Nonlocal optical response in metallic nanostructures

Topical Review

Nonlocal optical response in metallic nanostructures

Søren Raza[1,2], **Sergey I Bozhevolnyi**[3], **Martijn Wubs**[1,2] **and N Asger Mortensen**[1,2]

[1] Department of Photonics Engineering, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
[2] Center for Nanostructured Graphene (CNG), Department of Micro- and Nanotechnology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
[3] Department of Technology and Innovation, University of Southern Denmark, DK-5230 Odense, Denmark

E-mail: asger@mailaps.org

CrossMark

Abstract

This review provides a broad overview of the studies and effects of nonlocal response in metallic nanostructures. In particular, we thoroughly present the nonlocal hydrodynamic model and the recently introduced generalized nonlocal optical response (GNOR) model. The influence of nonlocal response on plasmonic excitations is studied in key metallic geometries, such as spheres and dimers, and we derive new consequences due to the GNOR model. Finally, we propose several trajectories for future work on nonlocal response, including experimental setups that may unveil further effects of nonlocal response.

Keywords: nanoplasmonics, nonlocal response, hydrodynamics

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

1. Introduction

The excitation of surface plasmons (SPs), i.e. the collective movement of conduction band electrons tightly bound to a metal-insulator interface, governs most of the phenomena observed in plasmonic studies. A few examples of the vast number of surface-plasmon related effects feature the large enhancement of the electric field in metal nanoparticles of close proximity [1] and sharp metal geometries [2], the squeezing of light beyond the diffraction limit [3, 4], and the tunability of the optical properties of metallic structures with size and shape [5]. These appealing properties have found applications in different fields, such as surface-enhanced Raman spectroscopy [6], biosensing [7, 8], plasmonic waveguiding [9], cancer therapy [10], and on-chip circuitry [11, 12].

The theoretical modelling of plasmonic phenomena is for the most part based on the macroscopic Maxwell's equations. In particular, the optical response of metals is described through the constitutive relations, which relate the response of the material to the applied field. In the linear regime, the constitutive relation relating the displacement field **D** to the electric field **E** is

$$\mathbf{D}(\mathbf{r}, \omega) = \varepsilon_0 \int d\mathbf{r}' \varepsilon(\mathbf{r}, \mathbf{r}', \omega) \mathbf{E}(\mathbf{r}', \omega), \tag{1}$$

where ε_0 is the vacuum permittivity and $\varepsilon(\mathbf{r}, \mathbf{r}', \omega)$ denotes the nonlocal permittivity of the metal, here assumed scalar. In a *homogeneous* medium, the nonlocal permittivity depends spatially on $\mathbf{r} - \mathbf{r}'$, turning equation (1) into a convolution which in **k**-space becomes the product

$$\mathbf{D}(\mathbf{k}, \omega) = \varepsilon_0 \varepsilon(\mathbf{k}, \omega) \mathbf{E}(\mathbf{k}, \omega). \tag{2}$$

We see that nonlocal response corresponds to a **k**-dependent dielectric function. For an isotropic response, we find dependence only on the length of **k**, and not its direction.

In the simpler local-response approximation (LRA), nonlocal effects are neglected and we write $\varepsilon(\mathbf{r}, \mathbf{r}', \omega) =$

$\delta(\mathbf{r} - \mathbf{r}')\varepsilon(\omega)$, which allows us to straightforwardly perform the integral in equation (1) as

$$\mathbf{D}(\mathbf{r}, \omega) = \varepsilon_0 \varepsilon(\omega)\mathbf{E}(\mathbf{r}, \omega). \quad (3)$$

Here, $\varepsilon(\omega)$ is the spatially constant permittivity of the metal, usually described by the Drude-like dielectric function [26–28]

$$\varepsilon(\omega) = \varepsilon_{\text{core}}(\omega) - \frac{\omega_{\text{p}}^2}{\omega(\omega + \mathrm{i}\gamma)}. \quad (4)$$

In equation (4), $\omega_{\text{p}}^2 = n_0 e^2/(\varepsilon_0 m)$ is the unscreened plasma frequency of the metal with n_0 denoting the equilibrium electron density of the free electrons, γ is the Drude damping rate, and $\varepsilon_{\text{core}}(\omega)$ is the response from the bound ions and electrons, which account for effects such as interband transitions. The importance of the LRA is accentuated by its prevalence in the plasmonic community, being the most commonly applied constitutive description [29]. The LRA has successfully described a plethora of plasmonic phenomena and experiments, such as optical far-field measurements [30–32], electron energy-loss spectroscopy (EELS) [33–40], cathodoluminescence experiments [41–43], near-field microscopy [44], and surprisingly, even effects in the two-dimensional material graphene [45] and plasmonic particles interspaced by nanometer-sized gaps [46].

Despite its success, the LRA has been challenged on a number of accounts. One example is the size-dependent SP linewidth broadening observed in metal clusters and small nanoparticles [47–49], which has to be phenomenologically accounted for in the LRA through an increased damping rate as a consequence of surface screening and scattering [50–52]. Size-dependent resonance shifts of the SP in noble metal nanoparticles have also been observed in both optical measurements [53, 54] and EELS (figure 1(a)) [13–16]. Another example is the multipole plasmon which, besides the usual surface-plasmon polariton, can be supported by the simple geometry of a metal-vacuum interface [55] as a direct consequence of the spill-out of free electrons beyond the classical metal boundary [56–58]. Thin metal films have also been shown to support resonant excitations above the plasma frequency due to confined longitudinal waves [59, 60], which are not taken into account in the LRA. Recently, several experiments on metal dimers with particles in mutual subnanometer proximity have shown plasmonic effects clearly going beyond the LRA (figures 1(b)–(d)) [17–25]. A theoretical description of the metal based on *ab initio* approaches such as density-functional theory (DFT) [61, 62] or similar theories [63] seem to capture all of the observed non-classical effects [64–67]. However, due to the computational demand of such approaches, only very small system sizes (few nanometers) can be considered [65], which puts serious constraints on the feasibility of these approaches for a generic plasmonic system, which is usually tens of nanometers or more. Another simpler and computationally less demanding path is to go beyond the LRA by taking into account nonlocal response through a hydrodynamic approach [68]. The hydrodynamic approach has been able to describe size-dependent resonance shifts of noble metal nanoparticles and
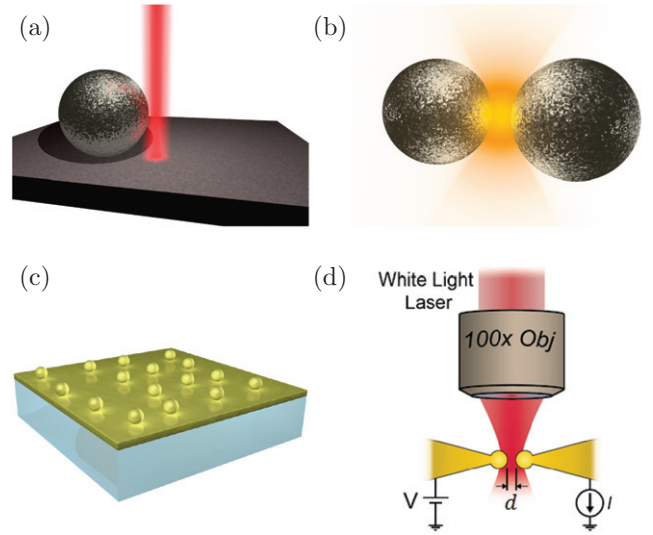
Figure 1. Schematic illustrations of (a) electron beam interacting with silver nanoparticle on thin substrate [13–16], (b) electromagnetic interaction of two nearby metal nanoparticles [17–23], (c) gold nanoparticles in close proximity of gold film [24], and (d) optical measurements of two gold-coated atomic-force microscopy tips in close proximity [25]. (c) and (d) courtesy of C Ciracì and J Baumberg, respectively.

gap-dependent resonance shift in a particle-film system [24], and can now with the inclusion of electron diffusion [69] also describe size-dependent damping and the optical spectra of closely-spaced dimers. Besides taking into account retardation effects and being physically transparent, significant analytical progress is also possible with the hydrodynamic approach. Many of these properties in the hydrodynamic approach are beneficial in the theoretical studies of generic plasmonic systems with large (>10 nm) feature sizes.

The aim of this Topical Review is to give a comprehensive overview of the real-space formulation of nonlocal response in Maxwell's equations. In particular, we review the hydrodynamic [68] and the generalized nonlocal optical response (GNOR) models [69], which are two examples of nonlocal response theories. We show that nonlocal response in metals manifests itself through the presence of longitudinal waves. The influence of nonlocal response in metallic nanostructures is exemplified by considering a silver nanosphere, a silver dimer with Ångstrom-sized gaps, and a core-shell cylinder with a nanometer-thin silver shell. Finally, we discuss future theoretical and experimental directions to unveil the effects of nonlocal response.

2. Phenomenological theory of nonlocal response

We begin by considering the real-space formulation of Maxwell's equations taking into account nonlocal response. In the absence of external sources and using equation (1), the nonlocal wave equation with assumed scalar nonlocal response reads

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r}, \omega) = \left(\frac{\omega}{c}\right)^2 \int \mathrm{d}\mathbf{r}' \varepsilon(\mathbf{r}, \mathbf{r}', \omega)\mathbf{E}(\mathbf{r}', \omega), \quad (5)$$

where we have furthermore assumed the material to be non-magnetic, i.e. $\mathbf{B}(\mathbf{r}, \omega) = \mu_0 \mathbf{H}(\mathbf{r}, \omega)$, which is applicable for plasmonic metals. We note that in equation (5) the electromagnetic response of the material is described only through the displacement field $\mathbf{D}(\mathbf{r}, \omega)$, which accounts for effects both from the bound and free charges of the metal. Recognizing that the LRA accounts successfully for many of the observed plasmonic phenomena, we proceed by writing the nonlocal permittivity as [70, 71]

$$\varepsilon(\mathbf{r}, \mathbf{r}', \omega) = \varepsilon(\omega)\delta(\mathbf{r} - \mathbf{r}') + f(\mathbf{r} - \mathbf{r}', \omega), \quad (6)$$

where $f(\mathbf{r} - \mathbf{r}', \omega)$ is the scalar nonlocal response function associated with a homogeneous medium. We make the justifiable assumptions that $f(\mathbf{r} - \mathbf{r}', \omega)$ is symmetric and short-ranged [70]. We can express these assumptions mathematically through the moments of the function

$$\int \mathbf{r} f(\mathbf{r}, \omega) d\mathbf{r} = 0, \quad (7a)$$

$$\int \mathbf{r}^2 f(\mathbf{r}, \omega) d\mathbf{r} = 2\xi^2, \quad (7b)$$

where $\mathbf{r}^2 \equiv (x^2, y^2, z^2)$ and we have introduced the length scale ξ as the range (or width) of the nonlocal response function. Since the response function $f(\mathbf{r} - \mathbf{r}', \omega)$ is short-ranged (i.e. ξ in equation (7b) is small on the length scale of the variations of the electric field), we Taylor expand the electric field in the integrand of equation (5) around $\mathbf{r}$. To second order we find (omitting the frequency dependency)

$$E_i(\mathbf{r}') \simeq E_i(\mathbf{r}) + [\nabla E_i(\mathbf{r})] \cdot (\mathbf{r}' - \mathbf{r})$$
$$+ \frac{1}{2}(\mathbf{r}' - \mathbf{r})^T \cdot \left[\hat{\mathbf{H}} E_i(\mathbf{r})\right] \cdot (\mathbf{r}' - \mathbf{r}), \quad (8)$$

where the Hessian matrix $\hat{\mathbf{H}}$ has the elements $H_{ij} = \partial^2/(\partial_i \partial_j)$ with $i, j = x, y, z$. With these considerations in mind and using the assumptions in equations (7)–(8), we can perform the integral in equation (5) and find

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}, \omega) = \left(\frac{\omega}{c}\right)^2 \left[\varepsilon(\omega) + \xi^2 \nabla^2\right] \mathbf{E}(\mathbf{r}, \omega), \quad (9)$$

where we have absorbed the zeroth-order moment integral of $f(\mathbf{r}, \omega)$ (i.e. $\int f(\mathbf{r}, \omega) d\mathbf{r}$) into our definition of $\varepsilon(\omega)$. Interestingly, equation (9) shows that *scalar* nonlocal response manifests itself through the Laplacian term in the wave equation, seemingly irrespective of the microscopic or semiclassical origin, and with a strength given by ξ, which relates to the width of the nonlocal response function through equation (7b). This general analysis of nonlocality also suggests the possibility of several nonlocal mechanisms playing in concert and adding up to an effective ξ^2 [72]. Furthermore, we have transformed the integro-differential equation of equation (5) into a regular differential equation, where the presence of the Laplacian operator does not give rise to increased numerical difficulties than the already present double-curl operator. As we will see in sections 3 and 4, the nonlocal hydrodynamic and GNOR models, respectively, can also be rewritten into the wave-equation form of equation (9), although the Laplacian operator is replaced by a gradient-of-the-divergence operator $[\nabla(\nabla \cdot)]$.

3. Hydrodynamic model

We discuss now the hydrodynamic model for the free-electron gas, which will allow us to determine the nonlocal length scale ξ introduced in equation (9). The idea of modeling the free-electron gas in a hydrodynamic formulation was first introduced by Bloch in a seminal paper in 1933 [73]. In 1974, Ying [74] extended Bloch's non-retarded approach to a more general density-functional formalism, allowing to go beyond the Thomas–Fermi ground state, which lacked information about the correlation and exchange energies of the electron gas. Shortly after Eguiluz and Quinn [75] included retardation effects. Despite the generalization by Ying [74], the Thomas–Fermi description of the electron gas in the hydrodynamic model remained popular, and was extensively used in the field of solid-state physics in the 1970s and the beginning of the 1980s. The effect of electron density inhomogeneity and spatial dispersion in planar interfaces [57, 76–78], multilayered structures [79, 80], spherical particles [81–83] and voids [84], and cylindrical particles [85, 86] was given a considerable attention. Especially, the results obtained for homogeneous planar surfaces showed good agreement with experiment [68].

Recently, interest in the hydrodynamic model was rekindled when a finite-element numerical implementation of the hydrodynamic equations was presented [87–89], which was subsequently utilized to study the plasmonic cylindrical dimer [88], surface-enhanced Raman spectroscopy [90], and waveguiding in metallic nanostructures [91]. Simultaneously, application of transformation optics to the hydrodynamic model allowed for analytical solutions of several non-trivial plasmonic structures, even some containing singular geometric features [92–94]. Additionally, numerous different metallic geometries and plasmonic effects have been studied using the hydrodynamic model, such as scattering and mode analysis of cylindrical structures, including nanotubes [91, 95–100], roughness effects on plasmonic tips [101, 102], nonlinear effects in nonlocal media [103–105], scattering of light off three-dimensional nanostructures [24, 106–110], surface plasmon propagation in metal-insulator, metal-insulator-metal, insulator-metal-insulator, and hourglass waveguides [111–115], epsilon-near-zero and perfect imaging effects [116–119], influence of nonlocal response on the Casimir force [120], studies of hyperbolic metamaterials and periodic media [121–126], investigations of nonlocal effects in EELS [127, 128], and coupling of dipole emitters to plasmonic structures [128–130]. Theoretical work has also been done to compare the hydrodynamic approach with more advanced approaches such as density-functional theory [65, 131, 132]. Finally, other nonlocal models besides the hydrodynamic model have also been utilized to study different geometries [51, 133–140].

As a detailed derivation of the hydrodynamic model has been reported before [68, 73, 75, 141], we focus here on the most important steps of the derivation and extract the essential physics of the hydrodynamic model. The basic assumption of the hydrodynamic model is that the many-electron energy and dynamics can be expressed in terms of a scalar field

and a velocity field, the electron density $n(\mathbf{r}, t)$ and the hydrodynamic velocity $\mathbf{v}(\mathbf{r}, t)$, respectively. The energy of the electron plasma is a functional of the electron density and velocity. The dynamics of these variables under the influence of macroscopic electromagnetic fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ is obtained by functional differentiation of the energy (i.e. Hamilton's principle). Functional differentiation with respect to the velocity field gives the hydrodynamic equation of motion [68]

$$[\partial_t + \mathbf{v} \cdot \boldsymbol{\nabla}] \, \mathbf{v} = -\frac{e}{m} \, [\mathbf{E} + \mathbf{v} \times \mathbf{B}] - \frac{1}{m} \boldsymbol{\nabla} \frac{\delta G[n]}{\delta n} - \gamma \mathbf{v}, \quad (10)$$

while functional differentiation with respect to the electron density gives the continuity equation

$$\partial_t n = -\boldsymbol{\nabla} \cdot (n\mathbf{v}), \quad (11)$$

expressing charge conservation. On the right-hand side of equation (10), the first term is the Lorentz force, while the second term can take into account the correlation, exchange and the internal kinetic energy of the electron gas, if an appropriate functional $G[n]$ is chosen [74]. Here, $\delta G[n]/\delta n$ denotes the functional derivative. The last term of equation (10), which represents damping in terms of the bulk damping rate γ, cannot be obtained from the energy functional approach and is therefore added phenomenologically.

The most common, and also the simplest, approach [56, 68] is to use the Thomas–Fermi model for the functional $G[n]$, given as

$$G[n(\mathbf{r}, t)] = \int \frac{3h^2}{10m} \left(\frac{3}{8\pi}\right)^{\frac{2}{3}} n^{\frac{5}{3}}(\mathbf{r}, t) \mathrm{d}\mathbf{r}, \quad (12)$$

which describes only the internal kinetic energy of the electron gas. The functional derivative of equation (12) can now be performed

$$\frac{\delta G[n]}{\delta n} = \frac{h^2}{2m} \left(\frac{3}{8\pi}\right)^{\frac{2}{3}} n^{\frac{2}{3}}(\mathbf{r}, t), \quad (13)$$

which upon insertion in equation (10) finally gives

$$[\partial_t + \mathbf{v} \cdot \boldsymbol{\nabla}] \, \mathbf{v} = -\gamma \mathbf{v} - \frac{e}{m} \, [\mathbf{E} + \mathbf{v} \times \mathbf{B}] - \frac{\beta^2}{n} \boldsymbol{\nabla} n. \quad (14)$$

In analogy with fluid hydrodynamics, the last term of equation (14) represents the pressure, which is proportional to $\beta^2 = 1/3 v_F^2$ with v_F denoting the Fermi velocity, and describes a force that will act to homogenize any inhomogeneity in the electron density. The pressure term gives rise to nonlocal response in the hydrodynamic model. The prefactor $1/3$ connecting β^2 to v_F^2 applies only in the low-frequency limit ($\omega \ll \gamma$) and must be multiplied by an additional factor of $9/5$ in the high-frequency limit ($\omega \gg \gamma$) [142]. Hence, we use the relation $\beta^2 = 3/5 v_F^2$ in our subsequent calculations [143].

In the spirit of linear-response theory, we now follow the usual approach [68, 141] to solve equations (11) and (14), by expanding the physical fields in a non-oscillating term (e.g. n_0 is the homogeneous equilibrium electron density),

Table 1. Plasma frequencies ω_P, Drude damping rates γ, Fermi velocities v_F and diffusion constants D for the metals Na, Ag, Au and Al.

| | $\hbar\omega_P$ [eV] | $\hbar\gamma$ [eV] | v_F [10^6 m s^{-1}] | D [10^{-4} m^2 s^{-1}] | |
				$A = 0.5$	$A = 1$
Na	6.04	0.16	1.07	1.08	2.67
Ag	8.99	0.025	1.39	3.61	9.62
Au	9.02	0.071	1.39	1.90	8.62
Al	15.8	0.6	2.03	1.86	4.59

Note: The method used for determining D is described by Raza [177].

and a small (by assumption) first-order dynamic term, as in perturbation theory, thereby linearizing the equation of motion and continuity equation. In the frequency domain, we obtain the coupled electromagnetic equations [88, 96]

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r}, \omega) = \left(\frac{\omega}{c}\right)^2 \varepsilon_{\text{core}} \mathbf{E}(\mathbf{r}, \omega) + \mathrm{i}\omega\mu_0 \mathbf{J}(\mathbf{r}, \omega), \quad (15a)$$

$$\frac{\beta^2}{\omega(\omega + \mathrm{i}\gamma)} \boldsymbol{\nabla}[\boldsymbol{\nabla} \cdot \mathbf{J}(\mathbf{r}, \omega)] + \mathbf{J}(\mathbf{r}, \omega) = \sigma \mathbf{E}(\mathbf{r}, \omega), \quad (15b)$$

where $\mathbf{J}(\mathbf{r}, \omega) = -en_0\mathbf{v}(\mathbf{r}, \omega)$ is the induced current density, and $\sigma = \varepsilon_0 \mathrm{i}\omega_p^2/(\omega + \mathrm{i}\gamma)$ is the Drude conductivity and relates to the Drude permittivity (equation (4)) as $\varepsilon(\omega) = \varepsilon_{\text{core}} + \mathrm{i}\sigma/(\varepsilon_0\omega)$. We see that in the LRA limit of $\beta \to 0$, equation (15b) reduces to Ohm's law. Now, by combining equations (15a) and (15b), we can rewrite the governing equations in the hydrodynamic model as

$$\boldsymbol{\nabla} \times \boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r}, \omega) = \left(\frac{\omega}{c}\right)^2 \left[\varepsilon(\omega) + \xi_H^2 \boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot)\right] \mathbf{E}(\mathbf{r}, \omega), \quad (16)$$

where we find the nonlocal parameter in the hydrodynamic model to be [69, 91]

$$\xi_H^2(\omega) = \frac{\varepsilon_{\text{core}}(\omega)\beta^2}{\omega(\omega + \mathrm{i}\gamma)}. \quad (17)$$

In the absence of interband effects and bulk damping mechanisms ($\varepsilon_{\text{core}} = 1$ and $\gamma = 0$), the nonlocal parameter is simply $\xi_H = \beta/\omega$, which is a purely real-valued quantity. We also see that the nonlocal parameter increases with the Fermi velocity (through β). In table 1 we list values for ω_p, v_F, and γ for relevant plasmonic metals. In addition, the response from the bound electrons $\varepsilon_{\text{core}}(\omega)$ can be determined from the measured bulk dielectric functions $\varepsilon_{\text{exp}}(\omega)$ using the recipe $\varepsilon_{\text{core}}(\omega) = \varepsilon_{\text{exp}}(\omega) + \omega_p^2/(\omega^2 + \mathrm{i}\gamma\omega)$ [138].

Comparing the hydrodynamic wave equation (equation (16)) with the wave equation from the phenomenological nonlocal model (equation (9)), we see that the mathematical operator responsible for the nonlocal effects is the gradient-of-the-divergence and not the Laplacian. In fact, a closer $\mathbf{k}$-space analysis of equation (9) reveals that the Laplacian operator introduces a weak spatial dispersion in the transverse part of the electric field [91], which is in contrast with the hydrodynamic wave equation where spatial dispersion only affects the longitudinal part of the electric field [96]. The Laplacian operator in the phenomenological model stems from the *scalar* nonlocal response function in equation (6).

Besides the hydrodynamic wave equation in equation (16), an additional enlightening way of writing up the nonlocal electromagnetic equations for the electric field is in terms of the divergence and curl of the electric field [76, 96, 112]

$$\left(\nabla^2 + k_{\mathrm{L}}^2\right) \boldsymbol{\nabla} \cdot \mathbf{E}(\mathbf{r}, \omega) = 0, \tag{18a}$$

$$\left(\nabla^2 + k_{\mathrm{T}}^2\right) \boldsymbol{\nabla} \times \mathbf{E}(\mathbf{r}, \omega) = 0. \tag{18b}$$

Here, $k_{\mathrm{L}}^2 = \varepsilon(\omega)/\xi_{\mathrm{H}}^2$ and $k_{\mathrm{T}}^2 = (\omega/c)^2 \varepsilon(\omega)$ are the wave vectors of the longitudinal and transversal electric field, respectively. In the Fourier domain (**k**-space), it becomes clear that equation (18a) describes the longitudinal part of the electric field, while equation (18b) corresponds to the transversal part of the electric field. The longitudinal and tranversal electric fields are two different types of waves, which in a homogeneous medium are uncoupled, but can in the presence of an interface be coupled by means of the electromagnetic boundary conditions. We stress that the difference between the LRA and the inclusion of nonlocal response in the hydrodynamic approach is the presence of the longitudinal wave, which will be responsible for all nonlocal effects.

3.1. Additional boundary condition

Within the LRA, Maxwell's boundary conditions, commonly derived using pill-box and current loop arguments, are sufficient to determine the amplitudes of the transversal electric and magnetic fields. However, the presence of an additional wave due to nonlocal response will require an additional boundary condition (ABC) to determine the amplitude of the longitudinal wave [77]. Inspired by the many discussions on the appropriate boundary conditions in the nonlocal hydrodynamic model [68, 77, 144–147], we will assume that the equilibrium free-electron density n_0 of the metal has a step profile, i.e. n_0 is constant inside the metal and abruptly drops to zero at the metal-dielectric interface. The consequence of this assumption is that the induced charge density ρ will vanish at the metal-dielectric boundary, at which a pill-box argument on the continuity equation $\boldsymbol{\nabla} \cdot \mathbf{J} = \mathrm{i}\omega\rho$ reveals the additional boundary condition

$$\mathbf{J} \cdot \hat{\mathbf{n}} = 0, \tag{19}$$

stating that the normal component of the induced current density $\mathbf{J}$ vanishes at the metal boundary. The step profile of n_0 is also at the heart of the LRA, but in contrast to equation (19), the normal component of the induced current at the metal boundary will in general not vanish, due to the LRA constitutive relation $\mathbf{J}(\mathbf{r}, \omega) = \sigma(\omega)\mathbf{E}(\mathbf{r}, \omega)$. We note that the ABC in equation (19) will not allow us to include the quantum mechanical effect of spill-out of electrons occurring due to the finite potential difference at the metal-dielectric interface. In fact, the ABC corresponds to the assumption of an infinite work function in more microscopic theories.

We add that the ABC stated in equation (19), which is valid for a metal-dielectric interface, can be rewritten in terms of the normal components of the electric field as [68, 91]

$$\varepsilon_{\mathrm{core}}\mathbf{E}_{\mathrm{m}} \cdot \hat{\mathbf{n}} = \varepsilon_{\mathrm{d}}\mathbf{E}_{\mathrm{d}} \cdot \hat{\mathbf{n}}, \tag{20}$$

where ε_{d} is the permittivity of the dielectric, $\varepsilon_{\mathrm{core}}$ is the response due to the bound charges in the metal (see equation (4)), and $\mathbf{E}_{\mathrm{m}}$ and $\mathbf{E}_{\mathrm{d}}$ are the electric fields in the metal and dielectric, respectively. Thus, in the special case of a metal-vacuum interface with a metal that has no response due to bound charges ($\varepsilon_{\mathrm{core}} = 1$), then the normal component of the electric field is continuous. For completeness, we add that a second ABC is needed at a metal-metal interface [146–148], but we will not discuss such interfaces in the following.

4. Generalized nonlocal optical response

A hitherto disregarded effect in the discussion of metallic nanostructures and nonlocal response is the classical phenomenon of electron diffusion [72, 149]. While the hydrodynamic model incorporates the convective current due to the pressure term in equation (14), it completely neglects any currents due to diffusion. The GNOR model expands the hydrodynamic theory to also take into account electron diffusion. We now consider the mathematical description of this effect. The inclusion of electron diffusion alters the continuity equation, which in its linearized form now reads

$$-\mathrm{i}\omega e n(\mathbf{r}, \omega) = D\nabla^2[en(\mathbf{r}, \omega)] + \boldsymbol{\nabla} \cdot [-en_0\mathbf{v}(\mathbf{r}, \omega)]$$
$$= \boldsymbol{\nabla} \cdot \mathbf{J}(\mathbf{r}, \omega), \tag{21}$$

also known as the convection-diffusion equation. Here, D is the diffusion constant, and the induced current density, given by Fick's law, now has a diffusive contribution

$$\mathbf{J}(\mathbf{r}, \omega) = -en_0\mathbf{v}(\mathbf{r}, \omega) + eD\boldsymbol{\nabla}n(\mathbf{r}, \omega). \tag{22}$$

Combining the convection-diffusion equation and Fick's law for the current density with the linearized form of the hydrodynamic equation (equation (14)) [69], we eventually arrive at the following constitutive relation for the current density

$$\left[\frac{\beta^2}{\omega(\omega + \mathrm{i}\gamma)} + \frac{D}{\mathrm{i}\omega}\right]\boldsymbol{\nabla}[\boldsymbol{\nabla} \cdot \mathbf{J}(\mathbf{r}, \omega)] + \mathbf{J}(\mathbf{r}, \omega) = \sigma\mathbf{J}(\mathbf{r}, \omega), \tag{23}$$

which we immediately recognize to have the same form as the hydrodynamic constitutive relation (equation (15b)). The difference lies in the prefactor of the first term, which we can rewrite as

$$\frac{\beta^2}{\omega(\omega + \mathrm{i}\gamma)} + \frac{D}{\mathrm{i}\omega} = \frac{\beta^2 + D(\gamma - \mathrm{i}\omega)}{\omega(\omega + \mathrm{i}\gamma)} \equiv \frac{\eta^2}{\omega(\omega + \mathrm{i}\gamma)}, \tag{24}$$

where we have defined the parameter

$$\eta^2 \equiv \beta^2 + D(\gamma - \mathrm{i}\omega). \tag{25}$$

Comparing equation (23) with equation (15b), we see that the mathematical considerations from the hydrodynamic model can be mapped directly to the GNOR model using the simple substitution $\beta^2 \to \eta^2$. Thus, we straightforwardly find the GNOR nonlocal parameter to be

$$\xi_{\mathrm{GNOR}}^2 = \frac{\varepsilon_{\mathrm{core}}(\omega)[\beta^2 + D(\gamma - \mathrm{i}\omega)]}{\omega(\omega + \mathrm{i}\gamma)}. \tag{26}$$

and similarly, the longitudinal wave vector in the GNOR model is $k_L^2 = \varepsilon(\omega)/\xi_{GNOR}^2$.

Considering the specific case with no interband effects and no bulk damping, we see that the diffusion constant only contributes to the imaginary part of ξ_{GNOR}^2, making the nonlocal parameter complex-valued (in contrast to the hydrodynamic model). Additionally, β^2 contributes only to the real part of ξ_{GNOR}^2 (as in the hydrodynamic model). In the general case, the hydrodynamic parameter $\beta \propto v_F$ (or the convective current) is still the main contributor to the real part of ξ_{GNOR}, while the imaginary part of ξ_{GNOR} is mostly characterized by the diffusion constant D (or the diffusive current). Table 1 lists values for D for relevant plasmonic metals.

A more elaborate description of the response of the free electrons in metals based on the Boltzmann equation has been reported by Lindhard [150], which includes Landau damping but not diffusive currents. Inclusion of diffusion in the Boltzmann description is possible by ensuring the perturbed electron density to relax to the local electron density [151] (rather than to the unperturbed electron density as done in the Lindhard description). This correction to the Boltzmann dielectric function was provided by Mermin [152]. Interestingly, Halevi [143] compared the dielectric function of the hydrodynamic model with the Boltzmann–Mermin dielectric function in the small k-limit and showed that

$$\eta_{Halevi}^2 = \frac{\frac{3}{5}\omega + i\frac{1}{3}\gamma}{\omega + i\gamma} v_F^2, \qquad (27)$$

to ensure agreement between the two models, which is in contrast to the usual $\eta = \beta$ in the hydrodynamic model. In particular, the result by Halevi shows that the Mermin correction (i.e. inclusion of diffusion in the Boltzmann equation) renders the usual hydrodynamic β parameter complex-valued, exactly as in the GNOR theory where the substitution $\beta^2 \to \eta^2$ is crucial. Comparison of the imaginary parts of equations (25) and (27) provides an estimate for the diffusion constant D as

$$D = \frac{4}{15} \frac{\gamma}{\omega^2 + \gamma^2} v_F^2. \qquad (28)$$

Inserting appropriate values for the different parameters in equation (28), see table 1, we find that $D \approx 10^{-6}\,\mathrm{m^2\,s^{-1}}$ at optical frequencies, which is approximately two orders of magnitude smaller than the values for D for nanoparticles shown in table 1. This discrepancy of values of the diffusion constant shows that the diffusion in plasmonic nanoparticles cannot be explained by bulk diffusion. On the theoretical side, it is important to see that inclusion of diffusion in the Boltzmann equation leads to a complex-valued η just as the inclusion of diffusion in the hydrodynamic model (i.e. GNOR theory) also results in a complex-valued η. Observed diffusion constants in nanoparticles are thus mainly due to surface effects (such as increased Landau damping), but we are not aware of a derivation based on the Boltzmann equation of surface diffusion with the correct magnitude.

5. Nonlocal effects in plasmonic systems

This section is devoted to studying some of the relevant plasmonic systems, which exhibit features due to nonlocal response. We begin in section 5.1 by giving an overview of the electromagnetic response of a single spherical nanoparticle, where the important length scale is the particle diameter. Next in section 5.2 we consider the plasmonic dimer, consisting in this case of two infinitely long cylinders, and study the dependence of the optical spectrum on the gap size. Finally in section 5.3, we also take a look at the interesting properties of a core-shell nanowire, consisting of an insulating core and a nanometer-sized metallic shell. We will mainly compare the GNOR model with the LRA, while comparison with the hydrodynamic model will also be displayed when relevant.

5.1. Spherical particle

The metallic spherical particle represents an archetypical geometry studied in plasmonics due to the support of localized surface plasmons and the presence of analytical solutions. The electromagnetic scattering problem of a metal sphere of radius R and Drude permittivity $\varepsilon(\omega)$, which is excited by a plane monochromatic wave and homogeneously embedded in a material with dielectric constant ε_B, was first analytically solved by Mie in the LRA [153]. The exact solution provided by Mie is commonly named Mie theory [154, 155] and takes into account retardation effects. Later, Ruppin [82, 83] extended the Mie theory to include nonlocal response in the metallic sphere by taking into account the longitudinal wave. In the simpler non-retarded limit, the multipolar polarizability of the metal sphere was extended to include nonlocal effects [137, 156–159] and used to study the optical properties of very small particles $R < 10\,\mathrm{nm}$, where retardation effects can for the most part be safely neglected.

In this section, we present the nonlocal dipolar polarizability, which in the LRA is described by the Clausius–Mossotti factor [29], to study the extinction cross section of spheres with diameters below 20 nm. In this size range and under the excitation of a plane wave, the effect of retardation is small and the response of the metal sphere is dominated by the dipolar mode, as we will see in the comparison with fully retarded calculations in section 5.1.3. By considering the poles of the nonlocal dipolar polarizability, we directly show that the nonlocal parameters β and D relate to the SP resonance energy and linewidth, respectively.

The derivation of the nonlocal dipolar polarizability α_{NL} is detailed elsewhere [16], so here we simply state the result for the non-retarded limit

$$\alpha_{NL} = 4\pi R^3 \frac{\varepsilon - \varepsilon_B (1 + \delta_{NL})}{\varepsilon + 2\varepsilon_B (1 + \delta_{NL})}, \qquad (29a)$$

where the nonlocal correction is given as

$$\delta_{NL} = \frac{\varepsilon - \varepsilon_{core}}{\varepsilon_{core}} \frac{j_1(k_L R)}{k_L R j_1'(k_L R)}. \qquad (29b)$$

Here, j_l is the spherical Bessel function of the first kind of angular-momentum order l, and the prime denotes differentiation with respect to the argument. In the LRA,

$\delta_{\text{NL}} \rightarrow 0$ and equation (29) simplifies to the LRA dipolar polarizability, described by the well-known Clausius–Mossotti factor $(\varepsilon - \varepsilon_{\text{B}})/(\varepsilon + 2\varepsilon_{\text{B}})$. We note that nonlocal effects enter the Clausius–Mossotti factor as an elegant and simple rescaling of either the metal permittivity [51] from ε to $\tilde{\varepsilon} = \varepsilon(1 + \delta_{\text{NL}})^{-1}$ or of the background permittivity [160] from ε_{B} to $\tilde{\varepsilon}_{\text{B}} = \varepsilon_{\text{B}}(1 + \delta_{\text{NL}})$.

The SP resonance energy follows theoretically from the Fröhlich condition, i.e., we must consider the poles of equation (29)

$$\varepsilon + 2\varepsilon_{\text{B}}(1 + \delta_{\text{NL}}) = 0, \tag{30}$$

which will be given by a complex-valued resonance frequency $\omega = \omega' + i\omega''$. The real part ω' gives the SP resonance frequency, while the imaginary part ω'' is related to the SP resonance linewidth. In the following analytical analysis based on the GNOR model, we consider for simplicity the case of a particle in vacuum ($\varepsilon_{\text{B}} = 1$) with no interband effects ($\varepsilon_{\text{core}} = 1$) and find (to first order in 1/R) [69]

$$\omega' = \frac{\omega_{\text{p}}}{\sqrt{3}} + \frac{\sqrt{2}\beta}{2R}, \tag{31a}$$

$$\omega'' = -\frac{\gamma}{2} - \frac{\sqrt{6}}{12}\frac{D\omega_{\text{p}}}{\beta R}. \tag{31b}$$

In equation (31a), the first term is the common size-independent local-response Drude result for the SP resonance that also follows from the poles of the LRA polarizability, and the second term gives the size-dependent blueshift due to the hydrodynamic pressure. In equation (31b), we have again the size-independent LRA term, while the second term shows a size-dependent linewidth due to diffusion. We can also clearly see that only a size-dependent blueshift is present in the hydrodynamic model, while the GNOR also accounts for a size-dependent SP linewidth. With the inclusion of interband effects, this clear distinction becomes somewhat blurred, since the hydrodynamic model will also show an extremely weak size-dependent linewidth. The origin to the size dependence in nonlocal response is from the smearing of the induced surface charges over a finite distance (few Å) [88, 90, 97, 161, 162]. Yan [97] showed explicitly how the smearing of induced charges into the metal leads to a size-dependent blueshift in the hydrodynamic model. In contrast, the delta-function behaviour of the induced surface charges in the LRA leads to no dependence on size. We point out that a 1/R dependence on the blueshift and the linewidth of the SP resonance energy of small Ag nanoparticles has been experimentally observed using optical spectroscopy [47, 50, 53, 54, 163–166].

With the nonlocal polarizability we can determine the extinction cross section C_{ext}, which is the sum of the scattering and absorption cross sections, of a metal sphere using the relation [155]

$$C_{\text{ext}} = \frac{k_{\text{B}}^4}{6\pi}|\alpha_{\text{NL}}|^2 + k_{\text{B}}\text{Im}(\alpha_{\text{NL}}), \tag{32}$$

where $k_{\text{B}} = (\omega/c)\sqrt{\varepsilon_{\text{B}}}$ is the wave vector in the homogeneous background dielectric medium. Equation (32) allows for a more quantitative assessment, based on an observable quantity, of the size-dependent blueshift and linewidth broadening of
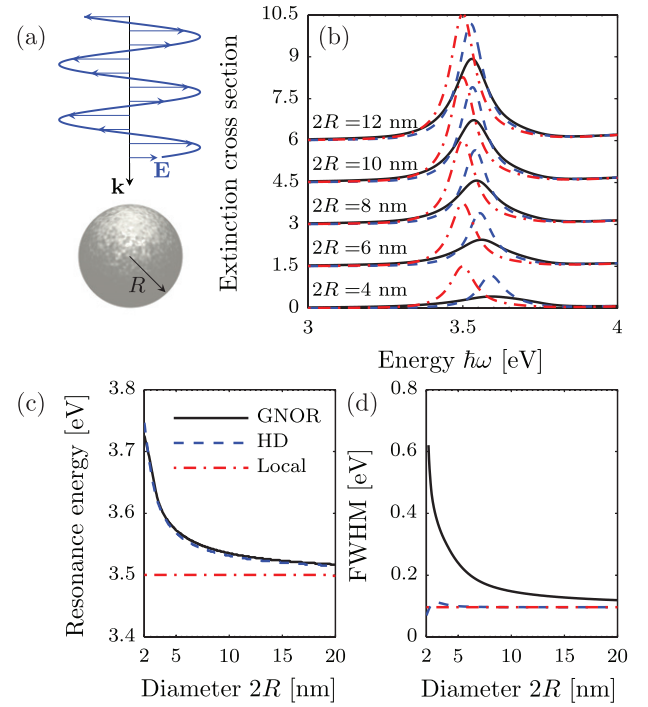
Figure 2. (a) Sketch of a plane wave impinging on a silver sphere of radius R. (b) Extinction cross section (in units of the geometrical cross section πR^2 of the sphere) of a silver sphere in vacuum for decreasing sphere diameter calculated using the GNOR model (black solid lines), hydrodynamic model (blue dashed lines) and LRA (red dash-dotted lines). For clarity, each spectrum is displaced vertically by 1.5 normalized unit. (c)–(d) Resonance energy and full-width at half maximum (FWHM) of the dipole mode of a silver sphere as a function of diameter 2R, respectively.

the dipolar SP resonance anticipated from the approximate analytical relations in equation (31). In figure 2(b) we show the extinction cross section of a small Ag sphere, sketched in figure 2(a), with diameter varying from 4 nm to 12 nm, calculated in the GNOR theory (black line), hydrodynamic model (blue line) and LRA (red line). As expected, the LRA shows no change in resonance energy or linewidth of the dipolar SP with size. The hydrodynamic model shows a blueshift of the SP with decreasing diameter and a slightly smaller SP amplitude as a consequence of small, but finite, surface absorption due to bulk material losses and interband transitions. The GNOR model shows both blueshift and linewidth broadening of the SP resonance with decreasing particle size. Quantitatively, we see in figures 2(c) and (d) that a blueshift of ∼0.2 eV and an increased linewidth broadening of ∼0.5 eV is seen when the sphere diameter decreases from 20 nm to 2 nm. We note that the size-dependent blueshift exceeds the 1/R-dependency, given by the analytical relations in equation (31), for diameters below 10 nm and higher-order terms become important [15]. The hydrodynamic model shows the same blueshift of the SP as the GNOR model, but no significant increase in linewidth. In particular, a weak size-dependent linewidth in the hydrodynamic model can be seen in figure 2(d) for diameters below 5 nm, which is due to the inclusion of bulk losses (i.e. $\gamma \neq 0$) and interband effects.
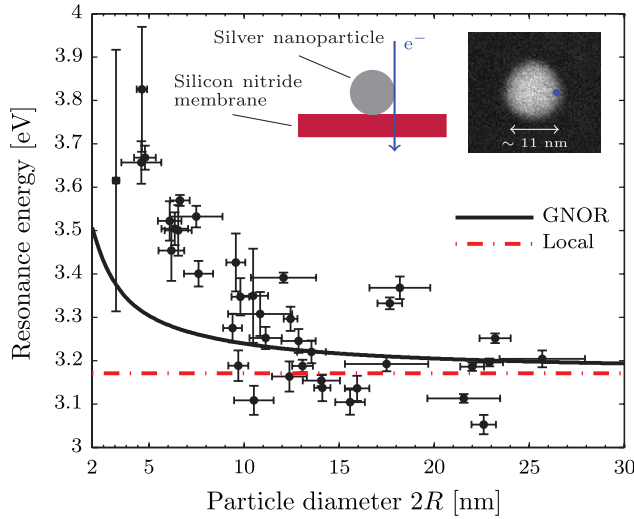
Figure 3. Resonance energy as function of particle diameter $2R$ of silver nanoparticles dispersed on a 10 nm silicon nitride membrane measured with EELS. The EELS data is collected by positioning the electron beam at the surface of the silver particle to strongly excite surface plasmons (see insets) [15]. The curves show calculations of a silver sphere in a homogeneous background using equation (29) in the GNOR model (black solid line) and LRA (red dash-dotted line). From the average resonance energy of the largest particles $2R > 20$ nm, we fit $\varepsilon_B = 1.98$.

5.1.1. Size-dependent resonance shift.

The size-dependent shift of the SP resonance energy of small metal particles has been experimentally observed on several occasions [13, 47, 53, 54, 163, 167, 168]. It has been found that alkali metals, such as potassium, redshift with decreasing particle size, while noble metals, such as silver and gold, blueshift with decreasing particle size. While the redshift in alkali metals is explained to be a consequence of the spill-out of free electrons, the blueshift in noble metals is mainly attributed to the screening from lower-lying band electrons, such as the d-band in silver [169], although additional sources for the blueshift are also present [18, 163, 170].

Recently, EELS measurements on silver nanoparticles dispersed on thin ($\leqslant 10$ nm) substrates have shown a strong blueshift of the SP resonance energy, when the particle diameter decreases from approximately 25 nm down to 2 nm [14–16]. In figure 3 we show EELS measurements of the resonance energy of silver nanoparticles dispersed on a 10 nm silicon nitride membrane as a function of diameter [15]. We see that the SP resonance energy shows a significant increase from approximately 3.2–3.7 eV in the diameter range considered. Additionally, we show in figure 3 calculations of the SP resonance energy using the dipolar polarizability (equation (29)) in the GNOR model (black line) and in the LRA (red line). As a crude approximation we may account for the effect of the substrate by incorporating it into an effective homogeneous background permittivity ε_B. In this approximation, one uses the average resonance energy of the largest particles $2R > 20$ nm to fit the background permittivity ε_B to ensure agreement with the LRA in the large-diameter range [14, 15]. The assumption behind this approach is that the effective background permittivity does
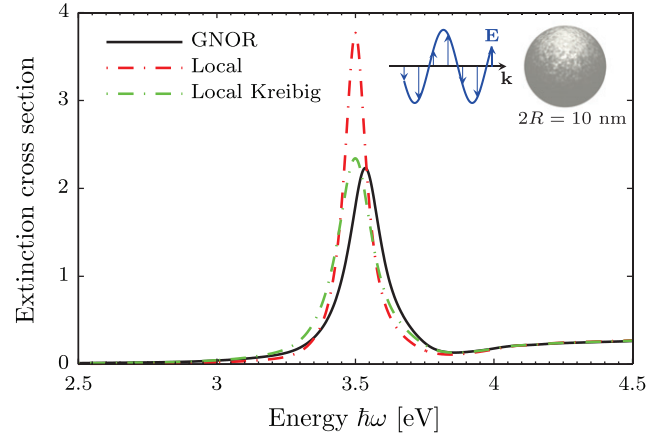
Figure 4. Extinction cross section (in units of the geometrical cross section πR^2 of the sphere) of a silver sphere in vacuum with diameter $2R = 10$ nm as a function of energy $\hbar\omega$, calculated within the GNOR (black solid line), LRA (red dash-dotted line), and Kreibig (green dash-dotted line) approaches, see section 5.1.2. The value $A = 0.5$ has been used for the Kreibig approach.

not depend on the radius of the particle, but more detailed theoretical calculations by Yan [97] show variations of 20% for particle radii varying from 2 to 18 nm. As expected from the discussion of figure 2, the SP resonance energy in the LRA is size-independent, while the GNOR model shows a blueshift, which is in qualitative agreement with the experiments. However, the experimentally-measured blueshift is larger. This difference was initially conjectured to be related to the inaccurate modeling of the substrate [15], however, a proper inclusion of the electromagnetic effects of the thin substrate has interestingly been shown not to be able to account for the discrepancy [16]. Instead the explanation for the discrepancy may be related to more complicated phenomena in silver, such as an inhomogeneous equilibrium electron density due to Friedel oscillations and electron spill-out [15] or changes in the electronic band structure [163].

5.1.2. Size-dependent damping.

The phenomenon of size-dependent damping in metal nanoparticles has been extensively observed in experiments [47, 48, 50, 163, 171–173]. The theoretical approach to account for this effect in the LRA, proposed by Kreibig [50] and adopted widely by researchers in the field [174], has been to phenomenologically modify the Drude bulk damping parameter γ as

$$\gamma \to \gamma + A \frac{v_F}{R}, \tag{33}$$

which is only valid for spherical particles of radius R. Here, A is a constant, which is related to the probability of the free electrons scattering off the surface of the particle. Experimental observations and advanced theoretical calculations have been compared to this approach, resulting in most cases in a value for A close to unity [48, 51, 175, 176]. In the following, we denote the method described in equation (33) as the Kreibig approach. In figure 4, we compare the extinction cross section of a silver sphere with diameter $2R = 10$ nm, calculated within the GNOR (black line), LRA (red line), and

Kreibig (green line) approaches. As we see from figure 4 the Kreibig approach displays a size-dependent broadening as in the GNOR theory (in contrast to the LRA), but no size-dependent resonance shift (in agreement with the LRA). The SP linewidths are practically the same in the GNOR and Kreibig calculations due to the chosen value for D.

The Kreibig approach has been quite successful in describing the increased linewidth of SPs for different metals and geometries using a value for A that is close to unity. It is therefore desirable to ensure that the value for the diffusion constant D in the GNOR model gives rise to the same SP linewidth broadening as with the Kreibig approach. An estimate for D can be given by comparing the size-dependent term in equation (31*b*) with the size-dependent term using the Kreibig approach, i.e. $\omega_k'' \simeq -\gamma/2 - Av_F/(2R)$. Here we find the simple relation

$$D = \frac{3\sqrt{10}}{5} A \frac{v_F^2}{\omega_p} \propto An_0^{\frac{1}{6}}, \qquad (34)$$

which shows a linear scaling with the Kreibig parameter A and a weak scaling with the equilibrium electron density n_0. Using appropriate material values, equation (34) provides a quite accurate estimate for D (to within a factor of 2 for simple metals) compared to more thorough numerical investigations [177] for the plasmonic metals listed in table 1. However, due to the lack of inclusion of the bound electrons (i.e. $\varepsilon_{\text{core}} \neq 1$) in equation (34), metals with similar plasma frequency and Fermi velocity, such as gold and silver, give rise to the same value for D, which is only approximately correct. We have therefore used a numerical routine described by Raza [177] to ensure that the value for D in the GNOR model gives rise to identical SP linewidth broadening for a spherical particle as the Kreibig approach. The values for D (appropriate for $A = 0.5$ and $A = 1$) using the numerical routine are summarized in table 1 for several plasmonic metals.

From equation (33), it is clear that the Kreibig approach is only a correction to first order in $1/R$. In contrast, the GNOR model contains corrections of order $1/R$, $1/R^2$, and so on. Thus, a signature of the GNOR model could be to find the linewidth broadening of the SP resonance to exceed the $1/R$-dependency given by the Kreibig approach. Such measurements could also be used to determine the appropriate value for D. Finally, we stress that size-dependent damping in non-spherical particles is also described by the GNOR model.

5.1.3. Retardation effects. In the previous sections, we have discussed the nonlocal optical response of a metal spherical particle in the nonretarded approximation, thus neglecting retardation effects. Here, we discuss the validity of the nonretarded approximation and the importance of retardation effects for different metals. One of the advantages of describing nonlocal response in the hydrodynamic approach is the ability to take into account retardation effects (in contrast to *ab initio* calculations such as DFT) [68, 75]. As an example, the retarded multipolar response of a sphere including nonlocal response has been determined by Ruppin [82, 83] by extending Mie theory to take into account longitudinal waves. In the
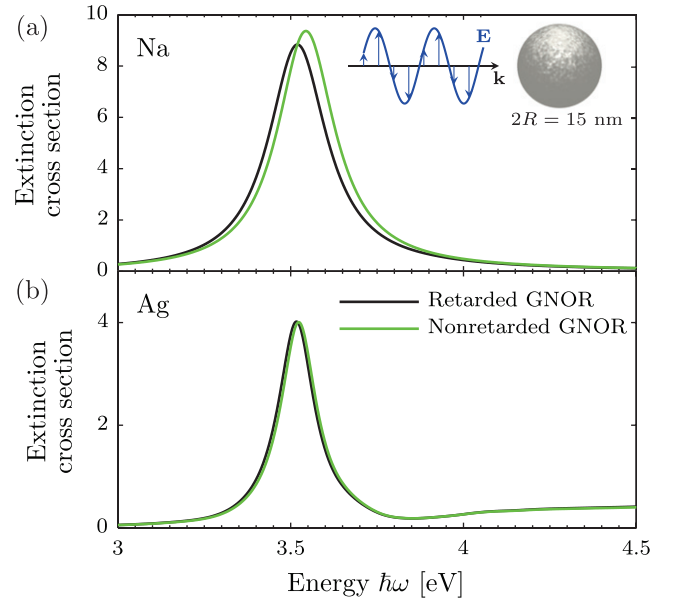
Figure 5. (*a*) Extinction cross section (in units of the geometrical cross section πR^2 of the sphere) of a sodium sphere in vacuum with diameter $2R = 15$ nm as a function of energy $\hbar\omega$, calculated within the GNOR model using the fully-retarded multipolar response (black line) and the nonretarded dipolar response (green line). (*b*) Same as (*a*) but for a silver sphere.

retarded framework, the extinction cross section of a sphere in a homogeneous background can be determined using the relation [82, 83, 155]

$$C_{\text{ext}} = -\frac{2\pi}{k_B^2} \sum_{l=1}^{\infty} (2l + 1) \operatorname{Re}\left(t_l^{\text{TE}} + t_l^{\text{TM}}\right), \qquad (35)$$

where l denotes the angular momentum. Here, the nonlocal Mie scattering coefficients are given as [82, 83, 106, 128]

$$t_l^{\text{TE}} = \frac{-j_l(x_T)[x_B j_l(x_B)]' + j_l(x_B)[x_T j_l(x_T)]'}{j_l(x_T)[x_B h_l^{(1)}(x_B)]' - h_l^{(1)}(x_B)[x_T j_l(x_T)]'}, \qquad (36a)$$

$$t_l^{\text{TM}} = \frac{-\varepsilon j_l(x_T)[x_B j_l(x_B)]' + \varepsilon_B j_l(x_B)\{[x_T j_l(x_T)]' + \Delta_l\}}{\varepsilon j_l(x_T)[x_B h_l^{(1)}(x_B)]' - \varepsilon_B h_l^{(1)}(x_B)\{[x_T j_l(x_T)]' + \Delta_l\}}, \qquad (36b)$$

where $x_B = k_B R$, $x_T = k_T R$, and $h_l^{(1)}$ denotes the spherical Hankel function of the first kind. The nonlocal correction Δ_l to the Mie coefficients is given as

$$\Delta_l = l(l + 1) j_l(x_T) \frac{\varepsilon - \varepsilon_\infty}{\varepsilon_\infty} \frac{j_l(x_L)}{x_L j_l'(x_L)}, \qquad (36c)$$

with $x_L = k_L R$. We note that for $l = 1$ (dipole mode) the nonlocal correction in equation (36*c*) has a similar form as δ_{NL} in the nonlocal Clausius–Mossotti factor, equation (29) [16].

To examine the influence of retardation, we show in figure 5 the extinction cross section of a metal sphere in the GNOR model with a diameter $2R = 15$ nm calculated with retardation (black line) and in the nonretarded limit (green line). In particular, figure 5(*a*) shows the results for a sodium sphere, which is a prototypical metal studied with DFT [65, 66, 131, 132, 178]. Here, we see that retardation

effects already set in and give rise to a redshift of the SP resonance energy and a decrease in the extinction cross section amplitude for $R = 7.5$ nm. Interestingly, this shows that retardation is important for individual nanoparticles at a size scale where nonlocal response still plays a role (see figure 2) [179]. Retardation becomes even more paramount in larger structures such as the dimer, which we discuss in section 5.2. In contrast to this result on the sodium sphere, retardation shows no significant influence on a silver sphere of the same size, see figure 5(b). For silver spheres, larger diameters must be considered ($2R > 20$ nm) before retardation effects show up, which justifies our nonretarded approach in the previous sections. The validity of the nonretarded limit for the silver particle is extended to larger sizes than the sodium particle due to the permittivity of silver. This interplay between material permittivity and particle size can be clearly seen from the criterion of validity for the quasistatic approximation $\lambda \gg 2\pi R|\sqrt{\varepsilon}|$, where λ denotes the free-space wavelength [155].

5.1.4. Multipolar response. We have so far focused on nonlocal effects on the dipole mode of a small sphere and shown that this mode dominates the extinction spectrum for diameters below 20 nm. However, other excitation sources, such as a swift electron (used in EELS measurements), produce significantly more inhomogeneous electric-field distributions than the plane-wave excitation used in determining the extinction cross section [128, 138]. When in the vicinity of a metal sphere, such sources can excite higher-order multipoles, even in spheres with diameters below 20 nm [128, 180]. Thus, it is relevant to consider the nonlocal resonance condition for all multipoles given in the nonretarded limit as [128]

$$l\varepsilon + \left(l + 1 + \Delta_l^{\text{nr}}\right)\varepsilon_{\text{B}} = 0, \tag{37a}$$

where Δ_l^{nr} is the nonretarded limit of Δ_l of equation (36c), given as

$$\Delta_l^{\text{nr}} = l(l+1)\frac{\varepsilon - \varepsilon_{\text{core}}}{\varepsilon_{\text{core}}}\frac{j_l(x_{\text{L}})}{x_{\text{L}} j_l'(x_{\text{L}})}. \tag{37b}$$

We see that for $l = 1$, equation (37) reduces to the dipole resonance condition stated in equation (30). As with the condition for the dipole resonance given in equation (30), we can determine an approximate solution to equation (37) in the simple case of $\varepsilon_{\text{B}} = \varepsilon_{\text{core}} = 1$. We then find the complex resonance frequencies $\omega_l = \omega_l' + i\omega_l''$ to order $1/R$ to be

$$\omega_l' = \frac{\omega_{\text{p}}}{\sqrt{1 + (l+1)/l}} + \sqrt{l(l+1)}\frac{\beta}{2R}, \tag{38a}$$

$$\omega_l'' = -\frac{\gamma}{2} - \frac{l\sqrt{l+1}}{4\sqrt{2l+1}}\frac{D\omega_{\text{p}}}{\beta R} \tag{38b}$$

which shows a clear dependence on the angular momentum l. Specifically, equation (38) shows that the size-dependent resonance shift and linewidth broadening increases with higher order of angular momentum. This means that e.g. the quadrupolar mode of the sphere experiences a larger blueshift and linewidth broadening due to nonlocality than the dipolar mode. Higher-order modes have electric fields which are more
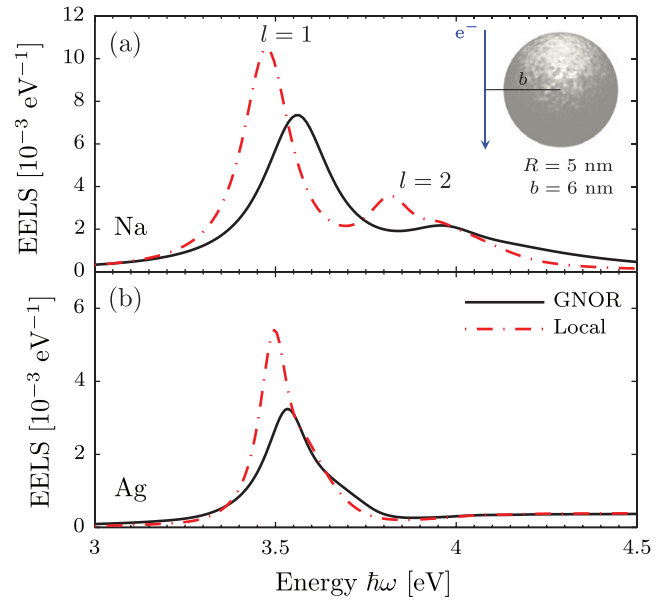
Figure 6. (*a*) Electron energy-loss probability of a sodium sphere in vacuum with diameter $2R = 10$ nm as a function of energy $\hbar\omega$, calculated within the GNOR model (black line) and in the LRA (red dash-dotted line). (*b*) Same as (*a*) but for a silver sphere. The electron beam follows a straight-line trajectory with a distance of $b - R = 1$ nm from the surface of the particle (see the schematic inset). The kinetic energy of the electron is 120 keV. The calculation is based on code by Christensen [128].

strongly localized to the surface of the particle, and therefore are more affected by the nonlocal smearing of the induced charges. Yan [97] provided the first theoretical observation and explanation of a l-dependent blueshift in the hydrodynamic model, while the l-dependent broadening, which is inherent for the GNOR model, is discussed here for the first time.

To exemplify the importance of higher-order modes and the l-dependent behavior of nonlocal response anticipated from equation (38), we consider in figure 6 the electron energy-loss (EEL) probability [43, 128] of a swift electron following a straight-line trajectory near the surface of a spherical particle with diameter $2R = 10$ nm. The EEL probability is directly comparable to the signal measured in EELS experiments. In figure 6(*a*), we consider the EEL probability from a sodium sphere, where the low-energy peak is due to the excitation of the dipole mode ($l = 1$) while the high-energy peak is due to the excitation of the quadrupole mode ($l = 2$) in both the GNOR (black line) and LRA (red line) calculations. As observed in the extinction cross section, the dipole mode blueshifts and broadens in the GNOR model. Interestingly, we see that the quadrupole mode experiences a larger blueshift and linewidth broadening than the dipole mode, in accordance with our discussion from the approximate relation in equation (38). In figure 6(*b*) a silver sphere is considered, where the presence of strong interband losses unfortunately dampens the higher-order modes, leaving only a resonance peak due to the excitation of the dipole mode. Other metals (in particular, Al) or settings where nonlocal response effects on higher-order plasmonic modes could be seen were discussed by Christensen [128].
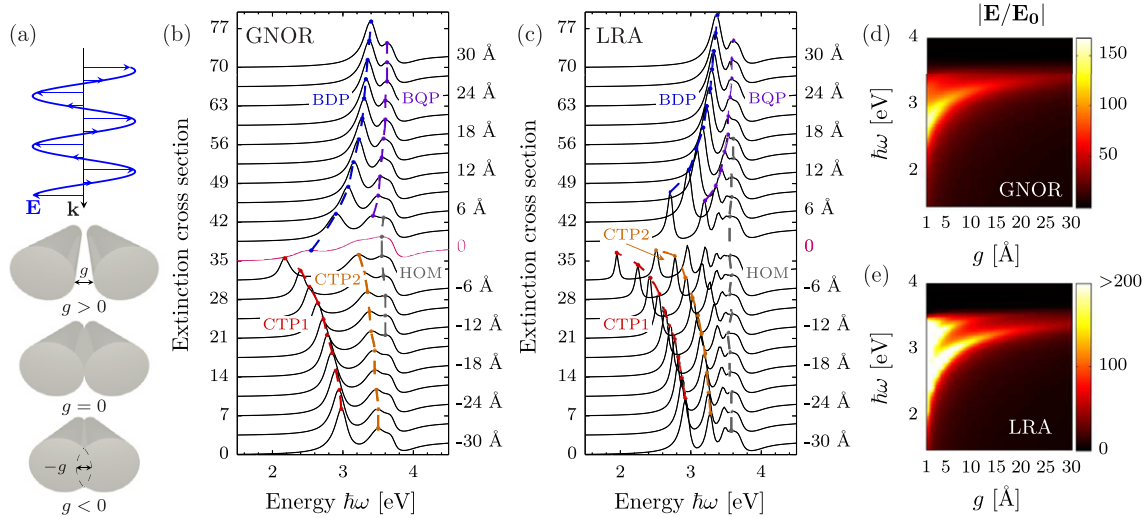
Figure 7. (*a*) Sketch of an incident plane wave, which is polarized along the dimer axis, impinging on a dimer consisting of two identical silver cylinders with radii R and gap size g. (*b*)–(*c*) Extinction cross section (in units of the cylinder diameter $2R$) of a silver dimer in vacuum with radius $R = 15$ nm for gap sizes varying from $g = 30$ Å (separated) to $g = -30$ Å (overlapping) in steps of 3 Å calculated using the GNOR model and in the LRA, respectively. For clarity, each spectrum is displaced vertically by 3.5 normalized units. (*d*)–(*e*) Electric-field enhancement in the center of the dimer gap for dimers with $R = 15$ nm as a function of energy $\hbar\omega$ and gap size g calculated using the GNOR model and in the LRA, respectively.

5.2. Cylindrical dimer

The plasmonic dimer, which consists of two metallic nanoparticles in close proximity, has attracted a lot of interest due to the plasmon hybridization [181, 182] occurring between the two closely-spaced nanoparticles, which gives rise to strongly gap-dependent resonance energies and electric-field enhancements [183]. Such features have been utilized in e.g. surface-enhanced Raman spectroscopy [184] and the plasmon ruler effect [185]. The dimer has been subject to intense theoretical and experimental studies. In the simple LRA, diverging field enhancements in the gap of the dimer are encountered in the extreme case of touching dimers (i.e. no gap), which sets no limit to the number of hybridized plasmon modes [177], thereby exciting an continuum of modes [92, 93, 186]. These unphysical attributes of the LRA are corrected in DFT [65, 67, 131, 132, 178] and hydrodynamic [88, 92, 93] approaches due to the inclusion of nonlocal response and electron spill-out (only DFT). Recent measurements on dimers with subnanometer gaps using both optical techniques [19, 22, 23, 25] and EELS [18, 187] show lack of agreement with the LRA, and, in the touching case, also display limits on the resonance energies of the bonding plasmon modes. However, due to the indirect nature of the measurements the explanation for the discrepancy between LRA and the observed measurements is not conclusive with possible interpretations being provided from both quantum tunneling [65, 131] and nonlocal response [69] perspectives.

In this section, we consider a specific dimer geometry consisting of two identical silver cylinders and vary the gap from separated via touching to overlapping configurations, see figure 7(*a*) for a schematic illustration. The dimer consists of cylinders with radii of $R = 15$ nm and is excited by a plane wave which is polarized along the dimer axis to strongly excite the hybridized modes. We numerically calculate the extinction

cross section and the field enhancement in the gap of the dimer in the frameworks of the LRA and the GNOR model by using the freely available COMSOL implementation of the nonlocal equations (www.nanopl.org) [88].

To clearly convey the results of the optical spectra in figures 7(*b*) and (*c*), we first discuss the plasmon hybridization occurring in the dimer system. When two cylinders are positioned in close proximity, their modes hybridize to form new plasmon modes, which can show up as resonances in the extinction cross section [182]. For large separation distances, the first modes to hybridize are the individual dipolar modes (angular momentum $l = 1$) of the cylinders to produce a lower-energy (with respect to the individual dipolar mode) bonding dipolar mode (BDP) and a higher-energy antibonding dipolar mode. Since the net dipole moment of the antibonding mode is zero, this mode will be optically dark and not show up in the extinction cross section. We therefore leave out further discussion of the antibonding modes. As the dimer separation decreases the plasmon hybridization increases and the resonance energy of the bonding dipolar mode decreases. Furthermore, with decreasing separation distance higher-order modes of the individual cylinders (i.e. $l > 1$) begin to hybridize as well. As an example the quadrupole mode of the individual cylinders hybridize to form bonding and antibonding quadrupole modes. Thus, in nanometer-proximity the dimer spectra can be quite complex and show multiple resonances due to the hybridization between many modes of the individual cylinders.

With this plasmon hybridization picture in mind, we consider now in detail the extinction cross section of a silver dimer in the LRA and GNOR theory, see figures 7(*b*) and (*c*), respectively. We vary the dimer gap from $g = 30$ Å (separated dimer) via touching configuration to $g = -30$ Å (overlapping dimer). We begin our discussion by considering the LRA results (figure 7(*c*)) for a dimer separated by a gap

of 30 Å. The lowest-energy and strongest peak is due to the BDP, while the next peak is actually due to two spectrally-close modes, the bonding quadrupole mode (BQP) and a higher-order mode (HOM). Of these two modes, the BQP has the largest amplitude and lowest energy. The electric field of the HOM is concentrated at the edges of the dimer (and not in the gap, like the BDP and BQP), making it spectrally insensitive to the gap size. As the gap decreases, the BDP and BQP redshift and additional bonding-mode resonances appear due to the increased plasmon hybridization. In fact as $g \rightarrow 0$ the bonding modes continue to redshift and the number of bonding modes increases without bound till a continuum of modes is found in the touching configuration $g = 0$ [92, 93]. The extinction cross section calculation for $g = 0$ does not converge in the LRA, which is why the spectrum is not present in figure 7(c). When the dimers overlap the nature of the resonant modes changes and can no longer be considered as bonding modes. In particular, the induced charges pile up at the sharp edges of the junction of the overlapping dimer [186]. The interaction between the induced charges gives rise to the several resonances seen for e.g. $g = -6$ Å, which are denoted charge transfer plasmons (CTPs). As the overlap increases, the sharp edges at the junction smoothen and the interaction between the surface charges decreases, leading to a blueshift of the resonances. The spectrum of the overlapping dimer begins to increasingly resemble that of an elongated particle.

Turning our attention now to the results of the GNOR model in figure 7(b), we find that the separated dimers show less redshift with decreasing gap size compared to the results in the LRA. The size-dependent blueshift and linewidth broadening of the plasmon resonances observed in a spherical nanoparticle (see section 5.1) translates into a *gap*-dependent blueshift and linewidth broadening for the bonding modes of the dimer. The gap-dependent blueshift counteracts the redshift due to plasmon hybridization, leading to a finite resonance energy and finite number of bonding modes in the touching limit $g = 0$ (pink curve in figure 7(b)). Furthermore, the increase in blueshift and linewidth broadening with angular momentum observed for the spherical particle (see section 5.1.4) manifests itself in the dimer spectra as an increased blueshift and broadening for higher-order bonding modes. Thus, the BQP experiences a stronger blueshift and broadening than the BDP, leading to a weaker plasmon hybridization for the BQP than the BDP. This is also the reason for the lower number of resonances in the GNOR spectra of close-proximity and short-overlap dimers compared to those of the LRA. For the overlapping dimers ($g < 0$) the strength of the resonant modes is weaker in the GNOR calculations than in the LRA due to the nonlocal smearing of the surface charges at the geometrically sharp edges of the dimer junction.

Besides examining the extinction cross section of the dimer, we also study the electric-field enhancement present in the center of the dimer gap in figures 7(d) and (e). Here, we find that the GNOR model shows a significant decrease in the electric-field enhancement compared to the LRA [88]. While the field enhancement in the LRA increases without bound with decreasing gap size [186], the GNOR model shows an amplitude- and frequency-dependence of the field

enhancement in agreement with DFT simulations [65, 131, 177] and other models based on quantum tunneling [188–191] for gaps above 5 Å.

Many of the recent experimental observations on dimers in subnanometer proximity have been interpreted in a quantum tunneling framework [18, 19, 25] by comparing measured spectra with theoretical spectra simulated using DFT [65] (or other quantum tunneling based models [188, 189]). But as discussed in this section, the GNOR model (which does not take into account electron spill-out) produces very similar far-field spectra (e.g. extinction cross section) as DFT-based modeling. Thus, the experimental spectra could easily also be interpreted as signatures of nonlocal response. To actually distinguish the GNOR model from quantum tunneling models (or in general theoretical models which include electron spill-out and the overlap of such), one is required to study the extreme near-field, such as the electric-field enhancement, to positively separate the effects. As far as we know, only the experimental observations of Zhu [22] and Hajisalem [23] have shown signs of a decrease in the electric-field enhancement when the dimer separation is smaller than 5 Å, which seems to be in agreement with the onset of the overlap of electron spill-out [131]. At these narrow gap sizes, where the overlap of electron spill-out is significant, the approximation of a hard-wall boundary condition in the GNOR model is no longer accurate, thereby setting a limit for the applicability of the model (approximately 5 Å for a vacuum gap).

5.3. Core-shell nanowire

By modifying the structure of the metal nanoparticle to have a dielectric core with a metal shell, an increased tunability of the localized SP resonances (LSPRs) is achieved due to the plasmon hybridization of the inner and outer surfaces of the metal. Especially the spherical core-shell structure has received a considerable amount of attention [192–195] due to its excellent and tunable sensing properties [196], which have been utilized in biological studies such as cancer therapy [197]. The plasmon hybridization allows one to position the LSPR of the nanoshell as desired by simply varying the core size R_1 and/or outer radius R_2 appropriately [198].

The hybridization of the inner and outer surface plasmons increases when the metal shell becomes thinner [198], which gives rise to significantly altered LSPRs compared to usual homogeneous metal nanoparticles, such as the sphere in section 5.1. In section 5.2 we showed that the effect of nonlocal response increases with decreasing gap size (i.e. increasing hybridization). We would therefore anticipate a strong signature of nonlocal response in the core-shell particle, since it features an ultra-thin metallic shell with resulting strong plasmon hybridization.

To study the impact of nonlocal effects in the core-shell geometry, we consider an infinite cylindrical nanowire with a dielectric core and a thin metal shell excited by a plane wave, see figure 8(a) for an illustration. By extending the nonlocal Mie theory for cylinders [199] to core-shell structures, we can analytically determine the extinction cross section taking into account nonlocal response in the thin metal shell and retardation effects [98].
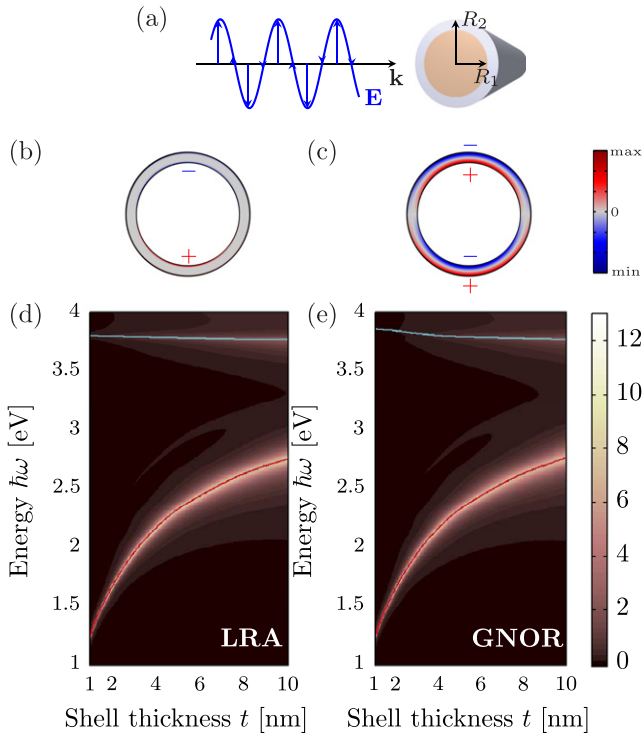
Figure 8. (*a*) Sketch of a plane wave impinging on a core-shell cylinder of inner radius R_1 and outer radius R_2. The permittivity of the core is $\varepsilon_{core} = 1.5^2$, applicable for silica, and the metal shell is silver. (*b*)–(*c*) GNOR result of the imaginary part of the induced charge density of a silica-silver cylinder with $R_1 = 15$ nm and $R_2 = 18$ nm at the resonance energies $\hbar\omega = 1.96$ eV and $\hbar\omega = 3.81$ eV corresponding to the bonding and antibonding dipole modes, respectively. (*d*)–(*e*) Color plots of the extinction cross section (in units of outer cylinder diameter $2R_2$) of a silica-silver cylinder in vacuum as a function of energy $\hbar\omega$ and shell thickness $t = R_2 - R_1$ calculated using the LRA and GNOR models, respectively. The inner radius is $R_1 = 15$ nm.

We focus on a particular design, consisting of an insulating core with permittivity $\varepsilon_{core} = 1.5^2$ corresponding to silica, inner radius of $R_1 = 15$ nm, and a thin silver shell. We are interested in studying the optical response of the core-shell nanowire when varying the shell thickness $t = R_2 - R_1$ by changing the outer radius R_2. Figures 8(*d*) and (*e*) shows the extinction cross section as a function of energy and shell thickness in the LRA and GNOR model, respectively. Considering first the LRA result, we see that the extinction cross section is dominated by two resonances: a strong low-energy resonance due to the bonding dipole mode (red line) and a weaker high-energy resonance due to the antibonding dipole mode (blue line) [198]. The bonding mode redshifts strongly for decreasing shell thickness, while the antibonding mode blueshifts only slightly. Both shifts are due to the increase in plasmon hybridization with decreasing shell thickness. We note that a strong tunability of the bonding mode with shell thickness is present, allowing for tailoring of the optical response. When we consider the result from the GNOR theory (figure 8(*e*)), we find quite surprisingly the same optical response as in the LRA for the bonding mode (in stark contrast to the dimer geometry in section 5.2). In particular, we see no significant size-dependent resonance shift or linewidth

broadening for the bonding mode as encountered for the sphere and dimer geometries, even in the extreme case of a 1 nm thin shell. However, the antibonding mode shows an increased blueshift and a size-dependent broadening effect due to nonlocal response.

To find an answer to this surprising lack of presence of nonlocal response in the bonding mode, we consider in more detail the plasmon hybridization occurring in the core-shell geometry. In particular, we are interested in how the induced charges of the bonding and antibonding dipole modes in figures 8(*d*) and (*e*) are distributed. Figure 8(*b*) displays the induced surface charge distribution of the bonding dipole mode of the core-shell cylinder, which shows that the negative and positive induced charges are isolated to each side of the cylinder, thus separated by a distance of approximately the inner cylinder diameter, i.e. $2R_1 = 30$ nm for the design considered in figure 8. The smearing of induced charges over Ångstrom length scales due to nonlocal response will therefore not play a significant role on the optical response of the bonding plasmon mode, since the positive and negative induced charges are separated by much greater distances than the smearing length scale. In contrast, the strong effect of nonlocal response in spheres and dimers occur due to the induced positive and negative charges coming in close proximity when the particle diameter and dimer gap, respectively, are decreased. However, the antibonding mode of the core-shell geometry has induced positive and negative charges on each side of the thin metal shell (figure 8(*c*)), which explains why nonlocal effects play a prominent role for this mode when the metal shell is sufficiently thin.

6. Outlook

6.1. Electron spill-out effect

The nonlocal hydrodynamic and the GNOR models are the natural immediate extension to the usual Drude model for the theoretical modeling of metals. We emphasize that the difference between the nonlocal response models (that is, with hard-wall ABC and homogeneous equilibrium electron density n_0) and the LRA is how we model the induced charges, i.e. the charges occurring due to an exciting electric field. In the LRA the induced charges reside only on the geometric surface of the metal structure, while the inclusion of nonlocal response serves to smear out the induced surface charges on the Ångstrom length scale. The main shortcoming of the nonlocal models is the inaccurate treatment of the free electrons at the metal surface in the absence of an exciting electric field, i.e. the ground-state equilibrium electron density. As discussed, the free electron density is modeled as being constant inside the metal and then abruptly dropping to zero outside the metal (i.e. step profile). From the pioneering work on density-functional theory by Lang and Kohn [200], we know that the equilibrium electron density should be smoothly varying at the metal-vacuum interface, with Friedel oscillations inside the metal and electron spill-out just outside the metal [58]. The strength of current DFT treatments of metals is the inclusion of a self-consistent treatment of the equilibrium electron density.

Studying the effects of a spatially-varying electron density in the LRA has been performed [201], however with the neglect of nonlocal gradient effects [202]. Recently, David [203], Toscano [179] and Yan [204] have shown that it is also possible to take into account a smoothly-varying equilibrium electron density in the hydrodynamic model, thus overcoming the discussed limitations of the hard-wall nonlocal models. Expanding this approach to include diffusion should not pose significant complications, but has yet to be done.

6.2. Metamaterials

Metamaterials are man-made engineered materials that can manipulate and mold electromagnetic waves in ways that are not attainable with naturally available materials [205]. A few examples of the functionalities that have been attained with metamaterials include the cloaking of macroscopic objects and negative refraction [32, 206–208]. Such properties have been designed using the theory of homogenization. Effects due to nonlocal response could come into play for optical metamaterials on two levels: firstly, in the response of the individual elements (metaatoms) as their overall sizes have to be considerably smaller than the wavelength of light, and secondly, in the arrangement of these elements at deeply subwavelength distances which enhances their nonlocal electromagnetic interaction.

While the homogenization procedure is quite complicated even in the LRA [209], some metamaterials have shown to be able to exhibit effects due to spatial dispersion stemming from the homogenization procedure (and not from the individual metaatoms) [210–214], yet the connection to the description of the intrinsic spatial dispersion found in metals has not been made. It would be beneficial to be able to use the real-space constitutive relations derived in sections 3 and 4 to also describe metamaterials with artificial spatial dispersion. This may allow a pathway for engineering spatial dispersion using metamaterials on a length scale significantly larger than the nonlocal length scale found in naturally occurring metals.

6.3. Doped semiconductors and 2D materials

Other free-carrier systems than metals, such as doped semiconductors [215] or the 2D material graphene [216, 217], provide an alternative powerful platform to study nonlocal effects of plasmons as these materials typically allow for easier tuning of the free-carrier density n_0. We note that the hydrodynamic nonlocal length scale scales as (for 3D materials) $\xi \propto v_F/\omega_p \propto n_0^{-1/6}$, which signifies a strong dependence of nonlocal effects with the free-carrier density. While nonlocal effects in excitons in semiconductors is a well-known topic [218–221], the search for nonlocal effects in plasmons of graphene and doped semiconductors has just been initiated [222–225].

6.4. Experimental ideas

Although there have been quite some studies of surface plasmon resonances in individual nanoparticles, showing e.g. a blueshift and broadening of the LSPR with decreasing particle size, the focus has mainly been on noble metals on substrates or in inert gases. We suggest to expand this study to include nanoparticles of different metals (e.g. Al) embedded in insulating materials. For studying the size-dependent resonance behavior, EELS would be a suitable technique due to the excellent spatial and spectral resolution.

Another interesting system to see features of nonlocal response is the metallic dimer. In this system, care must be taken since Ångstrom-sized gaps (more precise, below 5 Å) show signs of decreasing field enhancement in the gap of the dimer [22, 23], which is in contrast to expectations from the GNOR theory (see section 5.2). However, strong nonlocal effects are still present for gap sizes above 5 Å, where tunneling is not expected to play a role, so this makes the dimer indeed a good candidate. Additionally, the possible use of EELS to study dimers gives access to the dark modes of the system, which are also strongly gap-dependent and influenced by nonlocal response. In any case, it would also be interesting to reinterpret the existing experimental studies on dimers [18, 19, 22, 23, 25] using numerical calculations from the GNOR model to verify the applicability and accuracy of the nonlocal model.

The study of nonlocal effects in propagating SP modes is also of interest. In particular, continuous and homogeneous metal films support long-range SP modes [226, 227] which can be affected by nonlocal response for very small thicknesses [111, 112]. Ultra-thin metal films supporting SP modes with long propagation lengths constitute an interesting system for measuring effects due to nonlocal response, since additional effects due to quantum tunneling, as in the dimer geometry, can be immediately ruled out.

7. Conclusions

We have provided a comprehensive overview of the current status of both experimental and theoretical studies on nonlocal response in plasmonic nanostructures. The real-space constitutive relations relating the current density in metals to the electric field have been derived and discussed for the hydrodynamic and GNOR models. The GNOR model has been shown to be an extension of the hydrodynamic model by including currents due to electron diffusion. The main feature of nonlocal response is the inclusion of longitudinal waves in the metal, which gives rise to smearing of the induced charge density on the Ångstrom length scale. We have shown that, regardless of the geometry, only plasmonic modes which have negative and positive induced charges separated by a distance comparable to the smearing length scale, are affected by nonlocal response. In particular, the resonant plasmonic Mie modes supported by a spherical particle experience a blueshift and linewidth broadening as the particle diameter decreases. The same mechanism gives rise to a gap-dependent blueshift and linewidth broadening of the bonding modes of the dimer. In relation to experiments on dimers with nanometer-sized gaps, we find that the GNOR model describes the experimentally-measured spectra, without the need of invoking the quantum mechanical effect of tunneling. For the core-shell geometry, we find that only the antibonding (and not

the bonding) dipole mode is affected by nonlocal response. Finally, we have presented several theoretical approaches and experimental setups to unveil and measure further effects due to nonlocal response.

Acknowledgments

References

[1] Hao E and Schatz G C 2004 *J. Chem. Phys.* **120** 357
[2] Stockman M I 2004 *Phys. Rev. Lett.* **93** 137404
[3] Gramotnev D K and Bozhevolnyi S I 2010 *Nat. Photon.* **4** 83
[4] Gramotnev D K and Bozhevolnyi S I 2014 *Nat. Photon.* **8** 14
[5] Liz-Marzán L M 2006 *Langmuir* **22** 32
[6] Moskovits M 1985 *Rev. Mod. Phys.* **57** 783
[7] Homola J, Yee S S and Gauglitz G 1999 *Sensors Actuators* B **54** 3
[8] Anker J N, Hall W P, Lyandres O, Shah N C, Zhao J and Van Duyne R P 2008 *Nat. Mater.* **7** 442
[9] Bozhevolnyi S I, Volkov V S, Devaux E, Laluet J Y and Ebbesen T W 2006 *Nature* **440** 508
[10] Lal S, Clare S and Halas J 2008 *Acc. Chem. Res.* **41** 1842
[11] Zia R, Schuller J A, Chandran A and Brongersma M L 2006 *Mater. Today* **9** 20
[12] Huang K C Y, Seo M K, Sarmiento T, Huo Y, Harris J S and Brongersma M L 2014 *Nat. Photon.* **8** 244
[13] Ouyang F, Batson P and Isaacson M 1992 *Phys. Rev.* B **46** 15421
[14] Scholl J A, Koh A L and Dionne J A 2012 *Nature* **483** 421
[15] Raza S, Stenger N, Kadkhodazadeh S, Fischer S V, Kostesha N, Jauho A P, Burrows A, Wubs M and Mortensen N A 2013 *Nanophotonics* **2** 131
[16] Raza S, Yan W, Stenger N, Wubs M and Mortensen N A 2013 *Opt. Express* **21** 27344
[17] Kern J, Grossmann S, Tarakina N V, Häckel T, Emmerling M, Kamp M, Huang J S, Biagioni P, Prangsma J and Hecht B 2012 *Nano Lett.* **12** 5504
[18] Scholl J A, Garcia-Etxarri A, Koh A L and Dionne J A 2013 *Nano Lett.* **13** 564
[19] Cha H, Yoon J H and Yoon S 2014 *ACS Nano* **8** 8554
[20] Tan S F, Wu L, Yang J K, Bai P, Bosman M and Nijhuis C A 2014 *Science* **343** 1496
[21] Kravtsov V, Berweger S, Atkin J M and Raschke M B 2014 *Nano Lett.* **14** 5270
[22] Zhu W and Crozier K B 2014 *Nat. Commun.* **5** 5228
[23] Hajisalem G, Nezami M S and Gordon R 2014 *Nano Lett.* **14** 6651
[24] Ciracì C, Hill R T, Mock J J, Urzhumov Y, Fernández-Domínguez A I, Maier S A, Pendry J B, Chilkoti A and Smith D R 2012 *Science* **337** 1072
[25] Savage K J, Hawkeye M M, Esteban R, Borisov A G, Aizpurua J and Baumberg J J 2012 *Nature* **491** 574
[26] Johnson P B and Christy R W 1972 *Phys. Rev.* B **6** 4370
[27] Rakić A D, Djurišić A B, Elazar J M and Majewski M L 1998 *Appl. Opt.* **37** 5271
[28] Hao F and Nordlander P 2007 *Chem. Phys. Lett.* **446** 115
[29] Maier S A 2007 *Plasmonics: Fundamentals and Applications* (New York: Springer)
[30] Aizpurua J, Hanarp P, Sutherland D S, Käll M, Bryant G W and García de Abajo F J 2003 *Phys. Rev. Lett.* **90** 057401
[31] García de Abajo F J 2007 *Rev. Mod. Phys.* **79** 1267
[32] Ergin T, Stenger N, Brenner P, Pendry J B and Wegener M 2010 *Science* **328** 337
[33] Ritchie R H 1957 *Phys. Rev.* **106** 874
[34] Nelayah J, Kociak M, Stephan O, García de Abajo F J, Tence M, Henrard L, Taverna D, Pastoriza-Santos I, Liz-Marzan L M and Colliex C 2007 *Nat. Phys.* **3** 348
[35] Bosman M, Keast V J, Watanabe M, Maaroof A I and Cortie M B 2007 *Nanotechnology* **18** 165505
[36] Schaffer B, Hohenester U, Trügler A and Hofer F 2009 *Phys. Rev.* B **79** 041401
[37] Koh A L, Fernández-Domínguez A I, McComb D W, Maier S A and Yang J K W 2011 *Nano Lett.* **11** 1323
[38] Nicoletti O, Wubs M, Mortensen N A, Sigle W, van Aken P A and Midgley P A 2011 *Opt. Express* **19** 15371
[39] Husnik M, von Cube F, Irsen S, Linden S, Niegemann J, Busch K and Wegener M 2013 *Nanophotonics* **2** 241
[40] Raza S, Stenger N, Pors A, Holmgaard T, Kadkhodazadeh S, Wagner J B, Pedersen K, Wubs M, Bozhevolnyi S I and Mortensen N A 2014 *Nat. Commun.* **5** 4125
[41] Hofmann C E, Vesseur E J R, Sweatlock L A, Lezec H J, García de Abajo F J, Polman A and Atwater H A 2007 *Nano Lett.* **7** 3612
[42] Kuttge M, Vesseur E J R, Koenderink A F, Lezec H J, Atwater H A, García de Abajo F J and Polman A 2009 *Phys. Rev.* B **79** 113405
[43] García de Abajo F J 2010 *Rev. Mod. Phys.* **82** 209
[44] Schnell M, García-Etxarri A, Huber A J, Crozier K, Aizpurua J and Hillenbrand R 2009 *Nat. Photon.* **3** 287
[45] Chen J *et al* 2012 *Nature* **487** 77
[46] Duan H, Fernández-Domínguez A I, Bosman M, Maier S A and Yang J K W 2012 *Nano Lett.* **12** 1683
[47] Kreibig U and Genzel L 1985 *Surf. Sci.* **156** 678
[48] Kreibig U and Vollmer M 1995 *Optical Properties of Metal Clusters* (Berlin: Springer)
[49] Baida H *et al* 2009 *Nano Lett.* **9** 3463
[50] Kreibig U and von Fragstein C 1969 *Z. Phys.* **224** 307
[51] Apell P and Penn D R 1983 *Phys. Rev. Lett.* **50** 1316
[52] Uskov A V, Protsenko I E, Mortensen N A and O'Reilly E P 2014 *Plasmonics* **9** 185
[53] Tiggesbäumker J, Köller L, Meiwes-Broer K H and Liebsch A 1993 *Phys. Rev.* A **48** R1749
[54] Charlé K P, König L, Nepijko S, Rabin I and Schulze W 1998 *Cryst. Res. Technol.* **33** 1085
[55] Tsuei K D, Plummer E W, Liebsch A, Kempa K and Bakshi P 1990 *Phys. Rev. Lett.* **64** 44
[56] Bennett A 1970 *Phys. Rev.* B **1** 203
[57] Boardman A, Paranjape B and Teshima R 1975 *Surf. Sci.* **49** 275
[58] Bochterle J, Neubrech F, Nagao T and Pucci A 2012 *ACS Nano* **6** 10917
[59] Lindau I and Nilsson P O 1970 *Phys. Lett.* A **31A** 352
[60] Anderegg M, Feuerbacher B and Fitton B 1971 *Phys. Rev. Lett.* **27** 1565
[61] Onida G, Reining L and Rubio A 2002 *Rev. Mod. Phys.* **74** 601
[62] Zhang P, Feist J, Rubio A, García-González P and García-Vidal F J 2014 *Phys. Rev.* B **90** 161407
[63] Guidez E B and Aikens C M 2014 *Nanoscale* **6** 11512
[64] Lermé J, Baida H, Bonnet C, Broyer M, Cottancin E, Crut A, Maioli P, Del Fatti N, Vallée F and Pellarin M 2010 *J. Phys. Chem. Lett.* **1** 2922
[65] Teperik T V, Nordlander P, Aizpurua J and Borisov A G 2013 *Phys. Rev. Lett.* **110** 263901

[66] Andersen K, Jacobsen K W and Thygesen K S 2012 *Phys. Rev. B* **86** 245129

[67] Zuloaga J, Prodan E and Nordlander P 2009 *Nano Lett.* **9** 887

[68] Boardman A 1982 *Electromagnetic Surface Modes. Hydrodynamic Theory of Plasmon-Polaritons on Plane Surfaces* (Chichester: Wiley)

[69] Mortensen N A, Raza S, Wubs M, Søndergaard T and Bozhevolnyi S I 2014 *Nat. Commun.* **5** 3809

[70] Mortensen N A 2013 *Photon. Nanostruct.* **11** 303

[71] Ginzburg P and Zayats A 2013 *ACS Nano* **7** 4334

[72] Landau L D, Lifshitz E M and Pitaevskii L P 1984 *Electrodynamics of Continuous Media* (*Landau and Lifshitz Course on Theoretical Physics* vol 8) 2nd edn (Oxford: Butterworth Heinemann)

[73] Bloch F 1933 *Z. Phys. A* **81** 363

[74] Ying S C 1974 *Nuovo Cimento* **23** 270

[75] Eguiluz A and Quinn J 1976 *Phys. Rev. B* **14** 1347

[76] Boardman A D, Paranjape B V and Nakamura Y O 1976 *Phys. Status Solidi* **75** 347

[77] Melnyk A R and Harrison M J 1970 *Phys. Rev. B* **2** 835

[78] Eguiluz A, Ying S and Quinn J 1975 *Phys. Rev. B* **11** 2118

[79] Mochán W L, del CastilloMussot M and Barrera R G 1987 *Phys. Rev. B* **35** 1088

[80] Mochán W L and del CastilloMussot M 1988 *Phys. Rev. B* **37** 6763

[81] Boardman A D and Paranjape B V 1977 *J. Phys. F: Met. Phys.* **7** 1935

[82] Ruppin R 1973 *Phys. Rev. Lett.* **31** 1434

[83] Ruppin R 1975 *Phys. Rev. B* **11** 2871

[84] Aers G, Paranjape B and Boardman A 1979 *J. Phys. Chem. Solids* **40** 319

[85] Aers G C, Boardman A D and Paranjape B V 1980 *J. Phys. F: Met. Phys.* **10** 53

[86] Ruppin R 1989 *J. Opt. Soc. Am. B* **6** 1559

[87] McMahon J M, Gray S K and Schatz G C 2009 *Phys. Rev. Lett.* **103** 097403

[88] Toscano G, Raza S, Jauho A P, Mortensen N A and Wubs M 2012 *Opt. Express* **20** 4176

[89] Hiremath K R, Zschiedrich L and Schmidt F 2012 *J. Comput. Phys.* **231** 5890

[90] Toscano G, Raza S, Xiao S, Wubs M, Jauho A P, Bozhevolnyi S I and Mortensen N A 2012 *Opt. Lett.* **37** 2538

[91] Toscano G, Raza S, Yan W, Jeppesen C, Xiao S, Wubs M, Jauho A P, Bozhevolnyi S I and Mortensen N A 2013 *Nanophotonics* **2** 161

[92] Fernández-Domínguez A I, Wiener A, García-Vidal F J, Maier S A and Pendry J B 2012 *Phys. Rev. Lett.* **108** 106802

[93] Fernández-Domínguez A I, Zhang P, Luo Y, Maier S A, García-Vidal F J and Pendry J B 2012 *Phys. Rev. B* **86** 241110

[94] Fernández-Domínguez A I, Luo Y, Wiener A, Pendry J B and Maier S A 2012 *Nano Lett.* **12** 5946

[95] Villó-Pérez I and Arista N R 2009 *Surf. Sci.* **603** 1

[96] Raza S, Toscano G, Jauho A P, Wubs M and Mortensen N A 2011 *Phys. Rev. B* **84** 121412

[97] Yan W, Mortensen N A and Wubs M 2013 *Phys. Rev. B* **88** 155414

[98] Raza S, Toscano G, Jauho A P, Mortensen N A and Wubs M 2013 *Plasmonics* **8** 193

[99] Moradi A and Ebrahimi E 2014 *Plasmonics* **9** 209

[100] Ben X and Park H S 2012 *J. Phys. Chem. C* **116** 18944

[101] Wiener A, Fernández-Domínguez A I, Horsfield A P, Pendry J B and Maier S A 2012 *Nano Lett.* **12** 3308

[102] Ruppin R 005 *Phys. Lett.* A **340** 299

[103] Ciracì C, Poutrina E, Scalora M and Smith D R 2012 *Phys. Rev. B* **85** 201403

[104] Ciracì C, Poutrina E, Scalora M and Smith D R 2012 *Phys. Rev. B* **86** 115451

[105] Li H, Zhu N, Zhang H, Chen Z and Mei T 2015 *Opt. Commun.* **339** 177

[106] David C and García de Abajo F J 2011 *J. Phys. Chem.* C **115** 19470

[107] Ciracì C, Urzhumov Y and Smith D R 2013 *J. Opt. Soc. Am.* B **30** 2731

[108] Trivedi R, Thomas A and Dhawan A 2014 *Opt. Express* **22** 19970

[109] Luo Y, Zhao R and Pendry J B 2014 *Proc. Natl Acad. Sci.* **111** 18422

[110] Huang Y and Gao L 2014 *J. Phys. Chem. C* **118** 30170

[111] Moreau A, Ciracì C and Smith D R 2013 *Phys. Rev. B* **87** 045401

[112] Raza S, Christensen T, Wubs M, Bozhevolnyi S I and Mortensen N A 2013 *Phys. Rev. B* **88** 115401

[113] Wiener A, Fernández-Domínguez A I, Pendry J B, Horsfield A P and Maier S A 2013 *Opt. Express* **21** 27509

[114] Xiao Y, Li J, Pan T and Xu G 2015 *Phys. Status Solidi* B **252** 187

[115] Kang N, Mišković Z, Zhang Y Y, Song Y H and Wang Y N 2014 *Nanoscale Syst.: Math. Model. Theory Appl.* **3** 44

[116] Larkin I A and Stockman M I 2005 *Nano Lett.* **5** 339

[117] de Ceglia D, Campione S, Vincenti M A, Capolino F and Scalora M 2013 *Phys. Rev. B* **87** 155140

[118] David C, Mortensen N A and Christensen J 2013 *Sci. Rep.* **3** 2526

[119] Yanai A and Levy U 2014 *Phys. Rev. B* **90** 075107

[120] Sun J, Huang Y and Gao L 2014 *Phys. Rev. A* **89** 012508

[121] Yan W, Wubs M and Mortensen N A 2012 *Phys. Rev. B* **86** 205429

[122] Yan W, Mortensen N A and Wubs M 2013 *Opt. Express* **21** 15026

[123] Yanai A, Mortensen N A and Levy U 2013 *Phys. Rev. B* **88** 205120

[124] Benedicto J, Pollès R, Ciracì C, Centeno E, Smith D R and Moreau A 2014 arXiv:1410.4996

[125] Dechaux M, Tichit P H, Ciracì C, Benedicto J, Pollès R, Centeno E, Smith D R and Moreau A 2015 arXiv:1501.05765

[126] Intravaia F and Busch K 2014 arXiv:1412.4110

[127] Wiener A, Duan H, Bosman M, Horsfield A P, Pendry J B, Yang J K W, Maier S A and Fernández-Domínguez A I 2013 *ACS Nano* **7** 6287

[128] Christensen T, Yan W, Raza S, Jauho A P, Mortensen N A and Wubs M 2014 *ACS Nano* **8** 1745

[129] Fuchs R and Barrera R G 1981 *Phys. Rev. B* **24** 2940

[130] Filter R, Bösel C, Toscano G, Lederer F and Rockstuhl C 2014 *Opt. Lett.* **39** 6118

[131] Teperik T V, Nordlander P, Aizpurua J and Borisov A G 2013 *Opt. Express* **21** 27306

[132] Stella L, Zhang P, García-Vidal F J, Rubio A and García-González P 2013 *J. Phys. Chem. C* **117** 8941

[133] Fuchs R and Kliewer K L 1969 *Phys. Rev.* **185** 905

[134] Kliewer K L 1974 *Phys. Rev. Lett.* **33** 900

[135] Xue H J, Wu R L and Yu Y 2014 *J. Phys. Chem. C* **118** 18257

[136] Khurgin J 2014 *Faraday Discuss.* doi:10.1039/C4FD00193A

[137] Apell P and Ljungbert Å 1982 *Phys. Scr.* **26** 113

[138] García de Abajo F J 2008 *J. Phys. Chem. C* **112** 17983

[139] Moradi A 2015 *Phys. Plasmas* **22** 022119

[140] Zhang Y Y, An S B, Song Y H, Kang N, Mišković Z L and Wang Y N 2014 *Phys. Plasmas* **21** 102114

[141] Pitarke J M, Silkin V M, Chulkov E V and Echenique P M 2007 *Rep. Prog. Phys.* **70** 1

[142] Fetter A L 1973 *Ann. Phys.* **81** 367

[143] Halevi P 1995 *Phys. Rev. B* **51** 7497

[144] Sauter F 1967 *Z. Phys.* **203** 488

[145] Forstmann F and Stenschke H 1977 *Phys. Rev. Lett.* **38** 1365

[146] Boardman A D and Ruppin R 1981 *Surf. Sci.* **112** 153
[147] Jewsbury P 1981 *J. Phys. F: Met. Phys.* **11** 195
[148] Forstmann F and Stenschke H 1978 *Phys. Rev.* B **17** 1489
[149] Hanson G W 2010 *IEEE Antennas Propag. Mag.* **52** 198
[150] Lindhard J 1954 *Dan. Mat. Fys. Medd.* **28** 8
[151] Warren J L and Ferrell R A 1960 *Phys. Rev.* **117** 1252
[152] Mermin N D 1970 *Phys. Rev.* B **1** 2362
[153] Mie G 1908 *Ann. Phys.* **25** 377
[154] van de Hulst H C 1981 *Light Scattering by Small Particles* (New York: Dover)
[155] Bohren C F and Huffman D R 1983 *Absorption and Scattering of Light by Small Particles* (New York: Wiley)
[156] Apell P and Ljungbert Å 1982 *Solid State Commun.* **44** 1367
[157] Dasgupta B B and Fuchs R 1981 *Phys. Rev.* B **24** 554
[158] Fuchs R and Claro F 1987 *Phys. Rev.* B **35** 3722
[159] Crowell J and Ritchie R H 1968 *Phys. Rev.* **172** 436
[160] Luo Y, Fernández-Domínguez A I, Wiener A, Maier S A and Pendry J B 2013 *Phys. Rev. Lett.* **111** 093901
[161] Feibelman P J 1982 *Prog. Surf. Sci.* **12** 287
[162] Pendry J B, Aubry A, Smith D R and Maier S A 2012 *Science* **337** 549
[163] Genzel L, Martin T P and Kreibig U 1975 *Z. Phys.* B **21** 339
[164] Charlé K P, Frank F and Schulze W 1984 *Ber. Bunsenges. Phys. Chem.* **88** 350
[165] Charlé K P, Schulze W and Winter B 1989 *Z. Phys.* D **12** 471
[166] Charlé K P, König L, Rabin I and Schulze W 1996 *Z. Phys.* D **36** 159
[167] Zhao J, Zhang H and Wang G 1996 *J. Phys. Chem. Solids* **57** 225
[168] Bréchignac C, Cahuzac P, Kebïli N, Leygnier J and Sarfati A 1992 *Phys. Rev. Lett.* **68** 3916
[169] Liebsch A 1993 *Phys. Rev.* B **48** 11317
[170] Monreal R C, Antosiewicz T J and Apell S P 2013 *New J. Phys.* **15** 083044
[171] Gaudry M *et al* 2003 *Phys. Rev.* B **67** 155409
[172] Scaffardi L and Tocho J 2006 *Nanotechnology* **17** 1309
[173] Kolwas K and Derkachova A 2013 *J. Quant. Spectrosc. Radiat. Transfer* **114** 45
[174] Myroshnychenko V, Rodríguez-Fernández J, Pastoriza-Santos I, Funston A M, Novo C, Mulvaney P, Liz-Marzán L M and García de Abajo F J 2008 *Chem. Soc. Rev.* **37** 1792
[175] Ljungbert Å and Apell P 1983 *Solid State Commun.* **46** 47
[176] Apell P, Monreal R and Flores F 1984 *Solid State Commun.* **52** 971
[177] Raza S, Wubs M, Bozhevolnyi S I and Mortensen N A 2015 *Opt. Lett.* **40** 839
[178] Andersen K, Jensen K L, Mortensen N A and Thygesen K S 2013 *Phys. Rev.* B **87** 235433
[179] Toscano G, Rockstuhl C, Evers F, Xu H, Mortensen N A and Wubs M 2014 arXiv:1408.5862
[180] García de Abajo F J 1999 *Phys. Rev.* B **59** 3095
[181] Prodan E, Radloff C, Halas N and Nordlander P 2003 *Science* **302** 419
[182] Nordlander P, Oubre C, Prodan E, Li K and Stockman M I 2004 *Nano Lett.* **4** 899
[183] Thongrattanasiri S, Koppens F H L and García de Abajo F J 2012 *Phys. Rev. Lett.* **108** 047401
[184] Kneipp K, Wang Y, Kneipp H, Perelman L T, Itzkan I, Dasari R R and Feld M S 1997 *Phys. Rev. Lett.* **78** 1667
[185] Jain P K, Huang W and El-Sayed M A 2007 *Nano Lett.* **7** 2080
[186] Romero I, Aizpurua J, Bryant G W and García de Abajo F J 2006 *Opt. Express* **14** 9988
[187] Kadkhodazadeh S, Wagner J B, Kneipp H and Kneipp K 2013 *Appl. Phys. Lett.* **103** 083103
[188] Esteban R, Borisov A G, Nordlander P and Aizpurua J 2012 *Nat. Commun.* **3** 825
[189] Haus J W, de Ceglia D, Vincenti M A and Scalora M 2014 *J. Opt. Soc. Am.* B **31** 259
[190] Haus J W, de Ceglia D, Vincenti M A and Scalora M 2014 *J. Opt. Soc. Am.* B **31** 6
[191] Scalora M, Vincenti M A, de Ceglia D and Haus J W 2014 *Phys. Rev.* A **90** 013831
[192] Brongersma M L 2003 *Nat. Mater.* **2** 296
[193] Raschke G *et al* 2004 *Nano Lett.* **4** 1853
[194] Nehl C L, Grady N K, Goodrich G P, Tam F, Halas N J and Hafner J H 2004 *Nano Lett.* **4** 2355
[195] Tam F, Moran C and Halas N J 2004 *J. Phys. Chem.* B **108** 17290
[196] Kabashin A, Evans P, Pastkovsky S, Hendren W, Wurtz G, Atkinson R, Pollard R, Podolskiy V and Zayats A 2009 *Nat. Mater.* **8** 867
[197] Bardhan R, Lal S, Joshi A and Halas N J 2011 *Acc. Chem. Res.* **44** 936
[198] Prodan E and Nordlander P 2004 *J. Chem. Phys.* **120** 5444
[199] Ruppin R 2001 *Opt. Commun.* **190** 205
[200] Lang N D and Kohn W 1970 *Phys. Rev.* B **1** 4555
[201] Öztürk Z F, Xiao S, Yan M, Wubs M, Jauho A P and Mortensen N A 2011 *J. Nanophoton.* **5** 051602
[202] Ichikawa M 2011 *J. Phys. Soc. Japan* **80** 044606
[203] David C and García de Abajo F J 2014 *ACS Nano* **8** 9558
[204] Yan W 2015 *Phys. Rev.* B **91** 225416
[205] Smith D R, Pendry J B and Wiltshire M C K 2004 *Science* **305** 788
[206] Shelby R A, Smith D R and Schultz S 2001 *Science* **292** 77
[207] Chen X, Luo Y, Zhang J, Jiang K, Pendry J B and Zhang S 2011 *Nat. Commun.* **2** 176
[208] Zhang J, Liu L, Luo Y, Zhang S and Mortensen N A 2011 *Opt. Express* **19** 8625
[209] Pors A, Tsukerman I and Bozhevolnyi S I 2011 *Phys. Rev.* E **84** 016609
[210] Belov P A, Marqués R, Maslovski S I, Nefedov I S, Silveirinha M, Simovski C R and Tretyakov S A 2003 *Phys. Rev.* B **67** 113103
[211] Silveirinha M G 2006 *Phys. Rev.* E **73** 046612
[212] Elser J, Podolskiy V A, Salakhutdinov I and Avrutsky I 2007 *Appl. Phys. Lett.* **90** 19
[213] Menzel C, Paul T, Rockstuhl C, Pertsch T, Tretyakov S and Lederer F 2010 *Phys. Rev.* B **81** 035320
[214] Wells B M, Zayats A V and Podolskiy V A 2014 *Phys. Rev.* B **89** 035111
[215] Zhang H, Kulkarni V, Prodan E, Nordlander P and Govorov A O 2014 *J. Phys. Chem.* C **118** 16035
[216] Koppens F H L, Chang D E and García de Abajo F J 2011 *Nano Lett.* **11** 3370
[217] Christensen J, Manjavacas A, Thongrattanasiri S, Koppens F H L and García de Abajo F J 2012 *ACS Nano* **6** 431
[218] Pekar S I 1958 *Sov. Phys.—JETP* **7** 813
[219] Hopfield J J 1958 *Phys. Rev.* **112** 1555
[220] Hopfield J J and Thomas D G 1963 *Phys. Rev.* **132** 563
[221] Ruppin R 1981 *J. Opt. Soc. Am.* **71** 755
[222] Wang W and Kinaret J M 2013 *Phys. Rev.* B **87** 195424
[223] Fallahi A, Low T, Tamagnone M and Perruisseau-Carrier J 2015 *Phys Rev.* B **91** 121405
[224] Christensen T, Wang W, Jauho A P, Wubs M and Mortensen N A 2014 *Phys. Rev.* B **90** 241414
[225] Schimpf A M, Thakkar N, Gunthardt C E, Masiello D J and Gamelin D R 2014 *ACS Nano* **8** 1065
[226] Burke J J, Stegeman G I and Tamir T 1986 *Phys. Rev.* B **33** 5186
[227] Berini P 2009 *Adv. Opt. Photon.* **1** 484