when the task can be divided into Instructions with different resource requirements, in which case the separation may save computational time or resources. In particular, this is useful if a recalculation of a subset of the generated data is required.

The *input arguments* of an Instruction comprise all the information required to specify its task. When calls to an external simulation code are involved, the input arguments include a code specification, the computational parameters like k -point density, basis set specification, or exchange–correlation (xc) functional, as well the atomic structure.

An Instruction carries a *version number* to facilitate data migrations, i.e. transformations of the values or organisation of data produced by the Instruction. This may be required for backward compatibility when Instructions are updated, see Section 8.

3.2. Dependencies

It often happens that an Instruction can benefit from the functionality implemented by other Instructions. An example is the main Instruction of the “band structure” Recipe which calls an Instruction of the “ground state” Recipe to compute the electron density that the band structure should be based on. The caching layer logs whenever an Instruction requests data from another Instruction and uses that information to build a list of data dependencies. The data dependency list is stored in the Record object making it possible to trace what other pieces of data were used in the construction of the current result.

Implementation of data-dependencies in Recipes requires no extra coding. Whenever an Instruction calls another Instruction, the caching layer will *automatically* intercept the call and (1) determine if there exists a matching Record (cache hit/miss); (2) log the data dependency by registering the unique IDs and revision UIDS (see Section 8) of any dependent Records.

3.3. The Record object

The Record object is the basic data unit of ASR. It stores the results of Instructions together with metadata documenting how the results were obtained, and is used by the cache system to identify already performed Instruction calls. The Record object contains the following information:

- Result object (see Section 3.4)
- Input arguments, if relevant including
 - Atomic structure
 - Simulation code specification
 - Computational parameters
- Instruction version (see Section 8)
- External codes versions
- Randomly generated unique ID
- Dependencies (see Section 3.2)
- Revision History (see Section 8)
- Execution time and resources (number of cores)

To identify a cache hit/miss when evaluating an Instruction, the caching layer searches the cache for Records with matching Instruction name, version, and input arguments. A cache hit is then defined as the existence of a matching Record. A recursive comparison is used to compare input arguments with those from existing Records within a small numerical tolerance for floating point numbers. Any later evaluations of the Instruction with identical arguments will result in a cache hit.

3.4. The Result object

To store and document the result produced by a Recipe, ASR offers a Result object that wraps the actual result data (stored as a Python dictionary) in a simple data structure that also contains specification of the result data types along with short explanatory descriptions of the data. In addition, the Result object may implement methods to present itself in different formats, see below. Using the Result object is optional, but in practice all instructions that return more than a simple object or value utilizes a Result object for improved data documentation.

3.5. Presentation of results

The Result object may implement presentation options of the result data in various formats, for example text to terminal, figures, and web panels. The ASR Database and App modules draw on the Recipes' web panel implementation to create web pages for presenting, browsing, and distributing databases containing collected Result objects, see sections 5 and 6. This provides an efficient way of inspecting and sharing data as it is generated, which is highly practical for projects involving multiple collaborators.

3.6. General principles for Recipe development

To maintain and exploit the modular structure of ASR, the development of new Recipes should follow a few general design principles. First, the task performed by a Recipe should be well defined and clearly bounded to make it easy to use in different contexts. It should always be considered whether the Recipe could be split into smaller independent Recipes that could be useful individually. Additionally, it is encouraged that Recipes are designed/programmed so as to be as broadly applicable as possible, e.g. with respect to the type of material (structure dimensionality, chemical

composition, magnetic/non-magnetic, metallic/insulating, etc.). Any information required to define the simulation task should be included in the input argument of the Recipe, i.e. hard coding of parameters should be avoided. This should be done to ensure a flexible use and enhance the data provenance (input arguments are stored in the Records). Recipes should employ conservative parameter settings as default to ensure that the results are numerically well converged independent of the application, e.g. material type. Finally, in order to keep ASR Recipes simple and easy-to-read, and in order to enhance the modularity, code-extensive functionalities should be separated out into ASE functions and called from ASR whenever it is possible and sensible, i.e. when the ASE function is useful in other contexts than the specific Recipe.

4. The Recipe library

The ASR currently provides more than 40 complete Recipes allowing users to perform a broad range of materials simulation tasks ranging from construction and analysis of crystal structures over DFT calculations of thermodynamic, mechanical, electronic, magnetic, and optical properties to many-body methods for evaluating response functions, quasiparticle band structures, and collective excitations. A non-exhaustive list of available Recipes is provided in Table 1. It should be stressed that the list constitutes a snapshot of the current state of the Recipe library, which is continuously expanding. For example, we are currently developing Recipes for creating and modeling layered van der Waals structures and point defects in semiconductors.

Most of the currently implemented Recipes rely specifically on the GPAW[49] electronic structure code. As previously mentioned, we are currently working on a generalisation of the ASE Calculator interface to make the Recipes – or a large portion of them – simulation code-independent. Until then, usage of ASR with other simulation codes than GPAW is possible by porting of existing Recipes or development of new ones. The amount of work involved will depend on the type of Recipe and the state of the ASE interface for the specific simulation code.

A few specific examples of Recipe implementations are given in Section 6 where we outline the main computational steps and the final output of the `asr.bandstructure` and `asr.emasses` Recipes, respectively.

5. The ASR Database and App modules

The ASR Database and web App modules make it possible to package, inspect, share, and present ASR-driven projects easily and efficiently. The main tools and opportunities provided by these modules are described in more detail below.

5.1. Database

The ASR Database module can be used to collect Record objects from a directory tree into an ASE database. This is achieved by the command `asr database fromtree`. The procedure assumes a “one-material-per-folder” structure, relying on the existence of an atomic structure file in each folder to select Records pertaining to that atomic structure. The Database module proceeds to collect atomic structure-Record data sets and assign them to a particular row of an ASE database. We shall refer to such a database as an ASR database. Once an ASR database has been collected, it is possible to define key–value-pairs and relate property data to specific atomic structures.

The Database module also enables the reverse operation, that is, unpacking an existing ASR database to a directory tree containing Record objects. This is achieved by the command `asr database totree`. The function is useful when continuing a project, e.g. because existing data must be updated or new data must be added, for which the database is available but not the original directory tree. Moreover, it provides tools for merging and splitting databases.

It is possible to collect a database for any number of materials/Record objects – even for a single material – and thereby take advantage of the App tools for presenting and inspecting results in a browser with no extra efforts. However, collecting databases is obviously most powerful in cases involving many materials/properties where the database makes it possible to search and filter the data via the defined key–value-pairs.

The easy installation of ASE through the standard PyPI Python package manager makes the ASE database format highly accessible. Furthermore, the portability of an ASE database (via several backends, e.g. SQLite, PostgreSQL, MariaDB and MySQL) enables easy packaging and distribution of data among different parties.

5.2. Web App

The ASE provides a flexible and easily extensible database web application making it possible to present and inspect the content of an ASE database in a browser. ASR leverages this ASE functionality to customize the web application layout and provide more sophisticated features such as the automatic generation of web panels, generation of figures, and documentation of the presented data by utilizing the web panel data structures encoded in the `Result` objects. Normally a Recipe generates one web panel. However, panels gathering data from several Recipes may be created. One example of the latter is the “Summary” panel of the C2DB web pages discussed in the next section. In this case, a number of Recipes write data to a web panel data structure named “Summary” in their `Result` object. This information is stored in the database when collected. When generating the C2DB web pages from the C2DB database, the App constructs all web panels that are defined in the data pertaining to a particular material. If several Recipes have written to the same web panel, the data will be combined in an order controlled by a priority keyword written together with the web panel data.

5.2.1. Adding information fields

To enhance the accessibility of the data, it is possible to add an explanatory description to specific data entries, i.e. key–value pairs and data files, of an ASR database. These descriptions will appear as text boxes when clicking a “?”-icon placed next to the data on the web panels, see Fig. 5. General information boxes for web panels are always generated by ASR. They contain a customised field that can be manually edited, e.g. providing a short explanation of the data presented in the panel and/or links to relevant literature, and an automatically generated field listing the ASR Recipes that have produced data for the web panel and the key input parameters for the calculations. An example of such an information box is shown in Fig. 5.

5.2.2. Linking rows of databases

ASR provides functionality to create links between rows of the same, or different, ASR databases. This allows the developer to connect relevant materials when designing web panels such that the end user can move swiftly between them when browsing databases. For example, the `asr.convex_hull` Recipe creates the convex hull phase diagram of a material using an ASR reference database of stable materials (originally from from the OQMD[28]), and creates a table with links to all the materials on the phase diagram. Other examples, could be to link different defective versions of the same crystalline material or different isomers of the same material/molecule.

The links are defined in `links.json` files in the folders of the relevant materials. These files may be generated manually or automatically using the `Recipe.asr.database.treelinks`. When collecting the database, ASR reads the `links.json` file for each folder and stores the information in the Data dictionary of the corresponding row. The `Recipe.asr.database.crosslinks` then creates links between rows of the collected database and rows of other databases that are given as input to the Recipe. When generating the web panels, ASR uses this information to generate hyperlinks in HTML format and present them in the web application for each material.

6. High-throughput example: The C2DB

In this section we present an example of what can be accomplished by the ASR in the realm of data intensive high-throughput applications, showcase some examples of ASR-generated web panels, and discuss two specific Recipe implementations.

Historically, the ASR evolved in a symbiotic relationship with the Computational 2D Materials Database (C2DB) — an extensive database project organising various properties of more than 4000 two-dimensional (2D) materials[32, 48].

The C2DB distinguishes itself from existing computational databases of bulk[28, 29, 30] and low-dimensional[50, 15, 51] materials by the large number of physical properties available. These include convex hull diagrams, stiffness tensors, phonons (at high-symmetry points), projected density of states, electronic band structures with spin-orbit effects, effective masses, band topology indices, work functions, Fermi surfaces, plasma frequencies, magnetic anisotropies, magnetic exchange couplings, Bader charges, Born charges, infrared polarisabilities, optical absorption spectra, Raman spectra, and second harmonics generation spectra. The use of beyond-DFT theories for excited state properties (GW band structures and BSE absorption for selected materials) and Berry-phase techniques for band topology and polarization quantities (spontaneous polarization, Born charges, piezoelectric tensors), are other unique features of the C2DB.

Building the first version of C2DB without a fully functioning workflow framework was a long and painstaking endeavour, but absolutely critical for the successful development of the ASR. Today, the entire C2DB project can be generated by a single (MyQueue) Python workflow script comprising a sequence of ASR Recipe calls and simple Python code for controlling and directing the workflow via statements like `“if band_gap > 0:”`. Relying on the MyQueue task scheduler (see Section 7.3), generation of the C2DB is accomplished by the single command `“mq workflow c2db_workflow.py tree/**/**/”`, which will submit the C2DB workflow in folders matching the pattern `tree/**/**/`. With the current C2DB workflow, this statement will launch up to 23 unique Instructions for each of the 4047 materials amounting to a total of 59822 individual aiES calculations (some Recipes like phonon and stiffness calculations launch multiple aiES calculations). When the current workflow is run with the GPAW code, about 258 calculations are unsuccessful (most often due to convergence errors in the self-consistency DFT cycle) corresponding to a success rate of 99.5%.

Apart from the data provenance control that ensures the documentation and reproducibility of the data, there are two aspects of the ASR that are particularly crucial for making high-throughput computations work efficiently in practice. First, the caching functionality ensures that Recipes which have already been performed are automatically skipped by ASR (unless something in the input for a Recipe has changed since it was last executed). This means that only a single workflow script needs to be maintained and submitted every time something has been changed, e.g., new materials have been added, the workflow script has been updated, it has been decided to rerun certain tasks with new parameters, or a Recipe has been modified. Such functionality is essential because running and maintaining high-throughput projects inevitably requires that subsets of calculations are repeated at different points in time. Secondly, the carefully designed and well tested Recipes minimise the number of unsuccessful calculations and the risk of human errors.

6.1. Recipe and web page examples

Below we present a few examples of output generated by the ASR-C2DB workflow (for a full impression we refer the reader to the C2DB website).

6.1.1. Search page

Fig. 4 shows the C2DB search page, which consists of a search/filtering section followed by a list of the database rows presented by a selected number of key-value pairs. Clicking one of the highlighted key names once (twice) will sort the rows in increasing (decreasing) order of that key. Which keys should be shown by default can be customized,

Computational 2D materials database



Example: 'MoS2' OR 'gap>0,ehull<0.1'

Stoichiometry (A, AB2, A2B3, ...):

Material class:

Dynamically stable (phonons):

Dynamically stable (stiffness):

Thermodynamic stability: -

Is magnetic:

Band gap range [eV]: -

Help with constructing advanced search queries ...
Toggle list of keys ...

Displaying rows 1-25 out of 1583 (direct link) Rows: 25 Add Column

Formula X	Space group X	Magnetic X	Heat of formation X	Band gap X	Crystal type X
Ca4As4	P2_1/c	False	-0.743	0.998	AB-14-e
Mn2Se2	P4/nmm	True	-0.314	0.000	AB-129-bc
O8Te4	P-1	False	-1.087	2.667	AB2-2-i
Ru2F8	P2_1/c	True	-1.855	0.628	AB4-14-aa

Figure 4: The search page of C2DB with the first few rows of the database shown below. The default web page generated by ASR includes only the top most search field, but the panel can be customized by additional fields and buttons for more convenient data filtering.

but the user can always add extra keys via the “Add column” button. By default, the search page generated by the ASR App module will contain only the search field in the upper section, but additional fields or buttons may be added for easy filtering according to the most relevant parameters.

6.1.2. “Summary” panel

Fig. 5 shows the C2DB web page for monolayer MoS₂. All the web panels produced by the various Recipes of the workflow are seen, but only the “Summary” panel is unfolded. This panel is designed to provide an overview of the most basic properties of the material, and gathers data from the Result objects generated by the following Recipes: `asr.gs`, `asr.gw`, `asr.hse`, `asr.phonons`, `asr.magstate`, `asr.stiffness`, `asr.convex_hull`, and `asr.structureinfo`.

Fig. 5 also shows the information box of the “Effective masses” web panel. It contains a short explanation of the effective mass tensor and how it is evaluated by the Recipe as well as a link to a relevant paper. The automatically generated part shows that the panel contains data generated by the `asr.emasses` Recipe. The two fields at the top of the page “Download raw data” and “Browse raw data” provide access to the entire data set comprised by all Result objects of the specific material entry of the database.

6.1.3. “Band structure” Recipe

As another example, Fig. 6 shows the “Electronic band structure” panel for monolayer CrW₃S₈ as calculated and presented by the Recipe `asr.bandstructure`. The band structure is calculated with the PBE xc-functional including spin-orbit interactions. The out-of-plane spin projections of the states is shown by the color code. The main computational steps carried out by this Recipe are:

- Perform a self-consistent ground state calculation (by calling the `calculate` Instruction of the ground state Recipe `asr.gs`) to obtain a converged electron density.

MoS₂Download raw data Browse raw data 

Summary

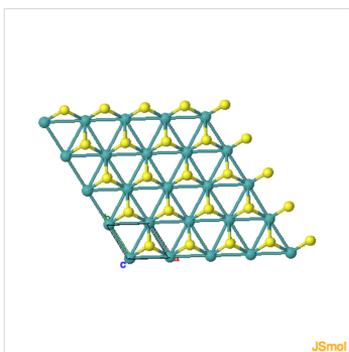
Structure info	Value
Crystal type 	AB2-187-bi
Material class 	TMDC-H
Space group 	P-6m2
Space group number 	187
Point group 	-6m2
Related ICSD id 	38401
Related COD id 	9007661
Reported DOI 	10.1103/PhysRevLett.105.136805

Stability

Thermodynamic 	HIGH
Dynamical (phonons) 	HIGH
Dynamical (stiffness) 	HIGH

Electronic properties

Magnetic 	False
Band gap (PBE) 	1.58 eV
Band gap (HSE) 	2.09 eV
Band gap (G0W0) 	2.53 eV



Download  Unit cell 

Axis	x (Å)	y (Å)	z (Å)	Periodic
1	3.184	0.000	0.000	True
2	-1.592	2.757	0.000	True
3	0.000	0.000	18.127	False

Lengths (Å):	3.184	3.184	18.127
Angles (°):	90.000	90.000	120.000

Thermodynamic stability

Stiffness tensor

Phonons

Basic electronic properties (PBE)

Electronic band structure (PBE)

Projected band structure and DOS (PBE)

Effective masses (PBE)

Electronic band structure (HSE)

Electronic band structure (G0W0)

Born charges

Optical polarizability (RPA)

Infrared polarizability (RPA)

Raman spectrum

Optical absorption (BSE and RPA)

Bader charges

Piezoelectric tensor

General panel information

The effective mass tensor represents the second derivative of the band energy w.r.t. wave vector at a band extremum. The effective masses of the valence bands (VB) and conduction bands (CB) are obtained as the eigenvalues of the mass tensor. The latter is determined by fitting a 2nd order polynomial to the band energies on a fine k-point mesh around the band extrema. Spin-orbit interactions are included. The "parabolicity" of the band is quantified by the mean absolute relative error (MARE) of the fit to the band energy in an energy range of 25 meV.

Relevant article(s):

- S. Hastrup et al. The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals, 2D Mater. 5 042002 (2018).

Relevant recipes

This panel contains information calculated with the following ASR Recipes:

- asr.emasses

Figure 5: Screenshot of the web page for monolayer MoS₂ from the C2DB project (only the "Summary" panel is unfolded). The panel presents data from the Result objects generated by the following Recipes: `asr.gs`, `asr.gw`, `asr.hse`, `asr.phonons`, `asr.magstate`, `asr.stiffness`, `asr.convex_hull`, `asr.structureinfo`.

- Determine crystal symmetries and corresponding band path (uses ASE functionalities).
- Calculate the Kohn–Sham eigenvalues along the band path. For magnetic materials, this step calls the Recipe `asr.magnetic_anisotropy` to obtain the magnetic easy axis for evaluating spin projections.
- Call the main Instruction of the ground state Recipe to get the Fermi level (in 3D) or the vacuum level (in <3D) for use as zero-point energy for the band structure.

In addition to these computational steps, the main Instruction of the Recipe formats two figures to present the band structure itself and the Brillouin zone with the band path and the positions of the valence band maximum (VBM) and conduction band minimum (CBM). Note that the position of the VBM and CBM, as well as a number of other properties like the band gap and band edge energies (not shown), are determined by the Recipe `asr.gs`, which is called by `asr.bandstructure`.

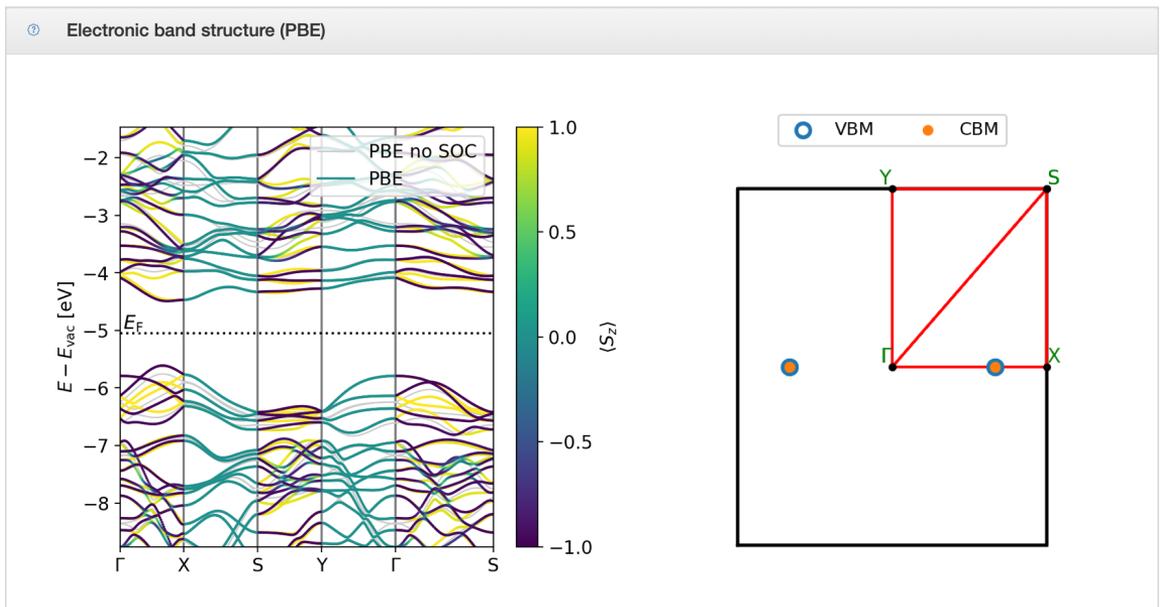


Figure 6: Screenshot of the “Band structure” panel for monolayer CrW_3S_8 from the C2DB project. The web panel contains data computed by the `asr.bandstructure` Recipe.

6.1.4. “Effective masses” Recipe

Fig. 7 shows a screenshot of the “Effective masses” panel for monolayer CrW_3S_8 generated by the Recipe `asr.emasses`. The effective mass tensor is calculated with the PBE xc-functional including spin–orbit interactions. The color code represents the spin projections along the z-axis. In addition to the effective masses themselves, the Recipe evaluates a “band parabolicity” parameter defined as the mean absolute relative error (MARE) between the parabolic fit and the true bands in an energy range of 25 meV. The main computational steps carried out by this Recipe involve three subsequent k -point grid refinements; specifically:

- Perform a self-consistent ground state calculation on a uniform k -point grid (by calling the `calculate` Instruction of the Recipe `asr.gs`) to obtain a converged electron density as well as Kohn–Sham band energies.
- Locate the preliminary positions of the VBM and CBM and calculate band energies on a higher-density k -point grid around the VBM and CBM to locate the VBM and CBM positions with higher accuracy.
- Define final high-density k -point grids in the vicinity of the VBM and CBM points, and calculate band energies.

- Locate VBM and CBM and fit bands by second-order polynomial using band energies in an energy range of 1 meV from the band extremum.
- Calculate band structures for the web panel and evaluate the “parabolicity parameter”.

It should be noted that even though effective mass calculations appear to be a simple task, it is surprisingly tricky to design a scheme that performs efficiently, robustly, and accurately across all types of band structures including flat bands, highly dispersive bands, highly anisotropic bands, and bands exhibiting complex spin-orbit effects like Rashba splittings.

6.1.5. General comments

In contrast to the “Summary” panel, which has been customized for the C2DB project (that is, the web panel sections of the relevant Recipes have been appropriately adjusted), the “Electronic band structure” and “Effective masses” panels are the default web panels produced by the `asr.bandstructure` and `asr.emasses` Recipes, respectively.

The examples given here concern two-dimensional (2D) materials. However, the Recipes `asr.bandstructure` and `asr.emasses` (like all other Recipes of the current ASR library) apply also to 1D and 3D materials, as well as 0D where it is meaningful. As mentioned in Section 3.6, this kind of generality should always be strived for when designing Recipes. Achieving this may be straightforward or more involved depending on the Recipe. The Recipe for the stiffness tensor represents an easy case, where the dimensionality merely dictates the number of axes along which the material must be strained. The Recipe for the band structure is more involved in this regard, as the determination of the band path requires separate treatments in 2D and 3D as does the determination of the spin projection axis (in 2D the out-of-plane direction is a natural choice while in 3D the magnetic easy axis is more appropriate).

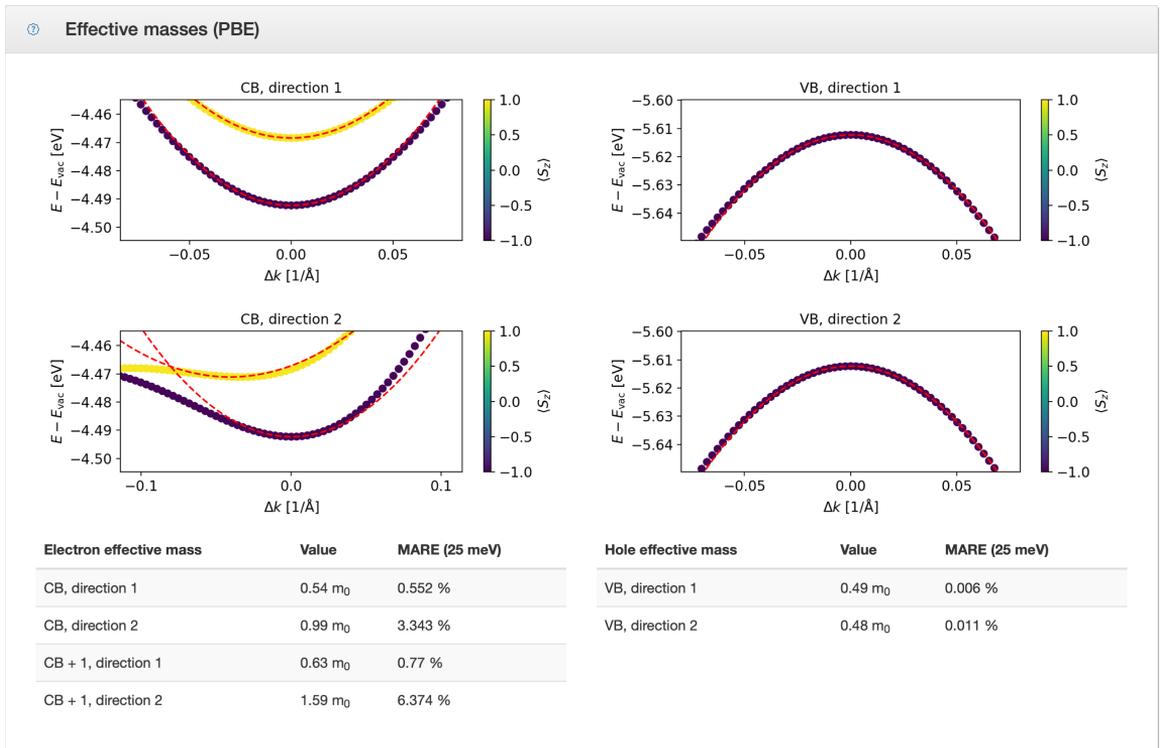


Figure 7: Screenshot of the “Effective masses” panel for monolayer CrW_3S_8 from the C2DB project. The panel contains data computed by the `asr.emasses` Recipe.

7. User interfaces

The ASR can be used via four different interfaces, c.f. Fig. 1: A command line interface (CLI), a Python interface, a task scheduling front-end, and an app-based interface. Below we describe each interface in more detail.

7.1. The CLI

The CLI provides convenient commands for easy interaction with ASR via the `cache` and `run` subcommands. The `cache` subcommand allows inspection of the `Records` stored in the cache, in particular their `Result` data. For example, `$ asr cache ls name=asr.gs` will list all `Records` produced by the “ground state” Recipe. The `run` subcommand can be used to execute Recipes directly from the command line. For example, `$ asr run asr.gs` will run the *ground state* Recipe.

7.2. Python interface

The Python scripting interface allows inspection of `Records` and execution of Recipes directly from Python. This makes it possible to implement more complex logic and integrate directly with ASE and any other tools in the user’s Python toolkit.

7.3. MyQueue interface

For high-throughput computations, ASR can be used in combination with a workflow manager that can handle the interaction with the scheduler of the supercomputer, such as Fireworks[40] or MyQueue[47]. The latter is a personal, decentralized, and lightweight front-end for schedulers (currently supporting SLURM, PBS, and LSF), which has been co-designed with ASR. MyQueue has a command line interface, which allows for submission of thousands of jobs in one command and provides easy-to-use tools for generating an overview of the status of jobs (‘done’, ‘queued’, ‘failed’ etc.). It also has a Python interface that can be used to define workflows. A Python script defines a dependency tree of *tasks* that MyQueue will submit without user involvement. The dependencies take the form: “if task X is done then submit task Y”. MyQueue works directly with folders and files, which makes it transparent and easy to use. Together ASR and MyQueue provide a powerful and extremely flexible toolkit for high-throughput materials computations.

Individual Instructions of the Recipes may be defined as separate MyQueue tasks, such that computational resources can be specifically dedicated each Instruction ensuring a flexible and efficient execution of any workflow. It is, however, not a requirement to specify resources on a per Instruction basis, in which case the resources specified for the main Instruction will apply to all Instructions of the Recipe.

7.4. App interface

The App interface is a web-based read-only interface that allows the user to present and inspect the data stored in an ASE database on a local or public network. Distributing the data on a local network is convenient for larger projects and/or projects involving several users, as it allows for easy sharing and monitoring of the data as the project evolves. Once a project is finalized, the App may be used as a platform to present the data to the world via web pages. The data presentation used by the App is defined in the `Result` object of the Recipes.

8. Data maintenance

It sometimes happens that a Recipe, or one of its Instructions, has to be updated, e.g. because a bug has been detected or it has been found appropriate to store additional metadata. Such updates may imply that previously generated `Records` are no longer consistent with the current implementation of the Recipe. Depending on the nature of the change made to the Recipe, it may be possible to update the `Record` objects without rerunning the Recipe (data migration) or it may be necessary to rerun the entire Recipe or some of its Instructions (data regeneration).

To support the migration of data, ASR implements a simple versioning system for Instructions. An Instruction is associated with a integer version number which is stored in the `Record` and identifies the version of the Instruction at the time of creation. When an Instruction is changed, its version number may be increased by the developer. Since the caching layer matches the current Instruction version number against `Records` in the cache (see Section 3.3), older `Records` would no longer yield cache hits and are then said to be invalidated.

To facilitate the migration of invalidated `Records`, it is possible to specify `Migrations` that can be associated with an Instruction and thereby provide a way to bring old `Records` up to date. In practice, a `Migration` bundles a `Record` transformation function, a unique migration ID and a human readable description of the effect of the migration, see

Fig. 8. In general, a Record transformation function induces a change to a Record. For example, this could be to convert a Record of version n to a later version $n + 1$ without rerunning the Instruction, but in general the effect of the transformation could be anything. Use of transformation functions is typically possible when the update involves changes to metadata and/or data restructuring while the actual result of the Instruction is unchanged.

When a Migration is applied to a Record, a Revision object is produced. A Revision contains a randomly generated UID, the UID of the applied Migration, an explanatory description of the changes made to the Record, and an automatically generated list of the Record entries that were changed, added or deleted. The auto-generated list of changes is constructed by comparing the Record returned by the transformation function to the input Record.

Upon migration of a Record, a revision history is updated by the latest Revision and stored in the migrated Record. The revision history can be inspected by users to learn which revisions, if any, have previously been applied to a given Record.

A Selector is used to identify the Records to be migrated, e.g. based on the Instruction name and version number. The Selector is bundled together with a Migration into a MigrationSelector, which can determine whether a particular Record matches the selection criteria of the Selector. To migrate a Record, ASR searches through all Recipes to collect their MigrationSelectors (if they have any) and apply them to the Record to find a “migration strategy”, i.e., which Migrations to be applied and in which order. The migration strategy is then encoded in a MigrationStrategy, which couples a particular Record to an ordered list of Migrations. The particular MigrationStrategy can then be applied to the Cache to execute the migration of the associated Record.

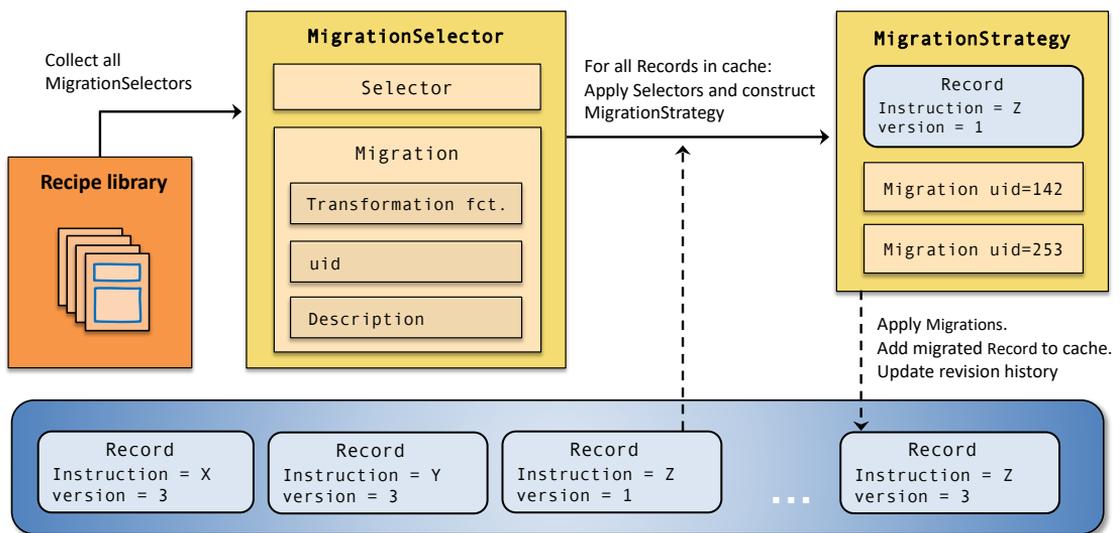


Figure 8: To support data maintenance, ASR provides migration tools for bringing Records up-to-date with the latest version of the Instructions that produced them thereby avoiding recalculations whenever possible. The ASR migration procedure consists of the main steps: (1) Collect all MigrationSelectors from all available Instructions. (2) Select the migratable Records of the cache. (3) Determine a migration strategy (an ordered list of Migrations) for each migratable Record. (4) Apply transformation functions to migrate the Records and add them to the cache. (5) Update the revision history by a Revision object that documents the effect of the migration.

ASR provides a simple CLI, via `asr cache migrate`, to analyse existing Records in the cache, and identify migratable Records.

Whenever the `asr` version used in a given project is upgraded, a project participant should identify migratable Records, migrate them and then rerun the project workflow. Up-to-date Records will then be taken directly from the cache, whereas the Instructions with invalidated Records and no associated Migration, i.e. Records that the

developer cannot migrate directly to the newest version, will be rerun.

In order to minimize the computational cost of bringing data up-to-date with ASR, developers are strongly advised to supply `Migrations` with their `Recipe` updates whenever possible.

To provide the best conditions for the long term deployment of ASR-generated data, the `asr` version of important projects should be upgraded regularly and the project workflow rerun. Obviously, this action may induce changes in the data. Whether this is acceptable or not is ultimately a strategic decision. However, for dynamic data projects, a regular version upgrade not only ensures that the data is of the highest quality, it also makes it easier for other parties to deploy the data because existing results (`Records`) can be reused directly with the newest version of ASR without having to rerun `Recipes` to bring the data up-to-date.

9. Data provenance

Simply stated, data provenance is the documentation of the circumstances under which a piece of data came into existence. This includes how the data originally was constructed, how the data has changed over time (also known as data-lineage) and a documentation of relevant system specifications such as architecture, operating system, important system packages, executables etc. If data provenance is handled perfectly, then data will in principle be reproducible, i.e. given access to exactly the same systems and software, any piece of data can be reproduced. In a scientific context, where reproducibility is key, data provenance is naturally very important.

In ASR, the basic unit of data is the `Record` object, which connects the result of an `Instruction` with various pieces of contextual metadata, see Section 3.3. Taken together, the metadata tell the story of how the original `Record` came into existence (`Instruction` name/version and input arguments), which other `Records` were implicitly used for the construction of this `Record` (`Dependencies`), what external package versions were used, and how the `Record` has transformed over time (`Revision` history). For simplicity, since it would be outside the scope of ASR, system information is not stored with the `Record`, which, in our experience, is not practically relevant for the purposes of ASR. As such, we characterize ASR as practically, but not perfectly, data provenant.

10. Documentation

ASR itself is documented on Read the Docs. The data is documented through the `Record` and `Result` objects, see previous Section on data provenance.

11. Technical specifications

Some technical specifications are listed in Table 2. ASR can be installed via pip using the command `pip install asr`.

ASR requires or is normally used with the following software:

- Python libraries: ASE, numpy, matplotlib, plotly, flask, click
- Computational and workflow software: GPAW or other ASE codes, MyQueue (SLURM/PBS/LSF)
- Optional extras: spglib, phonopy, and pymatgen (for `Recipes`); jinja, mysql or other ASE database backends

For community support see <https://asr.readthedocs.io/en/latest/src/contact.html>.

12. Summary and outlook

This article has introduced The Atomic Simulation Recipes (ASR) as an open source Python framework for developing materials simulation workflows and managing the data they produce.

To facilitate the transition to a paradigm of data-intensive science, ASR was designed to support the development of materials simulation workflows that operate in accordance with the FAIR data principles, by providing tools and concepts that are general enough that they do not restrict the user whilst being concrete enough to make a real difference. The ASR achieves this through the notion of a `Recipe`: a general Python script that performs a well defined simulation task and is wrapped in a caching layer that logs all relevant metadata without involving the user. This construction places essentially no restrictions on the developer's freedom to design and control the workflow, but resolves the critical

and complex issue of keeping track of the data provenance. We stress that the core of ASR, i.e. the Recipe concept and the caching system, is fully simulation code independent. In particular, it is not tied to materials simulations and could potentially be useful in other areas of computational science.

Beyond the built-in data documentation, there are many benefits of using standardized, well tested, and well documented Recipes. For example, it saves time and promotes a more sustainable scripting culture by reducing the need for individual researchers to write and maintain their own personal scripts (which can be hard for other to read and are often lost when the developer leaves the group). Furthermore, it reduces the risk of human errors and lowers the barrier for researchers to undertake simulation tasks with which they have little prior experience.

The fact that Recipes are independent units with own data provenance control implies that they can be freely combined to create advanced workflows using Python scripting for maximal flexibility. Such workflows can be executed on supercomputers using a workflow management software that supports a Python interface. To this end, we have developed the MyQueue[47] task manager that works as a front-end to the most common schedulers (currently SLURM, PBS, and LSF). While MyQueue will resubmit jobs that have timed out or crashed due to lack of memory, code-related failures must be handled manually. In the future, ASR should integrate more closely with MyQueue to permit that errors from the simulation codes are automatically analysed and reacted upon. Along the same lines, an automated estimation of the HPC resources (time/memory/nodes) required by individual tasks could limit the number of failed jobs and improve the utilization of resources.

The current Recipe library already covers a wide range of materials simulation tasks and more are continuously being added. Of special importance are Recipes for advanced beyond-DFT calculations where the benefits in terms of a lowered user barrier, improved data quality, and increased utilization of computing resources, are particularly large. The Recipe concept should also be advantageous for implementation of machine learning methods that could integrate with ASR databases and “standard” Recipes to make for more intelligent and computationally efficient workflows.

The ASR makes extensive use of the Atomic Simulation Environment (ASE) as a toolkit to process atomistic calculations. In particular, ASE is used as a front-end for ASR to communicate with external simulation codes. This has the clear advantage that ASR can become decoupled from the simulation codes. This decoupling is currently not in place, and the majority of the existing Recipe implementations contain code parts that are specific to the GPAW electronic structure code. To make ASR fully simulation code-independent, the ASE Calculator interfaces must be further generalized. This includes extensions of the interfaces to access outputs of calculations as well as a systematic mechanism to control multi-step tasks. The adaptation of this interface to multiple codes will eventually require a community effort that we hope many code developers will take part in. Until then, Recipes must to some extent be code specific.

13. Acknowledgments

We acknowledge funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program Grant No. 773122 (LIMA) and Grant agreement No. 951786 (NOMAD CoE).

References

- [1] Greeley J, Jaramillo T F, Bonde J, Chorkendorff I and Nørskov J K 2006 *Nature Materials* **5** 909–913
- [2] Madsen G K 2006 *Journal of the American Chemical Society* **128** 12140–12146
- [3] Curtarolo S *et al.* 2013 *Nature Materials* **12** 191–201
- [4] Kirklin S, Meredig B and Wolverton C 2013 *Advanced Energy Materials* **3** 252–262
- [5] Ørnsø K B, Garcia-Lastra J M and Thygesen K S 2013 *Physical Chemistry Chemical Physics* **15** 19478–19486
- [6] Zhang Z *et al.* 2019 *ACS Omega* **4** 7822–7828
- [7] Chen W *et al.* 2016 *J. Mater. Chem. C* **4** 4414–4426
- [8] Hachmann J, Olivares-Amaya R, Atahan-Evrenk S, Amador-Bedolla C, Sánchez-Carrera R S, Gold-Parker A, Vogt L, Brockway A M and Aspuru-Guzik A 2011 *The Journal of Physical Chemistry Letters* **2** 2241–2251
- [9] Bhattacharya S and Madsen G K 2015 *Physical Review B* **92** 085205
- [10] Castelli I E *et al.* 2012 *Energy and Environmental Sciences* **5** 5814–5819
- [11] Hautier G, Miglio A, Ceder G, Rignanese G M and Gonze X 2013 *Nature Communications* **4** 1–7
- [12] Yu L and Zunger A 2012 *Physical Review Letters* **108** 068701
- [13] Kuhar K, Pandey M, Thygesen K S and Jacobsen K W 2018 *ACS Energy Letters* **3** 436–446
- [14] Aykol M, Kim S, Hegde V I, Snyder D, Lu Z, Hao S, Kirklin S, Morgan D and Wolverton C 2016 *Nature communications* **7** 1–12
- [15] Mounet N, Gibertini M, Schwaller P, Campi D, Merkys A, Marrazzo A, Sohier T, Castelli I E, Cepellotti A, Pizzi G and Marzari N 2018 *Nature Nanotechnology* **13** 246–252

- [16] Chen L Q, Chen L D, Kalinin S V, Klimeck G, Kumar S K, Neugebauer J and Terasaki I 2015 *npj Computational Materials* **1** 1–2
- [17] Rupp M, Tkatchenko A, Müller K R and Von Lilienfeld O A 2012 *Physical Review Letters* **108** 058301
- [18] Lee J, Seko A, Shitara K, Nakayama K and Tanaka I 2016 *Physical Review B* **93** 115104
- [19] Xie T and Grossman J C 2018 *Physical Review Letters* **120** 145301
- [20] Ghiringhelli L M, Vybiral J, Levchenko S V, Draxl C and Scheffler M 2015 *Physical Review Letters* **114** 105503
- [21] Jørgensen P B, del Río E G, Schmidt M N and Jacobsen K W 2019 *Physical Review B* **100** 104114
- [22] Ghosh K, Stuke A, Todorović M, Jørgensen P B, Schmidt M N, Vehtari A and Rinke P 2019 *Advanced Science* **6** 1801367
- [23] Deringer V L and Csanyi G 2017 *Physical Review B* **95** 094203
- [24] Lorenz S, Gross A and Scheffler M 2004 *Chemical Physics Letters* **395** 210–215
- [25] Behler J and Parrinello M 2007 *Physical Review Letters* **98** 146401
- [26] Artrith N and Urban A 2016 *Computational Materials Science* **114** 135–150
- [27] Thygesen K S and Jacobsen K W 2016 *Science* **354** 180–181
- [28] Saal J E, Kirklin S, Aykol M, Meredig B and Wolverton C 2013 *JOM* **65** 1501–1509
- [29] Jain A *et al.* 2013 *APL Materials* **1** 011002
- [30] Curtarolo S, Setyawan W, Hart G L, Jahnatek M, Chepulskii R V, Taylor R H, Wang S, Xue J, Yang K, Levy O *et al.* 2012 *Computational Materials Science* **58** 218–226
- [31] Draxl C and Scheffler M 2019 *Journal of Physics: Materials* **2** 036001
- [32] Hastrup S, Strange M, Pandey M, Deilmann T, Schmidt P S, Hinsche N F, Gjerding M N, Torelli D, Larsen P M, Riis-Jensen A C *et al.* 2018 *2D Materials* **5** 042002
- [33] Borysov S S, Geilhufe R M and Balatsky A V 2017 *PLoS ONE* **12** e0171501
- [34] Winther K T, Hoffmann M J, Boes J R, Mamun O, Bajdich M and Bligaard T 2019 *Scientific Data* **6** 1–10
- [35] Talirz L, Kumbhar S, Passaro E, Yakutovich A V, Granata V, Gargiulo F, Borelli M, Uhrin M, Huber S P, Zoupanos S *et al.* 2020 *Scientific Data* **7** 1–12
- [36] Armiento R 2020 *Machine Learning Meets Quantum Physics* 377–395
- [37] Himanen L, Geurts A, Foster A S and Rinke P 2019 *Advanced Science* **6** 1900808
- [38] Kohn W and Sham L J 1965 *Physical Review* **140** A1133
- [39] Wilkinson M D, Dumontier M, Aalbersberg I J, Appleton G, Axton M, Baak A, Blomberg N, Boiten J W, da Silva Santos L B, Bourne P E *et al.* 2016 *Scientific Data* **3** 1–9
- [40] Jain A *et al.* 2015 *Concurrency and Computation* **27** 5037–5059
- [41] Pizzi G, Cepellotti A, Sabatini R, Marzari N and Kozinsky B 2016 *Computational Materials Science* **111** 218–230
- [42] Mathew K, Montoya J H, Faghaninia A, Dwarakanath S, Aykol M, Tang H, Chu I h, Smidt T, Bocklund B, Horton M *et al.* 2017 *Computational Materials Science* **139** 140–152
- [43] Larsen A H, Mortensen J J, Blomqvist J, Castellì I E, Christensen R, Dułak M, Friis J, Groves M N, Hammer B, Hargus C *et al.* 2017 *Journal of Physics: Condensed Matter* **29** 273002
- [44] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2013 *Computational Materials Science* **68** 314–319
- [45] Togo A and Tanaka I 2018 *arXiv:1808.01590*
- [46] Togo A, Chaput L and Tanaka I 2015 *Physical Review B* **91** 094306
- [47] Mortensen J, Gjerding M and Thygesen K 2020 *The Journal of Open Source Software* **5** 1844
- [48] Gjerding M, Taghizadeh A, Rasmussen A, Ali S, Bertoldo F, Deilmann T, Holguin U, Knøsgaard N, Kruse M, Manti S *et al.* 2021 *arXiv:2102.03029*
- [49] Enkovaara J, Rostgaard C, Mortensen J J, Chen J, Dułak M, Ferrighi L, Gavnholt J, Glinsvad C, Haikola V, Hansen H *et al.* 2010 *Journal of Physics: Condensed Matter* **22** 253202
- [50] Ashton M, Paul J, Sinnott S B and Hennig R G 2017 *Physical Review Letters* **118** 106101
- [51] Zhou J, Shen L, Costa M D, Persson K A, Ong S P, Huck P, Lu Y, Ma X, Chen Y, Tang H *et al.* 2019 *Scientific Data* **6** 1–10

Table 1

List of Recipes currently implemented in the ASR library. Most of the Recipes depend explicitly on the GPAW electronic structure code. The Recipes are grouped under thematic headings and listed in alphabetic order.

Recipe name	Description
Atomic structure	
asr.database.duplicates	Remove duplicate structures from a database
asr.database.rmsd	Root mean square distance between structures
asr.dimensionality	Dimensionality of covalently bonded substructures of a material
asr.push	Push atoms along specific phonon mode
asr.relax	Relax atomic structure
asr.setup.defects	Generate native point defects
asr.setup.displacements	Generate structures with a single displaced atom
asr.setup.magnetize	Initialize atomic magnetic moments
asr.setup.reduce	Reduce supercell to primitive cell
asr.setup.symmetrize	Symmetrize an atomic structure
asr.structureinfo	Extract structural information
Thermodynamic properties	
asr.chc	Constrained convex hull stability analysis
asr.convex_hull	Convex hull stability analysis
asr.defectformation	Formation energy of neutral point defect
asr.fere	Define elemental reference energies
Mechanical properties	
asr.phonopy	Phonon band structure and dynamical stability
asr.piezoelectrictensor	Piezoelectric tensor
asr.stiffness	Stiffness tensor
Electronic properties	
asr.bader	Bader charge analysis
asr.bandstructure	Kohn-Sham band structure
asr.berry	Various band topology invariants
asr.borncharges	Born effective charge tensor
asr.deformationpotentials	Deformation potentials (only for 2D)
asr.dos	Density of states
asr.emasses	Effective masses
asr.fermisurface	Fermi surface
asr.formalpolarization	Formal polarization phase
asr.gs	Electronic ground state
asr.gw	G_0W_0 quasiparticle band structure
asr.hse	HSE06 band structure
asr.pdos	Orbital projected density of states
asr.projected_bandstructure	Orbital projected Kohn–Sham band structure
Magnetic properties	
asr.exchange	Magnetic exchange coupling
asr.magnetic_anisotropy	Magnetic anisotropy
asr.magstate	Determine magnetic state
Optical properties	
asr.bse	Optical absorption from Bethe–Salpeter Equation (BSE)
asr.infraredpolarizability	Infrared polarizability (caused by vibrations)
asr.plasmafrequency	Plasma frequency (from intraband transitions)
asr.polarizability	Optical polarizability (caused by electrons)
asr.raman	Raman spectrum (first-order)
asr.shg	Second harmonics generation
asr.shift	Shift current

Source code	https://gitlab.com/asr-dev/asr
Releases	https://pypi.org/project/asr/
License	GNU GPLv3 or newer (free software)
Documentation	https://asr.readthedocs.io/en/latest/

Table 2
Technical specifications

A WLDA

A.1 Optimizations

A.1.1 WLDA

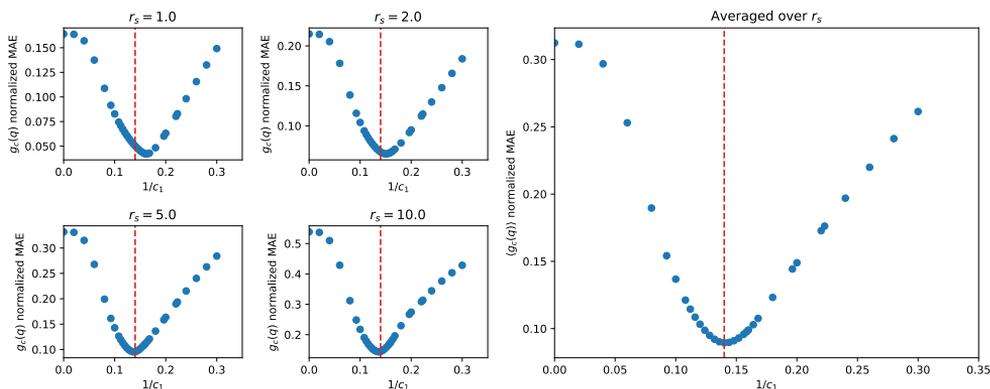


Figure A.1: The normalized mean absolute error (NMAE, defined in main text) between the WLDA functional using the exponential weight function for various c_1 values. The right plot shows the error averaged over all r_s . The vertical dashed, red lines shown the location of the minimum of the averaged errors in all cases. We see that the optimal value of the average error is close to the minima in all cases. The optimal c_1 value is 7.14286 with NMAE 0.089473.

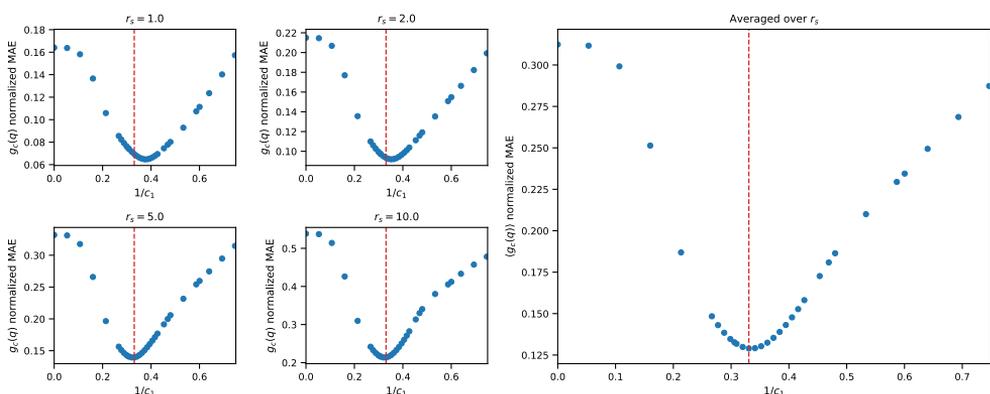


Figure A.2: The normalized mean absolute error (NMAE, defined in main text) between the WLDA functional using the Gaussian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 3.02419 with NMAE 0.128982.

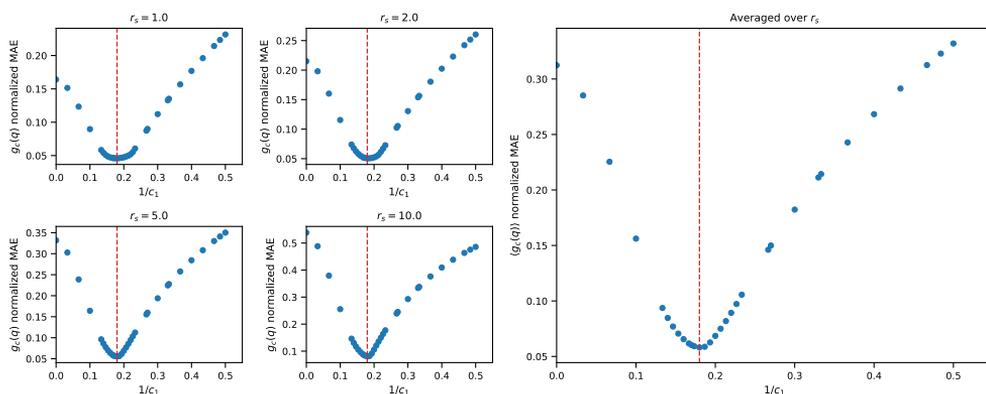


Figure A.3: The normalized mean absolute error (NMAE, defined in main text) between the WLDA functional using the Lorentzian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 5.5556 with NMAE 0.058134.

A.1.2 WLDax

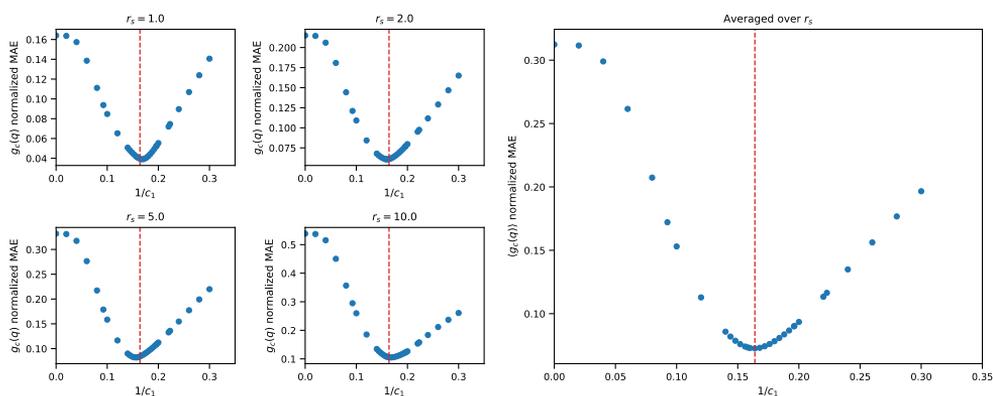


Figure A.4: The normalized mean absolute error (NMAE, defined in main text) between the WLDax functional using the exponential weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 6.09756 with NMAE 0.0725738.

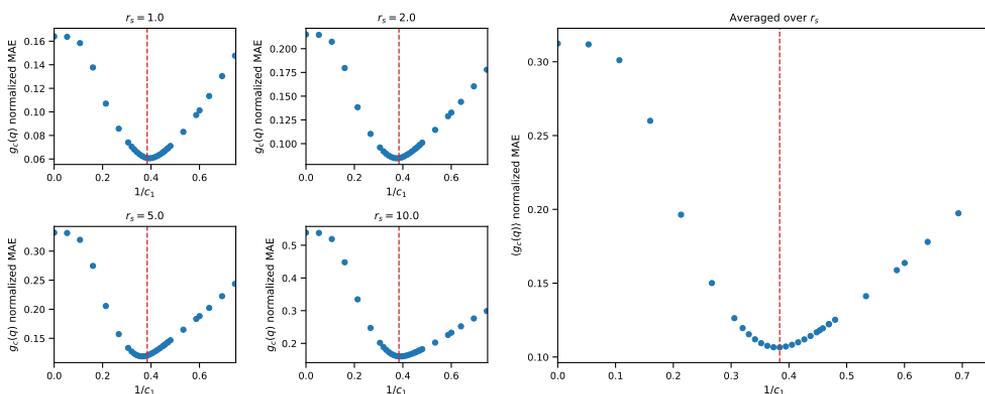


Figure A.5: The normalized mean absolute error (NMAE, defined in main text) between the WLDax functional using the Gaussian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 2.60417 with NMAE 0.10648.

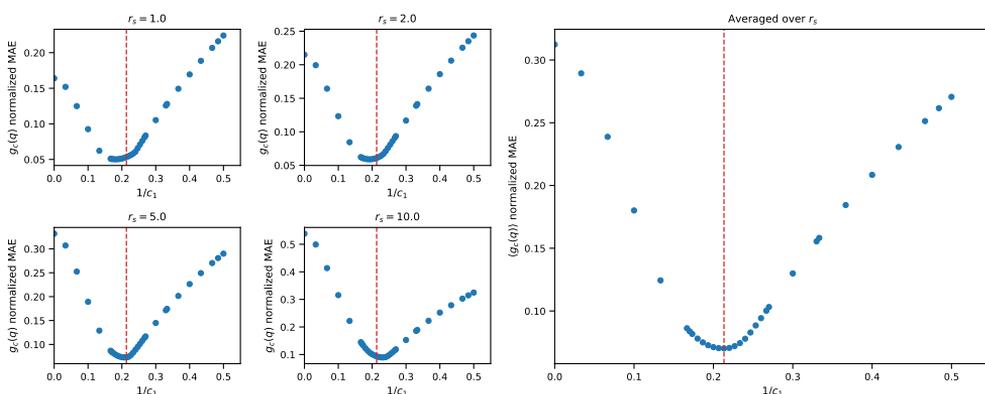


Figure A.6: The normalized mean absolute error (NMAE, defined in main text) between the WLDax functional using the Lorentzian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 4.6875 with NMAE 0.07015287.

A.1.3 rWLDA

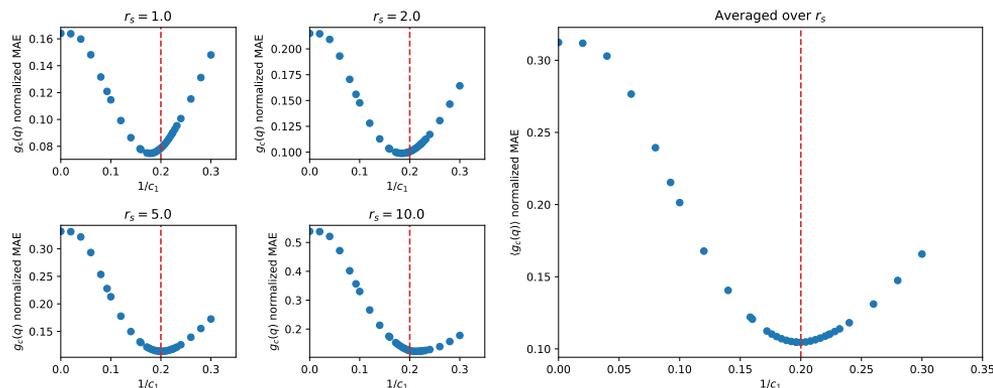


Figure A.7: The normalized mean absolute error (NMAE, defined in main text) between the rWLDA functional using the exponential weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 5.0 with NMAE 0.104596.

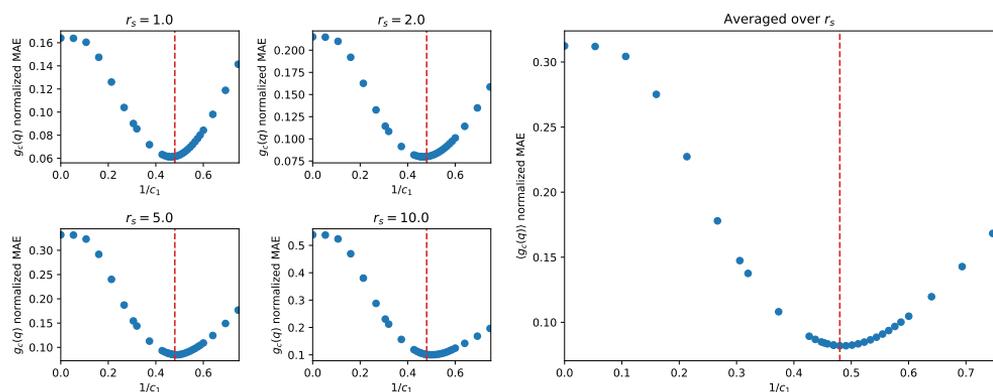


Figure A.8: The normalized mean absolute error (NMAE, defined in main text) between the rWLDA functional using the Gaussian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 2.08333 with NMAE 0.081950.

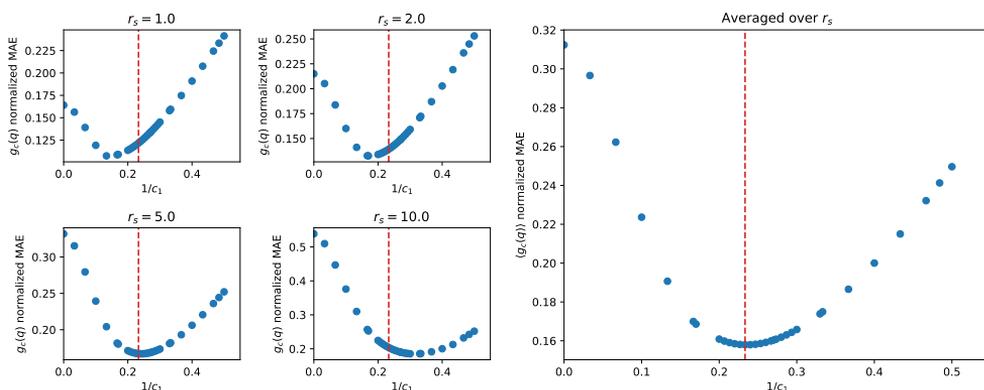


Figure A.9: The normalized mean absolute error (NMAE, defined in main text) between the rWLDA functional using the Lorentzian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 4.28571 with NMAE 0.157958.

A.1.4 rWLDax

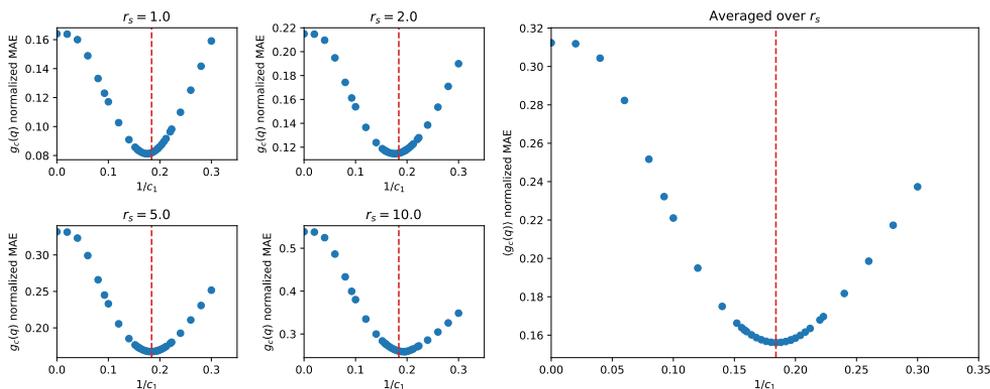


Figure A.10: The normalized mean absolute error (NMAE, defined in main text) between the rWLDax functional using the exponential weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 5.43478 with NMAE 0.1560827.

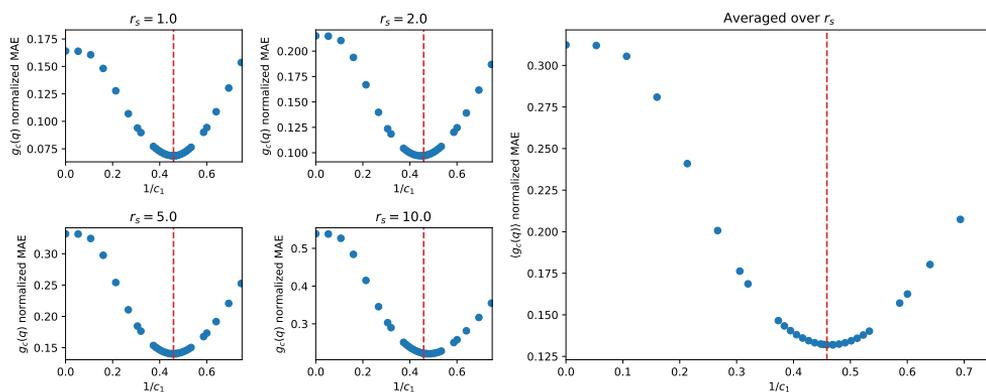


Figure A.11: The normalized mean absolute error (NMAE, defined in main text) between the rWLDax functional using the Gaussian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 2.18023 with NMAE 0.131995.

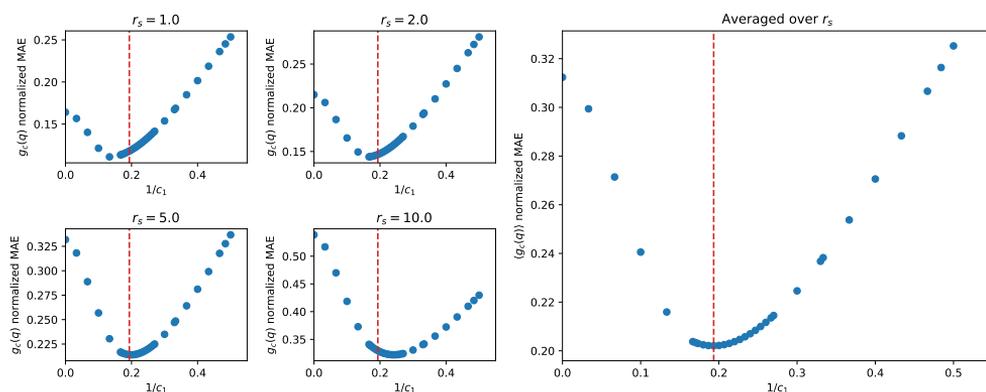


Figure A.12: The normalized mean absolute error (NMAE, defined in main text) between the rWLDax functional using the Lorentzian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 5.17241 with NMAE 0.202066.

A.1.5 fWLDA

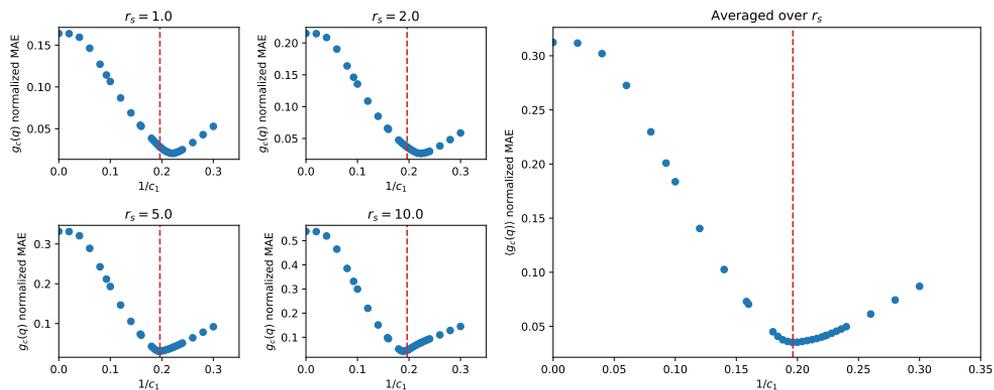


Figure A.13: The normalized mean absolute error (NMAE, defined in main text) between the fWLDA functional using the exponential weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 5.09393 with NMAE 0.035367.

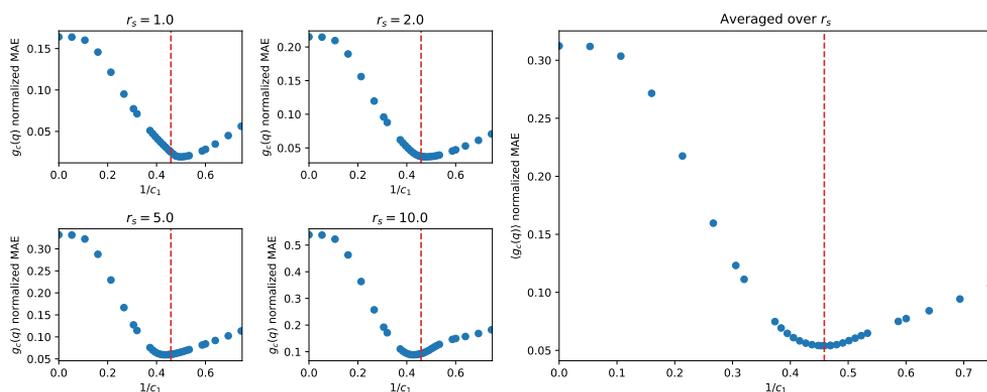


Figure A.14: The normalized mean absolute error (NMAE, defined in main text) between the fWLDA functional using the Gaussian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 2.08333 with NMAE 0.081950.

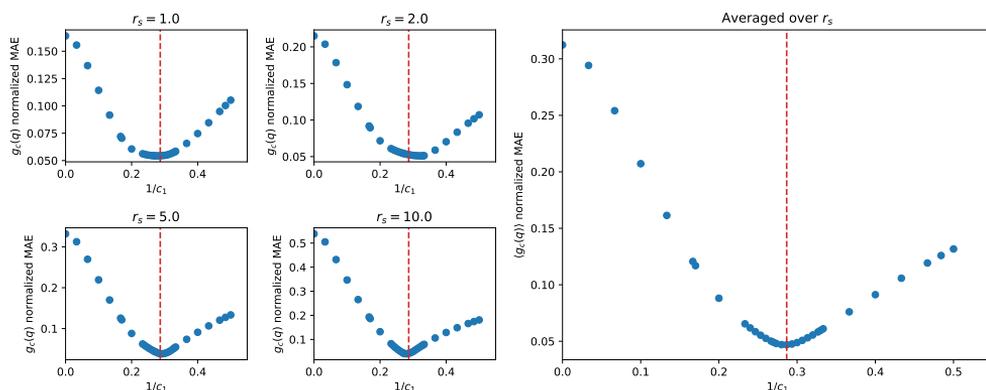


Figure A.15: The normalized mean absolute error (NMAE, defined in main text) between the fWLDA functional using the Lorentzian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 3.48837 with NMAE 0.046904.

A.1.6 fWLDAx

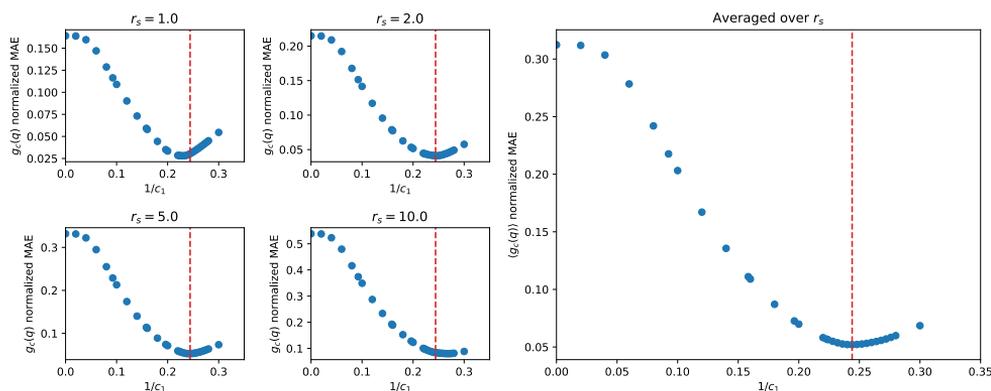


Figure A.16: The normalized mean absolute error (NMAE, defined in main text) between the fWLDAx functional using the exponential weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 4.09836 with NMAE 0.0520419.

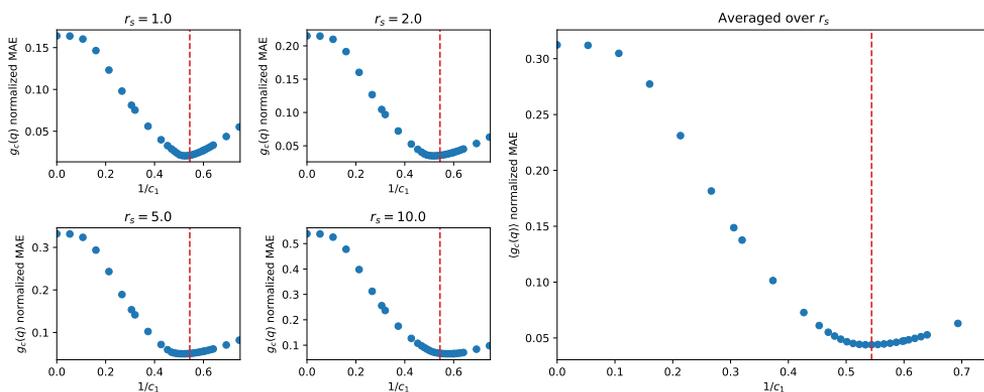


Figure A.17: The normalized mean absolute error (NMAE, defined in main text) between the fWLDax functional using the Gaussian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 1.83824 with NMAE 0.0439024.

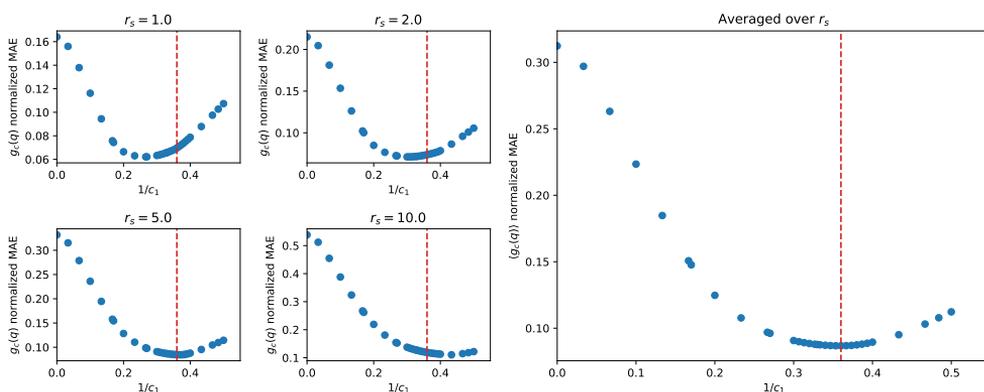


Figure A.18: The normalized mean absolute error (NMAE, defined in main text) between the fWLDax functional using the Lorentzian weight function for various c_1 values. See fig. A.1 for definitions. The optimal c_1 value is 2.77778 with NMAE 0.0867764.

A.2 All-electron plane-wave convergence

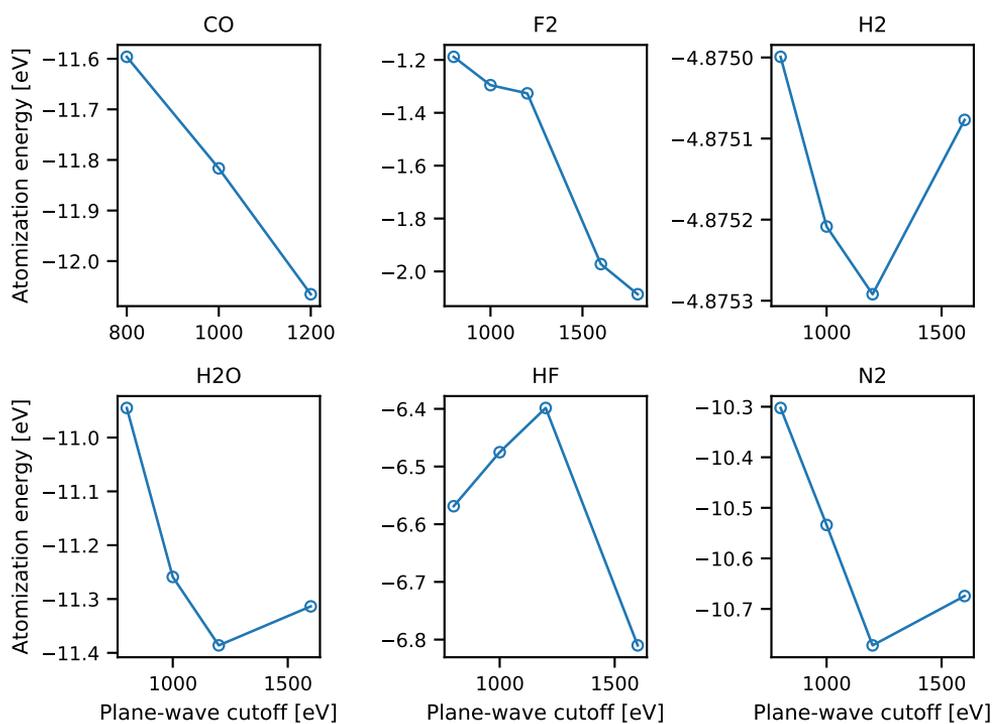


Figure A.19: Plane-wave convergence of WLDA for various small molecules using the all-electron density.

A.3 Atomization Results - All Variants

	ME	MAE	MARE
LDA	84.18	84.19	0.1984
LDAx	-20.62	36.41	0.1065
PBE	26.23	26.87	0.08238
WLDxA@Lor.	261 (329.7)	261 (329.7)	0.5688 (0.7163)
WLDA@Lor.	207.9 (257.7)	207.9 (257.7)	0.4551 (0.5613)
WLDxA@Exp.	45.15 (45.22)	45.4 (45.51)	0.1107 (0.1089)
WLDA@Exp.	60.05 (57.89)	60.05 (57.89)	0.1427 (0.1374)
WLDxA@Gau.	54.53 (49.95)	54.53 (49.95)	0.1328 (0.1228)
WLDA@Gau.	68.18 (63.97)	68.18 (63.97)	0.1618 (0.1531)
rWLDxA@Exp.	21.95 (28.38)	26.2 (32.84)	0.07088 (0.08081)
rWLDA@Exp.	21.32 (29.85)	26.54 (35.28)	0.07085 (0.08454)
rWLDxA@Gau.	26.8 (24.32)	28.04 (25.71)	0.07489 (0.06813)
rWLDA@Gau.	21.9 (20.74)	24.16 (23.2)	0.06555 (0.06098)
fWLDxA@Exp.	45 (63.52)	48.81 (67.32)	0.1144 (0.1473)
fWLDA@Exp.	50.98 (55.79)	51.35 (56.2)	0.1205 (0.1277)
fWLDxA@Gau.	35.87 (38.77)	37.3 (40.33)	0.09261 (0.0956)
fWLDA@Gau.	53.02 (51.37)	53.02 (51.37)	0.127 (0.1222)

Table A.1: Subset of G2/97 atomization energies (142 molecules) in kcal / mol. The values are calculated using the optimal c_1 parameters for each functional variant, see table 3.2. Results for spin-neutral Hartree is shown in parenthesis and other results are calculated with Hartree as exchange. This table includes all variants of WLDA.

B Bilayer Screening

B.1 Indirect to Direct Materials

We here show the full list of monolayers that have a energy above the convex hull of less than 50 meV and where the monolayer has an indirect gap and all bilayers have a direct gap. The monolayers are listed below with their C2DB [6] unique identifier (uid) through which they can be inspected at <https://cmrdb.fysik.dtu.dk/c2db/>.

1. As4-ff6eb167410f
2. Bi2Br6-2a1d71d82f8a
3. Bi2I6-433fccc74b5d
4. Ga2S2-6f3b2e422ac9
5. Sb2Te3-3e3e0acdba55
6. PbI2-82db29775962
7. SrCl2-67727fec854b
8. Zr2Br4-112846d9ccdd
9. BiBrSe-de5756e4fbfa
10. BiBrTe-304bc6a92d82
11. BiClSe-a80866a2c6b4
12. BiISe-70cbc0e44d36
13. Br2Ga2O2-f06d0758fdea
14. Br2Ir2S2-c0fadae216ee
15. BrIV-2da05db34435
16. Br2O2Sc2-7a1e1bf3f7d9
17. Br2O2TI2-c3c952d61f44
18. BrSbTe-18e62ba75259
19. Cl2Ga2O2-d5722b10ff79
20. Cl2Ir2S2-4064ccdd7b40
21. ClSbSe-0c0fbdaf8f4a
22. ClSbTe-da5fd2bb47af
23. Cl2S2Sc2-8dd7b2e2ca68
24. PbSb2Te4-8e5a0876861b

B.2 High Effective Mass Changes

Here we show the list of monolayers where one or more pairs of stackings are easily switchable (dipole change greater than 10^{-13} C/m), have an effective mass change of greater than 100% for at least one mass, and where the thermodynamic stability of the material is judged to be very high (energy above convex hull of less than 50 meV).

1. Ag₂Br₂Te₄-ce52fbc99f30
2. Ag₂Cl₂Se₄-55e92c5d7b58
3. SeBi₂S₂-5db44a94f3a6
4. Ag₂Br₂Se₄-c7e3d505588c
5. FeGa₂S₄-362cc8f15083
6. BrClV-878148656c6a
7. Ag₂Cl₂Te₄-d2ed9275e89c
8. Pt₂S₄-216f8f2dd8c1
9. BiBrS-49b7be14f786

B.3 Large Switchable Dipole Moment

Here we list the 28 monolayers that have at least one pair of stackings with a switchable dipole moment greater than 10^{-12} C/m and where the monolayers are highly thermodynamically stable.

1. SrCl₂-67727fec854b
2. PtO₂-6f43cd15097d
3. Ag₂Br₂Se₄-c7e3d505588c
4. HfSe₂-7a708c5759cf
5. Cl₂Rh₂Te₂-fb4fb079660b
6. Au₂Br₂Te₄-5e64ed168c9e
7. CaCl₂-643810e6bb76
8. MoSeTe-42eb12e7b656
9. MoSe₂-f61b14d398c7
10. ZrSe₂-7dc2a0a57b42
11. Ag₂Br₂Te₄-ce52fbc99f30
12. Ag₂Cl₂Se₄-55e92c5d7b58
13. MoSSe-de7ac5fc6945
14. BN-4a5edc763604

15. SrBr2-a17f01139eff
16. PbCl2-ec34fbc566ef
17. MoTe2-38a53176109a
18. WSe2-1cfbe6183886
19. FeO2-e25173381887
20. NiI2-0e87e38a2099
21. Fe2Cl6-a7111f341a6a
22. BaI2-080cc67ae6c9
23. WTe2-3c87365bc48c
24. SSeW-001e03f2c095
25. Ag2Cl2Te4-d2ed9275e89c
26. Ag2I2Te4-27e83673114d
27. SeTeW-6e2a4c6f4f57
28. MoS2-b3b4685fb6e1

B.4 Switchable Magnetic Order

The two materials mentioned in the main text that are easily switchable (dipole change greater than 10^{-13} C/m), and where the magnetic order changes between two stackings.

1. BrClV-878148656c6a
2. NiI2-0e87e38a2099