Plasmons in tunnel-coupled graphene layers: Backward waves with quantum cascade gain

D. Svintsov,[1,*] Zh. Devizorova,[2,3] T. Otsuji,[4] and V. Ryzhii[4]

[1]*Laboratory of 2d Materials' Optoelectronics, Moscow Institute of Physics and Technology, Dolgoprudny 141700, Russia*

[2]*Department of Physical and Quantum Electronics, Moscow Institute of Physics and Technology, Dolgoprudny 141700, Russia*

[3]*Kotelnikov Institute of Radio Engineering and Electronics, Russian Academy of Science, Moscow 125009, Russia*

[4]*Research Institute of Electrical Communication, Tohoku University, Sendai 980-8577, Japan*

(Received 16 March 2016; revised manuscript received 17 August 2016; published 6 September 2016)

We theoretically demonstrate that graphene-insulator-graphene tunnel structures can serve as plasmonic gain media due to the possibility of stimulated electron tunneling accompanied by emission of plasmons under application of interlayer voltage. The probability of plasmon-assisted tunneling is resonantly large at certain values of frequency and interlayer voltage corresponding to the transitions between chiral electron states with collinear momenta, which is a feature unique to the linear bands of graphene. The plasmon dispersion develops an anticrossing with the resonances in tunnel conductivity and demonstrates negative group velocity in several frequency ranges.

DOI: 10.1103/PhysRevB.94.115301

I. INTRODUCTION

The ultrarelativistic nature of electrons in graphene gives rise to the uncommon properties of their collective excitations—surface plasmons (SPs) [1–3]. The deep subwavelength confinement [2], the unconventional density dependence of frequency [3,4], and the absence of Landau damping [4] are their most well-known features. Among more sophisticated predictions there stand the existence of transverse electric plasmon modes [5] and quasineutral electron-hole sound at the charge neutrality [6,7]. It was not until the discovery of van der Waals heterostructures that the low-loss SPs supported by graphene could be observed [8]. The reported propagation length to wavelength ratio reaching 25 looks to be the fundamental limit of SP quality factor at room temperature governed by the electron-phonon interaction [9], which hinders further experimental studies of plasmonic effects. The perception of this fact has motivated the search for the methods to provide the gain of SPs in graphene [10–14].

Here, we theoretically demonstrate that the resonant tunneling structures composed of parallel graphene layers can act as plasmonic gain media by themselves. The very presence of negative differential resistance (NDR) in the static current-voltage characteristics of these structures [15] gives rise to the self-oscillation in electrical circuits [16] and might potentially lead to to the self-excitation of plasmons [17–19]. However, the *static* NDR of graphene tunnel diodes is insufficient to replenish the plasmon losses [20], which calls for the stability of electron plasma.

In this paper, we show that the dynamic and nonlocal effects in the tunnel conductivity radically change the picture of plasmon propagation. The calculated dynamic tunnel conductivity of a double graphene layer biased by voltage V has a negative real part at frequencies $\omega < eV/\hbar$, even if the static NDR is absent in the structure. This can be viewed as a consequence of the "interlayer population inversion." Surprisingly, the dynamic nonlocal conductivity possesses sharp resonances at certain frequencies and wave vectors q due to the prolonged tunneling interaction between chiral states with collinear momenta in neighboring layers. The singularities in the tunnel conductivity emerge at a series of lines on the ω-q plane, whose pattern is especially rich in twisted layers. At finite bias V, the dispersion of acoustic SPs does not develop a low-frequency gap, as opposed to SPs in coupled layers of massive electrons in equilibrium [21]. Instead, the SP spectrum develops an anticrossing with the tunnel resonances and demonstrates the parts with negative group velocity. At the same time, the dispersion is quite close to the tunnel resonances, and the tunnel gain can exceed the SP loss due to both inter- and intraband absorption.

The paper is organized as follows. In Sec. II we derive the dispersion relation for plasmons in tunnel-coupled graphene layers. Section III is devoted to the calculation of high-frequency nonlocal conductivity, which possesses both tunneling and in-plane components. Both components of conductivity affect the spectrum and damping (or gain) of acoustic plasmons, which is analyzed in Sec. IV. Section V is devoted to the effect of interlayer twist on high-frequency conductivity and SP dispersion. Possible experimental manifestations of the predicted effects are discussed in Sec. VI. Some cumbersome calculations are relegated to the Appendices.

II. DISPERSION RELATION FOR PLASMONS IN TUNNEL-COUPLED LAYERS

We consider the propagation of plasmons in the graphene-insulator-graphene structure shown in Fig. 1(a). The application of interlayer voltage V results in the electrical doping of layers; the corresponding filling of the bands is shown in Fig. 1(b). In the absence of built-in voltage, the density of induced electrons in the top layer equals the density of holes in the bottom one. This results in equal in-plane conductivities of the layers $\sigma_{\parallel}$.

The method of obtaining the SP dispersion relies on the solution of Poisson's equation and the carrier transport equations. The interlayer tunnel current, $\delta J_{\perp}$, appears as a source term in the continuity equation for the charge densities $\delta Q_{t,b}$ on the top and bottom layers:

$$-i\omega\delta Q_{t,b} = -i\mathbf{q}\delta\mathbf{j}_{t,b} \mp \delta J_{\perp}. \quad (1)$$

*svintcov.da@mipt.ru

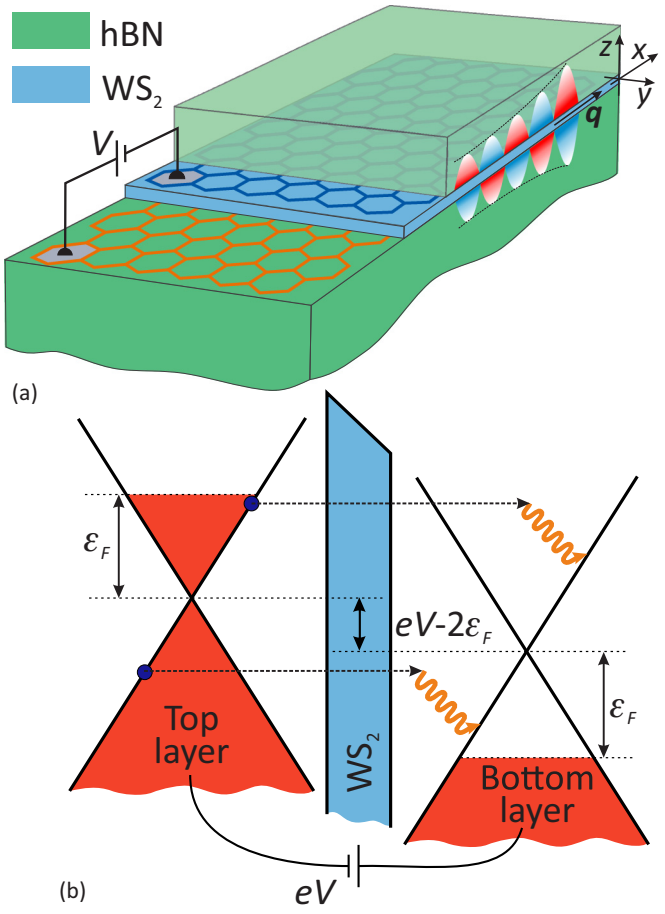

FIG. 1. (a) Schematic view of the double graphene layer encapsulated in hexagonal boron nitride (hBN) overlaid by the image of acoustic SP amplified by the tunneling. (b) Band diagram of the structure biased by voltage V. Inelastic interlayer electron tunneling transitions accompanied plasmon emission (wavy arrows) are shown schematically.

Here $\delta\mathbf{j}_{t,b}$ are the in-plane current densities which are related to the electric potentials $\delta\varphi_{t,b}$ via $\delta\mathbf{j}_{t,b} = -i\mathbf{q}\sigma_\parallel(\mathbf{q},\omega)\delta\varphi_{t,b}$. The interlayer current density is

$$\delta J_\perp = G_\perp(\mathbf{q},\omega)(\delta\varphi_t - \delta\varphi_b),$$

where $G_\perp(\mathbf{q},\omega)$ is tunnel conductivity. By matching the solutions of the Poisson's equation (κ is the background dielectric constant)

$$-q^2\delta\varphi(z) + \frac{\partial^2\delta\varphi(z)}{\partial z^2} = -\frac{4\pi}{\kappa}[\delta Q_t\delta(z-d/2) + \delta Q_b\delta(z+d/2)] \quad (2)$$

on graphene planes located at $z = \pm d/2$, we arrive at the plasmon dispersion equation which can be written as

$$\epsilon(\mathbf{q},\omega) \equiv \left[1 + \frac{2\pi i q}{\omega\kappa}\sigma_\parallel(1+e^{-qd})\right] \times \left[1 + \frac{2\pi i q}{\omega\kappa}\left(\sigma_\parallel + \frac{2G_\perp}{q^2}\right)(1-e^{-qd})\right] = 0. \quad (3)$$

The zeros of the first and second terms in the dielectric function $\epsilon(\mathbf{q},\omega)$ yield the dispersions of the optical and acoustic SPs, respectively. It is intuitive that the presence of tunneling does not affect the optical mode, while it can modify the spectrum of the acoustic branch considerably. The reason is that the average interlayer field in the optical mode is zero, and it cannot stimulate the interlayer carrier transfer, contrary to the strong field in the acoustic mode. We note here that the ratio of transverse and in-plane electric fields in the acoustic mode is approximately $2(qd)^{-1} \gg 1$, which also speaks in favor of the strong tunneling effects.

In addition to solving Eq. (3) for the mode dispersions $\omega(q)$, we shall study the loss function [22], $-\mathrm{Im}\epsilon^{-1}(\mathbf{q},\omega)$ (also referred to as plasmon spectral weight [23]). Within the RPA treatment, this quantity is proportional to the dynamical structure factor measured with inelastic scattering experiments. To focus on acoustic plasmons, we define the acoustic contribution to the loss function:

$$-\mathrm{Im}\epsilon^{-1}_{\mathrm{ac}}(\mathbf{q},\omega) = -\mathrm{Im}\left\{1 + \frac{2\pi i q}{\omega\kappa}\left(\sigma_\parallel + \frac{2G_\perp}{q^2}\right)(1-e^{-qd})\right\}^{-1}. \quad (4)$$

The peaks in $-\mathrm{Im}\epsilon^{-1}_{\mathrm{ac}}(\mathbf{q},\omega)$ correspond to the acoustic SPs, while the sign of the loss function determines whether the plasmons are damped or amplified.

III. HIGH-FREQUENCY NONLOCAL TUNNEL CURRENT

The only missing ingredient required for the analysis of surface plasmon modes is the expression for the high-frequency nonlocal tunnel conductivity $G_\perp(\mathbf{q},\omega)$. The theoretical studies of the latter have been limited to the dc [24,25] or local ($q = 0$) ac cases [26]. Here, we consider the linear response of voltage-biased graphene layers to the propagating acoustic plasmon whose distribution of electric potential $\delta\varphi(z)e^{iqx-i\omega t}$ is highly nonuniform (see Appendix A for explicit expressions). The electrons in tunnel-coupled graphene layers are described with the tight-binding Hamiltonian

$$\hat{H}_0 = \begin{pmatrix} \hat{H}_{G+} & \hat{T} \\ \hat{T}^* & \hat{H}_{G-} \end{pmatrix}, \quad (5)$$

where the blocks $\hat{H}_{G\pm} = v_0\boldsymbol{\sigma}\hat{\mathbf{p}} \pm \hat{I}\Delta/2$ stand for isolated graphene layers, $v_0 = 10^6$ m/s is the Fermi velocity, Δ is the voltage-induced energy spacing between the Dirac points, $\hat{\mathbf{p}}$ is the in-plane momentum operator, $\hat{I}$ is the identity matrix, and $\hat{T} = \Omega\hat{I}$ is the tunneling matrix. Such model of tunnel coupling applies to the AA-aligned graphene bilayer [27,28] and has proved to be useful for the description of dc tunneling in van der Waals heterostructures [24,29]. The effects of the small interlayer twist will be addressed at the end of the paper.

The eigenstates of $\hat{H}_0$ can be labeled by the in-plane momentum $\mathbf{p}$, the index $s = \{c,v\}$ for the conduction and valence bands, respectively, and the number $l = \pm 1$ governing the z localization of the wave function. At strong bias, $\Delta \gg \Omega$, the state with $l = +1$ (-1) is localized primarily on the top (bottom) layer, while at zero bias the states $l = -1$ and $l = +1$ are the symmetric and antisymmetric states of the coupled quantum wells, respectively. The states' energies are $\varepsilon^{ls}_{\mathbf{p}} = sv_0p + l\tilde{\Delta}/2$, where $\tilde{\Delta} = \sqrt{4\Omega^2 + \Delta^2}$ is the level spacing governed by the application of voltage and the tunneling repulsion [30].

The electron interaction with ac plasmon field is described by the Hamiltonian $\delta\hat{V}$ whose matrix elements are the overlap integrals between the wave functions of $\hat{H}_0$ eigenstates $|\mathbf{p}sl\rangle$ and the potential of the acoustic plasmon. To explicitly estimate these matrix elements, we switch from a tight-binding to a continuum description of electron states, and model each graphene layer as a one-dimensional delta well (see Appendix B for details). The strength of the well is chosen to provide the correct value of the work function from graphene to the surrounding dielectric U_b (0.4 eV for graphene embedded in WS_2). Proceeding this way, we are able to present the matrix elements of electron-plasmon interaction as

$$\langle \mathbf{p}sl|\,\delta\hat{V}\,|\mathbf{p}'s'l'\rangle = e\delta\varphi_0 u^{ss'}_{\mathbf{pp}'} S_{ll'}\delta_{\mathbf{p},\mathbf{p}'-\mathbf{q}}, \tag{6}$$

where $\delta\varphi_0$ is the plasmon potential amplitude at the graphene layer, $u^{ss'}_{\mathbf{pp}'}$ is the overlap between chiral wave functions of bands s and s', and $S_{ll'}$ are the dimensionless overlap integrals. Due to the antisymmetry of the plasmon mode distribution, $S_{++} = -S_{--}$ and $S_{\pm} = S_{\mp}$.

The evaluation of interlayer conductivity is based on the solution of the quantum Liouville equation for the electron density matrix $\hat{\rho}$. Being interested in the linear response to the plasmon field, we solve it in the form

$$i\hbar\frac{\partial\delta\hat{\rho}}{\partial t} = [\hat{H}_0,\delta\hat{\rho}] + [\delta\hat{V},\hat{\rho}^{(0)}], \tag{7}$$

where $\hat{\rho}^{(0)}$ is the density matrix in the absence of the ac field (but in the presence of strong dc tunneling), and $\delta\hat{\rho}$ is the sought-for linear correction. The solution of Eq. (7) in the basis of $\hat{H}_0$ eigenstates is immediate, though it relies on a particular choice of $\hat{\rho}^{(0)}$. When the frequency of interlayer tunneling $\Omega/\hbar$ exceeds the energy relaxation frequency in a layer ν_ε, the electron is "collectivized" by the two layers, and the states $|\mathbf{p}sl\rangle$ have a well-defined occupancy [31]. For this reason, we choose $\hat{\rho}^{(0)}$ to be diagonal in this basis, and the diagonal elements are the Fermi functions $f^{sl}_{\mathbf{p}}$ with quasi-Fermi levels shifted by eV for different l [32]. The subsequent calculation is based on the statistical averaging of current operator, which is the time derivative of the charge operator $\hat{Q}_t$:

$$\frac{\partial\delta Q_t}{\partial t} = -\frac{i}{\hbar}\mathrm{Tr}([\hat{Q}_t,\hat{H}_0]\delta\hat{\rho}). \tag{8}$$

Equation (8) with the density matrix obtained from (7) lets us evaluate both in-plane and tunnel conductivities at once, in accordance with the two terms on the right-hand side of the continuity equation (1). A lengthy but straightforward calculation leads to

$$\sigma_\parallel(\mathbf{q},\omega) = -ig\frac{e^2}{\hbar}S_{++}\cos\theta_M \times\sum_{ss'\mathbf{p}}\frac{\left|\mathbf{v}^{ss'}_{\mathbf{pp}'}\right|^2}{\varepsilon^s_{\mathbf{p}_-}-\varepsilon^s_{\mathbf{p}_+}}\frac{f^s_{\mathbf{p}_+}-f^{s'}_{\mathbf{p}_-}}{\varepsilon^s_{\mathbf{p}_+}-\varepsilon^s_{\mathbf{p}_-}-(\hbar\omega+i\delta)}, \tag{9}$$

$$G_\perp(\mathbf{q},\omega) = -ig\frac{e^2}{2\hbar}S_\pm\sin\theta_M\sum_{\substack{l\neq l'\\ss'\mathbf{p}}}\left|u^{ss'}_{\mathbf{pp}'}\right|^2 \times\frac{\varepsilon^{s'l'}_{\mathbf{p}_-}-\varepsilon^{sl}_{\mathbf{p}_+}}{\varepsilon^{sl}_{\mathbf{p}_+}-\varepsilon^{s'l'}_{\mathbf{p}_-}-(\hbar\omega+i\delta)}\left(f^{sl}_{\mathbf{p}_+}-f^{s'l'}_{\mathbf{p}_-}\right). \tag{10}$$

Above, $g = 4$ is the spin-valley degeneracy factor, θ_M is the "mixing angle" characterizing the strength of coupling, $\sin\theta_M = 2\Omega/\tilde{\Delta}$; $\mathbf{p}_\pm \equiv \mathbf{p} \pm \hbar\mathbf{q}/2$, and $\mathbf{v}^{ss'}_{\mathbf{pp}'}$ is the matrix element of the velocity operator between chiral states $|\mathbf{p}s\rangle$ and $|\mathbf{p}'s'\rangle$.

The expressions for conductivities (9) and (10) are the main results of this section. Though several limiting cases for these equations have been studied previously, the treatment of nonlocal effects with nonequilibrium population has been missing up to now. We first note that in the limit of large bias, $\Delta \gg \Omega$, Eq. (9) naturally yields the conductivity of a single graphene layer [33]. In the same limit, the factors $S_\pm$ and $\sin\theta_M$ are each proportional to the small tunneling exponent $e^{-k_b d}$, where k_b^{-1} is the decay length of the electron wave function. In the opposite limit $\Omega \lesssim \Delta$, the electron states of individual layers are highly mixed (the layer is not a good quantum number), and the distinction between in-plane and tunnel conductivities loses its meaning. In the ultimate case of zero bias, the antisymmetric plasmon field cannot induce the transitions between states of the same symmetry, which is reflected in the fact that $\cos\theta_M S_{++} = 0$. In this limit, $\sigma_\parallel = 0$, and the spectrum of acoustic plasmons is governed fully by the tunneling [21].

In the local limit, the tunnel conductivity becomes

$$G_\perp(0,\omega) = 2\frac{e^2}{2\hbar}S_\pm\sin\theta_M\hbar\omega[n_- - n_+] \times\left[\pi\delta(\tilde{\Delta}-\hbar\omega) - i\frac{2\tilde{\Delta}}{\tilde{\Delta}^2-(\hbar\omega)^2}\right], \tag{11}$$

where n_+ and n_- are the carrier densities in the states with $l = +1$ and $l = -1$. This result has been first obtained by Kazarinov and Suris in the theory of inelastic tunneling in superlattices [31] and later rederived in the case of tunneling in double-layer [26] and AA-stacked bilayer graphene [28]. In equilibrium, the low-energy symmetric state has larger population, $n_- > n_+$, the imaginary part of tunnel conductivity is negative at $\hbar\omega < 2\Omega$, while the real part is positive. The situation at nonequilibrium and at finite wave vector is radically different.

The first peculiarity of Eq. (10) in the presence of voltage bias is the negative value of the real part of tunnel conductivity at frequencies $\omega < eV/\hbar$. This negativity implies that the interlayer transitions accompanied by the emission of the quantum $(\omega,\mathbf{q})$ are more probable than the inverse absorptive transitions. The band filling providing the negative tunnel conductivity [Fig. 1(b)] can be viewed as an interlayer population inversion similar to that in quantum cascade lasers [34]. The frequency and wave vector dependence of $2\mathrm{Re}G_\perp/q^2$ is shown in Fig. 2(a), where the "cold" colors stand for the negative and "warm" colors for the positive conductivity. An analysis of the energy-momentum conservation reveals distinct regions on the frequency-wave vector plane, where different types of radiative and absorptive tunnel transitions are relevant; see Fig. 2(c). Among those, the most pronounced is the interlayer intraband emission allowed within the quadrant $qv_0 \geqslant |\tilde{\Delta}/\hbar - \omega|$. The interband transitions are generally weaker due to the small overlap of chiral wave functions of different bands [29].

Another distinct feature of the tunnel conductivity is its large absolute value near the series of lines

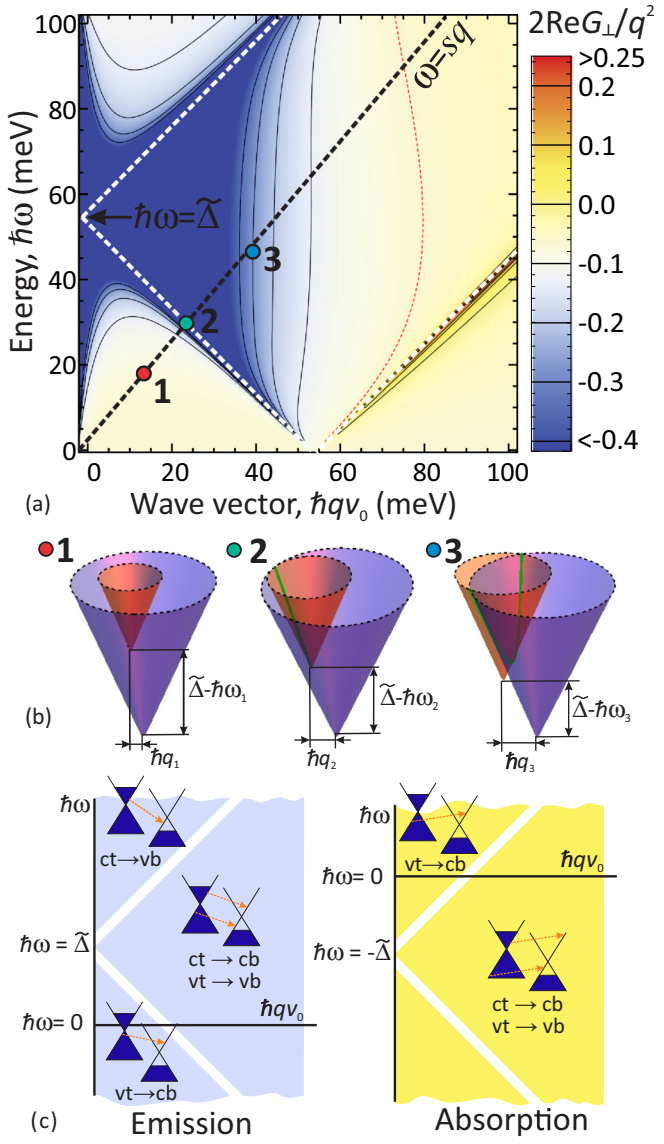

FIG. 2. (a) Color map of the tunnel conductivity, $2\mathrm{Re}G_\perp/q^2$ (in units of $e^2/4\hbar$), calculated at temperature $T = 77$ K and interlayer voltage $V = 0.2$ V. Dielectric layer is 3 nm WS_2 (effective mass $m^* = 0.28m_0$, conduction band offset to graphene $U_b = 0.4$ eV [35]). Red dashed line corresponds to the zero conductivity; black dashed line shows the dispersion of acoustic SP in the absence of tunneling. (b) Band diagrams illustrating available electron states for plasmon-assisted tunneling at different frequencies and wave vectors. (c) Map of the frequency and wave vector ranges, for which various inelastic interlayer electron transitions are possible.

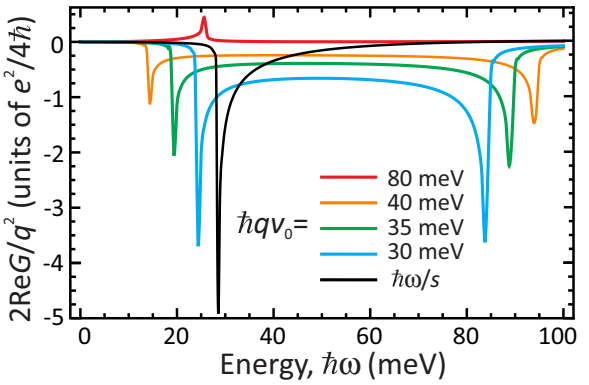

FIG. 3. Frequency dependence of the tunnel conductivity $\mathrm{Re}2G_\perp/q^2$ at certain wave vectors, i.e., cuts of the plot of Fig. 2 along $\hbar q v_0 = 80, 40, 35, 30$ meV, as well as along the unperturbed dispersion of acoustic SP $q = \omega/s$, where $s \approx 1.1v_0$. Weak singularity corresponding to the absorptive transition between like bands is seen in the red curve; at other wave vectors the transitions corresponding to tunneling emission are more pronounced.

$qv_0 = |\omega \pm \tilde{\Delta}/\hbar|$. The effect is demonstrated also in Fig. 3, where we show the frequency dependence of tunnel conductivity at certain wave vectors. The origin of these resonances can be explained by analyzing the possible electron states involved in plasmon-assisted tunneling at different frequencies and wave vectors, Fig. 2(b). To be precise, we focus on the interlayer intraband tunneling. Above the resonance, at $qv_0 > \omega - \tilde{\Delta}/\hbar$, the electrons capable of tunneling occupy a hyperbolic cut of the mass shell in graphene (case B3 in Fig. 2). With decreasing the frequency and wave vector, the hyperbola degenerates into a line (case B2), and the tunneling occurs between states with collinear momenta and equal velocities, whose interaction lasts for an infinitely long time in the absence of scattering. Alternatively, the singularities in the tunnel conductivity can be traced back to the van Hove singularities in the joint density of states [24]. At even lower frequencies (case B1), the intraband transitions are impossible, but the weaker interband tunneling sets in. Without the carrier scattering, the collinear tunneling singularities are square root, and the real part of the intraband tunnel conductivity is given by

$$\mathrm{Re}G_\perp^{\mathrm{intra}} = -\frac{e^2}{\hbar}\frac{q^2}{2\pi}\omega\left\{\frac{\mathcal{I}\left(\frac{\hbar q v_0}{2kT}, \frac{eV-\hbar\omega}{2kT}\right)}{\sqrt{q^2v_0^2 - (\tilde{\Delta}/\hbar - \omega)^2}}\right.$$
$$\left. - \frac{\mathcal{I}\left(\frac{\hbar q v_0}{2kT}, \frac{\hbar\omega+eV}{2kT}\right)}{\sqrt{q^2v_0^2 - (\tilde{\Delta}/\hbar + \omega)^2}}\right\}, \qquad (12)$$

where we have introduced an auxiliary integral $\mathcal{I}(\alpha,\beta) = \int_1^\infty dt\sqrt{t^2-1}[f_F(\alpha t - \beta) - f_F(\alpha t + \beta)]$ and the dimensionless "Fermi function" $f_F(\zeta) = [1 + e^\zeta]^{-1}$. The singularities in Eq. (12) are similar to the those of in-plane conductivity at the onset of the Landau damping [3]:

$$\mathrm{Re}\sigma_\parallel^{\mathrm{intra}} \propto \left[q^2v_0^2 - \omega^2\right]^{-1/2}. \qquad (13)$$

The actual value of the resonant conductivity in clean samples is limited by electron-acoustic phonon scattering [9]. We account for it by replacing the delta-peaked spectral functions of individual electrons in Eqs. (9) and (10) with Lorentz functions whose width is proportional to the imaginary part of electron self-energy [24]. With the scattering rate $\tau_{\mathrm{tr}}^{-1} \simeq (2\text{–}8) \times 10^{-11}$ s^{-1} at $T = 77\text{–}300$ K [36,37] and electron density $n = 5 \times 10^{11}$ cm^{-2}, the tunnel resonances are pronounced up to the room temperature.

From a practical point of view, it is important that the real part of net "effective conductivity," $\sigma_\parallel + 2G_\perp/q^2$, which enters the SP dispersion is negative in a wide range of

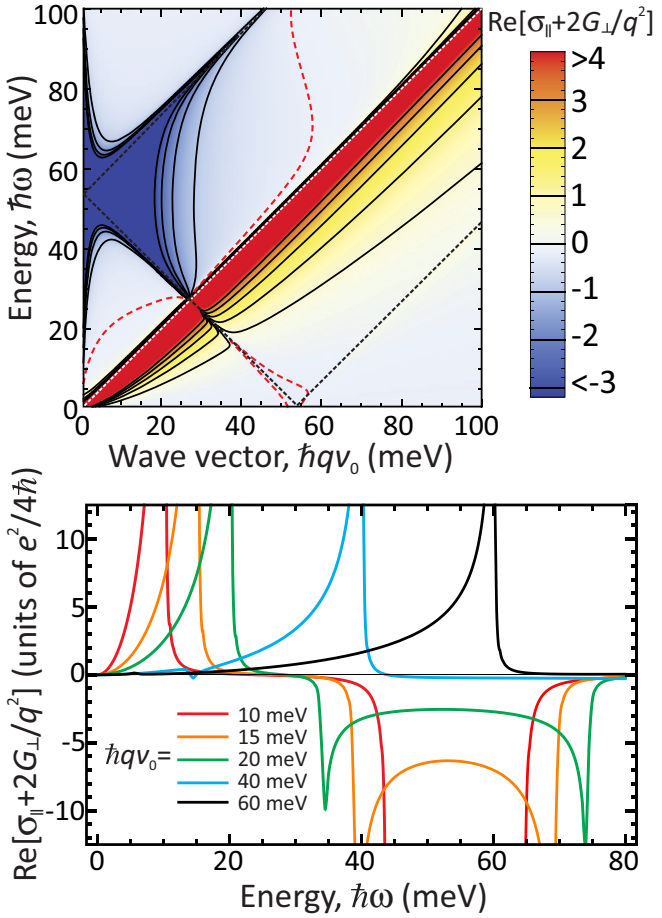

FIG. 4. Space-time dispersion of the effective conductivity $\mathrm{Re}[\sigma_\parallel + 2G_\perp/q^2]$ (normalized by $e^2/4\hbar$) governing the damping (or gain) of acoustic SPs. The structure parameters are the same as in Fig. 2. The contour of zero conductivity is highlighted with red dashed line. The bottom panel shows the cuts of the upper plot along certain wave vectors.

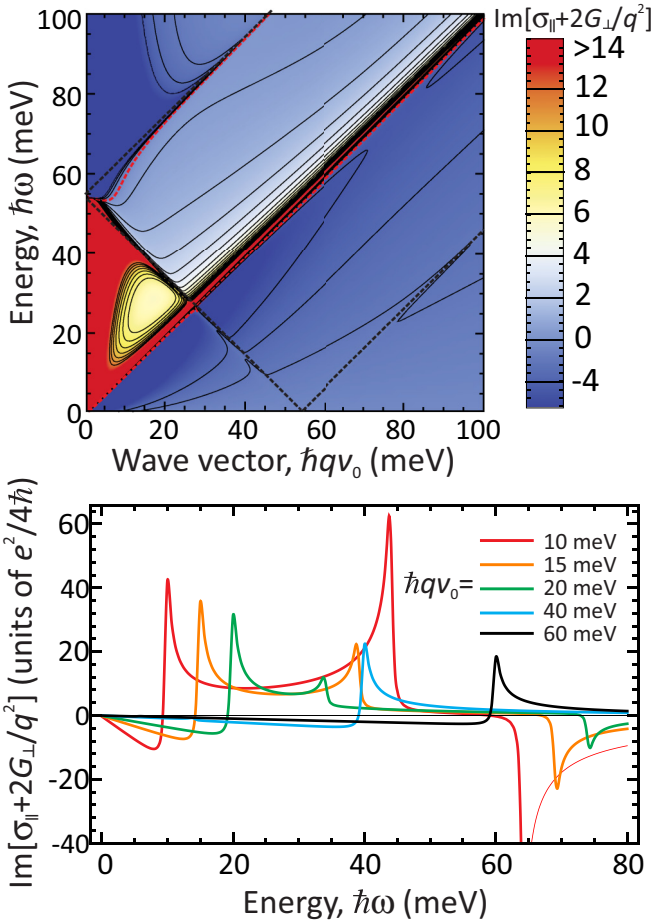

FIG. 5. Space-time dispersion of the effective conductivity $\mathrm{Im}[\sigma_\parallel + 2G_\perp/q^2]$ (normalized by $e^2/4\hbar$) governing the spectrum renormalization of acoustic SPs. The structure parameters are the same as in Fig. 2. The contour of zero conductivity is highlighted with red dashed line. The bottom panel shows the cuts of the upper plot along certain wave vectors.

frequencies and wave vectors (Fig. 4). This implies the possibility of amplified propagation plasmons instead of their damping. The sign of the net effective conductivity is governed by the competition of three processes: (1) emission of surface plasmons upon interlayer tunneling, (2) interband plasmon absorption, and (3) intraband (free-carrier) plasmon absorption. The latter two processes are intensified due to the finiteness of the SP wave vector, and use of the well-known local limit for the in-plane conductivity would lead to a considerable underestimate of the absorption (see Ref. [9] and Appendix C for details). Our numerical calculations show that the net negative conductivity is still possible as the collinear tunneling resonance can be tuned into the frequency range corresponding to the transparency of graphene, i.e., the frequencies where the Drude absorption is low while the interband absorption is suppressed by the Pauli blocking.

The square-root singularities in $\mathrm{Re}G_\perp^{\mathrm{intra}}$ above the threshold of interlayer intraband transitions are mirrored in the singularities in $\mathrm{Im}G_\perp^{\mathrm{intra}}$ below the threshold. This situation is illustrated in Fig. 5, where we show the color map of the imaginary part of full "effective" conductivity, $\mathrm{Im}[2G_\perp/q^2 + \sigma_\parallel]$, and its cuts at certain wave vectors. The presence of singularity in the imaginary part of tunnel conductivity at $\omega = \tilde{\Delta}/\hbar - qv_0$ suggests the strong renormalization of the acoustic SP dispersion, which will be shown in the next section. We note here that the mentioned singularities are absent in the case of tunneling between the layers of massive 2d electrons. In the massive case, the electron states with collinear momenta do not necessarily have the same velocity, and their interaction does not necessarily last for a very long time. Loosely speaking, for massive particles one should average the tunnel conductivity (12) over the absolute values of particle velocity v_0, which would wash the singularities out.

IV. PLASMONS IN THE PRESENCE OF TUNNELING

The long-wavelength part of the acoustic SP mode in coupled graphene layers at equilibrium does not qualitatively differ from that for coupled massive 2d electrons. The SP dispersion develops a gap, such that

$$(\hbar\omega_{\mathrm{ac}})^2(q \to 0) = (2\Omega)^2 + 2\Omega E_C, \qquad (14)$$

where $E_C = 8\pi g d e^2 (n_- - n_+)/(\hbar\kappa)$. It is possible to show that the formation of such gap is generic for coupled layers, once the in-plane dispersion of electrons is not renormalized by tunneling [38]. Indeed, the real part of tunnel conductivity is positive for the photon energy equal to the tunnel splitting of levels, 2Ω. Using the Kramers-Kronig relations, one obtains that the imaginary part of conductivity is negative at $\hbar\omega < 2\Omega$, which prohibits the existence of TM plasmons. Apart from 2Ω, an extra contribution to the plasmon gap in Eq. (14) comes from the electrostatic energy of electrons, E_C.

At finite bias, both the tunneling and in-plane electron motion govern the plasmon spectrum. The imaginary part of in-plane conductivity $\mathrm{Im}\sigma_\parallel$ is positive and singular above the domain of Landau damping, i.e., at $\omega \to q v_0$. Such behavior originating from the linearity of carrier dispersion leads to the existence of acoustic mode even in the limit $d \to 0$ [4,39], contrary to the case of parallel 2d layers of massive carriers in the absence of tunneling [40]. Moreover, the imaginary part of tunnel conductivity at nonequilibrium is also positive below the threshold of interlayer transitions. Hence, already at a very small bias, the condition $\mathrm{Im}[\sigma_\parallel + 2G_\perp/q^2] > 0$ can be fulfilled, and long-wavelength linear dispersion of SPs is restored. At higher frequencies, the tunneling resonance crosses the unperturbed SP dispersion $\omega = sq$. At this point, the effect of tunneling is the strongest one.

This is illustrated in the plot of the acoustic mode contribution to the loss function, $\mathrm{Im}\epsilon_{\mathrm{ac}}^{-1}(\mathbf{q},\omega)$, Fig. 6(a). The plasmon peak in the loss function develops an anticrossing with the tunnel resonance, and the bending of the SP dispersion is such that its group velocity increases and changes its sign at some point, passing through infinity. The "locking" of the long-wavelength SP dispersion in the domain $(\omega \geqslant q v_0) \cup (\omega \leqslant \tilde{\Delta}/\hbar - q v_0)$ comes, formally, from the interplay of two singular conductivities, in-plane and out-of-plane, which are positive and singularly large at the threshold of the respective transitions (see Fig. 5).

The interaction of unperturbed SP mode with tunnel resonance leading to abnormally large and negative group velocities is radically different from the interaction of plasmon modes with polar phonons [8], cyclotron resonances [41], etc. In the two latter cases, the interaction leads to the decrease in group velocity. The origin of the upward bending of SP dispersion is, actually, the inverted population between two layers [42]. Similar group velocity enhancement and divergence occurs for light interacting with a gas of inverted two-level systems [43], and can be analyzed with a simple Lorentz oscillator model [44].

The effects of tunneling on SP dispersion are more pronounced at low levels' spacing $\tilde{\Delta}$ and high carrier density n. By fixing $\tilde{\Delta}$ and increasing the carrier density (which can be achieved by extra gates), one can observe a large enhancement of the SP velocity below the resonance, as shown in Fig. 6(b). At some critical density the long-wavelength branch of SP dispersion can disappear entirely. At large level spacing, the effects of plasmon spectrum renormalization are relevant just in a narrow vicinity of the threshold of intraband tunneling, and further broadening of the tunnel resonance by the carrier scattering can wash out the renormalization effects. However, our calculations show that the predicted effects of tunneling on plasmons can survive up to the room temperature and

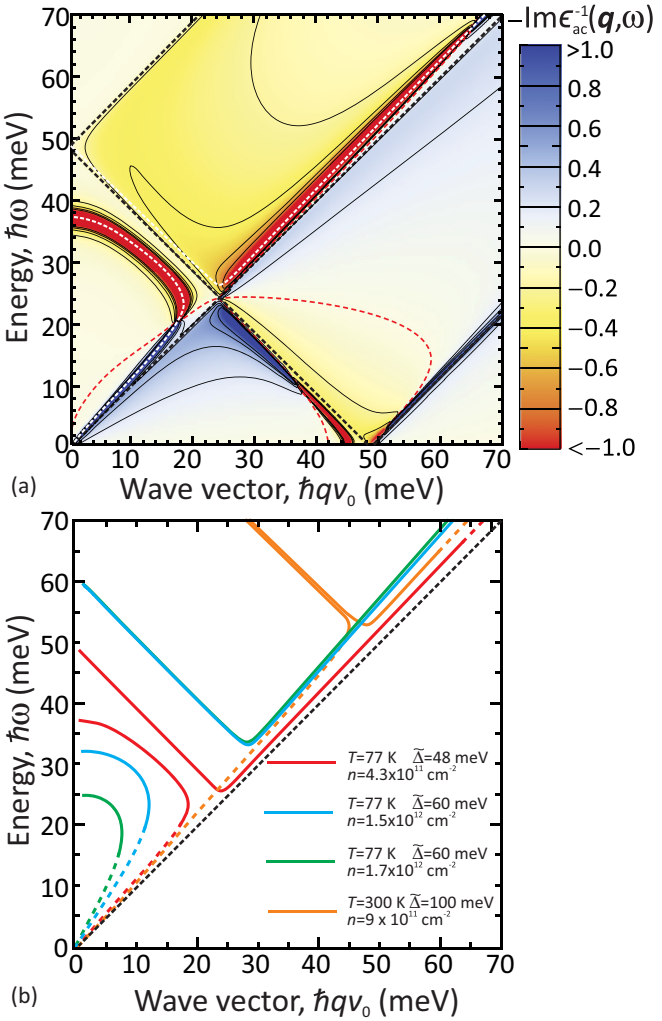

FIG. 6. (a) Acoustic contribution to the loss function of double graphene layer calculated for 2.5 nm WS_2 barrier, $T = 77$ K and $V = 0.2$ V. The plasmon spectrum develops an anticrossing with the collinear tunneling resonance. (b) Plasmon spectra calculated for different temperatures, electron densities, and level spacing of tunnel coupled layers $\tilde{\Delta}$. Dashed parts of the spectra correspond to the damped plasmons and solid parts to the amplified plasmons. Black dashed line is $\hbar\omega = q v_0$.

relatively high interlayer voltage. The plasmon dispersion at $T = 300$ K and $V = 0.3$ V shown by the orange line in Fig. 6(b) still demonstrates the backward-wave behavior as well as the possibility of plasmon gain.

The net gain of surface plasmons (i.e., negative real part of the net effective conductivity $\sigma_\parallel + 2G_\perp/q^2$) is possible both below and above the tunnel resonance at $\omega = \tilde{\Delta}/\hbar - q v_0$. Below the resonance, the gain is due to the emission of SPs upon tunneling from the valence band of top layer to the conduction band of the bottom one [see Fig. 2(c)]. In this domain, the renormalization of SP dispersion leads to the reduction of the wave vector at given frequency, and hence, to the reduction of in-plane loss. Above the resonance, the gain is due to the emission of SPs upon interlayer intraband transitions; the linear dispersion of SPs is almost unaffected by tunneling in this case.

V. EFFECTS OF INTERLAYER TWIST

The effects of rotational twist of graphene layers in the tunneling heterostructures manifests itself in the relative shift of their Dirac points by the vectors $\Delta\mathbf{q}_i$ in the reciprocal space ($i = 1 \ldots 6$; see the inset in Fig. 5). With the neglect of small off-diagonal elements of the $\mathcal{T}$ matrix, the tunnel conductivity of twisted layers $G_\perp^T(\mathbf{q},\omega)$ is related to the tunnel conductivity of the aligned layers $G_\perp$ via

$$G_\perp^T(\mathbf{q},\omega) = \frac{1}{6}\sum_{i=1}^{6} G_\perp(\mathbf{q}+\Delta\mathbf{q}_i,\omega). \tag{15}$$

In the presence of twist, the locus of collinear scattering singularities on the ω-q plane breaks down into six hyperbolas (or less, for particular angles between $\mathbf{q}$ and $\Delta\mathbf{q}$). The acoustic plasmon dispersion develops an anticrossing with each of the hyperbolas, demonstrating several frequency ranges with negative group velocity and gain. An example of the loss function for the wave vector parallel to the twist vector $\Delta\mathbf{q}$ in one pair of valleys is shown in Fig. 7 for $\hbar|\Delta\mathbf{q}|v_0 = 18$ meV (twist angle $\theta_T = 0.57^\circ$). In this example, there exist four curves corresponding to the singular plasmon gain and four for the singular absorption. Remarkably, the plasmon gain in twisted layers for certain directions of propagation can be greater than that in aligned layers, because the tunnel resonances can come closer to the unperturbed SP dispersion. Generally, the spectrum of plasmons in twisted layers becomes anisotropic with sixfold rotational symmetry.

VI. DISCUSSION AND CONCLUSIONS

The renormalization of SP dispersion due to the tunneling can be observed with the recently introduced tool of scattering-type near-field microscopy [45]. A more practical and readily observable effect of tunneling is the enhancement of SP propagation length due to the partial compensation of in-plane absorption by the tunneling emission. Moreover, our calculations show that under proper choice of the barrier layer (2 . . . 3 nm WS_2), the full compensation of plasmon losses and the net plasmon gain are possible in the far-infrared range $\hbar\omega \approx 20 \ldots 60$ meV. The predicted effect of the net gain opens up the possibility of creating graphene-based sources of coherent plasmons—spasers—and nanoscale sources of photons based, e.g., on the evanescent coupling of photonic waveguides and graphene-based active media for plasmons.

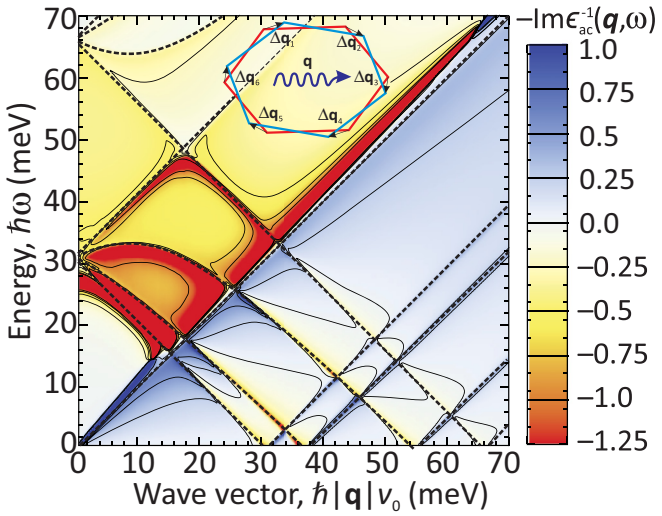

FIG. 7. Acoustic contribution to the loss function in twisted layers (twist angle $\theta_T = 0.57^\circ$); the wave vector is parallel to the misalignment vector in one pair of valleys. Inset shows the positions of K points in the reciprocal space for twisted graphene layers.

The observation of net plasmon gain in coupled graphene layers poses strong constraints on the tunnel transparency of the barrier material. At the same time, the spontaneous emission of SPs upon tunneling [46,47] is observable for a wide class of dielectrics. The tunneling SP emission with their subsequent conversion into free-space electromagnetic modes upon scattering might explain the terahertz electroluminescence from graphene-hBN-graphene diodes observed recently [48]. The presence of luminescence in [48] correlates with the magnitude of NDR in the static $I(V)$ curve supporting the tunneling origin of the emission.

The tunneling assisted by the spontaneous emission of plasmons can also manifest itself as an extra peak in the bias-dependent differential conductivity dG/dV [49]. Such peaks have not yet been identified with tunneling spectroscopy of graphene-based structures [50]. The extension of the presented calculation of tunnel current to the case of spontaneous emission of surface plasmons would reveal the favorable conditions for plasmon-assisted resonant tunneling and guide further experimental work.

In conclusion, we have theoretically demonstrated a number of unique properties of surface plasmons in tunnel-coupled voltage-biased graphene layers, including the amplified propagation due to the resonant tunneling under interlayer population inversion, and a strong renormalization of dispersion law. The pronounced effect of tunneling on both spectrum and damping of plasmons results from singularities in the tunnel conductivity which are, in turn, inherited from the linear bands of graphene.

ACKNOWLEDGMENTS

The work of D. S. was supported by Grant No. 16-19-10557 of the Russian Science Foundation. The work at RIEC was supported by Japan Society for Promotion of Science KAKENHI Grants No. 23000008 and No. 16H06361. The authors thank V. Vyurkov, S. Fillipov, A. Dubinov, A. Arsenin, and D. Fedyanin for helpful discussions.

APPENDIX A: PLASMON MODES SUPPORTED BY THE DOUBLE LAYER

In this section, we review the general properties of plasmons in double-graphene-layer structures [22,51] in the absence of tunneling. The acoustic SP dispersion in double layers is equivalent to the plasmon dispersion in graphene with a perfectly conducting gate [4,39]. The low-frequency part of the SP spectrum can be obtained with the neglect of interband conductivity, while the intraband conductivity can

be calculated in the Boltzmann limit ($\hbar q v_0 \ll \varepsilon_F$, $\hbar\omega \ll \varepsilon_F$):

$$\sigma_{\parallel}^{\text{intra}} \approx ig\frac{e^2}{\hbar}\frac{\tilde{\varepsilon}_F}{2\pi\hbar}\frac{\omega}{q^2 v_0^2}\left[\frac{\omega}{\sqrt{\omega^2 - q^2 v_0^2}} - 1\right], \tag{A1}$$

where $\tilde{\varepsilon}_F = T\ln(1 + e^{\varepsilon_F/T})$. The dispersion relations of plasmons (3) in the absence of tunneling ($G_\perp = 0$) can be solved analytically with conductivity given by Eq. (A1). This yields the linear dispersion of the acoustic mode

$$\omega_{\text{ac}} = v_0\frac{1 + 4\alpha_c q_F d}{\sqrt{1 + 8\alpha_c q_F d}}q, \tag{A2}$$

and the square-root dispersion of the optical mode

$$\omega_{\text{opt}} \approx v_0\sqrt{4\alpha_c q q_F}. \tag{A3}$$

Here, we have introduced the Fermi wave vector $q_F = \tilde{\varepsilon}_F/\hbar v_0$, the coupling constant $\alpha_C = e^2/\hbar\kappa v_0$, and neglected the spatial dispersion of conductivity for the optical mode. This is possible as the phase velocity of acoustic SPs greatly exceeds the Fermi velocity. On the other hand, the velocity of the acoustic mode always exceeds the Fermi velocity, but can be arbitrarily close to it. For this reason, the spatial dispersion of conductivity cannot be neglected for acoustic modes even in the formal long-wavelength limit $q \ll q_F$, $q \ll T/\hbar v_0$ as the inequality $q v_0 \ll \omega$ cannot be generally fulfilled.

The distinct dispersions of optical and acoustic plasmon modes are seen from the loss functions shown in Fig. 8. The acoustic part of the loss function is given by Eq. (4), while the optical part is given by

$$\text{Im}\epsilon_{\text{opt}}^{1}(\mathbf{q},\omega) = \text{Im}\left\{1 + \frac{2\pi i q\sigma_{\parallel}}{\omega\kappa}(1 + e^{-qd})\right\}^{-1}. \tag{A4}$$

To account for the plasmon damping one has to go beyond Eq. (A1) and calculate the real parts of the in-plane conductivity (see Appendix C for analytical expressions). As seen from Fig. 8, the plasmon peak in the spectral function becomes broadened at $\hbar\omega \approx 2\varepsilon_F - \hbar q v_0$, which corresponds to the onset of interband transitions. Below the threshold of interband transitions, the plasmon damping is governed by the free-carrier absorption which can be calculated from the Boltzmann equation with electron-phonon scattering taken into account.

The spatial distribution of the plasmon potential $\delta\varphi(z)$ in acoustic mode is governed solely by the wave vector q and can be obtained from the Poisson's equation (2). It is convenient to present it as

$$\delta\varphi(z) = \delta\varphi_0 s(z), \tag{A5}$$

where φ_0 is the electric potential on the top layer, and $s(z)$ is the dimensionless "shape function" having the following form:

$$s(z) = \begin{cases} e^{-q(z+d/2)}, & z < -d/2, \\ -\frac{\sinh(qz)}{\sinh(qd/2)}, & |z| < d/2, \\ -e^{-q(z-d/2)}, & z > d/2. \end{cases} \tag{A6}$$

The spatial dependence of the shape functions for acoustic and optical modes is shown in the inset of Fig. 8.

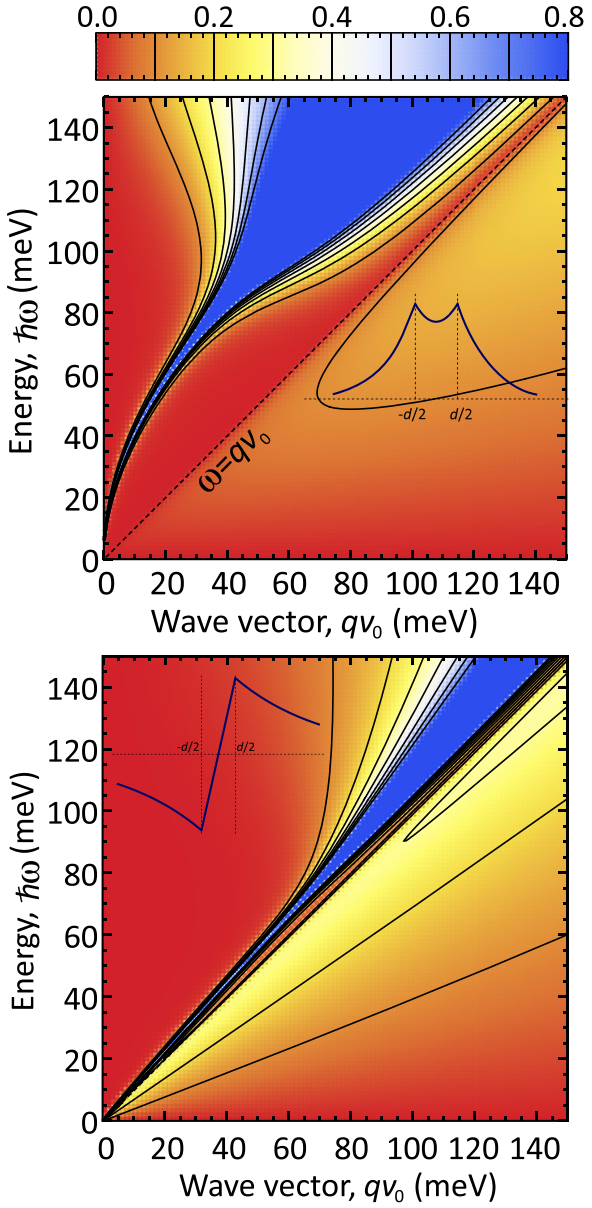

FIG. 8. Optical (top) and acoustic (bottom) contributions to the loss function for double graphene layer in the absence of tunneling. Fermi energy $\varepsilon_F = 100$ meV, temperature $T = 300$ K, insulator thickness $d = 3$ nm, dielectric constant $\kappa = 5$.

APPENDIX B: ESTIMATE OF THE TIGHT-BINDING PARAMETERS

To estimate the tight-binding parameters Ω and $S_{ll'}$, we switch to the continuum description of electron states in the z direction. We model each graphene layer with a delta well [52]

$$U_{t,b}(z) = 2\sqrt{\frac{\hbar^2 U_b}{2m^*}}\delta(z - z_{t,b}), \tag{B1}$$

where the potential strength is chosen to provide a correct value of electron work function U_b from graphene to the surrounding dielectric, and m^* is the effective electron mass in the dielectric. The effective Schrödinger equation in the presence of voltage bias Δ/e between graphene layers takes

on the following form:

$$-\frac{\hbar^2}{2m^*}\frac{\partial^2\Psi(z)}{\partial z^2}+[U_t(z)+U_b(z)+U_F(z)]\Psi(z)=E\Psi(z), \tag{B2}$$

where U_F is the potential energy created by the applied field

$$U_F(z)=\frac{\Delta}{2}\begin{cases}1, & z<-d/2,\\ 2z/d, & |z|<d/2,\\ -1, & z>d/2.\end{cases} \tag{B3}$$

The solutions of the effective Schrödinger equation represent decaying exponents at $|z|>d/2$, and a linear combination of Airy functions in the middle region $|z|<d/2$:

$$\Psi_M(z)=C\mathrm{Ai}(-z/a+\varepsilon)+D\mathrm{Bi}(-z/a+\varepsilon), \tag{B4}$$

where $\varepsilon=2m^*|E|a^2/\hbar^2$ is the dimensionless energy and $a=(\hbar^2 d/2m^*\Delta)^{1/3}$ is the effective length in the electric field. A straightforward matching of the wave functions at the graphene layers yields the dispersion equation

$$\det\begin{pmatrix} e^{-k_1 d/2} & -\mathrm{Ai}(d/2a+\varepsilon) & -\mathrm{Bi}(d/2a+\varepsilon) & 0\\ (2k_b-k_1)e^{-k_1 d/2} & -\frac{1}{a}\mathrm{Ai}'(d/2a+\varepsilon) & -\frac{1}{a}\mathrm{Bi}'(d/2a+\varepsilon) & 0\\ 0 & -\mathrm{Ai}(-d/2a+\varepsilon) & -\mathrm{Bi}(-d/2a+\varepsilon) & e^{-k_2 d/2}\\ 0 & -\frac{1}{a}\mathrm{Ai}'(-d/2a+\varepsilon) & -\frac{1}{a}\mathrm{Bi}'(-d/2a+\varepsilon) & (2k_b-k_2)e^{-k_2 d/2}\end{pmatrix}=0, \tag{B5}$$

where $k_b=\sqrt{2m^*U_b/\hbar^2}$ is the decay constant of the bound state wave function in a single delta well, $k_1=\sqrt{2m^*(E+\Delta/2)/\hbar^2}$, $k_2=\sqrt{2m^*(E-\Delta/2)/\hbar^2}$. Equation (B5) yields two energy levels E_l ($l=\pm1$) which can be found only numerically [see Fig. 9(a)]. The respective wave functions are shown in Fig. 9(b); at strong bias they are *almost* the wave functions localized on the different layers (see the discussion below). Despite the complexity of Eq. (B5), the dependence of E_l on the energy separation between layers Δ can be accurately modeled by

$$E_l(\Delta)=-U_b+\frac{l}{2}\sqrt{(E_{+1,\Delta=0}-E_{-1,\Delta=0})^2+\Delta^2}. \tag{B6}$$

The same functional dependence of energy levels on Δ is naturally obtained by diagonalizing the block Hamiltonian (5),

$$E_l(\Delta)=-U_b+l\sqrt{\Omega^2+\frac{\Delta^2}{4}}. \tag{B7}$$

This allows us to estimate the tunnel coupling Ω as half the energy splitting of states in the double-graphene-layer well in the absence of applied bias:

$$\Omega=\tfrac{1}{2}[E_{+1,\Delta=0}-E_{-1,\Delta=0}]. \tag{B8}$$

The wave functions corresponding to a relatively strong bias $\Delta=200$ meV are shown in Fig. 9. It is simple to relate the true eigenfunctions $\Psi_+(z)$ and $\Psi_-(z)$ to the functions located on the top and bottom layers $\Psi_t(z)$ and $\Psi_b(z)$:

$$\Psi_b=\cos\alpha\Psi_-+\sin\alpha\Psi_+, \tag{B9}$$

$$\Psi_t=-\sin\alpha\Psi_-+\cos\alpha\Psi_+, \tag{B10}$$

where

$$\cos\alpha=\frac{2\Omega}{\sqrt{(2\Omega)^2+(\Delta-\tilde{\Delta})^2}}. \tag{B11}$$

Knowing the wave functions of the coupled layers, we can estimate the matrix elements of electron-plasmon interaction and present them in the following form:

$$\langle \mathbf{p}sl|\,\delta\hat{V}\,|\mathbf{p}'s'l'\rangle=\delta_{\mathbf{p},\mathbf{p}'-\mathbf{q}}u^{ss'}_{\mathbf{p}\mathbf{p}'}e\delta\varphi_0 S_{ll'}. \tag{B12}$$

Here we have introduced the shorthand notations for the overlap factors of dimensionless plasmon potential and eigenfunctions of coupled layers,

$$S_{++}=\int_{-\infty}^{\infty}\Psi^*_{+1}(z)s(z)\Psi_{+1}(z), \tag{B13}$$

$$S_{\pm}=\int_{-\infty}^{\infty}\Psi^*_{+1}(z)s(z)\Psi_{-1}(z), \tag{B14}$$

and, obviously, $S_{--}=-S_{++}$.

The dependence of the overlap factors S_{++} and $S_{\pm}$ on the interlayer potential drop Δ is shown in Fig. 10. We note that these overlap factors weakly depend on the plasmon wave vector q as far as it is much smaller than electron wave function decay constant k_b. In this approximation, one can set $s(z)\approx 2z/d$ for $|z|<d/2$, $s(z)\approx 1$ at $|z|>d/2$.

APPENDIX C: ANALYTICAL RESULTS FOR THE CONDUCTIVITY: IN-PLANE CONDUCTIVITY

Despite the complex structure of Eqs. (9) and (10), several analytical approximations can be made in the frequency range of interest $\hbar\omega<2\varepsilon_F$, where the plasmons are weakly damped—at least, for the real part of conductivity that determines absorption or gain. For brevity, in this section we work with "God-given units" $\hbar=v_0\equiv1$. We start with the evaluation of in-plane interband conductivity [the term with $s=v$ and $s'=c$ in Eq. (9)].

The interband velocity matrix element reads $\langle c\mathbf{p}_-|\,\hat{v}_x\,|v\mathbf{p}_+\rangle=i\sin[(\theta_{\mathbf{p}+}+\theta_{\mathbf{p}-})/2]$, where $\theta_{\mathbf{p}-}$ and $\theta_{\mathbf{p}+}$ are the angles between the momenta of initial and final states with the x axis. The subsequent calculations are conveniently performed in the elliptic coordinates

$$\mathbf{p}=\frac{q}{2}\{\cosh u\cos v,\sinh u\sin v\}. \tag{C1}$$

In these coordinates $|\mathbf{p}_\pm|=(q/2)[\cosh u\pm\cos v]$, $|\langle c\mathbf{p}_-|\hat{v}_x|v\mathbf{p}_+\rangle|^2dp_xdp_y=(q^2/4)\cosh^2u\sin^2v\,du\,dv$. This

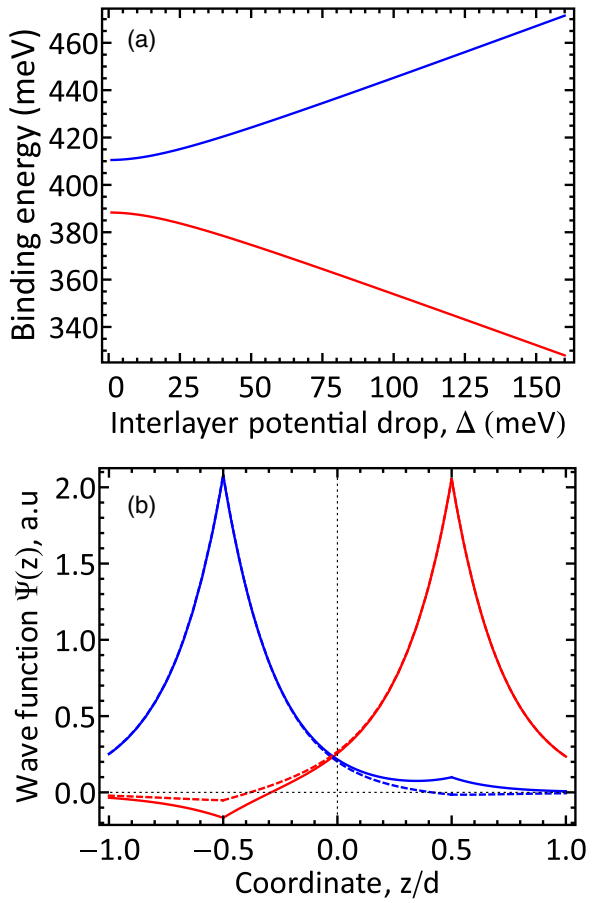

FIG. 9. (a) Dependence of energy levels in coupled graphene layers on the interlayer potential drop Δ. (b) Wave functions of the coupled layers calculated for $\Delta = 200$ meV. The barrier is 2.5 nm WS_2. Solid lines in (b) show the wave functions corresponding to $l = +1$ (red) and $l = -1$ (blue), while the dashed lines show the wave functions of the top and bottom layers obtained as a linear combination (B9) of the eigenfunctions.

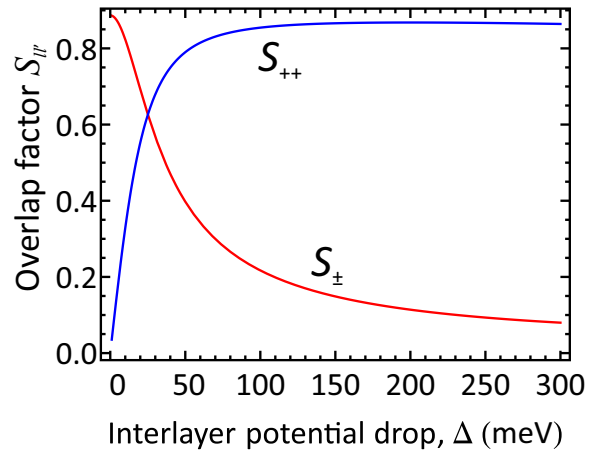

FIG. 10. Dependence of the overlap factors S_{++} and $S_{\pm}$ on the interlayer voltage drop Δ calculated for the 2.5 nm WS_2 dielectric layer.

leads us to (we henceforth omit the factor $S_{++}\cos\theta_M$)

$$\mathrm{Re}\sigma_{\parallel}^{v\to c} = \frac{e^2}{2\pi}\frac{\omega}{\sqrt{\omega^2-q^2}}\int_0^{\pi} dv\sin^2 v \times\left\{f^v\left[-\frac{\omega}{2}+\frac{q}{2}\cos v\right]-f^c\left[\frac{\omega}{2}+\frac{q}{2}\cos v\right]\right\}. \tag{C2}$$

To proceed further, we note that in the domain of interest $\omega > q$ one always has $q\cos v < \omega$. Due to this fact, the difference of distribution functions is a smooth function of v, while the prefactor $\sin^2 v$ varies strongly. This allows us to integrate $\sin^2 v$ exactly, and replace the difference of distribution functions with its angular average. This leads us to

$$\mathrm{Re}\sigma^{v\to c} \approx \frac{e^2}{4}\frac{T\omega}{q}\chi(q,\omega)\ln\frac{\cosh\frac{\varepsilon_F}{T}+\cosh\frac{\omega+q}{2T}}{\cosh\frac{\varepsilon_F}{T}+\cosh\frac{\omega-q}{2T}}, \tag{C3}$$

where we have introduced a resonant factor

$$\chi(q,\omega) = \frac{\theta(\omega)}{\sqrt{\omega^2-q^2}}. \tag{C4}$$

Clearly, the neglect of spatial dispersion in the case of acoustic SPs with velocity slightly exceeding the Fermi velocity results in an underestimation of the real part of the interband conductivity and, hence, of the plasmon damping.

We now pass to the in-plane conductivity associated with the intraband transitions. Here, we can restrict ourselves to the classical description of the electron motion justified at frequencies $\omega \ll \varepsilon_F$, $q \ll q_F$; otherwise, strong interband SP damping takes place. Clearly, one could work out the terms with $s = s'$ and $l = l'$ in Eq. (9); however, the accurate inclusion of carrier scattering in such equations is challenging. Instead, we use the kinetic equation to evaluate $\sigma_{\parallel}^{c\to c}$; this formalism allows an inclusion of carrier scattering in a consistent manner. One should, however, keep in mind that in the nonlocal case $q \neq 0$ a simple τ_p approximation is not particle conserving. A particle-conserving account of collisions is achieved with the Bhatnagar-Gross-Krook collision integral [53] in the right-hand side of the kinetic equation,

$$-i\omega\delta f(\mathbf{p}) + i\mathbf{q}\mathbf{v}\delta f(\mathbf{p}) + ie\mathbf{q}\mathbf{v}\delta\varphi\frac{\partial f_0}{\partial\varepsilon} = -\nu\left[\delta f(\mathbf{p}) + \frac{d\varepsilon_F}{dn}\frac{\partial f_0}{\partial\varepsilon}\delta n\right]. \tag{C5}$$

Here $\delta f(\mathbf{p})$ is the sought-for field-dependent correction to the equilibrium electron distribution function f_0, $\delta n_{\mathbf{q}}$ is the respective correction to the electron density, $\mathbf{v} = \mathbf{p}/p$ is the quasiparticle velocity, and ν is the electron collision frequency which is assumed to be energy-independent. Solving Eq. (C5) and recalling the continuity equation $\omega\delta n = q\delta\mathbf{j}$ we obtain the in-plane intraband conductivity:

$$\sigma_{\parallel}^{\mathrm{intra}} = \frac{ige^2\tilde{\varepsilon}_F}{(2\pi)^2 q}\frac{J_2\left(\frac{\omega+i\nu}{q}\right)}{1-\frac{i\nu}{2\pi\omega}J_1\left(\frac{\omega+i\nu}{q}\right)}, \tag{C6}$$

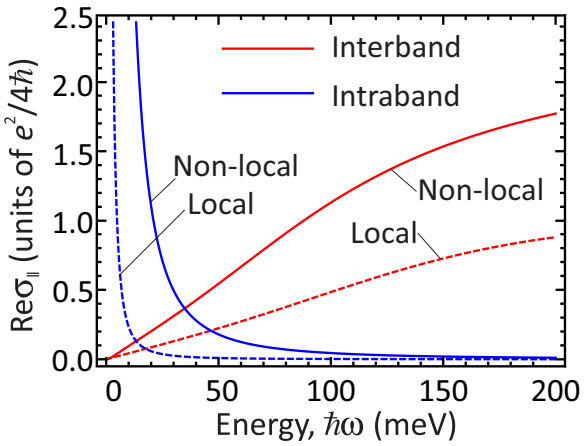

FIG. 11. Comparison of the real parts of the interband (red) and intraband (blue) conductivities of a single graphene layer evaluated in the local limit (dashed) and at finite wave vector corresponding to the acoustic SP dispersion $q = \omega/s$ (solid). The parameters used in the calculation are $\varepsilon_F = 100$ meV, $T = 300$ K, $s = 1.2v_0$. Acoustic phonons are considered as the main carrier relaxation mechanism.

where

$$J_n(x) = \int_0^{2\pi} \frac{\cos^n\theta d\theta}{x - \cos\theta}. \tag{C7}$$

Similarly to the real part of the interband absorption, the intraband absorption is generally larger in the nonlocal case $q \neq 0$ compared to the local case. This difference is illustrated in Fig. 11, where the local ($q = 0$) and nonlocal expressions at the acoustic plasmon dispersion ($q = \omega/s$) are compared. This result is in agreement with the recent measurements of plasmon propagation length in graphene on hBN: the local Drude formula underestimated the plasmon damping, and the account of nonlocality was crucial to explain the experimental data [9].

APPENDIX D: TUNNEL CONDUCTIVITY

Approximations similar to those used in deriving Eq. (C3) can be made to evaluate the interlayer interband conductivity. The only difference is that electrons in different layers have different chemical potentials. We present these results without derivation:

$$\frac{2G_\perp^{v\to c}}{q^2} = -e^2\frac{T\omega}{2q}\left\{\chi(q,\tilde{\Delta}-\omega)\ln\frac{\cosh\frac{q+eV-\omega}{4T}}{\cosh\frac{q-eV+\omega}{4T}} - \chi(q,\tilde{\Delta}+\omega)\ln\frac{\cosh\frac{q+eV+\omega}{4T}}{\cosh\frac{q-eV-\omega}{4T}}\right\}, \tag{D1}$$

$$\frac{2G_\perp^{c\to v}}{q^2} = -e^2\frac{T\omega}{2q}\left\{\chi(q,\omega-\tilde{\Delta})\ln\frac{\cosh\frac{q+eV-\omega}{4T}}{\cosh\frac{q-eV+\omega}{4T}} - \chi(q,-\tilde{\Delta}-\omega)\ln\frac{\cosh\frac{q+eV+\omega}{4T}}{\cosh\frac{q-eV-\omega}{4T}}\right\}. \tag{D2}$$

An analytical estimate of the tunnel conductivity associated with intraband transitions is possible only in the limit $\varepsilon_F \gg T$, which is not always the case in our calculations. Again, passing to the elliptic coordinates in Eq. (10) and keeping the terms with $s = s'$ one readily finds the result of Eq. (12). The value of the auxiliary integral at zero temperature is

$$\mathcal{I}(\alpha,\beta) = \frac{\theta(\beta)}{2}\left[\frac{\beta}{\alpha}\sqrt{\frac{\beta^2}{\alpha^2}-1} - \ln\left(\frac{\beta}{\alpha}+\sqrt{\frac{\beta^2}{\alpha^2}-1}\right)\right] - \{\beta\to-\beta\}. \tag{D3}$$

Finally, to obtain physically reasonable results one has to estimate the actual value of the resonant conductivity in the presence of carrier scattering. The result for in-plane intraband conductivity is given essentially by Eq. (C6). For all other terms, the classical approximation used in (C6) is invalid. To account for the scattering in those terms, we replace the delta-peaked spectral functions of individual particles in the expressions for conductivity with Lorentz-type functions using the following rule:

$$\sum_{\mathbf{p}} \frac{1}{\omega + i\delta - \left(\varepsilon_{\mathbf{p}}^{sl} - \varepsilon_{\mathbf{p}'}^{s'l'}\right)} \Rightarrow \frac{1}{(2\pi)^2}\int d\varepsilon d\varepsilon' \sum_{\mathbf{p}} \frac{\mathcal{A}_{sl}(\mathbf{p},\varepsilon)\mathcal{A}_{s'l'}(\mathbf{p}',\varepsilon)}{\omega + i\delta - (\varepsilon - \varepsilon')}. \tag{D4}$$

The spectral function is given by

$$\mathcal{A}_{sl}(\mathbf{p},\varepsilon) = \frac{2\gamma}{\left(\varepsilon - \varepsilon_{\mathbf{p}}^{sl}\right)^2 + \gamma^2}. \tag{D5}$$

In our calculations, the broadening factor γ equals the imaginary part of the self-energy due to the electron-phonon collisions evaluated at the Fermi surface [36]:

$$\gamma = \mathrm{Im}\Sigma(p_F,\varepsilon_F) = \frac{\varepsilon}{T}\frac{D^2T^2}{4\rho s^2 v_0^2}\bigg|_{\varepsilon=\varepsilon_F}. \tag{D6}$$

The approximation (D4) corresponds to the neglect of vertex corrections in the current-current correlator represented by the bubble diagram. The effect of the vertices is typically to reduce the collision frequency; hence, the approximation (D4) underestimates the resonant conductivity. All the more, careful calculations of the tunnel conductivity show that interference of the carrier scattering events in different layers leads to a further decrease of the effective scattering rate [31]. We leave the determination of the effective scattering rate γ to a future work, and use its upper estimate in the present paper.

[1] A. Grigorenko, M. Polini, and K. Novoselov, Nat. Photonics **6**, 749 (2012).

[2] F. H. L. Koppens, D. E. Chang, and F. J. Garcia de Abajo, Nano Lett. **11**, 3370 (2011).

[3] E. H. Hwang and S. Das Sarma, Phys. Rev. B **75**, 205418 (2007).

[4] V. Ryzhii, A. Satou, and T. Otsuji, J. Appl. Phys. **101**, 024509 (2007).

[5] S. A. Mikhailov and K. Ziegler, Phys. Rev. Lett. **99**, 016803 (2007).

[6] D. Svintsov, V. Vyurkov, S. Yurchenko, T. Otsuji, and V. Ryzhii, J. Appl. Phys. **111**, 083715 (2012).

[7] S. Gangadharaiah, A. M. Farid, and E. G. Mishchenko, Phys. Rev. Lett. **100**, 166802 (2008).

[8] A. Woessner, M. B. 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).

[9] A. Principi, M. Carrega, M. B. Lundeberg, A. Woessner, F. H. L. Koppens, G. Vignale, and M. Polini, Phys. Rev. B **90**, 165408 (2014).

[10] A. Tomadin and M. Polini, Phys. Rev. B **88**, 205426 (2013).

[11] D. Svintsov, V. Vyurkov, V. Ryzhii, and T. Otsuji, Phys. Rev. B **88**, 245444 (2013).

[12] A. A. Dubinov, V. Y. Aleshkin, V. Mitin, T. Otsuji, and V. Ryzhii, J. Phys.: Condens. Matter **23**, 145302 (2011).

[13] F. Rana, IEEE T. Nanotechnol. **7**, 91 (2008).

[14] M. Sabbaghi, H.-W. Lee, T. Stauber, and K. S. Kim, Phys. Rev. B **92**, 195429 (2015).

[15] L. Britnell, R. Gorbachev, A. Geim, L. Ponomarenko, A. Mishchenko, M. Greenaway, T. Fromhold, K. Novoselov, and L. Eaves, Nat. Commun. **4**, 1794 (2013).

[16] A. Mishchenko, J. Tu, Y. Cao, R. Gorbachev, J. Wallbank, M. Greenaway, V. Morozov, S. Morozov, M. Zhu, S. Wong, F. Withers, C. R. Woods, Y.-J. Kim, K. Watanabe, T. Taniguchi, E. E. Vdovin, O. Makarovsky, T. Fromhold, V. Fal'ko, A. Geim, L. Eaves, and K. Novoselov, Nat. Nanotechnol. **9**, 808 (2014).

[17] M. Feiginov and V. Volkov, JETP Lett. **68**, 662 (1998).

[18] V. Ryzhii and M. Shur, Jpn. J. Appl. Phys. **40**, 546 (2001).

[19] B. Sensale-Rodriguez, Appl. Phys. Lett. **103**, 123109 (2013).

[20] V. Ryzhii, A. Satou, T. Otsuji, M. Ryzhii, V. Mitin, and M. S. Shur, J. Phys. D: Appl. Phys. **46**, 315107 (2013).

[21] S. Das Sarma and E. H. Hwang, Phys. Rev. Lett. **81**, 4216 (1998).

[22] E. H. Hwang and S. Das Sarma, Phys. Rev. B **80**, 205405 (2009).

[23] E. H. Hwang and S. Das Sarma, Phys. Rev. B **64**, 165409 (2001).

[24] L. Brey, Phys. Rev. Applied **2**, 014003 (2014).

[25] F. T. Vasko, Phys. Rev. B **87**, 075424 (2013).

[26] V. Ryzhii, A. A. Dubinov, V. Y. Aleshkin, M. Ryzhii, and T. Otsuji, Appl. Phys. Lett. **103**, 163507 (2013).

[27] R. Bistritzer and A. H. MacDonald, Proc. Natl. Acad. Sci. U.S.A. **108**, 12233 (2011).

[28] C. J. Tabert and E. J. Nicol, Phys. Rev. B **86**, 075439 (2012).

[29] M. Greenaway, E. Vdovin, A. Mishchenko, O. Makarovsky, A. Patanè, J. Wallbank, Y. Cao, A. Kretinin, M. Zhu, S. Morozov, V. I. Fal'ko, K. Novoselov, A. Geim, T. Fromhold, and L. Eaves, Nat. Phys. **11**, 1057 (2015).

[30] F. T. Vasko and A. V. Kuznetsov, *Electronic States and Optical Transitions in Semiconductor Heterostructures* (Springer Science & Business Media, 2012).

[31] R. Kazarinov and R. Suris, Sov. Phys. Semicond. **6**, 148 (1972).

[32] In the other limiting case, $\Omega \ll \nu_\varepsilon$, the electron states of individual layers have a well-defined occupancy. The effects of plasmon spectrum modification in this case, however, persist. For the parameters used in the calculation, $\hbar\Omega \approx 10$ meV, which exceeds both electron-phonon and electron-electron scattering rate at nitrogen temperature.

[33] L. A. Falkovsky and A. A. Varlamov, Eur. Phys. J. B **56**, 281 (2007).

[34] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, Science **264**, 553 (1994).

[35] H. Shi, H. Pan, Y.-W. Zhang, and B. I. Yakobson, Phys. Rev. B **87**, 155304 (2013).

[36] F. T. Vasko and V. Ryzhii, Phys. Rev. B **76**, 233404 (2007).

[37] K. I. Bolotin, K. J. Sikes, J. Hone, H. L. Stormer, and P. Kim, Phys. Rev. Lett. **101**, 096802 (2008).

[38] The result can be applied to AA-stacked graphene bilayer, but not to the AB-stacked one. The carrier dispersion for AB stacking does not represent two parallel Dirac cones, and the expression for dynamic conductivity has a different structure [54] .

[39] A. Principi, R. Asgari, and M. Polini, Solid State Commun. **151**, 1627 (2011).

[40] A. Chaplik, Sov. Phys. JETP **35**, 395 (1972).

[41] R. Roldan, M. O. Goerbig, and J.-N. Fuchs, Phys. Rev. B **83**, 205406 (2011).

[42] L. J. Wang, A. Kuzmich, and A. Dogariu, Nature (London) **406**, 277 (2000).

[43] R. Y. Chiao, Phys. Rev. A **48**, R34 (1993).

[44] When relaxation processes are properly taken into account, the two parts of the split dispersion $\mathrm{Re}\omega(q)$ should merge through the resonance. The divergence in the group velocity, defined as $\partial \mathrm{Re}\omega/\partial q$, does not occur in this case. For simplicity of calculation, we have ignored the real part of conductivity when plotting Fig. 6, and sought for the real solutions of the dispersion equation.

[45] Z. Fei, E. G. Iwinski, G. X. Ni, L. M. Zhang, W. Bao, A. S. Rodin, Y. Lee, M. Wagner, M. K. Liu, S. Dai, M. D. Goldflam, M. Thiemens, F. Keilmann, C. N. Lau, A. H. Castro-Neto, M. M. Fogler, and D. N. Basov, Nano Lett. **15**, 4973 (2015).

[46] J. Lambe and S. L. McCarthy, Phys. Rev. Lett. **37**, 923 (1976).

[47] M. Parzefall, P. Bharadwaj, A. Jain, T. Taniguchi, K. Watanabe, and L. Novotny, Nat. Nanotechnol. **10**, 1058 (2015).

[48] D. Yadav, S. Tombet, T. Watanabe, V. Ryzhii, and T. Otsuji, in *73rd Annual Device Research Conference (DRC)* (IEEE, Piscataway, 2015), pp. 271–272.

[49] C. Zhang, M. L. F. Lerch, A. D. Martin, P. E. Simmonds, and L. Eaves, Phys. Rev. Lett. **72**, 3397 (1994).

[50] E. E. Vdovin, A. Mishchenko, M. T. Greenaway, M. J. Zhu, D. Ghazaryan, A. Misra, Y. Cao, S. V. Morozov, O. Makarovsky, T. M. Fromhold, A. Patanè, G. J. Slotman, M. I. Katsnelson, A. K. Geim, K. S. Novoselov, and L. Eaves, Phys. Rev. Lett. **116**, 186603 (2016).

[51] D. Svintsov, V. Vyurkov, V. Ryzhii, and T. Otsuji, J. Appl. Phys. **113**, 053701 (2013).

[52] A. A. Dubinov, V. Y. Aleshkin, V. Ryzhii, M. S. Shur, and T. Otsuji, J. Appl. Phys. **115**, 044511 (2014).

[53] P. L. Bhatnagar, E. P. Gross, and M. Krook, Phys. Rev. **94**, 511 (1954).

[54] E. J. Nicol and J. P. Carbotte, Phys. Rev. B **77**, 155409 (2008).