A brief review on optical properties of planar metallic interfaces and films: from classical view to quantum description

Alexey V Krasavin

Department of Physics and London Centre for Nanotechnology, King's College London, Strand, London WC2R 2LS, United Kingdom
E-mail: alexey.krasavin@kcl.ac.uk

Keywords: metal optical properties, nonlocality, Landau damping, ultrathin films

Abstract

With the rapid development of nanophotonics, metals play an ever increasing role in modern optics. At the same time, optical properties of metals are more complex than it is usually assumed in basic models. Such behaviour leads to intriguing phenomena at the nanoscale and also has a profound impact on conventional optical effects. In this brief topical review basic concepts and approaches for the description of an optical response of metals will be overviewed for the case of planar metallic interfaces and films, with an aim to provide a complete and inter-related prospective and also give a general guidance on the level of description needed to treat an optical system at hand.

1. Introduction

The progress in nanotechnology gave rise to a new field of plasmonics, which utilises metallic nanostructures to break the diffraction limit of light and manipulate optical signals at subwavelength scales. The resulting extreme field localisation enabled ultracompact optical devices [1, 2], high-resolution lithography [3] and high-density data storage [4], while in the complementary areas it led to a progress in subwavelength optical microscopy and imaging [5, 6]. The related strong field enhancement opened new horizons in nonlinear and ultrafast optics [7–9], hot electron driven photocatalysis [10], plasmonic-enhanced fluorescence [11] and sensing [6, 12].

The range of plasmonic structures, realising these phenomena and devices, spans from microscale objects and structured films to nanoparticles of just ~1 nm in size. At the same time, optical properties of metal constituting such diverse objects are different for technological and most importantly fundamental physical reasons. Thus, with a substantial knowledge accumulated on this topic, it is very helpful and timely now to summarise it, understand intricate interconnections between the involved theories and trace the evolution from the macroscopic classical description to nanoscale quantum dynamics.

In the following brief review, we overview the development of the theories of an optical response of metals starting from a standard Drude–Lorentz model and through its classical corrections and semi-classical models will arrive to the level of essentially quantum approaches and calculations from first principles. Particularly, without a loss of generality in regard of the involved theoretical descriptions we consider the evolution of optical properties of metallic films, starting with bulk materials and optically-thick slabs and finishing with ultrathin films with a thickness down to an atomic monolayer. From technological point of view the discussion will span from the case of highly polycrystalline samples to single crystals.

The structure of the review is the following. In section 2 optical properties of metals are discussed in a context of a generalised optical response of a linear medium, after which using explicit assumptions a classical Drude–Lorentz model is introduced. Section 3 presents corrections which are possible to make within the Drude approach, particularly related to the change of the scattering rate due to the polycrystalline structure of the metal and/or a finite thickness of metallic films. In the next sections in a step-by-step manner it is presented the development of the theory of the optical response of metallic structures towards more elaborated microscopic description of the electron gas, defining the optical properties of the metal.
Particularly, within a semi-classical description based on Boltzmann formalism, nonlocality of the metal optical response at the interfaces is introduced and discussed in section 4. Section 5 considers various quantum approaches progressively taking into account single-particle excitations (Landau damping), variation of electron density near the interface, electron–electron interactions and exchange-correlation effects in the quantum electron gas. This is followed by the discussion of optical properties of plasmonic 2D materials in section 6. The overviewed theoretical descriptions are put in a context of the state-of-the-art experimental research in section 7. Finally, section 8 presents concluding remarks.

2. From general optical response to Drude–Lorentz model

2.1. General optical response of linear medium in various approximations

As it follows from the theory of electromagnetism, the interaction of electromagnetic waves with a medium is defined by dielectric permittivity and magnetic permeability of the latter. As permeability of natural materials at high frequencies is equal to unity, their permittivity fully defines the optical response. Particularly, it determines the constitutive relation between the electric field and dielectric displacement in the material. Making the only assumption that the medium is linear one can express this relation as [13–16]:

$$\mathbf{D}(\mathbf{r}, t) = \varepsilon_0 \int \hat{\varepsilon}(\mathbf{r}, \mathbf{r}', t - t') \mathbf{E}(\mathbf{r}', t') \, d\mathbf{r}' \, dt',$$

where $\hat{\varepsilon}(\mathbf{r}, \mathbf{r}', t - t')$ is the dielectric permittivity tensor. As one can see, the material response at a given point depends on the value of the electric field not only at this particular point, but also at all other points. In other words, generally the optical response of the medium is nonlocal, opposed to being local when the reaction of the material at a given point would depend only on fields at this point. Equation (1) also takes into account all surface-related phenomena, which in the case of metals include surface-related nonlocal effects already present within a semi-classical description or electron spill-out and quantum-size effects appearing in quantum treatment. The extended discussion of the latter approaches will be given in the further sections of the review. In this section we will discuss optical properties of metals in a view of a classical Drude–Lorentz formalism. On the basis of the general response given by equation (1) it is very convenient to expose all the implicit assumptions which are made in Drude–Lorentz derivation:

(a) The medium is considered to be isotropic in terms of the optical response defined by its crystalline structure.

(b) Surface- and size-related effects in the material response are neglected, at all points of an object the medium is considered to have bulk and homogeneous properties, while for bulk and homogeneous medium the dielectric permittivity $\varepsilon(\mathbf{r} - \mathbf{r}', t - t')$ depends only on the difference between $\mathbf{r}$ and $\mathbf{r}'$, but not their particular values. Performing Fourier transforms over temporal and spatial coordinates in equation (1) under this assumption

$$\mathbf{D}(\mathbf{k}, \omega) = \varepsilon_0 \mathbf{E}(\mathbf{k}, \omega),$$

one can see that $\varepsilon(\mathbf{k}, \omega)$ in this case apart from being a function of frequency is also a function of wavevector, in other words it is spatially dispersive.

(c) The optical response is considered to be local, depending only on the value of the electric field at the considered point of space, thus the constitutive relation is further simplified:

$$\mathbf{D}(\omega) = \varepsilon_0 \varepsilon(\omega) \mathbf{E}(\omega).$$

2.2. Drude–Lorentz formula

After just a few years after the discovery of an electron, Drude proposed a model successfully describing electrical and thermal conductivities of metals, representing their structure as a lattice of positively charged atomic cores, which is populated with a gas of freely moving electrons. Illuminating such system with an electromagnetic wave, and considering its polarisation produced by the free electrons due to their displacement induced by the electric field present at their positions (local approximation) and adding the polarisation response of the lattice ions, one can readily obtain the Drude–Lorentz frequency-dependent expression for dielectric permittivity of metal [17]:

$$\varepsilon(\omega) = \varepsilon_{\text{inter}}(\omega) - \frac{\omega_p^2}{\omega^2 + \gamma \omega},$$
where $\varepsilon_{\text{inter}}(\omega)$ describes the optical behaviour of the atoms, $\omega_p = \sqrt{\frac{ne^2}{\varepsilon_0 m^*}}$ is the plasma frequency, depending on the free electron density $n$ and the effective mass $m^*$ of the electron, ($e$ is the elementary charge), and $\gamma$ is the phenomenological scattering rate. The plasma frequency defines the overall magnitude of the electronic optical response, while the scattering rate mostly determines its absorptive part. Scattering of the electrons happens through various channels and its total rate can be presented using Matthiessen’s rule:

$$\gamma = \gamma_{\text{bulk}} = \gamma_{\text{def}} + \gamma_{\text{e-ph}} + \gamma_{\text{e-e}},$$

where $\gamma_{\text{def}}, \gamma_{\text{e-ph}},$ and $\gamma_{\text{e-e}}$ are the rates corresponding to scattering of electrons on defects/impurities, phonons and other electrons, respectively.

The Drude–Lorentz model has had a tremendous success in describing the optical properties of metals and metallic objects, from the reflectivity of mirrors to a spectrally-resonant response of subwavelength metallic particles. As we will see in the following, with a certain degree of precision it can even be used to describe in a phenomenological way the optical behaviour of truly nanoscale metallic objects with a size of $\sim 10$ nm, taking into account new nanoscale phenomena through the modification of the scattering rate related to the increase of electron scattering on the nano-object surface. But at a length scale of $\sim 1$ nm a purely quantum description of the optical response becomes a must.

3. Classical correction approaches

3.1. Nature and length scale of corrections

Staying within the classical model of a free electron gas introduced by Drude and looking at the Drude–Lorentz formula, one can ask a question: while studying the material optical properties of smaller and smaller objects, what parameter within the model needs to be adjusted first and at which length scale it should happen? It is logical to assume that the change of the atomic response $\varepsilon_{\text{inter}}(\omega)$ happens at a scale of a few inter-atomic distances near the metallic surface, while the change in $\omega_p$, related to the modification of the electron density happens in the region of an order of the Fermi wavelength or screening length [18], both of these distances lying at an angstrom length scale. The adjustment of $\omega_p$ can be done phenomenologically using e.g. a simple function allowing analytical treatment of the problem [19], but calculation of a realistic profile of the ground state electron density near the interface requires quantum approaches and will be discussed below in section 5. Surface-related correction of $\varepsilon_{\text{inter}}(\omega)$ is generally beyond Drude–Lorentz approximation and is treated in advanced jellium-based quantum models [20] or in ab-initio calculations [31]. At the same time, the third parameter $\gamma$ can be treated completely within the classical model. It is related to a mean free path $l = v_F/\gamma$ ($v_F$ is the electron Fermi velocity), the distance the electron spans between scattering events. Thus, if any of the geometrical dimensions of a metallic object approach the mean free path in bulk $l_{\text{bulk}}$, due to collisions of the electrons with the object boundaries the effective mean free path will decrease, and the scattering rate $\gamma$ in the Drude–Lorentz formula needs to be adjusted.

3.2. Polycrystalline materials and thin films

The first calculations on the scattering rate correction in granulated and/or confined metallic structures, which can be utilised to correct the Drude–Lorentz response, were made in the studies of DC resistance of metallic films. This was commonly done by solving Boltzmann equation in the presence of scattering on grains and/or boundaries. As a starting point, we will discuss the corrections related to the granular structure of a bulk polycrystalline metal, related to electron scattering on the grains. The estimation of the scattering rate increase in this case was initially performed in a one-dimensional case [22], and then it was extended to a three-dimensional block-grain structure [23]. The conductivity of the metallic polycrystalline structure $\sigma_{\text{bulk}+\text{gr}} = f_{\text{gr}}(d/\text{bulk})$ is logically expressed in terms of a ratio between the grain size $d$ and the bulk mean free path $l_{\text{bulk}}$. Expectedly, it also depends on an electron reflection coefficient of the boundaries $R$, either calculated quantum-mechanically [22] or introduced phenomenologically [23]. The obtained corrected conductivity can be easily recast into the effective scattering rate $\gamma_{\text{bulk}+\text{gr}} = \sigma_{\text{bulk}}/\sigma_{\text{bulk}+\text{gr}} \cdot \gamma_{\text{bulk}}$. Reiss et al. [24] extended the approach on statistically varied grain sizes, performing quantum mechanical calculations using a transfer-matrix method. An interesting observation was that in the general case the dependence of the conductivity on the grain size is more intricate than being just a function of $d/l_{\text{bulk}}$, but at practically relevant grain sizes larger than 3 nm, this dependence is restored. It was also proposed an easy and practical way to introduce the grain-induced mean free path correction in standard tabulated data for metal permittivity [25].

The first calculations of the mean free path correction in confined systems were done in a seminal work by Fuchs [26], who derived an effective conductivity a metallic film with a thickness comparable to the electron mean free path for the case of purely inelastic (diffuse) electron scattering on the film boundaries.
The resulting integral formula $\sigma_{\text{bulk+film}} = f_{\text{film}} \left( t / b_{\text{bulk}} \right)$ is understandably expressed in terms of a ratio between the film thickness $t$ and the bulk mean free path $b_{\text{bulk}}$. In the further development, Sondheimer extended this approach to the case of partially diffuse scattering, described by a phenomenological parameter $p$, which represents a statistical ratio between elastic and inelastic scattering events and varies from 0 to 1 [27]. The value of 1 corresponds to purely elastic (specular) scattering, while 0 corresponds to purely inelastic (diffuse) scattering. The resulting expression for boundary scattering is proportional to $(1 - p)$. Particularly, this means that in the case of specular scattering there is no extra scattering loss, which is logical as there is no loss of the electron momentum in the direction of the current (in the plane of the film). Lucas extended the Sondheimer’s calculations to the case of two unlike film surfaces [28].

Finally, Mayadas and Shatzkes unified the above approaches, deriving the conductivity of a polycrystalline metallic film for the case of a ‘columnar’ structure, in which the film is divided into grains by two orthogonal sets of parallel planes perpendicular to the film, with spacings obeying the Gaussian distribution and one set of planes parallel to the applied electric field [29]. Thus, the grain columns were presented by blocks protruding throughout the film thickness. The reflection from the boundaries parallel to the applied field was considered to be elastic, while for the perpendicular boundaries it was calculated quantum-mechanically, presenting the boundary as a delta-function potential barrier of a certain strength related to its reflection coefficient. Interestingly, the obtained expression for conductivity (or the scattering rate) cannot be interpreted as a sum of corresponding values for scattering on grains $\gamma_{\text{gr}}$ and on surface $\gamma_{\text{surf}}$ (leading to a deviation from Matthiessen’s rule), following which it was argued that the mean free path for a polycrystalline film cannot be defined, but at the same time the calculated result can be used as its effective value. Further development of this method was done in Ref. [30]. Warkusz extended the analysis to the films with a granulated structure in all three dimensions [31], assuming different phenomenological transmission/reflection coefficients for the three directions and partly-diffusive reflection from the film surfaces [28]. Now let us understand the level of scattering rate corrections which could be inflicted by the polycrystalline structure of metal and the film boundaries. Considering gold as an example of a plasmonic metal and taking its single-crystal scattering rate $\gamma_{\text{bulk}} = 3.8 \times 10^{13}$ s$^{-1}$, one can estimate that the polycrystalline structure with a typical grain size of 30 nm and electron-grain reflection coefficient $R = 0.4$ [32] will result in the scattering rate increase in $\sim 2.2$ times, while for a diffusely-scattering 40 nm thick film the increase will be 1.4 times. Thus, the corrections can be quite essential and should be considered.

For highly granulated films with many grains across the film thickness, which usually happens for thicker $t > 100$ nm films, it is possible to make a more crude approximation incorporating the grain scattering into the bulk scattering rate, which does not, though, assume any (e.g. columnar) grain shape. Then, one can apply the procedure of surface scattering calculation to the obtained effective material conductivity of metallic films $\sigma_{\text{bulk+gr+film}} = f_{\text{film}} \left( t / b_{\text{bulk+gr}} \right)$, from which as was mentioned above the effective scattering rate can be calculated. For the practical application of all above calculations, it is particularly helpful that there is an experimentally observed correlation between the film thickness and the grain size in metallic polycrystalline films [33].

In relation to the prospective for a further development of the classical scattering-related corrections it is instructive to have a closer look at the assumptions made in the implemented models. In calculations of the scattering on the film surfaces the diffuse scattering is considered to be isotropic, furthermore, the parameter $p$, representing the statistical ratio of elastic/inelastic scattering events is independent on the angle of incidence [27]. As for the precise models of scattering on the polycrystalline structure, the grains are presented by potential barriers in the planes perpendicular to the principal axes of the film (slicing the space in three dimensions with grids of orthogonal planes with equal or randomised spacings), with further assumptions on their reflective properties and in some models on the direction of the field [22, 29, 31]. Developing more elaborated models seems to be unreasonable, as it will lead to overcomplication of already bulky expressions containing integrals which cannot be solved analytically, and still will offer only an incremental step towards the representation of real structures. Interestingly, a high amount of disorder makes the situation more straightforward to tackle, as a theoretical description on the basis of an effective medium theory can be applied [34, 35]. In the case of fragmented island films a model built on the interaction of coupled metallic nanoparticles or clusters can be used [36, 37].

4. Semi-classical models of nonlocal optical response

While making a step towards the consideration of the properties of metal at high frequencies, one needs to make an important observation. Generally, the definition of the refractive index of metals in this case is less straightforward than it is commonly assumed. The electromagnetic field penetrates into a metal to a certain distance, namely the skin depth, which for good conductors at room temperature is of the same order of magnitude as the electron mean free path. Therefore, while moving between the collisions, the electrons
experience a nonuniform electric field, or in other words the induced current density and therefore the overall optical response start to depend on the electric field distribution in the neighbouring spatial regions [38]. This leads to a breakdown of a usually assumed local relation that the current at a given point is proportional only to the value of the electric field at this particular point, which is used in the derivation of the standard Drude–Lorentz model. In other words, the optical response becomes nonlocal, in agreement with the general expression given in equation (1). The resulting electromagnetic field in the metal in this case has a non-exponential decay and the common definition of refractive index loses its validity [38]. We note that such nonlocal behaviour is already obtained from quite a straightforward model, particularly by solving Boltzmann equation for electrons treated as classical ideal gas (though with Fermi–Dirac statistics), scattered on an infinite barrier representing the surface and interacting in a self-consistent way with electromagnetic fields produced by the incident wave. Furthermore, it was observed for both pure specular (semi-classical infinite barrier model, SCIB) and pure diffuse scattering, as well as any intermediate case [39]. Another very important consequence of the nonlocal description is the implementation of a continuous change of the electric field component across the interface [18, 40, 41], instead of its discontinuous change due to a delta-function charge layer in the classical description. On the other hand, given the overwhelming evidence of successful application of the local Drude–Lorentz description, it must be meaningful and valid in the vast majority of cases with a reasonable precision. The apparent mismatch between the two concepts and the factors leading to their reconciliation will be discussed below.

A well-known illustration of the breakdown of the standard local permittivity concept is an anomalous skin effect, related to dramatically increased resistivity of metals in a microwave spectral region and at low temperatures compared to the predictions based on the theory of an ordinary skin effect. As for the frequencies as high as optical (\(\omega/\gamma \gg 1\)), in one of the first theoretical descriptions by Reuter and Sondheimer [38] it was made a plausible suggestion that as a criterion for the corrections it is not the mean free path, but an electron displacement during one field oscillation should be compared with the field inhomogeneity length. However, in a following study using a more scrupulous analysis this was put under question [42]. Following these works, an extended theory of the anomalous skin effect was derived for the frequency range from microwave to ultraviolet. Various boundary conditions for electron surface scattering were also considered, from pure specular (\(p = 1\)) to pure diffuse (\(p = 0\)), as well as arbitrary intermediate \(0 \leq p \leq 1\) cases [39, 42–45]. Here, one needs to note that the phenomenological parameter \(p\) in principle can be measured. Particularly, it can be derived analysing the transmission or absorption of thin metallic films [46, 47]. It is important to note, that in the general case the surface scattering and nonlocal effects become entangled and inseparable.

For the normal incidence at near-infrared and lower visible frequencies it was found that although the refractive index is usually introduced in the view of an exponential decay of the fields into the metal (which is incorrect when the mean free path is compared with the penetration depth, as discussed above), it still has a definite meaning through its connection with surface admittance, with standard formulas for reflectance and absorbance keeping their validity [43]. For the oblique incidence the situation becomes more complex even for the case of specular scattering [46, 48–53]. Effective permittivity can still be introduced, but as the consequence that the optical properties are nonlocal now and depend on the field distribution (or in other words on the way the object is illuminated), there will be two dielectric functions, one for \(s\) and another for \(p\) polarisation [49]. In this work, Kliewer and Fuchs first noticed that the description of Reuter–Sondheimer and Dingle [38, 43] for oblique \(p\)-polarised illumination was incorrect due to the disregard of \(\nabla \cdot \mathbf{E} \neq 0\) fields which can produce a charge imbalance penetrating the metal [49]. Interaction of electrons with this charge imbalance can lead to additional absorption losses. As it follows from the results, \(p\)-polarised illumination will excite charge fluctuations if the frequency of the electromagnetic wave is below the plasma frequency \(\omega < \omega_p\) [49], and generate longitudinal bulk plasmons if \(\omega > \omega_p\) [54, 55].

The theory was further extended to the case of thin metallic films [46, 47, 50–52, 56–60], which led to the discovery of size effects in the spectral dependences of absorption. One of the mechanism of such oscillations is related to the increased absorption when the electron scatters on diffusive boundaries after obtaining a maximum energy from an integer number of field oscillations [46, 47, 50]. This leads to a breakdown of a usually assumed local relation that the current at a given point is proportional only to the value of the electric field at this particular point, which is used in the derivation of the standard Drude–Lorentz model. In other words, the optical response becomes nonlocal, in agreement with the general expression given in equation (1). The resulting electromagnetic field in the metal in this case has a non-exponential decay and the common definition of refractive index loses its validity [38]. We note that such nonlocal behaviour is already obtained from quite a straightforward model, particularly by solving Boltzmann equation for electrons treated as classical ideal gas (though with Fermi–Dirac statistics), scattered on an infinite barrier representing the surface and interacting in a self-consistent way with electromagnetic fields produced by the incident wave. Furthermore, it was observed for both pure specular (semi-classical infinite barrier model, SCIB) and pure diffuse scattering, as well as any intermediate case [39]. Another very important consequence of the nonlocal description is the implementation of a continuous change of the electric field component across the interface [18, 40, 41], instead of its discontinuous change due to a delta-function charge layer in the classical description. On the other hand, given the overwhelming evidence of successful application of the local Drude–Lorentz description, it must be meaningful and valid in the vast majority of cases with a reasonable precision. The apparent mismatch between the two concepts and the factors leading to their reconciliation will be discussed below.

A well-known illustration of the breakdown of the standard local permittivity concept is an anomalous skin effect, related to dramatically increased resistivity of metals in a microwave spectral region and at low temperatures compared to the predictions based on the theory of an ordinary skin effect. As for the frequencies as high as optical (\(\omega/\gamma \gg 1\)), in one of the first theoretical descriptions by Reuter and Sondheimer [38] it was made a plausible suggestion that as a criterion for the corrections it is not the mean free path, but an electron displacement during one field oscillation should be compared with the field inhomogeneity length. However, in a following study using a more scrupulous analysis this was put under question [42]. Following these works, an extended theory of the anomalous skin effect was derived for the frequency range from microwave to ultraviolet. Various boundary conditions for electron surface scattering were also considered, from pure specular (\(p = 1\)) to pure diffuse (\(p = 0\)), as well as arbitrary intermediate \(0 \leq p \leq 1\) cases [39, 42–45]. Here, one needs to note that the phenomenological parameter \(p\) in principle can be measured. Particularly, it can be derived analysing the transmission or absorption of thin metallic films [46, 47]. It is important to note, that in the general case the surface scattering and nonlocal effects become entangled and inseparable.

For the normal incidence at near-infrared and lower visible frequencies it was found that although the refractive index is usually introduced in the view of an exponential decay of the fields into the metal (which is incorrect when the mean free path is compared with the penetration depth, as discussed above), it still has a definite meaning through its connection with surface admittance, with standard formulas for reflectance and absorbance keeping their validity [43]. For the oblique incidence the situation becomes more complex even for the case of specular scattering [46, 48–53]. Effective permittivity can still be introduced, but as the consequence that the optical properties are nonlocal now and depend on the field distribution (or in other words on the way the object is illuminated), there will be two dielectric functions, one for \(s\) and another for \(p\) polarisation [49]. In this work, Kliewer and Fuchs first noticed that the description of Reuter–Sondheimer and Dingle [38, 43] for oblique \(p\)-polarised illumination was incorrect due to the disregard of \(\nabla \cdot \mathbf{E} \neq 0\) fields which can produce a charge imbalance penetrating the metal [49]. Interaction of electrons with this charge imbalance can lead to additional absorption losses. As it follows from the results, \(p\)-polarised illumination will excite charge fluctuations if the frequency of the electromagnetic wave is below the plasma frequency \(\omega < \omega_p\) [49], and generate longitudinal bulk plasmons if \(\omega > \omega_p\) [54, 55].

The theory was further extended to the case of thin metallic films [46, 47, 50–52, 56–60], which led to the discovery of size effects in the spectral dependences of absorption. One of the mechanism of such oscillations is related to the increased absorption when the electron scatters on diffusive boundaries after obtaining a maximum energy from an integer number of field oscillations [56, 57]. Another mechanism exists even for elastically scattering boundaries and is related to the resonant excitation of the charge fluctuations (at \(\omega < \omega_p\)) and longitudinal plasmons (at \(\omega > \omega_p\)), discussed above [50]. In the latter case the change of the reflectivity at the resonances in comparison with classical local description can be many 10s of per cent [48, 50].

Of course, the most vivid implication of the anomalous skin effect lies in the region of low temperatures with larger mean free paths, specifically at microwave frequencies, where it fixes a two order of magnitude mismatch in absorbance between the experimental measurements and the values predicted by a theory of the ordinary skin effect [44]. At the same time, it has been shown that even for visible frequencies and room temperature the anomalous skin effect can lead to an essential ~10s of per cent change in the mean free path.
retrieved from fitting the measured reflectance curves with the classical Drude–Lorentz model [49, 58, 60]. These agrees with theoretical calculations in other works [46, 48, 50] (note that a 1% change in absorption (and consequently reflection) can lead to a ~20% change in the derived underlying scattering rate). We want to stress here, that due to the dependence of the optical response on the field distribution, the effective scattering rate \( \gamma \) essentially depends on the illumination method, particularly on the polarisation and the illumination angle, as well as the character of the surface scattering (specular or diffuse) [46, 48, 49, 59]. The general observation is that at near-infrared and visible frequencies (but differently already in mid-infrared and further) for specular scattering (\( p = 1 \)) at normal incidence and in the case of s-polarised oblique illumination, there is no scattering rate correction related to nonlocality [49, 59]. (This disagrees with [43], most probably due to different assumptions used.) Such correction appears for \( p \)-polarisation at oblique incidence [49]. For the film with diffusive or partly diffusive scattering the electron scattering rate correction due to the nonlocality exists even for normal incidence [58–60]. Theye explicitly showed that the Drude material parameters can be directly derived from experimentally measured spectral dependencies [59]. More on the theoretical insight on this matter and also a suggestion on the way to optically measure the needed scattering parameter \( p \) was reported in Ref. [47]. Importantly, Sotelo et al. [61] established a very interesting connection between the grain scattering correction in the classical DC case described in section 3 and in the nonlocal case at higher frequencies considered here, making the link between the two approaches.

5. Optical response in quantum description

In their following works [48, 49] Kliewer and Fuchs noticed that the previously used approach based on the Boltzmann equation treats the single-particle (electron–hole, e–h) excitations not entirely correctly, as it is not valid for wavevectors larger than the Fermi wavevector of electrons \( k_s \). Instead, they presented the analysis on the basis of a quantum-corrected SCIB model, which treats the situation properly using Lindhard description of a quantum electron gas taking into account both nonlocal effects and single-particle excitations [48, 62]. However, in a further investigation the formulas derived in this paper were found to be not quite correct, following which the proper expressions for the permittivities were presented [54, 63]. An overview of other approaches one can find in Refs. [18, 64] and overall explanatory discussion is given in Ref. [54]. Finally, we note that a quantum-corrected SCIB model for the case of diffuse scattering has also been created, though it was customised for a specific case of electron energy-loss spectroscopy [65]. In practical terms, if the full description of single-particle excitations will be ignored, this will result in a mistake in the scattering rates obtained analysing reflectivity data using the classical Drude model of the order of \( \sim 10\% \) [48]. It is interesting to note that the reverse process of plasmon excitation by electrons through electron-plasmon coupling is also possible. Plasmonic excitations can be generated via inelastic tunnelling [66, 67] through gaps in clustered films under application of voltage [68] or in a more controlled nanoantenna array geometry [69], where this can be used to engineer electrically-driven light emission. Another example of electron-plasmon coupling is the excitation of plasmonic modes by electron beams [70], observed in diverse structural geometries, from clustered films [71] to individual nanoparticles [72].

The next level of the model development is implemented in a quantum infinite barrier (QIB) approach, in which the metal interface is still treated as an infinite potential barrier like in the SCIB model, but the quantum-mechanical variation of the electron density in its vicinity is introduced, taking a form of a spill-in effect. At the same time this comes at a price. The quantum-corrected SCIB model allows an introduction of exchange-correlation effects (certainly with some approximations). Apart from a more realistic description of the electron gas, the most vivid consequence of this treatment is the appearance of an electron spill-out effect. At the same time this comes at a price. The quantum-corrected SCIB model allows an introduction of a realistic bulk dielectric function. Particularly it includes a straightforward mechanism for the inclusion of bulk relaxation processes. Thus, the conductivity in the surface region is completely specified within the quantum-corrected SCIB in terms of the bulk processes [18]. RPA treats the surface potential more realistically, but the introduction of the energy bands and bulk relaxation time in RPA is a much more difficult task [18]. Comparing RPA and quantum-corrected SCIB models in the case of light reflection form a metal surface, one can see that the metal surface absorbance can be \( 10–20 \) times different (for RPA it is higher) [18], though note that the absorbance is of the order of just 1%. In a further development of finite-potential models, it is possible to incorporate diffuse scattering [76], while the inclusion of an
atomistic structure in a form of a harmonically varied potential inside the metal was done in Refs. [77, 78]. Generally, one can say that in the above DFT and RPA models the atomic core background has quite an approximated form (although justified within a model of almost free electrons), particularly within the jellium model with a constant core potential within the material or harmonic potential as in the last works. At the same time, the distribution of the fields is either completely self-consistent or self-consistent to a good extent with the material response. It was also developed an alternative approach, when the field is calculated from the local models and then corrections to it in a form of is spatial distribution (or mode dispersion) and/or loss characteristics are found calculating the exact band structure of the metal, with an inclusion of direct and phonon-assisted loss mechanisms, as well as Landau damping (single-particle excitations) [79]. To ease the numerical complexity of the calculations, hybrid methods merging classical and quantum descriptions within the same model have been developed, e.g. [80].

Finally, it is interesting to note, that the nonlocal and single-particle effects generally modify the permittivity at the optical field wavevectors of the order of \( k \sim \omega / \nu_F \sim 1 \ \text{nm}^{-1} \) [54], so seemingly only the modes localised at a scale of 1 nm will be affected. But the later conclusion is misleading and incorrect, as with careful calculation (or even with a simple estimation) one can show that they influence the modes having a much larger characteristic length scale, such as surface plasmon polaritons [15, 79, 81], as the latter have in their wavevector spectrum a sufficient presence of the required large-magnitude wavevector components [15, 81]. Another vivid example is the modified absorption/reflection of plane waves discussed above, as the fields present in the metal in this case are also commonly localised at a \( \sim 15–30 \ \text{nm} \) scale.

The quantum research was further extended to the case of ultrathin films [82-85]. The most interesting phenomena in this case happen when the thickness of the metallic film approaches \( \sim 1 \ \text{nm} \) and becomes comparable with an electron Fermi wavelength \( \lambda_F \), which leads to the quantisation of the electronic states across the film thickness [77, 86–89]. Generally, the film supports symmetric and antisymmetric collective surface plasmon modes in respect to the symmetry of a charge distribution across the film thickness (note the notation here is opposite to that in ‘macroscale’ plasmonics, where it is defined in terms of the symmetry of the electric field). In the calculated results these modes can be identified from the plots of obtained charge density across the film. It was found that the symmetric mode experiences oscillations in its spectral width with the change of the film thickness. The period of the oscillations corresponds to the quantisation of the electronic levels and e–h pairs [87, 89]. Thus, Landau damping for this mode can be effectively tuned by varying the film thickness. Logically, its asymmetric counterpart disappears in the limit of a zero thickness, with a few e–h excitation peaks appearing instead [90]. Additionally, using an \textit{ab-initio} approach it was found that in the case of a monolayer, an acoustic plasmon mode with sound-like dispersion appears at small energies [88]. The authors further extended the approach from the jellium model to the inclusion of a full band structure and found a mode resulting from strong hybridisation of a surface plasmon with the single-particle excitations.

Here it is a good point to mention quantum-corrected or fully quantum descriptions of systems having lower dimensions, particularly metallic nanoparticles. A good review of classical, SCIB, RPA and fully-quantum jellium-based models as well as \textit{ab initio} models for metallic nanoparticles is given in a book of Kreibig and Vollmer [91]. The state of the art of atomistic \textit{ab initio} calculations of metallic nanoparticles up to a few nm in size can be found in Refs. [21, 92], while approximate atomistic models which can treat larger nanostructures of 1–15 nm in size are presented in Refs. [93, 94].

So far various approaches considered the modification of permittivity resulting from the change of the electron gas quantum response, applying various approximations to the gas behaviour and the background potential. Here, we make an important note on the other factors which can play a substantial role within the electronic response mechanism or beyond it, such as: (a) exchange-correlation potentials can be different from the bulk ones near the interface [18], (b) with the change of the electron density in the spill-out regions the electron wave-functions are also different from that in the bulk, (c) the polarisability of the lattice atoms in this region can also change, (d) electron–phonon scattering rate can be influenced by the interface, the phonon spectrum there can be modified [95–97], while strong fields at the interface can induce strong (Coulomb) electron–phonon interaction [91], (e) special surface phonons may occur [91], and (f) boundary morphology can play an important role [91]. Finally, effects related to the surrounding medium can make a substantial influence on the system [91, 98].

6. 2D materials

The era of 2D materials started in 2004 with pioneering experimental research on graphene by Andre Geim and Kostya Novoselov, for which they received a Nobel Prize in Physics in 2010 [99, 100]. Graphene is formed by a monoatomic layer of carbon atoms arranged is a honeycomb lattice. It has very unusual electronic and optical properties owing to Dirac-cone carrier dispersion. Carriers in this case have a linear relation between
their wavevector and frequency and can be viewed as massless Dirac fermions. The dispersion cones for electrons and holes meet at their tips at the Dirac point. Thus, graphene is a semi-metal with a zero bandgap.

In pristine graphene the valence band is fully occupied, while the conduction band is vacant. A prominent feature of graphene, however, is that free carriers can be easily introduced by application of external gating voltage. Moreover, it is possible to switch the conduction mechanism from electron- to hole-based applying voltages of opposite polarities. Also, carriers can be added in a more traditional way by doping.

The optical properties of graphene are governed by its dynamic frequency-dependant surface conductivity. Generally, the conductivity is nonlocal and can be derived using Kubo formalism [101]. However, in the long-wavelength limit it is transformed into a local form, which can be further separated into an interband term related to transitions between the valence and conduction bands and an intraband term related to the transitions within the conduction band for electrons or the valence band for holes. Notably, following the modification of the carrier density and consequently conductivity, the optical properties of graphene can be continuously tuned by application of gate voltage. For high frequencies \( \hbar \omega > 2 \mu \) (\( \mu \) is the gating- or doping-defined chemical potential), the optical response is dominated by the interband transitions causing absorption [102]. In the frequency range \( \hbar \omega < 2 \mu \) the interband absorption disappears and the intraband term gradually takes over towards lower frequencies. Finally, for the frequencies \( \hbar \omega \ll 2 \mu \) ranging from THz to mid-infrared, the intraband term becomes dominant and takes a Drude-like form.

Owing to two-dimensional confinement of the electron gas and low absorption in the long-wavelength region, graphene can support highly-localised (\( \sim 1/100 \) of the free space wavelength) low-loss plasmonic modes, which can be furthermore actively tuned by application of gating voltage or engineered by doping. In the simplest case of the low-frequency limit and the Drude-like surface conductivity expression one can obtain the dispersion of a transverse magnetic (TM)-polarised graphene plasmonic mode from local (at \( k \ll k_F \)) semi-classical calculations [16, 101–104]. Alternatively, the plasmon dispersion can be obtained using a hydrodynamic approach [105, 106]. In the large-wavevector region \( k \sim k_F \), a nonlocal RPA approximation should be employed [101, 102, 105, 107, 108], which can explicitly give the spectral region of the plasmon existence \( \hbar \omega < 1.667 \mu \) (for simplicity estimated for a collisionless case and at zero temperature). In either case a square-root plasmon dispersion is obtained \( \omega_{\text{GSP}} \sim \sqrt{F} \), which is characteristic to plasmons in 2D electron gas (2DEG) systems, e.g. in semiconductor quantum wells.

The carrier concentration dependence for graphene plasmon is \( \omega_{\text{GSP}} \sim \sqrt{\hbar \nu} \), which can explicitly give the spectral region of the peculiar linear carrier dispersion in graphene. Interestingly, the same type of plasmon dispersion \( \omega \sim \sqrt{F} \) can be obtained for an odd (in terms of the electric field) TM plasmonic mode in an infinitesimally thin metal slab with classical Drude permittivity, which allows to draw an analogy between graphene plasmonics and its classical counterpart, but also to expose the differences between them due to unlike electron and therefore optical dispersions of the materials [109, 110]. Additionally, graphene can also support TE-polarised (transverse electric) plasmonic waves [109, 111] and plasmons related to interband transitions (\( \pi \) and \( \sigma + \pi \)) [112]. Overall, due to its unique optical properties graphene can be used to realise emission, transmission, modulation and detection of light signals within a single material platform.

The discovery of graphene led to explosive research on 2D materials, which have shown diverse electronic and optical properties and can be dielectric, semiconductor or metallic. Employing an RPA approach it was found that doped semiconductor 2D materials with buckled honeycomb lattices, such as silicene and germanene (formed from Si and Ge atoms, respectively) also support plasmonic modes, however in this case they are influenced by spin–orbit interaction [113, 114]. Two-dimensional semiconductors possessing lattice anisotropy, such as phosphorene (a monolayer of black phosphorus) provide directionally-sensitive propagation of 2D plasmons [115]. Furthermore, such 2D anisotropic materials hold a promise of the realisation of hyperbolic dispersion for the plasmonic modes and therefore achieving highly-directional hyperbolic plasmons [116]. Finally, arranging 2D materials in layered structures opens a new area of artificial van der Waals heterostructures with engineered unique electronic and optical properties [117, 118].

### 7. State-of-the-art experimental research

Correct description of metal optical properties has become particularly important with the development of the field of experimental plasmonics, where the nanoscale size of the metallic structures can make a profound effect on the optical response of the metal at various levels of theoretical description discussed above, starting from the influence of electron scattering on the nanostructure boundaries and finishing with quantum effects. In this section we will give some examples of the state-of-the-art experimental developments in this field making a link to the theoretical descriptions considered above.

Drude–Lorentz permittivity with a corrected scattering rate due to the electron scattering on the grain boundaries was successfully employed to explain the quality factors of plasmonic resonances in
polycrystalline nanoantennas [119]. Analogous correction to the Drude scattering rate due to increased scattering on the surfaces was used to fit the optical properties of ultrathin metallic films with the thickness of just a few nanometres [120–122].

A particularly interesting direction in the plasmonic research is the investigation of localised plasmonic resonances in metallic nanoparticles with sizes approaching truly nanoscale dimensions. When the size of the nanoparticles approaches few nanometres, all the phenomena discussed in the review, such as nonlocality, Landau damping, electron spill-out and quantum-size effects start to be extremely important and affect both spectral positions and widths of the plasmonic resonances. What makes this area particularly interesting is that various effects result in different trends of the spectral position of the resonance and different levels of its broadening with the decrease of the nanoparticle size [91, 123–127]. Therefore, now it has come a crucial moment when the recently advanced experimental techniques give opportunity to test various semi-classical and quantum theories. In this respect, one can particularly emphasize the experimental findings in Refs. [128, 129], where evolution of the plasmonic resonance position was successfully explained by an interplay of quantisation of the electronic states in the confined geometry of the nanoparticle and the size-related increase of the electron scattering rate. Moreover, in the second work it was found that such evolution can be non-monotonous. In recent research, Campos et al. employing experimental measurements in conjunction with a comprehensive quantum model of the optical response of the electron gas showed that the surroundings can play an important role in the change of the resonance position through the modification of a relative weight between various quantum effects, particularly electron spill-out and screening of d-band electrons near the surface [130]. Furthermore, on the basis of these findings it was resolved an apparent controversy that optical absorption and electron-loss experiments on the same system can give different results, explained by the fact that in the latter case the electron beam vastly changes the porosity of the matrix around the nanoparticle.

8. Conclusions

With quite a complex behaviour of optical properties of metallic films and objects together with various levels of their theoretical description, it is important to understand in which situations which approach is appropriate. Careful treatment on the basis of SCIB, QIB, RPA or quantum-confinement models is generally needed in the case of single crystals or single crystal structures, particularly having nanoscale dimensions. At the same time, a lot depends on: (a) the particular characteristic which is studied, e.g. field or fluorescence enhancement, modal loss or modal dispersion, and (b) the precision with which this characteristic needs to be estimated. For example, one can envisage that the scattering characteristics of metal nanoparticles or nanoflakes can be evaluated with some precision with classical corrections [91], especially if this is accompanied and checked with experimental results. At the same time, for calculation of the fluorescence of a molecule positioned near their surface the situation is different and the choice of the approach (classical, nonlocal SCIB or QIB) largely depends on the distance between the molecule and the metal surface [54]. Then, at the next level of surface inhomogeneity, with an appearance of surface roughness supplying momentum in the in-plane direction and facilitating photon absorption, via the introduction of diffuse scattering (still within a model of a flat surface) the extra absorption mechanisms need to be taken into account. This can be treated using quantum-corrected or semiclassical models. Furthermore, in the case of polycrystalline films commonly used in the experiments the structural inhomogeneity smeared all other effects and the classical Drude–Lorentz description with a corrected (or phenomenologically introduced) scattering rate can be used. This explains the fact that in an overwhelming amount of observations the classical local approach was valid. Interestingly, this also includes experimental studies specifically devoted to the measurement of optical properties of metals, the nonlocal and single-particle excitation effects in this case are hidden inside the obtained effective Drude–Lorentz parameters. Considering this question in practical terms, this, however, can be justified, as in the further studies which will use these data, the fabricated film in any occasion will be different from the reported, which will substantially influence its optical properties on the top. As an illustration, the obtained tabulated optical data vary from study to study due to different fabrication procedures. Thus, in many cases Drude–Lorentz description will suffice giving a reasonable level of precision, as long as the used approximation is kept in mind. Finally, in the case of films with extreme irregularity effective medium or island-based models will be appropriate.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.
ORCID iD

Alexey V Krasavin  ●  https://orcid.org/0000-0003-2522-5735

References

[1] Ma R-M and Oulton R F 2019 Nat. Nanotechnol. 14 12
[2] Heni W et al 2019 Nat. Commun. 10 1694
[3] Fang N, Lee H, Sun C and Zhang X 2005 Science 308 534
[4] Stipe B C et al 2010 Nat. Photon. 4 484
[5] Wang M Y, Li M Q, Jiang S, Gao T J and Xi P 2020 Micron 137 102916
[6] Koker T, Tang N, Tian C, Zhang W, Wang X D, Martel R and Pinaud F 2018 Nat. Commun. 9 607
[7] Kauranen M and Zavats A V 2012 Nat. Photon. 6 737
[8] Panoiu N C, Sha W E I, Lei D Y and Li G-C 2018 J. Opt. 20 083001
[9] MacDonald K F, Samson Z L, Stockman M I and Zel'dovich N I 2009 Nat. Photon. 3 55
[10] Gargiulo J, Berte R, Li Y, Maier S A and Cortes E 2019 Acc. Chem. Res. 52 2525
[11] Giannini V, Fernandez-Dominguez A I, Heck S C and Maier S A 2011 Chem. Rev. 111 3888
[12] Nie S M and Emery S R 1997 Science 275 1102
[13] Agranovich V M and Ginzburg V L 1984 Crystal Optics with Spatial Dispersion, and Excitations (Berlin: Springer)
[14] Agranovich V M and Mills D L 1982 Surface Polaritons (Amsterdam: North-Holland Publishing Company)
[15] Wubs M and Mortensen N A 2017 Nonlocal response in plasmonic nanostructures Quantum Plasmonics ed S I Bozhevolnyi, I. Martin-Moreno and F Garcia-Vidal (Cham: Springer)
[16] Christensen T 2017 From Classical to Quantum Plasmonics in Three and Two Dimensions (Cham: Springer)
[17] Novotny L and Hecht B 2012 Principles of Nano-Optics (Cambridge: Cambridge University Press)
[18] Feibelman P J 1982 Prog. Surf. Sci. 12 287
[19] Bagchi A, Kar N and Barrera R G 1978 Phys. Rev. Lett. 40 803
[20] Liebsch A 1993 Phys. Rev. B 48 11317
[21] Rossi T P, Erhart P and Kusma M 2020 ACS Nano 14 9963
[22] Mayadas A F, Shatzkes M and Janak J F 1969 Appl. Phys. Lett. 14 345
[23] Pichard C R, Tellier C R and Tossel A J 1980 Phys. Status Solidi b 99 353
[24] Reiss G, Vancea J and Hoffmann H 1986 Phys. Rev. Lett. 56 2100
[25] Liberstiber P H and Nelson R G 1974 Thin Solid Films 21 159
[26] Fuchs K 1938 Proc. Camb. Phil. Soc. 34 100
[27] Sondheimer E H 1952 Adv. Phys. 1 1
[28] Lucas M S P 1965 J. Appl. Phys. 36 1632
[29] Mayadas A F and Shatzkes M 1970 Phys. Rev. B 1 1382
[30] Dudek J C 1986 Thin Solid Films 137 11
[31] Warkusz F 1988 Surf. Sci. 200 394
[32] Zhu Y F, Lang Y X, Zheng W T and Jiang Q 2010 ACS Nano 4 3781
[33] Zhang X, Song X H, Zhang X-G and Zhang D L 2011 Europhys. Lett. 96 17010
[34] Shubin V A, Sarychev A K, Clerc J P and Shalaev V M 2000 Phys. Rev. B 62 11230
[35] Aspnes D E, Kinsbron E and Bacon D D 1980 Phys. Rev. B 21 3290
[36] Dobierzewska-Mozrzymas E, Bieganski P and Pieciul E 2001 Opt. Appl. 31 53
[37] Fedotov V A, Emel`yantov V I, MacDonald K F and Zel’dovich N I 2004 J. Opt. A 6 155
[38] Reuter G E H and Sondheimer E H 1948 Proc. R. Soc. A 195 336
[39] Dingel R B 1953 Physica 19 729
[40] Klierwer K L 1980 Surf. Sci. 101 57
[41] Klierwer K L 1976 Phys. Rev. B 14 1412
[42] Holstein T 1952 Phys. Rev. 88 1427
[43] Dingle R B 1953 Physica 19 311
[44] Dingle R B 1953 Physica 19 348
[45] Dingle R B 1953 Physica 19 1187
[46] Hutchison F E and Hansen W N 1979 Phys. Rev. B 20 4069
[47] Szczyrbowski J, Dryzek J and Czapla A 1984 Thin Solid Films 112 175
[48] Fuchs R and Klierwer K L 1969 Phys. Rev. 185 905
[49] Klierwer K L and Fuchs R 1968 Phys. Rev. 172 607
[50] Jones W E, Klierwer K L and Fuchs R 1969 Phys. Rev. 178 1201
[51] Takimoto N 1968 J. Phys. Soc. Japan 25 390
[52] Maniv T and Metiu H 1982 J. Chem. Phys. 76 696
[53] Klierwer K L 1977 Phys. Rev. B 15 3759
[54] Ford G W and Weber W H 1984 Phys. Rep. 113 195
[55] Melnyk A R and Harrison M J 1970 Phys. Rev. B 2 835
[56] Szczyrbowski J, Schmalzbauer K and Hoffmann H 1985 Phys. Rev. B 32 763
[57] Dryzek J and Czapla A 1986 Phys. Status Solidi a 94 665
[58] Hutchison F E and Hansen W N 1979 Phys. Rev. B 20 4076
[59] Theye M L 1967 Phys. Lett. A 25 764
[60] Theye M L 1970 Phys. Rev. B 2 3060
[61] Sotelo J, Edert J and Niklasson G 2003 Phys. Rev. B 67 195106
[62] Klierwer K L and Fuchs R 1969 Phys. Rev. 181 352
[63] Mermin N D 1970 Phys. Rev. B 2 2362
[64] Mukhopadhyay G and Lundqvist S 1978 Phys. Scr. 17 69
[65] Nazarov V U 1997 Phys. Rev. B 56 2198
