DTU Library

Calculating excitons, plasmons, and quasiparticles in 2D materials and van der Waals heterostructures

Topical Review

Thygesen, Kristian Sommer

Link back to DTU Orbit

Calculating excitons, plasmons, and quasiparticles in 2D materials and van der Waals heterostructures

You may also be interested in:

Band structure engineering in van der Waals heterostructures via dielectric screening: the GW method
Kirsten T Winther and Kristian S Thygesen

Light–matter interaction in transition metal dichalcogenides and their heterostructures
Ursula Wurstbauer, Bastian Miller, Eric Parzinger et al.

Plasmonics in Dirac systems: from graphene to topological insulators
Tobias Stauber

2D Materials Advances: From Large Scale Synthesis and Controlled Heterostructures to Improved Characterization Techniques, Defects and Applications
Zhong Lin, Amber McCreary, Natalie Briggs et al.

Two-dimensional hexagonal semiconductors beyond grapheme
Bich Ha Nguyen and Van Hieu Nguyen

A theoretical review on electronic, magnetic and optical properties of silicene
Suman Chowdhury and Debnarayan Jana

Photonics and optoelectronics of two-dimensional materials beyond graphene
Joice Sophia Ponraj, Zai-Quan Xu, Sathish Chander Dhanabalan et al.

Time-dependent density-functional theory for extended systems
Silvana Botti, Arno Schindlmayr, Rodolfo Del Sole et al.

Excitonic effects in the optical properties of 2D materials:an equation of motion approach
A J Chaves, R M Ribeiro, T Frederico et al.

2D Materials

TOPICAL REVIEW

Calculating excitons, plasmons, and quasiparticles in 2D materials and van der Waals heterostructures

Kristian Sommer Thygesen

CAMD and Center for Nanostructured Graphene (CNG), Department of Physics, Technical University of Denmark, 2800, Kgs. Lyngby, Denmark

E-mail: thygesen@fysik.dtu.dk

Abstract

Atomically thin two-dimensional (2D) materials host a rich set of electronic states that differ substantially from those of their bulk counterparts due to quantum confinement and enhanced many-body effects. This Topical Review focuses on the theory and computation of excitons, plasmons and quasiparticle band structures in 2D materials and their heterostructures. The general theory is illustrated by applications to various types of 2D materials including transition metal dichalcogenides, graphene, phosphorene, and hexagonal boron nitride. The weak and highly non-local dielectric function of atomically thin crystals is shown to be responsible for many of the unique properties exhibited by the 2D materials such as the formation of strongly bound, non-Hydrogenic excitons, large band gap renormalization effects, and the different signatures of excitons and plasmons in electron energy loss spectroscopy (EELS). Among other topics covered are spin–orbit coupling, trions, interlayer excitons, exciton dissociation, and environmental screening. Technical issues associated with the application of the many-body GW method and the Bethe–Salpeter equation (BSE) to 2D materials are also discussed. A combined quantum/classical method is introduced and used throughout to account for dielectric screening and self-energy effects from substrates and van der Waals heterostructures including the difficult case of non-matching lattices.

1. Introduction

Atomically thin two-dimensional (2D) materials have created a revolution in materials science. The family of 2D materials, which started with graphene [1], has expanded rapidly over the past few years and now includes insulators, semiconductors, semimetals, metals, and superconductors [2]. This development has prompted an explosion in envisioned applications ranging from battery electrodes and catalysis to photovoltaics and electronics [3]. At the same time, 2D materials research has been perhaps the most important breeding ground for new concepts and phenomena in condensed matter physics this decade. This is not least due to the distinctive nature of the electronic excitations in atomically thin crystals.

In terms of materials, the most well studied 2D systems beyond graphene, are the insulator hexagonal boron-nitride (hBN) [4, 5], the anisotropic semiconductor phosphorene [6, 7], and the semiconducting Mo- and W-based transition metal dichalcogenides (TMDs) with the chemical formula MX_2 (M: transition metal; X: chalcogen) [8]. In addition to the Mo- and W-based

compounds, about a dozen other semiconducting and metallic TMDs have been synthesized in mono- or few-layer form. Apart from graphene (C) and phosphorene (P) a host of other elemental 2D materials including silicene (Si) [9–11], germanene (Ge) [12], stanene (Sn) [13], and borophene (B) [14] have been realized on different substrates. Additionally, group III–V compounds [15, 16], transition metal carbides and -nitrides (MXenes) [17], transition metal oxides [18] and -halides [19], as well as organic-inorganic hybrid perovskites [20, 21], represent new developing classes of 2D materials.

An important characteristic of the 2D materials is the weak and highly non-local way in which they screen electric fields [22, 23]. In fact, this feature is responsible for many of the peculiar opto-electronic properties of the 2D materials including the existence of tightly bound electron–hole complexes [24, 25] and plasmons with strongly wave vector-dependent energies and coupling strengths [26–28]. The weak intrinsic screening also enhances self-energy effects and renders the quasiparticle band structure of 2D materials

sensitive to screening from the environment, e.g. substrates [29, 30]. This state of affairs makes it essential to go beyond the single-particle approximation and explicitly account for many-body effects when modelling the electronic and optical properties of 2D materials.

Although the 2D materials are interesting in their own right, an even larger potential lies in the possibility of reassembling different 2D crystals into new layered compounds [31]. Such designer materials have been coined van der Waals heterostructures (vdWHs) with reference to the weak van der Waals forces holding the 2D crystal planes together. The concept provides an ideal platform for controlling the unique 2D electronic states with atomic scale precision, and offers an enormous amount of possibilities to broaden the versatility of 2D materials and to achieve superior and unusual material properties. In contrast to conventional heterostructure fabrication methods, which involve complex and expensive ultrahigh-vacuum techniques to epitaxially grow the single-crystalline layers, vdWHs can be stacked in ambient conditions with no requirements of lattice matching. Furthermore, the weak interlayer bonding in vdWHs leads to clean and atomically well defined interfaces, thus reducing detrimental interface scattering effects [32, 33].

The vast size of the vdWH compound space can be illustrated by considering a heterostructure containing 20 atomic layers (which is similar in size to the most complex structures made to date [34]). If each layer is picked from a pool of ten different 2D materials this yields 10^{20} different stacking combinations. In reality, the space of presently available 2D materials comprises several hundreds of compounds [35] and the possibilities for combining them are essentially endless. To harness the full potential of vdWHs it is essential to develop efficient and reliable computational methods that can scan the vdWH compound space and guide experimentalists towards the most interesting materials before undertaking the time consuming task of synthesising specific heterostructures.

The computational modelling of vdWHs, or substrate supported 2D materials, is complicated by the incommensurable nature of the interfaces. Solid state computational codes apply periodic boundary conditions which imply the use of large super cells to avoid straining the different 2D layers of the heterostructure. However, such an approach quickly becomes prohibitive for anything but the simplest few-layer structures. Consequently, reliable modelling of realistic, incommensurable heterostructures requires novel approaches that combine the quantum description of the individual layers with a more coarse grained description of the effect of interlayer interactions. Recently, a few such methods have appeared addressing different aspects of the interactions including the effect of interlayer hybridisation [36] and dielectric screening [37].

This review focuses on the theory and computational methods used to describe elementary electronic excitations in 2D materials and heterostructures. To limit the scope, only pristine crystals are considered, although defects are known to play an important role for the chemical and physical properties of 2D materials [38–40]. The review is organised into four main sections devoted to the topics: dielectric screening (section 2), quasiparticle band structures (section 3), excitons (section 4), and plasmons (section 5). Each section is structured such that the first part focusses on isolated 2D materials while the last part extends the discussion to van der Waals heterostructures and effects of substrate interactions with an emphasis on the role of the dielectric screening.

2. Dielectric screening

The macroscopic dielectric constant of a conventional bulk material is defined as the limiting value of $\epsilon(\mathbf{q}, \omega)$ as $\mathbf{q} \to 0$. For a 2D material this definition cannot be straightforwardly adopted since $\epsilon_{2D}(\mathbf{q} = 0, \omega) = 1$. In fact, the dielectric function of a 2D material is strongly $\mathbf{q}$-dependent and a more elaborate treatment of screening is required [41]. In the simplest approximation, the 2D material can be modelled as an infinitely thin isotropic slab. This approximation leads to a dielectric function of the form

$$\epsilon_{2D}(\mathbf{q}, \omega) = 1 + 2\pi\alpha(\omega)q, \qquad (1)$$

where α is the 2D polarizability of the layer [42]. Using that the 2D Fourier transform of the $1/r$ Coulomb interaction equals $2\pi/q$, the screened interaction can be written

$$W_{2D}(\mathbf{q}, \omega) = \frac{2\pi}{q\epsilon_{2D}(\mathbf{q}, \omega)}. \qquad (2)$$

This simple form of the screened interaction has the merit of leading to an analytical expression in real space and has been used by several authors to study excitons and charged impurities in various 2D semiconductors [23, 42–44]. We note that the form $2\pi/q$ for the Coulomb interaction and equation (1) for the dielectric function are consistent approximations which both become exact in the limit of vanishing thickness of the material, i.e. the strict 2D limit. In this section, the strict 2D screening model is first compared to more realistic models taking the finite thickness of the material into account. Next, we consider the effects of support or van der Waals heterostructuring on the dielectric properties.

2.1. Quasi-2D Coulomb interaction
To understand the nature of dielectric screening in atomically thin materials, we begin by considering the form of the bare Coulomb interaction. It is instructive to compare the case of a strict 2D material to the physically relevant case of a 2D material with a finite thickness. To this end we consider the interaction energy between two normalized line charges, ρ_1 and ρ_2, located at the in-plane positions $\mathbf{r}_{\parallel,1}$ and $\mathbf{r}_{\parallel,2}$, respectively,

and extending over the 2D material in the out-of-plane direction (see figure 1(b)),

$$V_{Q2D}(\mathbf{r}_{\parallel,1}, \mathbf{r}_{\parallel,2}) = \int d\mathbf{r} d\mathbf{r}' \frac{\rho_1(\mathbf{r})\rho_2(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \quad (3)$$

The (2D) Fourier transform of the quasi-2D interaction becomes

$$V_{Q2D}(\mathbf{q}_{\parallel}) = \frac{4\pi}{dq_{\parallel}^2}\left[1 - \frac{2}{q_{\parallel}d}e^{-q_{\parallel}d/2}\sinh\left(\frac{q_{\parallel}d}{2}\right)\right]. \quad (4)$$

Physically, $V_{Q2D}(\mathbf{q}_{\parallel})$ equals the Coulomb energy of a quasi-2D charge density wave with wave vector $\mathbf{q}_{\parallel}$ and thickness d. We note that for in-plane wavelengths much larger than the thickness of the material $V_{Q2D}(\mathbf{q}_{\parallel})$ equals the strict 2D potential while for smaller wave lengths it equals the 3D potential, i.e.

$$V_{Q2D}(\mathbf{q}_{\parallel}) = \begin{cases} \dfrac{2\pi}{q_{\parallel}}, & q_{\parallel}d \ll 1 \\ \dfrac{4\pi}{dq_{\parallel}^2}, & q_{\parallel}d \gg 1 \end{cases} \quad (5)$$

2.2. Dielectric function: 2D versus 3D

For a material of any dimensionality, the (inverse) microscopic dielectric function gives the total potential to first order in the applied external potential,

$$V_{tot}(\mathbf{r}) = \int d\mathbf{r}' \epsilon^{-1}(\mathbf{r},\mathbf{r}')V_{ext}(\mathbf{r}'), \quad (6)$$

where we have neglected the frequency dependence for ease of notation. For periodic systems it is often more convenient to use a plane wave representation. Taking $\mathbf{q}$ to represent a wave vector in the first Brillouin zone (BZ) and $\mathbf{G}$ a reciprocal lattice vector, the dielectric matrix of a crystal evaluated within the random phase approximation (RPA), takes the form

$$\epsilon_{GG'}(\mathbf{q}) = \delta_{GG'} - V(\mathbf{q}+\mathbf{G})\chi^0_{GG'}(\mathbf{q}), \quad (7)$$

where $V(\mathbf{q}+\mathbf{G})$ is the 3D Fourier transform of the Coulomb potential and χ^0 the non-interacting response function,
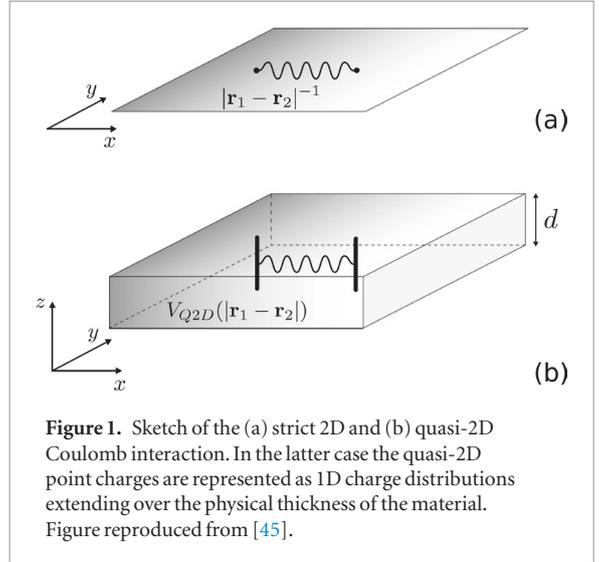
$$\chi^0_{GG'}(\mathbf{q},\omega) = \sum_{\mathbf{k}}^{BZ}\sum_{n,n'} \frac{f_{n\mathbf{k}}-f_{n'\mathbf{k}+\mathbf{q}}}{\omega+\varepsilon_{n\mathbf{k}}-\varepsilon_{n'\mathbf{k}+\mathbf{q}}+i\eta}$$
$$\times \langle\psi_{n\mathbf{k}}|e^{i(\mathbf{G}+\mathbf{q})\cdot\mathbf{r}}|\psi_{n'\mathbf{k}+\mathbf{q}}\rangle\langle\psi_{n\mathbf{k}}|e^{i(\mathbf{G}'+\mathbf{q})\cdot\mathbf{r}}|\psi_{n'\mathbf{k}+\mathbf{q}}\rangle^* \quad (8)$$

For a 3D periodic system, the total potential resulting from a plane wave external potential $V_{ext}(\mathbf{r}) = V_0 e^{i\mathbf{q}\cdot r}$ has the form

$$V_{tot}(\mathbf{r}) = \tilde{V}_{\mathbf{q}}(\mathbf{r})e^{i\mathbf{q}\cdot r}, \quad (9)$$

where $\tilde{V}_{\mathbf{q}}(\mathbf{r})$ is a lattice periodic function. Since usually we are interested in macroscopic fields, we define the 3D macroscopic dielectric function as

$$\frac{1}{\epsilon_{3D}(\mathbf{q})} \equiv \frac{\langle\tilde{V}_{\mathbf{q}}(\mathbf{r})\rangle_\Omega}{V_0} = \epsilon_{00}^{-1}(\mathbf{q}), \quad (10)$$

Figure 1. Sketch of the (a) strict 2D and (b) quasi-2D Coulomb interaction. In the latter case the quasi-2D point charges are represented as 1D charge distributions extending over the physical thickness of the material. Figure reproduced from [45].

where $\langle...\rangle_\Omega$ denotes a spatial average over a unit cell. Note that in general $\epsilon_{3D}(\mathbf{q}) \neq \epsilon_{00}(\mathbf{q})$ due to local field effects [46].

First-principles codes developed for bulk crystals use super cells with a finite vacuum region of thickness L to separate the periodically repeated 2D layers. When such codes are used direct application of equation (10) leads to $\epsilon_{3D}(\mathbf{q},\omega) = 1 + \mathcal{O}(1/L)$ [41]. This is a consequence of an averaging region much larger than the effective extension of the electron density around the material. The standard definition in equation (10) becomes meaningless in this case, which is the reason why relatively different values for ϵ_{3D} have been reported for monolayer materials in the recent literature [47–49].

From the first equality in equation (10), it is natural to substitute an average over the entire unit cell in the out of plane direction with an average over a confined region corresponding to the actual extension of the electronic density. In practice, we average the in-plane coordinates $(\mathbf{r}_{\parallel})$ over the unit cell area, A, and the out-of-plane coordinate from $z_0 - d/2$ to $z_0 + d/2$, where z_0 denotes the center of the material and d its width. The macroscopic dielectric function then becomes:

$$\frac{1}{\epsilon_{Q2D}(\mathbf{q}_{\parallel},\omega)} \equiv \frac{\langle\tilde{V}_{\mathbf{q}}(\mathbf{r},\omega)\rangle_{A,d}}{V_0}$$
$$= \frac{2}{d}\sum_{G_\perp} e^{iG_\perp z_0}\frac{\sin(G_\perp d/2)}{G_\perp}\epsilon_{G_\perp 0}^{-1}(\mathbf{q}_{\parallel},\omega), \quad (11)$$

with $\epsilon_{GG'}^{-1}(\mathbf{q}_{\parallel},\omega)$ calculated from $\chi^0_{GG'}(\mathbf{q}_{\parallel},\omega)$ according to the RPA expression in equation (7).

We stress that it is essential to use a truncated Coulomb potential in equation (7) in order to decouple the layers in neighboring supercells [41]. Note that we used the label Q2D in equation (11) since this definition of the macroscopic dielectric function is consistent with the quasi-2D picture. As a rule of thumb we choose d to be the distance between the layers in the layered bulk, but the results are not very sensitive to this choice [41].

The calculated $\mathbf{q}_\parallel$-dependent (static) dielectric functions of monolayer hBN and MoS_2 are illustrated in figure 2. The dielectric function approaches unity for $\mathbf{q}_\parallel \to 0$ meaning that long range screening is completely absent. Mathematically, this follows from equation (7) and the fact that in the small $\mathbf{q}$ limit, $\chi^0_{00}(\mathbf{q}, \omega) \propto q^2$ while the *truncated* Coulomb interaction, $V_{\mathrm{trunc}} \propto 1/q$. In addition to the Q2D dielectric function, figure 2 shows the linear expression in equation (1) as well as the dielectric function of the 3D layered crystals. As was the case for the bare quasi-2D Coulomb interaction, see equation (5), the quasi-2D dielectric function follows the strict 2D result for $q_\parallel d \ll 1$ and the bulk result for $q_\parallel d \gg 1$.

2.3. Dielectric screening in van der Waals heterostructures

The incommensurate nature of vdWHs presents a great challenge for first-principles calculations because it is generally not possible to represent the heterostructure in a computational cell that is small enough to allow the calculation to be performed without straining one or more of the layers and thereby alter its electronic properties. The problem is particularly severe for many-body calculations for which the computational cost grows rapidly with system size. For example, the hetero-bilayer MoS_2/WSe_2, which is studied in greater detail in section 4.6, requires a unit cell of at least 100 atoms to reduce the strain to 1% rendering standard plane wave based many-body calculations highly challenging even for this simple system.

To overcome the problem, the quantum electrostatic heterostructure (QEH) model computes the dielectric function of a general vdWHs from the *ab initio* response functions of the constituent 2D layers coupled only via their electrostatic interactions, see figure 3. A minimal representation of the response functions of the 2D layers (the *dielectric building blocks*), consisting of the $(\mathbf{q}_\parallel, \omega)$-resolved monopole and dipole components (out of plane), is employed without compromising the accuracy. As shown in later sections, the QEH model allows us to evaluate the screened interaction entering the theories of excitons, plasmons, and quasiparticle band structures, for general incommensurate vdWHs. A database containing the pre-calculated dielectric building blocks of a collection of 2D materials is available on-line (https://cmr.fysik.dtu.dk). From here the data files can be downloaded together with a Python module for constructing the dielectric building block of any 2D material and solving the coupled electrostatic equations for the heterostructure dielectric function. QEH model calculations for vdWHs containing a few hundred layers can be performed on a standard PC. Details on the QEH model can be found in [37].

As a simple illustration of the QEH model, figure 4 shows the calculated static dielectric function of multilayer MoS_2 slabs of varying thickness as a function of

in-plane wave vector. The dielectric function measures the ratio between an applied potential of the form $V_0 \exp(i\mathbf{r} \cdot \mathbf{q}_\parallel)$ and the resulting total potential averaged over the slab, i.e. the macroscopic dielectric function of the multilayer as defined in equation (11). For large $q_\parallel$ the dielectric functions show similar behavior. However, whereas $\epsilon(0) = 14$ for the bulk, the dielectric functions of the slabs decrease sharply to 1 for small $q_\parallel$. This demonstrates that the dielectric properties of a vdWH of thickness d are 2D-like for $q_\parallel \ll 1/d$ and 3D-like for $q_\parallel \gg 1/d$. The QEH model describes the change in the dielectric function from monolayer to bilayer very accurately in spite of the well known differences between their electronic band structures [47]. This shows that hybridisation driven band structure effects, i.e. quantum confinement, have rather weak influence on the dielectric properties of a vdWH and is the main reason for the success of the QEH model.

3. Quasiparticle band structures

The electronic band structure of a 2D material can be routinely calculated using density functional theory (DFT) adopting one of the many existing approximations for the exchange-correlation (xc) functional. It is well known that semi-local xc-functionals underestimate band gaps of semiconductors (including 2D semi-conductors [50]) while hybrid functionals, such as the HSE [51], perform better albeit at higher computational cost. It is perhaps less well known that the single-particle DFT band structure, even when calculated with the exact xc-functional, is not supposed to be exact. The exact band gap can in principle be obtained by adding the so-called derivative discontinuity, Δ_{xc}, to the exact DFT single-particle gap, however, most functionals do not evaluate Δ_{xc} or incorrectly predict it to be zero. One exception is the GLLBSC functional, which has been shown to predict fairly accurate band gaps for a range of materials [52].

3.1. GW approximation

The true band energies of a solid can be obtained as the poles of the single-particle Green function. For weakly correlated materials, these poles are well defined, i.e. the so-called renormalization factor Z is close to unity, and can be obtained by solving the quasiparticle (QP) eigenvalue equation where the xc-potential of the DFT Hamiltonian is replaced by the non-local and energy-dependent self-energy operator, $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$. Physically, the QP energies describe the energy cost of adding electrons/holes to the neutral ground state of the material. They are thus defined in terms of the many-body total energies of the neutral and charged systems,

$$\varepsilon^{\mathrm{QP}}_{n\mathbf{k}} = E_{n\mathbf{k}}(N+1) - E_0(N), \text{ for empty states}$$

(12)

$$\varepsilon^{\mathrm{QP}}_{n\mathbf{k}} = E_0(N) - E_{n\mathbf{k}}(N-1), \text{ for occupied states}$$

(13)

Figure 2. Dielectric functions of (a) hBN and (b) MoS$_2$. The bulk (dashed), the Q2D (green) and strict 2D (blue) static dielectric functions are illustrated. The condition $q_\| d = 1$ divides the 2D and 3D screening regimes. Figure reproduced from [45].

Figure 3. The QEH model. (a) The density response function and dielectric function of the heterostructure are calculated from the dielectric building blocks of the individual layers assuming a purely electrostatic interaction between the layers. The dielectric building blocks are calculated *ab initio* for the isolated layers. They comprise the monopole and dipole components of the density response function, $\tilde{\chi}_{M/D}$, together with the spatial shape of the electron density, $\rho_{M/D}(z)$, induced by a constant and linear potential of the forms $\exp(i\mathbf{q}_\| \cdot \mathbf{r})$ and $z \exp(i\mathbf{q}_\| \cdot \mathbf{r})$, respectively. (b) Monopole and dipole induced densities (blue) together with the associated potentials (red) for monolayer MoS$_2$. Figure reproduced from [37].

The QP energies can be measured by photo-emission or transport spectroscopy which probe the transition energies between states with N and $N \pm 1$ electrons. The QP gap is defined as the difference between the highest occupied state (ionisation potential) and lowest unoccupied state (electron affinity). It differs from the optical gap by the exciton binding energy which can be substantial for atomically thin materials, see section 4.

A highly successful approach to predict QP energies is the so-called GW method, which approximates Σ to first order in the screened interaction, W. The GW approximation was introduced by Hedin [53] in 1965

and first applied to real solids in an *ab initio* framework by Hybertsen and Louie [54] and Godby *et al* [55]. Compared to mean field methods, like the DFT Kohn–Sham scheme, the GW method is significantly more involved both in terms of numerical implementation and the required computation time. In real space the GW self-energy takes the form

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = i \int d\omega' G(\mathbf{r}, \mathbf{r}', \omega' + \omega) W(\mathbf{r}, \mathbf{r}', \omega') \quad (14)$$

where G is the single-particle Green function. Within the simplest G$_0$W$_0$ approach, the Green function and
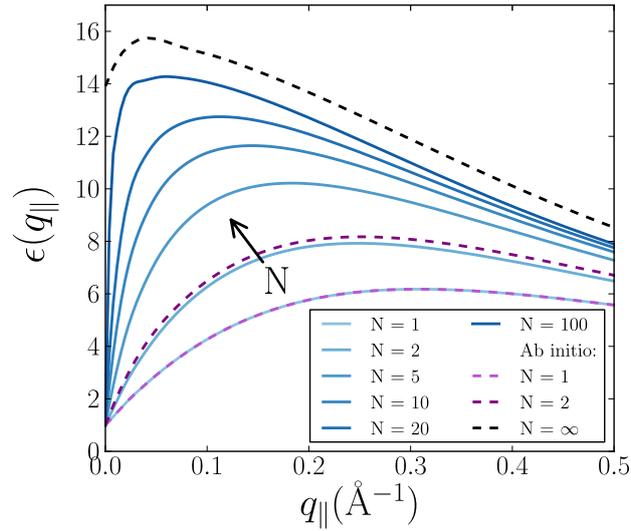
Figure 4. Static dielectric function of an N-layer MoS$_2$ stack as a function of the in-plane momentum transfer. $\epsilon(\mathbf{q}_\parallel)$ relates the total potential averaged over the slab to an external potential of the form $V_{\text{ext}}(\mathbf{r}) = V_0 \exp(i\mathbf{r} \cdot \mathbf{q}_\parallel)$. The response increases monotonically with N converging slowly towards the dielectric function of bulk MoS$_2$ (black dashed line) obtained from a first-principles calculation. Figure reproduced from [37].

screened interaction are evaluated from the eigenvalues and wave functions of the non-interacting Kohn–Sham Hamiltonian, H_0, with solutions

$$H_0 |\psi_{n\mathbf{k}}^0\rangle = \varepsilon_{n\mathbf{k}}^0 |\psi_{n\mathbf{k}}^0\rangle. \tag{15}$$

To obtain the QP energies one treats $\Sigma - V_{\text{xc}}$ using first-order perturbation theory which leads to the expression

$$\varepsilon_{n\mathbf{k}}^{\text{QP}} = \varepsilon_{n\mathbf{k}}^0 + Z_{n\mathbf{k}} \cdot \langle \psi_{n\mathbf{k}}^0 | \Sigma_{\text{xc}}(\varepsilon_{n\mathbf{k}}^0) - V_{\text{xc}} | \psi_{n\mathbf{k}}^0 \rangle, \tag{16}$$

where $Z_{n\mathbf{k}}$ is the renormalisation factor. For more details and background theory on the GW method we refer to the extensive literature on the topic.

### 3.2. GW calculations of freestanding 2D materials

The performance of the GW approximation has been thoroughly established for bulk materials [56–58] and more recently also for molecules [59, 60]. In comparison, systematic studies assessing the accuracy and numerical convergence of GW calculations for 2D materials are still few. Nevertheless, it has become clear that (i) it is essential to use a truncated Coulomb interaction when evaluating W in order to avoid long range screening between periodically repeated layers which otherwise leads to an unphysical reduction of the QP band gap, and (ii) when a truncated Coulomb interaction is used, the convergence of the GW calculation with respect to the number of in-plane $\mathbf{k}$-points becomes much slower than is the case for similar bulk systems [41, 50, 61, 62].

The slow $\mathbf{k}$-point convergence of the GW band structure is a direct consequence of the strong wave vector-dependence of the dielectric function of a 2D semiconductor, see figure 2, which implies that the number of $\mathbf{q}$-points required to obtain a proper sampling of the screened interaction $W(\mathbf{q})$ over the BZ is

much higher for the 2D material than what would be anticipated from the 3D case. For example, the band gap of bulk 2H-MoS$_2$ is converged to within 0.1 eV with an in-plane $\mathbf{k}$-point grid of 12×12 while the same accuracy for monolayer MoS$_2$ requires a grid of 40×40 when standard BZ sampling schemes are applied.

The $\mathbf{k}$-point convergence can be greatly improved by using an analytical expression for the small $\mathbf{q}$-limit of the 2D response function when performing the BZ integral over the critical region around $\mathbf{q} = 0$ [63]. In brief, an expression for the dielectric function similar to equation (1) is used to obtain a closed expression for $W_{\mathbf{G}\mathbf{G}'}(\mathbf{q})$ which can be analytically averaged around $\mathbf{q} = 0$. As seen in figure 6(a) the analytical expression for $W_{00}(\mathbf{q})$ is valid for surprisingly large in-plane $\mathbf{q}$-vectors. The use of the analytical model drastically reduces the requirements on the $\mathbf{k}$-point mesh. Figure 6(b) shows that the band gap of MoS$_2$ converges much faster than the traditional sampling scheme. The latter underestimates the screening around $\mathbf{q} = 0$ because $\epsilon(\mathbf{q} = 0) = 1$, and consequently no screening is picked up at this sampling point which results in a large over-estimation of the gap.

Table 1 shows the calculated QP band gaps of monolayer hBN, MoS$_2$, and phosphorene. These calculations were performed using a truncated Coulomb interaction, extrapolated as $1/N_G$ to the infinite plane wave basis limit, and converged with respect to $\mathbf{k}$-point grid. Spin-orbit interactions were not included in the reported values. Inclusion of spin–orbit interactions has no effect for hBN and phosphorene, but split the valence band of MoS$_2$ at the K point by 0.15 eV thereby lowering the QP gap by around 0.07 eV [50, 64] (see table 2 for the spin orbit splittings in the TMDs).
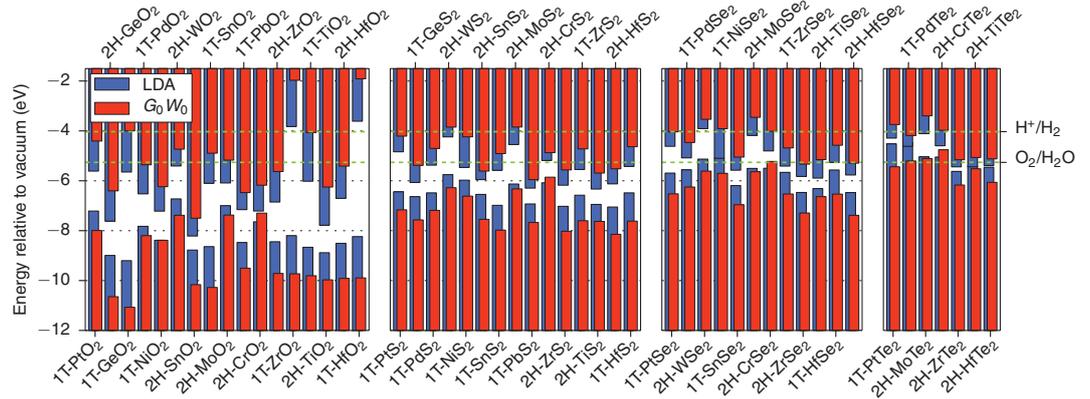
Figure 5. Position of the valence band maximum (VBM) and conduction band minimum (CBM) relative to the vacuum level (set to zero) for 51 transition metal dichalcogenides and oxides. The calculations are performed using the LDA and G_0W_0 methods and include spin–orbit interactions. The hydrogen and oxygen evolution potentials at pH 7 are shown by green dashed lines. Figure reproduced from [50].
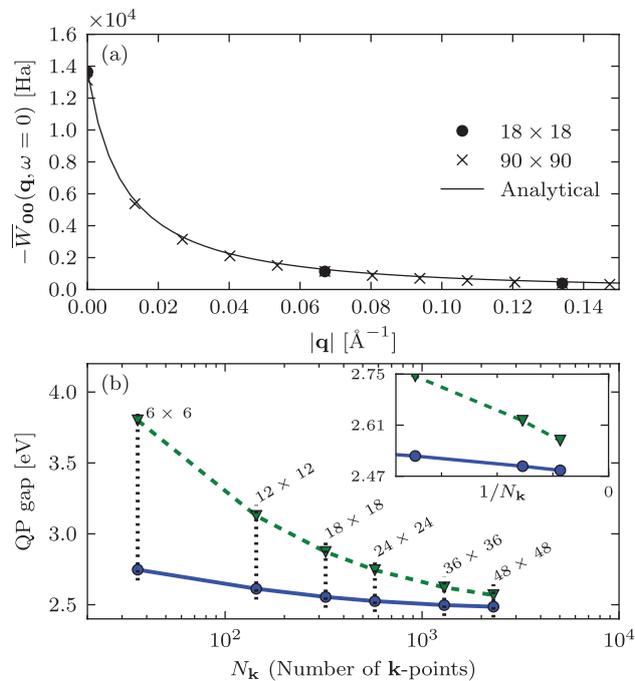
Figure 6. (a) Static screened potential subtracted the bare interaction, $\overline{W} = W - V$, for monolayer MoS_2 as a function of q along the $\Gamma \rightarrow M$ direction. The crosses and circles indicate numerical values obtained using two different uniform **q**-point grids. The analytical approximation to $W_{00}(\mathbf{q})$ (full curve) based on a lowest order expansion of $\chi^0_{GG'}(\mathbf{q})$ provides an excellent approximation to the numerical results. (b) Convergence of the QP band gap of monolayer MoS_2 calculated using two different treatments of the $\mathbf{q} = \mathbf{0}$ limit of $W_{00}(\mathbf{q})$. The dashed green line shows the result obtained when the $\mathbf{q} = \mathbf{0}$ contribution is simply neglected corresponding to the standard treatment used for 3D systems. The solid blue line shows the result obtained with the analytical expression for $W(\mathbf{q})$. The inset is a zoom illustrating the asymptotic $1/N_\mathbf{k}$ dependence of the QP gap. Figure adapted from [63].

For MoS_2 the converged G_0W_0@PBE band gap of 2.54 eV agrees well with our previously reported value of 2.48 eV (with spin–orbit coupling) obtained using a Wigner–Seitz truncated Coulomb interaction and 36×36 **k**-points [50]. Other reported G_0W_0 gaps range from 2.40–2.82 eV [47, 48, 49, 65, 66, 67]. However, these calculations were performed (i) without the use of a truncated Coulomb interaction (ii) using relatively small **k**-point grids of 6×6 to 18×18, and (iii) using different in-plane lattice constants vary-

ing between 3.15 and 3.19 Å. These different settings can affect the band gap by as much as 0.5 eV [41], and therefore we refrain from providing a detailed comparison of our result to these earlier calculations. An overview of previous GW results for MoS_2 can be found in [41, 62].

For many applications, not only the distance between the occupied and unoccupied bands, i.e. the band gap, but also the absolute position of the band edges relative to vacuum are of interest. For a 2D material the abso-

Table 1. Calculated band gaps (in eV) from PBE, G_0W_0@PBE and partially self-consistent GW_0@PBE for three monolayers in PBE-relaxed structures. The GW calculations were performed using a truncated Coulomb interaction and analytic integration of $W(\mathbf{q})$ around $\mathbf{q} = \mathbf{0}$. Spin–orbit interactions were not included. The $\mathbf{k}$-point grids were 18×18 (h-BN), 18×18 (MoS$_2$) and 10×14 (phosphorene). Table reproduced from [63].

	Transition	DFT-PBE	G_0W_0@PBE	GW_0@PBE
hBN	$K \rightarrow \Gamma$	4.64	7.06	7.49
	$K \rightarrow K$	4.72	7.80	8.25
2H-MoS$_2$	$K \rightarrow K$	1.65	2.54	2.65
Phosphorene	$\Gamma \rightarrow \Gamma$	0.90	2.03	2.29

Table 2. Spin-orbit induced splittings at the valence band maximum and conduction band minimum for some monolayer transition metal dichalcogenides and -oxides. The band splittings have been calculated by full diagonalization of the LDA and G_0W_0@LDA Hamiltonians, respectively [73]. The location of the band extrema in the BZ is indicated in parenthesis. Note that this can be different in LDA and GW. Table reproduced from [50].

Material	LDA ΔE_{vbm}^{soc}	LDA ΔE_{cbm}^{soc}	G_0W_0@LDA ΔE_{vbm}^{soc}	G_0W_0@LDA ΔE_{cbm}^{soc}
2H-CrS$_2$	0.07 (K)	0 (K)	0.07 (K)	0.01 (K)
2H-CrSe$_2$	0.09 (K)	0.02 (K)	0.1 (K)	0.02 (K)
2H-CrTe$_2$	0.12 (K)	0.02 (K)	0.13 (K)	0.03 (K)
2H-HfO$_2$	0 (T)	0.17 (T)	0 (T)	0.15 (T)
2H-HfS$_2$	0.03 (T)	0.0 (T)	0.02 (T)	0.09 (X)
2H-HfSe$_2$	0.13 (T)	0.1 (T)	0.12 (T)	0.11 (X)
2H-HfTe$_2$	0 (Γ)	0.15 (Γ)	0.48 (T)	0.18 (X)
2H-MoS$_2$	0.15 (K)	0 (K)	0.15 (K)	0 (K)
2H-MoSe$_2$	0.19 (K)	0.02 (K)	0.19 (K)	0.02 (K)
2H-MoTe$_2$	0.23 (K)	0.04 (K)	0.25 (K)	0.05 (X)
2H-TiO$_2$	0 (T)	0.02 (T)	0 (X)	0.02 (T)
2H-TiS$_2$	0.02 (T)	0 (T)	0 (Γ)	0 (Σ)
2H-TiTe$_2$	0 (Γ)	0 (Γ)	0.32 (T)	0 (Σ)
2H-WO$_2$	0 (Γ)	0.02 (Γ)	0 (Γ)	0 (K)
2H-WS$_2$	0.45 (K)	0.04 (K)	0.45 (K)	0.02 (K)
2H-WSe$_2$	0.49 (K)	0.04 (K)	0.49 (K)	0.03 (K)
2H-ZrO$_2$	0 (T)	0.05 (T)	0 (T)	0.05 (T)
2H-ZrS$_2$	0.02 (T)	0.02 (T)	0.02 (T)	0 (Σ)
2H-ZrSe$_2$	0.1 (T)	0.03 (T)	0.1 (T)	0 (Σ)
2H-ZrTe$_2$	0 (Γ)	0 (Γ)	0.28 (T)	0 (Σ)

lute band energies are obtained by referring them to the asymptotic value of the Hartree potential in the vacuum region. For bulk materials, the problem is much more delicate as it requires the use of thick slabs to represent both the bulk interior and its surface, and because the Hartree potential depends on the surface dipoles (on both sides of the slab), which makes the band alignment dependent on the surface termination. These problems are obviously not present for the 2D materials making them ideal as benchmark systems for the band alignment problem. Figure 5 shows the positions of the valence band maximum (VBM) and conduction band minimum (CBM) relative to the vacuum level for 51 MX$_2$ monolayer metal-

dichalcogenides and -oxides evaluated at both the LDA and G_0W_0 levels. For all materials, the effect of the G_0W_0 correction is to shift the conduction band up and the valence band down with respect to the LDA values. There are clear correlations between the band energies and chemical composition of the materials with the oxides showing the largest gaps and deepest lying band edges followed by the sulphides, selenides, and tellurides in that order. These trends can be explained in terms of the relative size and electronegativity of the O, S, Se, and Te atoms [50]. In section 4.6, the band alignment at the MoS$_2$/WSe$_2$ heterobilayer is analysed in greater detailed.
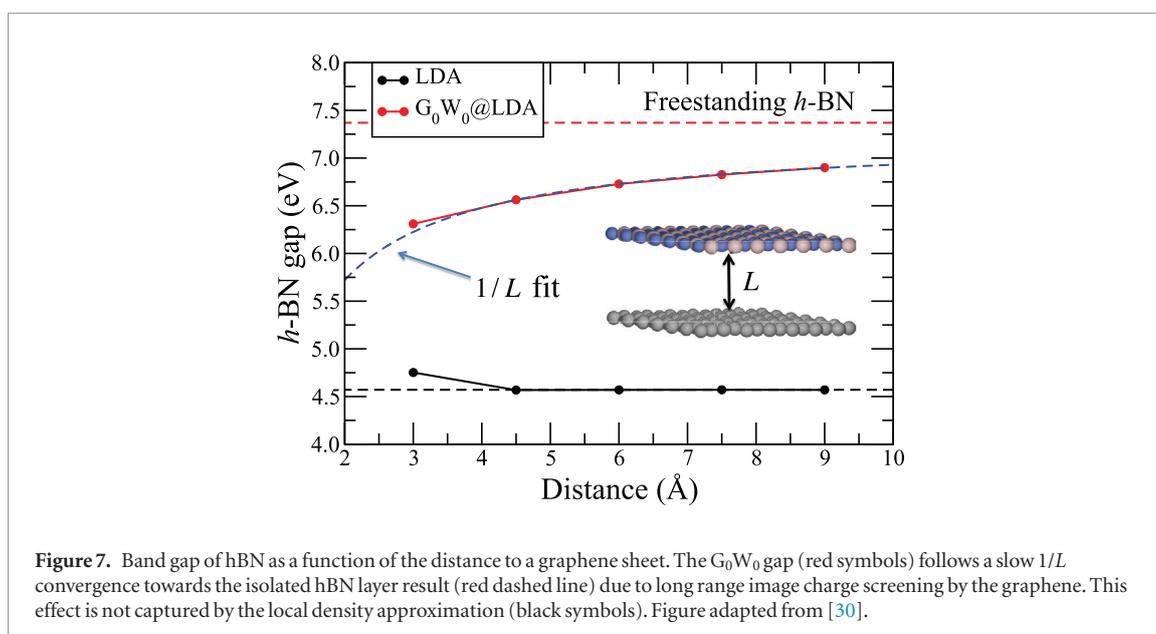
3.3. Spin–orbit interactions
Several of the interesting 2D materials, in particular the TMDs and the novel class of 2D metal-organic hybrid perovskites [20, 21], contain heavy elements for which spin–orbit (SO) interactions can have significant effects on the band structure. The SO coupling enters the Hamiltonian via the term [68]

$$H_{SO} = \frac{1}{2m^2c^2}\frac{1}{r}\frac{\partial V(r)}{\partial r}\mathbf{L} \cdot \mathbf{S} \qquad (17)$$

where c is the speed of light, m the electron mass, V the electron potential, and $\mathbf{S}$ and $\mathbf{L}$ are the electron spin- and angular momentum operators, respectively. The above form assumes a spherical potential. Even in a crystal lattice this is usually an excellent approximation because the SO coupling is completely dominated by the spherically symmetric regions close to the nuclei.

To compute the effect of SO coupling on the band structure, the single-particle Hamiltonian (including the H_{SO} term) must be diagonalised within the space of spin-up and spin-down eigenstates, $\{\psi_{n\mathbf{k},\sigma}\}$. Table 2 shows the spin–orbit induced splittings at the valence band maximum and conduction band minimum for some monolayer TMDs and TMOs. The splitting has been calculated by diagonalizing the LDA or the G_0W_0@LDA Hamiltonian, respectively. Clearly, the SO splitting is rather insensitive to the underlying band structures (LDA or GW) with large large differences occurring only in the cases where LDA and GW predict different locations of the band extrema within the BZ. Note that only monolayer TMDs in the 2H phase appear in table 2. This is because the combination of inversion symmetry and time reversal symmetry in the 1T phase inhibits spin splitting of the bands. The absence of band *splitting* does, however, not imply that SO coupling can be neglected for the 1T structures.

For the MoX$_2$ and WX$_2$ monolayers in the 2H-phase, the SO-induced splitting of the valence band at the K and K' points of the BZ, leads to a splitting of the lowest exciton into two excitons with opposite spin (the A and B excitons in figure 9). Since time reversal symmetry maps $|\sigma, K\rangle$ to $|-\sigma, K'\rangle$, the energetic ordering of the spin states is reversed in the K and K' valleys. This makes it possible to selectively excite excitons in one of the two valleys using circularly polarised light tuned to the

Figure 7. Band gap of hBN as a function of the distance to a graphene sheet. The G_0W_0 gap (red symbols) follows a slow $1/L$ convergence towards the isolated hBN layer result (red dashed line) due to long range image charge screening by the graphene. This effect is not captured by the local density approximation (black symbols). Figure adapted from [30].

frequency of the A or B exciton [69–71]. These discoveries have recently fueled the idea of using monolayer TMDs for valleytronics [72], i.e. technologies that exploit control over the valley degree of freedom.

3.4. Environmental screening

As described in the previous section, the QP energies depend on the screened interaction, $W = \epsilon^{-1} V$, via the electron self-energy. Owing to the long range nature of the Coulomb interaction and the weak intrinsic screening in 2D materials, the screened interaction, and consequently the QP energies, can be strongly affected by the dielectric environment outside the 2D material. For example, it was shown using non-linear optical spectroscopy and many-body calculations that the band gap of monolayer WS_2 can vary by 0.3 eV depending on the choice of substrate [29]. This effect can be utilised to control the band structure of 2D materials via the nonlocal screening by the substrate/embedding as demonstrated by GW calculations for both lateral [74] and vertical [75] heterostruture designs.
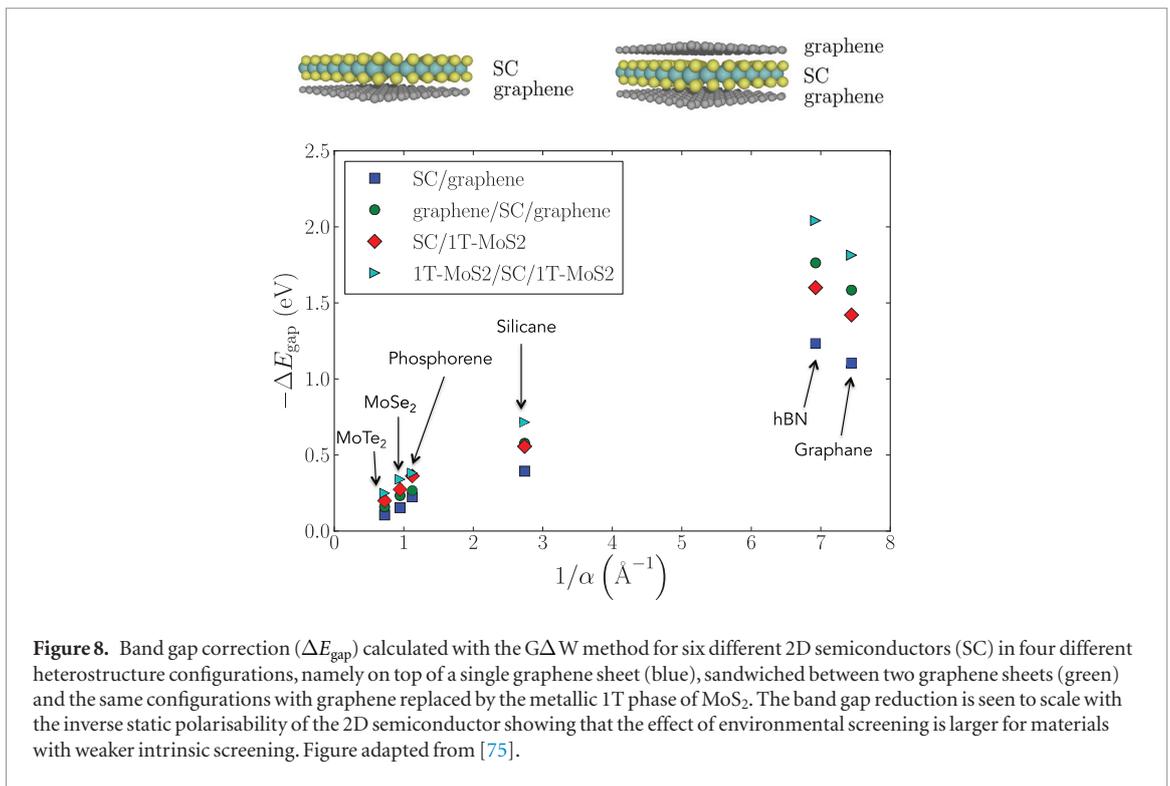
Figure 7 shows the band gap of an hBN layer as a function of distance, L, to a single graphene sheet. The G_0W_0 gap follows a $1/L$ dependence and at the equilibrium distance around $L = 3.5$ Å, the band gap is 1 eV lower than that of an isolated hBN layer despite the fact there is essentially no overlap between the wave functions of the two systems at the considered separations. This phenomenon is known as the image charge effect. It has been studied extensively for molecules on metal surfaces, where it can shift molecular energy levels by several electron volts [76, 77]. For metal-molecule-metal junctions a proper incorporation of the image charge effect is essential for correct prediction of energy level alignment and transport properties [78, 79].

In the case of a molecule on a metal surface, the origin of the image charge effect is easy to understand. Indeed, according to equations (12) and (13), the QP energies represent electron addition/removal energies.

When an electron is added or removed, the molecule is left in a ± 1 charge state. As a result of the electrostatic interaction between the charged molecule and its image charge in the metal surface, the QP energies will be lowered relative to those of the gas-phase molecule. Due to the different sign of occupied and unoccupied states, see equations (12) and (13), the image charge effect shifts occupied (empty) states up (down) in energy thereby closing the gap. For a 2D material, one might expect the image charge effect to vanish because the electrostatic interaction between a completely delocalised charge distribution of a Bloch state and its equally delocalised image charge, scales as $1/A$ where A is the area of the 2D material. However, this picture is in fact incorrect because the *electron* is not delocalised - only its probability density. Thus the image charge effect is a pure correlation effect.

Returning to figure 7, it can be seen that the LDA, in addition to underestimating the hBN band gap by 3 eV, does not capture the image charge effect. This is a common failure of all semi-local and hybrid xc-functionals [77] which are unable to detect changes in the single-particle energies when the (occupied) wave functions of the two subsystems do not overlap. This shows that quantitatively accurate modelling of band structures of supported 2D materials or vdWHs must employ many-body methods that are able to capture the non-local screening effects.

It is interesting to note that the effect of environmental screening on the QP band gap is highly material dependent. For example, GW calculations predict a reduction of the QP gap of a benzene molecule, an hBN monolayer, and an MoS_2 monolayer when placed on a graphite surface to be 3 eV, 1 eV, and 0.2 eV, respectively [30, 76, 80]. These differences cannot be explained by a simple image charge model that only accounts for the interaction between the bare electron/hole in the 2D material and its image charge in the substrate. Indeed, such a model would yield the same reduction of the QP

Figure 8. Band gap correction (ΔE_{gap}) calculated with the GΔW method for six different 2D semiconductors (SC) in four different heterostructure configurations, namely on top of a single graphene sheet (blue), sandwiched between two graphene sheets (green) and the same configurations with graphene replaced by the metallic 1T phase of MoS$_2$. The band gap reduction is seen to scale with the inverse static polarisability of the 2D semiconductor showing that the effect of environmental screening is larger for materials with weaker intrinsic screening. Figure adapted from [75].

gap for all three systems. The reason is that the change in QP band gap is determined by the *relative change* in screened interaction induced by the environment. Since the dielectric function of the 2D material has the form $\epsilon(\mathbf{q}) = 1 + 2\pi\alpha q$ one expects that materials with larger polarizability (α) are less affected by the environmental screening. Indeed, since the internal screening increases from benzene, to hBN to MoS$_2$, this argument provides an explanation for the observed differences in QP gap reduction.

The dependence of the band gap reduction on the intrinsic polarisability of the 2D material is clearly illustrated by figure 8 which shows the change in band gap (ΔE_{gap}) for six 2D semiconductors (SC) in four different heterostructure configurations: on top of a single graphene sheet, sandwiched between two graphene sheets, and the same configurations with graphene replaced by the metallic 1T phase of MoS$_2$. The change in band gap due to the environmental screening has been calculated using the GΔW method. In this method, the screened potential in the 2D material (layer i of a heterostructure (HS)) is split into two parts,

$$W_{i,HS}(\mathbf{r}, \mathbf{r}', \omega) = W_i(\mathbf{r}, \mathbf{r}', \omega) + \Delta W_i(\mathbf{r}, \mathbf{r}', \omega). \quad (18)$$
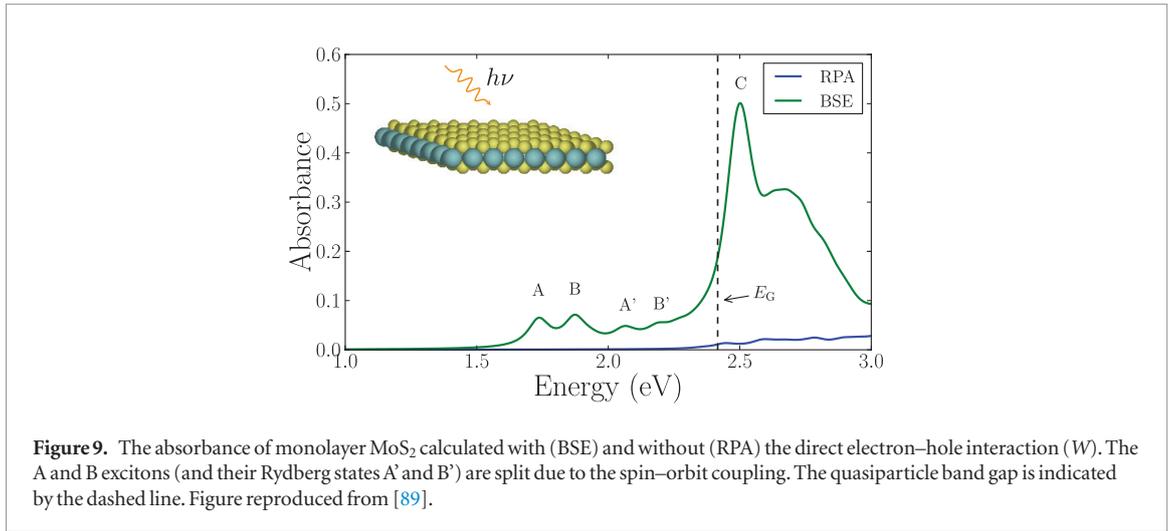
where W_i is the screened potential of the freestanding monolayer and ΔW_i is the change due to the surrounding layers. The latter term can be evaluated efficiently using the QEH method described in section 2.3. By combining equation (18) with equations (14) and (16) and assuming the Green function of the embedded layer to be the same as that of the freestanding monolayer (which is a good approximation due to the weak interlayer hybridisation), the change in QP energy can be readily computed. For details on the GΔW method

see [75]. Returning to figure 8, it is seen that the band gap correction is larger for the 1T-MoS$_2$ substrate than for graphene reflecting the increased screening from a metallic compared to a semi-metallic substrate. Moreover, as anticipated, the magnitude of the band gap correction is seen to scale with the inverse (static) polarisability of the 2D semiconductor.

The results of this section show that dielectric engineering can be used to manipulate the band gap of a 2D material and ultimately the vertical band structure profile of a vdWH. As another interesting application, it has been proposed that lateral control over the band edges within a homogeneous 2D material may be achieved by using a laterally structured dielectric environment [74].

4. Excitons

One of the hallmarks of the 2D semiconductors and insulators is the existence of strongly bound excitons with binding energies reaching up to 30% of the band gap. These excitons couple strongly to light and lead to a substantial modification of the optical spectrum both below and above the QP band gap. The large binding energies of excitons in 2D materials was initially predicted theoretically for monolayer hBN [81], graphane [23] and TMD monolayers [41, 82, 83], and has subsequently been confirmed experimentally [29, 84–86]. The reason for the large exciton binding energies in 2D materials is a combination of geometric confinement [87] and reduced dielectric screening [23, 41, 45]. The latter implies that exciton binding energies in 2D materials can be controlled via the dielectric environment, e.g. by encapsulation or deposition on substrates. Furthermore, by stacking 2D semiconductors

Figure 9. The absorbance of monolayer MoS$_2$ calculated with (BSE) and without (RPA) the direct electron–hole interaction (W). The A and B excitons (and their Rydberg states A' and B') are split due to the spin–orbit coupling. The quasiparticle band gap is indicated by the dashed line. Figure reproduced from [89].

with staggered band alignment, interlayer excitons with the electron and hole residing on different layers, can be formed. Such heterostructures provide an ideal platform for studying charge separation- and recombination processes in atomically well defined crystal environments.

4.1. Excitons in freestanding 2D materials

The most accurate method for calculating interband excitations in semiconductors and insulators is the Bethe–Salpeter equation (BSE) based on many-body perturbation theory [46, 88]. The excitation energies of crystal momentum **q** are found by solving an eigenvalue problem of the form

$$\sum_{S'} \mathcal{H}(\mathbf{q})_{SS'} F_S^\lambda(\mathbf{q}) = E^\lambda(\mathbf{q}) F_S^\lambda(\mathbf{q}) \qquad (19)$$

where $\mathcal{H}(\mathbf{q})$ is the BSE two-particle Hamiltonian evaluated in a basis of e–h states $\psi_S(\mathbf{r}_h, \mathbf{r}_e) = \psi_{n\mathbf{k}}(\mathbf{r}_h)^* \psi_{m\mathbf{k}+\mathbf{q}}(\mathbf{r}_e)$. The BSE Hamiltonian reads

$$\mathcal{H}_{SS'}(\mathbf{q}) = (\varepsilon_{m\mathbf{k}+\mathbf{q}}^{QP} - \varepsilon_{n\mathbf{k}}^{QP})\delta_{SS'} - (f_{m\mathbf{k}+\mathbf{q}} - f_{n\mathbf{k}})K_{SS'}(\mathbf{q}). \qquad (20)$$

The kernel contains two terms, namely the electron–hole exchange interaction (V) and the direct screened electron–hole interaction (W),

$$K_{SS'}(\mathbf{q}) = 2V_{SS'}(\mathbf{q}) - W_{SS'}(\mathbf{q}). \qquad (21)$$

The factor 2 accounts for spin. Neglecting the direct interaction (W) results in the random phase approximation (RPA). For a 2D material it can be shown that the exchange term becomes negligible in the small **q** limit (see section 5.6). Thus the e–h exchange only reduces the binding energy of the lowest zero momentum exciton of monolayer hBN and MoS$_2$ by 0.08 eV and 0.02 eV, respectively, which amounts to less than 5% of the total binding energy. To illustrate the important role of the direct e–h interaction for the optical properties of 2D materials, figure 9 shows the absorbance of a single layer of MoS$_2$ calculated from the BSE with and without inclusion of W. Since the e–h exchange term is negligible in the

optical limit ($\mathbf{q} = \mathbf{0}$), the RPA spectrum is essentially equal to the non-interacting result consisting of the bare e–h transitions. Inclusion of the direct e–h term changes the spectrum completely, introducing distinct peaks below and just above the QP band gaps due to absorption by excitons. The A and B excitons (and their Rydberg states A' and B') are composed of transitions from the spin–orbit split valence band to the conduction band at the K-point of the BZ. The C exciton involves transitions around the Γ-point [83]. The spectrum has been numerically broadened to simulate the effect of coupling to phonons at finite temperatures.

In the following we specialise to $\mathbf{q} = \mathbf{0}$ and thus neglect $V_{SS'}$. Furthermore, we specialise to a single valence and conduction band, i.e. $S = (v/c, \mathbf{k})$. This two-band approximation is typically excellent for the lowest bound excitons as long as the valence and conduction bands are separated from other bands at the relevant points of the BZ. Assuming that the wave functions are of the form $\psi_{v/c,\mathbf{k}}(\mathbf{r}) = \phi_{v/c}(z)\exp(i\mathbf{k}\cdot\mathbf{r}_\parallel)$, the direct screened interaction becomes,

$$W_{\substack{v c \mathbf{k} \\ v c \mathbf{k'}}} \equiv W(\mathbf{k} - \mathbf{k'})$$

$$= \int\int d\mathbf{r}d\mathbf{r}' |\phi_v(z)|^2 e^{i(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{r}_\parallel - \mathbf{r}_\parallel')} W(\mathbf{r}, \mathbf{r}') |\phi_c(z')|^2 \qquad (22)$$

In analogy to the 3D case it can be shown for a 2D semiconductor that the so-called exciton envelope function, $F(\mathbf{r}_\parallel) = \sum_{\mathbf{k}_\parallel} e^{-i\mathbf{k}_\parallel\cdot\mathbf{r}_\parallel} F(\mathbf{k}_\parallel)$, where $F(\mathbf{k}_\parallel)$ are the BSE expansion coefficients in equation (19), satisfies the 2D Schrödinger equation

$$\left[-\frac{\nabla_{2D}^2}{2\mu_{ex}} + W(\mathbf{r}_\parallel) \right] F(\mathbf{r}_\parallel) = E_B F(\mathbf{r}_\parallel). \qquad (23)$$

Here E_B is the exciton binding energy given as the difference between the QP band gap and the lowest optical transition, and μ_{ex} is the exciton effective mass, calculated from the hole and electron masses according to $\mu_{ex}^{-1} = m_e^{-1} + m_h^{-1}$. The function $F(\mathbf{r}_\parallel)$ describes the

probability amplitude for the distance between the electron and hole.

The screened interaction can be evaluated at different levels of approximations. First, the strict 2D approximation with the screened potential in equation (2) admits a closed analytical expression in real space

$$W_{2D}(\mathbf{r}_\parallel) = \frac{1}{4\alpha}[H_0(x) - N_0(x)]_{x=r_\parallel/2\pi\alpha}, \quad (24)$$

where $H_0(x)$ and $N_0(x)$ are the Struve and Neumann functions, respectively. Next, taking the finite width of the electron and hole charge densities into account by approximating the out-of-plane wave functions by step functions yields the quasi-2D screened potential [45]

$$W_{Q2D}(\mathbf{q}_\parallel) = \epsilon_{Q2D}^{-1}(\mathbf{q}_\parallel)V_{Q2D}(\mathbf{q}_\parallel), \quad (25)$$

where ϵ_{Q2D} was defined in equation (11). Finally, the out-of-plane variation of the electron and hole wave functions can be calculated from the exact wave functions at the relevant point of the BZ, $\rho_{e,h}(z) = \int_A d\mathbf{r}_\parallel |\psi_{c,v}(\mathbf{r}_\parallel, z)|^2$. However, as shown below, the exciton binding energies obtained from the effective Hamiltonian equation (23) are rather insensitive to the precise form chosen for the out-of-plane electron and hole densities.

Table 3 shows the binding energy of the lowest exciton in monolayer MoS_2 and hBN obtained from equation (23) with W evaluated in three different ways, namely (i) from equation (22) with $\phi_{v/c}$ given by the exact wave functions at the direct band gap (K point for MoS_2 and Γ point for hBN) (ii) from equation (25) corresponding to $\phi_{v/c}$ being step functions, and (iii) equation (24) corresponding to a strict 2D Coulomb potential. It can be seen that all the results obtained from the effective Hamiltonian agree within a few percent. The reason for this is that the Fourier transformed exciton wave function, $F(\mathbf{k}_\parallel)$, is localised in the small $\mathbf{k}_\parallel$ regime where the linear approximation, $\epsilon(\mathbf{q}_\parallel) = 1 + \alpha q_\parallel$, is valid. For comparison, the results obtained from the BSE are listed in the first column. It should be stressed that BSE calculations for 2D semiconductors, performed with a truncated Coulomb interaction to avoid screening between periodically repeated layers, converge very slowly with respect to the number of $\mathbf{k}$-points, see figure 10. The reason is again the strong variation of the screening around $\mathbf{q} = \mathbf{0}$ as discussed in detail in section 3 in the context of GW calculations [41, 62].

4.2. Screened 2D hydrogenic model

Apart from the dimensionality, the main difference between the hydrogen model of 3D excitons and the effective Hamiltonian equation (23) is that the screened interaction entering the latter is not simply $1/\epsilon r$. This implies that a closed expression for the exciton energies and radii cannot be directly obtained in the 2D case. However, as shown below it is possible to define an effective 2D dielectric *constant* and thereby recover a 2D Hydrogenic model which admits analytical

Table 3. Binding energy (in eV) of the lowest bound exciton in monolayer h-BN and MoS_2 evaluated using four different methods. Both the BSE and the model calculations are based on LDA input. The electron–hole exchange interaction is excluded in all calculations.

		Model equation (23)		
	E_B^{BSE}	E_B^{wfs}	E_B^{Q2D}	E_B^{2D}
h-BN	2.05	2.29	2.35	2.34
MoS_2	0.43	0.59	0.61	0.60

solutions. The effective dielectric constant is obtained by averaging $\epsilon(\mathbf{q})$ over the $\mathbf{q}$-vectors up to $1/a$ where a is the radius of the exciton in real space,

$$\epsilon = \frac{a^2}{\pi} \int_0^{2\pi} \int_0^{1/a} d\mathbf{q}\,\epsilon(\mathbf{q}), \quad (26)$$

The rationale behind this definition is that the $\mathbf{q}$-vectors coming into play when evaluating the average interaction energy $\int |F(\mathbf{q})|^2 W(\mathbf{q})d\mathbf{q}$, are mainly those fulfilling $|\mathbf{q}| < 1/a$. For the 2D Hydrogen atom the Bohr radius is given by $a = \epsilon/(2\mu_{ex})$ and equation (26) has to be solved self-consistently for ϵ given an expression for $\epsilon(\mathbf{q})$. In a strictly 2D system the screening is linear in $\mathbf{q}$ and equation (26) can be solved to yield [90]

$$\epsilon = \frac{1}{2}\left(1 + \sqrt{1 + 32\pi\alpha\mu_{ex}/3}\right). \quad (27)$$

Using that the Hydrogenic binding energy in 2D is a factor of four larger than in 3D [87], we obtain

$$E_B^{2D} = \frac{8\mu_{ex}}{\left(1 + \sqrt{1 + 32\pi\alpha\mu_{ex}/3}\right)^2}. \quad (28)$$

This expression constitutes a long-sought-for 2D analog of the 3D Hydrogenic exciton model.

A remarkable property of the expression (28) is that it becomes independent of the effective mass if the polarizability is sufficiently large. More precisely

$$E_B^{2D} \approx \frac{3}{4\pi\alpha}, \qquad 32\pi\alpha\mu_{ex}/3 \gg 1. \quad (29)$$

It may come as a surprise that the binding energy becomes independent of mass, since a large mass gives rise to a localized exciton and the binding energy typically increases with localization. However, in 2D, short range interactions are screened more effectively than long range interactions. Thus, there are two opposing effects of the exciton mass and for large polarizabilities, the binding energy becomes independent of mass. In order to assert the applicability of the expressions (28) and (29), we have calculated the effective masses and static polarizabilities (in the random phase approximation) of the 51 semiconducting TMDs shown in figure 5. In figure 11 we compare the model binding energies with the full solution of equation (23). Using the expression (28), the agreement is seen to be on the order of 10%. With the
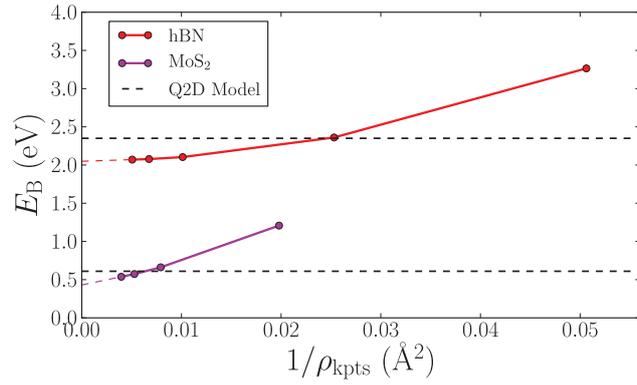
Figure 10. Convergence of the exciton binding energy obtained from the BSE with respect to the **k**-point density. The finest **k**-point grid corresponds to 60 × 60. Extrapolation to infinite **k**-point sampling is shown. The horizontal dashed lines show the results of the effective Hamiltonian (23) with the screened interaction W_{Q2D} of equation (25). Figure reproduced from [45].
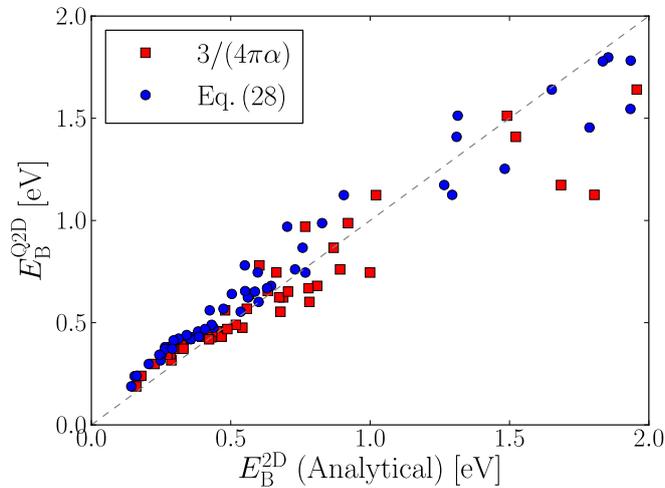
Figure 11. Exciton binding energies of the 51 monolayer transition metal dichalcogenides and -oxides listed in figure 5 calculated as the lowest eigenvalue of equation (23) with the screened interaction evaluated using $\epsilon_{Q2D}(\mathbf{q})$ (vertical axis) and the analytical Hydrogen model equations (28) and (29) (horizontal axis). Figure adapted from [90].

approximated expression (29), we also obtain excellent agreement for binding energies up to ∼0.5 eV, whereas the binding energies are somewhat underestimated for the more strongly bound excitons.

Recently, first principles calculations have indicated that exciton binding energies in 2D materials scale linearly with the band gaps [91]. In the present model, this behavior comes out naturally since the in-plane component of the polarizability, α, is roughly inversely proportional to the band gap as shown in [90]. Combining this with equation (29) thus gives $E_B^{2D} \propto E_{gap}$. However, in the present model the scaling originates solely from the screening and not the effective mass as previously proposed [91]. In fact, there is no direct correlation between the exciton binding energies and effective masses for this class of materials.

The screened Hydrogen model is not restricted to the lowest excitonic state, but can be used to account for the entire exciton series in 2D materials. In [92], the exciton series of graphene derivatives was predicted

to deviate from the 2D Rydberg series and in [93] this was experimentally demonstrated for a monolayer of WS$_2$ on SiO$_2$. The reason is that the effective screening depends on the n quantum number due to the increasing spatial extent of higher lying excitons. In [93] the authors defined an n-dependent effective dielectric constant, ϵ_n, which was determined by fitting each term in the Rydberg series to a 2D Hydrogen model. The Rydberg series is thus written

$$E_n^{2D} = -\frac{\mu_{ex}}{2\left(n - \frac{1}{2}\right)^2 \epsilon_n^2}. \tag{30}$$

We have previously shown that the Rydberg series of monolayer WS$_2$ can be accurately reproduced by the effective Hamiltonian equation (23) [37]. Thus the full solution of equation (23) is used as a reference for the Hydrogenic model in the following. We calculate the n-dependent effective screening by replacing a in equation (26) by an n-dependent characteristic extension of the state. To this end we note that for $l = 0$,
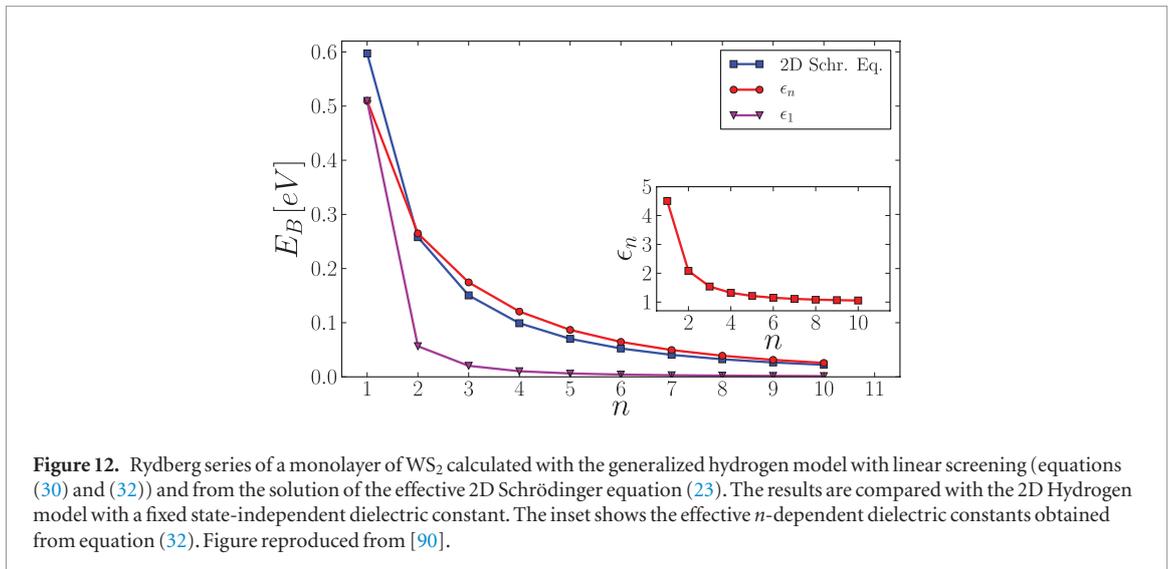
Figure 12. Rydberg series of a monolayer of WS_2 calculated with the generalized hydrogen model with linear screening (equations (30) and (32)) and from the solution of the effective 2D Schrödinger equation (23). The results are compared with the 2D Hydrogen model with a fixed state-independent dielectric constant. The inset shows the effective n-dependent dielectric constants obtained from equation (32). Figure reproduced from [90].

the radius of a state with principal quantum number n for a 2D Hydrogen atom with interaction $1/\epsilon r$ is [87]

$$a_n \equiv \langle n|\hat{r}|n\rangle = \epsilon[3n(n-1)+1]/(2\mu_{ex}), \quad (31)$$

where $\hat{r} = \sqrt{\hat{x}^2 + \hat{y}^2}$. With these more general definitions, the a defined previously is given by a_1 and E_B^{2D} is $-E_1^{2D}$. Using the 2D dielectric function $\epsilon(q) = 1 + 2\pi\alpha q$, the effective dielectric constant for state n then becomes

$$\epsilon_n = \frac{1}{2}\left(1 + \sqrt{1 + \frac{32\pi\alpha\mu_{ex}}{9n(n-1)+3}}\right). \quad (32)$$

Using a 2D polarizability of $\alpha = 5.25$ Å and $\mu_{ex} = 0.19m_0$, both obtained from first principles, we obtain the Rydberg series of a freestanding WS_2 monolayer shown in figure 12. Clearly the analytical results from the screened Hydrogenic model agree very well with the full solution of equation (23). In contrast, the pure 2D Hydrogen model with a fixed state-independent dielectric constant significantly underestimates the binding energies of higher lying states, since the decreased screening of extended states is not taken into account. We also note that the model binding energies of the $n = 1$ state agree very well with a full solution of the Bethe–Salpeter equation which yields an exciton binding energy of 0.54 eV [66]. The effective dielectric constants from equation (32) are shown in the inset and are in excellent agreement with the values fitted to experiments [93] (not shown).

It is straightforward to generalize the above expressions to $l \neq 0$ [87], which results in a larger value of the effective radius, a_{nl}, and thus $\epsilon_{n,l} < \epsilon_{n,l'}$ for $l > l'$. The energy is still given by equation (30) and for a given n, the higher angular momentum excitons will therefore have a larger binding energy. This effect was indeed observed experimentally for WS_2 monolayers [84].

Finally, it should be noted that, in addition to the nonlocal screening captured by the effective Hamiltonian equation (23) and the screened Hydrogen model, there are a number of other effects that can

lead to deviations from the Hydrogenic exciton series. These include the break down of the effective mass approximation, the two-band assumption as well as non-zero Berry curvature of the Bloch bands [94, 95]. All these effects are, however, related to the properties of the band structure rather than the dimensionality of the material and thus are not genuine 2D effects.

4.3. Finite momentum excitons

The above discussion has focussed on excitons with zero center of mass momentum, i.e. $\mathbf{q} = \mathbf{0}$. However, excitons with finite momentum, although optically inaccessible, are essential to describe exciton dynamics and the response of the electron system to momentum carrying probes such as x-rays or fast charged particles [96, 97]. In fact, the exciton $\mathbf{q}$-dispersion can be readily accessed by inelastic x-ray scattering or momentum resolved electron energy loss spectroscopy as discussed further in section 5.5.

In 2D materials, the electron–hole (e–h) exchange interaction entering the Bethe–Salpeter equation becomes proportional to $\mathbf{q}_{\parallel}$ for small momentum transfer, see discussion around equations (43)–(45). This has been shown to result in an anomalous linear (light-like) dispersion for excitons in some of the 2D semiconductors including MoS_2, phosphorene and hBN [98–100]. Only the singlet exciton exhibits linear dispersion because the exchange term vanishes for the triplet exciton where the electron and hole have opposite spin. Interestingly, the standard parabolic exciton dispersion that is inherited directly from the single-particle band energies, was found for the lowest singlet exciton of graphane. This indicates a suppression of the e–h exchange arising from a small spatial overlap of the electron and hole wave functions in this particular material [43]. It has been proposed that the difference in exciton dispersion, i.e. whether linear or parabolic, could be used to distinguish between Wannier- and Frenkel-like excitons in 2D semiconductors where the usual distinction based on the binding
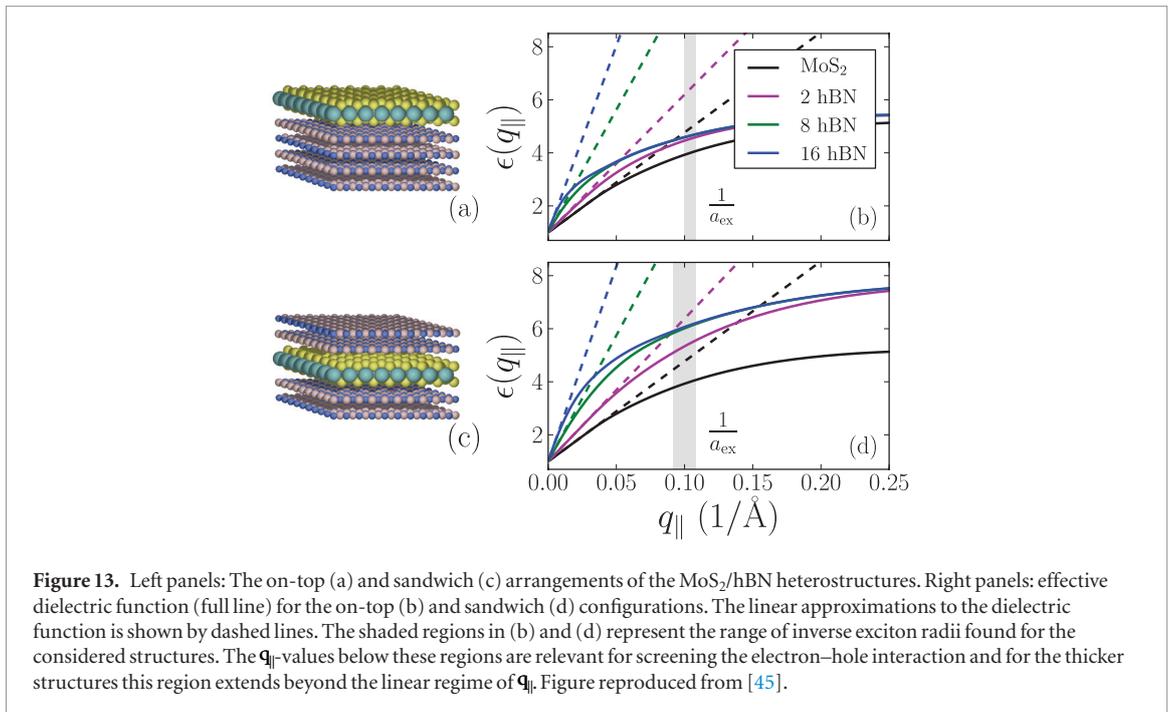
Figure 13. Left panels: The on-top (a) and sandwich (c) arrangements of the MoS$_2$/hBN heterostructures. Right panels: effective dielectric function (full line) for the on-top (b) and sandwich (d) configurations. The linear approximations to the dielectric function is shown by dashed lines. The shaded regions in (b) and (d) represent the range of inverse exciton radii found for the considered structures. The $\mathbf{q}_\parallel$-values below these regions are relevant for screening the electron–hole interaction and for the thicker structures this region extends beyond the linear regime of $\mathbf{q}_\parallel$. Figure reproduced from [45].

energy strength becomes difficult to maintain due to the universally large binding energies of 2D excitons [99].

4.4. Trions and biexcitons

Beyond excitons, the rich excitation spectrum of 2D materials features electron–hole complexes involving more than two particles. In doped semiconductors, free electrons (in n-doped regions) or holes (in p-doped regions) can interact with the neutral excitons forming three-particle complexes known as trions. Experimental binding energies of trions (relative to the free particle and the exciton) in monolayer TMDs are in the range 20–30 meV [24, 44, 101] while binding energies as high as 150 meV have been reported for phosphorene [102]. The size of the trion binding energy is thus larger than or comparable to the thermal energy at room temperature suggesting that trions should play active roles in the photophysics of monolayer semiconductors. For comparison, trion binding energies in conventional semiconductor quantum wells are in the range 1–5 meV [103].

The exciton model Hamiltonian equation (23) can be straightforwardly generalised to an N-particle 2D excitonic system

$$\hat{H}_N = -\frac{1}{2} \sum_{i=1}^{N} \frac{\nabla_i^2}{m_i^*} + \sum_{i<j}^{N} W(\mathbf{r}_\parallel^{(i)} - \mathbf{r}_\parallel^{(j)}) \qquad (33)$$

where $\mathbf{r}_\parallel^{(i)}$ and m_i^* are the position and effective mass of particle i. By adopting relative coordinates the problem reduces to an $(N-1)$-particle problem. For trions ($N = 3$), this type of Hamiltonian, with the analytical 2D potential of equation (24) and parameters (effective masses and screening lengths) obtained from DFT, has been solved for the MoX$_2$ and WX$_2$ monolayer TMDs by variational optimisation of simple trial

wave functions [44], path-integral Monte Carlo [104], and explicitly correlated Gaussians [105]. In all cases, reported binding energies for the lowest trion lie within a factor of two of the experimental values. Using similar methods, biexciton ($N = 4$) binding energies of around 20 meV were predicted, in reasonable agreement with experiments (50–70 meV) [106, 107].

Recently, a first-principles approach to trion excitations based on an extension of the BSE to three-particle states was developed and applied to carbon nanotubes [108]. The significantly larger size of the configurational space of trions as compared to excitons prohibits standard diagonalization of the three-body Hamiltonian. Fortunately, all Coulomb interaction terms are of two-particle nature and the resulting Hamiltonian is very sparse, allowing for iterative solutions of the corresponding eigenvalue problem [108].

4.5. Excitons in van der Waals heterostructures

It was shown in the previous section that the simple 2D model for dielectric screening, $\epsilon_{2D}(\mathbf{q}) = 1 + 2\pi\alpha q$, gives an accurate description of excitons in isolated atomically thin materials. The reason is that typical exciton radii in monolayer semiconductors and insulators are larger than the thickness of the material such that the screening is mainly governed by small $\mathbf{q}$-vectors where $\epsilon_{2D}(\mathbf{q})$ is accurate, see figure 2. This condition is bound to break down for thicker structures or supported/encapsulated 2D materials. In such cases the full non-linear $\mathbf{q}$-dependence of the dielectric function must be taken into account.

Figure 13 shows the static dielectric function of monolayer MoS$_2$ deposited onto (upper panel) and embedded inside (lower panel) an N-layer stack of hBN. The dielectric function of the MoS$_2$ monolayer has been calculated using the QEH model, see [45] for details. It is clear that the region of $\mathbf{q}_\parallel$-vectors where the strict
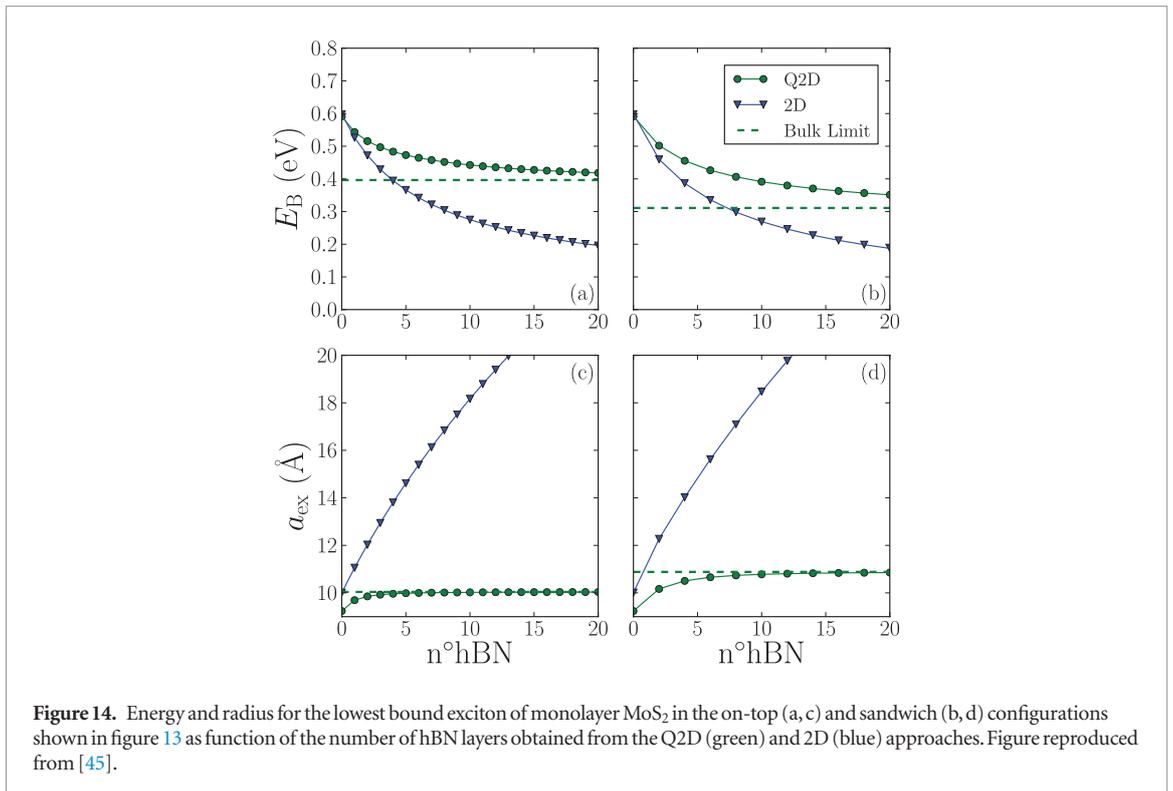
Figure 14. Energy and radius for the lowest bound exciton of monolayer MoS$_2$ in the on-top (a, c) and sandwich (b, d) configurations shown in figure 13 as function of the number of hBN layers obtained from the Q2D (green) and 2D (blue) approaches. Figure reproduced from [45].

2D screening model (dashed lines) is valid shrinks as the number of hBN layers increases. The exciton binding energies and radii obtained from the effective Hamiltonian equation (23) with a screened interaction evaluated using (i) the exact $q_\parallel$-dependent dielectric function and (ii) the linear 2D screening model, are shown in figure 14. The 2D model underestimates the binding energy and overestimates the radius significantly for more than just a few hBN layers and the results diverge in the limit of infinite hBN support. The reason is that the size of the exciton calculated using method (i), and indicated by the vertical grey lines, does not change much with the number of hBN layers and consequently the range of $q_\parallel$-vectors that are important for the screening also remains fairly constant.

As another example, we study the 2D to 3D transition of the exciton in MoS$_2$. In layered bulk materials, the exciton equation can be written as follow:

$$\left[-\frac{\nabla_\parallel^2}{2\mu_\parallel} - \frac{\nabla_\perp^2}{2\mu_\perp} + W(\mathbf{r}) \right] F(\mathbf{r}) = E_B F(\mathbf{r}). \quad (34)$$
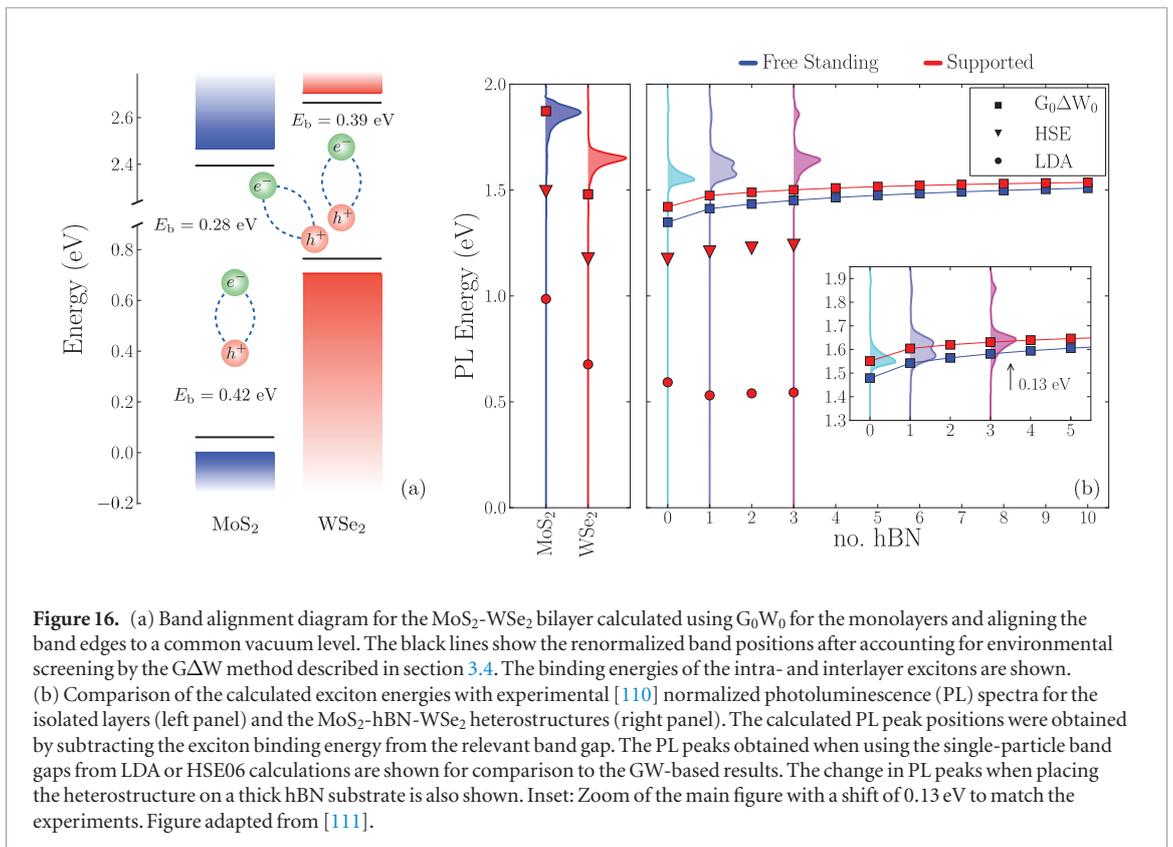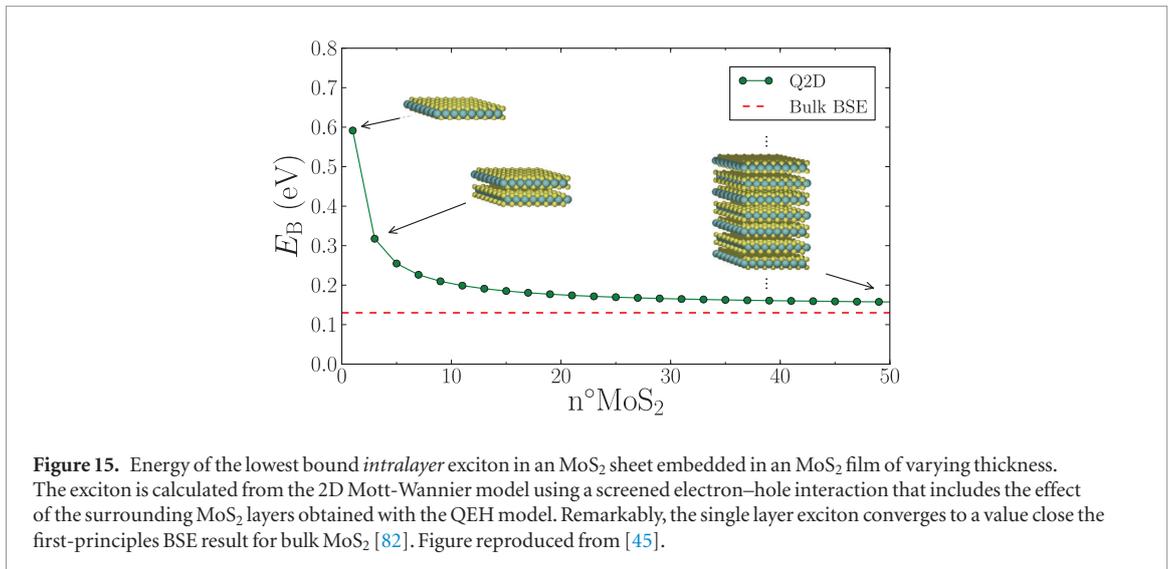
In layered materials like MoS$_2$, the exciton mass in the out of plane direction is typically much higher than in the in-plane directions ($\mu_\perp \gg \mu_\parallel$). Consequently, we can neglect the out of plane component of the kinetic energy and be left with the 2D exciton model. Moreover, the in-plane effective mass does not differ considerably between monolayer and bulk MoS$_2$ [109]. Therefore, one would expect the main difference between the physics of excitons in monolayer and layered bulk to be governed by the screened potential rather than the geometric confinement. To test this hypothesis we consider a multilayer MoS$_2$ structure and calculate the binding energy of an exciton localized in the central MoS$_2$ layer using the

effective Hamiltonian equation (23) with the screened potential calculated from the QEH model. The results for the binding energy as a function of the number of MoS$_2$ layers are plotted in figure 15. As expected, the reduction of the exciton binding energy is larger when the monolayer is embedded in MoS$_2$ compared to hBN (see figure 14 panel (b)). Remarkably, the binding energy converges towards a value of 0.16 eV only 0.03 eV higher than previously reported first principles BSE results for bulk MoS$_2$ [82]. This shows that the large difference in exciton binding energies of 2D and 3D (layered) bulk semiconductors comes mainly from the difference in screening while quantum confinement plays a minor role.

4.6. Interlayer excitons

In addition to the intralayer excitons discussed in previous sections, vdWHs can host more complex types of excitons where the electron and hole reside in different layers, so-called (spatially) indirect excitons or interlayer excitons. Because of the spatial separation of the electron and hole, interlayer excitons posses longer lifetimes [112, 113] than intralayer excitons, which make them ideal candidates for realization of exotic many-particle states such as room temperature Bose–Einstein condensates [114]. Indirect excitons also play a central role in the important charge separation process of excitonic solar cells [115, 116]. In this case, the binding energy of the indirect exciton determines the minimum band offset required to separate the electron and hole and thus the achievable output voltage.
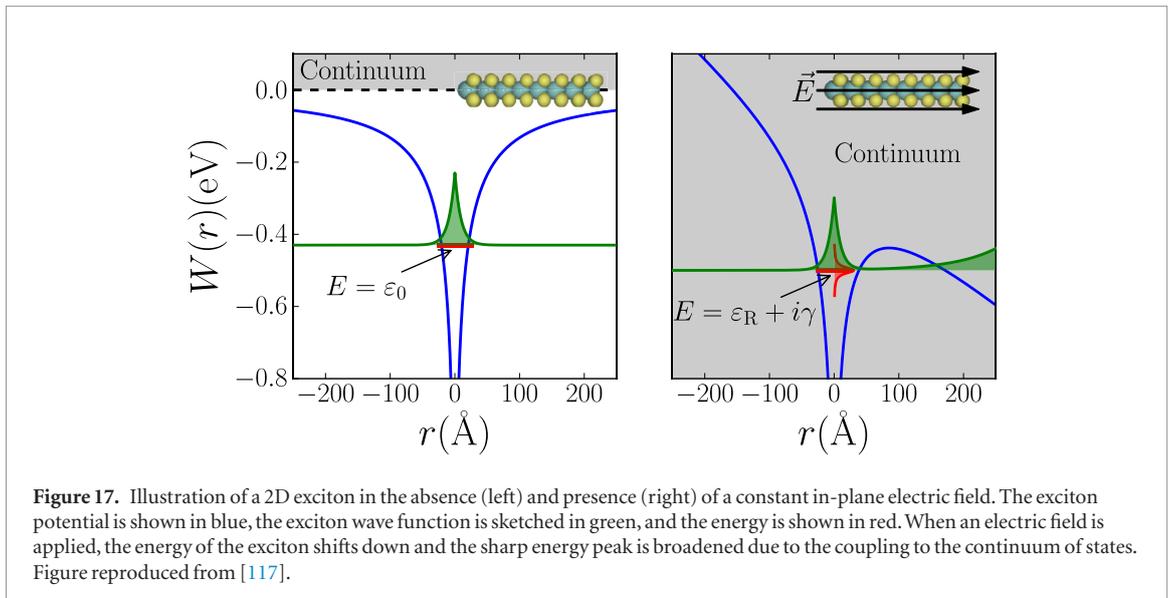
A realistic model of interlayer excitons in vdWHs must (i) start from an accurate description of the band alignment at the heterointerface and (ii) account for the

Figure 15. Energy of the lowest bound *intralayer* exciton in an MoS_2 sheet embedded in an MoS_2 film of varying thickness. The exciton is calculated from the 2D Mott-Wannier model using a screened electron–hole interaction that includes the effect of the surrounding MoS_2 layers obtained with the QEH model. Remarkably, the single layer exciton converges to a value close the first-principles BSE result for bulk MoS_2 [82]. Figure reproduced from [45].

Figure 16. (a) Band alignment diagram for the MoS_2-WSe_2 bilayer calculated using G_0W_0 for the monolayers and aligning the band edges to a common vacuum level. The black lines show the renormalized band positions after accounting for environmental screening by the $G\Delta W$ method described in section 3.4. The binding energies of the intra- and interlayer excitons are shown. (b) Comparison of the calculated exciton energies with experimental [110] normalized photoluminescence (PL) spectra for the isolated layers (left panel) and the MoS_2-hBN-WSe_2 heterostructures (right panel). The calculated PL peak positions were obtained by subtracting the exciton binding energy from the relevant band gap. The PL peaks obtained when using the single-particle band gaps from LDA or HSE06 calculations are shown for comparison to the GW-based results. The change in PL peaks when placing the heterostructure on a thick hBN substrate is also shown. Inset: Zoom of the main figure with a shift of 0.13 eV to match the experiments. Figure adapted from [111].

non-local screening of the electron–hole interaction by the 2D host materials. Below the importance of these effects are illustrated using the heterostructure MoS_2/WSe_2 as an example. As demonstrated experimentally in [110], by intercalating hBN spacer layers between the optically active TMD layers it is possible to vary the electron–hole separation with atomic precision making this system ideally suited for exploring the properties of interlayer excitons.

It was demonstrated in section 3, that QP band structures of isolated monolayers can be accurately calculated by the GW method. Moreover, it was shown how the effect of interlayer screening on the band structure of vdWHs can be included within a GW

framework by utilizing the QEH model to evaluate the change in screened interaction (ΔW) due to the surrounding layers. Figure 16(a) shows the G_0W_0@LDA calculated band alignment for the MoS_2/WSe_2 bilayer. The coloured bars represent the vacuum level aligned monolayer band structures (same as figure 6) while the black lines show the result when interlayer screening is included via the $G\Delta W$ method. The effect is rather small due to the modest screening provided by a single neighbouring layer and is comparable to the band gap reduction of ~0.2 eV found for MoS_2 on graphene in figure 8. It should be noted that DFT calculations performed for realistic MoS_2/WSe_2 heterostructure configurations employing large supercells to account for the

Figure 17. Illustration of a 2D exciton in the absence (left) and presence (right) of a constant in-plane electric field. The exciton potential is shown in blue, the exciton wave function is sketched in green, and the energy is shown in red. When an electric field is applied, the energy of the exciton shifts down and the sharp energy peak is broadened due to the coupling to the continuum of states. Figure reproduced from [117].

lattice mismatch, show that the hybridisation between the two layers is negligible, in particular around the relevant band extrema at the *K*-point [111].

An interlayer exciton can be modelled as a bound electron–hole pair with the particle motion confined to two normally displaced 2D planes. This implies that the binding energy of the interlayer exciton can be computed from the usual 2D exciton Hamiltonian equation (23) provided that the exciton mass is calculated from the effective electron and hole masses of the relevant layers and that the screened interaction takes the (constant) out-of-plane separation of the electron and hole into account. Figure 16(a) shows the binding energies of the intra- and interlayer excitons of the MoS_2/WSe_2 bilayer calculated in this way. The effective masses were obtained from the monolayer band structures and the screened interaction was calculated using the QEH model, see [111]. The binding energy of the interlayer exciton (0.28 eV) is only around 0.1 eV smaller than the binding energy of the intralayer excitons (0.39–0.42 eV). Upon insertion of hBN spacer layers, the binding energy of the interlayer exciton decreases as ∼1/*d* as the distance between the electron and hole (*d*) increases, while the intralayer excitons are only weakly affected [111].
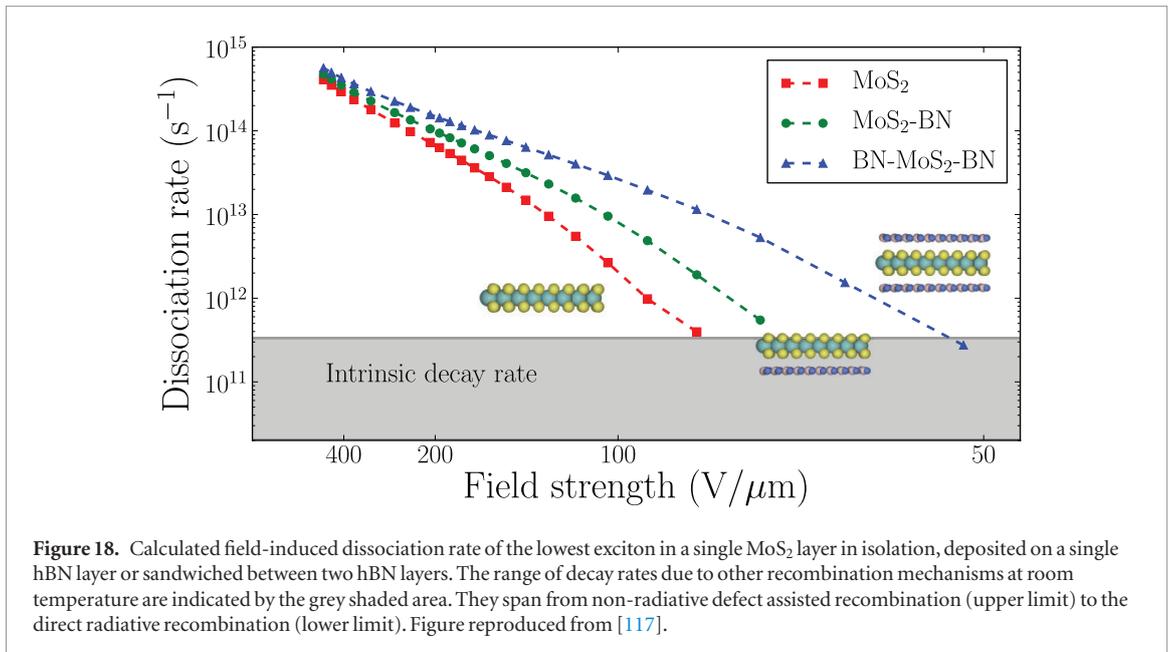
In [110] the photoluminescence (PL) due to exciton recombination in MoS_2/WSe_2-based heterostructures was measured. The measured PL spectra are reproduced in figure 16(b) for the individual MoS_2 and WSe_2 monolayers (left panel) and the hBN intercalated heterostructures (right panel). The position of the lowest optical transition calculated by subtracting the exciton binding energy from the QP band gap calculated using different methods is shown by the symbols. When the G_0W_0 band gap is used, the calculated PL peaks are in excellent agreement with the experiments for both the intralayer and interlayer excitons. The small underestimation (∼0.1 eV) of the interlayer exciton peak is also found for the WSe_2 intralayer exciton indicating that the origin is a too high positioning of the WSe_2 valence band by the G_0W_0 method.

4.7. Field-induced exciton dissociation

As illustrated by the previous sections, excitons in atomically thin 2D semiconductors can be extremely strongly bound. While this has some exciting fundamental perspectives [114], it is problematic for applications such as photodetectors and solar cells which rely on efficient conversion of photons into electrical currents. Indeed, photocurrent measurements on suspended MoS_2 show that large voltages are required to dissociate the excitons into free carriers [85]. On the other hand, the previous section showed that van der Waals heterostructuring could be employed as a strategy to weaken the electron–hole interaction and thereby lower the exciton binding energy without altering the band structure of the 2D semiconductor (apart from a reduction of the band gap due to image charge screening as described in section 3.4). As shown in the following, encapsulation of a single MoS_2 layer by just two layers of hBN is sufficient to increase exciton dissociation rates by an order of magnitude [117].

When a constant electric field is applied to an exciton, it will eventually decay into a free electron and hole [118, 119]. The application of a constant electric field changes the exciton from a bound state to a resonance with a finite spectral line width, $\gamma = \hbar/\tau$, where τ is the resonance life time or equivalently the inverse exciton dissociation rate, see figure 17. There are generally two approaches used to compute resonances. The so-called indirect methods identify resonances as the poles of the scattering amplitude analytically extended into the complex energy plane [120], while the direct methods obtain the resonance states directly as the eigenstates of a complex scaled non-hermitian Hamiltonian [121].

Figure 18 shows the dissociation rates of the lowest exciton in MoS_2 as a function of in-plane field strength with and without hBN encapsulation. The rates have been obtained by complex scaling of the exciton Hamiltonian equation (23) using the QEH model to include the screening from hBN. As expected, larger fields lead to shorter lifetimes and the dependence of the rate on the field strength is approximately $\propto \exp(-V_0/V)$ for the

Figure 18. Calculated field-induced dissociation rate of the lowest exciton in a single MoS_2 layer in isolation, deposited on a single hBN layer or sandwiched between two hBN layers. The range of decay rates due to other recombination mechanisms at room temperature are indicated by the grey shaded area. They span from non-radiative defect assisted recombination (upper limit) to the direct radiative recombination (lower limit). Figure reproduced from [117].

considered field strengths. It can also be seen that the dissociation rate can be tuned to a high degree by encapsulating the MoS_2 layer in hBN. This is as expected, since the dielectric screening from the hBN weakens the e–h interaction which makes the exciton dissociate more readily.

In a real device, the field-induced dissociation of excitons competes with other decay mechanisms such as direct radiative recombination [122], defect-assisted non-radiative recombination [66], and exciton-exciton annihilation [123]. The relative importance of these effects is highly dependent on temperature, defect concentration, and exciton density.

At very low temperatures, the direct radiative decay of zero momentum excitons dominates, with a characteristic lifetime of ∼200 fs. [122, 124] At room temperature, most of the excitons have non-vanishing momenta, and the radiative recombination lifetime is on average ∼1 ns [66, 122]. Under such conditions, defect-assisted recombination becomes the dominant mechanism reducing exciton lifetimes to 2–5 ps [66, 125]. Exciton-exciton annihilations become important only when the density of excitons in a sample is large. At a density of 1×10^{12} cm^{-2}, the effective lifetime due to annihilation is around 10 ps [123].

The calculations show that for field strengths larger than 100 V μm^{-1}, the dissociation lifetime is shorter than one picosecond for all the systems considered. This is shorter than the smallest characteristic lifetimes of the alternative decay channels at room temperature (indicated by the grey shaded region in figure 18) and thus field-induced dissociation should dominate. We note that a potential gradient of 100 V μm^{-1} is not unlikely to exist in the metal-MoS_2 contact region where charge transfer and interface dipole formation driven by Fermi level mismatch can lead to significant variations in the potential and band energies even in the absence of an applied bias voltage. Recently, a chemical treatment has successfully been used to eliminate the influence of defects on exciton decay, leaving radiative decay as

the main decay channel and leading to exciton lifetimes of 10 ns [40]. Under such conditions, field dissociation should become the dominant decay mechanism for field strengths as small as 10 V μm^{-1}. Recently, the complex scaling method was also used to study field-induced exciton dissociation in layered bulk TMDs [126] and electroabsorption in monolayer TMDs [127].

5. Plasmons

Plasmons in metals are quantised versions of the collective oscillatory modes of the electron gas against the background of positively charged ions [128]. By their collective nature, plasmons couple strongly to electromagnetic fields. This makes it possible to utilize metallic structures to confine, enhance, or guide light fields and thereby enable a range of applications from single-molecule spectroscopy [129, 130] to energy conversion [131, 132] and imaging [133, 134]. In this context, 2D materials present a number of unique properties which could lead to improved performances or new functionalities of plasmonic devices. For example, graphene plasmons are long lived, can be easily tuned via electrostatic or chemical doping, and can be used to confine light to extremely small volumes [135, 136]. In addition, van der Waals stacking of different 2D metals and dielectrics opens up the possibility of controlling plasmons on the atomic length scale and for creating deep-subwavelength metamaterials with unique optical properties that cannot be found in nature or achieved with conventional heterostructures.

5.1. The Drude model and its limitations
The semi-classical Drude model, which describes the electron system as a gas of classical particles subject to a constant empirical friction force, is the workhorse for describing electron dynamics in metal structures. Within the Drude model, the dielectric function of a metal takes the form

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega + i/\tau)} \qquad (35)$$

where $\omega_p = \sqrt{4\pi n/m^*}$ is the bulk plasma frequency (n the electron density and m^* the effective carrier mass) and τ is the empirical relaxation time. For simple bulk metals and surfaces, equation (35) provides an excellent description of the optical properties for frequencies below the onset of interband transitions and $1/q \gg \min\{v_F/\omega, v_F\tau\}$, where v_F is the Fermi velocity. The latter condition expresses that the wave length of the driving field is either longer than the distance travelled by an electron during one period of the field or the distance travelled by an electron between two collisions, i.e. the mean free path, $v_F\tau$. This condition defines the regime where the **q**-dependence of the dielectric function, also known as spatial dispersion or non-local response, can be neglected. For finite nano structures, the optical response naturally involves large **q** vectors and non-local effects become important [137–139]. Furthermore, surface scattering and spill-out effects must be taken into account in this regime [140]. For noble- and transition metals, the semi-classical picture must be modified to account for the less mobile d-electrons as well as the additional damping due to $s - d$ interband transitions [141]. For semiconductors, a full quantum theory is required to describe interband plasmons which do not have a direct classical analogue.

Quantum mechanically, the plasmon is a collective excited state which can be represented as a coherent superposition of electron–hole transitions (intraband or interband or a mix thereof). As such the plasmon state, in particular for interband plasmons, is very similar to the exciton state. However, while the energy of the exciton is lower than the energy of its constituent electron–hole pairs, the plasmon energy is higher. Thus one way to differentiate between the two types of excitations is to say that the collective nature of the states, i.e. the coherent rather than incoherent superposition of electron–hole pairs, leads to an increased (decreased) Coulomb energy for plasmons (excitons). As will be shown in section 5.6 the response of the two types of collective excitations to external probes is very different in 2D materials.

5.2. Plasmons in 2D metals

For an isotropic metallic film of charge density n and thickness d, solution of Maxwell's equations in the non-retarded regime (speed of light $c \to \infty$) yields two plasmons with energy dispersion [142]

$$\omega(\mathbf{q}) = \frac{\omega_p}{\sqrt{2}}\sqrt{1 \pm e^{-qd}} \qquad (36)$$

The two solutions correspond to tangential and normal oscillations of the electron gas. For $qd \gg 1$, the energy of both plasmon modes converge towards the surface plasmon energy of a semi-infinite metal, $\omega_p/\sqrt{2}$. This happens because the penetration depth of the field

associated with the plasmon becomes smaller for larger **q**, and consequently the density oscillations on the two surface become decoupled in this regime. For an atomically thin conductor, we can assume $qd \ll 1$, and the tangential solution becomes

$$\omega_{2D}(\mathbf{q}) = \omega_p\sqrt{qd/2} \qquad (37)$$

In contrast to 3D metals where the plasmon energy remains constant for $\mathbf{q} \to \mathbf{0}$, the 2D plasmon vanishes in this limit.

To understand the different behaviour of plasmons in 2D and 3D, it is instructive to consider the Dyson equation that links the long range ($\mathbf{G} = \mathbf{0}$) components of the irreducible and reducible response functions via the Coulomb interaction,

$$\chi_{00}(\mathbf{q}, \omega) = \chi_{00}^{irr}(\mathbf{q}, \omega) + \chi_{00}^{irr}(\mathbf{q}, \omega)V(\mathbf{q})\chi_{00}(\mathbf{q}, \omega) \quad (38)$$

The irreducible response function includes local field effects and e–h exchange-correlation effects. It satisfies the Dyson equation

$$\chi^{irr}(\mathbf{q}, \omega) = \chi^0(\mathbf{q}, \omega) + \chi^0(\mathbf{q}, \omega)[V^{sr}(\mathbf{q}) + f_{xc}(\mathbf{q}, \omega)]\chi^{irr}(\mathbf{q}, \omega) \quad (39)$$
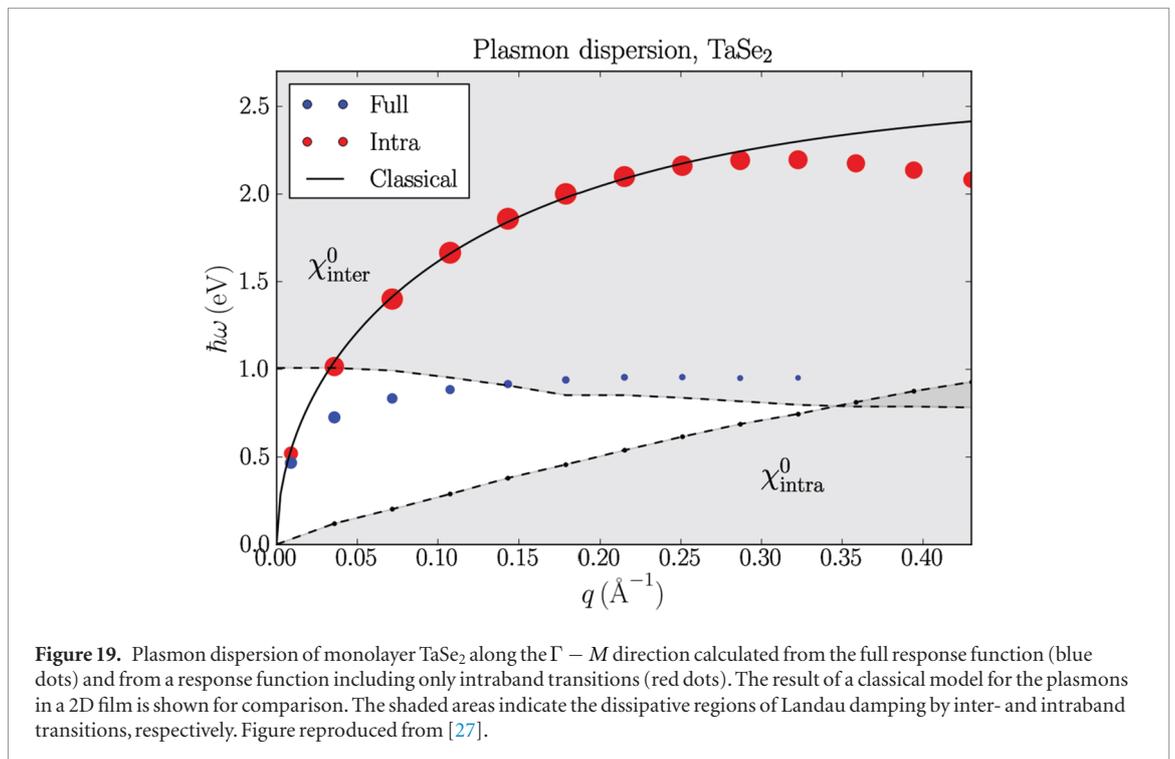
where matrix multiplication in the **G** and **G**′ indices is implied. χ^0 is the non-interacting response function from equation (8), $V_{\mathbf{GG'}}^{sr}(\mathbf{q}) = \delta_{\mathbf{GG'}}[V(\mathbf{G} + \mathbf{q}) - \delta_{\mathbf{G0}}V(\mathbf{q})]$ is the short range part of the Coulomb interaction, and $f_{xc}(\mathbf{q}, \omega)$ is the xc-kernel. For metals, the simplest adiabatic local density approximation (ALDA) to f_{xc} is often sufficient for describing the plasmonic properties and generally improves the RPA results.

Physically, the difference between χ_{00} and χ_{00}^{irr} is that the former includes long-range screening, i.e. the effect of potentials arising from variations in the induced density on length scales larger than the unit cell. In 3D metals, these are the potentials leading to plasmon formation, but in 2D, they are too weak to sustain the plasmon in the small **q** limit. In general, i.e. independent of system dimensionality, $\chi_{00}^{irr}(\mathbf{q}, \omega) \propto q^2$ in the small **q** limit (for metals this holds for $\omega \gg v_F q$). For a 2D material, since $V_{2D}(\mathbf{q}) \propto 1/q$ for $q \to 0$, it follows from equation (38) that

$$\chi_{00}^{2D}(\mathbf{q}, \omega) = \chi_{00}^{2D,irr}(\mathbf{q}, \omega), \text{ for } \mathbf{q} \to \mathbf{0} \qquad (40)$$

This means that also for metals, $\epsilon^{2D}(\mathbf{q}, \omega) = 1$ for $\mathbf{q} \to \mathbf{0}$ and finite ω. The fact that there is no difference between the total and external macroscopic fields means that there is no distinction between the loss spectrum and optical absorption spectrum of a 2D material.

It should be stressed that, in contrast to the metallic intraband plasmons, the formation of interband plasmons can be driven by short range Coulomb interactions and thus may be present, although weak, in the $\mathbf{q} \to \mathbf{0}$ limit. An example illustrating the different behavior of intraband and interband plasmons in 2D is provided by doped graphene. The doping introduces free carriers in the π or π^* bands which leads to the formation of a metallic intraband plasmon with the $\sqrt{q}$-dispersion characteristic of a 2D metal in the

Figure 19. Plasmon dispersion of monolayer TaSe$_2$ along the $\Gamma - M$ direction calculated from the full response function (blue dots) and from a response function including only intraband transitions (red dots). The result of a classical model for the plasmons in a 2D film is shown for comparison. The shaded areas indicate the dissipative regions of Landau damping by inter- and intraband transitions, respectively. Figure reproduced from [27].

long wavelength limit [26, 143]. In contrast, the $\pi - \pi^*$ interband plasmon approaches a finite energy for $\mathbf{q} \to 0$ [144]. Interband plasmons are discussed further in section 5.6.

5.3. Plasmons in metallic transition metal dichalcogenides

A highly attractive feature of the metallic plasmon in doped graphene is that its frequency can be tuned by controlling the density of free carriers e.g. via electrostatic gating. On the other hand, the energy of the metallic plasmon is restricted by the achievable carrier concentration and this limits applications of graphene plasmonics to the terahertz and infrared regimes. Metallic 2D materials have much higher charge carrier densities leading to plasmon energies up to 1 eV or above and could potentially extend 2D plasmonics to the near infrared or even optical regimes.

In [27] the (intraband) plasmons of six different metallic monolayer TMDs were explored by first-principles RPA calculations. As an example, figure 19 shows the plasmon dispersion of TaSe$_2$. The grey shaded areas indicate the dissipative regions of Landau damping formed by inter- and intraband transitions, respectively. The important role of the interband transitions for the plasmon energy and strength (indicated by the dot size) is evident from the large difference between the results of a full calculation based on the true response function which includes all possible transitions (blue dots), and a restricted calculation based on a response function which includes only intraband transitions (red dots). Clearly, the pure intraband plasmon is well described by the 2D free electron model equation (37) while the true plasmon is significantly redshifted by the coupling to interband transitions.

In addition to the extended sheet plasmons, certain 2D materials have been predicted to host one-dimensional (1D) plasmons. For example, first-principles calculations revealed the existence of such 1D plasmons along the metallic edges of MoS$_2$ nanoribbons [145]. Furthermore, as the metallic edge states have been proposed to be a general feature of polar 2D materials like the TMDs [146], 1D plasmons may be found at edges or domain boundaries of many other 2D materials.

5.4. Plasmons in graphene/hBN heterostructures

The prospects of graphene plasmonics were long obstructed by the strong damping of the its plasmons. Recently, it was demonstrated that by encapsulating graphene by two hBN films the losses can be significantly reduced, presumably down to the intrinsic limit set by Landau damping, electron-phonon and electron-electron interactions within the graphene itself [136]. Below we explore some properties of plasmons in graphene/hBN/graphene heterostructures with a focus on the possibility of tuning the plasmon energy *via* plasmon-plasmon hybridisation and dielectric screening.

The close to perfect lattice match between graphene and hBN allows full first-principles calculations to be performed for the thinnest heterostructures, see figure 20. Here we use doped graphene that has a finite density of states at the Fermi level, giving rise to metallic plasmons with energies in the range 0–2 eV depending in the in-plane momentum transfer. Note that such high plasmon energies require extremely high doping levels that are probably not achievable in practice. The plasmon energies go to zero in the optical limit as is characteristic for plasmons in 2D metals, see equation (36). We calculate the effect of hBN on the plasmons using the QEH model for up to 20 layers of
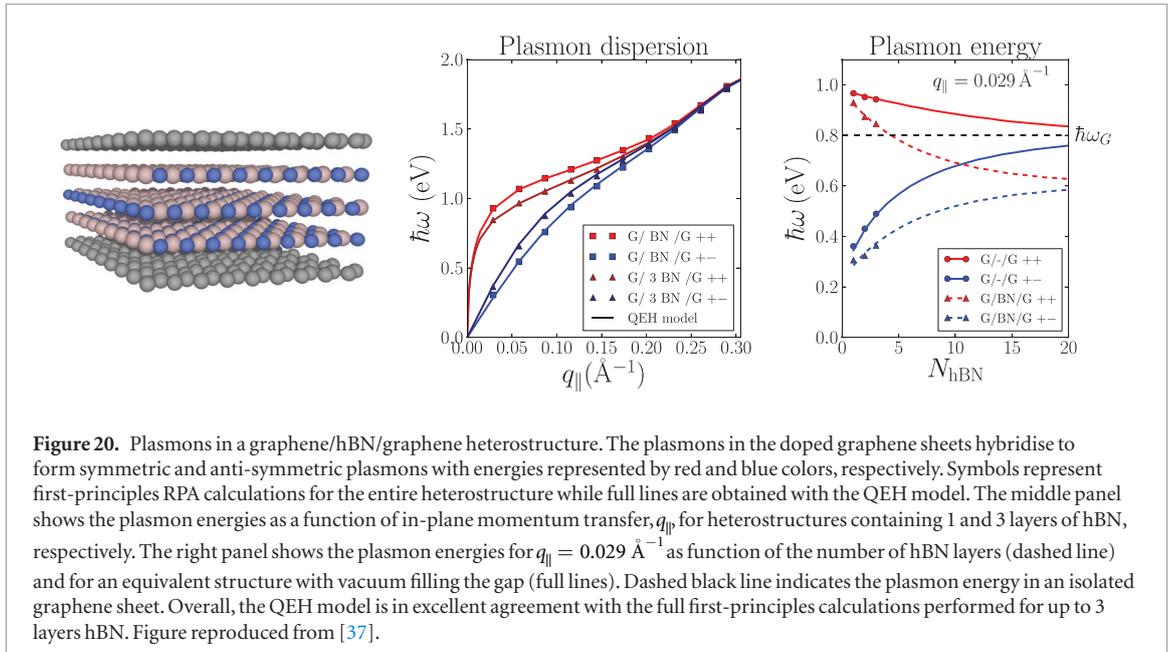
Figure 20. Plasmons in a graphene/hBN/graphene heterostructure. The plasmons in the doped graphene sheets hybridise to form symmetric and anti-symmetric plasmons with energies represented by red and blue colors, respectively. Symbols represent first-principles RPA calculations for the entire heterostructure while full lines are obtained with the QEH model. The middle panel shows the plasmon energies as a function of in-plane momentum transfer, $q_{\parallel}$, for heterostructures containing 1 and 3 layers of hBN, respectively. The right panel shows the plasmon energies for $q_{\parallel} = 0.029$ Å^{-1} as function of the number of hBN layers (dashed line) and for an equivalent structure with vacuum filling the gap (full lines). Dashed black line indicates the plasmon energy in an isolated graphene sheet. Overall, the QEH model is in excellent agreement with the full first-principles calculations performed for up to 3 layers hBN. Figure reproduced from [37].

hBN and compare to full first-principles calculations for 1–3 layers of hBN.

To identify the plasmons of the heterostructure we follow [147]. In brief, we compute the eigenvalues, $\epsilon_n(\omega)$, of the heterostructure dielectric function for each frequency point and identify a plasmon energy, $\hbar\omega_p$, from the condition $\mathrm{Re}\,\epsilon_n(\omega_p) = 0$. The corresponding eigenvector, $\phi_n(\omega_p)$, represents the potential associated with the plasmon oscillation (not shown). This analysis identifies two plasmons corresponding to the symmetric $(++)$ and antisymmetric $(+-)$ combinations of the graphene plasmons as previously found for two freestanding graphene sheets [148]. More details on these calculations can be found in [37].

The middle panel of figure 20 shows the calculated dispersion of the symmetric (red) and antisymmetric (blue) plasmons for 1 and 3 layers of hBN, respectively. For $q_{\parallel}d \gg 1$, the charge density oscillations in the two graphene sheets do not interact and both plasmon modes converge to that of a single freestanding graphene sheet. The right panel shows the evolution of the plasmon energies at a fixed in-plane momentum transfer as function of the distance between the two graphene sheets. The full (dashed) lines represent the case of vacuum (hBN) filling the gap. As expected, the hBN dielectric redshifts the plasmon energies by screening the interaction between the two charge densities. For 1–3 hBN layers, the QEH model perfectly reproduces the first-principles results for the plasmon energy.

5.5. Electron energy loss spectroscopy

Low loss electron energy loss spectroscopy (LL-EELS) is a commonly used technique to probe plasmons and interband transitions. In angular resolved LL-EELS the momentum loss is simultaneously recorded which allows the dispersion of the plasmon to be determined. As a complementary tool, spatially resolved LL-EELS can provide information about the energy loss at a specific point in space. Today, state of the art scanning transmission electron microscopes (STEM) provide loss spectra with simultaneous spatial and energy resolutions below 1 Å and 10 meV, respectively [149].

From Fermi's Golden rule and the spectral representation of the reducible response function χ, the power absorbed by the material when exposed to an external potential of the form $V_{\mathrm{ext}}(\mathbf{r}, t) = V_{\mathrm{ext}}(\mathbf{r})e^{i\omega t}$, becomes

$$P(\omega) = -\mathrm{Im} \int\int \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}'\, V_{\mathrm{ext}}(\mathbf{r})^* \chi(\mathbf{r}, \mathbf{r}', \omega) V_{\mathrm{ext}}(\mathbf{r}') \tag{41}$$

In angular resolved EELS, only excitations with a specific momentum $\mathbf{q}$ are probed, and thus the relevant loss function is $P(\mathbf{q}, \omega) = -\mathrm{Im}\,\chi_{00}(\mathbf{q}, \omega)$. For spatially resolved EELS, the relevant external potential is that of a point charge approaching the sample with constant velocity, $V_{\mathrm{ext}}(\mathbf{r}, t) = 1/|\mathbf{r} - \mathbf{v}_e t|$. Upon Fourier transforming this potential it can be shown that the loss spectrum of a 2D material for a normal incident electron beam is [150]

$$P_{\mathrm{local}}(\omega) = -\mathrm{Im} \int \mathrm{d}\mathbf{q}_{\parallel} \frac{(4\pi)^2}{(q_{\parallel}^2 + q_z^2)^2}$$
$$\int \mathrm{d}z \int \mathrm{d}z'(1 - iq_z z)\chi(\mathbf{q}_{\parallel}, z, z', \omega)$$
$$(1 + iq_z z')\delta(\omega - q_z v_e) \tag{42}$$

where a mixed $(\mathbf{q}_{\parallel}, z)$-representation of χ has been used. Discussions of the interpretation and calculation of electron energy loss spectroscopy for 2D materials can be found in [150, 151].

5.6. Interband plasmons versus excitons

The similar structure of interband plasmons and excitons has previously been alluded to. Below the difference between these two types of collective excitations is explored in more detail. It turns out that 2D materials present some novelty in this regard compared to the well

known case of bulk materials. The proper theoretical framework for discussing interband excitations is the Bethe–Salpeter equation (BSE) which was briefly introduced in section 4.1. From equations (19)–(21) it follows that the collective nature of the excitations, i.e. the mixing of the e–h pairs, is governed by the xc-kernel consisting of the e–h exchange (V) and direct screened interaction (W). These terms respectively increase and decrease the energy of the collective excitation relative to that of the constituent e–h pairs. Roughly speaking, this means that the e–h exchange interaction is responsible for plasmon formation while the direct e–h interaction is responsible for exciton formation.

Consider first the relevant e–h exchange matrix element in the BSE Hamiltonian (where **k** and **q** are vectors in the 2D Brillouin zone):

$$V_{SS'}(\mathbf{q}) = \langle \psi_S | \hat{V} | \psi_{S'} \rangle \tag{43}$$

$$= \int\int d\mathbf{r} d\mathbf{r}' \frac{\psi_{n\mathbf{k}}(\mathbf{r})\psi_{m\mathbf{k}+\mathbf{q}}(\mathbf{r})^*\psi_{n'\mathbf{k}'}(\mathbf{r}')^*\psi_{m'\mathbf{k}'+\mathbf{q}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{44}$$

It can be shown, e.g. using $k \cdot p$ perturbation theory, that for small momentum transfers

$$\psi_{n\mathbf{k}}(\mathbf{r})\psi_{m\mathbf{k}+\mathbf{q}}(\mathbf{r})^* = \left(qu_0(z) + \sum_{\mathbf{G} \neq \mathbf{0}}^{\mathrm{BZ}} u_{\mathbf{G}}(z)\mathrm{e}^{-i\mathbf{G}\cdot\mathbf{r}_\parallel} \right)\mathrm{e}^{-i\mathbf{q}\cdot\mathbf{r}_\parallel} \tag{45}$$

This equation shows that the long range ($\mathbf{G} = \mathbf{0}$) component of each of the wave functions ψ_S and $\psi_{S'}$ are proportional to q in the long wave length limit. In the same limit the (quasi) 2D Coulomb interaction takes the form $1/q$. Therefore the long range part of the Coulomb interaction does not contribute to $V_{SS'}(\mathbf{q})$ for small **q**. On the other hand, short range Coulomb interactions generally have little effect on the excitations due to the rapid decay of the Coulomb interaction with wave vector, see equation (5). We thus conclude that for a 2D material the e–h exchange, which drives plasmon formation, is very weak in the small **q** limit. Note that this is very different from the 3D case where the long range part of the Coulomb interaction dominates the e–h exchange for small **q**.

In contrast to the e–h exchange operator, the direct e–h interaction is relatively independent of **q**. In particular, the contribution to $W_{SS'}(\mathbf{q})$ from the long range component of the Coulomb interaction does not vanish in the $\mathbf{q} \to \mathbf{0}$ limit but in fact dominates the matrix element. Thus we must expect that the collective excitations of a 2D material are of excitonic nature for small **q** while plasmons become increasingly important for larger **q**. The example given below shows that this is indeed the case.
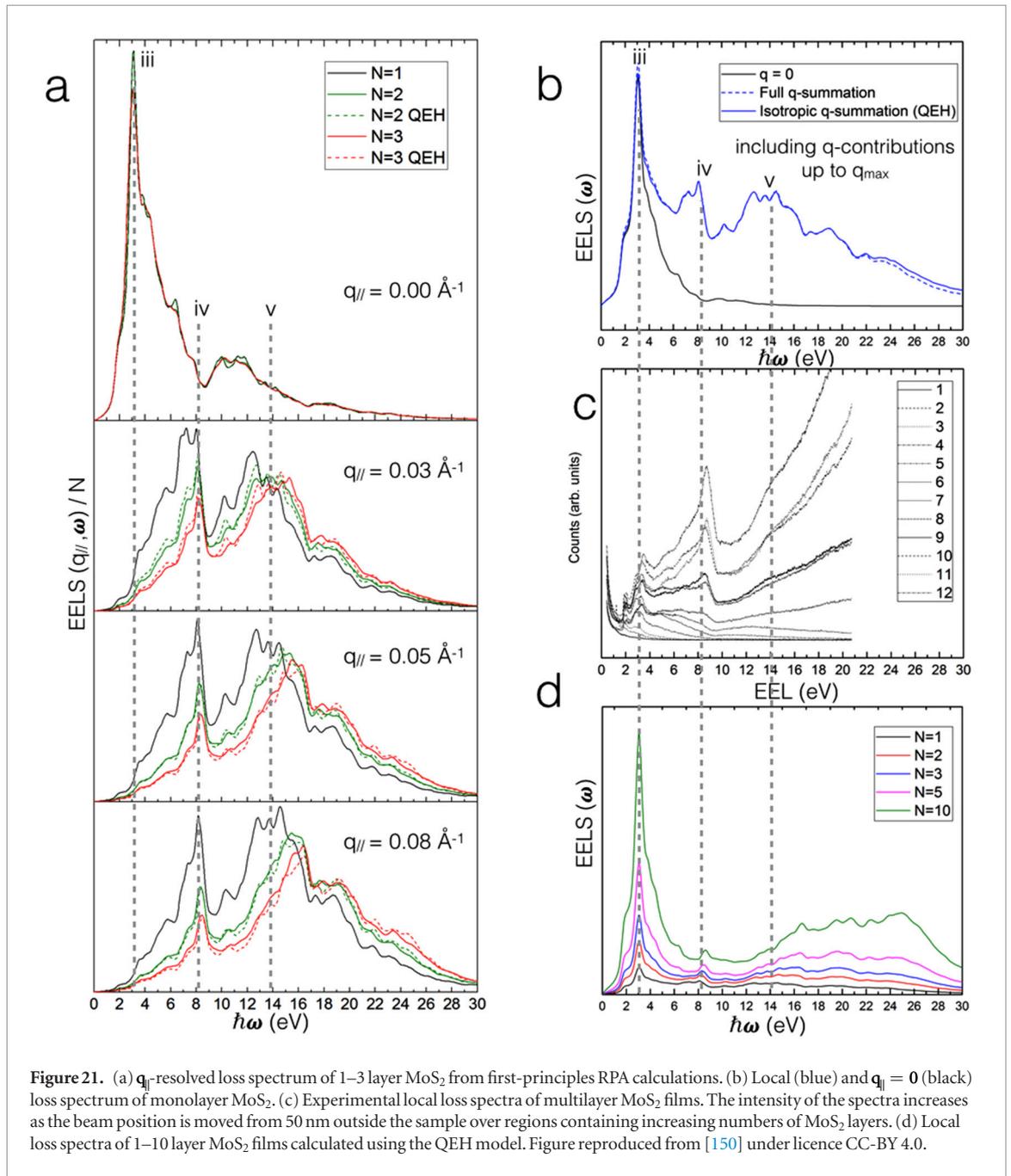
Figure 21(a) shows the loss spectrum of 1–3 layer MoS_2 for four different values of the in-plane momentum transfer **q**. All spectra are normalized by the number of layers. The response function has been calculated

at the RPA level using DFT-PBE wave functions for the multilayer films (full lines) or using the QEH model to include interlayer interactions (dashed lines). The excellent agreement between the full calculations and the QEH model demonstrates that interlayer hybridisation has negligible effect on the loss spectra.

A consistent description of plasmons and excitons should ideally be based on the BSE rather than the RPA. Indeed, the RPA neglects the direct e–h interaction (W) and thus cannot account for excitons. However, the computational cost of the BSE makes it impossible to describe the excitation spectrum accurately up to energies relevant for the plasmons (around 5–20 eV for MoS_2). On the other hand we argue that the low energy part of the RPA spectrum behaves qualitatively similar to the BSE spectrum with respect to the **q**-dependence.

Returning to figure 21(a) it can be seen that for $\mathbf{q} = \mathbf{0}$ the loss spectrum is independent of film thickness. This is a result of the absence of long range screening which leads to an effective decoupling of the layers. In fact, the RPA spectrum consists essentially of uncoupled single-particle transitions because the e–h exchange coupling vanishes (almost) for $\mathbf{q} = \mathbf{0}$. The peak at ~ 3 eV originates from transitions at the Γ-point of the BZ. These are the same transitions that form the C exciton giving rise to the peak at 2.5 eV in the BSE spectrum (see figure 9) [83]. Thus the 3 eV peak in the RPA spectrum represents an uncorrelated version of the C exciton. For finite **q**, plasmon peaks around 8 eV and 14 eV emerge and completely dominate the loss spectrum. The plasmon peaks are sensitive to both **q** and film thickness which again underlines the strongly **q**-dependent screening in 2D (an effect also referred to as spatial dispersion or non-local response). We conclude that excitonic states dominate the loss spectra for small **q** while plasmonic states dominate for larger **q**.

Figure 21(b) shows the local loss spectrum of monolayer MoS_2 (blue) obtained from equation (42). For comparison the $\mathbf{q} = \mathbf{0}$ spectrum is shown. The local spectrum involves a summation over all **q**-vectors and consequently exhibits peaks due to both excitons and plasmons. Physically, the localised electron beam probes the collective excitations at all momenta; it thus represents a sensitive but not very selective probe. Figure 21(c) shows experimentally measured local loss spectra obtained at different positions of an MoS_2 sample of varying thickness. The spectra increase in intensity as the beam is moved from 50 nm outside the sample over regions with increasing number of MoS_2 layers. Figure 21(d) shows the calculated local, i.e. **q**-summed, loss spectra for 1–10 layer MoS_2 calculated using the QEH model. The agreement with experiments is reasonable although the RPA calculations and experiments show the opposite trend regarding the relative intensity of the C-exciton and the 8 eV plasmon peaks. Interestingly, the experimental spectra recorded

Figure 21. (a) $q_\parallel$-resolved loss spectrum of 1–3 layer MoS_2 from first-principles RPA calculations. (b) Local (blue) and $q_\parallel = 0$ (black) loss spectrum of monolayer MoS_2. (c) Experimental local loss spectra of multilayer MoS_2 films. The intensity of the spectra increases as the beam position is moved from 50 nm outside the sample over regions containing increasing numbers of MoS_2 layers. (d) Local loss spectra of 1–10 layer MoS_2 films calculated using the QEH model. Figure reproduced from [150] under licence CC-BY 4.0.

up to 50 nm away from the MoS_2 sample are dominated by the 3 eV excitonic peak rather than the plasmons. A possible explanation is that when the beam is far away from the sample only the long wave length (small q) components of the probing beam potential reaches the MoS_2. Therefore, by varying the position of the electron beam relative to the sample it is possible, to some extent, to selectively excite collective excitations with different q.

6. Conclusions

This topical review has discussed some of the concepts, theories, and computational methods used to describe elementary electronic excitations in atomically thin 2D materials and layered van der Waals heterostructures. It was shown that the weak and highly non-local nature

of the dielectric screening in atomically thin crystals is responsible for a number of unique properties exhibited by the 2D materials. The non-conventional physics of excitons in 2D materials was illustrated and it was shown how the non-local screening can be incorporated into a 2D Hydrogenic model providing analytical expressions for the exciton properties with quantitative accuracy and simple explanations for the non-Hydrogenic Rydberg series, non-degeneracy of excitonic states with different l quantum numbers, and the (almost) linear dependence of the exciton binding energy with band gap. The problem of exciton dissociation in 2D materials by an in-plane electric field was discussed in the context of the complex scaling method. It was shown that the dissociation rate can be greatly enhanced by encapsulating the 2D material in a dielectric which screens the e–h interaction. For

plasmons, the absence of long range screening in 2D leads to acoustic dispersion relations for the metallic intraband plasmons and suppresses the formation of plasmonic states in general for small momentum transfers, **q**. Consequently, it was shown that the loss spectrum of a 2D material is dominated by excitonic states for small **q** and plasmonic states for larger **q**. The weak intrinsic screening in the 2D materials make them highly sensitive to the environment such as substrate or encapsulation. It was demonstrated that dielectric screening from the environment can shift the collective and single-particle excitation energies by up to 1 eV. The fact that standard density functional theory (DFT) methods are unable to account for these screening effects highlights the need to go beyond the independent particle approximation for band structure calculations. The application of the many-body GW method to 2D materials was briefly reviewed, and a method to overcome the slow **k**-point convergence of such calculations was described. The quantum electrostatic heterostructure (QEH) model was presented as a method for computing the dielectric properties of general, incommensurate van der Waals heterostructures with first-principles accuracy and minimal computational cost. The ability of the QEH model to provide an accurate and seamless connection between the screening properties of 2D and layered 3D materials was used to illustrate the transition of the exciton in MoS_2 from the monolayer to the bulk. These calculations demonstrated that the large exciton binding energy in 2D materials is a result of weak dielectric screening rather than quantum confinement.

Acknowledgments

References

[1] Novoselov K S, Geim A K, Morozov S V, Jiang D, Zhang Y, Dubonos S V, Grigorieva I V and Firsov A A 2004 Electric field effect in atomically thin carbon films *science* **306** 666–9

[2] Bhimanapati G R *et al* 2015 Recent advances in two-dimensional materials beyond graphene *ACS Nano* **9** 11509–39

[3] Ferrari A C *et al* 2015 Science and technology roadmap for graphene, related two-dimensional crystals and hybrid systems *Nanoscale* **7** 4598–810

[4] Ci L *et al* 2010 Atomic layers of hybridized boron nitride and graphene domains *Nat. Mater.* **9** 430–5

[5] Nagashima A, Tejima N, Gamou Y, Kawai T and Oshima C 1996 Electronic states of monolayer hexagonal boron nitride formed on the metal surfaces *Surf. Sci.* **357** 307–11

[6] Li L, Yu Y, Ye G J, Ge Q, Ou X, Wu H, Feng D, Chen X H and Zhang Y 2014 Black phosphorus field-effect transistors *Nat. Nanotechnol.* **9** 372–7

[7] Liu H, Neal A T, Zhu Z, Luo Z, Xu X, Tománek D and Ye P D 2014 Phosphorene: an unexplored 2d semiconductor with a high hole mobility *ACS Nano* **8** 4033–41

[8] Wang Q H, Kalantar-Zadeh K, Kis A, Coleman J N and Strano M S 2012 Electronics and optoelectronics of two-dimensional transition metal dichalcogenides *Nat. Nanotechnol.* **7** 699–712

[9] Lalmi B, Oughaddou H, Enriquez H, Kara A, Vizzini S, Ealet B and Aufray B 2010 Epitaxial growth of a silicene sheet *Appl. Phys. Lett.* **97** 223109

[10] Aufray B, Kara A, Vizzini S, Oughaddou H, Leandri C, Ealet B and Le Lay G 2010 Graphene-like silicon nanoribbons on ag (1 1 0): a possible formation of silicene *Appl. Phys. Lett.* **96** 183102

[11] Tao L, Cinquanta E, Chiappe D, Grazianetti C, Fanciulli M, Dubey M, Molle A and Akinwande D 2015 Silicene field-effect transistors operating at room temperature *Nat. Nanotechnol.* **10** 227–31

[12] Dávila M E, Xian L, Cahangirov S, Rubio A and Le Lay G 2014 Germanene: a novel two-dimensional germanium allotrope akin to graphene and silicene *New J. Phys.* **16** 095002

[13] Zhu F-F, Chen W-J, Xu Y, Gao C-L, Guan D-D, Liu C-H, Qian D, Zhang S-C and Jia J-F 2015 Epitaxial growth of two-dimensional stanene *Nat. Mater.* **14** 1020–5

[14] Mannix A J *et al* 2015 Synthesis of borophenes: anisotropic, two-dimensional boron polymorphs *Science* **350** 1513–6

[15] Sahin H, Cahangirov S, Topsakal M, Bekaroglu E, Akturk E, Tugrul Senger R and Ciraci S 2009 Monolayer honeycomb structures of group-iv elements and Iii-V binary compounds: First-principles calculations *Phys. Rev.* B **80** 155453

[16] Al Balushi Z Y *et al* 2016 Two-dimensional gallium nitride realized via graphene encapsulation *Nat. Mater.* **15** 1166–71

[17] Naguib M, Mashtalir O, Carle J, Presser V, Lu J, Hultman L, Gogotsi Y and Barsoum M W 2012 Two-dimensional transition metal carbides *ACS Nano* **6** 1322–31

[18] Sun Z, Liao T, Dou Y, Hwang S M, Park M-S, Jiang L, Kim J H and Dou S X 2014 Generalized self-assembly of scalable two-dimensional transition metal oxide nanosheets *Nat. Commun.* **5** 3813

[19] Lin S-H and Kuo J-L 2014 Towards the ionic limit of two-dimensional materials: monolayer alkaline earth and transition metal halides *Phys. Chem. Chem. Phys.* **16** 20763–71

[20] Dou L *et al* 2015 Atomically thin two-dimensional organic-inorganic hybrid perovskites *Science* **349** 1518–21

[21] Pandey M, Jacobsen K W and Thygesen K S 2016 Band gap tuning and defect tolerance of atomically thin two-dimensional organic–inorganic halide perovskites *J. Phys. Chem. Lett.* **7** 4346–52

[22] Keldysh L V 1978 *JETP Lett.* **29** 658

[23] Cudazzo P, Tokatly I V and Rubio A 2011 Dielectric screening in two-dimensional insulators: Implications for excitonic and impurity states in graphane *Phys. Rev.* B **84** 085406

[24] Mak K F, He K, Lee C, Lee G H, Hone J, Heinz T F and Shan J 2013 Tightly bound trions in monolayer mos2 *Nat. Mater.* **12** 207–11

[25] Splendiani A, Sun L, Zhang Y, Li T, Kim J, Chim C Y, Galli G and Wang F 2010 Emerging photoluminescence in monolayer MoS2 *Nano Lett.* **10** 1271–5

[26] Liu Y, Willis R F, Emtsev K V and Seyller T 2008 Plasmon dispersion and damping in electrically isolated two-dimensional charge sheets *Phys. Rev.* B **78** 201403

[27] Andersen K and Thygesen K S 2013 Plasmons in metallic monolayer and bilayer transition metal dichalcogenides *Phys. Rev.* B **88** 155128

[28] Koppens F H L, Chang D E and Javier Garcia de Abajo F 2011 Graphene plasmonics: a platform for strong light-matter interactions *Nano Lett.* **11** 3370–7

[29] Ugeda M M *et al* 2014 Giant bandgap renormalization and excitonic effects in a monolayer transition metal dichalcogenide semiconductor *Nat. Mater.* **13** 1091–5

[30] Falco Hüser, Olsen T and Thygesen K S 2013 Quasiparticle gw calculations for solids, molecules and two-dimensional materials *Phys. Rev.* B **87** 235132

[31] Geim A K and Grigorieva I V 2013 Van der waals heterostructures *Nature* **499** 419–25

[32] Cui X *et al* 2015 Multi-terminal transport measurements of mos2 using a van der waals heterostructure device platform *Nat. Nanotechnol.* **10** 534–40

[33] Lu C-P, Li G, Watanabe K, Taniguchi T and Andrei E Y 2014 Mos 2: choice substrate for accessing and tuning the electronic properties of graphene *Phys. Rev. Lett.* **113** 156804

[34] Withers F *et al* 2015 Light-emitting diodes by band-structure engineering in van der waals heterostructures *Nat. Mater.* **14** 301–6

[35] Mounet N *et al* 2016 Novel two-dimensional materials from high-throughput computational exfoliation of experimentally known compounds (arXiv: 1611.05234)

[36] Tritsaris G A, Shirodkar S N, Kaxiras E, Cazeaux P, Luskin M, Plecháč P and Cancès E 2016 Perturbation theory for weakly coupled two-dimensional layers *J. Mater. Res.* **31** 959–66

[37] Andersen K, Latini S and Thygesen K S 2015 Dielectric genome of van der waals heterostructures *Nano Lett.* **15** 4616–21

[38] Lin Z, Carvalho B R, Kahn E, Lv R, Rao R, Terrones H, Pimenta M A and Terrones M 2016 Defect engineering of two-dimensional transition metal dichalcogenides *2D Mater.* **3** 022002

[39] Komsa H-P, Kurasch S, Lehtinen O, Kaiser U and Krasheninnikov A V 2013 From point to extended defects in two-dimensional mos 2: evolution of atomic structure under electron irradiation *Phys. Rev.* B **88** 035301

[40] Amani M *et al* 2015 Near-unity photoluminescence quantum yield in mos2 *Science* **350** 1065–8

[41] Hüser F, Olsen T and Thygesen K S 2013 How dielectric screening in two-dimensional crystals affects the convergence of excited-state calculations: monolayer mos 2 *Phys. Rev.* B **88** 245309

[42] Cudazzo P, Attaccalite C, Tokatly I V and Rubio A 2010 Strong charge-transfer excitonic effects and the bose-einstein exciton condensate in graphane *Phys. Rev. Lett.* **104** 226804

[43] Pulci O, Gori P, Marsili M, Garbuio V, Del Sole R and Bechstedt F 2012 Strong excitons in novel two-dimensional crystals: silicane and germanane *Europhys. Lett.* **98** 37004

[44] Berkelbach T C, Hybertsen M S and Reichman D R 2013 Theory of neutral and charged excitons in monolayer transition metal dichalcogenides *Phys. Rev.* B **88** 045318

[45] Latini S, Olsen T and Thygesen K S 2015 Excitons in van der waals heterostructures: The important role of dielectric screening *Phys. Rev.* B **92** 245123

[46] Onida G, Reining L and Rubio A 2002 Electronic excitations: density-functional versus many-body green's-function approaches *Rev. Mod. Phys.* **74** 601–59

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

[48] Ramasubramaniam A 2012 Large excitonic effects in monolayers of molybdenum and tungsten dichalcogenides *Phys. Rev.* B **86** 115409

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

[50] Rasmussen F A and Thygesen K S 2015 Computational 2d materials database: electronic structure of transition-metal dichalcogenides and oxides *J. Phys. Chem.* C **119** 13169–83

[51] Heyd J and Scuseria G E 2004 Efficient hybrid density functional calculations in solids: assessment of the Heyd–Scuseria–Ernzerhof screened coulomb hybrid functional *J. Chem. Phys.* **121** 1187–92

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

[53] Hedin L 1965 New method for calculating the one-particle Green's function with application to the electron-gas problem *Phys. Rev.* **139** A796–823

[54] Hybertsen M S and Louie S G 1985 First-principles theory of quasiparticles: calculation of band gaps in semiconductors and insulators *Phys. Rev. Lett.* **55** 1418–21

[55] Godby R W, Schlüter M and Sham L J 1986 Accurate exchange-correlation potential for silicon and its discontinuity on addition of an electron *Phys. Rev. Lett.* **56** 2415–8

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

[57] Kotani T and van Schilfgaarde M 2002 All-electron GW approximation with the mixed basis expansion based on the full-potential LMTO method *Solid State Commun.* **121** 461–5

[58] Marini A, Hogan C, Grüning M and Varsano D 2009 Yambo: an *ab initio* tool for excited state calculations *Comput. Phys. Commun.* **180** 1392–403

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

[60] Blase X, Attaccalite C and Olevano V 2011 First-principles {GW} calculations for fullerenes, porphyrins, phtalocyanine and other molecules of interest for organic photovoltaic applications *Phys. Rev.* B **83** 115103

[61] Ismail-Beigi S 2006 Truncation of periodic image interactions for confined systems *Phys. Rev.* B **73** 233103

[62] Qiu D Y, da Jornada F H and Louie S G 2016 Screening and many-body effects in two-dimensional crystals: monolayer mos2 (arXiv:1605.08733)

[63] Rasmussen F A, Schmidt P S, Winther K T and Thygesen K S 2016 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 **94** 155406

[64] Zhu Z Y, Cheng Y C and Schwingenschlögl U 2011 Giant spin–orbit-induced spin splitting in two-dimensional transition-metal dichalcogenide semiconductors *Phys. Rev.* B **84** 153402

[65] Komsa H-P and Krasheninnikov A V 2012 Effects of confinement and environment on the electronic structure and exciton binding energy of mos_2 from first principles *Phys. Rev.* B **86** 241201

[66] Shi H, Pan H, Zhang Y-W and Yakobson B I 2013 Quasiparticle band structures and optical properties of strained monolayer mos_2 and ws_2 *Phys. Rev.* B **87** 155304

[67] Conley H J, Wang B, Ziegler J I, Haglund R F, Pantelides S T and Bolotin K I 2013 Bandgap engineering of strained monolayer and bilayer MoS2 *Nano Lett.* **13** 3626–30

[68] Martin R M 2004 *Electronic Structure: Basic Theory and Practical Methods* (Cambridge: Cambridge University Press)

[69] Xiao D, Liu G-B, Feng W, Xu X and Yao W 2012 Coupled spin and valley physics in monolayers of mos 2 and other group-vi dichalcogenides *Phys. Rev. Lett.* **108** 196802

[70] Mak K F, He K, Shan J and Heinz T F 2012 Control of valley polarization in monolayer mos2 by optical helicity *Nat. Nanotechnol.* **7** 494–8

[71] Jones A M *et al* 2013 Optical generation of excitonic valley coherence in monolayer wse2 *Nat. Nanotechnol.* **8** 634–8

[72] Schaibley J R, Yu H, Clark G, Rivera P, Ross J S, Seyler K L, Yao W and Xu X 2016 Valleytronics in 2d materials *Nat. Rev. Mater.* **1** 16055

[73] Olsen T 2016 Designing in-plane heterostructures of quantum spin hall insulators from first principles: 1 t- mos 2 with adsorbates *Phys. Rev.* B **94** 235106

[74] Rosner M, Steinke C, Lorke M, Gies C, Jahnke F and Wehling T O 2016 Two-dimensional heterojunctions from nonlocal manipulations of the interactions *Nano Lett.* **16** 2322–7

[75] Winther K T and Thygesen K S 2017 Band structure engineering in van der waals heterostructures via dielectric screening arXiv: 1703.03188

[76] Neaton J B, Hybertsen M S and Louie S G 2006 Renormalization of molecular electronic levels at metal-molecule interfaces *Phys. Rev. Lett.* **97** 216405

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

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

[79] Perrin M L, Verzijl C J O, Martin C A, Shaikh A J, Eelkema R, van Esch J H, van Ruitenbeek J M, Thijssen J M, van der Zant H S J and Dulić D 2013 Large tunable image-charge effects in single-molecule junctions *Nat. Nanotechnol.* **8** 282–7

[80] Jin C, Rasmussen F A and Thygesen K S 2015 Tuning the schottky barrier at the graphene/mos2 interface by electron doping: density functional theory and many-body calculations *J. Phys. Chem.* C **119** 19928–33

[81] Wirtz L, Marini A and Rubio A 2006 Excitons in boron nitride nanotubes: dimensionality effects *Phys. Rev. Lett.* **96** 126104

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

[83] Qiu D Y, da Jornada F H and Louie S G 2013 Optical spectrum of MoS2: many-body effects and diversity of exciton states *Phys. Rev. Lett.* **111** 216805

[84] Ye Z, Cao T, O'Brien K, Zhu H, Yin X, Wang Y, Louie S G and Zhang X 2014 Probing excitonic dark states in single-layer tungsten disulphide *Nature* **513** 214–8

[85] Klots A R *et al* 2014 Probing excitonic states in suspended two-dimensional semiconductors by photocurrent spectroscopy *Sci. Rep.* **4** 6608

[86] Hanbicki A T, Currie M, Kioseoglou G, Friedman A L and Jonker B T 2015 Measurement of high exciton binding energy in the monolayer transition-metal dichalcogenides WS2 and WSe2 *Solid State Commun.* **203** 16–20

[87] Yang X L, Guo S H, Chan F T, Wong K W and Ching W Y 1991 Analytic solution of a two-dimensional hydrogen atom. I. Nonrelativistic theory *Phys. Rev.* A **43** 1186–96

[88] Strinati G 1984 Effects of dynamical screening on resonances at inner-shell thresholds in semiconductors *Phys. Rev.* B **29** 5718–26

[89] Latini S 2016 Excitons in van der waals heterostructure—a theoretical study *PhD Thesis* (Technical University of Denmark)

[90] Olsen T, Latini S, Rasmussen F and Thygesen K S 2016 Simple screened hydrogen model of excitons in two-dimensional materials *Phys. Rev. Lett.* **116** 056401

[91] Choi J-H, Cui P, Lan H and Zhang Z 2015 Linear scaling of the exciton binding energy versus the band gap of two-dimensional materials *Phys. Rev. Lett.* **115** 066403

[92] Huang S, Liang Y and Yang L 2013 Exciton spectra in two-dimensional graphene derivatives *Phys. Rev.* B **88** 075441

[93] 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 Exciton binding energy and nonhydrogenic rydberg series in monolayer ws 2 *Phys. Rev. Lett.* **113** 076802

[94] Srivastava A and Imamoğlu A 2015 Signatures of bloch-band geometry on excitons: nonhydrogenic spectra in transition-metal dichalcogenides *Phys. Rev. Lett.* **115** 166802

[95] Zhou J, Shan W-Y, Yao W and Xiao D 2015 Berry phase modification to the energy spectrum of excitons *Phys. Rev. Lett.* **115** 166803

[96] Weissker H-C, Serrano J, Huotari S, Luppi E, Cazzaniga M, Bruneval F, Sottile F, Monaco G, Olevano V and Reining L 2010 Dynamic structure factor and dielectric function of silicon for finite momentum transfer: inelastic x-ray scattering experiments and ab initio calculations *Phys. Rev.* B **81** 085104

[97] Sharma S, Dewhurst J K, Sanna A, Rubio A and Gross E K U 2012 Enhanced excitonic effects in the energy loss spectra of lif and ar at large momentum transfer *New J. Phys.* **14** 053052

[98] Qiu D Y, Cao T and Louie S G 2015 Nonanalyticity, valley quantum phases and lightlike exciton dispersion in monolayer transition metal dichalcogenides: theory and first-principles calculations *Phys. Rev. Lett.* **115** 176801

[99] Cudazzo P, Sponza L, Giorgetti C, Reining L, Sottile F and Gatti M 2016 Exciton band structure in two-dimensional materials *Phys. Rev. Lett.* **116** 066803

[100] Wu F, Qu F and MacDonald A H 2015 Exciton band structure of monolayer mos2 *Phys. Rev.* B **91** 075310

[101] Ross J S *et al* 2013 Electrical control of neutral and charged excitons in a monolayer semiconductor *Nat. Commun.* **4** 1474

[102] Xu R *et al* 2016 Extraordinarily bound quasi-one-dimensional trions in two-dimensional phosphorene atomic semiconductors *ACS Nano* **10** 2046–53

[103] Huard V, Cox R T, Saminadayar K, Arnoult A and Tatarenko S 2000 Bound states in optical absorption of semiconductor quantum wells containing a two-dimensional electron gas *Phys. Rev. Lett.* **84** 187

[104] Kylänpää I and Komsa H-P 2015 Binding energies of exciton complexes in transition metal dichalcogenide monolayers and effect of dielectric environment *Phys. Rev.* B **92** 205418

[105] Zhang D K, Kidd D W and Varga K 2015 Excited biexcitons in transition metal dichalcogenides *Nano Lett.* **15** 7002–5

[106] You Y, Zhang X-X, Berkelbach T C, Hybertsen M S, Reichman D R and Heinz T F 2015 Observation of biexcitons in monolayer wse2 *Nat. Phys.* **11** 477–81

[107] Mai C, Barrette A, Yu Y, Semenov Y G, Kim K W, Cao L and Gundogdu K 2013 Many-body effects in valleytronics: direct measurement of valley lifetimes in single-layer mos2 *Nano Lett.* **14** 202–6

[108] Deilmann T, Drüppel M and Rohlfing M 2016 Three-particle correlation from a many-body perspective: trions in a carbon nanotube *Phys. Rev. Lett.* **116** 196804

[109] Peelaers H and Van de Walle C G 2012 Effects of strain on band structure and effective masses in MoS_2 *Phys. Rev.* B **86** 241401

[110] Fang H, Battaglia C, Carraro C, Nemsak S, Ozdol B, Kang J S, Bechtel H A, Desai S B, Kronast F and Unal A A 2014 Strong interlayer coupling in van der waals heterostructures built from single-layer chalcogenides *Proc. Natl. Acad. Sci.* **111** 6198–202

[111] Latini S, Winther K, Olsen T and Thygesen K S 2017 Interlayer excitons and band alignment in mos2/hbn/wse2 van der waals heterostructures *Nano Lett.* **17** 938–45

[112] Rivera P, Schaibley J R, Jones A M, Ross J S, Wu S, Aivazian G, Klement P, Seyler K, Clark G and Ghimire N J 2015 Observation of long-lived interlayer excitons in monolayer mose2–wse2 heterostructures *Nat. Commun.* **6** 6242

[113] Palummo M, Bernardi M and Grossman J C 2015 Exciton radiative lifetimes in two-dimensional transition metal dichalcogenides *Nano Lett.* **15** 2794–800

[114] Fogler M M, Butov L V and Novoselov K S 2014 High-temperature superfluidity with indirect excitons in van der waals heterostructures *Nat. Commun.* **5** 4555

[115] Lee C-H, Lee G-H, Der Zande V A M, Chen W, Li Y, Han M, Cui X, Arefe G, Nuckolls C and Heinz T F 2014 Atomically thin p–n junctions with van der waals heterointerfaces *Nat. Nanotechnol.* **9** 676–81

[116] Skinner B 2016 Interlayer excitons with tunable dispersion relation *Phys. Rev.* B **93** 235110

[117] Haastrup S, Latini S, Bolotin K and Thygesen K S 2016 Stark shift and electric-field-induced dissociation of excitons in monolayer mos2 and hbn/mos2 heterostructures *Phys. Rev.* B **94** 041401

[118] Keldysh L V 1957 Behaviour of non-metallic crystals in strong electric fields *J. Exp. Theor. Phys.* **33** 994–1003

[119] Franz W 1958 Einflu eines elektrischen feldes auf eine optische absorptionskante *13a* 484–9

[120] Taylor J R 1972 *Scattering Theory: the Quantum Theory on Nonrelativistic Collisions* (New York: Wiley)

[121] Balslev E and Combes J M 1971 Spectral properties of many-body schrodinger operators with dilatation- analytic interactions *Commun. Math. Phys.* **22** 280–94

[122] Palummo M, Bernardi M and Grossman J C 2015 Exciton radiative lifetimes in two-dimensional transition metal dichalcogenides *Nano Lett.* **15** 2794

[123] Sun D, Rao Y, Reider G A, Chen G, You Y, Brézin L, Harutyunyan A R and Heinz T F 2014 Observation of rapid exciton–exciton annihilation in monolayer molybdenum disulfide *Nano Lett.* **14** 5625–9

[124] Wang H, Zhang C, Chan W, Manolatou C, Tiwari S and Rana F 2016 Radiative lifetimes of excitons and trions in monolayers of the metal dichalcogenide mos2 *Phys. Rev.* B **93** 045407

[125] Lagarde D, Bouet L, Marie X, Zhu C R, Liu B L, Amand T, Tan P H and Urbaszek B 2014 Carrier and polarization dynamics in monolayer mos 2 *Phys. Rev. Lett.* **112** 047401

[126] Pedersen T G, Latini S, Thygesen K S, Mera H and Nikolić B K 2016 Exciton ionization in multilayer transition-metal dichalcogenides *New J. Phys.* **18** 073043

[127] Pedersen T G 2016 Exciton stark shift and electroabsorption in monolayer transition-metal dichalcogenides *Phys. Rev.* B **94** 125424

[128] Bohm D and Pines D 1953 A collective description of electron interactions: III. Coulomb interactions in a degenerate electron gas *Phys. Rev.* **92** 609

[129] Zijlstra P, Paulo P M R and Orrit M 2012 Optical detection of single non-absorbing molecules using the surface plasmon resonance of a gold nanorod *Nat. Nanotechnol.* **7** 379–82

[130] Liebermann T and Knoll W 2000 Surface-plasmon field-enhanced fluorescence spectroscopy *Colloids Surf.* A **171** 115–30

[131] Linic S, Christopher P and Ingram D B 2011 Plasmonic-metal nanostructures for efficient conversion of solar to chemical energy *Nat. Mater.* **10** 911–21

[132] Atwater H A and Polman A 2010 Plasmonics for improved photovoltaic devices *Nat. Mater.* **9** 205–13

[133] Liu Z, Durant S, Lee H, Pikus Y, Fang N, Xiong Y, Sun C and Zhang X 2007 Far-field optical superlens *Nano Lett.* **7** 403–8

[134] Kawata S, Ono A and Verma P 2008 Subwavelength colour imaging with a metallic nanolens *Nat. Photon.* **2** 438–42

[135] Ju L *et al* 2011 Graphene plasmonics for tunable terahertz metamaterials *Nat. Nanotechnol.* **6** 630–4

[136] Woessner A *et al* 2015 Highly confined low-loss plasmons in graphene–boron nitride heterostructures *Nat. Mater.* **14** 421–5

[137] Teperik T V, Nordlander P, Aizpurua J and Borisov A G 2013 Robust subnanometric plasmon ruler by rescaling of the nonlocal optical response *Phys. Rev. Lett.* **110** 263901

[138] Javier Garci F 2008 Nonlocal effects in the plasmons of strongly interacting nanoparticles, dimers and waveguides *J. Phys. Chem.* C **112** 17983–7

[139] Raza S, Toscano G, Jauho A-P, Wubs M and Asger Mortensen N 2011 Unusual resonances in nanoplasmonic structures due to nonlocal response *Phys. Rev.* B **84** 121412

[140] Toscano G, Straubel J, Kwiatkowski A, Rockstuhl C, Evers F, Xu H, Asger Mortensen N and Wubs M 2015 Resonance shifts and spill-out effects in self-consistent hydrodynamic nanoplasmonics *Nat. Commun.* **6** 7312

[141] Liebsch A 1993 Surface plasmon dispersion of ag *Phys. Rev. Lett.* **71** 145

[142] Pitarke J M, Silkin V M, Chulkov E V and Echenique P M 2006 Theory of surface plasmons and surface-plasmon polaritons *Rep. Prog. Phys.* **70** 1

[143] Hwang E H and Das Sarma S 2007 Dielectric function, screening and plasmons in two-dimensional graphene *Phys. Rev.* B **75** 205418

[144] Lu J, Loh K P, Huang H, Chen W and Wee A T S 2009 Plasmon dispersion on epitaxial graphene studied using high-resolution electron energy-loss spectroscopy *Phys. Rev.* B **80** 113410

[145] Andersen K, Jacobsen K W and Thygesen K S 2014 Plasmons on the edge of mos2 nanostructures *Phys. Rev.* B **90** 161410

[146] Gibertini M and Marzari N 2015 Emergence of one-dimensional wires of free carriers in transition-metal-dichalcogenide nanostructures *Nano Lett.* **15** 6229–38

[147] Andersen K, Jacobsen K W and Thygesen K S 2012 Spatially resolved quantum plasmon modes in metallic nano-films from first-principles *Phys. Rev.* B **86** 245129

[148] Hwang E H and Das Sarma S 2009 Plasmon modes of spatially separated double-layer graphene *Phys. Rev.* B **80** 205405

[149] Krivanek O L, Lovejoy T C, Murfitt M F, Skone G, Batson P E and Dellby N 2014 Towards sub-10 mev energy resolution stem-eels *J. Phys.: Conf. Ser.* **522** 012023

[150] Nerl H, Winther K T, Hage F S, Thygesen K S, Houben L, Backes C, Coleman J N, Ramasse Q M and Nicolosi V 2017 Probing the local nature of excitons and plasmons in few-layer mos2 *2D Mater. Appl.* **1** 2

[151] Nazarov V U 2015 Electronic excitations in quasi-2d crystals: what theoretical quantities are relevant to experiment? *New J. Phys.* **17** 073018