

Topical Review

Plasmonics in Dirac systems: from graphene to topological insulators

Tobias Stauber

Departamento de Física de la Materia Condensada, Instituto Nicolás Cabrera and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, E-28049 Madrid, Spain

E-mail: tobias.stauber@uam.es

Received 14 October 2013, revised 16 December 2013

Accepted for publication 14 January 2014

Published 6 March 2014

Abstract

Recent developments in the emerging field of plasmonics in graphene and other Dirac systems are reviewed and a comprehensive introduction to the standard models and techniques is given. In particular, we discuss intrinsic plasmon excitation of single- and bilayer graphene via hydrodynamic equations and the random phase approximation, but also comment on double and multilayer structures. Additionally, we address Dirac systems in the retardation limit and also with large spin–orbit coupling including topological insulators. Finally, we summarize basic properties of the charge, current and photon linear response functions in an appendix.

Keywords: optical properties of graphene, plasmonics, linear response

(Some figures may appear in colour only in the online journal)

Contents

1. Introduction	2	5.2. Damped plasmons and energy loss function	11
2. Hydrodynamic models	3	5.3. Double layer	12
2.1. Drude model	3	6. Plasmons including retardation effects	13
2.2. Semiclassical Boltzmann equation	4	6.1. Longitudinal or TM plasmons	13
3. Plasmons in single-layer graphene	4	6.2. Transverse or TE plasmons	14
3.1. Gapless Dirac fermions	5	7. Plasmons in Dirac systems with strong spin–orbit coupling	15
3.2. Gapped Dirac fermions	6	7.1. Graphene plasmons with spin–orbit coupling	15
3.3. Interband plasmons and EELS	7	7.2. Plasmons in MoS ₂	16
3.4. Magneto-plasmons and strain	7	7.3. Plasmons in 3D topological insulators	16
3.5. Dissipative effects	8	8. Summary and outlook	18
3.6. Beyond RPA	8	Acknowledgments	18
3.7. Plasmons in patterned graphene	8	Appendix. Linear response	18
4. Plasmons in bilayer graphene	9	A.1. Density response	18
4.1. Minimal stacked bilayer graphene	9	A.2. Current response	19
4.2. Twisted bilayer graphene	9	A.3. Photon propagator	20
5. Plasmons in general layered structures	10	A.4. Graphene–light coupling	21
5.1. Undamped plasmons	10	References	21

1. Introduction

The outstanding optical properties of two-dimensional (2D) carbon sheets were the key to the discovery of exfoliated graphene in 2004 [1–4], and its optoelectronic properties are arguably the most promising ones for applications [5–7]. In particular, the large intrinsic carrier mobilities and doping tunability have led to a number of proposals in which the engineering of long-lived graphene plasmons could play a major role [7–10].

Plasmon excitations are intrinsic collective charge or current oscillations coupled via the Coulomb interaction which constitutes the restoring force. Obviously, the group velocity of these oscillations cannot exceed the velocity of light. This means that the plasmon dispersion lies outside the light cone in the near-field (evanescent) regime. The group velocity only merges with the velocity of light for low energies $\hbar\omega \lesssim \alpha E_F$ due to retardation effects where

$$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137} \quad (1)$$

denotes the fine-structure constant with ϵ_0 the vacuum permittivity and $E_F = \hbar v_F k_F$ the Fermi energy of the doped graphene layer. At THz frequencies, the confined collective oscillations of electrons thus enable the manipulation of electromagnetic energy at sub-wavelength scales which is usually coined as plasmonics.

The field of plasmonics in noble metals has attracted a great deal of attention for the past 15 years [11–16], whereby the collective charge excitations on the metallic surface localize the electromagnetic field on sub-wavelength dimensions and can act as a strong dipole or antenna. Additionally, the field enhancement can become very large such that single molecules are detected by Raman scattering [17]. This has opened up the possibility to efficiently couple light to electrons, thus merging photonics and electronics at nanoscale dimensions [18], and has already given rise to a number of metamaterials [19–22].

Plasmons in graphene provide a suitable alternative to noble-metal surface plasmon polaritons because of atomistic confinement of the electrons and the accompanying electromagnetic fields, relatively long propagation lengths compared to the plasmon wavelengths and, most importantly, tunability [7–9]. Here, we review the recent advances in the emerging field of graphene plasmonics and give a comprehensive introduction to the basic theoretical models and techniques.

Plasmons cannot directly couple to propagating electromagnetic radiation because the conservation of momentum is not satisfied in the photon absorption process. To couple light to surface plasmon polaritons on noble metals, the Otto [23] or Kretschmann [24] configuration is used, i.e., the velocity of light of the incoming light c' is reduced by a factor of 2–10 due to an optically dense medium. The slope of the light cone is thus smaller ($c' < c$) and plasmons can be excited by incoming light under an appropriate incident angle, see figure 1(a).

For graphene on a typical dielectric substrate with relative dielectric constant $\epsilon \approx 3$ and covered by air, the plasmon

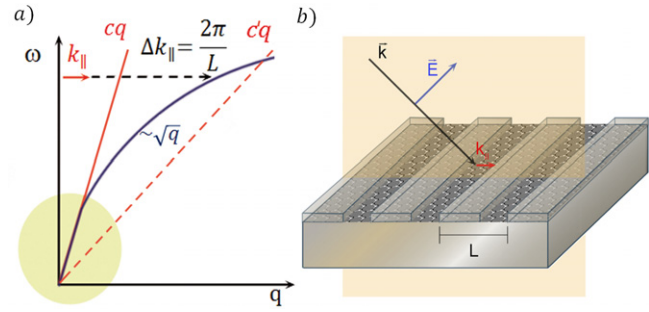


Figure 1. (a) Schematic 2D plasmon dispersion (blue curve) together with the light cone in vacuum (red solid line) and in an optically denser medium (red dashed line). The retardation regime is indicated by the circled region where strong light–matter interaction sets in. (b) Excitation of plasmonic modes by light is possible via an artificial sub-wavelength periodicity providing the missing momentum, Δk_{\parallel} , of the incident electromagnetic radiation with parallel momentum k_{\parallel} .

dispersion is shifted to larger q -vectors, approximated by the compact formula

$$\frac{q}{q_p} \approx \alpha \frac{E_F}{\hbar\omega} \quad (2)$$

where $\omega = cq$ is the energy dispersion of the vacuum light cone, q_p is the wavenumber of the plasmon and E_F the Fermi energy of the doped graphene layer. For typical Fermi energies of $E_F = 0.3$ eV, there is thus a strong reduction of the wavelength even in the THz regime ($\hbar\omega \gtrsim 4$ meV), and the Otto or Kretschmann configuration can usually not be used to detect graphene's intrinsic plasmonic excitations. Graphene plasmons were, therefore, first investigated by means of electron energy-loss spectroscopy (EELS) where the electronic beam was carrying the necessary momentum [25, 26].

One possibility to couple propagating light to graphene's density oscillations is to break the translational symmetry. The necessary (missing) momentum is then provided by a patterned 2D surface with a periodic sub-wavelength structure, see figure 1(b). For graphene, this was first achieved with a grated geometry revealing plasmon resonances with remarkably large oscillator strengths even at room temperature [27]. To this end, the absorption spectrum was obtained for two different polarizations: for the electric field parallel to the grating the usual Drude peak was seen whereas a (plasmon) resonance showed up at finite q in the case of perpendicular polarization [27, 28]. Alternatively, a graphene disk array can be used to excite plasmon resonances by incoming light [29].

Graphene plasmons can also be excited without periodic sub-wavelength patterning by converting the far-field modes into near-field modes via dipole scattering. By this, graphene plasmons were recently launched and detected directly by evanescent waves produced by illuminating an atomic force microscope (AFM) tip with propagating light in the infrared regime [30–32]. By this technique, the propagation of plasmons can also be studied and a new near-field scattering microscopy is developing which is likely to be superior to AFM scanning techniques for large sample areas.

Up to now, the experimental techniques can only access the low wavenumber regime $q \sim 10^4$ – 10^5 cm^{−1} for which

the standard theoretical tools such as hydrodynamic models or the random phase approximation (RPA) are well justified (assuming usual doping levels with $q/k_F \ll 1$). In this topical review, we want to summarize these basic theoretical concepts needed to describe the light–graphene interaction for various physical environments and conditions. We further intend to illustrate the same ideas from different perspectives and will thus derive the basic formulas using various approaches.

In section 2, we present the fundamental formulas concerning the 2D plasmon dispersion using phenomenological descriptions. In sections 3 and 4, we review various aspects of intrinsic plasmon excitations of single-layer and bilayer graphene, mostly based on the RPA. In section 5, we derive the plasmonic spectrum of electrostatically coupled graphene layers and in section 6 we discuss the effect of retardation on longitudinal (or transverse magnetic, TM) as well as on transverse (or transverse electric, TE) plasmons. We close with an account on plasmon excitations in Dirac systems with large spin–orbit coupling. In an appendix, we summarize basic concepts and results of linear response theory for Dirac fermions based on the hexagonal tight-binding model as well as for the electromagnetic gauge field.

2. Hydrodynamic models

Plasmons are collective density oscillations present in almost all electronic systems. They are straightforwardly obtained via a hydrodynamic description based on the continuity equation and linear response

$$-i\omega\rho = -\nabla \cdot \mathbf{j} = \chi_{jj}^+ \nabla \cdot \mathbf{A} = -\frac{\chi_{jj}^+}{i\omega} \nabla^2 \phi, \quad (3)$$

where we have introduced the (local) longitudinal current response function, $\chi_{jj}^+ = \chi_{jj}^+(\omega)$, defined in the appendix, equation (69), and used $\mathbf{E} = i\omega\mathbf{A} = -\nabla\phi$ with \mathbf{E} the electric field and \mathbf{A} , ϕ the vector and electrostatic potential, respectively. With the Fourier transform, $\phi(\mathbf{r}) = \phi_{\mathbf{q}} e^{i\mathbf{q} \cdot \mathbf{r}}$, and the Poisson equation which relates the potential and the charge density via $\phi_{\mathbf{q}} = v_{\mathbf{q}} \rho_{\mathbf{q}}$, we can write the above equation as

$$(\omega^2 - \chi_{jj}^+ v_{\mathbf{q}} q^2) \rho_{\mathbf{q}} = 0. \quad (4)$$

Collective density oscillations are thus defined by the dispersion relation $\omega_p^2 = \chi_{jj}^+ v_{\mathbf{q}} q^2$ which holds for all dimensions. This approach is well justified in the long-wavelength limit $q \rightarrow 0$ where the system can be described by an electron liquid. Including additionally pressure and shear forces will lead to the same plasmonic dispersion as obtained from the RPA, discussed in the next section [33].

We now limit the following discussion to two dimensions (2D) and write the dispersion relation in terms of the Drude weight defined by $D = e^2 \chi_{jj}^+(\omega \rightarrow 0)$. This yields the usual expression $D_{2\text{DEG}} = e^2 n/m$ for a 2D electron gas (2DEG) with particle density n and electron mass m . Note that we have taken the local approximation $q \rightarrow 0$ before the static limit $\omega \rightarrow 0$ since the electrons cannot establish local equilibrium and remain dynamical. The other order of limits would be

related to the density of states of the electron system and thus an equilibrium property.

With the 2D Coulomb interaction $v_{\mathbf{q}} = \frac{e^2}{2\epsilon_0 \epsilon q}$ and ϵ the relative (effective) dielectric constant, the plasmon dispersion for a general 2D system in the local approximation is thus given by

$$\omega_p = \sqrt{\frac{D}{2\epsilon_0 \epsilon}} q, \quad (5)$$

yielding the characteristic square-root dispersion in D . For graphene on the interface of two different dielectric media, one further has $\epsilon = (\epsilon_1 + \epsilon_2)/2$, see equation (30).

In the local limit, $q \rightarrow 0$, we can treat general isotropic systems with energy dispersion relation $E(\mathbf{k}) \sim |\mathbf{k}|^v$ on the same footing, see the appendix. For low temperature, one obtains the general result in terms of the chemical potential μ ,

$$\chi_{jj}^+ = \frac{g_s g_v v}{2} \frac{\mu}{2\pi \hbar^2}, \quad (6)$$

with g_s , g_v , the spin- and valley degeneracies, respectively.

The 2D particle density is independent of the energy dispersion and given by $n = \frac{g_s g_v}{4\pi} k_F^2$ with k_F the Fermi wavenumber. This results in a different density behavior of χ_{jj}^+ for different $E(\mathbf{k}) \sim |\mathbf{k}|^v$. Assuming further a dependence of the dispersion on the momentum $p = \hbar k$, it is clear that the Drude weight will depend on \hbar for all $v \neq 2$ [34]. Interestingly, for the two most prominent cases, monolayer graphene and a 2DEG, we have $\chi_{jj}^+ = \frac{\mu}{\pi \hbar^2}$ and thus the same plasmon dispersion in terms of the chemical potential; only the density behavior is different.

For Dirac fermions with $v = 1$, the Drude weight can be written as

$$D_{\text{Dirac}} = \frac{e^2 v_F}{\hbar} \sqrt{\frac{g_s g_v n}{4\pi}} = \frac{4\mu}{\pi \hbar} \sigma_0, \quad (7)$$

with $\sigma_0 = \frac{g_s g_v}{16} \frac{e^2}{\hbar}$ the universal conductivity. The general plasmon dispersion of equation (5) can further be expressed with respect to dimensionless quantities as

$$\frac{\hbar \omega_p}{\mu} = \sqrt{\frac{g_s g_v \alpha_g}{2\epsilon}} \sqrt{\frac{q}{k_F}}, \quad (8)$$

with graphene's fine-structure constant $\alpha_g = \alpha \frac{c}{v_F} \approx 2.2$. For $g_s = g_v = 2$ and $\epsilon = 2$, one obtains the formula of the Introduction, equation (2).

2.1. Drude model

Dissipative effects on the plasmon dispersion are most easily included within the phenomenological Drude model which provides the corresponding conductivity. We will first recapitulate the results for 2D surface plasmon polaritons emerging on the surface of noble metals. We then discuss genuine 2D plasmons, i.e. graphene plasmons.

2.1.1. Surface plasmon polaritons. On metal/insulator interfaces, i.e. on a metal surface with negative dielectric constant covered by a dielectric medium with $\epsilon > 0$, surface plasmon polaritons can exist up to the frequency $\omega_{spp} = \omega_p^{3D} / \sqrt{1 + \epsilon}$, with the volume plasma frequency $\omega_p^{3D} = \sqrt{ne^2 / \epsilon_0 m}$ [35]. Dissipation is introduced via the three-dimensional (3D) Drude model which leads to the following local dielectric function [36]:

$$\epsilon_{3D}(\omega) = 1 - \frac{(\omega_p^{3D})^2}{\omega(\omega + i\gamma)}, \quad (9)$$

where γ denotes the electronic relaxation rate. For silver one finds $\hbar\omega_p^{3D} = 9.176$ eV and $\hbar\gamma = 21$ meV, and $\hbar\omega_p^{3D} = 9.062$ eV and $\hbar\gamma = 70$ meV for gold [37]. For energies in the visible regime, the response of bound electrons and high-energy interband transitions also need to be taken into account [35]. Nevertheless, the plasmon dispersion is fixed by the bulk properties of the underlying 3D metal which cannot easily be changed. This is one of the main disadvantages compared to graphene's genuine 2D plasmons.

2.1.2. Graphene plasmons. Graphene's plasmons are intrinsic excitations of a truly 2D system. They are defined by the following 'local' 2D dielectric function obtained from the Maxwell equations:

$$\epsilon_{2D}(q, \omega) = \epsilon + \frac{i\sigma(\omega)q}{2\omega\epsilon_0}, \quad (10)$$

with the 2D (local) Drude conductivity given by $\sigma(\omega) = ie^2\chi_{jj}^+(\omega)/(\omega + i\gamma)$. Using equation (5), we thus have a similar expression for the dielectric function as in the 3D case,

$$\epsilon_{2D}(q, \omega) = \epsilon \left(1 - \frac{\omega_p^2}{\omega(\omega + i\gamma)} \right). \quad (11)$$

Genuine 2D plasmons are defined by $\epsilon_{2D}(q, \omega_p) = 0$ and with the electronic relaxation time $\tau = 1/\gamma$ this yields

$$\omega_p^\tau = -i\frac{1}{2\tau} + \sqrt{\omega_p^2 - \frac{1}{(2\tau)^2}}. \quad (12)$$

Given the relaxation time due to Coulomb [38] or resonant [39] scattering, the damping rate of graphene plasmons can be estimated. For graphene on a substrate, one usually sets $\hbar\gamma = 10$ meV. For suspended graphene, the main scattering mechanism at finite temperature is given by flexural phonons which can be eliminated by applying strain or placing graphene on a substrate [40]. We finally note that transport lifetimes are usually calculated in the local limit, but for high-frequency plasmons the q -dependence of the response function can become important [41].

2.2. Semiclassical Boltzmann equation

The above treatment holds for any 2D electronic system and graphene's characteristic properties only entered through the local response function, i.e. ω_p and the phenomenological

relaxation time τ . We will now discuss hydrodynamic (Euler) equations which explicitly take into account the linear Dirac dispersion. We can also allow for electron as well as for hole currents and include damping terms defined via microscopic collision integrals.

To derive the Euler equations for graphene, we follow [42]. For a general discussion including also a magnetic field and the full band structure, see [43, 44]. The starting point is the continuity equation for the semiclassical distribution function $\frac{d}{dt}f_{k(t)}(\mathbf{r}, t) + \nabla \cdot \mathbf{j}_k(\mathbf{r}, t) = 0$ with the current given by $\mathbf{j}_k = \mathbf{v}_k f_k$. For graphene, the velocity reads $\mathbf{v}_k = v_F \mathbf{k}/k$ and with the equation of motion $\dot{\mathbf{p}} = e\nabla\phi$ we arrive at the collision-free Boltzmann equation for graphene,

$$\frac{\partial f}{\partial t} + v_F \frac{\mathbf{p}}{p} \frac{\partial f}{\partial \mathbf{r}} + e \frac{\partial \phi}{\partial \mathbf{r}} \frac{\partial f}{\partial p} = 0. \quad (13)$$

Expanding the Fermi distribution, $f \approx f^0 - \frac{\partial f^0}{\partial \epsilon}(\mathbf{p} \cdot \mathbf{v})$, multiplying the above equation with \mathbf{p} and integrating over the phase space $d\Gamma_{\mathbf{p}} = \frac{g_s g_v d^2 p}{(2\pi\hbar)^2}$, we arrive at the Euler equations for graphene as discussed by Ryzhii and co-workers: [42]

$$\frac{3}{2v_F} \frac{\partial \langle p \rangle \mathbf{v}}{\partial t} + \frac{v_F}{2} \frac{\partial \langle p \rangle}{\partial \mathbf{r}} - ne \frac{\partial \phi}{\partial \mathbf{r}} = 0 \quad (14)$$

with $\langle p \rangle = \int d\Gamma_{\mathbf{p}} p f_0$. The Euler equation has to be solved together with the continuity equation $\frac{\partial}{\partial t}n + \frac{\partial}{\partial \mathbf{r}}(n\mathbf{v}) = 0$.

The above equations hold for electrons as well as for holes and the effect of disorder, phonons and/or Coulomb interaction can be included by appropriate collision integrals. By linearization, one obtains analytical solutions for two limits: the symmetric bipolar and the monopolar system. In both cases, the plasmon dispersion has the same analytic structure as in equation (12), but the relaxation time is now replaced by expressions involving the collision integrals. In the monopolar case, this leads to a square-root or linear plasmon dispersion, depending on the screening behavior (of the gate), see section 3.1.5. The symmetric bipolar system can support plasmons with a sound velocity $v_s \approx 0.6v_F$, emerging due to the co-directional motion of electrons and holes. A novel plasmonic mode in neutral graphene due to excitonic effects was also reported in [45], with sound-velocity $v_s = (1 - e^{-N})v_F$ and $N = 4$ the number of fermion flavors.

Within the above Euler equations, one can also discuss the generation of plasma waves by a dc current [46], which was first proposed by Dyakonov and Shur in a 2DEG [47], and recently investigated experimentally in the context of graphene which opens up the possibility of effective THz generation [48].

3. Plasmons in single-layer graphene

A plasmon is an oscillating charge density mode which is necessarily accompanied by a corresponding electric potential, neglecting retardation effects for the moment. Density and electric potential are related via the Poisson equation and the oscillations are thus sustained by the Coulomb interaction between electrons.

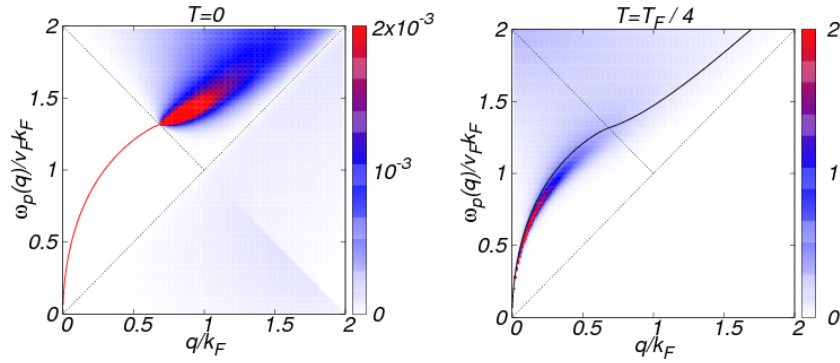


Figure 2. Generalized loss function $S(q, \omega) = -\text{Im} \chi_{\rho\rho}^{\text{RPA}}(q, \omega + i0)$ for doped graphene in units of E_F/\hbar^2 at zero temperature (left) and at $T = T_F/4$ (right). The region of undamped plasmons at $T = 0$ is defined by straight lines. The black curve on the right hand side corresponds to $T = 0$.

In order to describe plasmon excitations, the response of an electronic system to the total (screened) electric potential $\phi_{\text{total}}(\mathbf{r}, t)$ is needed, which shall be denoted by $\chi_{\rho\rho}$. But it is often more convenient to discuss the total response of the system ($\chi_{\rho\rho}^{\text{total}}$) to the external potential, $\phi_{\text{ext}}(\mathbf{r}, t)$. The long-range Coulomb interaction then needs to be treated self-consistently, leading to the following total density response:

$$\chi_{\rho\rho}^{\text{total}}(\mathbf{q}, \omega) = \frac{\chi_{\rho\rho}(\mathbf{q}, \omega)}{\epsilon(\mathbf{q}, \omega)} = \frac{\chi_{\rho\rho}(\mathbf{q}, \omega)}{1 - v_q \chi_{\rho\rho}(\mathbf{q}, \omega)}, \quad (15)$$

with the 2D Fourier transform of the electron–electron interaction $v_q = \frac{e^2}{2\epsilon_0 \epsilon q}$. The plasmonic excitations are then defined by the zeros of the dielectric function $\epsilon(\mathbf{q}, \omega)$, that also relates the total (screened) electric potential to the externally applied potential via $\phi_{\text{total}}(\mathbf{q}, \omega) = \phi_{\text{ext}}(\mathbf{q}, \omega)/\epsilon(\mathbf{q}, \omega)$. A plasmon is, therefore, a finite solution $\phi_{\text{total}}(\mathbf{q}, \omega)$ that requires no external driving, i.e. it is self-sustained.

So far, the above analysis is formally exact. Taking now the response function $\chi_{\rho\rho}$ as the bare response without including vertex corrections, this is usually coined as the RPA, $\chi_{\rho\rho}^{\text{total}} \rightarrow \chi_{\rho\rho}^{\text{RPA}}$, and represents the standard approximation to analyze the plasmonic spectrum within linear response theory. It is well justified in the high-density limit or for wave numbers $q \lesssim k_F$, with k_F the Fermi wavenumber [33, 49, 50].

Including retardation effects, the full current response needs to be considered. Assuming an isotropic system, the longitudinal (+) and transverse (−) current generated by a (total) gauge potential in linear response is given by $j^\pm = -q_e \chi_{jj}^\pm A_{\text{total}}^\pm$ with $q_e = -e$ the electron charge. The total gauge potential consists of the external potential and the field produced by the generated current, $A_{\text{total}}^\pm = A_{\text{ext}}^\pm + \Delta A^\pm$, and in linear response we have $\Delta A^\pm = -q_e d^\pm j^\pm$ with d^\pm the 2D photonic propagator, see the appendix. The total response defined through $j^\pm = -q_e \chi_{jj,\text{total}}^\pm A_{\text{ext}}^\pm$ is thus given by

$$\chi_{jj,\text{total}}^\pm = \frac{\chi_{jj}^\pm}{1 - q_e^2 d^\pm \chi_{jj}^\pm}. \quad (16)$$

Again, we have $\chi_{jj,\text{total}}^\pm \rightarrow \chi_{jj,\text{RPA}}^\pm$ in the case of a vertex-free (bare) current response.

In the following, we will summarize the basic results for the plasmonic excitations based on the RPA in single-layer graphene and also comment on various extensions. The underlying response functions are discussed in the appendix.

3.1. Gapless Dirac fermions

Graphene is a 2D crystal where the carbon atoms form a hexagonal lattice. The two equivalent atoms in the unit cell give rise to two electronic bands which touch each other at the corners of the Brillouin zone. They can be grouped together to two inequivalent K-points which are related via time-reversal symmetry and around these K-(Dirac) points, the energy dispersion is conical and isotropic with Fermi velocity $v_F \approx c/300$ [4]. Within this Dirac cone approximation the response functions are also isotropic and we may drop the vector character of \mathbf{q} to q .

The general, non-retarded plasmon dispersion including phenomenological damping is then defined by $\epsilon(q, \omega_p - i\gamma) = 0$, where γ is the decay rate of the plasmons [49]. For weak damping, the plasmon dispersion $\omega_p(q)$ and the decay rate γ are determined by

$$1 = v_q \text{Re} \chi_{\rho\rho}(q, \omega_p), \quad \gamma = \frac{\text{Im} \chi_{\rho\rho}(q, \omega_p)}{\frac{\partial}{\partial \omega} \text{Re} \chi_{\rho\rho}(q, \omega)|_{\omega_p}}. \quad (17)$$

Solutions to the first equation require $\text{Re} \chi_{\rho\rho} > 0$, which for single-layer graphene is only the case for finite doping $E_F > 0$ and $\omega > v_F q$. Furthermore, a stable solution demands $\text{Im} \chi_{\rho\rho} = 0$, which is the regime indicated by the white triangle of figure A.1. In RPA, $\epsilon \rightarrow \epsilon_{\text{RPA}}$, this yields stable δ -like excitations with an energy dispersion given in equation (8) for $q \lesssim k_F$ [51, 52]. For larger $q \gtrsim k_F$, the plasmon dispersion enters the regime of interband transitions (the violet region of figure A.1) where the plasmon becomes (Landau) damped due to dissipation into particle–hole excitations. This leads to a nonzero decay rate γ [51].

In figure 2, we plot the generalized loss function $S(q, \omega) = -\text{Im} \chi_{\rho\rho}^{\text{RPA}}(q, \omega)$ indicating intrinsic plasmon excitations, see section 5.2. The left-hand side shows the dispersion at zero temperature and the full red line stands for δ -like undamped excitations which merge into the Landau-damped regime of interband transitions.

3.1.1. Finite temperature. At finite temperature, plasmons can be sustained even by undoped graphene [53], i.e. the thermally activated charge density leads to coherent oscillations which are only weakly damped by the temperature induced interband transitions. There is a simple analytic expression for the energy dispersion using the formula of $\chi_{\rho\rho}$ for finite Fermi energy E_F at $T = 0$ by replacing $E_F \rightarrow 2 \ln 2 k_B T$ [53, 54]. This substitution also holds for bilayer graphene [55].

For finite chemical potential, no closed analytic formula for $\chi_{\rho\rho}$ is known. Still, there is a compact expression involving only a one-dimensional integral, first obtained for the density response [56]. The full current–current correlation was also derived, displaying a similar symmetry between the longitudinal and transverse channel as for the $T = 0$ result, see equation (70) [57].

On the right-hand side of figure 2, the energy loss function displaying the plasmonic resonances is shown at finite temperature $T = T_F/4$ with the Fermi temperature $T_F = E_F/k_B$. The black line indicates the plasmon dispersion at $T = 0$ obtained by equation (17) with $\gamma = 0$. The plasmonic resonances are red-shifted with respect to the $T = 0$ result, but for larger temperature $T \gtrsim T_F/2$, they become blue-shifted.

3.1.2. Local response. For small wave numbers $q \ll k_F$, the local response is sufficient for the description of the plasmonic excitations which is the case in most experimental setups. The current response function is thus often approximated by the constant Drude weight $D = e^2 \chi_{jj}^+(\omega \rightarrow 0)$ or, in terms of the conductivity, by $D = e^2 \lim_{\omega \rightarrow 0} \omega \text{Im} \sigma(\omega)$. This yields the expressions obtained from hydrodynamic models, see equation (5).

The local approximation can be improved by also including the frequency dependence of the local conductivity. One can then split up the contribution in intra- and interband processes:

$$\sigma(\omega) = \sigma_{\text{intra}}(\omega) + \sigma_{\text{inter}}(\omega). \quad (18)$$

Intraband processes lead to longitudinal plasmons, and interband processes to transverse plasmons. The local conductivity has been discussed by numerous authors in relation to magnetic fields, phenomenological disorder and finite temperature [54, 58–62].

3.1.3. Undoped graphene. The charge response function of undoped graphene was calculated in 1994, [63] and yields the characteristic square root singularity at the one-particle energy dispersion $\omega = v_F q$, discussed in the appendix. The conductivity is then given by

$$\sigma^{\mu=0, T=0}(\omega, q) = \sigma_0 \frac{\omega}{\sqrt{\omega^2 - (v_F q)^2}}, \quad (19)$$

with $\sigma_0 = \frac{g_s g_v e^2}{16\hbar}$ the universal conductivity. The conductivity σ is real for $\omega > v_F q$ and there are no plasmon excitations at zero temperature. But including vertex corrections in the polarizability leads to a positive imaginary part and a linear plasmon mode with sound velocity below the Fermi velocity emerges [45]. Undoped graphene can also sustain plasmonic oscillations when exposed to circularly polarized external electric fields [64].

3.1.4. Beyond the Dirac cone approximation. Up to now, the (bare) density response $\chi_{\rho\rho}$ was calculated within the Dirac cone approximation. But for large Fermi energies with $E_F \gtrsim 1$ eV, this must be extended to also include trigonal warping. More generally, the full hexagonal tight-binding model can be considered which is also suitable to treat chemical potentials around the van Hove singularity at ~ 3 eV.

The polarizability of the full tight-binding model has been discussed numerically [65] and within the semiclassical Boltzmann equation [44]. Interestingly, one can also obtain analytical results for $\chi_{\rho\rho}$ for small q -vectors in the high-symmetry direction Γ – M [66]. The analytical solution displays the characteristic square-root singularity at the one-particle dispersion $\omega = v_F q$ independent of the doping-dependent Fermi velocity which becomes zero at the van Hove singularity. For the general q -direction, this singularity splits into two peaks and acoustic plasmons were predicted due to different group velocities [67].

Large Fermi energies up to $E_F = 1.5$ eV are realized, e.g., in intercalated graphene [68]. But the inclusion of lattice effects has only little effect on the low-frequency plasmon dispersion with an induced anisotropy within 1%. Nevertheless, at energies close to $\hbar\omega \sim 3$ eV, i.e. the van Hove singularity, a linear dispersing damped plasmon mode emerges due to interband transitions [69]. Also for large wavevectors close to the corners of the hexagonal Brillouin zone, new low-frequency plasmon modes with a linear spectrum, so-called intervalley plasmons, emerge which are related to the transitions between the two nonequivalent Dirac cones [70].

3.1.5. Acoustic intraband plasmons. Apart from the above mentioned acoustic plasmons due to interband or intervalley scattering, the optical \sqrt{q} -plasmons can also be converted into charged acoustic (intraband) plasmons. This is due to the strong screening of a metallic gate [71] or of a substrate with a huge dielectric constant [55]. The sound velocity characterizing the acoustic plasmon dispersion $\omega = v_s q$ then reads as

$$v_s = \sqrt{4\alpha_g k_F z} v_F, \quad (20)$$

with z the distance of the graphene layer to the metallic gate or substrate. This approximation breaks down for small $k_F z$ since the sound velocity cannot become smaller than the Fermi velocity and a more careful analysis is needed [72, 73]. Linear collective dispersions are also found from a general analysis of the plasmon spectrum of graphene in the vicinity of a thick plasma-like substrate [74].

3.2. Gapped Dirac fermions

The spectrum of graphene on various substrates like boron nitride [75] or iridium [76] shows a one-particle gap. Gapped Dirac fermions can also approximately describe a number of new 2D crystals like molybdenum disulfide, MoS_2 , or other transition metal dichalcogenides [77, 78].

For undoped, but gapped graphene, the response is similar to the case of doped, but ungapped graphene by identifying

the gap parameter, Δ , with twice the Fermi energy, $2E_F$. For instance, the local conductivity for neutral graphene with one-particle gap Δ reads as

$$\text{Re } \sigma = \sigma_0 \frac{(\hbar\omega)^2 + \Delta^2}{(\hbar\omega)^2} \theta(\hbar\omega - \Delta), \quad (21)$$

$$\text{Im } \sigma = \frac{\sigma_0}{\pi} \left(\frac{2\Delta}{\hbar\omega} - \frac{(\hbar\omega)^2 + \Delta^2}{(\hbar\omega)^2} \ln \left| \frac{\Delta + \hbar\omega}{\Delta - \hbar\omega} \right| \right). \quad (22)$$

In the case of large particle gaps, the non-relativistic limit is obtained [79]. For the general, doped case, analytical results for the density response $\chi_{\rho\rho}$ [80], as well as for the current response, χ_{jj}^\pm [79], can be obtained and the corresponding plasmonic excitations were discussed within the RPA [81].

A gap in the one-particle spectrum can also be provoked artificially by graphene anti-dot lattices. In addition to the typical bulk plasmons in doped samples, interband plasmons also appear [82]. These are discussed in detail in the next section.

3.3. Interband plasmons and EELS

Before the technological advances to efficiently couple light to graphene by various near-field techniques, graphene plasmons were mainly investigated by means of high resolution EELS. Two regimes were discussed, i.e. acoustic interband plasmons at low and high energies, which will be addressed below.

3.3.1. Acoustic plasmons at low energies. For low energies $\hbar\omega \lesssim 0.5$ eV, EELS was first performed for graphene on SiC [25]. These experiments were repeated [83, 84] and extended to various metallic substrates including platinum(111) [85] and iridium(111) [86].

All experiments have in common a characteristic peak in the loss function with linear dispersion at larger energies, even though the systems are quite different. For example, graphene on SiC is doped with $E_F \approx 0.3$ eV and graphene on iridium is undoped and gapped with $\Delta \approx 0.1$ eV. Additionally, the width of the resonance shows linear behavior in all cases.

Both features, linear dispersion and linearly increasing line-width, can be captured by assuming the q -dependent conductivity of neutral graphene including only interband transitions, equation (19). The loss function $S = -\text{Im } \epsilon_{\text{RPA}}^{-1}$ is then given by

$$S(q, \omega) = \frac{x}{1 + x^2}, \quad (23)$$

$$\text{with } x = \frac{\pi\alpha_g}{2\epsilon} \frac{v_F q}{\sqrt{\omega^2 - (v_F q)^2}}$$

which shows a maximum at $x = 1$. This corresponds to a linear (acoustic) dispersion $\omega = v_s q$ with sound-velocity $v_s = \sqrt{1 + \left(\frac{\pi\alpha_g}{2\epsilon}\right)^2} v_F$. For $\epsilon \approx 3.5$, we obtain the experimentally observed sound velocity of $v_s \approx 1.4v_F$ for a SiC-substrate [25].

The above analysis is practically unchanged if we consider a lossy substrate with $\epsilon = \epsilon_R + i\epsilon_I$ and the substitution $\epsilon \rightarrow |\epsilon| = \sqrt{\epsilon_R^2 + \epsilon_I^2}$. For $|\epsilon| \approx 3.5$, we obtain the experimentally observed sound velocity of $v_s \approx 1.4v_F$ for iridium [86]. For a

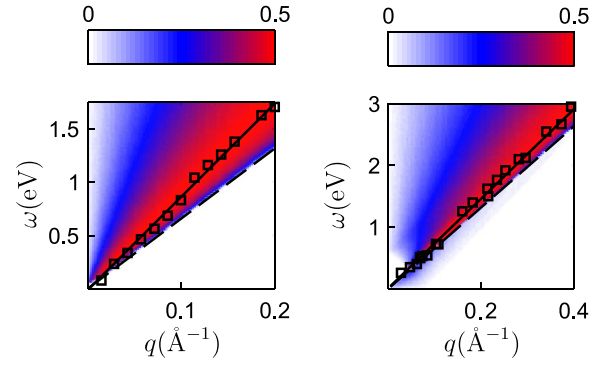


Figure 3. Loss function $S(q, \omega) = -\text{Im } \epsilon_{\text{RPA}}^{-1}(q, \omega)$ of graphene on an iridium (left) and a platinum (right) substrate due to interband transitions compared to the experimental data of [86] and [85] (squares), respectively. Also shown is the acoustic plasmon dispersion $\omega = v_s q$ with sound velocity $v_s \approx \sqrt{1 + \left(\frac{\pi\alpha_g}{2|\epsilon|}\right)^2} v_F$ for dielectric constants $|\epsilon| = 3.5$ (left) and $|\epsilon| = 6.1$ (right).

platinum substrate, one finds $v_s \approx 1.15v_F$ leading to a dielectric constant $|\epsilon| \approx 6.1$ [85].

In figure 3, the electron loss function $S = -\text{Im } \epsilon_{\text{RPA}}^{-1}$ is shown together with the experimental data of [86] (left) and [85] (right). As mentioned above, a good fit for the linear dispersion $\omega = v_s q$ is obtained for $|\epsilon| \approx 3.5$ (Ir) and $|\epsilon| \approx 6.1$ (Pt) which lies considerably below the values of the corresponding local dielectric constants. The screening behavior of metals thus seems strongly reduced at finite q which deserves further investigation.

From the above analysis, it is clear that these ‘interband plasmons’ are *not* collective excitations, but merely represent an enhanced charge resonance, i.e. they do not correspond to $\epsilon_{\text{RPA}} = 0$.

3.3.2. Acoustic plasmons at high energies. For large energies $\hbar\omega \approx 5$ eV, a peak in the loss function associated with $\pi \rightarrow \pi^*$ transitions around the van Hove singularity was first predicted by DFT studies [87], and later experimentally observed in suspended graphene by EELS [26]. These π -plasmons also display a linear dispersion. Within the hexagonal tight-binding model and RPA, i.e. without including correlation or renormalization effects, no zero of the dielectric function ϵ_{RPA} is obtained [65]. The absorption peak would thus be merely due to interband transitions enhanced by a band-structure effect. Nevertheless, in bi- or multilayer, $\epsilon_{\text{RPA}}(q, \omega)$ becomes zero around the M-point and genuine plasmons emerge [88].

3.4. Magneto-plasmons and strain

A magnetic field strongly alters the response of the electrons and thus the plasmonic excitations. The resulting magneto-plasmons have been studied, within different approaches, in [89–91]. They were observed in graphene epitaxially grown on SiC, where the Drude absorption is transformed into a strong terahertz plasmonic peak due to natural nanoscale inhomogeneities, such as substrate terraces and wrinkles [92]. Similar experiments were also performed in a graphene disk array [93] and graphene nanoribbons [94].

It was further shown that the excitation of the plasmon modifies dramatically the magneto-optical response and in particular the Faraday rotation [92]. The giant Faraday rotation due to magneto-plasmons in graphene microribbons was also recently analyzed theoretically [95].

Due to the linear dispersion of Dirac fermions, non-homogeneous strain and thus a variable hopping amplitude gives rise to pseudo-magnetic fields [96]. The influence of strain on the response function was discussed in [97] and on plasmons in [98].

3.5. Dissipative effects

Intrinsic dissipation such as one-particle scattering or temperature naturally damp plasmonic excitations and limit the propagation length of the light-like density waves. Experimentally, the damping rate seems to be larger than what would be expected from the Drude formalism [31]. This was traced back to the large absorption plateau of gated graphene for energies below the absorption threshold, $\hbar\omega \lesssim 2E_F$ [99]. Including impurity scattering due to short-range and Coulomb scatterers as well as electron–phonon interaction, the residual absorption could partially be explained [100–102], but important questions concerning the value of the plateau conductivity remain [103]. Recently, electron–electron interactions were included to address these discrepancies [104], and below we discuss this and additional lifetime limiting processes in more detail.

3.5.1. Phonons. Due to the low carbon mass, the energy of optical phonons of graphene is as large as 0.2 eV. Below this energy threshold, there are no other prominent decay channels and for THz frequencies, long-lived plasmon excitations seem possible [9]. But for large gate voltage, the plasmon dispersion hybridizes with the phonon modes which results in three new branches [28], leading to plasmon lifetimes of 20 fs or less when damping via the emission of graphene optical phonons is allowed. In [105], similar experiments with graphene nanodisks have been performed, yielding a larger lifetime approximately agreeing with the estimate coming from dc transport experiments.

Furthermore, surface polar phonons in the SiO₂ substrate under graphene nanostructures lead to a significantly modified plasmon dispersion and damping, in contrast to the case of a nonpolar diamond-like-carbon substrate [28]. Surface phonons can be treated by using a frequency dependent dielectric function. For a polar substrate, this is usually parametrized by

$$\epsilon(\omega) = \epsilon_\infty \left(1 + \frac{\omega_{LO}^2 - \omega_{TO}^2}{\omega_{TO}^2 - \omega(\omega + i\gamma)} \right), \quad (24)$$

with the phonon frequencies $\omega_{LO} = 1180 \text{ cm}^{-1}$, $\omega_{TO} = 1070 \text{ cm}^{-1}$ and the damping rate $\hbar\gamma \approx 1 \text{ meV}$ in the case of SiO₂.

3.5.2. Electron–electron interaction. The RPA is valid for wave numbers below the Thomas–Fermi screening length $\propto k_F$. Including vertex corrections in the bare charge response might lead to further dissipation channels, but the chiral nature of the Dirac carriers suppresses intrinsic plasmon losses when compared to parabolic band electrons in a 2DEG [104].

3.5.3. Nonlinear damping terms. Another possible intrinsic damping mechanism is due to nonlinear effects leading to an asymmetric broadening of the plasmon resonance [106]. Mathematically, this was traced back to the singularity in the Boltzmann equation at the neutrality point. Following this reasoning, this dissipation should vanish in the case of gapped graphene, but the final expressions of [106] are independent of a mass term. The effectiveness of this decay channel thus deserves more investigation, because this would question the general RPA-approach based on linear response.

3.6. Beyond RPA

The RPA has become a popular tool to analyze the screening and plasmonic properties of electronic systems mainly due to its simplicity. Obviously, it would be desirable to go beyond this first approximation by including more interaction terms which might have strong effects.

For undoped graphene, vertex corrections were included in the bare polarizability which leads to a novel plasmon mode in the region of intraband transitions [45, 107]. For doped graphene sheets, a diagrammatic perturbation theory to first order in the electron–electron interaction was performed and proves that the plasmon frequency and Drude weight of the electron liquid might be enhanced even in the long-wavelength limit [108].

Alternatively, the G_0W -approximation is employed where the self-energy is calculated within the Born approximation based on the bare electronic Green function G_0 and the RPA-dressed photon Green function W [109, 110]. With this approximation, angle resolved photoemission spectroscopy (ARPES) can be analyzed. APRES for epitaxially grown graphene, e.g., showed that interaction effects indeed lead to measurable changes in the energy spectrum [111]. These changes can be interpreted in terms of new quasi-particles, so-called plasmarons, that arise due to the interaction between charge carriers and plasmons [112].

3.7. Plasmons in patterned graphene

Plasmons cannot be directly excited by propagating electromagnetic radiation because the conservation of momentum is not satisfied in the photon absorption process. But periodically modulated sub-wavelength structures enable the direct coupling between propagating photonic modes and matter, see figure 1. For graphene, this has been achieved by a one-dimensional grating of nanoribbons [27, 28, 113], and also photonic-crystal-like structures [29, 105].

The plasmon dispersion in quasi-1D arrays depends on the width of the nanoribbon and the energy is lowered compared to 2D bulk plasmons due to the dipole–dipole interaction between the ribbons. The internal excitations in periodic structures have been investigated theoretically for photonic crystal-like structures based on disks [114] and anti-dots [115], nanoribbons [116], and modulated nanowires [117–119]. Also polarization-sensitive and gate-tunable photodetection in graphene nanoribbon arrays was demonstrated [120]. For more details, we refer to the pedagogical review of [121].

To numerically solve the Maxwell equations for general guided wave structures, 2D finite-difference time domain (FDTD) or finite-difference frequency domain (FDFD) techniques are widely used [122, 123]. Since retardation effects can often be neglected, self-consistent eigenvalue equation combining graphene's response with the Poisson equation yield similar results [113, 124]. For a linear stripe in the y direction, this can be formulated similarly to the hydrodynamic equation of equation (3):

$$\begin{aligned}\rho(x) &= \frac{\chi_{jj}^+}{\omega^2} (q^2 - \partial_x^2) f(x), \quad \text{with} \\ f(x) &= \frac{1}{2\pi} \int dx' K_0(q|x-x'|) \rho(x'),\end{aligned}\quad (25)$$

where ρ is the charge density, q the conserved momentum in the y -direction and K_0 denotes the modified Bessel function of the third kind. The above set of equations has to be solved self-consistently.

4. Plasmons in bilayer graphene

When exfoliating graphene by micromechanical cleavage (Scotch tape) techniques, one naturally produces graphene flakes with various number of layers N . These layers are normally Bernal or AB stacked and the unit cell contains $2N$ atoms in which half of all sites are vertically aligned. These multilayer graphene systems can be well described by including a interlayer hopping term changing the electronic spectrum and response. In this section, we will limit ourselves to bilayer graphene, $N = 2$, and discuss plasmonic excitations in AB-stacked, AA-stacked and also twisted bilayer graphene where the two layers are rotated with respect to an arbitrary angle.

4.1. Minimal stacked bilayer graphene

At low energies, a Bernal stacked bilayer can be described by two parabolic bands touching at the Dirac points, leading to a Berry phase of 2π [125]. The plasmonic spectrum shows a transition from Dirac to 2DEG plasmons [126], and analytical formulas for the effective parabolic two-band model were first presented in [127]. This was also recently discussed for finite temperature [128].

The full tight-binding model possesses four bands, including also the corresponding anti-bonding modes. Analytical formulas for the four-band model were given in [88, 129] and similar expressions can be found for graphene with spin-orbit coupling [130]. The simultaneous treatment of Coulomb interaction between and inside the layer was discussed in [131] where the formalism is naturally based on in-phase and out-of-phase excitations (see also section 5.3.1). Let us finally note that due to an optically active phonon mode and a resonant interband transition at infrared frequencies, the plasmonic properties of bilayer graphene can be strongly modified, leading to Fano-type resonances, giant plasmonic enhancement of infrared phonon absorption and a narrow window of optical transparency [132, 133].

The spectrum of bilayer graphene can become gapped by breaking the inversion symmetry between the two layers, e.g., by applying an interlayer bias [134]. The dispersion relation is then given by a Mexican hat dispersion and even though the ground state is still a Fermi liquid, the response is anomalous for small but finite energies due to the diverging density of states at the band edge [135]. This leads to novel plasmonic modes, present even for undoped biased bilayer graphene, but the physical origin of these genuine interband plasmons remains to be elucidated [136].

Apart from Bernal or AB-stacked graphene, also AA-stacked graphene can be obtained from folded graphene or twisted bilayer graphene with very small twist angle. In this configuration, all atoms are vertically aligned leading to a Fermi ring rather than a Dirac point at neutrality. The plasmon modes were discussed in [137] and have the curious property of being independent of the chemical potential in the energy region in which the two Dirac cones cross.

4.2. Twisted bilayer graphene

Apart from minimal stacked AB or AA bilayer graphene, there is also turbostratic (twisted) graphene naturally obtained from epitaxial graphene grown on the carbon-terminated face of SiC. But even with mechanical cleavage techniques, these 2D carbon systems with internal rotational disorder can be produced and transferred to virtually any substrate, e.g. BN.

For each valley, the electronic structure of twisted bilayer is defined by two Dirac points which are symmetrically separated by $\Delta K = 2|K|\sin(\theta/2)$, θ being the twist angle and $|K|$ the modulus of the two K-points [138]. The electronic spectrum is characterized by a van Hove singularity located in between the two Dirac points at energy $\epsilon_M \approx \hbar v_F \Delta K / 2$ [139], which is repeated at higher energies due to an approximate shell-structure.

The plasmon dispersion can be discussed numerically based on the local dielectric function of equation (10) by first calculating $\text{Re } \sigma$ [140], and then $\text{Im } \sigma$ by a subsequent Kramers–Kronig transformation [141]. This gives rise to four possible plasmonic modes or resonances. (i) There are undamped (conventional) graphene plasmons for chemical potentials with $\mu \ll \epsilon_M$ for which $\text{Re } \sigma = 0$ and which are governed by twice the Drude weight of single-layer graphene, $D = 2D_{\text{Dirac}}$. This energy window becomes smaller for decreasing twist angle since the van Hove singularity moves closer to the neutrality point and is only relevant for large twist angles. (ii) Due to the existence of several van Hove singularities, there are also interband ‘plasmons’, see section 3.3. These are especially dominant for large twist angle and low μ and can lead to a broad optical gap in the interband excitations where $\text{Im } \sigma < 0$, see figure 4(a). (iii) In the regime of large chemical potential, $\mu \gg \epsilon_M$, the conventional intraband plasmon is recovered, albeit Moiré-damped due to intrinsic (twist) disorder. The dispersion only depends slightly on the twist angle and becomes well-defined for large μ , extending into the Landau-damped region just as for the monolayer, see figure 4(d). (iv) Finally, due to the van Hove singularities, the imaginary part can become negative,

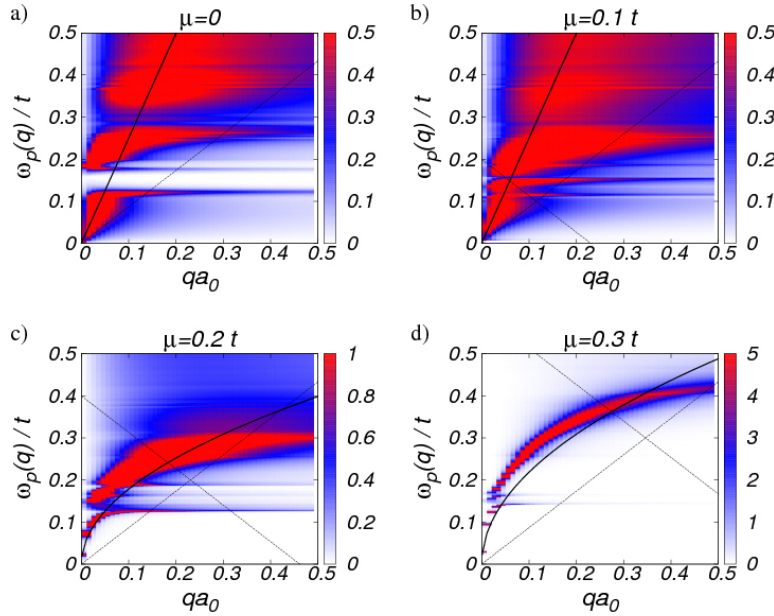


Figure 4. Loss function $S(q, \omega) = -\text{Im} \epsilon_{\text{RPA}}^{-1}(q, \omega)$ in the long-wavelength RPA for twist angle $\theta = 3.15^\circ$ and various chemical potentials $\mu/t = 0, 0.1, 0.2, 0.3, 0.4$ with $\epsilon = 2.4$ assuming a SiO_2 substrate. The straight full line in (a) and (b) corresponds to acoustic interband plasmons, the curved full line in (c) and (d) to undamped intraband plasmons—both for the decoupled bilayer. The region of undamped plasmons of the decoupled bilayer is defined by thin dashed lines.

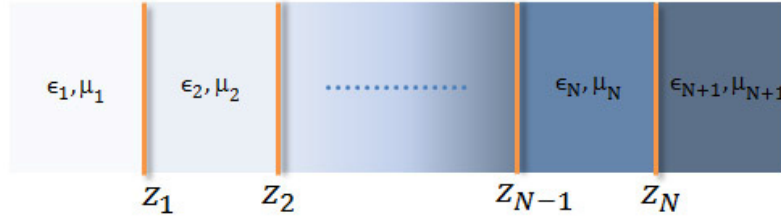


Figure 5. Schematic setup of the multilayer graphene structure. The graphene (2DEG) layers, characterized by graphene's (2DEG) current response, $\chi_{jj}^{\pm, i}$ at position z_i , are surrounded by different dielectric media characterized by the relative dielectric constants ϵ_i and the relative magnetic permeabilities μ_i .

$\text{Im} \sigma < 0$, opening up the existence of transverse plasmons, see section 6.2.

In figure 4 the loss function $S = -\text{Im} \epsilon_{\text{RPA}}^{-1}$ is shown for twist angle $\theta = 3.15^\circ$ and various chemical potential with $\epsilon = 2.4$ assuming a SiO_2 substrate. Similar results are obtained for smaller angles [141]. In all cases, the Dirac cone dispersion $\omega = v_F q$ and $\hbar\omega = 2E_F - \hbar v_F q$ are shown as dashed lines, indicating the onset of intra- and interband transitions (see also figure A.1). Also shown are the acoustic interband ‘plasmons’ (see section 3.3) with sound velocity $v_s \approx \pi \alpha_g v_F / \epsilon$ (solid line in (a) and (b)) and the optical \sqrt{q} plasmonic mode of equation (5) (solid line in (c) and (d)) for decoupled bilayer.

5. Plasmons in general layered structures

There is renewed focus on layered structures due to experimental advances in exfoliating a number of 2D materials and combining them in vertical stacks. Double-layer structures can thus be fabricated with relatively narrow and low energy barriers [142], leading to novel devices like broadband optical modulator [143] or vertical field-effect transistors [144].

Here, we present the basic steps to derive the plasmon dispersion of multi-layer graphene or other 2D electronic systems including full retardation. We further assume that the layers only interact among themselves via Coulomb interaction; to also include coherent interlayer hopping, the 2D response function must be written in matrix form and would contain non-diagonal entries [131, 136]. An alternative method based on a simple analytical transfer-matrix approach can be found in [145].

5.1. Undamped plasmons

To discuss electromagnetic bound states, it is convenient to work within the Weyl gauge, setting the electrostatic potential equal to zero, $\phi = 0$. We then only have to consider the vector field \mathbf{A} , and the coupling to a 2D electronic system (in the following, we will mainly discuss graphene) is entirely described by the current–current correlation function. We will treat the general multi-layer system of figure 5 and first discuss the longitudinal or p-polarization.

5.1.1. Longitudinal or p-polarization. For longitudinal polarization, the general vector field has a component parallel and normal (z) to the interface,

$$\mathbf{A}(\mathbf{r}, z) = \sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} \left(A^{\parallel}(\mathbf{q}, z) \mathbf{e}_{\mathbf{q}} + A^{\perp}(\mathbf{q}, z) \mathbf{e}_z \right). \quad (26)$$

The components of A^{\perp} can be obtained from the components of A^{\parallel} via the condition for a transverse field $\nabla \cdot \mathbf{A} = 0$. It thus suffices to discuss the parallel component which is continuous at the interfaces. With $A^{\parallel} = \sum_i A_i^{\parallel}$, we make the general ansatz for the gauge field in medium i ,

$$A_i^{\parallel}(\mathbf{q}, z) = a_i e^{-q'_i z} + b_i e^{q'_i z}, \quad z_{i-1} \leq z < z_i, \quad (27)$$

with the perpendicular wavevector $q'_i = \sqrt{q^2 - (\omega/c_i)^2}$ and $c_i = c/\sqrt{\epsilon_i \mu_i}$ the speed of light in the corresponding medium. The two boundary conditions at the i th interface are related to the continuity of the vector field and the discontinuity of the displacement field:

$$a_j e^{-q'_j z_i} + b_j e^{q'_j z_i} = a_{i+1} e^{-q'_{i+1} z_i} + b_{i+1} e^{q'_{i+1} z_i}, \quad (28)$$

$$\begin{aligned} q'_{i+1}(\epsilon_i - \alpha_i) a_j e^{-q'_j z_i} - q'_{i+1}(\epsilon_i + \alpha_i) b_j e^{q'_j z_i} \\ = \epsilon_{i+1} q'_i a_{i+1} e^{-q'_{i+1} z_i} - \epsilon_{i+1} q'_i b_{i+1} e^{q'_{i+1} z_i}, \end{aligned} \quad (29)$$

where $\alpha_i = e^2 \chi_{jj}^{+,i}(\mathbf{q}, \omega) \frac{q'_i}{\epsilon_0 \omega}$.

In the case of N graphene interfaces ($z_0 \rightarrow -\infty, z_{N+1} \rightarrow \infty$), we set $a_1 = b_{N+1} = 0$ and in the absence of dissipation ($\text{Im} \chi_{jj}^{+,i} = 0$) we have a homogeneous set of $2N$ linear (real) equations with $2N$ variables, $Mx = 0$. The condition $\det M = 0$ then yields N plasmon modes with positive wavenumber q . In the non-retarded limit [146] they split into one optical mode with square-root dispersion and $N - 1$ acoustic modes with linear dispersion for small q [147].

For a single layer, the boundary conditions yield the implicit plasmon dispersion,

$$\omega^2 = e^2 \frac{q'_1 q'_2}{\epsilon_0 (\epsilon_2 q'_1 + \epsilon_1 q'_2)} \chi_{jj}^{+}. \quad (30)$$

Neglecting retardation effects ($q'_1 = q'_2 = q$), and approximating the current response by the Drude weight $e^2 \chi_{jj}^{+} \rightarrow D$, we recover the familiar expression for the plasmon dispersion $\omega_p^2 = \frac{D}{\epsilon_0 (\epsilon_1 + \epsilon_2)} q$. This provides the usual substitution rule for the dielectric constant in equation (5), $\epsilon \rightarrow (\epsilon_1 + \epsilon_2)/2$. The electron–electron interaction of 2D electrons is thus equally mediated through the upper and lower dielectric medium.

5.1.2. Transverse or s-polarization. For transverse polarized light, only the parallel component is nonzero. We can thus write

$$A^{\parallel}(\mathbf{r}, z) = \sum_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} A^{\parallel}(\mathbf{q}, z) \quad (31)$$

and make the same ansatz as in equation (27):

$$A_i^{\parallel}(\mathbf{q}, z) = a_i e^{-q'_i z} + b_i e^{q'_i z}, \quad z_{i-1} < z < z_i. \quad (32)$$

The two boundary conditions at the i th interface are related to the continuity of the vector field and the discontinuity of the magnetic field. They are obtained from equation (28) by substituting $q'_i \rightarrow \mu_i, \epsilon_i \rightarrow q'_i, \omega \rightarrow c$ and $\chi_{jj}^{+} \rightarrow -\chi_{jj}^{-}$. For a single layer, the plasmon dispersion is then defined by

$$\mu_2 q'_1 + \mu_1 q'_2 + \mu_1 \mu_2 \mu_0 e^2 \chi_{jj}^{-}(\mathbf{q}, \omega) = 0, \quad (33)$$

whose possible solutions strongly depend on the surrounding dielectric media [148]. For a discussion on general double-layer graphene structures, see [57].

Note that the symmetry between p- and s-polarization is normally not present. But here, we base our discussion on the parallel field component which is continuous in both cases, in contrast to the (usually discussed) total field which is (dis)continuous for transverse or s (longitudinal or p) polarization. In the following, we will focus on the longitudinal polarization, but discuss in detail transverse plasmons in section 6.2.

5.2. Damped plasmons and energy loss function

In the presence of dissipation, the plasmon dispersion ceases to be well defined. Therefore, to characterize damped plasmons, one frequently relies on the energy loss function defined as $S(\mathbf{q}, \omega) = -\text{Im} \epsilon^{-1}(\mathbf{q}, \omega + i0)$. This is a measure of the spectral density of the intrinsic plasmonic excitations: a sharp peak in $S(\mathbf{q}, \omega)$ reveals long-lifetime plasmons; undamped plasmons, defined by $\epsilon(\mathbf{q}, \omega) = 0$, correspond to a delta peak in $S(\mathbf{q}, \omega)$.

This formalism needs to be extended for two or more (N) interfaces and graphene's response is then given by a $N \times N$ -matrix for each polarization (we will drop this index in the following). Within RPA, this gives the following matrix equation:

$$\chi_{\text{RPA}} = (\mathbf{1} - e^2 \chi_{jj} \mathbf{d})^{-1} \chi_{jj} \equiv \epsilon_{\text{RPA}}^{-1} \chi_{jj}. \quad (34)$$

The $N \times N$ -matrix χ_{jj} denotes the (bare) graphene response which is diagonal in the absence of (coherent) interlayer coupling. The $N \times N$ -matrix \mathbf{d} is the (bare) photon propagator in the absence of graphene ($\chi_{jj}^i = 0$), but with the dielectric geometry of figure 5. The entries of \mathbf{d} can be obtained from the standard matching conditions or, equivalently, using multiple scattering formalism. Diagonalizing the response matrix χ_{RPA} , one obtains the elementary excitations of the full system, i.e. in-phase and out-of-phase mode in the case of $N = 2$.

Let us now define the energy loss function for arbitrary multi-layer structures. We emphasize this point because the plasmonic spectrum was frequently discussed by $S = -\text{Im} \epsilon_{\text{RPA}}^{-1}$, where the (scalar) dielectric function was obtained by $\epsilon_{\text{RPA}} = \det \epsilon_{\text{RPA}}$ [149]. But this 'loss function' changes sign and can thus not be interpreted as a (positive definite) spectral density. Instead of the determinant, one rather needs to discuss the trace of the dielectric matrix [136]. But graphene's excitations correspond to the imaginary part of the full response, χ_{RPA} , and the relative response of the several layers might differ. It is thus more appropriate to define the following generalization of the energy loss function:

$$S(\mathbf{q}, \omega) = -\text{Im} \text{Tr} \chi_{\text{RPA}}(\mathbf{q}, \omega + i0). \quad (35)$$

Since $S(\mathbf{q}, \omega)$ is related to the imaginary part of a causal function, it is strictly positive and reveals the presence of the intrinsic excitations of the multi-layer system. It is further invariant with respect to unitary transformations between the several layers.

5.3. Double layer

For the special case of two graphene layers, the above matrix $\chi_{jj} = \text{diag}(\chi_{jj}^1, \chi_{jj}^2)$ represents the bare graphene's response in layer 1 (χ_{jj}^1) and layer 2 (χ_{jj}^2). The photon propagator \mathbf{d} for the two polarizations can be found in [72].

Undamped (longitudinal) plasmonic excitations are defined as usual by the zeros of the dielectric function $\det \epsilon_{\text{RPA}} = 0$. For two layers without retardation, this is often written in terms of the charge density response of the two layers, $\chi_{\rho\rho}^{1/2}$:

$$(1 - v_1 \chi_{\rho\rho}^1)(1 - v_2 \chi_{\rho\rho}^2) - v_{12}^2 \chi_{\rho\rho}^1 \chi_{\rho\rho}^2 = 0, \quad (36)$$

where $v_{1/2}(q)$ and $v_{12}(q)$ are the intra- and interlayer Coulomb interaction, respectively. For different dielectric media on the left (ϵ_1), center (ϵ_2) and right (ϵ_3), and $z = z_2 - z_1$ the distance between the two layers, the general expressions for the intra- and interlayer are given by [55, 73, 150] $v_{1/2} = [\cosh(qz) + (\epsilon_{3/1}/\epsilon_2) \sinh(qz)]v_{12}(q)$ and $v_{12} = e^2 \epsilon_2 / (\epsilon_0 q N)$ with $N = \epsilon_2(\epsilon_1 + \epsilon_3) \cosh(qz) + (\epsilon_1 \epsilon_3 + \epsilon_2^2) \sinh(qz)$.

Including retardation effects, we have to solve

$$\begin{aligned} & (q_2' \epsilon_1 + q_1' \epsilon_2 - q_2' \alpha_1)(q_2' \epsilon_3 + q_3' \epsilon_2 - q_3' \alpha_2) \\ & - (q_2' \epsilon_1 - q_1' \epsilon_2 - q_2' \alpha_1) \\ & \times (q_2' \epsilon_3 - q_3' \epsilon_2 - q_3' \alpha_2) e^{-2q_2' z} = 0, \end{aligned} \quad (37)$$

where $\alpha_i = e^2 \chi_{jj}^{+,i} q_i' / (\epsilon_0 \omega^2)$ and $z = z_2 - z_1$ again the distance between the two layers.

Below, we will discuss the two elementary modes of this system in more detail and also comment on near-field amplification at certain energies.

5.3.1. Optical and acoustic modes. The plasmonic spectrum of double-layer graphene is characterized by a bonding and anti-bonding mode due to the electrostatic coupling between the two layers [149]. In the case of longitudinal plasmons this leads to an ordinary (optical) 2D plasmon with \sqrt{q} -dispersion, but with larger energy since the charges of the two layers oscillate in phase. It also leads to a linear (acoustic) plasmon mode where the charges oscillate out of phase. In the case of transverse plasmons, there is no charge accumulation in the graphene layer and we find either one mode (for small layer separation) or two plasmon modes (for large layer separation) [72].

Usually, the plasmon dispersion is well separated from the light cone and we can set $q_1' = q_2' = q_3' = q$. Another approximation is given by the local response valid in the long-wavelength limit $qz \ll 1$, i.e., replacing the current response by the corresponding Drude weight, $e^2 \chi_{jj}^{+,i} = D_i$. The optical mode for $\omega \gg v_F q \rightarrow 0$ is then obtained as

$$\omega_{op}^2 = g_s g_v \alpha_g v_F^2 (k_F^1 + k_F^2) \frac{q}{\epsilon_1 + \epsilon_3}, \quad (38)$$

with graphene's fine-structure constant $\alpha_g = \alpha \frac{c}{v_F} \approx 2.2$. The acoustic mode reads for $(k_F^1 + k_F^2)z/\epsilon_2 \gg 1$

$$\omega_{ac}^2 = g_s g_v \alpha_g v_F^2 z \frac{k_F^1 k_F^2}{(k_F^1 + k_F^2) \epsilon_2} q^2. \quad (39)$$

The optical mode only depends on the sum of the outer dielectric media $\epsilon_1 + \epsilon_3$ whereas the acoustic sound velocity only depends on the dielectric medium in the center, ϵ_2 . This is a general result because for the optical (in-phase) mode the interfaces have the same homogeneous charge density in the limit $qz \rightarrow 0$, thus not polarizing the inner medium. For the acoustic (out-of-phase) mode in the same limit, there are opposite homogeneous charge densities on the two sheets just as for a capacitor which in turn does not polarize the surrounding media.

For general parameters, the acoustic mode must be obtained in terms of a Laurent–Taylor expansion including the full expression of the response function [151]. The square-root singularity of $\chi_{jj}^+(q, \omega)$ at $\omega = v_F q$ then guarantees that the sound velocity is always greater than the Fermi velocity, $v_s > v_F$ [55]. The general analytical expression has been obtained by Profumo *et al* [73].

The range of applicability of the analytical formula for the optical mode, equation (38), depends on the relative value of ϵ_1, ϵ_3 with respect to ϵ_2 and is only valid if they are of the same order. The general formula valid for $qz \ll 1$ is given by

$$\frac{\omega_+^2}{g_s g_v \alpha_g v_F^2 q} = \frac{\epsilon_2 (k_F^1 + k_F^2) + qz (\epsilon_1 k_F^2 + \epsilon_3 k_F^1) + \sqrt{R}}{2 [\epsilon_2 (\epsilon_1 + \epsilon_3) + qz (\epsilon_2^2 + \epsilon_1 \epsilon_3)]} \quad (40)$$

with $R = \epsilon_2^2 (k_F^1 + k_F^2)^2 - 2qz \epsilon_2 (k_F^1 - k_F^2) (\epsilon_1 k_F^2 - \epsilon_3 k_F^1) + (qz)^2 (\epsilon_1 k_F^2 - \epsilon_3 k_F^1)^2$. For a 3D topological insulator with $\epsilon_2 \gg \epsilon_1, \epsilon_3$ and in the case of equal densities $k_F^1 = k_F^2 = k_F$, this simplifies to

$$\omega_+^2 = \frac{2g_s g_v \alpha_g v_F^2 k_F q}{\epsilon_1 + \epsilon_3} \left[1 + \frac{qz \epsilon_2}{\epsilon_1 + \epsilon_3} \right]^{-1}. \quad (41)$$

In this case, equation (38) is only valid for $qz \epsilon_2 / (\epsilon_1 + \epsilon_3) \ll 1$.

The left-hand side of figure 6 shows the energy loss function defined in equation (35) at zero temperature and the full red lines stand for δ -like undamped excitations which merge into the Landau-damped regime of interband transitions. The right-hand side shows the loss function at finite temperature $T = T_F/4$ with the Fermi temperature $T_F = E_F/k_B$. The black curves indicate the plasmon dispersion at $T = 0$ obtained by equation (36) without dissipation, $\text{Im} \chi_{jj}^{+,i} = 0$. The finite-temperature plasmonic resonances are slightly red-shifted compared to the $T = 0$ dispersion as was the case in single-layer graphene.

5.3.2. Near-field amplification. In the case of two plasmon modes, there exists a frequency where the transmission is exponentially amplified, reminiscent to the situation of what happens in 'Pendry's perfect lens' [20]. The frequency lies in between the two plasmon frequencies and is pinned to the out-of-phase mode for small wavenumbers. For large

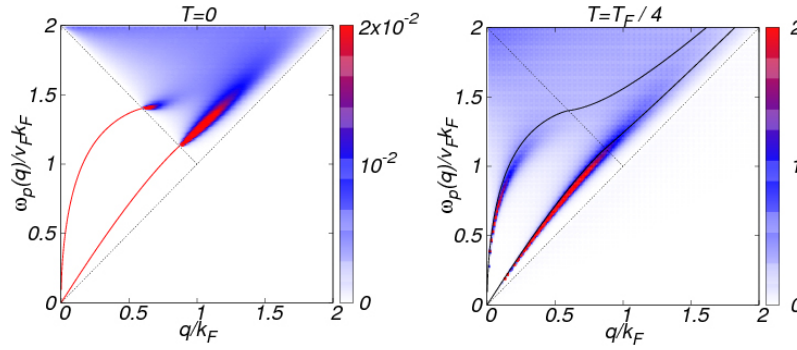


Figure 6. Generalized loss function $S(q, \omega) = -\text{Im Tr } \chi_{\text{RPA}}(q, \omega + i0)$ for doped graphene in units of E_F/\hbar^2 at zero temperature (left) and at $T = T_F/4$ (right). The region of undamped plasmons of the decoupled bilayer at $T = 0$ is defined by straight lines. The black curves on the right hand side correspond to $T = 0$.

wavenumbers or interlayer distances, the two plasmon modes merge and sandwich the frequency of exponential amplification. The exponential transmission

$$T_{ex} = \frac{q'_1 \epsilon_2 - q'_2 \epsilon_1 + 2q'_2 \alpha_1}{q'_3 \epsilon_2 - q'_2 \epsilon_3 - 2q'_3 \alpha_2} \frac{q'_3}{q'_1} e^{(q'_2 + q'_3)z}, \quad (42)$$

with $\alpha_i = e^2 \chi_{jj}^{+,i} q'_i / (\epsilon_0 \omega^2)$ and $z = z_2 - z_1$, is accompanied by zero reflection, $R_{ex} = 0$, and similar expressions hold for transverse plasmons [72]. For different densities in the two layers, the energy for near-field amplification depends on the arrangement of the layers.

The possibility of exponential amplification might be useful for near-field microscopies and deserves further investigation for general multi-layer structures including dissipation.

6. Plasmons including retardation effects

Due to their large momentum, 2D longitudinal plasmons do not easily couple to propagating electromagnetic radiation, and retardation effects can usually be neglected. But lowering the frequency, the unretarded square-root dispersion will finally cross the light cone, pronouncing the onset of retardation effects, indicated by the circled region of figure 1(a).

On the other hand, in order to discuss transverse plasmons, full retardation is always needed since the plasmon dispersion is closely pinned to the light cone. In section 6.2, we will discuss general aspects of these excitations which lead to broadband polarization in graphene waveguides [152]. They are also present in gapped and one-dimensional structures as we will show below.

6.1. Longitudinal or TM plasmons

The standard expression for a 2D plasmon, $\omega_p \propto \sqrt{q}$, assumes instantaneous Coulomb coupling between charges [51, 52]. Therefore, it cannot be correct when the nominal plasmon dispersion meets the light cone, $\omega_p \lesssim cq$. In this regime, even homogeneous graphene plasmons must couple strongly to (propagating) light and we will discuss the phenomena associated with the strong light–graphene coupling for single- and double-layer structures.

6.1.1. Single-layer graphene. We consider a single graphene sheet between two dielectrics with $\epsilon_1 > \epsilon_2$. Graphene plasmons are then obtained from equation (30) which includes retardation, and in the unretarded limit $c \rightarrow \infty$, this gives the known square-root dispersion. In contrast, the exact dispersion is linear below a characteristic crossover frequency ω_c , and for reasonable dielectric constants this scale is given by $\omega_c \sim \alpha \omega_F$ with $\omega_F = E_F/\hbar$. The plasmon dispersion thus merges with the light cone of the slower medium and one obtains the following asymptotic behavior: [153]

$$\left(\frac{\omega}{\omega_F}\right)^2 = \begin{cases} \left(\frac{4\alpha_g}{\epsilon_1 + \epsilon_2}\right) \left(\frac{q}{k_F}\right)^2, & \omega \gtrsim \omega_c \\ \left(\frac{c_1}{v_F}\right)^2 \left(\frac{q}{k_F}\right)^2, & \omega \lesssim \omega_c, \end{cases} \quad (43)$$

with $c_1 = c/\sqrt{\epsilon_1}$ the (slower) light velocity inside medium 1. The crossover between the two regimes takes place for frequencies which roughly correspond to the intersection of the unretarded plasmon and light-cone dispersion. This yields $\nu_c \sim 600$ GHz for doping level $n \sim 10^{13} \text{ cm}^{-2}$, reaching the technologically important THz regime for $n \sim 10^{14} \text{ cm}^{-2}$.

The linear regime $\omega \lesssim \omega_c$ is also the region of strong graphene–light coupling. This can be seen by looking at the reflection and transmission amplitudes for the (in-plane) longitudinal vector potential upon passing from medium i to j , given by [55].

$$r_{ij} = \frac{(\epsilon_i q'_j - \epsilon_j q'_i) \epsilon_0 \omega^2 + q'_i q'_j e^2 \chi_{jj}^+}{(\epsilon_i q'_j + \epsilon_j q'_i) \epsilon_0 \omega^2 - q'_i q'_j e^2 \chi_{jj}^+}, \quad (44)$$

and $t_{ij} = 1 + r_{ij}$.

For interband transitions with $\omega \gtrsim 2\omega_F$, graphene's response in equation (44) is small, leading to the universal 2.3% weak absorption in vacuum [154, 155]. On the other hand, for $\omega \lesssim \omega_c$, graphene response starts to dominate in equation (44) implying strong radiation–graphene coupling. For instance, the reflection amplitude becomes $r \sim -1$ for $\omega \ll \omega_c$, meaning (almost) perfect reflection for single-layer graphene. This perfect reflection is converted into perfect absorption when losses are allowed [156], providing a complementary and potentially simpler alternative to absorption enhancement based on periodic patterning.

6.1.2. Double-layer graphene. Enhanced light–matter interaction also leads to *extraordinary transmission* for a double-layer graphene arrangement. This term was originally coined to describe the enormous transmission experimentally observed through periodically perforated metal sheets, where naive expectation would assume just the opposite [11]. An explanation was provided in terms of the excitation of surface plasmons which results in enhanced (perfect without dissipative) transmission through a nominally opaque region [12].

In the case of double-layer graphene where the central dielectric is less than the surrounding ones $\epsilon_2 < \epsilon_1, \epsilon_3$, the resonant coherent excitations of the graphene layers also allow for the enhanced transmission of photons through the central, classically forbidden region for photons, in direct analogy with the metallic case.

The perfect transmission through the evanescent region is accompanied by a maximum in the spectral photonic density and is thus due to a plasmonic response of the double-layer graphene system. This enhanced light–matter interaction is also present in a general setup, in particular in the allowed (propagating) region where Fabry–Pérot resonances emerge. These Fabry–Pérot resonances become strongly quenched compared to the case without the graphene layers and the response, i.e. the spectral density, displays a typical Fano line-shape. We can interpret this as the formation of quasi-localized states between the doped graphene layers which slightly leak out and thus interact with the incoming (continuous) light field.

The sharp response of graphene in the absence of absorption also leads to enhanced absorption when losses are allowed, and the setup is similar to the previously suggested enhanced absorption of graphene placed in a (double) Fabry–Pérot cavity [157, 158]. We finally note that there is a critical layer separation for the emergence of a separated acoustic mode, z_c . For layer separations with $z > z_c$ both modes, the in-phase and out-of-phase mode, coincide with the light cone [153].

6.2. Transverse or TE plasmons

Collective charge density fluctuations are accompanied by collective longitudinal current fluctuations as dictated by the continuity equation. But there is also the possibility for collective transverse current fluctuations, see equation (16). Whereas longitudinal plasmons can only exist for a metallic response, $\text{Re } \chi_{jj}^+ > 0$, transverse plasmons require a dielectric response, $\text{Re } \chi_{jj}^- < 0$, because the photon propagator changes sign for the two polarization channels, see equation (74).

Transverse or TE plasmons are light-like excitations defined in equation (33) and were first discussed in [159] in the case of suspended graphene. Due to their transverse nature, they are closely pinned to the light cone which makes them only weakly confined to the graphene sheet. For graphene on the interface of two distinct dielectric media, though, no solution is found because the different (local) light cones are too much separated in energy to simultaneously host the transverse plasmon [148]. Also multi-layer structures with special dielectric media, e.g. superconductors with $\mu = 0$, that might localize TE plasmons, do not lead to a solution after a critical layer separation [57].

In the retardation regime, the local optical conductivity is usually sufficient to discuss TE plasmons. This yields the following equation for suspended graphene:

$$1 - \frac{i\omega\sigma(\omega)}{c^2\sqrt{q^2 - \omega^2/c^2}} = 0. \quad (45)$$

Writing the optical conductivity in terms of intra- and inter-band transitions, $\sigma = \sigma_{\text{intra}} + \sigma_{\text{inter}}$, with $\text{Im } \sigma_{\text{intra}} > 0$, transverse plasmons can only be sustained by doped graphene when interband transitions with $\text{Im } \sigma_{\text{inter}} < 0$ prevail. This is the case for energies $1.667 < \hbar\omega/E_F < 2$ where $\text{Im } \sigma < 0$. With the dimensionless constants $Q = \hbar cq/E_F$ and $\Omega = \hbar\omega/E_F$, the TE plasmon dispersion in suspended graphene then reads as [159]

$$\sqrt{Q^2 - \Omega^2} = 2\alpha \left(\frac{\Omega}{4} \ln \left| \frac{2 + \Omega}{2 - \Omega} \right| - 1 \right), \quad (46)$$

where α is the fine-structure constant.

Due to their small spectral weight, they have not been directly observed, yet. Nevertheless, strain will lead to field enhancement [97] and also in bilayer graphene this mode is expected to be much stronger [160]. In double-layer structures with a certain layer separation, the TE mode can split into a bonding and anti-bonding mode and in between these two modes, a region of exponential near-field amplification can be defined [72]. Another detection method is based on fluorescence quenching of a dye due to the presence of doped graphene and the non-radiative decay rate will entirely be defined by transverse plasmons at large distances [161].

6.2.1. Gapped graphene. Without a charge density, one cannot generate longitudinal plasmons, i.e. collective charge density fluctuations. But it is possible to generate transverse plasmons, i.e., collective current fluctuations which are transverse to the direction of wave propagation. TE plasmons are thus present for gapped graphene even when the chemical potential lies inside the gap, Δ . They are even more prominent compared to doped graphene ($2E_F \rightarrow \Delta$) due to the pure dielectric response of the system ($\sigma_{\text{intra}} = 0$) and exist in the whole energy window $0 < \hbar\omega/\Delta < 1$. These collective current fluctuations associated with energies below the gap are possible since the generated particle–hole pairs are lowered in energy due to the attractive long-ranged Coulomb interaction.

The dispersion of this mode, assuming local response, is pinned to the light cone but diverges logarithmically for energies close to the band gap in analogy to equation (46)

$$\sqrt{Q^2 - \Omega^2} = \alpha \left(\frac{\Omega^2 + 1}{2\Omega} \ln \left| \frac{1 + \Omega}{1 - \Omega} \right| - 1 \right), \quad (47)$$

where α is the fine-structure constant and the dimensionless constants are $Q = \hbar cq/\Delta$ and $\Omega = \hbar\omega/\Delta$. For a gap of $\Delta \approx 0.1$ eV, the well-defined plasmonic modes would thus correspond to approximately 24 THz.

6.2.2. One-dimensional transverse plasmons. There can also exist purely transverse excitations in a quasi one-dimensional nanowire which is surprising since one would always expect charge accumulation at the border due to the transverse current oscillations. But for energies exponentially pinned to the light cone, the transverse component of the propagator becomes negative, opening up the possibility of transverse excitations for a dielectric response, see the appendix.

If we model the (graphene) nanoribbon by a cylindrical dielectric nanowire, characterized by a local susceptibility $\chi_e(\omega) > 0$, the bare current response is given by $\chi_{jj}^- = -\omega^2 \chi_e$. The self-consistent equation defining the current response then reads as

$$\delta j^\alpha(\mathbf{r}) = -\chi_{jj}^- \left(A_{\text{ext}}^\alpha(\mathbf{r}) - \int d^2 r' \mathcal{D}^{\alpha,\beta}(\mathbf{r} - \mathbf{r}') \delta j^\beta(\mathbf{r}') \right), \quad (48)$$

with $\mathcal{D}^{\alpha,\beta}$ the photonic propagator, see the appendix. If the external field and the induced current distribution are homogeneous, we can average over the cylinder and obtain the following RPA-response function

$$\chi_{jj}^{\text{RPA},-} = \frac{\chi_{jj}^-}{1 - d_1^- \chi_{jj}^-}, \quad (49)$$

with d_1^- the one-dimensional transverse photon-propagator, see equation (79).

For a dielectric response $\chi_{jj}^- < 0$, there are collective modes only if $d_1^- < 0$. These modes are closely pinned to the light cone and with equation (79) we get

$$\left(q - \frac{\omega}{c} \right) \approx \frac{c}{2\omega a^2} \exp \left\{ 1 - \left(1 + \frac{2\pi a^2 \epsilon_0}{\chi_e} \right) \frac{2c^2}{\omega^2 a^2} \right\}, \quad (50)$$

where a is the radius of the cylinder.

We finally note that in three dimensions there are no collective transverse current oscillations which are separated from the light cone.

7. Plasmons in Dirac systems with strong spin-orbit coupling

The discovery of graphene triggered the search for other layered quasi-2D crystals like optically active 2D transition metal dichalcogenides MoS₂, MoSe₂ or WS₂. It also stimulated the search for new states of condensed matter resulting in a paradigmatic model for 2D topological insulators [162, 163]. Around the Dirac points, it is given by the graphene Hamiltonian with a positive and negative gap with respect to the two K-points which arises from an intrinsic spin-orbit coupling. Including also Rashba spin-orbit coupling, one can tune the system from the quantum spin-Hall to the normal phase which are separated by a quantum critical point.

Another system is represented by Hg(Cd)Te quantum wells [164] described by the Bernevig-Hughes-Zhang model [165]. In the quantum spin-Hall phase, electrons have intermediate properties between Dirac and Schrödinger fermions

which give rise to plasmonic resonances even in the undoped limit [166].

In this section, we will discuss the plasmonic spectrum for graphene with intrinsic and Rashba spin-orbit coupling and the direct band gap 2D semiconductor MoS₂. We close with a discussion on the recently measured plasmonic spectrum of the 3D topological insulators Bi₂Se₃ [167].

7.1. Graphene plasmons with spin-orbit coupling

Let us define the graphene Hamiltonian with spin-orbit coupling in the Dirac cone approximation [163]

$$H = v_F \vec{p} \cdot \vec{\tau} + \lambda_R (\vec{\tau} \times \vec{\sigma}) \cdot \vec{e}_z + \lambda_I \tau_z \sigma_z, \quad (51)$$

where $\vec{\sigma}$, $\vec{\tau}$ denote the Pauli matrices referring to the pseudo-spin and valley degrees of freedom, respectively. For a sufficiently large intrinsic coupling parameter, $\lambda_I > \lambda_R$, the system is in the spin quantum Hall phase with a characteristic band gap. For $\lambda_R > \lambda_I$ the gap in the spectrum is closed and the system behaves as an ordinary semi-metal. At $\lambda_R = \lambda_I$ a quantum phase transition occurs in the system.

Closed analytical expressions for the complex polarizability of the above model have been obtained in [168] and we will follow this discussion. Similar to the case of massive Dirac fermions [80], and bilayer graphene [129], this leads to several solutions of $\epsilon_{\text{RPA}}(q, \omega) = 0$ for nonzero SOC parameters. One of these solutions has an almost linear dispersion with a sound velocity close to the Fermi velocity which exhibits an ending point for $\lambda_R \sim \lambda_I$ associated with a double zero of $\text{Re } \epsilon_{\text{RPA}}$. This solution does not lead to a resonance in the loss function and thus does not resemble a plasmonic mode. In the case where the gap in the spectrum is closed ($\lambda_R > \lambda_I$), two additional zeros appear leading to potential high energy modes similar to bilayer graphene [88, 129]. However, these potential collective modes are damped by interband transitions and no clear signature is seen in the density plot.

We are thus left with the solution corresponding to the genuine 2D plasmonic mode. Its dispersion ω_p can be approximated in the long-wavelength limit by

$$\omega_p^2 = \frac{g_v \alpha_g v_F}{2\epsilon} \sum_{v=\pm 1} \frac{k_{Fv}^2}{\sqrt{k_{Fv}^2 + \lambda_{\pm}^2}} q, \quad (52)$$

where $\lambda_{\pm} = \lambda \pm \lambda_I$ and the Fermi wavenumber $k_{F\pm} = \sqrt{\tilde{\mu}(\tilde{\mu} \mp 2\lambda_R) \pm 2\lambda_R \lambda_I - \lambda_I^2}$ with $\tilde{\mu} = \mu/(\hbar v_F)$.

The numerical solution coincides with the long-wavelength solution for small momenta and then becomes red-shifted. But for two occupied conduction bands, there is an additional Landau-damped region which is due to interband transitions from the two conduction bands and the plasmon mode is disrupted at $q \approx 0.05\tilde{\mu}$. At this ‘pseudo-gap’, the group velocity of the collective excitations formally diverges at the entering and exit points and the spectral weight is eventually transferred from the lower to the upper band as momentum is increased.

The pseudo-gap of the plasmonic mode always emerges for $\lambda_R < 0.5\tilde{\mu}$, since the two bands are occupied independently

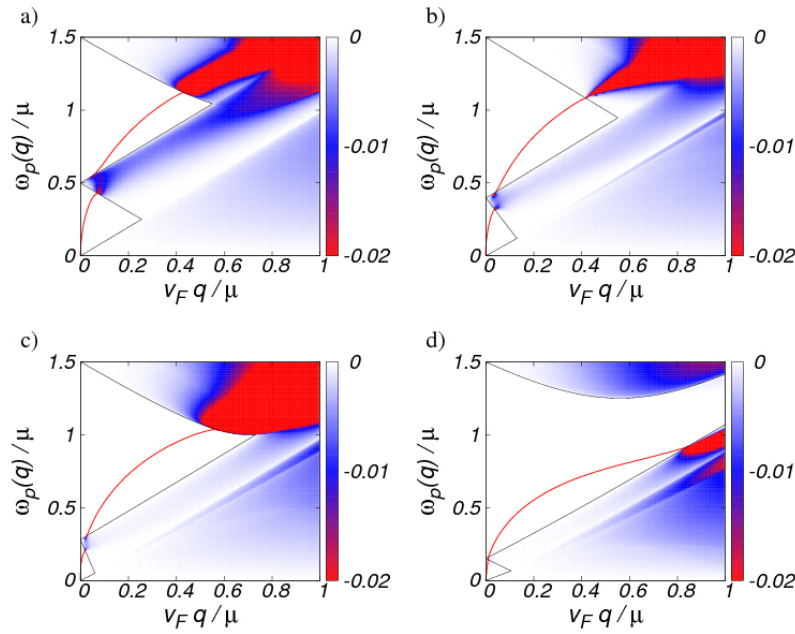


Figure 7. Energy loss function $S = -\text{Im} \epsilon_{\text{RPA}}^{-1}(q, \omega + i0)$ of graphene with intrinsic (λ_I) and Rashba (λ_R) spin-orbit coupling for (a) $(\lambda_R/\mu, \lambda_I/\mu) = (0.25, 0)$, (b) $(0.25, 0.25)$, (c) $(0.25, 0.5)$, and (d) $(0.25, 0.75)$. The straight red lines show the undamped plasmon modes. The black lines indicate the boundaries of the particle-hole continuum.

of the value of λ_I , but it decreases for larger λ_I . This is shown in figure 7, where the energy loss function for several values of $\lambda_{I,R}$ is plotted.

The spin-orbit coupling in graphene is small [169–171], but still there are proposals of how to enhance it [172], which might lead to interesting plasmonic systems when the above pseudo-gap region is reached.

7.2. Plasmons in MoS₂

Around the corners of the Brillouin zone, a monolayer of MoS₂ can be described by an effective two-band model for both spin ($s = \pm 1$) and valley ($\tau = \pm 1$) components. A large energy gap of $\Delta = 1.9$ eV separates the valence and conduction bands and the Hamiltonian for MoS₂ is similar to the one of gapped graphene. The effective model of [77, 78] also contains quadratic terms in the momentum which lead to different electron and hole masses. The plasmon dispersion in the long wavelength limit again displays the typical 2D behavior

$$(\hbar\omega_p)^2 = \frac{e^2}{8\pi\epsilon_0\epsilon} \sum_{\tau,s=\pm 1} k_F^{\tau,s} \left| \frac{\partial E_{\pm}^{\tau,s}}{\partial k} \right|_{k=k_F^{\tau,s}} q, \quad (53)$$

with the energy dispersion $E_{\pm}^{\tau,s}$ for valley τ and spin s given in [168] where the upper (lower) sign stands for the conduction (valence) band.

Due to the electron-hole symmetry in graphene, plasmons in n- and p-doped samples show the same dynamics for equal carrier concentrations. This is no longer true in monolayer MoS₂ as valence and conduction bands differ due to the strong spin-orbit coupling. On the left-hand side of figure 8, the plasmon dispersion and the intraband part of the electron-hole continuum are shown at carrier concentration $n = 10^{12} \text{ cm}^{-2}$

for electron (black) and hole (red) doping. The dotted-dashed lines resemble the long-wavelength result of equation (53) and are in good agreement for $a_0 q \leq 0.05$ ($a_0 = 0.184 \text{ nm}$). The plasmon dispersions and the electron-hole continuum for n and p doping clearly differ and is enhanced for larger carrier density $n = 5 \times 10^{13} \text{ cm}^{-2}$ as the difference in the electron and hole masses becomes more important, see right-hand side of figure 8.

Due to the large direct band gap in monolayer MoS₂, collective charge excitations enter the intraband electron hole continuum similar to 2D electron and hole gases with spin-orbit coupling [173–179]. The investigation of plasmonic effects in general transition metal dichalcogenides forms an active research area as they might determine the absorption and screening properties of single- and multilayer systems.

7.3. Plasmons in 3D topological insulators

Typical 3D topological insulators (TI) like Bi₂Se₃ or Bi₂Te₃ are layered materials with repeating unit cells of hexagonal structure consisting of five layers. Due to the strong spin-orbit coupling, they display protected surface states that are characterized by a single Dirac cone whereas the bulk states show a full insulating gap. Dirac carriers at the surface of a TI remind one of graphene, but in graphene, it is momentum and pseudo-spin that are constrained, whereas it is momentum and real spin which are locked in the case of these topologically protected edge states [180, 181]. The collective modes of this ‘helical metal’ were first discussed in [182] focusing on the curious fact that density fluctuations induce transverse spin fluctuations and vice versa. Spin-plasmons were also discussed in terms of the plasmon wavefunction [183].

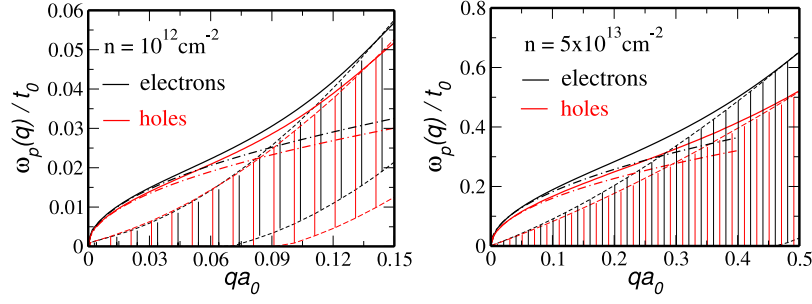


Figure 8. Plasmon dispersion of MoS₂ for electron (black full line) and hole (red full line) concentration of $n = 10^{12} \text{ cm}^{-2}$ (left) and $n = 5 \times 10^{13} \text{ cm}^{-2}$ (right) in units of $t_0 = 1.68 \text{ eV}$ and $a_0 = 0.184 \text{ nm}$. The dashed lines show the boundaries of the electron–hole continuum. The dotted-dashed lines are the long-wavelength results of equation (53).

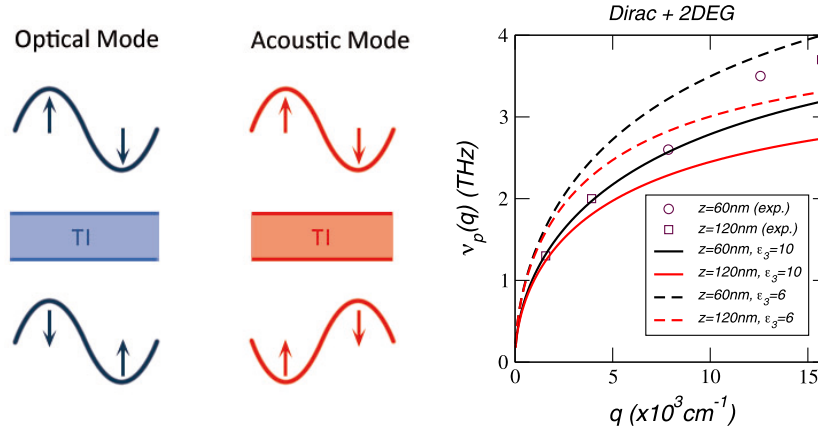


Figure 9. Left: schematic picture of the spin–charge separation. For the optical mode, the charge waves are in-phase and the spin waves are in different directions for the top and bottom layers. This leads to effective (pure) charge oscillations. For the acoustic mode, the charge modulations are out-of-phase and the spin waves point in the same direction for the top and bottom layers. This leads to effective (pure) spin oscillations. Right: comparison of the experimental data of [167] (symbols) with the optical mode of equation (41) considering the response of Dirac fermions and 2DEG for two slab widths $z = 60 \text{ nm}$ (black) and $z = 120 \text{ nm}$ (red) for $\epsilon_3 = 10$ (full lines) and $\epsilon_3 = 6$ (dashed lines).

7.3.1. Spin–charge separation. Dirac cones must come in multiples of two and a single Dirac cone on one surface naturally finds its pair on the opposite side. A thin topological insulator slab without wavefunction overlap thus seems to mimic double-layer graphene (see section 5.3), because the charge response of a helical metal is identical to the charge response of graphene apart from a factor 4 (for TI $g_s = g_v = 1$). But the Dirac cone on one TI surface is *not* an identical copy of the Dirac cone on the other surface because the sign of the Fermi velocity must be opposite for the two Dirac cones. This means that the spin locked to the charge momentum is polarized in opposite directions on the two surfaces which has the curious consequence that in-phase and out-of-phase oscillations can be purely charge- and spin-like, respectively, see left-hand side of figure 9 [184].

Let us now consider the electronic motion confined by a quasi-1D nanowire and assume that the ribbon with width a is wide enough to justify the local approximation of the response function. Thus, only the in-plane Coulomb interaction needs to be modified, see the appendix. The optical mode is then obtained as

$$\omega^2 = (\alpha_d v_F^2 k_F a / \pi) q^2 \ln(\sqrt{e}/(qa)). \quad (54)$$

Considering the logarithmic correction only as an additional factor, the optical mode also shows a linear dispersion.

Spin–charge separation and collective excitations with linear dispersion are the characteristics of the Tomonaga–Luttinger phenomenology for one-dimensional electron systems. Choosing the width and the length of the ribbon as $a \approx 100 \text{ nm}$ and $L \approx 10 \mu\text{m}$, we obtain $v_c \approx 10 v_F$ in the long-wavelength limit. For the sound (spin) velocity we set $v_s \approx v_F$, valid for small TI slab widths. This would correspond to 1D interacting electrons with Luttinger liquid parameter $K \approx 0.1$. By varying the ribbon width to $a \approx 10 \text{ nm}$, one can reach $K \approx 0.2$, and thus be able to tune the effective interaction. A helical Luttinger liquid in topological insulator nanowires was recently discussed in [185] and offers a microscopic theory for the observed spectrum.

7.3.2. Comparison with experiment. Optical plasmon excitations have recently been detected using infrared spectroscopy as discussed in the Introduction [167]. In order to explain their data based on a double-layer model, one not only needs to consider the charge response of the Dirac fermions giving rise to a 4×4 response matrix including spin and charge channels, but also the 2DEG trapped underneath the TI surface, resulting

in a 8×8 response matrix [186]. Taking the depletion layer into account might thus change the number and behavior of plasmon modes, found in typical double-layer structures.

Nevertheless, due to the closeness of the depletion layer to the TI surface, the only newly emerging modes would be charge-less acoustic-like excitations formed by superpositions of the Dirac carriers and the depletion layer on the same (top or bottom) TI surface. These modes are thus closely pinned to the particle–hole continuum and not observable. They also do not affect the modes obtained by the initial 4×4 matrix which can further be reduced to a 2×2 matrix since only the charge channels are coupled. We can, therefore, set the effective response as $\chi = \chi^{\text{Dirac}} + \chi^{\text{2DEG}}$ [184].

The full density response of a 2DEG was derived by Stern [187], but here the local approximation is sufficient, i.e. equation (6) with $v = g_s = 2$ and $g_v = 1$. We can thus use equation (41) with $\mu \rightarrow \mu^{\text{Dirac}} + 4\mu^{\text{2DEG}}$, and with $\mu^{\text{Dirac}} = 542$ meV and $\mu^{\text{2DEG}} = 60$ meV [186], we obtain a reasonable fit to the experimental data for low wave numbers $q \lesssim 10^4$ cm $^{-1}$. This can be seen on the right-hand side of figure 9, where we plot the resonant plasmon frequencies ν_p for slab widths $z = 60$ nm (black) and $z = 120$ nm (red) for a dielectric substrate with $\epsilon_3 = 10$ (full lines). We further assumed $\epsilon_1 = 1$ and $\epsilon_2 = 100$.

The two high-energy plasmon resonances with $q > 10^4$ cm $^{-1}$ cannot be well described by our fit and are blue-shifted. This is in contrast with our expectations because the dipole–dipole interaction between the patterned nanowires should lead to an additional red-shift compared to the analytic curves for samples with small periodicities [116, 124]; and this shift can be as large as 20% [113]. A possible blue-shift could be provided by including the frequency dependence of ϵ_B which might lead to smaller values $\epsilon_B(\omega) < 10$. We, therefore, also show curves with $\epsilon_B = 6$ (dashed lines) which value was measured for thin (15 nm) Al₂O₃-films [188]. Also a decrease of ϵ_{TI} would lead to a blue-shift for larger frequencies and further studies are needed to reconcile theory with experiment in this regime.

8. Summary and outlook

In this topical review, we have presented and discussed various aspect related to plasmonic excitations in graphene nanostructures and other Dirac systems. Our discussion was based on linear response theory and the RPA as well as on hydrodynamic approaches. The typical square-root dispersion of 2D electronic systems defined the plasmonic spectrum in most cases as suggested from phenomenological models. Still, retardation effects, strong screening and spin–orbit coupling can lead to a linear or disrupted spectrum. Also interband plasmons disperse linearly, even though these charge excitations often do not represent genuine plasmons defined by $\epsilon_{\text{RPA}} = 0$, but are only manifested by a peak in the energy loss function.

We also discussed plasmons in graphene-based heterostructures with inhomogeneous dielectric background. For intrinsic dissipation, we introduced the generalized loss function necessary to define the plasmonic multi-layer resonances. For double-layer structures, general analytical formulas in the

long-wavelength limit were presented, especially important in the context of 3D topological insulators. Also retardation effects were discussed leading to enhanced absorption and quenched Fabry–Pérot resonances, as well as to transverse plasmons even in undoped, gapped 2D and 1D materials.

Several aspects related to plasmonics were not discussed, such as quantum effects for plasmons in quantum junctions of graphene dimers [189], boundary effects giving rise to Mie resonances in conducting nanoparticles [190], and fluorescent quenching [191, 192]. For a recent review focusing on the wide-range potential applications of graphene with respect to plasmonic metamaterials, light harvesting, THz technology, biotechnology and medical sciences, see [193].

Another uncovered topic was the nonlinear response of graphene which is ten times larger than that of noble metals like gold [194]. This property can lead to an effective metamaterial with negative refractive index [195] or support the propagation of sub-wavelength optical solitons [196]. Nonlinear plasmonics based on graphene thus promises to become an important research field in the future [197].

Acknowledgments

We thank Nuno Peres, Paco Guinea, Luis Brey, Ángel Gutiérrez, Rafa Roldán, John Schliemann, Andreas Scholz, Reza Asgari, F J Gracia Vidal, Javier Gracia de Abajo and especially Guillermo Gómez-Santos for helpful discussions. This work has been supported by FCT under grant PTDC/FIS/101434/2008 and MICINN under grant no. FIS2010-21883-C02-02.

Appendix. Linear response

The linear response of a system to an external perturbation is related to correlation functions via the Kubo formula [33, 198]. For a general Hamiltonian $H = H_0 + \delta H$ and $\delta H = \lambda^\alpha \psi^\alpha$, the response of a quantum field ψ^α shall be defined by $\delta \langle \psi^\alpha \rangle = \chi_{\psi\psi}^{\alpha,\beta} \lambda^\beta$ with

$$\chi_{\psi\psi}^{\alpha,\beta}(\mathbf{q}, \omega) = -\frac{i}{\hbar} \int_0^\infty e^{i\omega t} \langle [\psi_{\mathbf{q}}^\alpha(t), \psi_{-\mathbf{q}}^\beta(0)] \rangle, \quad (55)$$

where summation over repeated indices is implied. Depending on the context, we will call the response function the retarded Green function or propagator of ψ^α . In the following, we will discuss the density, current and gauge field response. We will then point out the relation between the current and photon propagator.

A.1. Density response

We first discuss the density–density correlation function or polarizability, i.e. $\psi^\alpha \rightarrow \rho$ is the electronic density operator and $\lambda^\alpha \rightarrow \phi$ the electrostatic potential. In 3D, this is also called the Lindhard function. For a 2D electron gas (2DEG), $\chi_{\rho\rho}$ was first calculated by Stern [187]. For a 2D Dirac system, it was first discussed by Shung in the context of intercalated graphite [199].

The 2D polarizability is sometimes defined including a minus sign with respect to the convention in 3D, following the original work of Stern. Here, we will use the definition of equation (55), i.e. we have $\text{Im } \chi_{\rho\rho} < 0$ for $\omega > 0$, and the RPA is then defined as usual with a relative minus sign, see equation (15).

The density–density correlator or Lindhard function for the tight-binding model on a 2D (honeycomb) lattice is given by

$$\chi_{\rho\rho}(\mathbf{q}, \omega) = \frac{g_s}{(2\pi)^2} \int_{1.BZ} d^2k \sum_{s,s'=\pm} f_{s,s'}(\mathbf{k}, \mathbf{q}) \times \frac{n_F(E^s(\mathbf{k})) - n_F(E^{s'}(\mathbf{k} + \mathbf{q}))}{E^s(\mathbf{k}) - E^{s'}(\mathbf{k} + \mathbf{q}) + \hbar\omega + i0}, \quad (56)$$

with the eigenenergies $E^\pm(\mathbf{k}) = \pm t|\phi(\mathbf{k})|$ ($t \approx 2.78$ eV is the hopping amplitude), $n_F(E) = (e^{\beta(E-\mu)} + 1)^{-1}$ the Fermi function, $g_s = 2$ the spin degeneracy and $\phi(\mathbf{k}) = \sum_{\delta_i} e^{i\delta_i \cdot \mathbf{k}}$ the complex structure factor, with δ_i the three nearest neighbor vectors of the hexagonal tight-binding model [65]. Due to the two gapless bands, the above expression contains the band-overlap function

$$f_\pm(\mathbf{k}, \mathbf{q}) = \frac{1}{2} \left(1 \pm \text{Re} \left[\frac{\phi(\mathbf{k}) \phi^*(\mathbf{k} + \mathbf{q})}{|\phi(\mathbf{k})| |\phi(\mathbf{k} + \mathbf{q})|} \right] \right), \quad (57)$$

not present in the one-band 2DEG discussed by Stern. In the Dirac-cone approximation, the above integral can be solved analytically in terms of two analytic functions [51]. We present this solution in the context of the longitudinal current response, see equations (69)–(70).

To discuss the plasmonic dispersion, the local approximation ($q \rightarrow 0$) is frequently used which is given by [51, 52]

$$\chi_{\rho\rho}(\omega) = \frac{g_s g_v q^2}{8\pi \hbar \omega} \left[\frac{2\mu}{\hbar\omega} + \frac{1}{2} \ln \left| \frac{2\mu - \hbar\omega}{2\mu + \hbar\omega} \right| - i \frac{\pi}{2} \Theta(\hbar\omega - 2\mu) \right], \quad (58)$$

with $g_v = 2$ the valley degeneracy. From this expression, we obtain the universal conductivity of undoped graphene $\sigma_0 = \frac{\pi}{2} \frac{e^2}{h}$ via the continuity equation $\sigma = ie^2 \chi_{\rho\rho} \omega / q^2$.

Let us give a brief reminder on this universal conductivity. The local conductivity of graphene at zero temperature is not uniquely defined and can lead to two distinct universal expressions, i.e. both do not depend on any material constants, as was already noted in 1994 [200]. Its value depends on whether the artificially introduced phenomenological damping term $\gamma \rightarrow 0$ goes to zero before or after the (finite) frequency $\omega \rightarrow 0$ and reflects in some way the duality of graphene being a (semi)metal with zero density of state or a semiconductor with zero band-gap. This ambiguity is not a mathematical artifact, but can be interpreted physically whether or not metallic leads give rise to a finite broadening in the dc limit. In the case of transport measurements (first $\omega \rightarrow 0$, then $\gamma \rightarrow 0$), the universal conductivity $\sigma_{dc} = \frac{4}{\pi} \frac{e^2}{h}$ is thus observed [201], whereas in optical experiments (first $\gamma \rightarrow 0$, then $\omega \rightarrow 0$) $\sigma_0 = \frac{\pi}{2} \frac{e^2}{h}$ is seen [154, 155]. Clearly, it is the latter order

of limits which is relevant in the context of plasmonics and we will always assume this universal value to be taken.

The local charge response of equation (58) contains intra- as well as interband contributions. To discuss longitudinal plasmons, mainly intraband transitions need to be considered. In the local approximation, the band-overlap goes to one, $f_\pm \rightarrow 1$, and we can approximate for general isotropic dispersion

$$\chi_{\rho\rho} = \frac{g_s g_v}{(2\pi)^2} \int d^2k \frac{n_F(E(\mathbf{k})) - n_F(E(\mathbf{k} + \mathbf{q}))}{\hbar\omega + E(\mathbf{k}) - E(\mathbf{k} + \mathbf{q})}. \quad (59)$$

In the limit $q \rightarrow 0$, this becomes

$$\chi_{\rho\rho} = \frac{g_s g_v}{(2\pi)^2} \int d^2k \left(-\frac{\partial n_F(E(\mathbf{k}))}{\partial E(\mathbf{k})} \right) \left(\frac{\nabla E(\mathbf{k}) \cdot \mathbf{q}}{\hbar\omega} \right)^2, \quad (60)$$

which can be evaluated for low temperatures, leading to equation (6).

A.2. Current response

We now discuss the general current–current correlator. By this, we can treat longitudinal and transverse response on the same footing. We will thus use $\delta H = -q_e j^i A^i$ with $q_e = -e$ the electron charge ($e > 0$) and $\psi^\alpha \rightarrow j^i$, $\lambda^\alpha \rightarrow q_e A^i$.¹

The current operator for the full tight-binding model needs to be defined with care because a simple Peierls substitution on the lattice breaks gauge invariance. Within an adequate continuum model, we obtain the general expression for the paramagnetic current–current correlation function [202]

$$\chi^{P;i,j}(\mathbf{q}, \omega) = \left(\frac{te}{\hbar} \right)^2 \frac{g_s}{(2\pi)^2} \int_{1.BZ} d^2k \sum_{s,s'=\pm} f_{s,s'}^{i,j}(\mathbf{k}, \mathbf{q}) \times \frac{n_F(E^s(\mathbf{k})) - n_F(E^{s'}(\mathbf{k} + \mathbf{q}))}{E^s(\mathbf{k}) - E^{s'}(\mathbf{k} + \mathbf{q}) + \hbar\omega + i0}, \quad (61)$$

with the same definitions as for the density–density correlation function below equation (56), but the band-overlap is now given by

$$f_\pm^{i,j}(\mathbf{k}, \mathbf{q}) = \frac{1}{2} \left(\text{Re} \left[\tilde{\phi}^i(\mathbf{k}, \mathbf{q}) (\tilde{\phi}^j(\mathbf{k}, \mathbf{q}))^* \right] \pm \text{Re} \left[\tilde{\phi}^i(\mathbf{k}, \mathbf{q}) \tilde{\phi}^j(\mathbf{k}, \mathbf{q}) \frac{\phi^*(\mathbf{k}) \phi^*(\mathbf{k} + \mathbf{q})}{|\phi(\mathbf{k})| |\phi(\mathbf{k} + \mathbf{q})|} \right] \right), \quad (62)$$

$$\tilde{\phi}^i(\mathbf{k}, \mathbf{q}) = \sum_{\delta} \frac{\delta^i}{\mathbf{q} \cdot \delta} \left(e^{i(\mathbf{k} + \mathbf{q}) \cdot \delta} - e^{i\mathbf{k} \cdot \delta} \right). \quad (63)$$

The physical response, $\chi^{i,j}$, also includes the diamagnetic contribution

$$\chi^{i,j}(\mathbf{q}, \omega) = \chi^{P;i,j}(\mathbf{q}, \omega) + \chi_q^{D;i,j}, \quad \chi_q^{D;i,j} = \frac{e^2}{\hbar^2} h_{\text{bond}} \sum_{\delta} \delta^i \delta^j \frac{4}{(\mathbf{q} \cdot \delta)^2} \sin^2 \left(\frac{\mathbf{q} \cdot \delta}{2} \right), \quad (64)$$

where the energy per bond per unit area is given by

$$h_{\text{bond}} = \frac{g_s}{3(2\pi)^2} \int_{1.BZ} d^2k E^+(\mathbf{k}) \sum_{s=\pm} s \cdot n_F(E^{-s}(\mathbf{k})).$$

¹ We use Greek indices α, β for 3D and Latin indices i, j for 2D.

Charge conservation then implies

$$q_i \chi^{i,j}(\mathbf{q}, \omega) q_j = e^2 \omega^2 \chi_{\rho\rho}(\mathbf{q}, \omega). \quad (65)$$

Notice that the anisotropy of the response for finite \mathbf{q} requires the full tensorial structure of $\chi^{i,j}$. The conductivity tensor is defined by $\sigma^{i,j} = i \frac{e^2 \chi^{i,j}}{\omega + i0}$.

The system linearized around the Dirac point is rotationally invariant. We can thus decompose the response tensor $\chi^{i,j}$ into a longitudinal (χ_{jj}^+) and transverse (χ_{jj}^-) scalar component,

$$\begin{aligned} \chi^{i,j}(\mathbf{q}, \omega) &= \frac{q_i q_j}{|\mathbf{q}|^2} \chi_{jj}^+ (|\mathbf{q}|, \omega) \\ &+ \left(\delta_{ij} - \frac{q_i q_j}{|\mathbf{q}|^2} \right) \chi_{jj}^- (|\mathbf{q}|, \omega). \end{aligned} \quad (66)$$

The longitudinal component is thus directly related to the polarizability via equation (65), i.e. $\chi_{jj}^+ = \chi_{\rho\rho} \frac{\omega^2}{q^2}$. The transverse component, χ_{jj}^- , was first calculated in [203].

Following [51], the results can be written in compact form using two dimensionless, complex functions defined as

$$F^\pm(q, \omega) = \frac{g_s g_v}{16\pi} \frac{\hbar \omega}{t} \left[1 - \left(\frac{v_F q}{\omega} \right)^2 \right]^{\mp \frac{1}{2}}, \quad (67)$$

$$G^\pm(x) = x \sqrt{x^2 - 1} \mp \ln \left(x + \sqrt{x^2 - 1} \right), \quad (68)$$

with the Fermi velocity $\hbar v_F = \frac{3}{2} a t$ and carbon-carbon distance $a = 0.142$ nm. We then can write

$$\hbar^2 \chi_{jj}^\pm(q, \omega) = t \left[\tilde{\chi}_0^\pm(q, \omega) + \tilde{\chi}_\mu^\pm(q, \omega) \right], \quad (69)$$

where the dimensionless functions $\tilde{\chi}_0^\pm = -i\pi F^\pm(q, \omega)$ contain the response of the system at half-filling, i.e., genuine interband contributions. $\tilde{\chi}_\mu^\pm$ contains the additional contributions due to the finite chemical potential μ ,

$$\begin{aligned} \tilde{\chi}_\mu^\pm(q, \omega) &= \mp \frac{g_s g_v}{2\pi} \frac{\mu}{t} \frac{\omega^2}{(v_F q)^2} \pm F^\pm(q, \omega) \left\{ G^\pm(x_+) \right. \\ &\quad - \Theta(x_- - 1) [G^\pm(x_-) \mp i\pi] \\ &\quad \left. - \Theta(1 - x_-) G^\pm(-x_-) \right\}, \end{aligned} \quad (70)$$

where we defined $x_\pm = \frac{2\mu \pm \hbar \omega}{\hbar v_F q}$.

Writing equation (70) in real and imaginary expressions, the results are divided into six different zones, three above and three below the Dirac cone dispersion $\omega = v_F q$, see figure A.1. The imaginary part is zero in the white areas and the left triangle is the Pauli-protected region where long-lived plasmons exist.

A.3. Photon propagator

The retarded photon Greens function $\mathcal{D}^{\alpha\beta}$ for the gauge field is defined by $\psi^\alpha \rightarrow A^\alpha$ and $\lambda^\alpha \rightarrow q_e j^\alpha$. Within the Weyl gauge, i.e. setting the scalar potential to zero, $\phi = 0$, it reads as [55]

$$\varepsilon_0 \mathcal{D}^{\alpha\beta}(\mathbf{k}, \omega) = \frac{1}{\omega^2 - c^2 k^2} \left(\delta_{\alpha,\beta} - \frac{k^\alpha k^\beta}{\omega^2/c^2} \right). \quad (71)$$

We will now derive the representations useful for two- and one-dimensional geometries. Part of this discussion can also be found in [35]. In the context of 2D graphene, the main equations were first derived in [204].

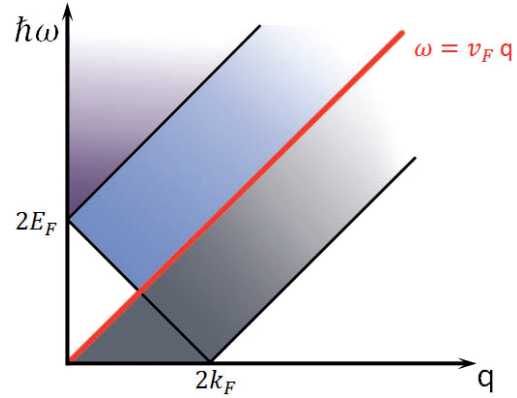


Figure A.1. Particle-hole continuum of pristine graphene. Gray shaded regions correspond to intraband transitions, whereas blue and violet shaded regions indicate interband transitions. The white triangle with $\hbar\omega < 2E_F$ indicates the area where long-lived plasmons may exist.

A.3.1. Two-dimensional geometries. In a layered structure, assumed to be perpendicular to the z axis, the components parallel to the interface, $\mathbf{q} = (q_x, q_y)$, will be preserved as a good quantum number. It is, therefore, convenient to employ the following representation for the Green function in a homogeneous medium:

$$\mathcal{D}^{\alpha\beta}(z, z'; \mathbf{q}, \omega) = \frac{1}{2\pi} \int dk_z e^{ik_z(z-z')} \mathcal{D}^{\alpha\beta}(\mathbf{k}, \omega), \quad (72)$$

with $\mathbf{k} = (\mathbf{q}, k_z)$.

In this representation, the tensor components have a different structure depending on whether $\alpha, \beta = i, j$ with $i, j = x, y$ or $\alpha, \beta = z$. For graphene plasmonics, we are mainly interested in the in-plane components. Decomposed into longitudinal and transverse contributions, we obtain with $q' = \sqrt{q^2 - \epsilon\mu(\omega/c)^2}$,

$$\mathcal{D}^{ij} = \left[\frac{q_i q_j}{q^2} d^+(\mathbf{q}, \omega) + \left(\delta_{ij} - \frac{q_i q_j}{q^2} \right) d^-(\mathbf{q}, \omega) \right] e^{-q'|z-z'|}, \quad (73)$$

where the in-plane longitudinal (d^+) and transverse (d^-) propagators are given by

$$d^+ = \frac{q'}{2\epsilon\epsilon_0\omega^2}, \quad d^- = -\frac{\mu\mu_0}{2q'} = -\frac{\mu}{2\epsilon_0 q' c^2}. \quad (74)$$

Above, we introduced the vacuum permittivity ϵ_0 and permeability μ_0 as well as the relative dielectric constant ϵ and relative permeability μ . Note that the negative transverse propagator can be obtained from the longitudinal one by the substitutions, $\epsilon \rightarrow q'$, $q' \rightarrow \mu$, and $\omega \rightarrow c$. This substitution is completed by $\chi_{jj}^+ \rightarrow -\chi_{jj}^-$ and holds for all in-plane quantities. By considering in-plane components, it thus usually suffices to discuss the longitudinal channel.

A.3.2. One-dimensional geometries. To discuss the propagation in a nanowire, it is convenient to employ the following

representation of the propagator:

$$\mathcal{D}^{\alpha,\beta}(\mathbf{r}-\mathbf{r}', q, \omega) = \frac{1}{(2\pi)^2} \times \int d^2p e^{i\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}')} D^{\alpha,\beta}(\mathbf{k}, \omega), \quad (75)$$

with $\mathbf{k} = (\mathbf{p}, q)$ and $\mathbf{R} = (\mathbf{r}, z)$. With the two-dimensional Green function

$$g(\mathbf{r}, q) = \frac{1}{(2\pi)^2} \int d^2p \frac{e^{i\mathbf{p}\cdot\mathbf{r}}}{p^2 + q^2} = \frac{1}{2\pi} K_0(qr), \quad (76)$$

where K_0 denotes the modified Bessel function of the third kind, the one-dimensional photon propagator is defined by g in the limit $r \rightarrow 0$. Since $K_0(x) \rightarrow -\ln x$, this limit is not well defined and a regularization procedure is needed. If we model the nanowire by a small cylinder with radius a and look for the average of the fields induced by a uniform perturbation, we need to consider

$$\mathcal{D}^{\alpha,\beta}(q, \omega) = \frac{1}{\pi a^2} \int d^2r \int d^2r' \mathcal{D}^{\alpha,\beta}(\mathbf{r}-\mathbf{r}', q, \omega). \quad (77)$$

This yields the following (well-defined) longitudinal ($d_1^+ = \mathcal{D}^{z,z}$) and transverse ($d_1^- = \mathcal{D}^{x,x} = \mathcal{D}^{y,y}$) one-dimensional propagator:

$$\varepsilon_0 d_1^+ = \frac{1}{2\pi} \frac{q'^2}{\omega^2} \ln \frac{\sqrt{e}}{q'a}, \quad (78)$$

$$\varepsilon_0 d_1^- = \frac{1}{2\pi a^2} \left[\frac{1}{\omega^2} - \left(\frac{a^2}{c^2} + \frac{(q'a)^2}{2\omega^2} \right) \ln \frac{\sqrt{e}}{q'a} \right], \quad (79)$$

with the retarded wavenumber $q' = \sqrt{q^2 - \omega^2/c^2}$. All other tensor components are zero.

The longitudinal propagator will give rise to plasmons which disperse as $q' \sqrt{-\ln(q'a)}$, [205] whereas the transverse propagator will predominantly give rise to transverse, but charged plasmons [124]. But in the limit $q' \rightarrow 0$, $d_1^- < 0$ and there will be purely light-like transverse plasmons. These are discussed in more detail in section 6.2.2.

A.4. Graphene–light coupling

Within the Dirac cone approximation, the linear response of the gauge field and the current decomposes into a longitudinal and transverse channel. We can thus write the response as

$$\delta A^\pm = e \chi_{jj}^\pm \delta j^\pm, \quad \delta j^\pm = e d^\pm \delta A^\pm, \quad (80)$$

with the photon propagator d^\pm and the current propagator χ_{jj}^\pm , respectively. Inserting one equation into the other, we obtain self-sustained oscillations for

$$1 - e^2 d^\pm \chi_{jj}^\pm = 0. \quad (81)$$

Within the Dirac cone approximation, longitudinal and transverse plasmonic excitations are thus decoupled.

The product of $e^2 d^\pm$ and χ_{jj}^\pm is dimensionless. The bosonic response function $e^2 d^\pm$ should therefore be related to

the inverse of the fermionic response function χ_{jj}^\pm . In the case of zero chemical potential and for the photonic propagator in vacuum, this relation becomes particularly clear. From the previous equations, we have

$$\chi_{jj}^\pm = \mp \frac{g_s g_v \omega}{16\hbar} \left[\left(\frac{v_F q}{\omega} \right)^2 - 1 \right]^{\mp 1/2}, \quad (82)$$

$$e^2 d^\pm = \pm \frac{2\pi\alpha\hbar}{\omega} \left[\left(\frac{cq}{\omega} \right)^2 - 1 \right]^{\pm 1/2}, \quad (83)$$

with α the fine-structure constant. We thus obtain a dualism between massless bosons and massless (Dirac) fermions, $e^2 d^\pm \leftrightarrow 1/\chi_{jj}^\pm$, by interchanging $c \leftrightarrow v_F$ and $2\pi\alpha \leftrightarrow -16/(g_s g_v)$. For a homogeneous dielectric medium with general ϵ and μ , the mapping is slightly different for longitudinal or transverse channel.

We finally comment on the case of propagating light $\omega < cq$ since taking the right branch cut requires some care. For $\omega \rightarrow \omega + i0$, we must have $\text{Im } d < 0$ following our convention. This yields

$$e^2 d^\pm = -i \frac{2\pi\alpha\hbar}{\omega} \left[1 - \left(\frac{cq}{\omega} \right)^2 \right]^{\pm 1/2}. \quad (84)$$

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 *Science* **306** 666
- [2] Novoselov K S, Jiang D, Schedin F, Booth T J, Khotkevich V V, Morozov S V and Geim A K 2005 *Proc. Natl Acad. Sci. USA* **102** 10451
- [3] Blake P, Hill E W, Neto A H C, Novoselov K S, Jiang D, Yang R, Booth T J and Geim A K 2007 *Appl. Phys. Lett.* **91** 063124
- [4] Castro Neto A H, Guinea F, Peres N M R, Novoselov K S and Geim A K 2009 *Rev. Mod. Phys.* **81** 109
- [5] Bonaccorso F, Sun Z, Hasan T and Ferrari A C 2010 *Nature Photon.* **4** 611
- [6] Avouris P 2010 *Nano Lett.* **10** 4285
- [7] Koppens F H L, Chang D E and García de Abajo F J 2011 *Nano Lett.* **11** 3370
- [8] Rana F 2008 *IEEE Trans. Nanotechnol.* **7** 91
- [9] Jablan M, Buljan H and Soljačić M 2009 *Phys. Rev. B* **80** 245435
- [10] Grigorenko A N, Polini M and Novoselov K S 2012 *Nature Photon.* **6** 749
- [11] Ebbesen T W, Lezec H J, Ghaemi H F, Thio T and Wolff P A 1998 *Nature* **391** 667
- [12] Garcia-Vidal F J, Martin-Moreno L, Ebbesen T W and Kuipers L 2010 *Rev. Mod. Phys.* **82** 729
- [13] Brongersma M L and Shalaei V M 2010 *Science* **328** 440
- [14] Stockman M I 2011 *Opt. Express* **19** 22029
- [15] Maier S A 2007 *Plasmonics: Fundamentals and Applications* (New York: Springer)
- [16] Klimov V V 2012 *Nanoplasmonics: Fundamentals and Applications* (Singapore: Pan Stanford Publ.)
- [17] Maier S A 2006 *Opt. Express* **14** 1957
- [18] Ozbay E 2006 *Science* **311** 189
- [19] Veselago V G 1968 *Sov. Phys.—Usp.* **10** 509
- [20] Pendry J B 2000 *Phys. Rev. Lett.* **85** 3966

- [21] Engheta N and Ziolkowski R 2006 *Metamaterials: Physics and Engineering Explorations* (Hoboken, NJ: Wiley-IEEE Press)
- [22] Zheludev N I and Kivshar Y S 2010 *Nature Mater.* **11** 917
- [23] Otto A 1968 *Z. Phys.* **216** 398
- [24] Kretschmann E 1971 *Z. Phys.* **241** 313
- [25] Liu Y, Willis R F, Emtsev K V and Seyller T 2008 *Phys. Rev. B* **78** 201403
- [26] Eberlein T, Bangert U, Nair R R, Jones R, Gass M, Bleloch A L, Novoselov K S, Geim A and Briddon P R 2008 *Phys. Rev. B* **77** 233406
- [27] Ju L *et al* 2011 *Nature Nanotechnol.* **6** 630
- [28] Yan H, Low T, Zhu W, Wu Y, Freitag M, Li X, Guinea F, Avouris P and Xia F 2013 *Nature Photon.* **7** 394
- [29] Yan H, Li X, Chandra B, Tulevski G, Wu Y, Freitag M, Zhu W, Avouris P and Xia F 2012 *Nature Nanotechnol.* **7** 330
- [30] Chen J *et al* 2012 *Nature* **487** 77
- [31] Fei Z *et al* 2012 *Nature* **487** 82
- [32] Fei Z *et al* 2011 *Nano Lett.* **11** 4701
- [33] Giuliani G and Vignale G 2005 *Quantum Theory of the Electron Liquid* (Cambridge: Cambridge University Press)
- [34] Das Sarma S and Hwang E H 2009 *Phys. Rev. Lett.* **102** 206412
- [35] Novotny L and Hecht B 2006 *Principles of Nano-Optics* (Cambridge: Cambridge University Press)
- [36] Ashcroft N W and Mermin N D 1976 *Solid State Phys.* (Philadelphia, PA: Saunders)
- [37] Johnson P B and Christy R W 1972 *Phys. Rev. B* **6** 4370
- [38] Adam S, Hwang E H, Galitski V M and Das Sarma S 2007 *Proc. Natl Acad. Sci. USA* **104** 18392
- [39] Stauber T, Peres N M R and Guinea F 2007 *Phys. Rev. B* **76** 205423
- [40] Castro E V, Ochoa H, Katsnelson M I, Gorbachev R V, Elias D C, Novoselov K S, Geim A K and Guinea F 2010 *Phys. Rev. Lett.* **105** 266601
- [41] Principi A, Vignale G, Carrega M and Polini M 2013 *Phys. Rev. B* **88** 121405
- [42] Ryzhii V, Satou A and Otsuji T 2007 *J. Appl. Phys.* **101** 024509
- [43] Roldán R, Fuchs J-N and Goerbig M 2013 *Solid State Commun.* **175/176** 114
- [44] Lovat G, Hanson G W, Araneo R and Burghignoli P 2013 *Phys. Rev. B* **87** 115429
- [45] Gangadharaiah S, Farid A M and Mishchenko E G 2008 *Phys. Rev. Lett.* **100** 166802
- [46] Tomadin A and Polini M 2013 *Phys. Rev. B* **88** 205426
- [47] Dyakonov M and Shur M 1993 *Phys. Rev. Lett.* **71** 2465
- [48] Vicarelli L, Vitiello M S, Coquillat D, Lombardo A, Ferrari A C, Knap W, Polini M, Pellegrini V and Tredicucci A 2012 *Nature Mater.* **11** 865
- [49] Fetter A L and Walecka J D 2003 *Quantum Theory of Many-Particle Systems* (New York: Dover)
- [50] Mahan G D 2000 *Many-Particle Physics* (New York: Plenum)
- [51] Wunsch B, Stauber T, Sols F and Guinea F 2006 *New J. Phys.* **8** 318
- [52] Hwang E H and Das Sarma S 2007 *Phys. Rev. B* **75** 205418
- [53] Vafeek O 2006 *Phys. Rev. Lett.* **97** 266406
- [54] Falkovsky L A and Varlamov A A 2007 *Eur. Phys. J. B* **56** 281
- [55] Stauber T and Gómez-Santos G 2012a *New J. Phys.* **14** 105018
- [56] Ramezanali M R, Vazifeh M M, Asgari R, Polini M and MacDonald A H 2009 *J. Phys. A: Math. Theor.* **42** 214015
- [57] Gutiérrez-Rubio A, Stauber T and Guinea F 2013 *J. Opt.* **15** 114005
- [58] Ando T, Zheng Y and Suzuura H 2002 *J. Phys. Soc. Japan* **71** 1318
- [59] Peres N M R, Guinea F and Castro Neto A H 2006 *Phys. Rev. B* **73** 125411
- [60] Gusynin V P and Sharapov S G 2006 *Phys. Rev. B* **73** 245411
- [61] Gusynin V P, Sharapov S G and Carbotte J P 2007 *J. Phys.: Condens. Matter* **19** 026222
- [62] Peres N M R and Stauber T 2008 *Int. J. Mod. Phys. B* **22** 2529
- [63] González J, Guinea F and Vozmediano M 1994 *Nucl. Phys. B* **424** 595
- [64] Busl M, Platero G and Jauho A-P 2012 *Phys. Rev. B* **85** 155449
- [65] Stauber T, Schliemann J and Peres N M R 2010 *Phys. Rev. B* **81** 085409
- [66] Stauber T 2010 *Phys. Rev. B* **82** 201404
- [67] Pisarra M, Sindona A, Riccardi P, Silkin V M and Pitarke J M 2013 arXiv:1306.6273
- [68] Khrapach I, Withers F, Bointon T H, Polyushkin D K, Barnes W L, Russo S and Craciun M F 2012 *Adv. Mater.* **24** 2844
- [69] Hill A, Mikhailov S A and Ziegler K 2009 *Europhys. Lett.* **87** 27005
- [70] Tudorovskiy T and Mikhailov S A 2010 *Phys. Rev. B* **82** 073411
- [71] Principi A, Asgari R and Polini M 2011 *Solid State Commun.* **151** 1627
- [72] Stauber T and Gómez-Santos G 2012 *Phys. Rev. B* **85** 075410
- [73] Profumo R E V, Asgari R, Polini M and MacDonald A H 2012 *Phys. Rev. B* **85** 085443
- [74] Horing N J M 2009 *Phys. Rev. B* **80** 193401
- [75] Hunt B *et al* 2013 *Science* **340** 1427
- [76] Rusponi S, Papagno M, Moras P, Vlaic S, Etzkorn M, Sheverdyaeva P M, Pacilè D, Brune H and Carbone C 2010 *Phys. Rev. Lett.* **105** 246803
- [77] Rostami H, Moghaddam A G and Asgari R 2013 *Phys. Rev. B* **88** 085440
- [78] Kormányos A, Zólyomi V, Drummond N D, Rakyta P, Burkard G and Fal'ko V I 2013 *Phys. Rev. B* **88** 045416
- [79] Scholz A and Schliemann J 2011 *Phys. Rev. B* **83** 235409
- [80] Pyatkovskiy P K 2009 *J. Phys.: Condens. Matter* **21** 025506
- [81] Qaiumzadeh A and Asgari R 2009 *Phys. Rev. B* **79** 075414
- [82] Yuan S, Jin F, Roldán R, Jauho A-P and Katsnelson M I 2013 *Phys. Rev. B* **88** 195401
- [83] Langer T, Baringhaus J, Pfnür H, Schumacher H W and Tegenkamp C 2010 *New J. Phys.* **12** 033017
- [84] Tegenkamp C, Pfnür H, Langer T, Baringhaus J and Schumacher H W 2011 *J. Phys.: Condens. Matter* **23** 012001
- [85] Politano A, Marino A R, Formoso V, Farías D, Miranda R and Chiarello G 2011 *Phys. Rev. B* **84** 033401
- [86] Langer T, Förster D F, Busse C, Michely T, Pfnür H and Tegenkamp C 2011 *New J. Phys.* **13** 053006
- [87] Kramberger C *et al* 2008 *Phys. Rev. Lett.* **100** 196803
- [88] Yuan S, Roldán R and Katsnelson M I 2011 *Phys. Rev. B* **84** 035439
- [89] Iyengar A, Wang J, Fertig H A and Brey L 2007 *Phys. Rev. B* **75** 125430
- [90] Shizuya K 2007 *Phys. Rev. B* **75** 245417
- [91] Roldán R, Fuchs J-N and Goerbig M O 2009 *Phys. Rev. B* **80** 085408
- [92] Crassee I, Orlita M, Potemski M, Walter A L, Ostler M, Seyller T, Gaponenko I, Chen J and Kuzmenko A B 2012 *Nano Lett.* **12** 2470

- [93] Yan H, Li Z, Li X, Zhu W, Avouris P and Xia F 2012b *Nano Lett.* **12** 3766
- [94] Poumirol J M *et al* 2013 *Phys. Rev. Lett.* **110** 246803
- [95] Tymchenko M, Nikitin A Y and Martín-Moreno L 2013 *ACS Nano* **7** 9780
- [96] Guinea F, Katsnelson M I and Geim A K 2010 *Nature Phys.* **6** 30
- [97] Pellegrino F M D, Angilella G G N and Pucci R 2011 *Phys. Rev. B* **84** 195407
- [98] Dugaev V K and Katsnelson M I 2012 *Phys. Rev. B* **86** 115405
- [99] Li Z Q, Henriksen E A, Jiang Z, Hao Z, Martin M C, Kim P, Stormer H L and Basov D N 2008 *Nature Phys.* **4** 532
- [100] Stauber T, Peres N M R and Castro Neto A H 2008 *Phys. Rev. B* **78** 085418
- [101] Peres N M R, Stauber T and Neto A H C 2009 *Europhys. Lett.* **86** 49901
- [102] Peres N M R 2010 *Rev. Mod. Phys.* **82** 2673
- [103] Buljan H, Jablan M and Soljavčić M 2013 *Nature Photon.* **7** 346
- [104] Principi A, Vignale G, Carrega M and Polini M 2013b *Phys. Rev. B* **88** 195405
- [105] Fang Z, Thongrattanasiri S, Schlather A, Liu Z, Ma L, Wang Y, Ajayan P M, Nordlander P, Halas N J and García de Abajo F J 2013 *ACS Nano* **7** 2388
- [106] Mikhailov S A and Beba D 2012 *New J. Phys.* **14** 115024
- [107] Sodemann I and Fogler M M 2012 *Phys. Rev. B* **86** 115408
- [108] Abedinpour S H, Vignale G, Principi A, Polini M, Tse W-K and MacDonald A H 2011 *Phys. Rev. B* **84** 045429
- [109] Polini M, Asgari R, Borghi G, Barlas Y, Pereg-Barnea T and MacDonald A H 2008 *Phys. Rev. B* **77** 081411
- [110] Hwang E H and Das Sarma S 2008 *Phys. Rev. B* **77** 081412
- [111] Bostwick A, Ohta T, Seyller T, Horn K and Rotenberg E 2007 *Nature Phys.* **3** 36
- [112] Bostwick A, Speck F, Seyller T, Horn K, Polini M, Asgari R, MacDonald A H and Rotenberg E 2010 *Science* **328** 999
- [113] Strait J H, Nene P, Chan W-M, Manolatu C, Tiwari S, Rana F, Kevek J W and McEuen P L 2013 *Phys. Rev. B* **87** 241410
- [114] Thongrattanasiri S, Koppens F H L and Garcia de Abajo F J 2012 *Phys. Rev. Lett.* **108** 047401
- [115] Nikitin A Y, Guinea F and Martin-Moreno L 2012 *Appl. Phys. Lett.* **101** 151119
- [116] Nikitin A Y, Guinea F, Garcia-Vidal F J and Martin-Moreno L 2012b *Phys. Rev. B* **85** 081405
- [117] Ferreira A and Peres N M R 2012 *Phys. Rev. B* **86** 205401
- [118] Peres N M R, Ferreira A, Bludov Y V and Vasilevskiy M I 2012 *J. Phys.: Condens. Matter* **24** 245303
- [119] Peres N M R, Bludov Y V, Ferreira A and Vasilevskiy M I 2013 *J. Phys.: Condens. Matter* **25** 125303
- [120] Freitag M, Low T, Zhu W, Yan H, Xia F and Avouris P 2011 *Nature Commun.* **4** 1951
- [121] Bludov Y V, Ferreira A, Peres N M R and Vasilevskiy M I 2013 *Int. J. Mod. Phys. B* **27** 1341001
- [122] Yee K 1966 *IEEE Trans. Antennas Propag.* **14** 302
- [123] jiu Zhao Y, Wu K-L and Cheng K-K 2002 *IEEE Trans. Microw. Theory Tech.* **50** 1844
- [124] Christensen J, Manjavacas A, Thongrattanasiri S, Koppens F H L and García de Abajo F J 2012 *ACS Nano* **6** 431
- [125] McCann E and Fal'ko V I 2006 *Phys. Rev. Lett.* **96** 086805
- [126] Wang X-F and Chakraborty T 2007 *Phys. Rev. B* **75** 041404
- [127] Sensarma R, Hwang E H and Das Sarma S 2010 *Phys. Rev. B* **82** 195428
- [128] Das Sarma S and Li Q 2013 *Phys. Rev. B* **87** 235418
- [129] Gamayun O V 2011 *Phys. Rev. B* **84** 085112
- [130] Scholz A, Stauber T and Schliemann J 2012 *Phys. Rev. B* **86** 195424
- [131] Borghi G, Polini M, Asgari R and MacDonald A H 2009 *Phys. Rev. B* **80** 241402
- [132] Yan H, Low T, Guinea F, Xia F and Avouris P 2013 arXiv:1310.4394
- [133] Low T, Guinea F, Yan H, Xia F and Avouris P 2013 arXiv:1310.4693
- [134] McCann E 2006 *Phys. Rev. B* **74** 161403
- [135] Stauber T, Peres N M R, Guinea F and Castro Neto A H 2007 *Phys. Rev. B* **75** 115425
- [136] Chuang Y-C, Wu J-Y and Lin M-F 2013 *Sci. Rep.* **3** 1038
- [137] Roldán R and Brey L 2013 *Phys. Rev. B* **88** 115420
- [138] Lopes dos Santos J M B, Peres N M R and Castro Neto A H 2007 *Phys. Rev. Lett.* **99** 256802
- [139] Li G, Luican A, Lopes dos Santos J M B, Castro Neto A H, Reina A, Kong J and Andrei E Y 2010 *Nature Phys.* **6** 109
- [140] Moon P and Koshino M 2013 *Phys. Rev. B* **87** 205404
- [141] Stauber T, San-Jose P and Brey L 2013a *New J. Phys.* **15** 113050
- [142] Kim S, Jo I, Nah J, Yao Z, Banerjee S K and Tutuc E 2011 *Phys. Rev. B* **83** 161401
- [143] Liu M, Yin X, Ulin-Avila E, Geng B, Zentgraf T, Ju L, Wang F and Zhang X 2011 *Nature* **474** 64
- [144] Britnell L *et al* 2012 *Science* **335** 947
- [145] Kaipa C S R, Yakovlev A B, Hanson G W, Padooru Y R, Medina F and Mesa F 2012 *Phys. Rev. B* **85** 245407
- [146] Profumo R E V, Polini M, Asgari R, Fazio R and MacDonald A H 2010 *Phys. Rev. B* **82** 085443
- [147] Zhu J-J, Badalyan S M and Peeters F M 2013 *Phys. Rev. B* **87** 085401
- [148] Kotov O, Kol'chenko M and Lozovik Y E 2013 *Opt. Express* **21** 13533
- [149] Hwang E H and Das Sarma S 2009 *Phys. Rev. B* **80** 205405
- [150] Badalyan S M and Peeters F M 2012 *Phys. Rev. B* **85** 195444
- [151] Santoro G E and Giuliani G F 1988 *Phys. Rev. B* **37** 937
- [152] Bao Q, Zhang H, Wang B, Ni Z, Lim C H Y X, Wang Y, Tang D Y and Loh K P 2011 *Nature Photon.* **5** 411
- [153] Gómez-Santos G and Stauber T 2012 *Europhys. Lett.* **99** 27006
- [154] Mak K F, Sfeir M Y, Wu Y, Lui C H, Misewich J A and Heinz T F 2008 *Phys. Rev. Lett.* **101** 196405
- [155] Nair R R, Blake P, Grigorenko A N, Novoselov K S, Booth T J, Stauber T, Peres N M R and Geim A K 2008 *Science* **320** 1308
- [156] Bludov Y V, Vasilevskiy M I and Peres N M R 2010 *Europhys. Lett.* **92** 68001
- [157] Furchi M *et al* 2012 *Nano Lett.* **12** 2773
- [158] Ferreira A, Peres N M R, Ribeiro R M and Stauber T 2012 *Phys. Rev. B* **85** 115438
- [159] Mikhailov S A and Ziegler K 2007 *Phys. Rev. Lett.* **99** 016803
- [160] Jablan M, Buljan H and Soljavčić M 2011 *Opt. Express* **19** 11236
- [161] Gómez-Santos G and Stauber T 2011 *Phys. Rev. B* **84** 165438
- [162] Haldane F D M 1988 *Phys. Rev. Lett.* **61** 2015
- [163] Kane C L and Mele E J 2005 *Phys. Rev. Lett.* **95** 226801
- [164] König M, Wiedmann S, Brüne C, Roth A, Buhmann H, Molenkamp L W, Qi X-L and Zhang S-C 2007 *Science* **318** 766

- [165] Bernevig B A, Hughes T L and Zhang S-C 2006 *Science* **314** 1757
- [166] Juergens S, Michetti P and Trauzettel B 2014 *Phys. Rev. Lett.* at press arXiv:1309.6083
- [167] Di Pietro P *et al* 2013 *Nature Nanotechnol.* **8** 556
- [168] Scholz A, Stauber T and Schliemann J 2013 *Phys. Rev. B* **88** 035135
- [169] Trauzettel B, Bulaev D V, Loss D and Burkard G 2006 *Nature Phys.* **3** 192
- [170] Huertas-Hernando D, Guinea F and Brataas A 2006 *Phys. Rev. B* **74** 155426
- [171] Min H, Hill J E, Sinitsyn N A, Sahu B R, Kleinman L and MacDonald A H 2006 *Phys. Rev. B* **74** 165310
- [172] Castro Neto A H and Guinea F 2009 *Phys. Rev. Lett.* **103** 026804
- [173] Pletyukhov M and Gritsev V 2006 *Phys. Rev. B* **74** 045307
- [174] Pletyukhov M and Korschuh S 2007 *Eur. Phys. J. B* **60** 29
- [175] Badalyan S M, Matos-Abiague A, Vignale G and Fabian J 2009 *Phys. Rev. B* **79** 205305
- [176] Schliemann J 2011 *Phys. Rev. B* **84** 155201
- [177] Kyrychenko F V and Ullrich C A 2011 *Phys. Rev. B* **83** 205206
- [178] Agarwal A, Chesi S, Jungwirth T, Sinova J, Vignale G and Polini M 2011 *Phys. Rev. B* **83** 115135
- [179] Scholz A, Dollinger T, Wenk P, Richter K and Schliemann J 2013 *Phys. Rev. B* **87** 085321
- [180] Hasan M and Kane C 2010 *Rev. Mod. Phys.* **82** 3045
- [181] Qi X-L and Zhang S-C 2011 *Rev. Mod. Phys.* **83** 1057
- [182] Raghu S, Chung S B, Qi X-L and Zhang S-C 2010 *Phys. Rev. Lett.* **104** 116401
- [183] Efimkin D, Lozovik Y and Sokolik A 2012 *Nanoscale Res. Lett.* **7** 163
- [184] Stauber T, Gómez-Santos G and Brey L 2013b *Phys. Rev. B* **88** 205427
- [185] Egger R, Zazunov A and Yeyati A L 2010 *Phys. Rev. Lett.* **105** 136403
- [186] Bansal N, Kim Y S, Brahlek M, Edrey E and Oh S 2012 *Phys. Rev. Lett.* **109** 116804
- [187] Stern F 1967 *Phys. Rev. Lett.* **18** 546
- [188] Kim S, Nah J, Jo I, Shahrjerdi D, Colombo L, Yao Z, Tutuc E and Banerjee S K 2009 *Appl. Phys. Lett.* **94** 062107
- [189] Thongrattanasiri S, Manjavacas A, Nordlander P and de Abajo F J G 2013 *Laser Photon. Rev.* **7** 297
- [190] Zhao Q, Zhou J, Zhang F and Lippens D 2009 *Mater. Today* **12** 60
- [191] Huidobro P A, Nikitin A Y, González-Ballester C, Martín-Moreno L and García-Vidal F J 2012 *Phys. Rev. B* **85** 155438
- [192] Gaudreau L, Tielrooij K J, Prawiroatmodjo G E D K, Osmond J, de Abajo F J G and Koppens F H L 2013 *Nano Lett.* **13** 2030
- [193] Luo X, Qiu T, Lu W and Ni Z 2013 *Mater. Sci. Eng. R* **74** 351
- [194] Hendry E, Hale P J, Moger J, Savchenko A K and Mikhailov S A 2010 *Phys. Rev. Lett.* **105** 097401
- [195] Harutyunyan H, Beams R and Novotny L 2013 *Nature Phys.* **9** 423
- [196] Nesterov M L, Bravo-Abad J, Nikitin A Y, García-Vidal F J and Martín-Moreno L 2013 *Laser Photon. Rev.* **7** L7
- [197] Kauranen M and Zayats A V 2012 *Nature Photon.* **6** 737
- [198] Wen X-G 2004 *Quantum Field Theory of Many-Body Systems* (Oxford: Oxford University Press)
- [199] Shung K W K 1986 *Phys. Rev. B* **34** 979
- [200] Ludwig A W W, Fisher M P A, Shankar R and Grinstein G 1994 *Phys. Rev. B* **50** 7526
- [201] Miao F, Wijeratne S, Zhang Y, Coskun U C, Bao W and Lau C N 2007 *Science* **317** 1530
- [202] Stauber T and Gómez-Santos G 2010 *Phys. Rev. B* **82** 155412
- [203] Principi A, Polini M and Vignale G 2009 *Phys. Rev. B* **80** 075418
- [204] Hanson G W 2008 *J. Appl. Phys.* **103** 064302
- [205] Brey L and Fertig H A 2007 *Phys. Rev. B* **75** 125434