

Technical University of Denmark



## Classification of scalar and dyadic nonlocal optical response models

**Wubs, Martijn**

*Published in:*  
Optics Express

*Link to article, DOI:*  
[10.1364/OE.23.031296](https://doi.org/10.1364/OE.23.031296)

*Publication date:*  
2015

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Wubs, M. (2015). Classification of scalar and dyadic nonlocal optical response models. Optics Express, 23(24), 31296-31312. DOI: 10.1364/OE.23.031296

## DTU Library

Technical Information Center of Denmark

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Classification of scalar and dyadic nonlocal optical response models

M. Wubs<sup>1,2,\*</sup>

<sup>1</sup>Department of Photonics Engineering, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

<sup>2</sup>Center for Nanostructured Graphene, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark

\*[mwubs@fotonik.dtu.dk](mailto:mwubs@fotonik.dtu.dk)

<http://www.fotonik.dtu.dk/SEM>

**Abstract:** Nonlocal optical response is one of the emerging effects on the nanoscale for particles made of metals or doped semiconductors. Here we classify and compare both scalar and tensorial nonlocal response models. In the latter case the nonlocality can stem from either the longitudinal response, the transverse response, or both. In phenomenological scalar models the nonlocal response is described as a smearing out of the commonly assumed infinitely localized response, as characterized by a distribution with a finite width. Here we calculate explicitly whether and how tensorial models, such as the hydrodynamic Drude model and generalized nonlocal optical response theory, follow this phenomenological description. We find considerable differences, for example that nonlocal response functions, in contrast to simple distributions, assume negative and complex values. Moreover, nonlocal response regularizes some but not all diverging optical near fields. We identify the scalar model that comes closest to the hydrodynamic model. Interestingly, for the hydrodynamic Drude model we find that actually only one third (1/3) of the free-electron response is smeared out nonlocally. In that sense, nonlocal response is stronger for transverse and scalar nonlocal response models, where the smeared-out fractions are 2/3 and 3/3, respectively. The latter two models seem to predict novel plasmonic resonances also below the plasma frequency, in contrast to the hydrodynamic model that predicts standing pressure waves only above the plasma frequency.

© 2015 Optical Society of America

**OCIS codes:** (260.3910) Metal optics; (260.2030) Dispersion; (120.6710) Susceptibility; (240.6680) Surface plasmons; (000.6800) Theoretical physics.

---

## References and links

1. M. Barbry, P. Koval, F. Marchesin, R. Esteban, A. G. Borisov, J. Aizpurua, and D. Sánchez-Portal, "Atomistic near-field nanoplasmonics: reaching atomic-scale resolution in nanooptics," *Nano Lett.* **15**(5), 3410–3419 (2015).
2. X. Chen, J. E. Moore, M. Zekarias, and L. Jensen, "Atomistic electrodynamics simulations of bare and ligand-coated nanoparticles in the quantum size regime," *Nature Commun.* **6**, 8921 (2015).
3. H. Zhang, V. Kulkarni, E. Prodan, P. Nordlander, and A. O. Govorov, "Theory of quantum plasmon resonances in doped semiconductor nanocrystals," *J. Phys. Chem. C* **118**, 16035–16042 (2014).
4. E. Forstmann and R. R. Gerhardts, *Metal Optics Near the Plasma Frequency* (Springer, 1986).
5. O. Keller, "Screened electromagnetic propagators in nonlocal metal optics," *Phys. Rev. B* **34**(6), 3883–3899 (1986).

6. S. Raza, S. I. Bozhevolnyi, M. Wubs, and N. A. Mortensen, "Nonlocal optical response in metallic nanostructures," *J. Phys.: Condens. Matter* **27**, 183204 (2015).
7. R. Ruppin, "Optical properties of a spatially dispersive cylinder," *J. Opt. Soc. Am. B* **6**(8), 1559–1563 (1989).
8. R. Carmina Monreal, T. J. Antosiewicz, and S. P. Apell, "Diffuse surface scattering in the plasmonic resonances of ultralow electron density nanospheres," *J. Phys. Chem. Lett.* **6**, 1847–1853 (2015).
9. A. Moradi, "Infrared absorption spectra of a spatially dispersive polar semiconductor nanowire," *Solid State Commun.* **212**, 10–13 (2015).
10. T. Christensen, W. Wang, A.-P. Jauho, M. Wubs, and N. A. Mortensen, "Classical and quantum plasmonics in graphene nanodisks: role of edge states," *Phys. Rev. B* **90**(24), 241414(R) (2014).
11. D. Correas-Serrano, J. S. Gomez-Diaz, M. Tymchenko, and A. Alù, "Nonlocal response of hyperbolic metasurfaces," *Opt. Express* **23**(23), 29434 (2015).
12. F. Bloch, "Bremsvermögen von Atomen mit mehreren Elektronen," *Z. Physik A* **81**, 363–376 (1933).
13. S. Raza, N. Stenger, S. Kadkhodazadeh, S. V. Fischer, N. Kostesha, A.-P. Jauho, A. Burrows, M. Wubs, and N. A. Mortensen, "Blueshift of the surface plasmon resonance in silver nanoparticles studied with EELS," *Nanophotonics* **2**(3), 161–166 (2013).
14. I. Lindau and P. O. Nilsson, "Experimental verification of optically excited longitudinal plasmons," *Phys. Scr.* **3**, 87–92 (1971).
15. A. I. Fernández-Domínguez, A. Wiener, F. J. García-Vidal, S. A. Maier, and J. B. Pendry, "Transformation-optics description of nonlocal effects in plasmonic nanostructures," *Phys. Rev. Lett.* **108**(10), 106802 (2012).
16. G. Toscano, S. Raza, A.-P. Jauho, N. A. Mortensen, and M. Wubs, "Modified field enhancement and extinction in plasmonic nanowire dimers due to nonlocal response," *Opt. Express* **20**(4), 4176–4188 (2012).
17. N. A. Mortensen, S. Raza, M. Wubs, T. Søndergaard, and S. I. Bozhevolnyi, "A generalized non-local optical response theory for plasmonic nanostructures," *Nature Commun.* **5**, 3809 (2014).
18. W. Yan, N. A. Mortensen, and M. Wubs, "Green's function surface-integral method for nonlocal response of plasmonic nanowires in arbitrary dielectric environments," *Phys. Rev. B* **88**(15), 155414 (2013).
19. T. Christensen, W. Yan, S. Raza, A.-P. Jauho, N. A. Mortensen, and M. Wubs, "Nonlocal response of metallic nanospheres probed by light, electrons, and atoms," *ACS Nano* **8**(2), 1745–1758 (2014).
20. S. Raza, S. Kadkhodazadeh, T. Christensen, M. Di Vece, M. Wubs, N. A. Mortensen, and N. Stenger, "Multipole plasmons and their disappearance in few-nanometer silver nanoparticles," *Nature Commun.* **6**, 8788 (2015).
21. K. J. Savage, M. M. Hawkeye, R. Esteban, A. G. Borisov, J. Aizpurua, and J. J. Baumberg, "Revealing the quantum regime in tunnelling plasmonics," *Nature* **491**, 574–577 (2012).
22. H. Jung, H. Cha, D. Lee, and S. Yoon, "Bridging the nanogap with light: continuous tuning of plasmon coupling between gold nanoparticles," *ACS Nano*, Article ASAP (2015).
23. R. Esteban, A. G. Borisov, P. Nordlander, and J. Aizpurua, "Bridging quantum and classical plasmonics with a quantum-corrected model," *Nature Commun.* **3**, 825 (2012).
24. S. Raza, M. Wubs, S. I. Bozhevolnyi, and N. A. Mortensen, "Nonlocal study of ultimate plasmon hybridization," *Opt. Lett.* **40**(5), 839–842 (2015).
25. W. Yan, M. Wubs, and N. A. Mortensen, "Projected-dipole model for quantum plasmonics," *Phys. Rev. Lett.* **115**(13), 137403 (2015).
26. G. Toscano, J. Straubel, A. Kwiatkowski, C. Rockstuhl, F. Evers, H. Xu, N. A. Mortensen, and M. Wubs, "Resonance shifts and spill-out effects in self-consistent hydrodynamic nanoplasmonics," *Nature Commun.* **6**, 7132 (2015).
27. W. Yan, M. Wubs, and N. A. Mortensen, "Hyperbolic metamaterials: nonlocal response regularizes broadband supersingularity," *Phys. Rev. B* **86**(20), 205429 (2012).
28. W. Yan, N. A. Mortensen, and M. Wubs, "Hyperbolic metamaterial lens with hydrodynamic nonlocal response," *Opt. Express* **21**(12), 15026 (2013).
29. A. Moradi, "Maxwell-Garnett effective medium theory: quantum nonlocal effects," *Phys. of Plasmas* **22**(4), 042105 (2015).
30. A. Yanai, N. A. Mortensen, and U. Levy, "Absorption and eigenmode calculation for one-dimensional periodic metallic structures using the hydrodynamic approximation," *Phys. Rev. B* **88**(20), 205120 (2013).
31. F. Intravaia and K. Busch, "Fluorescence in nonlocal dissipative periodic structures," *Phys. Rev. A* **91**(5), 053836 (2015).
32. P. Ginzburg and A. Zayats, "Localized surface plasmon resonances in spatially dispersive nano-objects: phenomenological treatise," *ACS Nano* **7**(5), 43344342 (2013).
33. N. A. Mortensen, "Nonlocal formalism for nanoplasmonics: phenomenological and semi-classical considerations," *Phot. Nanostr.* **11**(4), 303 (2013).
34. P. de Vries, D. V. van Coevorden, and A. Lagendijk, "Point scatterers for classical waves," *Rev. Mod. Phys.* **70**(2), 447 (1998).
35. D. P. Craig and T. Thirunamachandran, *Molecular Quantum Electrodynamics. An Introduction to Radiation Molecule Interactions* (Academic, 1984); see Sec. 3.1: Transverse and longitudinal  $\delta$ -dyadics.
36. C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg, *Photons and Atoms. Introduction to Quantum Electrodynamics* (Wiley, 1989); see pp. 36–42.

37. J. M. McMahon, S. K. Gray, and G. C. Schatz, "Nonlocal optical response of metal nanostructures with arbitrary shape," *Phys. Rev. Lett.* **103**(9), 097403 (2009).
  38. J. Lindhard, "On the properties of a gas of charged particles," *Kgl. Dansk Videnskab. Selskab, Mat. Fys. Medd.* **28**(8), 1–57 (1954).
  39. S. Raza, G. Toscano, A.-P. Jauho, M. Wubs, and N. A. Mortensen, "Unusual resonances in nanoplasmonic structures due to nonlocal response," *Phys. Rev. B* **84**(12), 121412(R) (2011).
- 

## 1. Introduction

In plasmonics research, conducting materials are described on many levels, from classically all the way to atomistically [1, 2]. Here we focus on effects of nonlocal response (also known as ‘spatial dispersion’) that for metals emerge on the few-nanometer scale, which is the intermediate regime where deviations from classical electrodynamics occur while atomistic descriptions are not needed yet. Nonlocal response means that the polarization field at a certain point depends not just on the electric field in that point, but rather on the electric field in its neighborhood. In Fourier space it is characterized by a wavevector-dependent dielectric response.

In microscopic descriptions of light-matter interactions, nonlocal optical response may ultimately occur due to the quantum mechanical position uncertainty of the particles or material excitations that the light interacts with. The optical properties of conductors is typically dominated by the free-carrier response, which is a collective effect (except for semiconductors at very low doping [3]) that allows semi-classical descriptions of nonlocal response, both for metals [4–6], for doped semiconductors [3, 7–9], and recently also for graphene [10, 11]. Indeed various nonlocal-response models for conductors have been proposed, predicting different phenomena. This article is not an attempt to review all of these models but rather to classify some of them, after introducing a few developments in this active subfield of plasmonics.

Probably the best known semiclassical model to describe nonlocal collective response is the hydrodynamic Drude model (HDM) that goes back to Bloch [4, 12] who included the dynamics of the electron gas in the Thomas-Fermi approximation. Although the HDM has been around for a while, it is relevant for current research both because new experiments become possible and because new effects are predicted with the HDM. Three nonclassical phenomena that are well known to be explained by the HDM are (i) size-dependent blueshifts of the dipolar resonance of noble metal nanoparticles as their sizes get smaller than typically 10 nanometers, see for example Ref. [13], (ii) standing bulk-plasmon wave resonances above the plasma frequency, first observed as anomalous absorption in Ag thin films [14], and (iii) somewhat reduced plasmonic field enhancements [15, 16].

A natural extension of the HDM to include nonlocality due to surface diffusion is the Generalized Nonlocal Optical Response (GNOR) model [6, 17]. Recently it was recognized that the HDM predicts that surface-plasmon resonances of higher multipolar order are blueshifted more [18, 19], and in the GNOR model also broadened more [6]. The latter nonlocal broadening was used very recently to explain experimental electron-loss spectra of embedded silver spheres, in particular the disappearance of higher-order plasmon resonances for spheres of radii smaller than 4 nanometers [20].

Very actively studied nowadays are the optical properties of plasmonic dimer systems with (sub-)nanometer-size gaps. Again an important role is played by semi-classical models. Even though they cannot explain all phenomena, to get insight it is important to find out which phenomena can and which cannot be explained by these relatively simple models. In Ref. [16] the HDM was used to show that large dimers with small gaps respond non-classically even though classical electrodynamics suffices to describe the response of the individual monomers. An intermediate nonlocal regime of gap sizes was anticipated to be well described by semiclassical models, with gaps too small for classical electrodynamics to be accurate, yet too large to allow short-circuiting due to electronic spill-out at both interfaces [16]. In some experiments this in-

intermediate nonlocal regime was not identified [21], while in others it was [22]. The observed broadening of optical spectra for smaller-gap dimers [21] is typically interpreted as due to quantum tunneling [23], so it is rather interesting that the semiclassical GNOR theory systematically neglects tunneling but still predicts an analogous spectral broadening of smaller-gap dimers as due to diffusive nonlocal response [6, 17, 24]. What is more, without further ingredients the GNOR theory also explains the observed spectral broadening for smaller monomers, which of course cannot be explained as due to quantum tunneling. Here we have the interesting situation that a semiclassical nonlocal-response theory suggests to reinterpret microscopic quantum calculations, because besides quantum tunneling there must be at least one other broadening mechanism at work in small-gap dimers. This is confirmed in Ref. [25] where it is shown that besides damping due to electron-hole pair creation in forward-scattering processes across the gap (i.e. tunneling) also backward scattering processes at metal-air interfaces are important, the latter processes both for monomers and dimers.

Another very recent generalization of the standard hydrodynamic Drude model is the self-consistent hydrodynamic model (with electron gas dynamics beyond the Thomas-Fermi approximation) that can describe electronic spill-out semiclassically [26]. It predicts nonlocal blueshifts for noble metals while for simple metals redshifts are predicted for smaller nanoparticles [26], both in agreement with experiments. By contrast, the usual (hard-wall) hydrodynamic model systematically excludes spill-out at the interfaces and consequently always predicts nonlocal blueshifts. Furthermore, the self-consistent hydrodynamic model is the first semiclassical theory to describe so-called multipole surface plasmons, also known as Bennett resonances, and which have been experimentally observed, as reviewed in [4].

As a final research trend, it deserves mentioning that effects of nonlocal response are being investigated for metamaterials [27–29], metasurfaces [11] and other periodic structures [30,31].

In this article we will classify several types of nonlocal-response models, both scalar and dyadic, as summarized in Table 1. We will derive their nonlocal response functions in real space,

Table 1. **Overview of Response Models, and Where in the Text they are Discussed.**

Name	Defined by	Where?
Scalar local response	$\epsilon_L(\omega) = \epsilon_T(\omega)$	Sec. 3.1
Nonlocal with local trace	$\epsilon_L(\omega) \neq \epsilon_T(\omega)$	Sec. 3.2
Longitudinal nonlocal response	$\epsilon_L(k, \omega)$ and $\epsilon_T(\omega)$	Sec. 4
Transverse nonlocal response	$\epsilon_L(\omega)$ and $\epsilon_T(k, \omega)$	Sec. 5
Scalar nonlocal response	$\epsilon_L(k, \omega) = \epsilon_T(k, \omega)$	Sec. 6.1
Non-scalar L- and T-nonlocal response	$\epsilon_L(k, \omega) \neq \epsilon_T(k, \omega)$	Sec. 6.2

and compare the results with phenomenological scalar models for nonlocal response [32, 33]. Aim is to obtain a sharper intuition about the models considered and the phenomena that they describe, especially the analogies and differences between scalar and dyadic models.

## 2. Preliminaries: nonlocal response in real space and in wavevector space

### 2.1. From real space to wavevector space: Fourier relations

In general, nonlocal response means that the response of a medium at a certain point  $\mathbf{r}$  depends on the exciting field in a neighborhood of that point, rather than on the exciting field at  $\mathbf{r}$  only. Here for simplicity we only consider dielectric (nonmagnetic) response of a medium to light, and assume the response to be linear. Then the dielectric function (or rather: the  $3 \times 3$  dielectric tensor  $\boldsymbol{\epsilon}$ ) depends on two positions rather than one, so that the displacement field at a certain

point depends on the electric field in a neighborhood of that point:

$$\mathbf{D}(\mathbf{r}_1, \omega) = \epsilon_0 \int d\mathbf{r}_2 \boldsymbol{\epsilon}(\mathbf{r}_1, \mathbf{r}_2, \omega) \cdot \mathbf{E}(\mathbf{r}_2, \omega), \quad (1)$$

where the dielectric function has become a nonlocal response function  $\boldsymbol{\epsilon}(\mathbf{r}_1, \mathbf{r}_2, \omega)$  that depends on two positions. One can also study its Fourier representation  $\boldsymbol{\epsilon}(\mathbf{k}_1, \mathbf{k}_2, \omega)$  but for arbitrarily shaped media this does not necessarily simplify the analysis. Although nonlocal response in finite geometries will be discussed in this article, we only calculate response functions for translationally invariant systems, where the nonlocal response can only depend on the spatial difference between the two position arguments,  $\boldsymbol{\epsilon}(\mathbf{r}_1, \mathbf{r}_2, \omega) = \boldsymbol{\epsilon}(\mathbf{r}_1 - \mathbf{r}_2, \omega) = \boldsymbol{\epsilon}(\mathbf{r}, \omega)$ , where we defined the position difference vector  $\mathbf{r} \equiv \mathbf{r}_1 - \mathbf{r}_2$ . Due to this translational invariance, the response assumes a simple form in  $k$ -space,  $\boldsymbol{\epsilon}(\mathbf{k}_1, \mathbf{k}_2, \omega) = (2\pi)^3 \delta^3(\mathbf{k}_1 - \mathbf{k}_2) \boldsymbol{\epsilon}(\mathbf{k}_1, \omega)$ , where the tensor function  $\boldsymbol{\epsilon}(\mathbf{k}_1, \omega)$  is the Fourier transform of  $\boldsymbol{\epsilon}(\mathbf{r}, \omega)$ . For translationally invariant systems, it is usually simpler to first find the dielectric function in  $k$ -space rather than in real space. By subsequent inverse Fourier transformation (see below in Sec. 2.3) one then finds the sought real-space form of the nonlocal response that features in the constitutive relation (1). The nonlocal response function of infinite systems so obtained is also useful for the analysis of finite systems, as an approximation for example [32], or as input for a phenomenological specular reflection model for conduction electrons at the metal surface [4].

## 2.2. Transverse and longitudinal response

The dielectric function is written as a tensor, and indeed in a few important models that we consider below, the response function is tensorial (or dyadic, meaning  $3 \times 3$  tensorial) rather than scalar. For isotropic bulk media there are only two fundamentally different directions in wavevector space, namely the longitudinal direction parallel to a given wavevector, and the transverse directions orthogonal to it. More specifically, given a wavevector  $\mathbf{k}$  of length  $k = |\mathbf{k}|$  in Fourier space,  $\hat{\mathbf{k}} \equiv \mathbf{k}/k$  is the corresponding longitudinal unit vector, where ‘longitudinal’ means that it points in the direction of  $\mathbf{k}$ . In addition, we can find two mutually orthogonal *transverse* unit vectors  $\mathbf{e}_{1,2}(\mathbf{k})$  that by definition are also orthogonal to  $\mathbf{k}$ . These three unit vectors together thus form an orthonormal basis for  $k$ -space. An arbitrary vector  $\mathbf{v}$  has a projection of length  $\mathbf{v} \cdot \hat{\mathbf{k}}$  in the longitudinal direction  $\hat{\mathbf{k}}$ . The longitudinal projection of the vector is thus given by  $(\mathbf{v} \cdot \hat{\mathbf{k}})\hat{\mathbf{k}}$ , which is the same as the inner product of  $\mathbf{v}$  with the dyadic product  $\hat{\mathbf{k}} \otimes \hat{\mathbf{k}}$ . Sometimes the symbol  $\otimes$  that signifies dyadic product is omitted, but here we keep it. The  $3 \times 3$  unit tensor  $\mathbf{I}$  in  $k$ -space can then be expanded as

$$\mathbf{I} = \hat{\mathbf{k}} \otimes \hat{\mathbf{k}} + (\mathbf{I} - \hat{\mathbf{k}} \otimes \hat{\mathbf{k}}) = \hat{\mathbf{k}} \otimes \hat{\mathbf{k}} + \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (2)$$

where on the right-hand side we wrote the matrix representation of  $\mathbf{I}$  in the  $\{\hat{\mathbf{k}}, \mathbf{e}_1(\mathbf{k}), \mathbf{e}_2(\mathbf{k})\}$  basis. For convenience we also introduce the longitudinal and transverse projectors

$$\mathbf{L}_{\mathbf{k}} \equiv \hat{\mathbf{k}} \otimes \hat{\mathbf{k}} \quad \text{and} \quad \mathbf{T}_{\mathbf{k}} \equiv \mathbf{I} - \hat{\mathbf{k}} \otimes \hat{\mathbf{k}}. \quad (3)$$

They are indeed projectors and together they project on the entire Fourier space because of their properties

$$\mathbf{L}_{\mathbf{k}} \cdot \mathbf{L}_{\mathbf{k}} = \mathbf{L}_{\mathbf{k}}, \quad \mathbf{T}_{\mathbf{k}} \cdot \mathbf{T}_{\mathbf{k}} = \mathbf{T}_{\mathbf{k}}, \quad \mathbf{L}_{\mathbf{k}} \cdot \mathbf{T}_{\mathbf{k}} = \mathbf{T}_{\mathbf{k}} \cdot \mathbf{L}_{\mathbf{k}} = \mathbf{0}, \quad \mathbf{L}_{\mathbf{k}} + \mathbf{T}_{\mathbf{k}} = \mathbf{I}. \quad (4)$$

In general the dyadic dielectric response in  $k$ -space will be different in the longitudinal and in the transverse directions. Even in isotropic media the dielectric response function need not be



a scalar. The most general response of an isotropic bulk medium is rather of the form

$$\boldsymbol{\varepsilon}(\mathbf{k}, \omega) = \boldsymbol{\varepsilon}_L(\mathbf{k}, \omega) + \boldsymbol{\varepsilon}_T(\mathbf{k}, \omega) = \varepsilon_L(k, \omega)\mathbf{L}_k + \varepsilon_T(k, \omega)\mathbf{T}_k = \begin{pmatrix} \varepsilon_L(k, \omega) & 0 & 0 \\ 0 & \varepsilon_T(k, \omega) & 0 \\ 0 & 0 & \varepsilon_T(k, \omega) \end{pmatrix}, \quad (5)$$

where the scalar functions  $\varepsilon_L(k, \omega)$  and  $\varepsilon_T(k, \omega)$  describe the response in the longitudinal and transverse directions, respectively. Because of the assumed isotropy of the medium, the scalar functions  $\varepsilon_{L,T}(k, \omega)$  depend only on the magnitude but not on the direction of the wavevector.

### 2.3. Real-space nonlocal response: from $k$ -space back to real space

For a given dielectric tensor in  $k$ -space, we can find the corresponding dielectric function in real space by inverse Fourier transformation:

$$\boldsymbol{\varepsilon}(\mathbf{r}, \omega) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k}_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}} \boldsymbol{\varepsilon}(\mathbf{k}_1, \omega). \quad (6)$$

We can do this inverse transformation for the longitudinal and transverse response separately. Let us first transform the longitudinal response  $\boldsymbol{\varepsilon}_L = \boldsymbol{\varepsilon}_L(\mathbf{k}, \omega)$  back to real space, making use of its property that the scalar function  $\varepsilon_L(k, \omega)$  depends only on the magnitude of the wavevector:

$$\boldsymbol{\varepsilon}_L(\mathbf{r}, \omega) = \frac{1}{(2\pi)^3} \int d^3\mathbf{k}_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}} \varepsilon_L(\mathbf{k}_1, \omega)\mathbf{L}_k = \frac{1}{(2\pi)^3} \int_0^\infty dk_1 k_1^2 \varepsilon_L(k_1, \omega) \int d^2\hat{\mathbf{k}}_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}} \hat{\mathbf{k}}_1 \otimes \hat{\mathbf{k}}_1. \quad (7)$$

So, because of the isotropic response, we can evaluate the three-dimensional integral in two steps: first do the  $4\pi$ -angular integral over wave vector directions  $\hat{\mathbf{k}}_1$ , and then the integral over the wave vector magnitude  $k_1$ . (Doing the magnitude integral first may also work in some cases.) The angular integral can be evaluated as follows (see the Appendix of Ref. [34]):

$$\int d^2\hat{\mathbf{k}}_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}} \hat{\mathbf{k}}_1 \otimes \hat{\mathbf{k}}_1 = -\frac{1}{k_1^2} \nabla \otimes \nabla \int d^2\hat{\mathbf{k}}_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}} = -\frac{4\pi}{k_1^2} \nabla \otimes \nabla \frac{\sin(k_1 r)}{k_1 r}. \quad (8)$$

By working out the spatial derivatives, one can show that this angular integral can be written in terms of spherical Bessel functions, see Appendix A and Ref. [34], but here we prefer instead to take the derivatives also out of the  $k_1$ -integral and with  $\varepsilon_L(-k_1, \omega) = \varepsilon_L(k_1, \omega)$  obtain

$$\boldsymbol{\varepsilon}_L(\mathbf{r}, \omega) = -\frac{1}{4\pi^2} \nabla \otimes \nabla \int_{-\infty}^\infty dk_1 \varepsilon_L(k_1, \omega) \frac{\sin(k_1 r)}{k_1 r}. \quad (9a)$$

Following analogous steps, the transverse response function in real space becomes

$$\boldsymbol{\varepsilon}_T(\mathbf{r}, \omega) = \varepsilon_T(\mathbf{r}, \omega)\mathbf{I} + \frac{1}{4\pi^2} \nabla \otimes \nabla \int_{-\infty}^\infty dk_1 \varepsilon_T(k_1, \omega) \frac{\sin(k_1 r)}{k_1 r}, \quad (9b)$$

where  $\varepsilon_T(\mathbf{r}, \omega)$  is the Fourier transform of the scalar function  $\varepsilon_T(k, \omega)$ .

We are interested in fully explicit forms of the real-space response functions  $\boldsymbol{\varepsilon}_{L,T}(\mathbf{r}, \omega)$ . The procedure to arrive at that in subsequent sections will be to assume specific model functions for  $\varepsilon_{L,T}(k_1, \omega)$ , to perform the  $k_1$ -integrals in Eqs. (9a) and (9b), and then to evaluate the double spatial derivatives. Here we make the general point that the result of doing the  $k_1$ -integral will be a radial function, i.e. depending on the length  $r$  rather than on the full position difference vector  $\mathbf{r}$ . That makes doing the spatial derivatives easier, since they follow immediately from the dyadic identity

$$\nabla \otimes \nabla f(r) = \frac{f'(r)}{r} (\mathbf{I} - \hat{\mathbf{r}} \otimes \hat{\mathbf{r}}) + f''(r) \hat{\mathbf{r}} \otimes \hat{\mathbf{r}} \quad (\text{for } r \neq 0), \quad (10)$$

which is derived and related to spherical harmonics in Appendix A. There is a sum rule for the real-space response functions that we will derive, namely that their integrals over all of space should give

$$\int d\mathbf{r} \boldsymbol{\varepsilon}_{L,T}(\mathbf{r}, \omega) = \int d^3\mathbf{k}_1 \left[ \frac{1}{(2\pi)^3} \int d\mathbf{r} e^{i\mathbf{k}_1 \cdot \mathbf{r}} \right] \boldsymbol{\varepsilon}_{L,T}(\mathbf{k}_1, \omega) = \boldsymbol{\varepsilon}_{L,T}(\mathbf{k} = \mathbf{0}, \omega), \quad (11)$$

where in the first equality we used Eq. (6) and in the second one we identified  $\delta(\mathbf{k}_1)$  in between the square brackets. The sum rule (11) will be a convenient consistency test of our results. On the right-hand side we see the  $k$ -space response in the limit  $k \rightarrow 0$ . This limit typically coincides with the local-response limit, and then the sum rule describes that the integral over the nonlocal response function equals that of the corresponding local-response model. In this sense, nonlocal-response models ‘smear out’ the response of the corresponding local model. There are infinitely many different functions that have the same  $k \rightarrow 0$  limit, and equally many ways to smear out the delta-function response of local models. Below we focus on specific response functions that in most cases are physically motivated.

#### 2.4. Relation between dyadic nonlocal response function and dyadic Green function

The dyadic Green function can conveniently be determined using the longitudinal and transverse projectors of Eq. (3). The retarded (dyadic) Green function for an infinite medium with nonlocal response is defined by the equation

$$-\nabla \times \nabla \times \mathbf{G}(\mathbf{r}, \omega) + (\omega/c)^2 \int d\mathbf{r}_1 \boldsymbol{\varepsilon}(\mathbf{r} - \mathbf{r}_1, \omega) \cdot \mathbf{G}(\mathbf{r}_1, \omega) = \mathbf{l}\delta(\mathbf{r}), \quad (12)$$

plus the boundary condition that  $\mathbf{G}(\mathbf{r}, \omega) \rightarrow 0$  for  $r \rightarrow \infty$ . The translation invariance means that it is best solved in  $k$ -space. Using the vector identity  $\hat{\mathbf{k}} \times \hat{\mathbf{k}} \times \mathbf{A} = (\hat{\mathbf{k}} \cdot \mathbf{A})\hat{\mathbf{k}} - (\hat{\mathbf{k}} \cdot \hat{\mathbf{k}})\mathbf{A} = -\mathbf{T}_{\mathbf{k}} \cdot \mathbf{A}$ , the defining equation becomes

$$-k^2 \mathbf{T}_{\mathbf{k}} \cdot \mathbf{G}(\mathbf{k}, \omega) + (\omega/c)^2 \boldsymbol{\varepsilon}(\mathbf{k}, \omega) \cdot \mathbf{G}(\mathbf{k}, \omega) = \mathbf{l}. \quad (13)$$

Now analogous to Eq. (5) for  $\boldsymbol{\varepsilon}$ , the Green function  $\mathbf{G}(\mathbf{k}, \omega)$  for isotropic translationally invariant media can be written as the sum of a transverse and a longitudinal part, where  $\mathbf{G}_{\mathbf{L}}(\mathbf{k}, \omega)$  equals  $G_{\mathbf{L}}(k, \omega)\mathbf{L}_{\mathbf{k}}$  and  $\mathbf{G}_{\mathbf{T}}(\mathbf{k}, \omega)$  equals  $G_{\mathbf{T}}(k, \omega)\mathbf{T}_{\mathbf{k}}$ , and where the  $G_{L,T}(k, \omega)$  are scalar functions depending only on the magnitude of the wave vector. Inserting this expansion into Eq. (13) and using the properties of the projectors  $\mathbf{L}_{\mathbf{k}}$  and  $\mathbf{T}_{\mathbf{k}}$  of Eq. (4), we obtain two scalar identities for the transverse and the longitudinal parts of the resulting equation separately, whereby the scalar Green functions can be expressed in terms of the dielectric response functions as

$$G_{\mathbf{L}}(k, \omega) = \frac{1}{(\omega/c)^2 \boldsymbol{\varepsilon}_{\mathbf{L}}(k, \omega)}, \quad G_{\mathbf{T}}(k, \omega) = \frac{-1}{k^2 - (\omega/c)^2 \boldsymbol{\varepsilon}_{\mathbf{T}}(k, \omega)}. \quad (14)$$

The poles of these Green functions give the dispersion relations for the excitations in the medium. Thus the longitudinal modes correspond to  $\boldsymbol{\varepsilon}_{\mathbf{L}}(k, \omega) = 0$  and the transverse modes to  $k^2 = (\omega/c)^2 \boldsymbol{\varepsilon}_{\mathbf{T}}(k, \omega)$ .

The corresponding real-space representations of the Green functions (14) are found by inverse Fourier transformation, where we can use the isotropy of the medium, analogous to Eqs. (9a) and (9b) for the longitudinal and transverse response functions. The result is therefore

$$\mathbf{G}_{\mathbf{L}}(\mathbf{r}, \omega) = -\frac{1}{4\pi^2} \nabla \otimes \nabla \int_{-\infty}^{\infty} dk_1 G_{\mathbf{L}}(k_1, \omega) \frac{\sin(k_1 r)}{k_1 r}, \quad (15a)$$

$$\mathbf{G}_{\mathbf{T}}(\mathbf{r}, \omega) = G_{\mathbf{T}}(\mathbf{r}, \omega)\mathbf{l} + \frac{1}{4\pi^2} \nabla \otimes \nabla \int_{-\infty}^{\infty} dk_1 G_{\mathbf{T}}(k_1, \omega) \frac{\sin(k_1 r)}{k_1 r}, \quad (15b)$$



where  $G_T(\mathbf{r}, \omega)$  is the three-dimensional inverse Fourier transform of the scalar function  $G_T(k, \omega)$ . The total Green function in real space  $\mathbf{G}(\mathbf{r}, \omega)$  is given by the sum of  $\mathbf{G}_L(\mathbf{r}, \omega)$  and  $\mathbf{G}_T(\mathbf{r}, \omega)$ . This shows how the total Green function can be computed once the scalar functions  $\varepsilon_{L,T}(k, \omega)$  are given.

### 3. $k$ -independent response [ $\varepsilon_L(\omega)$ and $\varepsilon_T(\omega)$ ]

#### 3.1. Local response [ $\varepsilon_L(\omega) = \varepsilon_T(\omega)$ ]

When the scalar functions  $\varepsilon_{L,T}(k, \omega)$  actually have no  $k$ -dependence, it may be expected that the corresponding response in real space will be local. In general this is incorrect, as shown below, but let us first consider the simplest case for which it is true: if the longitudinal and transverse responses are equal and independent of  $k$ ,  $\varepsilon_L(k, \omega) = \varepsilon_T(k, \omega) = \varepsilon(\omega)$ , then  $\boldsymbol{\varepsilon}(\mathbf{k}_1, \omega) = \varepsilon(\omega)\mathbf{I}$ . In this simple situation we do not need to consider longitudinal and transverse response separately since from Eqs. (2) and (6) it follows immediately that

$$\boldsymbol{\varepsilon}(\mathbf{k}, \omega) = \varepsilon(\omega)\mathbf{I} \quad \Leftrightarrow \quad \boldsymbol{\varepsilon}(\mathbf{r}, \omega) = \varepsilon(\omega)\delta^3(\mathbf{r})\mathbf{I}, \quad (16)$$

where the first  $\mathbf{I}$  denotes the  $3 \times 3$  unit matrix in three-dimensional  $k$ -space, while the second  $\mathbf{I}$  is the  $3 \times 3$  unit matrix in real space. Because of the delta function  $\delta^3(\mathbf{r})$ , we see that the response in this simple case is indeed local.

Local response models are employed in most of plasmonics research, and different metals are described by different permittivities  $\varepsilon(\omega)$ . No need to elaborate much on this here. For later reference we mention the widely used Drude-like models, with scalar local response functions of the form

$$\varepsilon_D(\omega) = \varepsilon_{\text{core}}(\omega) + \chi_D(\omega), \quad \text{with} \quad \chi_D(\omega) = -\frac{\omega_p^2}{\omega(\omega + i\gamma_D)}. \quad (17)$$

Here the free-carrier susceptibility  $\chi_D(\omega)$  features the plasma frequency  $\omega_p$  and Drude damping  $\gamma_D$ , while  $\varepsilon_{\text{core}}(\omega)$  (not further specified here) models the response from the bound ions and electrons, and includes effects of interband transitions.

#### 3.2. Nonlocal response model with local trace [ $\varepsilon_L(\omega)$ and $\varepsilon_T(\omega)$ ]

A slightly more complex situation occurs if the scalar transverse and longitudinal response functions are different, but still are independent of the wavevector, that is if we assume

$$\boldsymbol{\varepsilon}(\mathbf{k}_1, \omega) = \varepsilon_L(\omega)\mathbf{L}_k + \varepsilon_T(\omega)\mathbf{T}_k, \quad (18)$$

which has  $k$ -independent matrix elements in the basis of Eq. (2). The real-space response for this model will nevertheless be nonlocal, because the basis of Eq. (2) itself does depend on the wavevector, on its direction to be precise. We now calculate the spatial dispersion explicitly, starting with  $\boldsymbol{\varepsilon}_L(\mathbf{r}, \omega)$ :

$$\boldsymbol{\varepsilon}_L(\mathbf{r}, \omega) = \frac{\varepsilon_L(\omega)}{(2\pi)^3} \int d^3\mathbf{k}_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}} \mathbf{L}_k = \varepsilon_L(\omega)\boldsymbol{\delta}^{\parallel}(\mathbf{r}) = \varepsilon_L(\omega) \left[ \frac{1}{3}\delta^3(\mathbf{r})\mathbf{I} + \frac{1}{4\pi r^3}(\mathbf{I} - 3\hat{\mathbf{r}} \otimes \hat{\mathbf{r}}) \right], \quad (19)$$

in terms of the so-called longitudinal delta function  $\boldsymbol{\delta}^{\parallel}(\mathbf{r}) \equiv -\nabla \otimes \nabla(1/(4\pi r))$ , which is a dyadic, see the derivation in the Appendix and in Refs. [35,36]. Hence the longitudinal response does not solely give rise to a delta-function, but also to a ‘‘dipolar’’ nonlocal term scaling as  $1/r^3$ . Similarly, for the transverse part we find

$$\boldsymbol{\varepsilon}_T(\mathbf{r}, \omega) = \frac{\varepsilon_T(\omega)}{(2\pi)^3} \int d^3\mathbf{k}_1 e^{i\mathbf{k}_1 \cdot \mathbf{r}} \mathbf{T}_k = \varepsilon_T(\omega)\boldsymbol{\delta}^{\perp}(\mathbf{r}) = \varepsilon_T(\omega) \left[ \frac{2}{3}\delta^3(\mathbf{r})\mathbf{I} - \frac{1}{4\pi r^3}(\mathbf{I} - 3\hat{\mathbf{r}} \otimes \hat{\mathbf{r}}) \right], \quad (20)$$

in terms of the dyadic transverse delta function  $\delta^\perp(\mathbf{r})$  [35,36]. Again, besides the delta-function term that describes the local part of the response, there is a dipolar nonlocal term scaling with distance as  $1/r^3$ . From the above two expressions it is clear that the adjectives “longitudinal” and “transverse” refer to being parallel or perpendicular to the wavevector  $\mathbf{k}$  in Fourier space, rather than parallel or perpendicular to the position vector  $\mathbf{r}$  in real space. The total response of this model is the sum of its longitudinal and transverse parts, and is given by

$$\boldsymbol{\epsilon}(\mathbf{r}, \omega) = \frac{2\epsilon_T(\omega) + \epsilon_L(\omega)}{3} \delta^3(\mathbf{r}) \mathbf{I} + \frac{\epsilon_L(\omega) - \epsilon_T(\omega)}{4\pi r^3} (\mathbf{I} - 3\hat{\mathbf{r}} \otimes \hat{\mathbf{r}}). \quad (21)$$

Notice that in the special case of scalar response, i.e. if  $\epsilon_T(\omega) = \epsilon_L(\omega)$ , the nonlocal  $1/r^3$  terms of the transverse and longitudinal responses cancel exactly. Then Eq. (21) reduces as it should to the scalar local response found earlier in Eq. (16). On the other hand, when  $\epsilon_T(\omega)$  and  $\epsilon_L(\omega)$  differ, then the dyadic tensor  $\boldsymbol{\epsilon}(\mathbf{r}, \omega)$  is a non-scalar dyadic tensor. What does this mean in a bulk system? It means that the response in the direction defined by  $\mathbf{r}$  differs from the response in directions perpendicular to  $\mathbf{r}$ . In particular, along  $\mathbf{r}$  and perpendicular to it we find the diagonal tensor components

$$\hat{\mathbf{r}} \cdot \boldsymbol{\epsilon}(\mathbf{r}, \omega) \cdot \hat{\mathbf{r}} = \frac{2\epsilon_T(\omega) + \epsilon_L(\omega)}{3} \delta^3(\mathbf{r}) - \frac{\epsilon_L(\omega) - \epsilon_T(\omega)}{2\pi r^3}, \quad (22a)$$

$$\mathbf{e}_{1,2}(\mathbf{r}) \cdot \boldsymbol{\epsilon}(\mathbf{r}, \omega) \cdot \mathbf{e}_{1,2}(\mathbf{r}) = \frac{2\epsilon_T(\omega) + \epsilon_L(\omega)}{3} \delta^3(\mathbf{r}) + \frac{\epsilon_L(\omega) - \epsilon_T(\omega)}{4\pi r^3}, \quad (22b)$$

so the three-dimensional space is subdivided into a one-dimensional subspace parallel to  $\mathbf{r}$  and a two-dimensional subspace perpendicular to it. All off-diagonal tensor components are identically zero. We can partly characterize this dyadic response by a scalar, namely its average over all directions, in other words by taking one third of the trace of Eq. (21). The result is  $[2\epsilon_T(\omega) + \epsilon_L(\omega)]\delta^3(\mathbf{r})/3$ , which is a local quantity. In other words, the nonlocality of the dyadic response of this model does not show up in its trace. Related to this vanishing angle average, the nonlocal term also disappears when taking the volume average  $\int d\mathbf{r}$  of the dielectric response in Eq. (21), because the  $4\pi$  solid-angle integral over  $(\mathbf{I} - 3\hat{\mathbf{r}} \otimes \hat{\mathbf{r}})$  vanishes. In general one could hope that dyadic nonlocal response functions allow an approximately correct scalar description by taking the angle-averaged response, but the present model shows a qualitative change where taking the average turns a nonlocal model into a local one.

To summarize, in a model in which neither the  $k$ -longitudinal nor the  $k$ -transverse dielectric functions depend on the wavevector, we find that in real space there is nevertheless a concomitant spatial dispersion that scales as  $1/r^3$ , unless the two dielectric functions  $\epsilon_L(\omega)$  and  $\epsilon_T(\omega)$  are identical. This result also puts the commonly employed local response models in another perspective: in local-response models, neither the longitudinal nor the transverse responses are local; rather, the overall local response is the result of an exact cancellation between the longitudinal and the transverse nonlocal response functions.

### 3.3. Examples of nonlocal response models with local trace?

We have not come across situations in which different  $\epsilon_L(\omega)$  and  $\epsilon_T(\omega)$  are employed, so in that sense the nonlocal response model with local trace can be considered a toy model that at least gave some insight. It is a rather special model in the sense that the nonlocality does not involve a new length scale. Below we will see instead examples where either  $\epsilon_L$  or  $\epsilon_T$  has an explicit  $k$ -dependence, and where in the small- $k$  limit the two become equal. It is actually an interesting open question whether nonlocal response models with local trace may describe a physical state of affairs in some limit of other dyadic nonlocal response models. Toy models or not, below we will make use of the results obtained for this model in order to understand other nonlocal-response models that do have a clear physical motivation.

#### 4. Longitudinal nonlocal response [ $\epsilon_L(k, \omega)$ and $\epsilon_T(\omega)$ ]

Although we have just seen that even without explicit  $k$ -dependence in the longitudinal and transverse scalar functions the real-space response is typically still nonlocal, we now look at models where there *is* a  $k$ -dependence in the scalar longitudinal response response function. Let's call such models longitudinal nonlocal response models.

##### 4.1. Hydrodynamic Drude model

The hydrodynamic Drude model (HDM) in the Thomas-Fermi approximation in general describes a response of metals to light that is nonlinear [6, 12], but here we will only consider the linear-response limit. The linearized dynamics of the HDM can be described by two coupled equations of motion, in which the electric field is driven by the free-carrier current density and *vice versa*,

$$-\nabla \times \nabla \times \mathbf{E}(\mathbf{r}, \omega) + (\omega/c)^2 \epsilon_{\text{core}}(\omega) \mathbf{E}(\mathbf{r}, \omega) = -i\mu_0 \omega \mathbf{J}(\mathbf{r}, \omega) \quad (23a)$$

$$\frac{\beta^2}{\omega(\omega + i\gamma_D)} \nabla \otimes \nabla \cdot \mathbf{J}(\mathbf{r}, \omega) + \mathbf{J}(\mathbf{r}, \omega) = -i\epsilon_0 \omega \chi_D(\omega) \mathbf{E}(\mathbf{r}, \omega), \quad (23b)$$

where  $\beta$  is proportional to the Fermi velocity of the conductor. Our results will also apply to the mentioned generalized nonlocal optical response (GNOR) theory, a recent generalization of the HDM where due to diffusion the parameter  $\beta$  becomes complex-valued [17]. Mathematically, by taking the limit  $|\beta| \rightarrow 0$  the equation (23b) reduces to Ohm's law  $\mathbf{J} = \sigma \mathbf{E}$ . By inserting this into the first equation, we would obtain the Maxwell wave equation in a medium with local response as described by  $\epsilon_D(\omega)$  of Eq. (17). From the non-classical term proportional to  $\beta^2 \nabla \otimes \nabla \cdot \mathbf{J}$  in Eq. (23b) one can identify the nonlocal response in the HDM as due to non-classical pressure waves in the electron density; the observation of corresponding standing pressure waves in finite structures as in [14] is an important example of the predictive power of the HDM [19].

The above coupled equations indeed also describe finite systems when combined with the right boundary conditions. For infinite systems on the other hand, we can find the longitudinal and transverse response functions in  $k$ -space, where the above coupled equations turn into

$$-k^2 \mathbf{T}_{\mathbf{k}} \cdot \mathbf{E} + (\omega/c)^2 \epsilon_{\text{core}}(\omega) \mathbf{E} = -i\mu_0 \omega \mathbf{J} \quad \text{and} \quad -\frac{\beta^2 k^2}{\omega(\omega + i\gamma_D)} \mathbf{L}_{\mathbf{k}} \cdot \mathbf{J} + \mathbf{J} = -i\epsilon_0 \omega \chi_D \mathbf{E}, \quad (24)$$

where the variables  $(\mathbf{k}, \omega)$  were dropped for readability. Now both the electric field and the current density have longitudinal and transverse parts. By Eq. (4) we can decompose  $\mathbf{E}(\mathbf{k}, \omega)$  uniquely into the sum of  $\mathbf{E}_L = \mathbf{L}_{\mathbf{k}} \cdot \mathbf{E}$  and  $\mathbf{E}_T = \mathbf{T}_{\mathbf{k}} \cdot \mathbf{E}$ . This is essentially the Helmholtz decomposition carried out in  $k$ -space. Projection properties such as  $\mathbf{L}_{\mathbf{k}} \cdot \mathbf{E}_L = \mathbf{E}_L$  and  $\mathbf{T}_{\mathbf{k}} \cdot \mathbf{E}_L = \mathbf{0}$  etcetera follow immediately. An analogous decomposition exists for  $\mathbf{J}(\mathbf{k}, \omega)$ . By multiplying both equations in (24) from the left with  $\mathbf{T}_{\mathbf{k}}$ , one obtains two coupled equations involving only the transverse fields  $\mathbf{E}_T$  and  $\mathbf{J}_T$ , the solution of which gives the dispersion relation  $k^2 = \epsilon_T(k, \omega)(\omega/c)^2$ . Multiplying Eq. (24) from the left instead with  $\mathbf{L}_{\mathbf{k}}$  gives the longitudinal dispersion relation  $\epsilon_L(k, \omega) = 0$ . This way one obtains these response functions in the HDM as

$$\epsilon_T(\omega) = \epsilon_{\text{core}}(\omega) + \chi_D(\omega) = \epsilon_D(\omega), \quad (25a)$$

$$\epsilon_L(k, \omega) = \epsilon_{\text{core}}(\omega) - \frac{\omega_p^2}{\omega(\omega + i\gamma) - \beta^2 k^2}. \quad (25b)$$

In other words, the transverse response is the familiar Drude-like response well-known from local response models, while the longitudinal response is modified by a  $k$ -dependent term in

the denominator of the free-electron response. The parameter  $\beta$  is proportional to the Fermi velocity of the material. Since  $\epsilon_{\text{core}}(\omega)$  shows up both in the longitudinal and in the transverse response, the total response of the bound electrons is local in this model, whereas the response due to the free carriers is nonlocal, as we shall see. From Eq. (25) one can appreciate that the response in the HDM is both a generalization of the Drude-like local response of Eq. (17) and a special case of the general nonlocal response model (5). We note in passing that the self-consistent hydrodynamic model [26] that can describe electronic spill-out also belongs to the class of longitudinal nonlocal response models.

#### 4.2. Transverse response in real space in hydrodynamic model

From Eq. (25) it can be seen that the transverse dielectric function has no dependence on  $k$ , exactly as in the result Eq. (20) for the “nonlocal model with local trace” considered in Sec. 3.2. In fact, the integral is identical, and so is the result, namely that in the hydrodynamic Drude model the spatial dependence of the transverse response function is given by

$$\boldsymbol{\epsilon}_{\text{T}}(\mathbf{r}, \omega) = \epsilon_{\text{T}}(\omega) \boldsymbol{\delta}^{\perp}(\mathbf{r}) = \epsilon_{\text{D}}(\omega) \left[ \boldsymbol{\delta}(\mathbf{r}) \mathbf{I} + \nabla \otimes \nabla \left( \frac{1}{4\pi r} \right) \right] = \epsilon_{\text{D}}(\omega) \left[ \frac{2}{3} \boldsymbol{\delta}^3(\mathbf{r}) \mathbf{I} - \frac{\mathbf{I} - 3\hat{\mathbf{r}} \otimes \hat{\mathbf{r}}}{4\pi r^3} \right], \quad (26)$$

featuring the well-known Drude-like dielectric function of Eq. (17). Thus in the HDM we find terms in the spatial dependence of  $\boldsymbol{\epsilon}(\mathbf{r}, \omega)$  that fall off as  $1/r^3$ , originating from the transverse hydrodynamic response. Similar rather singular terms appeared in the total response of “non-local model with local trace”, and in the local-response limit they cancelled exactly against similar terms originating from the longitudinal response. It is thus interesting to see whether such cancellation also exists in the HDM, so let us now also consider the real-space version of the longitudinal response in the HDM.

#### 4.3. Longitudinal response in real space in hydrodynamic model

In order to obtain  $\boldsymbol{\epsilon}_{\text{L}}(\mathbf{r}, \omega)$  for the HDM, we insert into the integral Eq. (9a) the specific hydrodynamic form of  $\epsilon_{\text{L}}(k, \omega)$  as given in Eq. (25b), which has a bound-carrier part and a free-carrier part. For the first part we need the integral

$$\epsilon_{\text{core}}(\omega) \int_{-\infty}^{\infty} dk_1 \frac{\sin(k_1 r)}{k_1 r} = \frac{\pi \epsilon_{\text{core}}(\omega)}{r}, \quad (27)$$

while the integral over the hydrodynamic Drude part becomes

$$\int_{-\infty}^{\infty} dk_1 \left[ \frac{-\omega_p^2}{\omega(\omega + i\gamma) - \beta^2 k_1^2} \right] \frac{\sin(k_1 r)}{k_1 r} = \pi \left[ \frac{-\omega_p^2}{\omega(\omega + i\gamma)} \right] \frac{1 - e^{iqr}}{r} = \pi \chi_{\text{D}}(\omega) \frac{1 - e^{iqr}}{r}, \quad (28)$$

where  $q^2 \equiv \omega(\omega + i\gamma)/\beta^2$ . It is interesting to see the familiar local Drude susceptibility emerging here, but modified by the spatially dependent factor  $(1 - e^{iqr})$ . We can combine these two results and write the spatially dependent longitudinal response in the HDM as

$$\boldsymbol{\epsilon}_{\text{L}}(\mathbf{r}, \omega) = -\epsilon_{\text{D}}(\omega) \nabla \otimes \nabla \left( \frac{1}{4\pi r} \right) + \chi_{\text{D}}(\omega) \nabla \otimes \nabla \left( \frac{e^{iqr}}{4\pi r} \right), \quad (29)$$

with  $\epsilon_{\text{D}}(\omega)$  and  $\chi_{\text{D}}(\omega)$  as defined in Eq. (17).

#### 4.4. Total response in real space in hydrodynamic model

By adding up Eqs. (26) and (29), the same two terms cancel that cancel in local-response models, and we obtain the total real-space response of the HDM in the compact form

$$\boldsymbol{\varepsilon}(\mathbf{r}, \omega) = \varepsilon_D(\omega)\delta(\mathbf{r})\mathbf{I} + \chi_D(\omega)\nabla \otimes \nabla \left( \frac{e^{iqr}}{4\pi r} \right). \quad (30)$$

Here the first term on the right-hand side is the usual scalar local Drude-like response of Eq. (17). All effects of spatial dispersion are therefore described by the other term in Eq. (30). Sometimes nonlocal response is said to “smear out” the local delta-function response over a finite volume. We can now determine whether that is indeed what happens for the HDM, the archetypical model with nonlocal response. “Smearing out” would only occur if the second term of Eq. (30) besides smooth functions of  $r$  contains a delta-function against which some of the local free-carrier response of the first term cancels. Shortly we can and will work out the spatial derivatives of the second term using Eq. (10), but that identity is valid everywhere except in  $r = 0$ , and thus is blind for any delta-function contribution. But if for the moment it is only the delta-function contribution that we are interested in, then we can zoom in on  $\nabla \otimes \nabla [e^{iqr}/(4\pi r)]$  so close to the origin that  $|q|r \ll 1$  and find that the sought delta-function contribution indeed exists and equals that of the longitudinal delta function of Eq. (19). Thus we find that the second term in Eq. (30) was hiding a delta-function term  $-\chi_D(\omega)\delta(\mathbf{r})\mathbf{I}/3$ , which partly cancels the local response described by the first term of Eq. (30). It is interesting that the HDM thereby apparently “smears out” only one third of the local free-electron response over a finite spatial region, while leaving two thirds of the free-carrier local response and (of course) the entire local core response unaffected. We arrive at one of our main results, the explicit spatial response of the HDM as

$$\boldsymbol{\varepsilon}(\mathbf{r}, \omega) = \left[ \varepsilon_{\text{core}}(\omega) + \frac{2}{3}\chi_D(\omega) \right] \delta(\mathbf{r})\mathbf{I} + \frac{\chi_D(\omega)e^{iqr}}{4\pi r^3} [(iqr - 1)(\mathbf{I} - 3\hat{\mathbf{r}} \otimes \hat{\mathbf{r}}) - q^2 r^2 \hat{\mathbf{r}} \otimes \hat{\mathbf{r}}]. \quad (31)$$

We have not yet shown that the “one third nonlocal smearing out” story indeed holds, because we have only shown that one third of the local free-carrier response was removed, but we have not checked whether exactly this part was smeared out elsewhere. This balance can be tested by taking the volume integral of Eq. (31). Now the solid-angle integral  $\int_{\partial\Omega} \hat{\mathbf{r}} \otimes \hat{\mathbf{r}}/(4\pi)$  equals  $\mathbf{I}/3$ , so that the integral over terms proportional to  $(\mathbf{I} - 3\hat{\mathbf{r}} \otimes \hat{\mathbf{r}})$  vanishes. The last term in Eq. (31), the one proportional to  $\hat{\mathbf{r}} \otimes \hat{\mathbf{r}}$ , does contribute to the volume integral, and it does so in a neat way since it ensures that indeed  $\int d\mathbf{r} \boldsymbol{\varepsilon}(\mathbf{r}, \omega)$  equals  $\varepsilon_D(\omega)\mathbf{I}$ , the same outcome as for the local-response Drude-like model. This agrees with the sum rule (11) and it also implies that the nonlocal term in Eq. (30) had a vanishing volume integral. But most importantly, it entails that indeed in the HDM one third (no more and no less) of the free-carrier response is smeared out nonlocally. This interesting result is a direct consequence of the dyadic nature of the HDM. In  $k$ -space the response is only  $k$ -dependent in the longitudinal direction, not in the two other directions, which makes the HDM only one third non-local. This is a rather general argument to explain the factor 1/3 and we did not use any specifics of the hydrodynamic model other than that it belongs to the longitudinal nonlocal class. We therefore anticipate that the argument holds for the entire class, which for example includes models with a fourth-order polynomial of  $k$  in the denominator of  $\varepsilon_L(k, \omega)$  instead of the parabolic  $k$ -dependence for the HDM in Eq. (25b).

Intuitively one might have expected that nonlocal response gives rise to a smooth real-space response function, but for the HDM the expression Eq. (31) shows several diverging near-field terms. In the near field (i.e. for  $qr \ll 1$ ), the nonlocal terms are proportional to the longitudinal delta function, thus scaling as  $r^{-3}$ , both in directions parallel and perpendicular to  $\mathbf{r}$ . By

contrast, in the far field ( $qr \gg 1$ ) the dyadic response along  $\mathbf{r}$  scales with distance as  $e^{iqr}/r$ , while perpendicular to  $\mathbf{r}$  the response falls off faster, namely as  $e^{iqr}/r^2$ . So in that sense, the  $k$ -dependence of the  $\mathbf{k}$ -longitudinal response  $\epsilon_L(k, \omega)$  gives rise to a nonlocal response in real space that is also predominantly  $\mathbf{r}$ -longitudinal, stronger along  $\mathbf{r}$  than perpendicular to it.

The response function  $\boldsymbol{\epsilon}(\mathbf{r}, \omega)$  in Eq. (31) is an integration kernel in Maxwell's equations, in particular in the constitutive relation Eq. (1). Because of the diverging terms in the kernel, evaluating the integral may be numerically much more costly than in case of a Gaussian kernel, say. Instead of solving nonlocal-response problems as an integro-differential equation, it is therefore often simpler to solve coupled differential equations such as in Eq. (23). But one could try to approximate the kernel to make the integro-differential approach more manageable. A first approximation could be to neglect the dipole term out of the kernel (31), since its volume integral vanishes. A further approximation would be to map the dyadic HDM response onto the scalar response function that captures it best. We propose to use the angle-averaged response, obtained by taking one third of the trace of Eq. (31) and multiplying by  $\mathbf{I}$ . The result is

$$\boldsymbol{\epsilon}_{\text{av.}}(\mathbf{r}, \omega) = \frac{\mathbf{I}}{3} \text{Tr}[\boldsymbol{\epsilon}(\mathbf{r}, \omega)] = \left[ \epsilon_{\text{core}}(\omega) + \frac{2}{3} \chi_{\text{D}}(\omega) \right] \delta(\mathbf{r}) \mathbf{I} - \frac{\mathbf{I}}{3} \frac{q^2 \chi_{\text{D}}(\omega) e^{iqr}}{4\pi r}. \quad (32)$$

The first thing to appreciate is that tracing out the HDM response does give a nonlocal response, unlike for the nonlocal model in Sec. 3.2 that had a local trace. The advantages of this scalar model in capturing the HDM response would be that by construction its spatial average is the same as for the HDM, and secondly that here also only one third of the usual local free-electron response is smeared out, and in the third place the  $r$ -dependence is more similar to the HDM response than a Gaussian function would be. The numerical accuracy of this tailor-made scalar function in capturing hydrodynamic nonlocal response in phenomenological approaches such as in Ref. [32] would be interesting for further study. For more scalar models see Sec. 6.1 and for a warning about using them we refer to the Discussion in Sec. 7.

## 5. Transverse nonlocal response [ $\epsilon_L(\omega)$ and $\epsilon_T(k, \omega)$ ]

While in the previous section we considered models with  $k$ -dependent longitudinal and  $k$ -independent transverse scalar response functions, here we consider just the opposite: models defined by a  $\epsilon_L(\omega)$  in combination with a function  $\epsilon_T(k, \omega)$  that has an explicit dependence on the wavevector. We discuss some physical predictions of such models in Sec. 7 below, but first let us specify one such a model and study its response in real space. A main question is whether the real-space response function is qualitatively different from the one of the HDM.

We study the model from the recent work of Ref. [8], which in our notation becomes

$$\epsilon_L(\omega) = \epsilon_{\text{core}}(\omega) + \chi_{\text{free}}(\omega), \quad \text{with} \quad \chi_{\text{free}}(\omega) = -\frac{\omega_p^2}{\omega(\omega + i\gamma_b) - \Delta^2}, \quad (33a)$$

$$\epsilon_T(k, \omega) = \epsilon_{\text{core}}(\omega) - \frac{\omega_p^2}{\omega(\omega + i\gamma_b) - \Delta^2 - \beta_T^2 k^2}, \quad (33b)$$

and which was proposed as a physical model of the nonlocal response in semiconductor spheres with very low doping. The parameter  $\Delta$  is a resonance frequency and in the limit  $\Delta \rightarrow 0$ , the free-electron susceptibility  $\chi_{\text{free}}(\omega)$  reduces to the Drude response  $\chi_{\text{D}}(\omega)$  that we considered above. The corresponding real-space longitudinal response is simple and follows the longitudinal part of the "toy model", with a spatial dependence of the longitudinal delta function, see Eq. (19). The real-space transverse response requires a bit more work, and after adding up the



longitudinal and transverse contributions one finds the total response function

$$\begin{aligned} \boldsymbol{\varepsilon}(\mathbf{r}, \omega) &= \mathbf{l}\delta(\mathbf{r}) \left[ \varepsilon_{\text{core}}(\omega) + \frac{1}{3}\chi_{\text{free}}(\omega) \right] \\ &- \frac{\chi_{\text{free}}(\omega)e^{iQr}}{4\pi r^3} [(iQr-1)(\mathbf{l}-3\hat{\mathbf{r}}\otimes\hat{\mathbf{r}}) + Q^2r^2(\mathbf{l}-\hat{\mathbf{r}}\otimes\hat{\mathbf{r}})], \end{aligned} \quad (34)$$

where the parameter  $Q \equiv \sqrt{\omega(\omega + i\gamma_b) - \Delta^2}/\beta_T$ . So for vanishing  $\Delta$  and by replacing  $\beta_T$  by  $\beta$  this  $Q$  coincides with  $q$  as introduced earlier. As follows from the first line of this expression, in the transverse nonlocal model, two-thirds of the free-electron response is smeared out due to the nonlocality, whereas one third remains a delta-function response. So again we find that the nonlocal spreading is only partial, but this time twice as much as in the hydrodynamic model. This is directly related to the dyadic nature of this transverse nonlocal-response model, and since in  $k$ -space the response is nonlocal in two out of three independent directions, namely the two transverse directions perpendicular to  $\mathbf{k}$ , it starts to become intuitive that only two-thirds of the free-electron response becomes nonlocal. It can be shown that the integral over all of space of Eq. (34) indeed gives the local-response result  $\mathbf{l}[\varepsilon_{\text{core}}(\omega) + \chi_{\text{free}}(\omega)]$ , in accordance with the sum rule of Eq. (11).

## 6. Both longitudinal and transverse nonlocal response [ $\varepsilon_L(k, \omega)$ and $\varepsilon_T(k, \omega)$ ]

### 6.1. Scalar models [ $\varepsilon_L(k, \omega) = \varepsilon_T(k, \omega)$ ]

Scalar models for nonlocal response have also been proposed in the literature [7, 37], and most phenomenological models are also scalar [32, 33]. In the light of the previous models, the specialty about scalar models is that their longitudinal and scalar response functions are identical,

$$\varepsilon_L(k, \omega) = \varepsilon_T(k, \omega) = \varepsilon(k, \omega), \quad \text{so that} \quad \boldsymbol{\varepsilon}(\mathbf{k}, \omega) = \varepsilon(k, \omega)\mathbf{l}. \quad (35)$$

In the absence of an explicit  $k$ -dependence, such scalar models reduce to simple local-response models. Ref. [7] assumes  $\varepsilon(k, \omega)$  to have the functional form of  $\varepsilon_T(k, \omega)$  in Eq. (33b). Here we focus instead on the interesting scalar model of Ref. [37] where  $\varepsilon(k, \omega)$  is assumed to be of the same hydrodynamic form as  $\varepsilon_L(k, \omega)$  in Eq. (25b). The crucial difference with the real HDM is that in the scalar model also an explicit  $k$ -dependence of the transverse response is assumed. The model can be derived from two coupled equations analogous to the two equations for the HDM, the only difference being that in the second equation, that is in Eq. (23b), the  $\nabla \otimes \nabla$  is replaced by a Laplacian  $\nabla^2\mathbf{l}$ :

$$\left[ \frac{\beta^2}{\omega(\omega + i\gamma_D)} \nabla^2\mathbf{l} + \mathbf{l} \right] \cdot \mathbf{J}(\mathbf{r}) = -i\varepsilon_0\omega\chi_D(\omega)\mathbf{E}(\mathbf{r}) \Rightarrow \left[ -\frac{\beta^2k^2\mathbf{l}}{\omega(\omega + i\gamma_D)} + \mathbf{l} \right] \cdot \mathbf{J}(\mathbf{k}) = -i\varepsilon_0\omega\chi_D\mathbf{E}(\mathbf{k}), \quad (36)$$

The interpretation of nonlocal response as due to pressure waves in the electron density is lost after this replacement. Doing the same longitudinal and transverse projections as for the HDM in Sec. 4, we obtain that now indeed  $\varepsilon_L(k, \omega)$  and  $\varepsilon_T(k, \omega)$  are equal so that  $\boldsymbol{\varepsilon}(\mathbf{k}, \omega)$  is scalar and given by the right-hand side of Eq. (25b). This leads to the scalar real-space nonlocal response function

$$\varepsilon(\mathbf{r}, \omega) = \mathbf{l} \left[ \varepsilon_{\text{core}}(\omega)\delta(\mathbf{r}) + \frac{\omega_p^2}{\beta^2} \frac{e^{iqr}}{4\pi r} \right], \quad (37)$$

where  $q = \sqrt{\omega(\omega + i\gamma_D)}/\beta$  as before. This scalar model does not have the same strongly divergent terms scaling as  $r^{-3}$  for small  $r$  as we found for the HDM, but a  $1/r$  divergence still remains, and on top of that  $\varepsilon(\mathbf{r}, \omega)$  is again complex-valued. Unlike in the HDM, here *all*

free-carrier response is smeared out over a finite volume, because nonlocal response was assumed for all three independent directions in  $k$ -space. The volume integral of Eq. (37) indeed equals  $\epsilon_D(\omega)\mathbf{I}$ , in accordance with the general sum rule (11). The scalar model (37) that we just obtained can be contrasted with the traced-out HDM model of Eq. (32). Only the latter has the correct amount of smearing out of the local response, whereas Eq. (37) overestimates the amount of nonlocality by a factor of three, at least when assuming that the HDM is accurate.

### 6.2. Non-scalar models with L- and T-nonlocal response [ $\epsilon_L(k, \omega) \neq \epsilon_T(k, \omega)$ ]

In the general case that both the transverse and the longitudinal scalar response functions  $\epsilon_{L,T}(k, \omega)$  are explicitly  $k$ -dependent but different, the total real-space response functions can be calculated by adding Eqs. (9a) and (9b). In this case all of the local free-space response will be smeared out, one third as determined by the typical length scales in  $\epsilon_L(k, \omega)$ , and two thirds on the length scales of  $\epsilon_T(k, \omega)$ . When assuming a hydrodynamic form of the response functions, this would correspond to different  $\beta$  parameters for the longitudinal and for the transverse scalar responses. Then again the total integral of  $\epsilon(\mathbf{r}, \omega)$  will give the local Drude  $\epsilon_D(\omega)\mathbf{I}$ , because the sum rule (11) holds for longitudinal and transverse parts separately.

The microscopic RPA calculations for a free-electron gas by Lindhard [38] also lead to response functions  $\epsilon_{L,T}(k, \omega)$  that are both nonlocal (and different). The main difference with the hydrodynamic Drude model is obviously that the latter has a local transverse response. But also the wavevector dependence of the longitudinal response is different, with more pronounced differences for larger  $k$ . In the long-wavelength approximation on the other hand, Lindhard's  $\epsilon_{L,T}(k, \omega)$  both tend to the same classical Drude form, so we do not obtain a nonlocal model with local trace (see Sec. 3.2) by taking the  $k \rightarrow 0$  limit. For applications of the Lindhard response functions also for finite systems, see Ref. [4].

## 7. Discussion

We have compared scalar, longitudinal, and transverse models for nonlocal response and already for translationally invariant systems found qualitatively different predictions, for example whether all local free-carrier response gets nonlocally distributed or not. If we suppose that longitudinal nonlocal models describe metal nanospheres accurately and transverse nonlocal models apply to low-doped semiconductors, can one then still call scalar nonlocal response models phenomenological? It all depends how well they describe physical phenomena.

For finite structures the HDM predicts novel resonances (standing pressure waves of the electron density) due to nonlocal response only above the plasma frequency, and indeed these have been observed experimentally [14]. It was found for metals that scalar nonlocal models with hard-wall boundary conditions exhibit novel nonlocal standing-wave resonances also *below* the plasma frequency, for example in Ref. [37] (and follow-up papers) where a scalar hydrodynamic-like response function was assumed that in real space becomes Eq. (37). Likewise in Ref. [32] with assumed Gaussian scalar nonlocal response functions, near the dipole resonance new standing-wave patterns were predicted on a scale determined by the strength of the nonlocality. Since experimental observations of such resonances in metallic nanoparticles to date have not been reported, the longitudinal hydrodynamic model seems to be the better simple model of nonlocal response in metals. Further discussion and graphical comparisons of the resonances in scalar models as compared to the HDM are given in Ref. [39]. It should be mentioned that non-classical plasmonic resonances of another kind *have* been observed below the plasma frequency for some metals, namely the so-called multipole surface plasmons (reviewed in [4]), but they are related to electronic spill-out at interfaces which classical electrodynamics and the standard HDM both neglect; the first semiclassical model to describe these non-classical resonances is the self-consistent hydrodynamic model [26].

In Eq. (32) we constructed the scalar nonlocal response function that comes closest to the standard HDM. It is an interesting question for future studies how well it captures the predictions of the HDM. However, one concern should be that ‘scalarizing’ the HDM response may again lead to unphysical predictions of novel nonlocal standing-wave resonances in nanostructures. If so, then one should perhaps give up attempts to model nonlocal response phenomena in metals by scalar functions. This would still not rule out transverse nonlocal response in metals, but only the assumption that the transverse response function equals the longitudinal one.

The for metals not-so-physical novel nonlocal resonances below the plasma frequency calculated in Refs. [32, 37] are related to the fact that also the transverse response was assumed to be nonlocal [39]. Let us assume that Ref. [8] is right in modeling the nonlocal response of ultra-low doped semiconductor particles with a transverse nonlocal model. (But here the issue is not settled either, Ref. [8] should be compared with the scalar nonlocal model for doped semiconductors in Ref. [7] and with the longitudinal nonlocal model in Ref. [9]). Then the very interesting question arises whether this transverse model gives rise to new resonances below the plasma frequency, just like in the scalar models of Refs. [32, 37]. If so, then it is to be expected that new nonlocal resonances will show up below the plasma frequency when miniaturizing doped semiconductor nanoparticles. That would constitute a qualitative difference in the nonlocal response of doped semiconductors as compared to metals. Indeed, Ref. [8] features spectral asymmetries and oscillatory behavior on the high-energy side that could very well be due to such novel nonlocal resonances below the plasma frequency.

## 8. Conclusions

We considered different classes of scalar and dyadic nonlocal response models, and in particular their real-space behavior. We were inspired by the intuitive scalar models of Refs. [32, 33] that stressed the smearing out of the local response onto a finite region of space. That is what we also found here, both for scalar and for dyadic models. The simple sum rule Eq. (11) describes exactly this: the spatial integral of the nonlocal response functions equals that of the corresponding local-response model. So whatever delta-function response gets ‘missing’ due to nonlocality is balanced by the sum of the response elsewhere. This equality holds for longitudinal and transverse response functions separately.

What the sum rule does not describe is *how much* of the local response is smeared out in this way. The phenomenological models Refs. [32, 33] assumed implicitly that all free-carrier response is redistributed. That is indeed what we find for scalar models, but for longitudinal nonlocal models including the hydrodynamic Drude model we found to our initial surprise that only one third of the free-electron response is redistributed nonlocally, while for the class of transverse nonlocal models we find two thirds. In this sense scalar models are more nonlocal than transverse ones, which in turn are more nonlocal than longitudinal ones. We explained these properties of real-space response functions back in  $k$ -space, where longitudinal nonlocal response modified only one out of three directions (i.e. only along  $\mathbf{k}$ ), and transverse nonlocal response changed only the two other directions (only normal to  $\mathbf{k}$ ). These general arguments apply to the whole classes of models.

Another clear difference with phenomenological models is that the nonlocal response functions  $\boldsymbol{\epsilon}(\mathbf{r}_1 - \mathbf{r}_2)$  that we found for some important dyadic models are interesting and rather non-Gaussian: no simple non-negative distributions but instead complex-valued functions that diverge as  $r^{-3}$  for  $r \rightarrow 0$ . Even for the scalar version of the hydrodynamic model a  $1/r$  divergence remains. Although these divergencies are integrable, as guaranteed by the sum rule Eq. (11), it does explain that solving these nonlocal response problems as integro-differential equations can be slow, and slower for dyadic models such as the HDM than for scalar ones.

We finally suggested that novel resonances below the plasma frequency may arise in trans-

verse nonlocal models. If so, and if the transverse nonlocal model of Ref. [8] indeed applies to doped semiconductors, then effects of novel collective nonlocal resonances may be observed in semiconductor nanoparticles that have no counterpart in metal nanoparticles.

### A. Derivation of Eq. (10) and connection with spherical harmonics

By the chain rule, one obtains the identity for radial functions  $\nabla f(r) = f'(r)\nabla r$ , and by repeating this chain rule once more one obtains  $\nabla \otimes [\nabla f(r)] = \nabla \otimes [f'(r)\nabla r] = f''(r)\nabla r \otimes \nabla r + f'(r)\nabla \otimes \nabla r$ . By writing out in Cartesian coordinates it is easy to show that  $\nabla r = \mathbf{r}/r = \hat{\mathbf{r}}$ , and after only a little more work that  $\nabla \otimes \nabla r = (\mathbf{I} - \hat{\mathbf{r}} \otimes \hat{\mathbf{r}})/r$ . From the previous identities it then follows that

$$\nabla \otimes \nabla f(r) = \frac{f'(r)}{r} (\mathbf{I} - \hat{\mathbf{r}} \otimes \hat{\mathbf{r}}) + f''(r)\hat{\mathbf{r}} \otimes \hat{\mathbf{r}}, \quad (38)$$

which is the sought Eq. (10) of the main text. Applying this formula to  $f(r) = -1/(4\pi r)$  gives

$$-\nabla \otimes \nabla \left( \frac{1}{4\pi r} \right) = \left( \frac{\mathbf{I} - 3\hat{\mathbf{r}} \otimes \hat{\mathbf{r}}}{4\pi r^3} \right) + \frac{1}{3}\delta(\mathbf{r})\mathbf{I} = \boldsymbol{\delta}^{\parallel}(\mathbf{r}), \quad (39)$$

which gives the longitudinal delta function. Actually only the dipole terms proportional to  $r^{-3}$  follow from Eq. (38) that was derived for  $\mathbf{r} \neq \mathbf{0}$ , thereby missing that a delta-function term must be added, as shown in a few lines in Ref. [35]. To make the connection with spherical harmonics, notice that Eq. (38) can be rewritten as

$$\nabla \otimes \nabla f(r) = -\mathbf{I} \left[ (-r) \left( \frac{1}{r} \frac{d}{dr} \right) \right] f(r) + \hat{\mathbf{r}} \otimes \hat{\mathbf{r}} \left[ (-r)^2 \left( \frac{1}{r} \frac{d}{dr} \right)^2 \right] f(r). \quad (40)$$

In the special case that  $f(r) = \sin(k_1 r)/(k_1 r)$  this leads to the identity presented in Ref. [34],

$$-\frac{1}{k_1^2} \nabla \otimes \nabla \frac{\sin(k_1 r)}{k_1 r} = \frac{j_1(k_1 r)}{k_1 r} \mathbf{I} - j_2(k_1 r) \hat{\mathbf{r}} \otimes \hat{\mathbf{r}}, \quad (41)$$

in terms of the spherical Bessel functions  $j_n$  that are defined by

$$j_n(x) \equiv (-x)^n \left( \frac{1}{x} \frac{d}{dx} \frac{1}{x} \frac{d}{dx} \right)^n \frac{\sin x}{x}. \quad (42)$$

### Acknowledgments

Stimulating discussions in the Structural Electromagnetic Materials group at DTU Fotonik are gratefully acknowledged, in particular with Asger Mortensen, Thomas Christensen, Søren Raza, Nicolas Stenger, Giuseppe Toscano, and Wei Yan. Funding is acknowledged from the Danish Council for Independent Research, Grant No. 1323-00087. The Center for Nanostructured Graphene is sponsored by the Danish National Research Foundation, Project DNRF58.