Probing nonlocal effects in metals with graphene plasmons

Eduardo J. C. Dias,[1,*] David Alcaraz Iranzo,[2] P. A. D. Gonçalves,[3,4,5] Yaser Hajati,[6] Yuliy V. Bludov,[1] Antti-Pekka Jauho,[5,7] N. Asger Mortensen,[3,5,8] Frank H. L. Koppens,[2,9] and N. M. R. Peres[1,†]

[1]*Department of Physics and Center of Physics, and QuantaLab, University of Minho, PT–4710–057, Braga, Portugal*
[2]*ICFO - The Institute of Photonic Sciences, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*
[3]*Center for Nano Optics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark*
[4]*Department for Photonics Engineering, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark*
[5]*Center for Nanostructured Graphene, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark*
[6]*Department of Physics, Faculty of Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran*
[7]*Department of Micro- and Nanotechnology, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark*
[8]*Danish Institute for Advanced Study, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark*
[9]*ICREA - Institució Catalana de Recera i Estudis Avançats, Barcelona, Spain*

(Received 17 April 2018; published 11 June 2018)

In this paper, we analyze the effects of nonlocality on the optical properties of a system consisting of a thin metallic film separated from a graphene sheet by a hexagonal boron nitride (hBN) layer. We show that nonlocal effects in the metal have a strong impact on the spectrum of the surface plasmon-polaritons on graphene. If the graphene sheet is nanostructured into a periodic grating, we show that the resulting extinction curves can be used to shed light on the importance of nonlocal effects in metals. Therefore graphene surface plasmons emerge as a tool for probing nonlocal effects in metallic nanostructures, including thin metallic films. As a byproduct of our study, we show that nonlocal effects may lead to smaller losses for the graphene plasmons than what is predicted by a local calculation. Finally, we demonstrate that such nonlocal effects can be very well mimicked using a local theory with an effective spacer thickness larger than its actual value.

DOI: 10.1103/PhysRevB.97.245405

I. INTRODUCTION

Nanoplasmonics is a field of optics dealing with the interaction of electromagnetic radiation with metallic nanostructures and nanoparticles [1–4]. Over the last decade, the characteristic size of plasmonic structures has been steadily approaching the few-nanometer scale [5], with concomitantly ultraconfined plasmonic modes [6–9]. The most common description of the electrodynamics governing plasmonic systems typically ignores that the response of a metallic nanostructure is controlled by a nonlocal dielectric tensor [10,11]. Indeed, the general linear-response expression for the electric displacement vector reads [12]

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

where $\mathbf{D}(\mathbf{r},\omega)$ is the electric displacement vector, $\mathbf{E}(\mathbf{r},\omega)$ is the electric field, and $\overleftrightarrow{\epsilon}(\mathbf{r},\mathbf{r}',\omega)$ is the nonlocal dielectric tensor. For a translationally invariant system, we have $\overleftrightarrow{\epsilon}(\mathbf{r},\mathbf{r}',\omega) = \overleftrightarrow{\epsilon}(\mathbf{r}-\mathbf{r}',\omega)$. Equation (1) embodies the statement that the electric displacement field at point $\mathbf{r}$ depends on the electric field at all points $\mathbf{r}'$.

In general, and in particular for systems with translational invariance, it is convenient to transform Eq. (1) to momentum space, obtaining $\mathbf{D}(\mathbf{k},\omega) = \varepsilon_0 \overleftrightarrow{\epsilon}(\mathbf{k},\omega)\cdot\mathbf{E}(\mathbf{k},\omega)$, where translational invariance has been assumed and $\mathbf{k}$ denotes the wave vector. The local response approximation (LRA) is equivalent to neglecting the wave vector dependence of the dielectric tensor by taking the long wavelength limit ($\mathbf{k}\to 0$). However, when the system's characteristic length scales approach the few-nanometer range, the wave vector dependence of the dielectric tensor has profound consequences on the spectrum of plasmonic resonances. It follows from Maxwell's equations that the wave equation in momentum space reads [3]

$$\begin{aligned} -\mathbf{k}\times[\mathbf{k}\times\mathbf{E}(\mathbf{k},\omega)] &= -\mathbf{k}[\mathbf{k}\cdot\mathbf{E}(\mathbf{k},\omega)] + k^2\mathbf{E}(\mathbf{k},\omega) \\ &= \frac{\omega^2}{c^2}\overleftrightarrow{\epsilon}(\mathbf{k},\omega)\cdot\mathbf{E}(\mathbf{k},\omega). \end{aligned} \tag{2}$$

Equation (2) has two types of solutions. The divergence-free solution ($\mathbf{k}\cdot\mathbf{E}=0$) corresponds to the usual wave equation $k^2\mathbf{E} = (\omega/c)^2\epsilon(\mathbf{k},\omega)\cdot\mathbf{E}$ (where we have implicitly assumed that the dielectric tensor is isotropic); in this context, this solution is usually dubbed as the transverse mode. However, within the nonlocal framework, there is an additional curl-free solution ($\mathbf{k}\times\mathbf{E}=\mathbf{0}$), which, for a given frequency ω, is obtained when the condition $\varepsilon(\mathbf{k},\omega)=0$ is satisfied. This solution is traditionally referred to as the longitudinal mode.

The longitudinal solution is generally overlooked within the LRA, which may reveal itself as an inaccurate approximation when either the size of the metallic nanostructures or the separation between two metallic surfaces fall below a couple of tens of nanometers [7,13,14]. Here, we consider

*eduardo.dias@fisica.uminho.pt
†peres@fisica.uminho.pt

 ©2018 American Physical Society

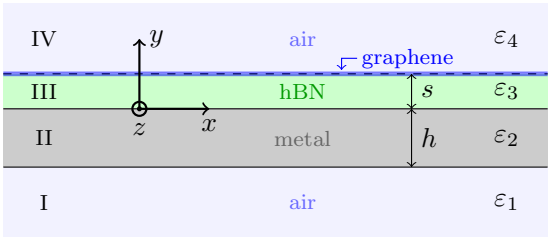

FIG. 1. Schematic representation of one type of layered heterostructure investigated in this work. We assume the system to be invariant in the x and z directions. We stress here that throughout this work we use the full nonlocal optical conductivity of graphene. For the metallic film, on the other hand, we study the cases where the metal optical properties are treated both locally and nonlocally.

a configuration in which a graphene sheet is placed parallel to and extremely close to a metal surface—see Fig. 1. The polaritonic spectrum of the aforementioned structures can be computed either analytically or numerically—depending of the geometry—and such calculations show that the resonances associated with the excitation of surface plasmon-polaritons (either localized or propagating) can exhibit strong deviations from the ones obtained within the LRA [15,16]. Additionally, the local response calculation predicts field enhancements which are larger than in the nonlocal case [17–19]. This is particularly true when we have two nanoparticles in close proximity [2,17,20–22]. Experimentally, one can investigate nonlocal effects using diverse techniques, such as electron energy-loss spectroscopy [8,20] and far-field spectroscopy [23]. Therefore a proper theoretical description of nonlocal plasmonics requires the calculation of a suitable nonlocal dielectric function (or, equivalently, the nonlocal conductivity). A possible approach is to use *ab initio* methods [24,25]. However, typical (albeit small) plasmonic structures involve a large number of atoms which render those methods difficult to use in a routinely fashion, since they quickly become time-consuming and computationally demanding. Hence it is natural to seek for an alternative approach to the calculation of the nonlocal dielectric function. As early as in the 1970s, it was observed that the plasmonic properties of thin metallic films did not follow the prediction of the local-response theory [26,27]. The need to reconcile theory and experiment required a description of an electron gas subjected to external fields introduced by Bloch [28,29] and Jensen [30]. This model of an electronic fluid came to be known as the hydrodynamic model [13,27,31–33]. This model combines Maxwell's equations with Newton's second law of motion for a charged particle supplemented with a term taking into account the statistical pressure in the electron gas due to charge inhomogeneity [13,26,27]. The revival of plasmonics unburied the hydrodynamic model and applied it to the optical response of metallic nanoparticles with great success [7,22,32]. As already mentioned, both the position and the width of the plasmon resonance, and the field enhancement measured experimentally do not agree with the approach using the LRA, whereas the hydrodynamic model is able to explain the experimental results [7]. For this reason, the hydrodynamic model has now become a popular approach for nonlocal optics [11,13,34].

More recently, the birth of 2D crystals [35] introduced another possibility for studying nonlocal effects in the optical response of materials. In particular, the emergence of graphene plasmonics [3] constitutes a new playground for studying these effects down to the one atom thick limit [9,36]. The combination of graphene with two-dimensional (2D) insulators—such as hexagonal Boron Nitride (hBN)—has allowed an unprecedented control of the distance that a 2D material can be placed in the vicinity of metallic films or nanostructures. In fact, the placement of a graphene sheet at a distance of a few nanometers away from a metal surface has recently been experimentally demonstrated [9], in a similar setting as depicted in Fig. 1. This geometry provides a suitable way to obtain deep subwavelength confinements down to the atomic limit. This is possible due to the interplay of graphene plasmons and the screening exerted by the nearby metal film. This approach has revealed that the nonlocal properties of graphene's conductivity have to be included in a proper account of the experimental data [9,37]. In graphene, it is more useful to describe the nonlocal response via the material's nonlocal dynamical conductivity, which within linear-response theory relates the surface current, $\mathbf{K}_s(\mathbf{r},\omega)$, and the in-plane electric field, $\mathbf{E}_\parallel(\mathbf{r},\omega)$, via [3]

$$\mathbf{K}_s(\mathbf{r},\omega) = \int d\mathbf{r}' \overleftrightarrow{\sigma}(\mathbf{r},\mathbf{r}',\omega) \cdot \mathbf{E}_\parallel(\mathbf{r}',\omega). \tag{3}$$

Notice that this statement is merely a reformulation of Eq. (1). Here, $\overleftrightarrow{\sigma}(\mathbf{r},\mathbf{r}',\omega)$ refers to the nonlocal (surface) conductivity tensor of graphene. This quantity can be probed experimentally using graphene plasmons when the material is in the proximity of a metallic surface, and varying the graphene-metal distance, thereby retrieving the Fourier transform of $\overleftrightarrow{\sigma}(\mathbf{r},\mathbf{r}',\omega)$, that is, $\overleftrightarrow{\sigma}(\mathbf{q},\omega)$ [9]. In the absence of strain, graphene is isotropic and thus $\overleftrightarrow{\sigma}(\mathbf{q},\omega) = \sigma(q,\omega)\mathbb{1}$, where $\mathbf{q}$ labels the in-plane wave vector.

If the graphene sheet is transformed into a grating (or an external grating is deposited on the graphene sheet), the properties of the plasmons depend sensitively on the length of the imposed period [3,38,39]. Therefore adjusting the grating period can be used to tune the graphene plasmons' frequency to the desired spectral range. Either graphene itself [38] or a grating may be patterned [39]. When graphene is placed in the vicinity of a metallic surface, e.g., as illustrated in Fig. 1, nonlocal effects arising from the metal's response may influence indirectly the behavior of the fields in the region between graphene and the metal, namely through a larger penetration of the field inside the metal due to a less effective screening induced by nonlocality. The quantification of the importance of such nonlocal effects constitutes the goal of this work.

Here, we investigate the influence of the nonlocal response of a metal on the optical response of systems based on graphene lying in close proximity to a metallic film or surface. In what follows, we consider both gold and titanium metals. We shall focus on the latter, as this metal exhibits a very strong nonlocal behavior (compared, for example, with gold). The nonlocal effects arising from the metal's response are evaluated both for the plasmonic (near-field physics) and for the optical properties (far-field spectroscopy) of these structures. Throughout the manuscript, graphene's conductivity is taken as being nonlocal

[3]. The aim of this work is therefore to use graphene plasmonics to probe nonlocal effects within the metal, in contrast to the study of nonlocality in graphene performed in a recent publication [9].

Specifically, we study the impact of nonlocal effects due to the metal as a function of the graphene-metal separation, and discuss its implications on the field distribution and plasmonic losses. Finally, we consider a system in which the graphene sheet is replaced by a graphene diffraction grating made of a periodic array of graphene ribbons. In this case, we compute the system's response in the far-field and determine the influence of the metal's nonlocality on the measured spectra. We take realistic values for the dielectric functions of the materials that constitute each system, which allows us to compare our results directly to the experiments.

II. MATHEMATICAL DETAILS

Throughout this work, we consider nonmagnetic ($\mu = 1$) media. The system under study consists in a multilayered heterostructure, with either dielectric or metallic layers stacked along the y direction, as portrayed in Fig. 1. The structure is further assumed to be infinite in the xz plane. We look for TM-modes with a harmonic time dependence in the form of $e^{-i\omega t}$, and thus the corresponding magnetic field may be written as

$$\mathbf{H}(\mathbf{r},t) = H^z(x,y)e^{-i\omega t}\hat{\mathbf{z}}. \tag{4}$$

The form of $H^z(x,y)$ follows from Maxwell's equations, and, in general, is different in each of the regions defined in Fig. 1. As we consider nonlocal effects only in graphene and in the metal, it is useful to distinguish the description of the fields in the dielectric(s) and metal regions.

A. Dielectric regions

In dielectric media (source-free), the magnetic field obeys Helmholtz's equation, $\nabla^2\mathbf{H} + \varepsilon k_0^2\mathbf{H} = 0$, where $k_0 = \omega/c$. Assuming a magnetic field in the form of Eq. (4), the wave equation admits the following (transverse) solution:

$$H_T^z(x,y) = (C^+e^{ik_T y} + C^-e^{-ik_T y})e^{iqx}, \tag{5}$$

where the in-plane wave vector, q (assumed to be along the x direction without loss of generality), and the perpendicular wave vector, k_T, are related by

$$k_T = \sqrt{\varepsilon k_0^2 - q^2}. \tag{6}$$

The "T" subscript was introduced to make explicit the transverse nature of these solutions. For uniaxial media (characterized by different permittivities in the xz plane, ε^x, and in the y direction, ε^y), such as hBN, q and k_T are connected through an alternative condition, i.e., $k_T = \sqrt{\varepsilon^x\omega^2/c^2 - q^2\varepsilon^x/\varepsilon^y}$. Maxwell's equation $\mathbf{E}_T = (-i\omega\varepsilon_0\varepsilon)^{-1}\nabla\times\mathbf{H}_T$ enables us to write the corresponding electric field components as

$$E_T^x(x,y) = \frac{-k_T}{\omega\varepsilon_0\varepsilon^x}(C^+e^{ik_T y} - C^-e^{-ik_T y})e^{iqx}, \tag{7}$$

$$E_T^y(x,y) = \frac{q}{\omega\varepsilon_0\varepsilon^y}(C^+e^{ik_T y} + C^-e^{-ik_T y})e^{iqx}. \tag{8}$$

Naturally, if the dielectric is isotropic, then $\varepsilon \equiv \varepsilon^x = \varepsilon^y$.

B. Metal region

Within the metal, we assume a Drude dielectric function ε_m of the form

$$\varepsilon_m(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{\omega^2 + i\gamma_m\omega}, \tag{9}$$

where ω_p and γ_m are the plasma and damping frequencies, respectively, and ε_∞ is a background permittivity to account for interband polarization effects. We take nonlocal effects in the metal within the framework of the hydrodynamic model—see Refs. [31,33] for details. When taking nonlocality into account, Ampère's law becomes [31,33]

$$\nabla\times\mathbf{H} = -i\omega\varepsilon_0\varepsilon_m[\mathbf{E} - \xi\nabla(\nabla\cdot\mathbf{E})], \tag{10}$$

with the parameter ξ defined as

$$\xi = \beta\sqrt{\omega_p^2/\varepsilon_\infty - \omega^2 - i\gamma_m\omega}, \tag{11}$$

where β is a nonlocal parameter proportional to the Fermi velocity $v_{F,m}$ of electrons in the metal. In this work, we will take the most usual definition [33] $\beta = \sqrt{3/5}v_{F,m}$. Naturally, the local limit is recovered upon setting $\beta = 0$.

As discussed above, Maxwell's equations can be shown to admit two kinds of solutions [31,33]: divergence-free and curl-free fields. The former are the usual transverse waves that exist in the local regime. Within the LRA, the electromagnetic fields in the metal are then described similarly to the fields in the dielectric case, upon replacing $\varepsilon \to \varepsilon_m(\omega)$—see Eqs. (5)–(8). On the other hand, curl-free waves do not have an associated magnetic field, as imposed by Faraday's law. However, unlike the local case, the electric field has a nontrivial solution given by the vanishing of the term in square parentheses in Eq. (10), equivalent to the wave equation $\nabla^2\mathbf{E} - \xi^{-1}\mathbf{E} = 0$. This equation describes longitudinal solutions of the form $\mathbf{E}_L = (E_L^x\hat{\mathbf{x}} + E_L^y\hat{\mathbf{y}})e^{-i\omega t}$, with

$$E_L^x(x,y) = (D^+e^{ik_L y} + D^-e^{-ik_L y})e^{iqx}, \tag{12}$$

$$E_L^y(x,y) = \frac{k_L}{q}(D^+e^{ik_L y} - D^-e^{-ik_L y})e^{iqx}, \tag{13}$$

with (longitudinal) wave vector

$$k_L = \sqrt{-\xi^{-2} - q^2} = \beta^{-1}\sqrt{\omega^2 + i\gamma_m\omega - \omega_p^2/\varepsilon_\infty - q^2}. \tag{14}$$

Furthermore, E_L^x and E_L^y are related by the curl-free condition, $\partial E_L^y/\partial x = \partial E_L^x/\partial y = iqE_L^x$. The general solution for the fields inside the metal is therefore $H^z = H_T^z$ and $E^{x/y} = E_T^{x/y} + E_L^{x/y}$.

C. Boundary conditions

The coefficients that characterize the fields in each layer are determined by the boundary conditions (BCs). For bound modes (like surface plasmons), we have $q > \sqrt{\varepsilon}\omega/c$. This renders k imaginary, and the signal must be chosen judiciously to ensure that the fields satisfy Sommerfeld's radiation condition. For an interface between two dielectrics, the usual BCs apply, that is, the continuity of the tangential component of the electric field (E^x) and the (dis)continuity of the magnetic field (H^z) in the (presence) absence of a surface current density at the

interface. The presence of a finite surface current density is needed in order to describe a graphene sheet (or any other 2D material) placed at an interface between the two media, and enters through Ohm's law $\mathbf{K}_s(q,\omega) = \sigma(q,\omega)E_x(q,\omega)\hat{\mathbf{x}}$; see also Eq. (3). Note that here $\sigma(q,\omega)$ entails both frequency and momentum dependencies in order to account for nonlocal effects in graphene. We employ graphene's nonlocal conductivity using Mermin's particle-conserving prescription, which is detailed in Appendix C.

Finally, for a dielectric/metal interface (without graphene[1]), although the BCs described above remain valid, these need to be augmented by an additional BC (due to the existence of a longitudinal mode within the metal). Typically, this additional BC dictates that the normal component of the polarization vector vanishes at metal's surface. The polarization vector is given by [31]

$$\mathbf{P} = (\mathrm{i}/\omega)\boldsymbol{\nabla} \times \mathbf{H} - \varepsilon_0\varepsilon_\infty \mathbf{E}. \tag{15}$$

Note that this polarization field refers to the one associated with the free electrons in the surface of the metal, responsible for the transport of electric currents. Therefore, from a physical perspective, this condition merely imposes that cannot exist currents flowing from the metal to the neighboring dielectric regions (e.g., the interface is a hard wall). In possession of this additional BC, the amplitudes of the fields may now be determined; see Appendices A and B for a more detailed account.

III. NONLOCAL EFFECTS IN THE PLASMONIC PROPERTIES

We consider a system composed by a thin metallic layer (with thickness h) with a graphene sheet lying at a distance s from its surface. In the spacer region, i.e., between graphene and the metal, we assume to have a slab of hBN; however, our formalism is general and thus allows for the consideration of different dielectric media. Below the metal and above the graphene, we assume to have air. For the sake of clarity, the system has been divided into four regions, I–IV, as depicted in Fig. 1. The calculation of the nonlocal plasmonic properties of the system follows the guidelines discussed in the previous section, and can be consulted with a greater degree of detail in the Appendix A. In order to assess the influence of the metal's nonlocal effects in the plasmonic properties of the system, we compare our nonlocal results with the corresponding predictions of the local-response theory.

A. Nonlocal effects in the plasmon dispersion

For the study of the metal's nonlocal effects in the dispersion relation of graphene plasmons, we consider here two different metals—gold (Au) and titanium (Ti)—described by the parameters presented in Table I. The dielectric function of hBN, on the other hand, is adopted from Refs. [36] (out-of-plane direction) and [40] (in-plane direction).

TABLE I. Drude model parameters used for titanium[a] (Ti) and gold (Au).

	Ti	Au
ω_p (eV)	2.80 [41]	8.84 [42]
γ_m (meV)	82.0 [41]	103.0 [42]
ε_∞	2.2 [43]	9.84 [44]
$v_{F,m}/c$	0.00597 [45]	0.00464 [46]

[a]Note that the authors of Ref. [41] described titanium using both Drude and Lorentz terms, but, for the purpose of this work, we only considered the Drude contribution, which nonetheless provides a very good approximation.

The influence of the nonlocal effects arising from the metal's response is investigated by computing the dispersion relation, $\omega(q)$, of graphene plasmons and comparing the ensuing spectra obtained assuming a local and a nonlocal response. The outcome is presented in Fig. 2 (white curves) for the allowed plasmonic modes in the structure pictured in Fig. 1. Since there is dissipation in the system (both in the metal, hBN, and graphene), either q or ω needs to be regarded as a complex quantity in order to fulfill the boundary conditions. In what follows, we have chosen q to be a real number and hence ω is complex. For the time being, we focus on the real part of the frequency, $\mathrm{Re}\{\omega\}$, and denoting it by ω for simplicity. The corresponding imaginary part, $\mathrm{Im}\{\omega\}$, associated with the plasmonic losses, will be discussed in Sec. III C.

It is apparent from Fig. 2 that the dispersion curves are not continuous, but rather present two asymptotes around 750 and 1350 cm^{-1}; this feature is shared in both local and nonlocal frameworks. These frequencies correspond to the beginning of the hBN's Reststrahlen bands, where this material is hyperbolic [3], and are associated with the excitation of surface phonon-polaritons (which in this case hybridize with graphene plasmons) [3].

Moreover, Fig. 2 clearly demonstrates that nonlocal effects (associated with the metal) impact the plasmon dispersion very differently depending on the metal: for Au, the nonlocal and local curves lie very close, whereas in the case of Ti, there is a significant blueshift of the dispersion of graphene plasmons due to the nonlocal effects of the Ti (we stress that graphene is treated as being nonlocal in both cases). Quantitatively, the respective blueshifts are around 2% and 20%. We have determined that this difference in the magnitude of the nonlocal effects originates from the cumulative influence of two main factors: the Ti's larger nonlocal parameter with respect to Au's ($\beta_{Ti}/\beta_{Au} \simeq 1.29$), and Ti's smaller plasma and relaxation frequencies (specially the former, $\omega_{p,Ti}/\omega_{p,Au} \simeq 0.32$). Both quantities—which are intrinsic to each metal—are crucial for the enhancement of the nonlocal effects in Ti. For this reason, we shall focus on titanium henceforth in order to illustrate a case in which nonlocal effects are pivotal.

It should be emphasized that, apart from the specific characteristics of the metal, the dielectric environment in its vicinity also plays a significant role on the impact of nonlocal effects. In the particular case considered here, the proximity between the metal and graphene (that is, the thickness s of the spacer)

[1]We do not consider in this work the case of a dielectric/metal interface separated by a graphene sheet. The main reason is that, under those conditions, doping the graphene in order to adjust its Fermi level would be impractical, since the metal would drain the free electrons in the graphene sheet.

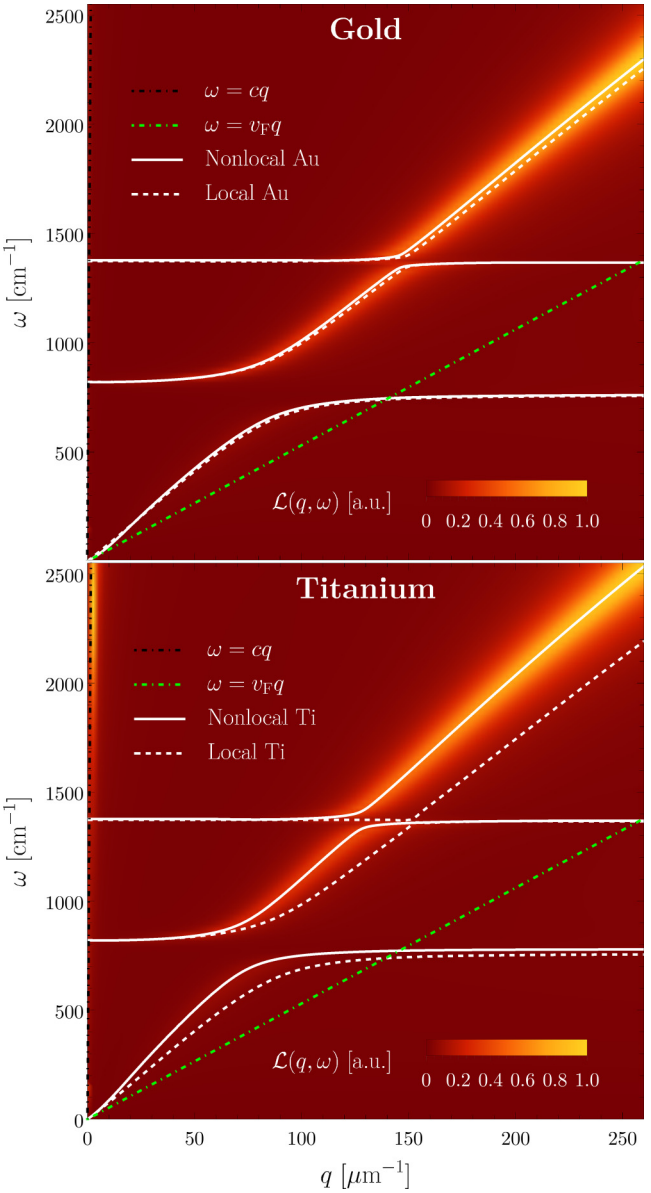

FIG. 2. Dispersion relation (DR) of SPPs in an air/metal/hBN/graphene/air configuration, both for a local (dashed white) and nonlocal (solid white) metal, in the case of gold (top) and titanium (bottom). The black dot-dashed line corresponds to the light dispersion in vacuum, whereas the green dot-dashed line corresponds to the graphene's electron-hole continuum boundary, $\omega = v_F q$, with $v_F \approx c/300$. The color plot shows the respective loss function (defined as minus the imaginary part of the reflection coefficient), obtained nonlocally (see Appendix B). The remaining parameters are $s = 1$ nm, $h = 10$ nm, $E_F = 0.5$ eV, and $\Gamma = 16$ meV.

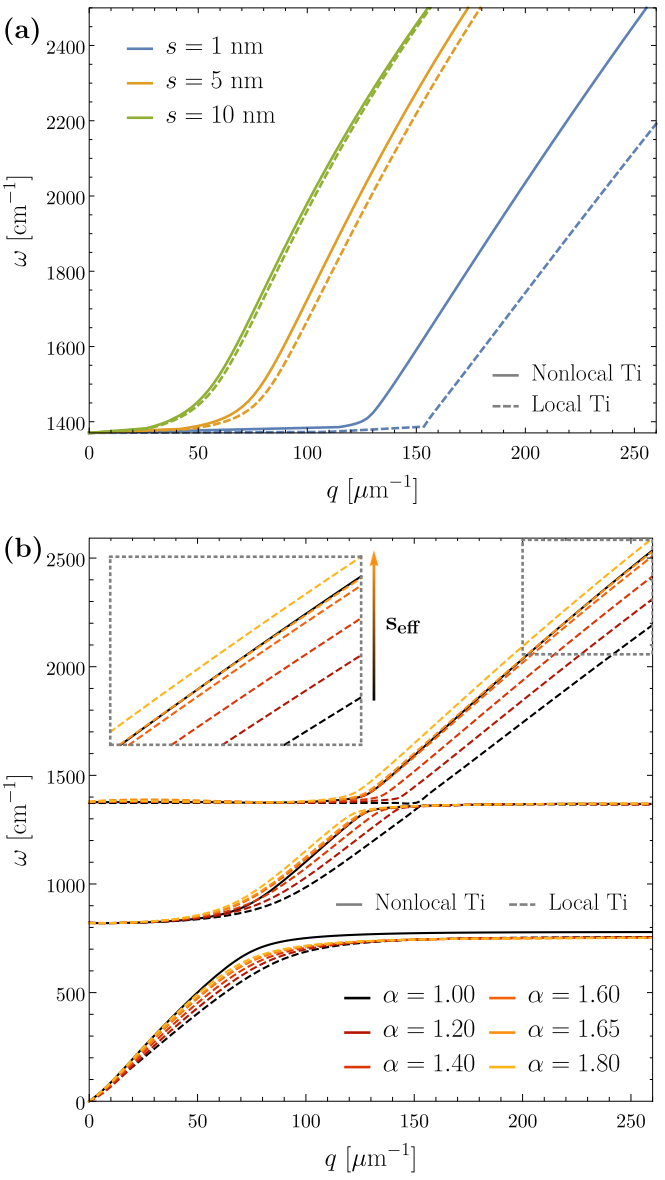

FIG. 3. (a) Uppermost branch of the dispersion relation ($\omega > 1350$ cm^{-1}) of the plasmonic modes in an air/Ti/hBN/graphene/air configuration, for different values of the hBN's slab thickness, s, calculated for a local (dashed) and nonlocal (solid) metal. (b) Dispersion relation in the same configuration, calculated for local and nonlocal metals for different values of the effective parameter $\alpha = s_{eff}/s$, with $s = 1$ nm. The remaining parameters, for both panels, are $h = 10$ nm, $E_F = 0.5$ eV, and $\Gamma = 16$ meV.

is determinant, as Fig. 3(a) plainly shows. The figure shows that the difference between the local and nonlocal calculations increases as the thickness of the spacer region is decreased. In particular, while for $s = 10$ nm the difference is negligible, nonlocal effects become clearly perceptible for $s = 5$ nm, and for $s = 1$ nm, nonlocal effects become of paramount importance for an accurate description of the system's plasmonic response. This behavior is due to a greater spatial confinement of the fields inside narrower spacers, which is enhanced owing to the screening of graphene plasmons by the metal, and in turn gives rise to an acousticlike graphene plasmons with very high momenta [3,9] and thus more susceptible to nonlocality.

Figure 3(a) also shows that an increment of the spacer thickness induces a blueshift in the dispersion relation of the plasmonic modes. For that reason, a naïve approach to account for nonlocal effects while carrying out a local calculation is to consider an *effective* spacer thickness, s_{eff}, larger than the actual value, s, in a similar fashion to what has been proposed in earlier works [44,47,48]. Although this method can indeed be regarded as a somewhat naïve version of quantum-corrected boundary conditions [49,50], it can mimic the proper nonlocal

calculation as illustrated in Fig. 3(b). In particular, the nonlocal dispersion can be reproduced using a local formalism with an effective parameter $\alpha = s_{\rm eff}/s$ between 1.75 and 1.80. However, it should be stressed that the value of the effective parameter highly depends both on the considered materials (particularly the metal's properties) and on the configuration itself. For this reason, this procedure is generally hard to implement because the particular value of $\alpha = s_{\rm eff}/s$ that reproduces the nonlocal effects is difficult to predict theoretically, which renders this approach unsuitable. Instead, it can be merely regarded as a fitting parameter when describing an experiment through local calculations.

Before concluding the present section, we investigate how nonlocal effects vary with the thickness of the metal film. To that end, in Fig. 4(a), we have plotted the plasmon dispersion relation for an air/Ti/hBN/graphene/air structure using three different values for the titanium thickness: 1, 10, and 100 nm. The plasmonic spectrum was obtained both nonlocally (left panel) and locally (right panel). In both cases, the graphene is treated as a nonlocal medium. The figure demonstrates that the 10- and 100-nm cases produce the same dispersion, whereas the 1-nm curve lies toward smaller frequencies (for the same q). These results suggest that, above a certain threshold thickness, the plasmonic properties of a system with a finite-thickness metal are equivalent to those of a configuration with a semi-infinite metal. This behavior becomes particularly evident upon inspection of Fig. 4(b), where the plasmon wave vector, for a given frequency, is shown as a function of the metal's thickness. Clearly, for $h \gtrsim 3$ nm, the metallic film is well approximated by a semi-infinite metal (this threshold of $\sim$3 nm is consistent with the penetration depth of the fields in the nonlocal regime, as will be discussed in Sec. III B).

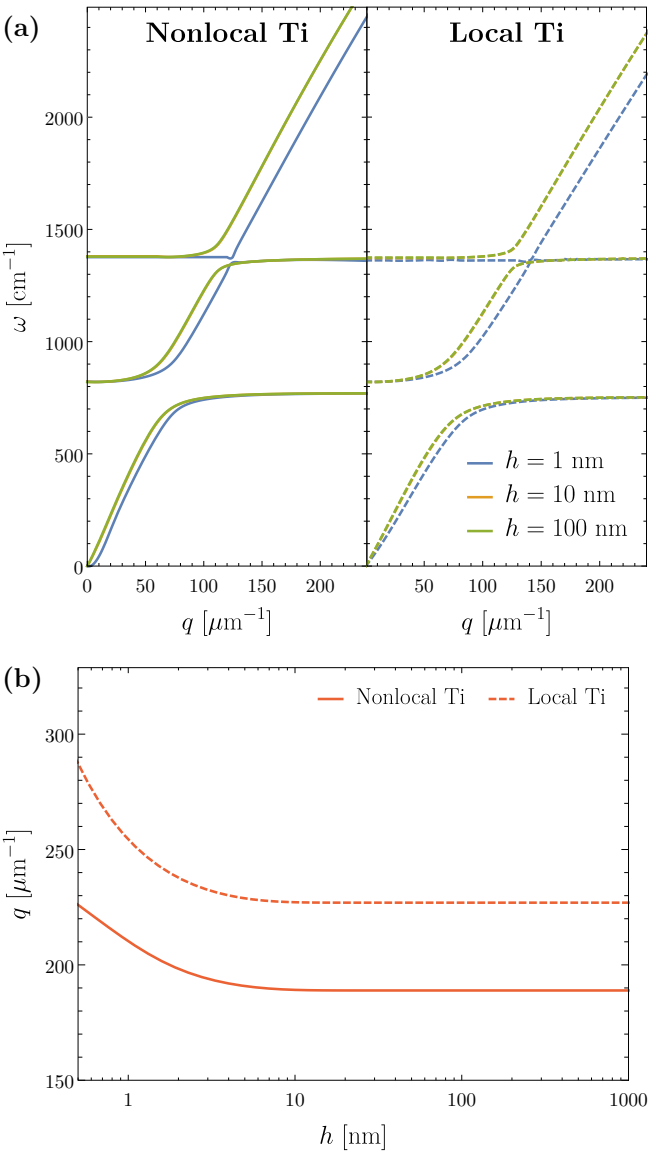

FIG. 4. Effect of the variation of the dispersion relation of the plasmonic modes in a air/Ti/hBN/graphene/air structure with the metal thickness, h. (a) Nonlocal (left, solid) and local (right, dashed) dispersions for three distinct values of the metal thickness (the orange curve is behind the green one). (b) Plasmon wave vector, for a fixed frequency $\omega = 2000$ cm^{-1} (dot-dashed on the top panel) as a function of the metal thickness. In both panels, the remaining parameters are $s = 1$ nm, $E_{\rm F} = 0.5$ eV, and $\Gamma = 16$ meV.

Comparison to the semi-infinite metal case

We have seen that for metal thicknesses $h \gtrsim 3$ nm, the thin film can be well approximated by a semi-infinite metal. Such a scenario is in fact the most relevant under realistic experimental conditions, which typically employ metals (acting as a gate) with thicknesses in excess of 10 nm [9,51]. This is convenient because it not only simplifies the analysis, but it also allows us to write a closed-form expression for the plasmon dispersion in the heterostructure—now a dielectric/graphene/dielectric/metal configuration—reading

$$\left(\frac{\varepsilon_4^x}{\kappa_{\rm T}^{(4)}}+\frac{\varepsilon_3^x}{\kappa_{\rm T}^{(3)}}+\frac{i\sigma}{\omega\varepsilon_0}\right)\left(1+\frac{\varepsilon_{\rm m}\kappa_{\rm T}^{(3)}}{\varepsilon_3^x\kappa_{\rm T}^{\rm m}}+\delta_{\rm nl}\right)$$
$$=\left(\frac{\varepsilon_4^x}{\kappa_{\rm T}^{(4)}}-\frac{\varepsilon_3^x}{\kappa_{\rm T}^{(3)}}+\frac{i\sigma}{\omega\varepsilon_0}\right)\left(-1+\frac{\varepsilon_{\rm m}\kappa_{\rm T}^{(3)}}{\varepsilon_3^x\kappa_{\rm T}^{\rm m}}-\delta_{\rm nl}\right)e^{-2\kappa_{\rm T}^{(3)}s}\,, \tag{16}$$

where $\kappa_{\rm T}^{(\nu)}=\sqrt{q^2\varepsilon_\nu^x/\varepsilon_\nu^y-\varepsilon_\nu^x k_0^2}$ for $\nu=\{3,4\}$, $\kappa_{\rm T}^{\rm m}=\sqrt{q^2-\varepsilon_{\rm m}k_0^2}$, and $\delta_{\rm nl}$ is a nonlocal correction term given by [33]

$$\delta_{\rm nl}=\frac{q^2}{\kappa_{\rm L}^{\rm m}\kappa_{\rm T}^{\rm m}}\frac{\varepsilon_{\rm m}-\varepsilon_\infty}{\varepsilon_\infty}\,, \tag{17}$$

with $\kappa_{\rm L}^{\rm m}=\sqrt{q^2-(\omega^2+i\gamma_{\rm m}\omega-\omega_{\rm p}^2/\varepsilon_\infty)/\beta^2}$. Indeed, our calculations demonstrate that Eq. (16) is able to reproduce extremely well the plasmon dispersion presented, for instance, in Figs. 2 and 3. Furthermore, it is instructive to note that by taking the $\varepsilon_{\rm m}\to\infty$ limit in Eq. (16), which corresponds to the case where the metal becomes a perfect conductor, one obtains (neglecting nonlocal effects)

$$\frac{\varepsilon_4^x}{\kappa_{\rm T}^{(4)}}\coth\left[\kappa_{\rm T}^{(3)}s\right]+\frac{\varepsilon_3^x}{\kappa_{\rm T}^{(3)}}+\frac{i\sigma}{\omega\varepsilon_0}=0\,. \tag{18}$$

Equation (18) coincides with the dispersion relation of the acoustic plasmon branch in double-layer graphene [3] in a symmetric dielectric environment, where the individual graphene layers are separated by a distance $2s$. This result

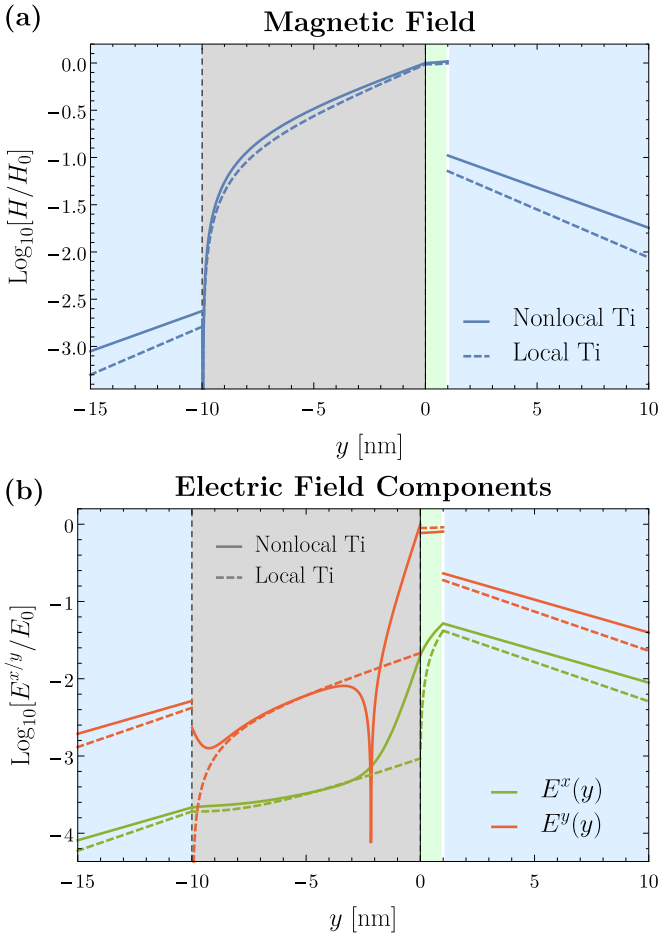

FIG. 5. Spacial distribution of (a) the magnetic field and (b) the electric field components, for an air (blue)/Ti (gray)/hBN (green)/graphene (white)/air (blue) structure, calculated using local and nonlocal metal models, for plasmonic modes with the same real part of the frequency $\omega = 2000\ \mathrm{cm}^{-1}$. H_0 and E_0 are respectively the nonlocal magnetic and total electric fields calculated at $x = 0$. The parameters used are $h = 10$ nm, $s = 1$ nm, $E_F = 0.5$ eV, and $\Gamma = 16$ meV.

reflects the scenario in which the screening exerted by the perfect conductor mirrors exactly the charges induced in the graphene sheet. Lastly, notice that for large separations, i.e., $s \to \infty$, the plasmon dispersion (16) reduces to that of an isolated graphene sheet between two dielectric media ε_4^x and ε_3^x.

B. Nonlocal effects in the field distributions

Having discussed the influence of nonlocal effects in the plasmon dispersion, we now study their impact in the spatial distribution of the fields associated with the plasmonic modes. The field amplitudes follow from the BCs, and we determine the y dependence of the fields at a given (q,ω) point, which satisfies the plasmon dispersion shown in the previous section. In this spirit, Fig. 5 depicts the variation of the absolute value of the magnetic and electric field (in logarithmic units), along the heterostructure, for a (real part of the) frequency $\omega = 2000\ \mathrm{cm}^{-1}$. It is apparent from Fig. 5(a) that the magnetic field distribution remains essentially unchanged under nonlocal corrections. This is a natural consequence of the hydrodynamic model, in which nonlocality only enters in the longitudinal components. Since the magnetic field is purely transverse, nonlocal effects do not influence it, and the only differences between the local and nonlocal cases arise from small differences between the coefficients (and the value of q for the same frequency) that describe the magnetic field on each case.

In contrast to the magnetic field, the inclusion of nonlocal effects in the metal's response renders significant changes in the spatial distribution of the electric field, as can be seen from Figs. 5(b)–5(c). Naturally, the differences between the local and nonlocal calculations arise mostly in fields within the metal region, as could be anticipated, due to the fact that we have allowed the existence of a longitudinal mode inside the metal, in the nonlocal case. The figure demonstrates that the nonlocality introduced by this additional longitudinal wave has profound implications in the eletric field's spatial distribution. Specifically, notice that the electric field inside the metal in the local case is significantly smaller than the field in the spacer region (as expected for a good metal). However, when nonlocal effects are taken into account, the electric field is increased (when compared with the LRA) by several orders of magnitude in the vicinity of the metal's surface. This feature is a consequence of the smearing of the electron density introduced by nonlocality, which translates into a larger penetration of the fields inside the metal. This penetration depth is practically negligible in the local approximation, but becomes important when taking nonlocality into account, as Fig. 5 shows. In the latter, the penetration length reaches a couple nanometers, and can become comparable to the metal's thickness for ultrathin films.

Another feature visible in Figs. 5(b)–5(c) is the presence, in the nonlocal case, of a sharp dip in the electric field magnitude around $y = -2.5$ nm, increasing again towards the metal/air interface. This is due to a node of its y component, which is caused by a destructive interference between the transversal and longitudinal modes inside the metal. Note that both the transversal and longitudinal modes have x and y components, with the total field being given by $E^{x/y} = E_{\mathrm{T}}^{x/y} + E_{\mathrm{L}}^{x/y}$; the node occurs when $E_{\mathrm{T}}^{y} + E_{\mathrm{L}}^{y} = 0$. For that reason, across the node there is a change of the sign of the y component of the electric field, dividing the regions where the longitudinal mode amplitude is higher (to the right of the node) or lower (to the left) than the transversal mode amplitude.

C. Nonlocal effects on the plasmonic losses

We conclude the study of the system portrayed in Fig. 1 with an analysis of the effect of the nonlocality in the metallic film in the losses of the plasmonic modes. Two different choices for studying the losses are possible: the complex frequency and the complex wave-number approaches. The latter is more convenient, for instance, when we are talking about propagating plasmons. On the other hand, for localized plasmons—e.g., like those in graphene ribbons—it may be more convenient to use the complex frequency approach. When the damping is small, both approaches are equivalent. As mentioned above, in the presence of losses and for a real-valued q, the ensuing condition

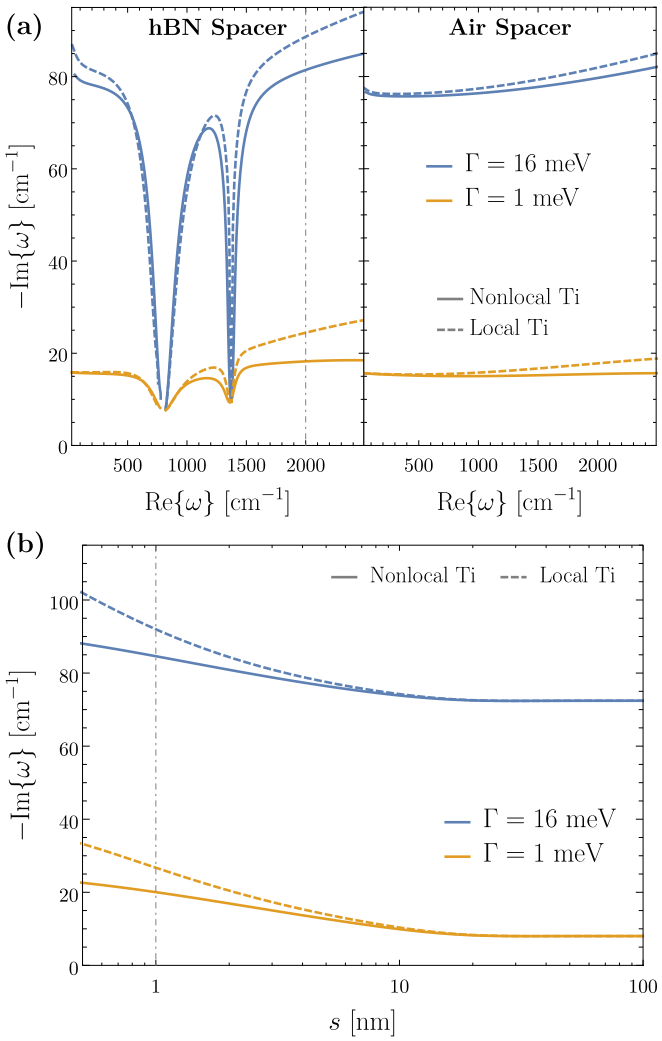

FIG. 6. (a) Imaginary part of the frequency of the SPPs in an air/Ti/spacer/graphene/air configuration, with the spacer being hBN (left) or air (right), calculated using local (dashed) and nonlocal (continuous) models, for two different values of the graphene's electric relaxation energy Γ, and for $s = 1$ nm (dot-dashed on the bottom panel). (b) Variation of the imaginary part of the SPPs, in a configuration with a hBN spacer, with the spacer width s, for a fixed frequency $\omega = 2000\ \mathrm{cm}^{-1}$ (dot-dashed on the top panel). The remaining parameters are $h = 10$ nm and $E_F = 0.5$ eV.

for the plasmon dispersion requires a complex-valued ω. So far, we have limited our discussing to its real part, but here we focus on its imaginary part, $\mathrm{Im}\{\omega\}$, since this quantity is intrinsically related with the plasmonic losses and plasmon life-time τ_p, in particular, $\tau_p^{-1} = -\mathrm{Im}\{\omega\}/2$. In Fig. 6, we have plotted $-\mathrm{Im}\{\omega\}$ as a function of the real part of the polariton frequency, $\mathrm{Re}\{\omega\}$, both within the local and nonlocal response formalism (for the metal; graphene is modeled as nonlocal in all cases).

We have considered two cases, one with a relatively high (for graphene) scattering rate of $\Gamma = 16$ meV, and another illustrating high-quality, low-loss graphene, in which $\Gamma = 1$ meV. We note that the change of graphene's electronic scattering rate does not significantly alter the dispersion relation presented in Fig. 2 (in fact, there is no visible difference to the eye). However, it significantly impacts the losses affecting graphene plasmons sustained in the heterostructure. As Fig. 6 shows, the differences between the calculation using a local and a nonlocal metal are modest. This suggests that the losses that the plasmons incur originate from the graphene, which is natural since the metal only participates indirectly—via screening—and the mode's spectral weight is essentially attributed to graphene's. Furthermore, Fig. 6(b) shows that this difference becomes more evident when the spacer width is reduced, since the losses, when calculated locally, strongly increase for small spacers, whereas nonlocally their increase is very small. This means that, by reducing the spacer size, an increasingly higher confinement can be achieve without a strong increase in the losses, which is a very important result.

Let us now understand the sudden decrease of the losses seen, in Fig. 6(a) around the hBN phonon frequencies, when large damping in graphene (16 meV) is considered. To that end, we recall that the damping of the phonon modes in hBN are 2.4 meV for the in-plane phonon [40] and 1.9 meV for the out-of-plane phonon [36]. At the frequency of the phonons, the polariton spectrum has, essentially, a phononic nature. Therefore the losses are essentially controlled by those due to phonons. Since these are much smaller than 16 meV, we see a sudden decrease in the losses. On the other hand, for a damping in graphene of 1 meV, the phonon damping is comparable to the damping in graphene. As a consequence, the curves of the losses in Fig. 6(a) do not present the sudden drop at the phonon frequencies. In conclusion, away from the phonon frequencies, the losses are essentially due to graphene, whereas near the phonon frequencies of hBN the losses are essentially controlled by the behavior of the imaginary part of the dielectric function of the hBN spacer.

IV. PROBING NONLOCALITY IN METALS USING A GRAPHENE NANORIBBON GRATING

A common approach to experimentally access the plasmonic properties of a system consists on performing reflectance and/or transmittance measurements of the far-field spectra, upon illuminating the sample. In this context, plasmon excitations appear as resonances in the spectra (as peaks or dips). Therefore one may investigate the influence of nonlocal effects by studying the resulting spectra and compare it to experimental data.

In the extended, continuous layered system studied in the previous section, we have seen that the metal's nonlocal response affects the plasmon properties of the modes supported in the system. However, the wave vector mismatch between the graphene plasmons in Fig. 1 structure and the one of a photon in free-space differs by more than two orders of magnitude, and thus it has been primarily investigated by near-field techniques, which are able to overcome this kinematic limitation [9,51].

In what follows we consider a different configuration. It is similar to the one considered in the previous section, but we now assume that the graphene monolayer has been patterned into a periodic array of graphene nanoribbons—see Fig. 7. These effectively act as a diffraction grating, whose Fourier components provide the necessary in-plane momentum to excite plasmons in the system [3,38]. Indeed, this mechanism not only allows the excitation of graphene plasmons by free-space photons but it also serves as a platform susceptible to

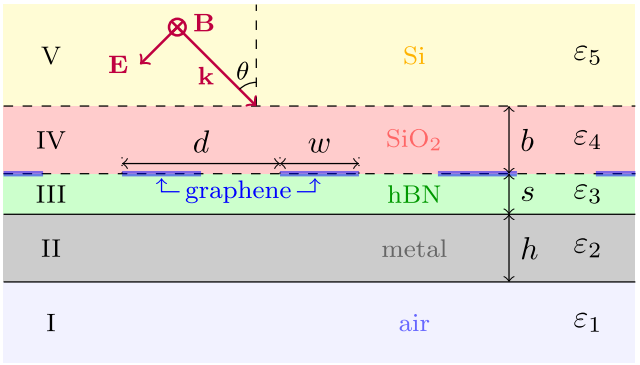

FIG. 7. Schematic representation of the layered system considered in this section. It is assumed to be periodic in the x direction and uniform in the z direction. All the different materials, dimensions and regions considered are marked in the figure. Graphene is considered as being nonlocal throughout and the metal is assumed to be either local or nonlocal.

enhance nonlocal effects, since the higher diffraction order can carry substantial momentum and thus promote nonlocal effects. Hence, in order to illustrate a case where nonlocality plays a major role on the optical properties of the system, we will replace the graphene sheet considered in the previous section by a graphene diffraction grating with a period d, and ribbon width w (see Fig. 7). We calculate the optical properties (namely, the reflectance $\mathcal{R}$, transmittance $\mathcal{T}$ and absorbance $\mathcal{A}$) of such structure when it is illuminated by a p-polarized plane wave coming from region V, with monochromatic frequency ω and incident angle θ, as depicted in Fig. 7. Apart from the introduction of the diffraction grating, we will also be considering henceforth a system where the ribbon array is encapsulated between the spacer (hBN, as before) and a thick (285 nm) layer of silicon dioxide (SiO_2). The latter's permittivity is taken from Ref. [52]. On top of the SiO_2, we further consider a semi-infinite layer of silicon (Si) described by an isotropic dielectric constant of $\varepsilon_{Si} = 11.66$ [53]. Our theoretical calculations for the grating system follow from the considerations outlined in the previous sections and they are based on a Fourier modal expansion of the electromagnetic fields described elsewhere [3,38]. To be self-contained, we provide the mathematical details in Appendix B.

Figure 8(a) shows the outcome of our computations for the reflectance, $\mathcal{R}$, transmittance, $\mathcal{T}$, and absorbance, $\mathcal{A}$, spectrum for a representative graphene ribbon diffraction grating placed near a metal, with period $d = 25$ nm and ribbon width of $d/2$. Normal incidence ($\theta = 0$) is assumed hereafter, but our formalism is valid for arbitrary incident angles (though the dependence on the impinging angle is weak since $\sin\theta \ll 2\pi/d$). Furthermore, as before, the impact of the nonlocal effects arising from the nearby metal's response are evaluated by comparing the local with the nonlocal case (while maintaining graphene treated at the nonlocal level). The spectral features visible in Fig. 8(a) are rich, but these can be split into two categories, detailed below. The first three peaks/dips—located approximately at 750, 1100, and 1350 cm^{-1}—do not present any variation due to the nonlocal effects. These peaks correspond to the excitation of optical

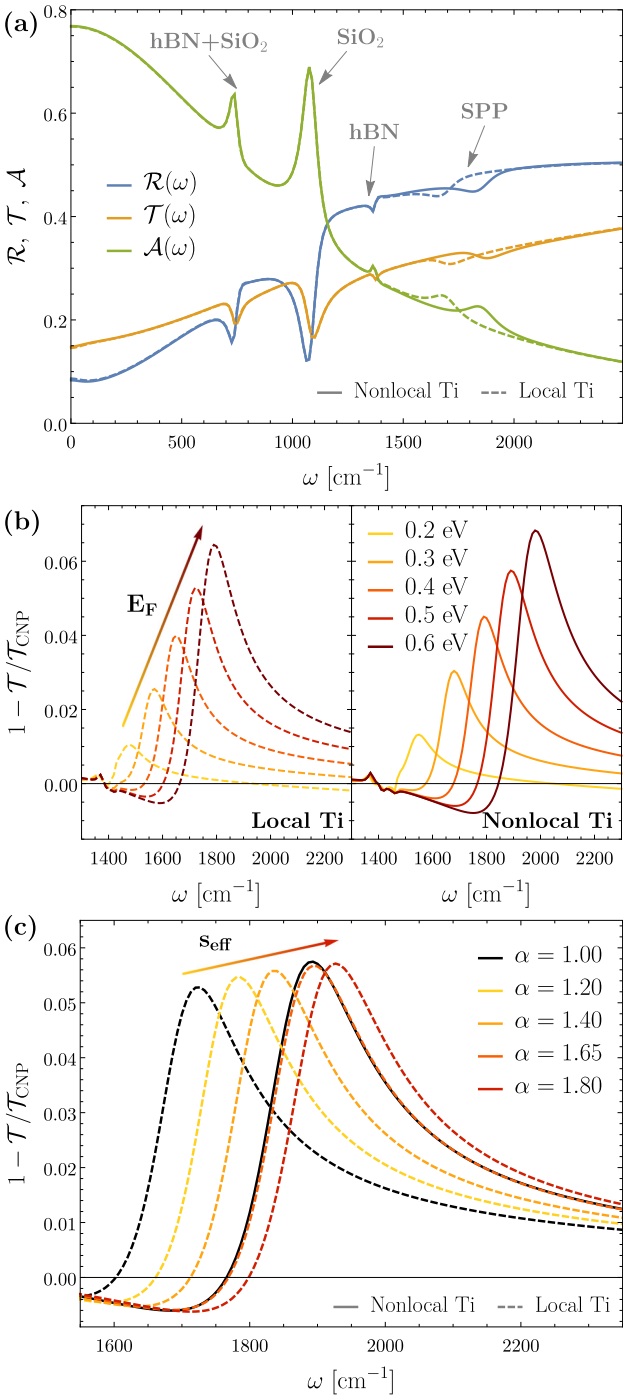

FIG. 8. (a) Reflectance ($\mathcal{R}$), transmittance ($\mathcal{T}$), and absorbance ($\mathcal{A}$) spectra of an air/Ti/hBN/graphene/SiO_2/Si system. Graphene ($E_F = 0.5$ eV, $\Gamma = 8$ meV) is treated as being nonlocal throughout, whereas the metal is considered both within the local and nonlocal frameworks. The character of each peak (phononic for the first three and plasmonic for the fourth) is indicated by the arrows. (b) Extinction spectra (see details in the main text) near the graphene plasmon resonance frequency for different Fermi levels of graphene. The local and nonlocal cases were plotted separately for clarity. Notice the blueshift of the nonlocal curves relative to their local counterparts. (c) Variation of the extinction curves (with a local description of the metal) of the same system with the effective parameter $\alpha = s_{eff}/s$, compared to the nonlocal counterpart. In all cases, the remaining parameters are $h = 10$ nm, $s = 1$ nm, $b = 285$ nm, $d = 25$ nm, $w = d/2$, and $\theta = 0°$.

phonons in either the hBN or the SiO_2 slabs [indicated in Fig. 8(a)]. Note that both hBN and SiO_2 support phononic modes around 750 cm^{-1}. On the other hand, the position of the fourth resonance shifts considerably to higher frequencies upon inclusion of nonlocal effects in the metal. Moreover, in Fig. 8(b), we have plotted the extinction spectra, defined as $1 - \mathcal{T}/\mathcal{T}_{CNP}$ (where $\mathcal{T}$ is the transmittance at some finite Fermi level and $\mathcal{T}_{CNP}$ is the transmittance at the graphene's charge neutrality point), for several values of the Fermi energy. We focus on the spectral window where the graphene plasmon lies. Clearly, the peak disperses towards higher frequencies with increasing Fermi level. This demonstrates that this resonance corresponds indeed to the excitation of graphene plasmons in the array. The advantage of exciting these plasmons using a grating is that their frequency is highly dependent on the value chosen for the period and ribbon width of the graphene array, with smaller values of the period yielding concomitant larger wave vectors and thus plasmon resonances at higher frequencies. For a Fermi level of 0.5 eV, the difference in the resonance position, between the local and nonlocal metal cases, is roughly 200 cm^{-1}, which in turn corresponds to a relative variation of more than 10%. This plainly shows that, in the case of a graphene sheet placed very close a metallic film, the proper account of nonlocal effects both in the metal and in graphene are key to properly model and interpret experimental results in such a system. A solution for this problem could be to consider the spacer width s to be an effective fitting parameter, s_{eff}, higher than the actual experimental value, as shown in Fig. 8(c).

Figure 8(c) shows the nonlocal curve for $s = 1$ nm overlaid by local curves for different values of the effective spacer thickness; for $s_{eff} = 1.75$ nm, the correspondence between the local and nonlocal curves is nearly perfect, proving the high efficacy of this method.

V. CONCLUSIONS

Nonlocality in metals is strongly influenced by several material-dependent parameters, namely the plasma frequency: the Fermi velocity, the momentum relaxation rate, and the background dielectric function of the core electrons in the metal. In bulk metals, nonlocal effects can be safely neglected since in that case the wave number of the radiation interacting with the metal is, in general, small ($q/k_F \ll 1$). However, when the wave number is large ($q/k_F \sim 1$), as it happens with graphene acoustic plasmons, and the metallic structures are small, compared to the wavelength of the radiation in free space, nonlocality becomes non-negligible and can change substantially, both qualitatively and quantitatively, the results predicted by a local calculation. Therefore a fully nonlocal calculation is essential when the above conditions are met, if one wants to have a quantitative description of the experimental data.

Nonlocal effects are particularly strong in titanium, due to its small plasma frequency, small scattering rate, and large Fermi velocity. Often titanium is used as an adhesion layer between gold and other materials, such as hBN; for that reason, nonlocality imparts a significant signature in composite systems when graphene, hBN, and titanium are all brought to close proximity to each other. In this paper, we have shown this to be the case, with the extinction spectrum of such a system showing a clear signature of titanium nonlocality. The extinction spectrum of the structure presents well defined resonances due to the excitation of surface plasmons in graphene whose spectral position depends strongly on the degree of nonlocality in the metal. Therefore extinction experiments in a metal/hBN/graphene system provide a viable route to examine nonlocal effects in metals, as we have shown here. In addition, nonlocal effects also promote larger propagation lengths of the graphene acoustic plasmons (created by the proximity of the metal) than those predicted by the local theory. We have shown, however, that the local calculation can be rescued if the spacer distance between graphene and the metal is taken as a fitting parameter. Indeed, we have shown that an increase of the thickness of the spacer is needed to account for the extinction curves of the structure if one insists in making a local calculation for describing the experimental data. Our work made a detailed analysis of all these aspects and can be used to interpret the optical properties of a graphene and metallic nanostructure in close proximity to each other.

ACKNOWLEDGMENTS

The authors thank Sébastien Nanot and Itai Epstein for valuable discussions and comments. E.J.C.D., Yu.V.B., and N.M.R.P. acknowledge support from the European Commission through the project Graphene-Driven Revolutions in ICT and Beyond (Ref. No. 785219), and from the Portuguese Foundation for Science and Technology (FCT) in the framework of the Strategic Financing UID/FIS/04650/2013. E.J.C.D. acknowledges FCT for the Grant No. CFUM-BI-14/2016. D.A.I. acknowledges the FPI Grant No. BES-2014-068504. F.H.L.K. acknowledges financial support from the Government of Catalonia trough the SGR Grant No. (2014-SGR-1535), and from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2015-0522), support by Fundacio Cellex Barcelona, CERCA Programme/Generalitat de Catalunya and the Mineco Grants Ramn y Cajal (RYC-2012-12281) and Plan Nacional (FIS2013-47161-P and FIS2014-59639-JIN). Furthermore, the research leading to these results has received funding from the European Union Seventh Framework Programme under Grant agreement No.696656 Graphene Flagship, the ERC starting Grant (307806, CarbonLight), and project GRASP (FP7-ICT-2013-613024-GRASP). N.A.M. is a VILLUM Investigator supported by VILLUM FONDEN (Grant No. 16498). Center for Nano Optics is financially supported by the University of Southern Denmark (SDU 2020 funding). Center for Nanostructured Graphene is supported by the Danish National Research Foundation (Project No. DNRF103).

APPENDIX A: CALCULATION OF THE PLASMONIC PROPERTIES

In this Appendix, we describe the procedure we employed to calculate the dispersion relation and all subsequent analysis of the system presented in Fig. 1. We will assume, with full generality, that all the dielectric media are described by two

dielectric functions along the in-plane (ε^x) and out-of-plane (ε^y) directions.

Following the guidelines discussed in Sec. II, we write the magnetic field in either of the 4 regions that compose the system as

$$H_{\rm I}^z(x,y) = \eta e^{\kappa_{\rm T}^{(1)} y} e^{iqx}, \tag{A1}$$

$$H_{\rm II}^z(x,y) = \left(\gamma^+ e^{\kappa_{\rm T}^{(2)} y} + \gamma^- e^{-\kappa_{\rm T}^{(2)} y}\right) e^{iqx}, \tag{A2}$$

$$H_{\rm III}^z(x,y) = \left(\alpha^+ e^{\kappa_{\rm T}^{(2)} y} + \alpha^- e^{-\kappa_{\rm T}^{(2)} y}\right) e^{iqx}, \tag{A3}$$

$$H_{\rm IV}^z(x,y) = \zeta e^{-\kappa_{\rm T}^{(4)} y} e^{iqx}, \tag{A4}$$

where $\eta, \gamma^\pm, \alpha^\pm$, and ζ are undetermined coefficients, q is the in-plane momentum of the modes, and $\kappa_{\rm T}^{(\nu)} = \sqrt{q^2 \varepsilon_\nu^x/\varepsilon_\nu^y - \varepsilon_\nu^x \omega^2/c^2}$ is the out-of-plane momentum of the transversal modes.

The electric fields, on the other hand, are given by the expressions

$$E_{\rm I}^x(x,y) = \frac{i\kappa_{\rm T}^{(1)}}{\omega\varepsilon_0\varepsilon_1^x} \eta e^{\kappa_{\rm T}^{(1)} y} e^{iqx}, \tag{A5}$$

$$E_{\rm III}^x(x,y) = \frac{i\kappa_{\rm T}^{(3)}}{\omega\varepsilon_0\varepsilon_3^x} \left(\alpha^+ e^{\kappa_{\rm T}^{(3)} y} - \alpha^- e^{-\kappa_{\rm T}^{(3)} y}\right) e^{iqx}, \tag{A6}$$

$$E_{\rm IV}^x(x,y) = \frac{-i\kappa_{\rm T}^{(4)}}{\omega\varepsilon_0\varepsilon_4^x} \zeta e^{-\kappa_{\rm T}^{(4)} y} e^{iqx}, \tag{A7}$$

for the dielectric regions, and by

$$E_{\rm II}^x(x,y) = \left[\frac{i\kappa_{\rm T}^{(2)}}{\omega\varepsilon_0\varepsilon_3^x}\left(\gamma^+ e^{\kappa_{\rm T}^{(2)} y} - \gamma^- e^{-\kappa_{\rm T}^{(2)} y}\right) + \left(\delta^+ e^{\kappa_{\rm L}^{(2)} y} + \delta^- e^{-\kappa_{\rm L}^{(2)} y}\right)\right] e^{iqx}, \tag{A8}$$

for the metallic region, with $\delta^\pm$ being two additional unknown coefficients that arise from the nonlocality, and $\kappa_{\rm L}^{(2)} = \sqrt{q^2 - (\omega^2 + i\gamma_{\rm m}\omega - \omega_{\rm p}^2/\varepsilon_\infty)/\beta^2}$ is the vertical momentum of the longitudinal modes. Furthermore, in region II, the normal component of the electric field is given by

$$E_{\rm II}^y(x,y) = \left[\frac{q}{\omega\varepsilon_0\varepsilon_3^x}\left(\gamma^+ e^{\kappa_{\rm T}^{(2)} y} + \gamma^- e^{-\kappa_{\rm T}^{(2)} y}\right) + \frac{\kappa_{\rm L}^{(2)}}{iq}\left(\delta^+ e^{\kappa_{\rm L}^{(2)} y} - \delta^- e^{-\kappa_{\rm L}^{(2)} y}\right)\right] e^{iqx}, \tag{A9}$$

which fully determines the polarization in the y direction, through Eq. (15), with the form $P_{\rm II}^y(x,y) = (q/\omega) H_{\rm II}^z(x,y) - \varepsilon_0\varepsilon_\infty E_{\rm II}^y(x,y)$.

There are then eight undetermined coefficients in the problem: $\eta, \gamma^\pm, \delta^\pm, \alpha^\pm$ and ζ, which are determined through the application of the boundary conditions of the problem:

$$H_{\rm I}^z(x,-h) = H_{\rm II}^z(x,-h), \tag{A10}$$

$$E_{\rm I}^x(x,-h) = E_{\rm II}^x(x,-h), \tag{A11}$$

$$P_{\rm II}^y(x,-h) = 0, \tag{A12}$$

$$H_{\rm II}^z(x,0) = H_{\rm III}^z(x,0), \tag{A13}$$

$$E_{\rm II}^x(x,0) = E_{\rm III}^x(x,0), \tag{A14}$$

$$P_{\rm II}^y(x,0) = 0, \tag{A15}$$

$$H_{\rm III}^z(x,s) - H_{\rm IV}^z(x,s) = -\sigma E_{\rm IV}^x(x,s), \tag{A16}$$

$$E_{\rm III}^x(x,s) = E_{\rm IV}^x(x,s), \tag{A17}$$

where $\sigma(q,\omega)$ is the conductivity of the graphene sheet, calculated using Mermin's formula in Appendix C.

It is straightforward to check that the eight equations (A10)–(A17) form an undetermined system of equations, meaning that it cannot be solved for all the unknown coefficients. In order to calculate the dispersion relation of the allowed plasmonic modes, we need therefore to define $\zeta \equiv H_0$ as a free parameter of the problem, and find the remaining coefficients as a function of this parameter. Using the boundary equations (A10)–(A16), we find the coefficients η, $\gamma^\pm$, $\delta^\pm$, and $\alpha^\pm$, normalized to H_0.

The remaining boundary condition (A17), on the other hand, can be solved in order to find the relation between q and ω, which ensures the solubility of the system of equations; it corresponds to the dispersion relation of the allowed plasmonic solution. This equation is given by

$$\frac{\kappa_{\rm T}^{(3)}}{\varepsilon_3^x}\left(\frac{\alpha^+}{H_0} e^{\kappa_{\rm T}^{(3)} s} - \frac{\alpha^-}{H_0} e^{-\kappa_{\rm T}^{(3)} s}\right) = -\frac{\kappa_{\rm T}^{(4)}}{\varepsilon_4^x} e^{-\kappa_{\rm T}^{(4)} s}, \tag{A18}$$

where we need to bear in mind that $\alpha^\pm/H_0$ are, at this point, totally determined coefficients (function of both q and ω, in general).

The process described above takes explicitly in consideration the nonlocal effects. The counterpart local dispersion relation, on its turn, can be calculated when explicitly considering $\delta^\pm = 0$ and ignoring the boundary conditions (A12) and (A15), proceeding analogously otherwise.

APPENDIX B: CALCULATION OF THE OPTICAL PROPERTIES

In this Appendix, we describe the procedure we employed to calculate the optical properties of the system presented in Fig. 7. We will be considering that the media in regions III and IV may be uniaxial, and we will thus describe it by two dielectric functions along the in-plane (ε^x) and out-of-plane (ε^y) directions. We limit this generalizations to these regions, because if media I and/or V were uniaxial, it would have consequences in the definitions of the reflectance and the transmittance.

Let us consider that the impinging light carries a momentum $\mathbf{k} = k_x\hat{\mathbf{x}} - k_y\hat{\mathbf{y}}$ with $k_x = k_0 \sin(\theta)$, $k_y = k_0 \cos(\theta)$, and $k_0 = \sqrt{\varepsilon_5}\omega/c$. Due to the grating, this field will be diffracted, so the scattered fields need to be described by a combination of several different diffraction modes n, written in each region as

$$H_{\rm I}^z(x,y) = H_0 \sum_n \tau_n e^{-ik_{{\rm T},n}^{(1)} y} e^{i\rho_n x}, \tag{B1}$$

$$H_{\rm II}^z(x,y) = H_0 \sum_n \left(\gamma_n^+ e^{ik_{{\rm T},n}^{(2)} y} + \gamma_n^- e^{-ik_{{\rm T},n}^{(2)} y}\right) e^{i\rho_n x}, \tag{B2}$$

$$H^z_{\rm III}(x,y) = H_0 \sum_n \left(\alpha_n^+ e^{ik^{(3)}_{{\rm T},n} y} + \alpha_n^- e^{-ik^{(3)}_{{\rm T},n} y}\right) e^{i\rho_n x}, \quad \text{(B3)}$$

$$H^z_{\rm IV}(x,y) = H_0 \sum_n \left(\phi_n^+ e^{ik^{(4)}_{{\rm T},n} y} + \phi_n^- e^{-ik^{(4)}_{{\rm T},n} y}\right) e^{i\rho_n x}, \quad \text{(B4)}$$

$$H^z_{\rm V}(x,y) = H_0 e^{ik_x x} e^{-ik_y y} + H_0 \sum_n r_n e^{ik^{(5)}_{{\rm T},n} y} e^{i\rho_n x}, \quad \text{(B5)}$$

where $\rho_n = k_x + 2n\pi/d$ is the in-plane momentum of the nth mode, as settled by the Bloch theorem, and $k^{(\nu)}_{{\rm T},n} = \sqrt{\varepsilon^x_\nu \omega^2/c^2 - \rho_n^2 \varepsilon^x_\nu/\varepsilon^y_\nu}$ is the out-of-plane momentum in the region ν for the transversal waves.

The electric field, in its turn, is written as

$$E^x_{\rm I}(x,y) = H_0 \left(\frac{1}{\omega\varepsilon_0 \varepsilon^x_4}\right) \sum_n \tau_n k^{(1)}_{{\rm T},n} e^{-ik^{(1)}_{{\rm T},n} y} e^{i\rho_n x}, \quad \text{(B6)}$$

and analogously for regions III, IV, and V, whereas, for region II, it is written as

$$E^x_{\rm II}(x,y) = \left[\frac{-k^{(2)}_{{\rm T},n}}{\omega\varepsilon_0\varepsilon^x_3}\left(\gamma^+ e^{ik^{(2)}_{{\rm T},n} y} - \gamma^- e^{-ik^{(2)}_{{\rm T},n} y}\right) + \left(\delta^+ e^{ik^{(2)}_{{\rm L},n} y} + \delta^- e^{-ik^{(2)}_{{\rm L},n} y}\right)\right] e^{i\rho_n x}, \quad \text{(B7)}$$

where $k_{{\rm L},n} = \sqrt{(\omega^2 + i\gamma_{\rm m}\omega - \omega_{\rm p}^2/\varepsilon_\infty)/\beta^2 - \rho_n^2}$ is the out-of-plane momentum for the longitudinal waves. The y-component $E^y_{\rm II}$ is defined analogously. Note that $k_{{\rm L},n}$ and $k_{{\rm T},n}$ were defined appropriately to describe propagating fields.

There is a total of ten undetermined coefficients in the problem for each mode n (τ_n, $\gamma_n^\pm$, $\delta_n^\pm$, $\alpha_n^\pm$, $\phi_n^\pm$ and r_n), which must be determined through the boundary conditions of the system. Boundary conditions (A10)–(A15) and (A17) still hold, to which we need to add two more conditions,

$$H^z_{\rm IV}(x, s+b) = H^z_{\rm V}(x, s+b), \quad \text{(B8)}$$

$$E^x_{\rm IV}(x, s+b) = E^x_{\rm V}(x, s+b). \quad \text{(B9)}$$

To illustrate the next step, let us consider boundary condition (A10), with the form

$$H_0 \sum_n \tau_n e^{ik^{(1)}_{{\rm T},n} h} e^{i\rho_n x} = H_0 \sum_n \left(\gamma_n^+ e^{-ik^{(2)}_{{\rm T},n} h} + \gamma_n^- e^{ik^{(2)}_{{\rm T},n} h}\right) e^{i\rho_n x}. \quad \text{(B10)}$$

Multiplying the previous equation, on both sides, by $e^{-i\rho_\ell x}$ and integrating the resulting equation in one unit cell of the system ($|x| < d/2$), we obtain integrals with the form $\int_{-d/2}^{d/2} dx e^{i(\rho_n - \rho_\ell)x} = d\delta_{n\ell}$, which then yield

$$\tau_\ell e^{ik^{(1)}_{{\rm T},\ell} h} = \gamma_\ell^+ e^{-ik^{(2)}_{{\rm T},\ell} h} + \gamma_\ell^- e^{ik^{(2)}_{{\rm T},\ell} h}. \quad \text{(B11)}$$

Notice that the previous equation, for each mode ℓ, relates only the coefficients of the same index ℓ —or, in other words, the boundary condition is obeyed by each individual mode that composes the total field, and not only by the total field itself, which strongly simplifies the problem. This property is equally obeyed by the remaining (A11)–(A15), (A17) and (B8)–(B9) conditions, meaning that, proceeding as illustrated above, we arrive at nine equations which relate all the coefficients for some particular mode ℓ. These form a determined system of equations which allows for the calculation of nine out of the ten different coefficients, as a function of the remaining one. We will henceforth take the undetermined coefficient to be the r_ℓ; in that case, one finds that all the other coefficients are related to r_ℓ by a linear relation of the form

$$\alpha_\ell^\pm = \theta_0^\pm \delta_{\ell 0} + r_\ell \psi_\ell^\pm, \qquad \phi_\ell^\pm = \lambda_0^\pm \delta_{\ell 0} + r_\ell \chi_\ell^\pm, \quad \text{(B12)}$$

and analogously for τ_ℓ, $\gamma_\ell^\pm$, and $\delta_\ell^\pm$.

The remaining unknown coefficient, r_ℓ, must now be determined using the remaining boundary condition, which corresponds to the discontinuity of the magnetic field across the graphene grating, due to the presence of surface currents in the graphene. This condition, analogous to the condition expressed in Eq. (A16), has the explicit form

$$\sum_n \left(\alpha_n^+ e^{ik^{(3)}_{{\rm T},n} s} + \alpha_n^- e^{-ik^{(3)}_{{\rm T},n} s}\right) e^{i\rho_n x} - \sum_n \left(\phi_n^+ e^{ik^{(4)}_{{\rm T},n} s} + \phi_n^- e^{-ik^{(4)}_{{\rm T},n} s}\right) e^{i\rho_n x} = \sum_n \left(\frac{\sigma_n k^{(4)}_{{\rm T},n}}{\omega\varepsilon_0\varepsilon^x_4}\right) \left(\phi_n^+ e^{ik^{(4)}_{{\rm T},n} s} - \phi_n^- e^{-ik^{(4)}_{{\rm T},n} s}\right) e^{i\rho_n x}, \quad \text{(B13)}$$

where $\sigma_n = \sigma(\rho_n, \omega)$ is the conductivity of the graphene. The big difference between this condition and all the others is that graphene's conductivity is only nonzero for $|x| < w/2$; this means that, when multiplying both sides of the equation by $e^{-i\rho_\ell x}$ and integrating it on the unit cell of the system, we obtain

$$\tilde\alpha_\ell^+ + \tilde\alpha_\ell^- - \tilde\phi_\ell^+ - \tilde\phi_\ell^- = \sum_n \left(\frac{\sigma_n k^{(4)}_{{\rm T},n}}{\omega\varepsilon_0\varepsilon^x_4}\right) (\tilde\phi_n^+ - \tilde\phi_n^-) S_{\ell n}, \quad \text{(B14)}$$

with $\tilde\alpha_n^\pm \equiv \alpha_n^\pm e^{\pm ik^{(3)}_{{\rm T},n} s}$ (and equivalently for $\tilde\phi_n^\pm$), and

$$S_{\ell n} = \int_{-w/2}^{w/2} dx e^{i(\rho_n - \rho_\ell)x} = \frac{\sin[\pi w(n-\ell)/d]}{\pi(n-\ell)}. \quad \text{(B15)}$$

This equation, unlike all the others, relate coefficients of all the allowed diffraction orders n. To solve it, we employ equations (B12) and, upon some mathematical manipulation, we arrive at an expression of the form $\sum_n M_{\ell n} r_n = -F_\ell$, where

$$F_\ell = (\tilde\theta_0^+ + \tilde\theta_0^- - \tilde\lambda_0^+ - \tilde\lambda_0^-)\delta_{\ell 0} - \left[\frac{\sigma_0 k^{(4)}_{{\rm T},0}}{\omega\varepsilon_0\varepsilon^x_4}\right] (\tilde\lambda_0^+ - \tilde\lambda_0^-) S_{\ell 0}, \quad \text{(B16)}$$

$$M_{\ell n} = (\tilde\psi_n^+ + \tilde\psi_n^- - \tilde\chi_n^+ - \tilde\chi_n^-)\delta_{\ell n} - \left[\frac{\sigma_n k^{(4)}_{{\rm T},n}}{\omega\varepsilon_0\varepsilon^x_4}\right] (\tilde\chi_n^+ - \tilde\chi_n^-) S_{\ell n}. \quad \text{(B17)}$$

This equation may be written in the matrix form as $\mathbb{M}\cdot\mathbb{R} = -\mathbb{F}$, with $\mathbb{M}$ being a square matrix with elements $[\mathbb{M}]_{\ell n} = M_{\ell n}$, and $\mathbb{R}$ and $\mathbb{F}$ being columns with elements $[\mathbb{R}]_\ell = r_\ell$ and $[\mathbb{F}]_\ell = F_\ell$. The r_ℓ are hence determined through the solution of that matrix equation, written for a sufficiently large matrix dimension to ensure the convergence of the solution. The remaining coefficients are then determined by equations (B12) and analogously for τ_n, $\gamma_n^{\pm}$, and $\delta_n^{\pm}$.

After this process, we can find the reflectance and transmittance of the system by the equations

$$\mathcal{R} = \sum_{n\in\mathrm{PM}} \mathrm{Re}\left[\frac{k_{\mathrm{T},n}^{(5)}}{\varepsilon_5}\right]\mathrm{Re}\left[\frac{\varepsilon_5}{k_y}\right]|r_n|^2, \tag{B18}$$

$$\mathcal{T} = \sum_{n\in\mathrm{PM}} \mathrm{Re}\left[\frac{k_{\mathrm{T},n}^{(1)}}{\varepsilon_4^x}\right]\mathrm{Re}\left[\frac{\varepsilon_5}{k_y}\right]|\tau_n|^2, \tag{B19}$$

where the summations are performed over the propagating modes (PM). The absorbance, on the other hand, is defined as $\mathcal{A} = 1-\mathcal{R}-\mathcal{T}$. Once again, the optical properties for the local case are calculated when repeating the same procedure whilst setting $\delta_n^{\pm} = 0$ and ignoring boundary conditions (A12) and (A15).

Apart from the optical properties, it is interesting to note that, from the reflectance amplitudes r_n, one can calculate the loss function $L(\omega,q) = -\sum_n \mathrm{Im}[r_n]$ (where k_x is substituted by an arbitrarily changeable momentum q), which is an alternative way to calculate the dispersion relation of the allowed bound modes of the problem. This is shown in Fig. 2, where the loss function was overlaid by the dispersion relation calculated as detailed in Appendix A, and the correspondence between the two plots is excellent.

APPENDIX C: GRAPHENE'S NONLOCAL CONDUCTIVITY

For the conductivity of the graphene sheet, we have used Mermin's formula [54,55] at zero temperature, which includes nonlocal effects. Let $x \equiv q/k_{\mathrm{F}}$ and $y \equiv \hbar\omega/E_{\mathrm{F}}$ be dimensional variables constructed from q and ω, respectively. E_{F} refers to the graphene's Fermi energy, $k_{\mathrm{F}} = E_{\mathrm{F}}/(\hbar v_{\mathrm{F}})$ is the Fermi momentum ($v_{\mathrm{F}} \approx c/300$ is the Fermi speed), and Γ is the material's relaxation energy. The formula we have used was retrieved from Gonçalves and Peres [3],

$$\sigma(q,\omega) = 4\mathrm{i}\sigma_0\frac{\hbar\omega}{q^2}\chi_\tau\left(\frac{q}{k_{\mathrm{F}}},\frac{\hbar\omega}{E_{\mathrm{F}}}\right), \tag{C1}$$

with $\sigma_0 \equiv e^2/(4\hbar)$ and

$$\chi_\tau(x,y) = \frac{\left(1+\mathrm{i}\frac{\Gamma}{yE_{\mathrm{F}}}\right)\chi_{\mathrm{g}}\left(x,y+\mathrm{i}\frac{\Gamma}{E_{\mathrm{F}}}\right)}{1+\mathrm{i}\frac{\Gamma}{yE_{\mathrm{F}}}\chi_{\mathrm{g}}\left(x,y+\mathrm{i}\frac{\Gamma}{E_{\mathrm{F}}}\right)\Big/\chi_{\mathrm{g}}(x,0)}. \tag{C2}$$

The function $\chi_{\mathrm{g}}(x,y)$ is calculated differently according to the region where it is calculated in the xy-phase space represented in Fig. 9. It can be written as

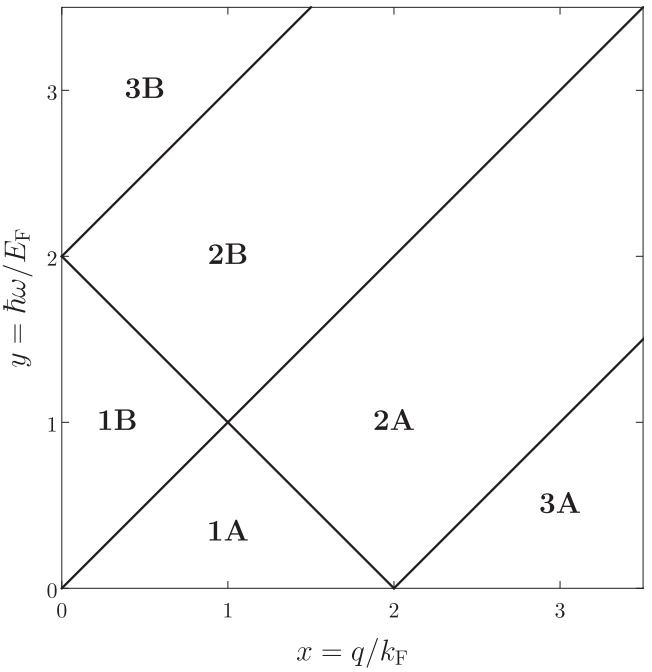

FIG. 9. Regions for the calculation of the graphene susceptibility in the xy-phase space.

$$\chi_{\mathrm{g}}(x,y) = \begin{cases} \chi_{\mathrm{B}}^{(1)}(x,y), & \mathrm{Re}[y] > x \wedge \mathrm{Re}[y] < 2-x, \\ \chi_{\mathrm{B}}^{(2)}(x,y), & \mathrm{Re}[y] > x \wedge \mathrm{Re}[y] > 2-x, \\ \chi_{\mathrm{B}}^{(3)}(x,y), & \mathrm{Re}[y] > x+2, \\ \chi_{\mathrm{A}}^{(1)}(x,y), & \mathrm{Re}[y] < x \wedge \mathrm{Re}[y] < 2-x, \\ \chi_{\mathrm{A}}^{(2)}(x,y), & \mathrm{Re}[y] < x \wedge \mathrm{Re}[y] > 2-x, \\ \chi_{\mathrm{A}}^{(3)}(x,y), & \mathrm{Re}[y] < x-2, \end{cases} \tag{C3}$$

where $\mathrm{Re}[y]$ stands for the real part of y, and each of the functions in the previous expression are given by

$$\chi_{\mathrm{B}}^{(1)}(x,y) = -\frac{2}{\pi}\frac{E_{\mathrm{F}}}{(\hbar v_{\mathrm{F}})^2} + \frac{1}{4\pi}\frac{E_{\mathrm{F}}}{(\hbar v_{\mathrm{F}})^2}\frac{x^2}{\sqrt{y^2-x^2}} \times\left[F\left(\frac{y+2}{x}\right) - F\left(\frac{2-y}{x}\right)\right], \tag{C4}$$

$$\chi_{\mathrm{B}}^{(2)}(x,y) = -\frac{2}{\pi}\frac{E_{\mathrm{F}}}{(\hbar v_{\mathrm{F}})^2} + \frac{1}{4\pi}\frac{E_{\mathrm{F}}}{(\hbar v_{\mathrm{F}})^2}\frac{x^2}{\sqrt{y^2-x^2}} \times\left[F\left(\frac{y+2}{x}\right) + \mathrm{i}G\left(\frac{2-y}{x}\right)\right], \tag{C5}$$

$$\chi_{\mathrm{B}}^{(3)}(x,y) = -\frac{2}{\pi}\frac{E_{\mathrm{F}}}{(\hbar v_{\mathrm{F}})^2} + \frac{1}{4\pi}\frac{E_{\mathrm{F}}}{(\hbar v_{\mathrm{F}})^2}\frac{x^2}{\sqrt{y^2-x^2}} \times\left[-\mathrm{i}\pi + F\left(\frac{y+2}{x}\right) - F\left(\frac{y-2}{x}\right)\right], \tag{C6}$$

$$\chi_{\mathrm{A}}^{(1)}(x,y) = -\frac{2}{\pi}\frac{E_{\mathrm{F}}}{(\hbar v_{\mathrm{F}})^2} - \frac{\mathrm{i}}{4\pi}\frac{E_{\mathrm{F}}}{(\hbar v_{\mathrm{F}})^2}\frac{x^2}{\sqrt{x^2-y^2}} \times\left[F\left(\frac{y+2}{x}\right) - F\left(\frac{2-y}{x}\right)\right], \tag{C7}$$

$$\chi_{\rm A}^{(2)}(x,y) = -\frac{2}{\pi}\frac{E_{\rm F}}{(\hbar v_{\rm F})^2} + \frac{\rm i}{4\pi}\frac{E_{\rm F}}{(\hbar v_{\rm F})^2}\frac{x^2}{\sqrt{x^2-y^2}}\left[{\rm i}\pi - F\left(\frac{y+2}{x}\right) + {\rm i}G\left(\frac{2-y}{x}\right)\right], \tag{C8}$$

$$\chi_{\rm A}^{(3)}(x,y) = -\frac{2}{\pi}\frac{E_{\rm F}}{(\hbar v_{\rm F})^2} + \frac{1}{4\pi}\frac{E_{\rm F}}{(\hbar v_{\rm F})^2}\frac{x^2}{\sqrt{x^2-y^2}}\left[-\pi + G\left(\frac{y+2}{x}\right) - G\left(\frac{y-2}{x}\right)\right] \tag{C9}$$

with the functions $F(x) \equiv x\sqrt{x^2-1} - \text{arccosh}(x)$ and $G(x) \equiv x\sqrt{1-x^2} - \text{arccosh}(x)$.

[1] S. A. Maier, *Plasmonics: Fundamentals and Applications* (Springer, 2007).
[2] M. Pelton and G. W. Bryant, *Introduction to Metal-Nanoparticle Plasmonics* (Wiley, 2013).
[3] P. A. D. Gonçalves and N. M. R. Peres, *An Introduction to Graphene Plasmonics*, 1st ed. (World Scientific, Singapore, 2016).
[4] V. Klimov, *Nanoplasmonics* (CRC Press, 2014).
[5] A. I. Fernández-Domínguez, F. J. García-Vidal, and L. Martín-Moreno, Nat. Photonics **11**, 8 (2017).
[6] F. Benz, M. K. Schmidt, A. Dreismann, R. Chikkaraddy, Y. Zhang, A. Demetriadou, C. Carnegie, H. Ohadi, B. de Nijs, R. Esteban, J. Aizpurua, and J. J. Baumberg, Science **354**, 726 (2016).
[7] C. Ciracì, R. Hill, J. J. Mock, Y. Urzhumov, A. I. Fernández-Domínguez, S. A. Maier, J. B. Pendry, A. Chilkoti, and D. R. Smith, Science **337**, 1072 (2012).
[8] S. Raza, S. Kadkhodazadeh, T. Christensen, M. Di Vece, M. Wubs, N. A. Mortensen, and N. Stenger, Nat. Commun. **6**, 8788 (2015).
[9] M. B. Lundeberg, Y. Gao, R. Asgari, C. Tan, B. Van Duppen, M. Autore, P. Alonso-González, A. Woessner, K. Watanabe, T. Taniguchi, R. Hillenbrand, J. Hone, M. Polini, and F. H. L. Koppens, Science **357**, 187 (2017).
[10] S. Raza, G. Toscano, A.-P. Jauho, M. Wubs, and N. A. Mortensen, Phys. Rev. B **84**, 121412(R) (2011).
[11] N. A. Mortensen, S. Raza, M. Wubs, T. Søndergaard, and S. I. Bozhevolnyi, Nat. Commun. **5**, 3809 (2014).
[12] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. (Wiley, New York, 1998).
[13] S. Raza, T. Christensen, M. Wubs, S. I. Bozhevolnyi, and N. A. Mortensen, Phys. Rev. B **88**, 115401 (2013).
[14] C. David and F. J. García de Abajo, J. Phys. Chem. C **115**, 19470 (2011).
[15] A. I. Fernández-Domínguez, A. Wiener, F. J. García-Vidal, S. A. Maier, and J. B. Pendry, Phys. Rev. Lett. **108**, 106802 (2012).
[16] J. A. Scholl, A. L. Koh, and J. A. Dionne, Nature (London) **483**, 421 (2012).
[17] J. Khurgin, W.-Y. Tsai, D. P. Tsai, and G. Sun, ACS Photonics **4**, 2871 (2017).
[18] F. J. García de Abajo, J. Phys. Chem. C **112**, 17983 (2008).
[19] C. David and F. J. García de Abajo, AIP Conf. Proc. **1291**, 43 (2010).
[20] A. Wiener, H. Duan, M. Bosman, A. P. Horsfield, J. B. Pendry, J. K. W. Yang, S. A. Maier, and A. I. Fernández-Domínguez, ACS Nano **7**, 6287 (2013).
[21] O. Schnitzer, V. Giannini, R. V. Craster, and S. A. Maier, Phys. Rev. B **93**, 041409 (2016).
[22] G. Toscano, S. Raza, A.-P. Jauho, N. A. Mortensen, and M. Wubs, Opt. Express **20**, 4176 (2012).
[23] H. Shen, L. Chen, L. Ferrari, M.-H. Lin, N. A. Mortensen, S. Gwo, and Z. Liu, Nano Lett. **17**, 2234 (2017).
[24] L. Stella, P. Zhang, F. J. García-Vidal, A. Rubio, and P. García-González, J. Phys. Chem. C **117**, 8941 (2013).
[25] A. Varas, P. García-González, J. Feist, F. J. García-Vidal, and A. Rubio, Nanophotonics **5**, 409 (2016).
[26] A. D. Boardman, B. V. Paranjape, and R. Teshima, Surf. Sci. **49**, 275 (1975).
[27] A. D. Boardman, *Electromagnetic Surface Modes* (Wiley, 1982).
[28] F. Bloch, Ann. Phys. **408**, 285 (1933).
[29] F. Bloch, Z. Phys. A **81**, 363 (1933).
[30] H. Jensen, Z. Phys. **106**, 620 (1937).
[31] A. Moreau, C. Ciracì, and D. R. Smith, Phys. Rev. B **87**, 045401 (2013).
[32] C. Ciracì, J. B. Pendry, and D. R. Smith, Chem. Phys. Chem. **14**, 1109 (2013).
[33] S. Raza, S. I. Bozhevolnyi, M. Wubs, and N. A. Mortensen, J. Phys.: Condens. Matter **27**, 183204 (2015).
[34] G. Toscano, J. Straubel, A. Kwiatkowski, C. Rockstuhl, F. Evers, H. Xu, N. A. Mortensen, and M. Wubs, Nat. Commun. **6**, 7132 (2015).
[35] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, Proc. Natl. Acad. Sci. USA **102**, 10451 (2005).
[36] A. Woessner, M. Lundeberg, Y. Gao, A. Principi, P. Alonso-González, M. Carrega, K. Watanabe, T. Taniguchi, G. Vignale, M. Polini, J. Hone, R. Hillenbrand, and F. H. L. Koppens, Nat. Mater. **14**, 421 (2015).
[37] A. Principi, M. Carrega, M. B. Lundeberg, A. Woessner, F. H. L. Koppens, G. Vignale, and M. Polini, Phys. Rev. B **90**, 165408 (2014).
[38] P. A. D. Gonçalves, E. J. C. Dias, Y. V. Bludov, and N. M. R. Peres, Phys. Rev. B **94**, 195421 (2016).
[39] E. J. C. Dias and N. M. R. Peres, ACS Photonics **4**, 3071 (2017).
[40] V. W. Brar, M. S. Jang, M. Sherrott, S. Kim, J. J. Lopez, L. B. Kim, M. Choi, and H. Atwater, Nano Lett. **14**, 3876 (2014).
[41] A. D. Rakić, A. B. Djurišić, J. M. Elazar, and M. L. Majewski, Appl. Opt. **37**, 5271 (1998).
[42] T. Brandt, M. Hövel, B. Gompf, and M. Dressel, Phys. Rev. B **78**, 205409 (2008).
[43] D. Barchiesi and T. Grosges, J. Nanophotonics **8**, 083097 (2014).
[44] A. Derkachova, K. Kolwas, and I. Demchenko, Plasmonics **11**, 941 (2016).

[45] J. S. Lehtinen, T. Sajavaara, K. Y. Arutyunov, M. Y. Presnjakov, and A. L. Vasiliev, Phys. Rev. B **85**, 094508 (2012).
[46] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, 1976).
[47] Y. Luo, A. I. Fernández-Domínguez, A. Wiener, S. A. Maier, and J. B. Pendry, Phys. Rev. Lett. **111**, 093901 (2013).
[48] Y. Luo, R. Zhao, and J. B. Pendry, Proc. Natl. Acad. Sci. USA **111**, 18422 (2014).
[49] W. Yan, M. Wubs, and N. A. Mortensen, Phys. Rev. Lett. **115**, 137403 (2015).
[50] T. Christensen, W. Yan, A.-P. Jauho, M. Soljačić, and N. A. Mortensen, Phys. Rev. Lett. **118**, 157402 (2017).
[51] P. Alonso-González, A. Y. Nikitin, Y. Gao, A. Woessner, M. B. Lundeberg, A. Principi, N. Forcellini, W. Yan, S. Vélez, A. J. Huber, K. Watanabe, T. Taniguchi, F. Casanova, L. E. Hueso, M. Polini, J. Hone, F. H. L. Koppens, and R. Hillenbrand, Nat. Nanotechnol. **12**, 31 (2017).
[52] E. D. Palik, *Handbook of Optical Constants of Solids* (Academic Press, 1998), Vol. 3.
[53] A. Dargys and J. Kundrotas, *Handbook on Physical Properties of Ge, Si, GaAs and InP* (Science and Encyclopedia, 1994).
[54] N. D. Mermin, Phys. Rev. B **1**, 2362 (1970).
[55] A. K. Das, J. Phys. F **5**, 2035 (1975).